Information Environment and the Geography of Firms and Investors*

Gennaro Bernile, *Singapore Management University*

Shimon Kogan, *IDC Herzliya and the University of Texas at Austin*

Johan Sulaeman, *National University of Singapore*

February 27, 2015

Abstract – We develop a model linking investor attention, asset ownership, and asset returns in a rational setting, and test the model's prediction that investors' attention to a particular firm depends jointly on their proximity to the firm's locations and on the firm's information environment. The empirical evidence shows that the quality and quantity of publicly available information are strongly related to investors' propensity to hold and trade stocks; however, the direction of these relations depends crucially on the investors' geographic proximity to the firm. The holdings and trading propensity of investors located near the firm decrease with the quality and quantity of public information, but the pattern is reversed for investors located away from the firm. Tests exploiting exogenous shocks to the firm's information environment indicate that these relations are causal. At the firm level, we find that the cross-sectional variations in firms' information environment and proximity to potential investors jointly explain the variation in the geographic concentration of firm ownership and in stock returns.

*We thank Douglas Skinner (the Editor) and an anonymous referee for their valuable guidance. We also benefited from comments and suggestions of Sandro Andrade, Diego Garcia, George Korniotis, Tim Loughran, Evgeny Lyandres, Jacob Sagi, Kumar Venkataraman, participants at Financial Management Association 2014 Meeting, Financial Intermediation Research Society 2014 Meeting, 2012 China International Conference in Finance, 2012 University of Washington Summer Finance Conference, and seminars at Singapore Management University, Southern Methodist University, and University of Miami. We are responsible for all remaining errors and omissions. We are grateful to Ambrus Kecskés for generously providing the data on the mergers and closings of brokerage houses, and Irem Demirci for excellent research assistance with the model. This paper was previously circulated as "Do Exposure and Disclosure Affect Ownership Structure and Stock Returns?"

Please address all correspondence to Gennaro Bernile, Department of Finance, Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Room 4066, Singapore 178899; *Phone:* +65 6808-5478; *email:* `gbernile@smu.edu.sg`. Shimon Kogan can be reached at +972 9960-2741 or `skogan@idc.ac.il`. Johan Sulaeman can be reached at +65 6516-1403 or `sulaeman@nus.edu.sg`.

Information Environment and the Geography of Firms and Investors

Abstract – We develop a model linking investor attention, asset ownership, and asset returns in a rational setting, and test the model's prediction that investors' attention to a particular firm depends jointly on their proximity to the firm's locations and on the firm's information environment. The empirical evidence shows that the quality and quantity of publicly available information are strongly related to investors' propensity to hold and trade stocks; however, the direction of these relations depends crucially on the investors' geographic proximity to the firm. The holdings and trading propensity of investors located near the firm decrease with the quality and quantity of public information, but the pattern is reversed for investors located away from the firm. Tests exploiting exogenous shocks to the firm's information environment indicate that these relations are causal. At the firm level, we find that the cross-sectional variations in firms' information environment and proximity to potential investors jointly explain the variation in the geographic concentration of firm ownership and in stock returns.

1. Introduction

A growing literature documents the tendency of investors to invest "close to home" – both internationally and within the U.S. The U.S.-based evidence (e.g., Coval and Moskowitz 1999, 2001, Huberman, 2001, Ivković and Weisbenner, 2005, Baik, Kang, and Kim, 2010, Seasholes and Zhu, 2010, and Bernile, Kumar, and Sulaeman, 2015) is particularly interesting. Given the absence of institutional barriers to investing across U.S. state borders, why would investors exhibit local bias? If investors have to focus on a limited set of stocks due to limited cognitive ability, they may rely on *familiarity* as a screening device and therefore overweight more familiar local stocks. Alternatively, investors may overweight local stocks if geographic proximity provides them with an *informational advantage*.

Both of these explanations suggest that we may observe portfolio holdings that differ across investors based on their respective location. Both explanations also have similar asset pricing implications as they predict that stock returns would be higher for firms with large ownership concentration of local investors. However, the economic forces that give rise to these implications across the two scenarios are quite distinct. Under the limited attention explanation, higher returns compensate shareholders for the lack of risk-sharing as in Merton (1987). In contrast, under the asymmetric information explanation, higher returns compensate uninformed investors for adverse selection risk of trading with potentially informed investors. This distinction has important implications for the policies that govern and regulate how firms disseminate information to the public.

In this paper we examine the links between a firm's public information environment, the geographic proximity between the firm and its potential investor base, the firm ownership structure, and its stock returns. In particular, we exploit features of the firm's information environment – the quantity and quality of public information – to separate out the two explanations above. While firm's proximity to investors may proxy for both attention and access to private information, the availability and precision of public information about the firm affect only the value of private information. As such, we expect that if private information drives the preference for local stocks, then the improvement in information environment will result in weaker local bias, a more dispersed ownership structure, and a weaker link between the firm's proximity to potential investors and stock returns. These predictions are consistent with a fully-rational model in which investors have symmetric access to public information, but asymmetric access to private information as a result of their relative proximity to the firm. To test these predictions, we employ novel proxies for the independent variables of interest – firms' proximity to investors and public information environment.

We use two complementary measures of geographic proximity that reflect the multiple dimensions of firm locations. Our first measure addresses a shortfall in the way proximity is measured in the extant literature, which uses the distance between a firm and its *actual shareholders*, e.g., the local ownership measure employed in Coval and Moskowitz (2001) and Baik, Kang, and Kim (2010). This measure is a *result* of – rather than an *input* into – investors' decision problem and therefore cannot be viewed as exogenous in our context. Instead, we measure proximity as the distance between the firm's headquarters location and its *potential investor base*.[1] Our second measure is intended to capture the proximity of potential investors to locations where the firm has economic interests beyond its headquarters. Specifically, we use the firm's geographic dispersion as defined in Garcia and Norli (2012). The choice of this measure is motivated by the evidence that institutional investors overweight and display an informational advantage for firms with local non-HQ operations (Bernile, Kumar, and Sulaeman (2015)). Consequently, more geographically dispersed firms are likely to be "local" to a larger number of potential investors.

To capture features of the firm's information environment, we use two distinct measures of public information quality and quantity. First, we employ a novel measure of disclosure quality intended to reflect the informativeness of firms' annual reports. Financial statements are an important vehicle for public dissemination of information but measuring their quality is difficult. Related to a growing number of studies (see Li (2011) and Miller and Skinner (2015) for surveys of the literature), we focus on the text in the Management Discussion and Analysis section (MD&A) of annual reports, where managers discuss past performance and future directions and challenges. In particular, we follow the empirical method of Kogan, Routledge, Sagi, and Smith (2011) to measure MD&A's informativeness by mapping its text into subsequent return volatility and proxying for disclosure quality by the degree to which the MD&A helps predict *out-of-sample* volatility. Our second measure is based on extensive research indicating that information produced by equity research analysts is value-relevant (e.g., Stickel (1991), Barber et al. (2001), Jegadeesh et al. (2004), and Loh and Stulz (2011)) and affects investors'

[1] Davis and Henderson (2008) provides a thorough discussion of the role of corporate headquarters in the information exchange among firms. Malloy (2005) and Pirinsky and Wang (2006) examine the effects of corporate HQ location on analyst coverage and stock returns, respectively. See Aarland, Davis, Henderson, and Ono (2007) and Liberti and Mian (2009) for more discussion on the role of corporate headquarters in information aggregation.

portfolio decisions (e.g., Womack (1996), Juergens and Lindsey (2009)). Specifically, we use analyst coverage as a measure of availability of public information about the firm.

Although the two measures of information environment are desirable because of their availability for a wide cross-section of firms, they are subject to potential concerns regarding reverse causality or omitted variable bias. For example, firms with more complicated operations may provide higher quality disclosures, firms with more concentrated ownership may provide higher quality disclosures or be followed by more analysts, analysts may choose to follow more complicated firms, etc. To address these potential concerns and shed more light on the causal effects of (changes in) information environment, we identify two sets of shocks to firms' information environment unlikely to be driven by firm-specific conditions or managerial decisions: (1) the changes in regulation that affected the informativeness of MD&A sections in 2002–2003, as in Kogan, Routledge, Sagi, and Smith (2011), and (2) the loss of analyst coverage due to brokerage houses' closures and mergers, as in Hong and Kacperczyk (2010), Kelly and Ljungqvist (2012), and Derrien and Kecskés (2013). The regulation changes affect the quality and reliability of information disclosure across *all* firms, while brokerage house disappearances reduce analyst coverage for *some* firms. Both of these "shocks" are exogenous in that they are not driven by individual firms' decisions. Moreover, the exogenous loss of analyst coverage provides an additional layer of variation – i.e., random assignment across firms – that allows for differences-in-differences tests between affected firms and a matched control sample of non-affected firms.

Our evidence supports the adverse selection-based predictions regarding the joint effects of firm's proximity to investors and information environment on stock ownership and trading activity. First, consistent with earlier studies, we document excess ownership concentration around firms' headquarters as well as other locations important for the firm. Second, the geographic dispersion of a firm's ownership structure increases with both the firm's overall geographic dispersion and the quantity and quality public information. Third, the marginal impact of one factor on ownership dispersion decreases as the other factor increases. Fourth, the intensity of investors' trading (i.e., the volatility of investors' holdings) increases with investors' proximity to the firm's locations, and this effect decreases significantly as the firm's geographic dispersion increases or its information environment improves. Finally, consistent with the ownership results, we find that the effects of firm's geographic dispersion and information environment on trading temper each other. Importantly, the empirical evidence we obtain from exploiting exogenous changes in firms' information environment suggests these relations are causal.

Our empirical evidence is also consistent with the predictions regarding the joint effects of firm's proximity to investors and information environment on stock returns. In particular, we find that the concentration of potential investors around the firm's HQ has an economically large and statistically significant effect on asset returns. Firms whose headquarters are located far from the pool of potential investors have higher stock returns than those located near potential investors by a considerable margin: the monthly abnormal return differential ranges between $0.27 - 0.37$ percent for equal-weighted portfolios and $0.37 - 0.68$ percent for value-weighted portfolios. We also find that the link between potential investors' proximity to the firm's HQ and stock returns is weakened by the firm's geographic dispersion and information environment. These empirical results are robust to the inclusion of controls for firm characteristics such as past returns, firm size, market-to-book ratio, liquidity, and volatility.

It is worth noting that the relation we document between the concentration of potential investors around firms' HQ and stock returns is distinct from the effect of our second proxy of investor proximity, firm's geographic dispersion, on returns previously documented in Garcia and Norli (2012). In fact, while we confirm the finding in Garcia and Norli (2012) that firm's geographic dispersion has a negative relation with stock returns, our results suggest that this link is largely driven by the interaction between a firm's HQ-remoteness and the geographic dispersion of its economics interests. That is, firms headquartered in more remote locations also tend to be smaller, more geographically concentrated, and at the same time located far away from potential investors.

Our paper makes significant contributions to three strands of the literature. First, it contributes to the literature on the effect of firms' information environment on ownership structure and cost of capital.[2] Prior studies link investors' propensity to invest abroad to the perceived quality of foreign firms' information environment, including their accounting choices (e.g, Gillan and Starks, 2003; Bradshaw, Bushee, and Miller, 2004). The most closely related paper to ours in this literature is Corvig, DeFond, and Hung (2007), which documents a direct link between ownership by foreign mutual funds and firms' voluntary adoptions of International Accounting Standards (IAS).

Our paper extends this literature in several directions. We examine *within*-country variations of firm ownership structures. This allows us to sidestep the issue of *cross*-country variations in legal systems

[2]Theoretical works in this literature have examined the effects of disclosure on cost of capital both in single-asset (see Verrecchia, 2001 for a survey) and multiple-asset (e.g., Lambert, Leuz, and Verrecchia, 2007) settings. Both approaches suggest that disclosure has a direct effect of reducing cost of capital since it reduces return variance (or increases precision regarding the asset value, in the single-asset framework) or covariance with other assets (in the multiple-asset framework).

and other institutional details, including languages, conventions, and accounting rules.[3] Moreover, our setting allows us to alleviate potential concerns about reverse causality and omitted variables by providing us with plausible exogenous shocks that affect either *all* firms – i.e., the regulatory changes in 2002–2003 – or *some* firms but not due to the firms' own choice – i.e., the loss of analyst coverage due to changes in the brokerage industry. Lastly, the institutional investors in our sample account for a large fraction of firms' ownership – on average close to 50 percent throughout our sample period and even higher in the later period. These features make our experimental setting well-suited to test various theories that highlight the role of access to private and public information in determining asset prices and cost of capital, such as those developed by Diamond and Verrecchia (1991) and Baiman and Verrecchia (1996). Our analysis of the links between information environment, firm locations, ownership structure, and stock returns complements earlier empirical tests of those models, such as Leuz and Verrecchia (2000) and Dai, Parwada, and Zhang (2015).

Second, we contribute to the literature on attention, asymmetric information, and asset prices by formulating a hypothesis linking potential determinants of investor attention and asset prices. In testing theories that link attention and asset prices – such as the Merton (1987) model, there are two significant and related issues: (1) there is no direct measure of investor attention, and (2) investor attention is not exogenously determined. Most indirect proxies for investor attention are developed from potential determinants of attention: extreme returns (Barber and Odean, 2008), news and headlines (Barber and Odean, 2008; Yuan, 2011; Fang and Peress, 2009), advertising expense (Chemmanur and Yan, 2009; Grullon, Kanatas, and Weston, 2004; and Lou, 2009), price limit (Seasholes and Wu, 2007), trading volume (Barber and Odean, 2008; Gervais, Kaniel, and Mingelgrin, 2001; Hou, Peng, and Xiong, 2009). Another strand of literature uses proxies of revealed attention based on online search frequency (Da, Engelberg, and Gao, 2011; and Brown, Stice, and White, 2015). However, as the proxies developed by these studies tend to be short-lived, it is difficult to examine whether the effect of attention on stock prices is due to rational or non-rational behavior. Moreover, these studies have largely ignored how features of firms' information environment affect investors' attention allocation decisions.

In contrast, our paper focuses on the potentially long-term determinants of investors' access to pri-

[3]Indeed, previous evidence suggests that information environment may have significant within-country effects on investors' attention. For example, using AIMR ranking as a proxy of disclosure quality, Bushee and Noe (2000) find that firms with higher disclosure quality have greater institutional ownership and experience more trading by "transient" investors.

vate information that are unlikely to be driven by (the desire for) investor attention (van Nieuwerburgh and Veldkamp, 2009): firm HQ location and geographic dispersion. In particular, firm locations are likely to be driven by operational needs and opportunities, rather than the desire to attract investors' attention. As such, these characteristics are likely to be exogenous factors vis-à-vis investor attention, which allow us to test whether these factors affect both investor attention and its links with stock prices and returns. In this sense, our paper complements recent analyses on the effect of investor attention on stock returns, such as Garcia and Norli (2012). Importantly, we expand this literature by examining explicitly how the availability and quality of public information affect the link between investors' access to private information and the firm ownership structure and stock returns.

Third, we contribute to the empirical literature on the effect of firm locations on asset prices and returns. Hong, Kubik, and Stein (2008) document the relation between asset prices (but not returns) and firm headquarter location as a function of the demand and supply of local assets.[4] Pirinsky and Wang (2006) document the comovement of stock returns of firms sharing the same headquarters location. We complement these studies by documenting the link between the cross-sectional variations in stock *returns* and the variations in corporate locations – headquarters and otherwise.

In the next section, we summarize the intuition and hypotheses at the core of our empirical tests, while Appendix A describes the model that we use to formalize the intuition and derive the predictions. Section 3 describes the data. Sections 4 and 5 examine the effects of information environment and firm geography on ownership structure and investors' portfolio decisions, respectively. Section 6 examines the joint effects of firms' information environment and proximity to potential investors on stock returns. Section 7 concludes.

2. Hypothesis Development

To derive testable hypotheses, we use a formal model that links firms' information environment to investors' portfolio decisions and the resulting asset prices. In particular, we develop a fully-rational model, following Easley and O'Hara (2004), in which investors have symmetric access to public information about firms' values but asymmetric access to private information. It assumes that agents are endowed with varying precision of private information about the firm depending on their proximity to

[4] In their model, the demand and supply are static. As such, while prices are affected by demand and supply, the "changes" in prices are not.

the firm, while firms differ by the amount and quality of public information available. For the sake of brevity, we present the formal model and derivations in Appendix A. In this section, we discuss three sets of predictions derived from the model that we test in our empirical analysis.

The first set of hypotheses pertains to the geography of firms' ownership.

1a. Investors located near firm locations hold relatively larger stakes of the firm than investors located away from the firm.

Intuitively, because local investors have access to private information, the firm's stock would seem less risky to these (risk-averse) investors than (similarly risk-averse) nonlocal investors. Consequently, the nearby investors would hold relatively larger stakes.

The second prediction related to the geography of firms' ownership is that:

1b. The aggregate *excess* holdings of nearby investors decrease as the firm's geographic dispersion increases.

An increase in the firm's geographic dispersion is accompanied by an increase in the size of the firm's privately informed investor base, which has two implications. On the one hand, a firm's greater geographic dispersion increases the aggregate demand from informed investors as more investors compete to take advantage of their private information about the firm. On the other hand, increased competition among privately informed investors results in more informative stock prices, which lowers the risk for the uninformed investors and consequently increases their demand for the firm stock. Our model formally shows that the latter effect dominates the former. Therefore, in expectation, an increase in geographic dispersion of the firm would be associated with an increase in the holdings of uninformed investors vis-à-vis informed investors.

The third prediction related to the geography of firms' ownership is that:

1c. The aggregate *excess* holdings of nearby investors decrease with the availability and precision of public information about the firm.

An increase in the quantity and quality of information available to all investors tilts the playing field in the direction of the (privately uninformed) nonlocal investors. In turn, this increases their demand relative to the informed local investors' demand.

The last prediction related to the geography of firms' ownership is that:

1d. The marginal effect of geographic dispersion (public information) on the excess holdings of nearby investors decreases with an improvement in public information environment (an increase in geographic dispersion).

For firms with better public information environment, the effect of an increase in geographic dispersion is relatively weaker because the potential profit from access to private information is smaller and therefore the resulting competition is not as fierce. Similarly, for a firm with larger base of (privately) informed investors, the effect of an improvement in information environment is relatively smaller because competition among (privately) informed investors already results in more informative stock prices. Therefore, firms' geographic dispersion and public information are substitutes at the margin.

The second set of hypotheses pertains to investors' trading intensity. To the best of our knowledge, we are the first to derive (and test) these predictions.

2a. Investors located near firm locations trade relatively more than investors located away from the firm.

2b. The excess *trading* of nearby investors decreases with the firm's geographic dispersion and the improvement of the public information environment.

2c. The marginal effect of geographic dispersion (public information) on the excess trading of nearby investors decreases with an improvement in information environment (an increase in geographic dispersion).

The effects on nearby investors' trading activity are similar to the effects on their holdings: they will trade less aggressively when they face stiffer competition among themselves (i.e., higher geographic dispersion) and when the public information environment is less favorable. These effects are substitutes at the margin.

The last set of hypotheses pertains to expected returns and is directly related to the predictions on ownership and trading:

3a. Firms with a larger base of privately informed investors have lower expected returns.

3b. Firms with greater availability and quality of public information have lower expected returns.

3c. The marginal effect of geographic dispersion (public information) on expected returns decreases with an improvement in information environment (an increase in geographic dispersion).

As the size of the firm's privately informed investor base increases, its stock price becomes more informative and the aggregate demand increases, which in turn lower the expected stock returns or cost of capital. An improvement in the information environment has a similar effect, and the two effects are substitutes at the margin.

3. Data Sources and Variables Description

Our sample contains all firms in the CRSP/Compustat merged database that have available 10-K filings in Electronic Data Gathering, Analysis, and Retrieval system (EDGAR), with fiscal years ending between 1998 and 2006. There are ultimately 25,207 firm-year observations in our sample for which we can obtain all necessary variables described below and used in the analysis.

Our first main data source is the quarterly common stock holdings of 13(f) institutions compiled by Thomson Reuters. We identify the institutional location (zip code) using the *Nelson's Directory of Investment Managers* and by searching the SEC documents and web sites of institutional managers. We use the institution location data to measure the geographic dispersion of ownership as well as the state investor holdings levels and changes for each firm in our sample.

In particular, in our firm-level analysis, *Actual Shareholder Distance* is defined as the share-weighted average distance of the firm's shareholder base from the firm's headquarters location. This variable decreases with the excess holdings of shareholders located closer to the firm headquarters. In our firm-state investor-level analysis, we define $Firm - State - Qtr\ IO\ Share$ as the firm's institutional ownership by investors in each state, and $|\Delta Firm - State - Qtr\ IO\ Share|$ as the absolute value of the change in $Firm - State - Qtr\ IO\ Share$.

3.1. Proxy for Proximity to HQ Location: Potential Investor Distance

Using the 13(f) dataset, we construct our first proxy of geographic proximity: *Potential Investor Distance*, defined as the investor portfolio value-weighted average distance from the firm's headquarter state. The sample of potential investors includes all institutional investors located in the U.S., independent of whether they own any shares in the firm. A higher value of Potential Investor Distance implies lower proximity of potential institutional investors to the firm headquarters.

Our focus on the investor proximity to corporate headquarters locations is motivated by the burgeoning literature on the role of corporate headquarters in information aggregation within firms (e.g., Aarland, Davis, Henderson, and Ono, 2007; Liberti and Mian, 2009), information exchange among firms (e.g., Davis and Henderson, 2008), and between firms and capital market participants (e.g., Coval and Moskowitz 1999, 2001; Malloy, 2005; Pirinsky and Wang, 2006).

3.2. Proxy for Proximity to Non-HQ Locations: Geographic Dispersion of Firm's Economic Interests

To construct two of the main variables in our analysis, we rely on the information contained in the mandated annual financial reports (Form 10-K) filed by firms with the SEC and stored on EDGAR. Form 10-K is a standardized form with a predetermined number of sections and structure, containing a comprehensive summary of a companys performance and operations. In addition to financial data, the form typically includes information on the evolution of the firms operations during that year, details on its organizational structure, executive compensation, competition, and regulatory issues. Annual 10-K filings also require information on each firm's properties, such as factories, warehouses, and sales offices. We use a computer-based parsing program to create our second measure of geographic proximity as well as the informativeness of the firm's annual report (discussed in the next subsection).

We use a firm-level measure of geographic dispersion of economic interests as our second proxy of proximity. Our focus on this measure is motivated by the recent literature on the importance of economic locations away from firm headquarters (Giroud, 2011; Garcia and Norli, 2012; Bernile, Kumar, and Sulaeman, 2015). To construct the geographic dispersion measure, we count the number of states (and Washington D.C.) cited in the firm's annual 10-K filing. For each annual filing, we count the occurrence of geographic locations in sections "Item 1: Business", "Item 2: Properties", "Item 6: Consolidated Financial Data", and "Item 7: Managements Discussion and Analysis." These citation counts provide a measure of the geographic locations of the firms economic interests, including the firms plant and store locations, office spaces, acquisition activities, pending litigations, and other means through which the firm can have an economic presence in a state.

3.3. Information Environment Measure: MD&A Informativeness

We use two proxies to capture the firms' public information environment. First, we use the disclosure quality measure developed by Kogan, Routledge, Sagi, and Smith (2011), which is based on the MD&A informativeness. This measure quantifies the usefulness of the textual content in each firm's annual 10-K filing's "Item 7: Managements Discussion and Analysis" (MD&A) section for predicting the firm's future volatility. Specifically, we estimate a statistical model that determines how various terms in the reports contribute to increases or decreases in firm volatility. We then take these term weights and apply them *out of sample* to forecast annual volatility.

The MD&A informativeness is measured by subtracting the absolute value of the estimation error of the text-based model from the absolute value of the estimation error of a statistical model that uses the firm's past year volatility as the next year's forecast. Thus, the MD&A informativeness measure increases when the text in the MD&A section is more useful for – i.e., has lower estimation error in – predicting firm volatility.

The data used to construct this measure are from the MD&A sections available in annual reports filed during the 1996–2006 period. The firm-year observations for which this procedure results in a meaningful size report are then matched with CRSP return data. On a rolling basis, a computer algorithm uses the past two years' reports and realized volatility to estimate word weights. Then, it applies the weights to the current report to forecast the firm's next year volatility. This procedure produced an MD&A informativeness measure for firm-years in the 1998–2006 period.[5] Appendix B provides a detailed discussion of how this measure compares to other predictive measures of future volatility. In sum, the MD&A-based measure performs better in predicting future volatility, and it does so by capturing firm dimensions that cannot be easily captured using other approaches.

3.4. Information Environment Measure: Analyst Coverage

Our second measure of the public information environment is based on analyst coverage. In particular, we count the number of unique equity analysts providing stock recommendations for the firm during the 12-month period ending on the fiscal year-end. We employ this measure for two reasons. First, equity analysts are an important source of public information about the firms that they cover as

[5]Because the algorithm requires that the MD&A section contains at least 1,000 words, we assign the minimum informativeness measure to missing observations that do not satisfy this criterion, i.e., -0.6.

attested by the evidence that they produce value-relevant information (e.g., Stickel (1991), Barber et al. (2001), Jegadeesh et al. (2004), and Loh and Stulz (2011)) and affect investors' portfolio decisions (e.g., Womack (1996), Juergens and Lindsey (2009)). Second, using analyst coverage provides us with a unique opportunity to exploit exogenous shocks to the firm information environment associated with merger and closure activities in the brokerage industry (Hong and Kacperczyk (2010); Kelly and Ljungqvist (2012); Derrien and Kecsks (2013)). This allows us to potentially identify the causal effect of firms' information environment on investors' portfolio allocation decisions as well as its causal effect on the relation between investor-firm proximity and investor portfolio decisions.

3.5. Other Data Sources

In addition to the main data sources described above, we use several other standard data sets. We obtain price, volume, return, and industry membership data from the Center for Research on Security Prices (CRSP). The firm headquarters location data are from the CRSP-Compustat merged (CCM) file, corrected for HQ moves. We obtain the performance benchmarks for computing characteristic-adjusted stock returns from Russell Wermers' web site.[6] We use the Central Index Key (CIK) to match SEC's EDGAR data with CRSP and Compustat data on stock returns and various firm characteristics.

We also use several state-level time-varying variables to absorb the variation in state characteristics. We use state-level Presidential elections data to identify the state political orientation.[7] We obtain additional state-level demographic characteristics from the U.S. Census Bureau. Specifically, we consider state population density and education level (the proportion of the population above age 25 with a bachelor's degree or higher) in our regressions. Further, using data from the "Churches and Church Membership" files available through the American Religion Data Archive (ARDA), we compute the proportion of Catholics (CATH) and Protestants in the state (PROT), and the resulting Catholic-Protestant ratio (CPRATIO). We also measure the overall religiosity of the state. Lastly, we use quarterly data on state economic activity from Korniotis and Kumar (2010). The economic activity index is defined as the equal weighted average of the standardized values of state income growth, state housing collateral (Lustig and van Nieuwerburgh, 2005 and 2010), and the negative value of standardized relative state unemployment. Summary statistics of the main measures used in the paper are

[6] The web site is `http://www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm`.

[7] The election data are obtained from David Leip's web site: `www.uselectionatlas.org`.

reported in Table 1.

4. Firm-level Geographic Dispersion of Ownership

In this section, we discuss the evidence from our tests of the hypotheses pertaining to the geography of firms' ownership summarized below:

1a. Investors near firm locations hold relatively larger stakes of the firm than investors located away from the firm.

1b. The aggregate *excess* holdings of nearby investors decrease as the firm's geographic dispersion increases.

1c. The aggregate *excess* holdings of nearby investors decrease with the availability and precision of public information about the firm.

1d. The marginal effect of geographic dispersion (public information) on the excess holdings of nearby investors decreases with an improvement in public information environment (an increase in geographic dispersion).

The first prediction has been tested extensively in the literature using geographic proximity to firm headquarters. For example, Coval and Moskowitz (1999) and Ivković and Weisbenner (2005) document that U.S. mutual fund managers and U.S. retail investors tend to hold disproportionately more local stocks in their portfolios. Thus, existing evidence implies that shareholders tend to be located closer to the firm headquarters than non-shareholders. In our context, this hypothesis translates to a prediction that *Actual Shareholder Distance*, the share-weighted average distance of the firm's shareholders from the firm's headquarters location, is less than *Potential Investor Distance*.

The second prediction implies that *Actual Shareholder Distance* is an increasing function of *Geographic Dispersion*, controlling for potential shareholders' distance. A firm with more geographically diversified economic interests is exposed to more investors near those locations (see Bernile, Kumar, and Sulaeman, 2015). Therefore, its ownership should be more dispersed and the average distance of the firm's shareholders from the firm's headquarters location should increase. The third hypothesis implies that *Actual Shareholder Distance* increases with *MD&A Quality* and *Analyst Coverage*, our proxies for the firm's public information environment, controlling for potential shareholder distance.

In Table 2, we report the results from univariate sorting analyses. The first two rows in Panel A report the mean *Potential Investor Distance* and *Actual Shareholder Distance*, and the last row their mean difference. We find that Actual Shareholder Distance is smaller and significantly different from the potential distance, consistent with institutional excess ownership increasing with investor proximity to the firm's headquarters. This evidence is consistent with prior literature and our first hypothesis.

In Panel B, we use the difference in the last row of Panel A as the basic construct to examine the effect of firm geographic dispersion on ownership dispersion. In particular, the model predicts that the difference between Actual Shareholder and Potential Investor Distance increases with firm's Geographic Dispersion. The univariate evidence is consistent with this prediction.

To test the full set of predictions on the geography of firm ownership, we estimate the following regression model:

$$\begin{aligned}(Log)Actual\ Shareholder\ Distance_{i,t} &= a + b \cdot (Log)Potential\ Investor\ Distance_{i,t} \\ &+ c \cdot Geo.\ Disp._{i,t} + d \cdot MD\&A\ Quality_{i,t} \\ &+ e \cdot Geo.Disp._{i,t} \cdot MD\&A\ Quality_{i,t} \\ &+ F \cdot Controls_{i,t} + \epsilon_{i,t},\end{aligned}$$

where $Geo.\ Disp._{i,t}$ is the firm geographic dispersion tercile (0 for low, 1 for medium, and 2 for high) and $MD\&A\ Quality_{i,t}$ is the firm MD&A informativeness tercile. This allows us to account for the joint effect of firm geographic dispersion, information environment, and their interaction with shareholder geographic dispersion, while also controlling for other factors that may affect the geographic dispersion of ownership.

We report the log-distance version in Model (1) of Table 3 and the raw distance version in Model (2). The coefficient estimate for Potential Investor Distance is significantly less than 1, consistent with our first hypothesis. Moreover, the estimates for Geographic Dispersion and MD&A Informativeness are significantly positive, consistent with our second and third hypotheses. All else equal, shareholders' distance from the firm headquarters tends to increase with firm's geographic dispersion and MD&A informativeness. Lastly, the interaction term is negative and significant, suggesting that the effects of disclosure quality and geographic dispersion are substitutes, consistent with prediction (1d).

In the last model of Table 3, we use an alternative measure of ownership dispersion: the proximity-weighted Herfindahl index of ownership concentration. The expected signs of the parameters in this analysis are the opposite of those in the other models since the ownership concentration index decreases as the firm's shareholders are more dispersed and located further away from the headquarters. The empirical results based on the alternative measure of ownership dispersion support predictions 1a-1d from the model, in line with the other evidence in the same table.

Table 4 repeats the analysis after we replace *MD&A Quality* with *Analyst Coverage*:

$$\begin{aligned}(Log)Actual\ Shareholder\ Distance_{i,t} &= a + b \cdot (Log)Potential\ Investor\ Distance_{i,t} \\ &+ c \cdot Geo.\ Disp._{i,t} + d \cdot Analyst\ Coverage_{i,t} \\ &+ e \cdot Geo.Disp._{i,t} \cdot Analyst\ Coverage_{i,t} \\ &+ F \cdot Controls_{i,t} + \epsilon_{i,t},\end{aligned}$$

The inferences from this table are identical to those obtained in Table 3. In particular, the estimate for *Analyst Coverage* is positive, consistent with prediction (1c): shareholders' distance from the firm headquarters tends to increase with the number of equity analysts following the firm. Moreover, the estimate for the interaction term is also negative, suggesting that the effects of analyst coverage and geographic dispersion are substitutes, consistent with prediction (1d).[8]

5. State-level Institutional Ownership Levels and Changes

In this section we change the unit of observation to the firm-state investor level in contrast to the tests reported in Tables 3 and 4, which are performed at the firm level. Aggregating institutional ownership at the firm level is desirable to alleviate concerns about within-firm clustering of institutional ownership. However, this level of aggregation does not allow us to examine the levels and changes of ownership at the investor location level, which is more appropriate given our empirical predictions. Therefore,

[8] In untabulated analysis, we perform similar tests conditional on institutional investor types using the classification from Bushee (1998): Dedicated, Quasi-Index, or Transient. The evidence indicates that our aggregate results are driven by dedicated and quasi-index investors. Conversely, the degree of local bias of transient investors and its relation with information environment and geographic dispersion are statistically and economically weaker. These results may not be surprising to the extent that transient investors obtain information on as many firms as possible to facilitate their investment strategies.

in this section, we focus instead on institutional ownership aggregated at the state level. Specifically, we examine whether the quarterly levels and changes of firm-state ownership are related to firm-state investor proximity, the firm's overall geographic dispersion, and its public information environment.

5.1. Firm-State Panel Analysis

We begin with the analysis of state investor ownership levels to assess whether our firm-level ownership results are robust to the change in the unit of observation. In particular, we estimate tobit specifications of the following multivariate regression models, with constant upper and lower bounds equal to 1 and 0 respectively:

$$\begin{aligned}
\text{Firm-State-Qtr IO Share}_{i,s,t} &= a + b \cdot \mathit{Exposure}_{i,s,t} + \ c \cdot \mathit{MD\&A\ Quality}_{i,t} + \ d \cdot \mathit{Geo.\ Disp.}_{i,t} \\
&\quad + e \cdot \mathit{Exposure}_{i,s,t} \cdot \mathit{MD\&A\ Quality}_{i,t} \\
&\quad + f \cdot \mathit{Exposure}_{i,s,t} \cdot \mathit{Geo.\ Disp.}_{i,t} \\
&\quad + g \cdot \mathit{Exposure}_{i,s,t} \cdot \mathit{MD\&A\ Quality}_{i,t} \cdot \mathit{Geo.\ Disp.}_{i,t} \\
&\quad + Z \cdot \mathit{Controls}_{i,t} + \epsilon_{i,t},
\end{aligned}$$

and

$$\begin{aligned}
\text{Firm-State-Qtr IO Share}_{i,s,t} &= a + b \cdot \mathit{Exposure}_{i,s,t} + \ c \cdot \mathit{Analyst\ Coverage}_{i,t} + \ d \cdot \mathit{Geo.\ Disp.}_{i,t} \\
&\quad + e \cdot \mathit{Exposure}_{i,s,t} \cdot \mathit{Analyst\ Coverage}_{i,t} \\
&\quad + f \cdot \mathit{Exposure}_{i,s,t} \cdot \mathit{Geo.\ Disp.}_{i,t} \\
&\quad + g \cdot \mathit{Exposure}_{i,s,t} \cdot \mathit{Analyst\ Coverage}_{i,t} \cdot \mathit{Geo.\ Disp.}_{i,t} \\
&\quad + Z \cdot \mathit{Controls}_{i,t} + \epsilon_{i,t},
\end{aligned}$$

where *Firm-State-Qtr IO Share*$_{i,s,t}$ is firm i's institutional ownership by institutional investors in state s, *Geo. Disp.*$_{i,t}$ is the firm geographic dispersion tercile (0 for low, 1 for medium, and 2 for high), $\mathit{MD\&A\ Quality}_{i,t}$ is the firm's MD&A informativeness tercile, and *Analyst Coverage*$_{i,t}$ is the firm's analyst coverage tercile. We include state fixed effects as control variables to take into account the geographic variation in the presence of institutional ownership (e.g., large fractions of institutional investors are located in New York and Boston).

We use three distinct variables to capture the proximity of state s' investors to firm i, $Exposure_{i,s,t}$. The first variable is a *HQ State Indicator*, which takes the value of 1 if the firm's HQ is located in state s, and 0 otherwise. The second exposure measure is a distance based variable: $1/1 + (Log(HQ\ Distance_{i,s}))$. $HQDistance_{i,s}$ is the distance between state s and firm i's HQ state, in 100 miles. Same-state distances are set equal to 100 miles (this seemed reasonable since no cross-state population-weighted centroids distance is less than 100 miles in the sample). Hence, the second exposure variable takes the value of 1 if the firm's HQ is located in the state, and is between 0 and 1 otherwise. The last exposure variable is a *Contact State Indicator*, which takes the value of 1 if firm i mentions state s in its 10-K filing, and 0 otherwise.

Tables 5 and 6 report the results based on the MD&A- and analyst-based measures of public information environment, respectively. Consistent with the evidence in Table 3, the results in columns 1 and 3 of Tables 5 and 6 indicate that institutional holdings increase with proximity to the firm HQ. The evidence is similar when investors' proximity is measured using locations mentioned in the firm's 10-K, in column 5 of each table. This supports the logic for using the firm overall geographic dispersion as a proxy for average investor proximity in our earlier firm level tests.

The remaining results in Tables 5 and 6 add important texture to our baseline firm-level evidence and lend direct support to our empirical predictions. In particular, $Geo.\ Disp._{i,t}$ and $MD\&A\ Quality_{i,t}$ (or *Analyst Coverage*) are positively associated with state investor ownership levels when the state investors are not directly exposed to the firm. This is consistent with the notion that better public information environment or more widely diffused private information increases (privately) uninformed investors' incentives to hold the stock. Moreover, the estimate for their interaction, $GEO \cdot MDA$, indicates that the two effects are substitutes at the margin, consistent with an adverse selection-based theory.

In stark contrast, and consistent with our predictions, all of the aforementioned effects go in the opposite direction for state investors that are more likely to have access to private information, i.e., those located in close proximity to firm locations. Nearby investors' excess ownership decreases with the level of firm overall exposure (geographic dispersion) and improvements in the firm public information environment (MD&A informativeness and Analyst Coverage), and again the two effects are substitutes at the margin.

We now turn our attention to the evidence from our tests of the hypotheses pertaining to investors' trading intensity conditional on their proximity to the firm summarized below:

2a. Investors near firm locations trade relatively more than investors located away from the firm.

2b. The excess *trading* of nearby investors decreases with the firm's geographic dispersion and the improvement of the public information environment.

2c. The marginal effect of geographic dispersion (public information) on the excess trading of nearby investors decreases with an improvement in information environment (an increase in geographic dispersion).

In particular, to test predictions 2a–2c, we estimate tobit specifications of the following multiple regression model, with lower bounds always equal to 0 and firm-state-quarter specific upper bound varying with previous quarter firm-state holdings level:[9]

$$\begin{aligned}|\Delta \textit{Firm-State-Qtr IO Share}_{i,s,t}| \;=\; & a + b \cdot \textit{Exposure}_{i,s,t} + \; c \cdot \textit{MD\&A Quality}_{i,t} + \; d \cdot \textit{Geo. Disp.}_{i,t} \\ & + e \cdot \textit{Exposure}_{i,s,t} \cdot \textit{MD\&A Quality}_{i,t} \\ & + f \cdot \textit{Exposure}_{i,s,t} \cdot \textit{Geo. Disp.}_{i,t} \\ & + g \cdot \textit{Exposure}_{i,s,t} \cdot \textit{MD\&A Quality}_{i,t} \cdot \textit{Geo. Disp.}_{i,t} \\ & + Z \cdot \textit{Controls}_{i,t} + \epsilon_{i,t},\end{aligned}$$

and

$$\begin{aligned}|\Delta \textit{Firm-State-Qtr IO Share}_{i,s,t}| \;=\; & a + b \cdot \textit{Exposure}_{i,s,t} + \; c \cdot \textit{Analyst Coverage}_{i,t} + \; d \cdot \textit{Geo. Disp.}_{i,t} \\ & + e \cdot \textit{Exposure}_{i,s,t} \cdot \textit{Analyst Coverage}_{i,t} \\ & + f \cdot \textit{Exposure}_{i,s,t} \cdot \textit{Geo. Disp.}_{i,t} \\ & + g \cdot \textit{Exposure}_{i,s,t} \cdot \textit{Analyst Coverage}_{i,t} \cdot \textit{Geo. Disp.}_{i,t} \\ & + Z \cdot \textit{Controls}_{i,t} + \epsilon_{i,t},\end{aligned}$$

where $|\Delta \textit{ Firm-State-Qtr IO Share}_{i,s,t}|$ is the absolute value of the change in *Firm-State-Qtr IO Share* as defined above.

[9] For each quarter observation, we set the upper limit of the absolute change in institutional firm-state share equal to Max(Lag(Firm-State-Qtr IO Share); 1-Lag(Firm-State-Qtr IO Share)), because the firm-state-quarter share can at most go up to 1 or go down to 0 from its level at the beginning of the quarter.

Columns 2, 4, and 6 of Tables 5 and 6 report the parameter estimates from these regressions. The evidence provides strong and consistent support for the model's predictions. First, the coefficient on exposure is positive throughout all six models. This is consistent with investors located near the firm (HQ and/or other locations mentioned in the firm's 10-K filings) trading more intensely and, thus, their ownership share experiencing a larger quarter-to-quarter change. Second, the coefficient estimates on firm's geographic dispersion and information environment are consistent with both factors having a positive and significant impact on the trading intensity of investors located away from headquarters and other locations mentioned in the 10-K. Lastly, the coefficients on the interaction terms lend further support to the model's predictions. Namely, we find that greater firm geographic dispersion and improvement in the information environment are negatively related with the trading intensity of investors located near firm locations. In addition, the evidence supports the notion that there is substitution between the effects of more widely available private information (geographic dispersion) and improvements in public information environment (MD&A informativeness or analyst coverage) on the trading intensity of both nearby and distant investors.

5.2. Tests based on Exogenous Shocks to the Information Environment

The predictions of the theory rely on the assumption that the geographic distribution of private information and the public information environment are exogenous to firms and investors. Although our measures of information environment are useful because of their availability for a wide cross-sections of firms, they may be endogenous with respect to the geographic distribution of firms and investors. For example, firms with more complicated operations may provide higher quality disclosure, firms with more concentrated ownership may provide higher quality disclosure and/or followed by more analysts, analysts may choose to follow more complicated firms, etc. Therefore, our baseline tests raise potential concerns regarding reverse causality or omitted variable bias.

To address these potential concerns and attempt to identify the causal effects of firms' information environment on (1) investors' portfolio decisions and (2) how investors' proximity to firm locations affects those decisions, we exploit two different shocks that are unlikely to be driven by specific firm's conditions or managerial decisions. Each of these exogenous shocks is directly related to our baseline measures of information environment: MD&A informativeness and analyst coverage, respectively.

First, we exploit the regulation changes that occurred at the turn of the century, which intended to improve the reliability and informativeness of firm disclosures. Namely, various sections of the Sarbanes-Oxley Act (e.g., Section 302) of 2002 as well as new rules, guidance, and enforcement actions by the U.S. Securities and Exchange Commission between 2001 and 2003 would likely affect MD&A reliability and informativeness.[10] Consistent with the stated objective of these regulatory actions, Kogan, Routledge, Sagi, and Smith (2011) indeed find that the average informativeness of MD&A disclosures increases following the new rules of 2002-2003. Building on this insight, we replace the MD&A informativeness variable (which each individual firm can arguably influence) with an indicator variable for the passage of the new rules (which is not a firm-specific decision). In this analysis, the post-regulation improvement in the public information environment is an exogenous factor common across firms that is not due to firm-specific factors or choices.

We incorporate the regulatory regime shift in our baseline tests as follows.

$$
\begin{aligned}
\textit{Firm-State-Qtr IO Share}_{i,s,t} \text{ or } |\Delta \textit{Firm-State-Qtr IO Share}_{i,s,t}| \quad = \\
a \quad + \quad & b \cdot \textit{Exposure}_{i,s,t} + \; c \cdot \textit{PostReg.}_{i,t} + \; d \cdot \textit{Geo. Disp.}_{i,t} \\
+ \quad & e \cdot \textit{Exposure}_{i,s,t} \cdot \textit{Geo. Disp.}_{i,t} + \; f \cdot \textit{Exposure}_{i,s,t} \cdot \textit{PostReg.}_{i,t} \\
+ \quad & g \cdot \textit{PostReg.}_{i,t} \cdot \textit{Geo. Disp.}_{i,t} + \; h \cdot \textit{Exposure}_{i,s,t} \cdot \textit{PostReg.}_{i,t} \cdot \textit{Geo. Disp.}_{i,t} \\
+ \quad & Z \cdot \textit{Controls}_{i,t} + \epsilon_{i,t},
\end{aligned}
$$

where *PostReg.* is an indicator that takes the value of 1 on or after 2003, and 0 otherwise.

The parameter estimates from these regressions are reported in Table 7. The inferences from this table are very similar to those we obtain from Table 5. In particular, proximity to firm locations has a positive effect on state investors' holdings and trading intensity. Moreover, the firm's overall geographic dispersion and the regulatory regime shift have positive and significant effects on the holdings and trading intensities of investors located away from headquarters and other locations mentioned in the 10-K. The interaction terms also lend support to our predictions. First, geographic dispersion and the post-regulation regime negatively affect the holdings and trading intensity of investors located near the

[10] In December 2001, the Commission encouraged public companies to include explanations of their "critical accounting policies" in their MD&A and the likelihood that materially different amounts would be reported under different conditions (Release No. 33-8040 (Dec. 12, 2001)). In January 2002, the Commission issued explicit interpretive guidance on certain MD&A topics considered material to an understanding of companies' operations. In January 2003, the Commission adopted additional disclosure requirements regarding off-balance sheet arrangements and aggregate contractual obligations reported in the MD&A. (Release No. 33-8182 (Jan. 28, 2003)). Finally, in December 2003, issued further guidance to elicit more meaningful disclosure in a number of areas of the MD&A (Release Nos. 33-8350, 34-48960 (Dec. 29, 2003)).

firm's HQ or other locations mentioned in its 10-K. In an interesting contrast, the effects are reversed for investors located away from the firm. Second, the firm overall geographic dispersion and the new regulatory regime appear to be substitutes with respect to the holding and trading decisions of both nearby and distant investors.

Our second experiment exploits the exogenous variation in analyst coverage that results from mergers and closures of brokerage houses (Hong and Kacperczyk, 2010, Kelly and Ljungqvist, 2012, and Derrien and Kecskés, 2013).[11] The basic premise of prior studies using these events is that the resulting reduction in coverage for firms previously covered by affected analysts worsens the firms' information environment and increases information asymmetry. As in the case of the regulatory regime shift, the shock to the information environment is not due to firm-specific factors or choices. Moreover, given the largely random nature of the variation in affected firms, this shock lends itself to a difference-in-difference comparison between the affected group and a matched-control group of non-affected firms.

Following the logic in Derrien and Kecskés (2013), we limit our analysis to the sample of firm-quarters immediately surrounding the exogenous loss of analyst coverage. More precisely, we skip the quarter in which the merger/closing event occurs, and compare quarters -1 and +1 in our analysis. For each firm in our sample that experiences an exogenous loss of analyst coverage in quarter 0, we identify a control firm using the methodology in Derrien and Kecskés (2013). The focus of our tests is on how the changes in investors' holding and trading decisions around brokerage merger/closing events differ for the treatment and control samples.

We incorporate the exogenous loss of analyst coverage in our baseline tests as follows.

$$\textit{Firm-State-Qtr IO Share}_{i,s,t} \text{ or } |\Delta \textit{Firm-State-Qtr IO Share}_{i,s,t}| =$$

[11] We are grateful to Derrien and Kecskés for sharing the data used to identify analyst coverage terminations.

$$
\begin{aligned}
a \quad + \quad & b \cdot Exposure_{i,s,t} + c \cdot Geo.\ Disp._{i,t} + d \cdot Exposure_{i,s,t} \cdot Geo.\ Disp._{i,t} \\
+ \quad & e \cdot PostQuarter_{i,t} + \ f \cdot AnalystLoss_{i,t} \\
+ \quad & g \cdot PostQuarter_{i,t} \cdot AnalystLoss_{i,t} \\
+ \quad & h \cdot PostQuarter_{i,t} \cdot AnalystLoss_{i,t} \cdot Geo.\ Disp._{i,t} \\
+ \quad & i \cdot PostQuarter_{i,t} \cdot AnalystLoss_{i,t} \cdot Exposure_{i,s,t} \\
+ \quad & j \cdot PostQuarter_{i,t} \cdot AnalystLoss_{i,t} \cdot Exposure_{i,s,t} \cdot Geo.\ Disp._{i,t} \\
+ \quad & Z \cdot Controls_{i,t} + \epsilon_{i,t},
\end{aligned}
$$

where *PostQuarter* is an indicator variable that takes the value of 1 for observations in the quarter following the merger/closing, and 0 for observations in the quarter before the merger/closing. *AnalystLoss* is an indicator variable that takes the value of 1 for the treatment sample, i.e., firms that experience analyst loss due to merger/closing, and 0 for the matched-control sample. The difference-in-difference effects are captured by the parameter estimates of variables that contain the $PostQuarter_{i,t} \cdot AnalystLoss_{i,t}$ interaction term: g, h, i, and j.

The results of these tests are reported in Table 8. The inferences from this table reinforce those derived from Table 6. Again, proximity to firm locations has a positive effect on state investors' holdings and trading intensity, and the firm's overall geographic dispersion has a positive and significant effect on the holding and trading intensity of investors located away from the firm. More importantly, the sign of the interaction terms lend strong support to predictions of the theory. First, reduced analyst coverage results in a reduction of holdings and trading by state-investors away from firm locations (line 5 in Table 8). Second, this effect is tempered for more dispersed firms (line 6). Third, state-investors near firm locations actually increase their holdings and trading intensity in stocks that experience an exogenous loss of analyst coverage (line 7). This effect is tempered for more dispersed firms (line 8), again suggesting that analyst coverage and firm geographic dispersion are substitutes from the perspective of a nearby investors.

6. Stock Returns

In this section, we test the hypotheses pertaining to the relation between firms' returns, on the one hand, and investor proximity and information environment, on the other.

3a. Firms with a larger base of privately informed investors have lower expected returns.

3b. Firms with greater availability and quality of public information have lower expected returns.

3c. The marginal effect of geographic dispersion (public information) on expected returns decreases with an improvement in information environment (an increase in geographic dispersion).

Table 9 reports the results from univariate tests of prediction 3a when we sort stocks into terciles based on their Potential Investor Distance. Panel A reports the equal-weighted average monthly return of each tercile portfolio, while Panel B reports the value-weighted average return. We also report the abnormal return of a zero-cost portfolio with long positions in firms located far from potential investors and short positions in firms located near potential investors. Abnormal returns are computed by controlling for standard risk factors (using 4-factor Carhart (1997) model) or firm characteristics (using Daniel, Grinblatt, Titman, and Wermers (1997) characteristic adjustment model; henceforth DGTW).

The univariate evidence shows that firms located far from potential investors outperform those located near potential investors by a considerable margin. The annualized abnormal performance differential ranges between $3.3 - 4.4$ percent for the equal-weighted portfolios and $4.4 - 8.3$ percent for the value-weighted portfolios. The relatively larger differential for the value-weighted portfolios suggests that the proximity-based results are not driven by small-cap stocks unlike most return differentials documented in the literature.

To also account for the joint effects of firm's geographic dispersion and information environment while controlling for other stock characteristics, we estimate multiple regression models of monthly stock returns. The dependent variable is the monthly DGTW characteristic-adjusted return. In these models, we include stock return in the prior month to account for potential short-term contrarian effects (Jegadeesh, 1990), effective spread to control for potential liquidity effects, as well as idiosyncratic volatility and skewness.

Table 10 and 11 report the Fama and MacBeth (1973) estimates of the following two models of monthly returns when we proxy for the firm information environment by the MD&A informativeness and analyst coverage, respectively.

$$
\begin{aligned}
Return_{i,t} \quad = \quad & a + b \cdot Potential\ Investor\ Distance_{i,t} \\
& + c \cdot Geo.\ Disp._{i,t} + d \cdot MD\&A\ Quality_{i,t} \\
& + e \cdot Potential\ Investor\ Distance_{i,t} \cdot Geo.\ Disp._{i,t} \\
& + f \cdot Potential\ Investor\ Distance_{i,t} \cdot MD\&A\ Quality_{i,t} \\
& + g \cdot Potential\ Investor\ Distance_{i,t} \cdot Geo.\ Disp._{i,t} \cdot MD\&A\ Quality_{i,t} \\
& + H \cdot Controls_{i,t} + \epsilon_{i,t}
\end{aligned}
$$

and

$$
\begin{aligned}
Return_{i,t} \quad = \quad & a + b \cdot Potential\ Investor\ Distance_{i,t} \\
& + c \cdot Geo.\ Disp._{i,t} + d \cdot AnalystCoverage_{i,t} \\
& + e \cdot Potential\ Investor\ Distance_{i,t} \cdot Geo.\ Disp._{i,t} \\
& + f \cdot Potential\ Investor\ Distance_{i,t} \cdot AnalystCoverage_{i,t} \\
& + g \cdot Potential\ Investor\ Distance_{i,t} \cdot Geo.\ Disp._{i,t} \cdot AnalystCoverage_{i,t} \\
& + H \cdot Controls_{i,t} + \epsilon_{i,t}
\end{aligned}
$$

The first column of each table reports the specification that includes our HQ proximity measure, (Log) Potential Investor Distance, the firm overall geographic dispersion tercile, its disclosure quality tercile (in Table 10) or its analyst coverage tercile (in Table 11), and other firm characteristic as control variables. The coefficient estimate on Potential Investor Distance is positive and statistically significant, and of a similar economic magnitude to that obtained in the univariate setting. The estimates from alternative specifications reported in other columns of each table continue to indicate that Potential Investor Distance explains a significant fraction of the variation in the cross-section of monthly returns. Moreover, the estimated coefficients on firms' geographic dispersion and MD&A informativeness (in Table 10) have the predicted signs. Nonetheless, the MD&A variable alone does not appear to be significantly related to returns.

The last two columns in each table report the full model estimates, based on two alternative measures of Potential Investor Distance from the HQ – the raw log-distance measure and the corresponding firm tercile rank in the prior calendar year, respectively. The evidence from the full models largely supports the predictions of the model. First, the estimate on Potential Investor Distance continues to be positive and significant, both statistically and economically. Second, we find that the parameter estimate for the interaction between Potential Investor Distance and geographic dispersion is significantly negative. That is, consistent with a wider distribution of private information reducing information asymmetry and therefore adverse selection risk for (privately) uninformed investors, potential investors' distance has a weaker effect on returns when the firm's economic interests are more geographically dispersed. Third, we find similar results for the interaction between Potential Investor Distance and improvements in the firm public information environment – i.e., MD&A informativeness or analyst coverage. That is, consistent with the availability and precision of public information reducing adverse selection risk for (privately) uninformed investors, potential investors' distance has a weaker effect on stock returns when the information environment improves.

The estimates reported in the last column of Tables 10 and 11, based on potential investors' distance terciles, allow a direct comparison of the economic magnitudes of the effects of Potential Investor Distance and those of the other variables of interest: geographic dispersion, MD&A informativeness, and analyst coverage. The evidence suggests that there are economically strong substitution effects between Potential Investor Distance, on one hand, and firm geographic dispersion and information environment, on the other - predictions 3a and 3b. For firms in the bottom geographic dispersion *and* MDA informativeness terciles, the effect of Potential Investor Distance on returns is economically large: moving from the bottom to the top tercile of Potential Investor Distance corresponds with 0.822 (0.708) percent increase in abnormal monthly returns. This contrasts sharply with the effect of Potential Investor Distance for firms in the top geographic dispersion or MDA informativeness terciles, for which Potential Investor Distance has no significant effect on subsequent returns. Lastly, consistent with prediction 3c, the sign and magnitude of the triple interaction terms in both tables suggest that firm geographic dispersion and information environment have substituting effects on the relation between Potential Investor Distance and stock returns.

Overall, the evidence in Tables 10 and 11 indicates that Potential Investor Distance is a strong predictor of stock returns, and this relation is strongest for firms that have more geographically concentrated economic interests or poor public information environment.

7. Conclusion

Theory suggests that investors' potential access to private information and firms' public information environment determine the distribution of information across investors and, thus, firms' dispersion of ownership and stock returns. Based on this insight, and using a number of novel measures for investors' proximity and the quality of public information, we derive and test several empirical predictions that link a firm's geographic distribution of ownership and stock returns to investors' proximity to firm locations and the firm's public information environment. The empirical evidence provides strong and consistent support for the predictions of the theory.

First, we find that the geographic dispersion of a firm's institutional shareholders increases with both the geographic dispersion of the firm's potential investors and improvements in its public information environment, and that at the margin there is substitution between these two determinants of the geographic distribution of firm ownership. Second, the intensity of investors' trading in a firm's stock is strongly related to the same factors that determine the firm's ownership dispersion and, again, there is evidence of substitution between firm's geographic dispersion and public information environment. Third, tests that exploit plausibly exogenous variation in firms' information environment indicate that its effects on ownership and trading dispersion, as well as on their relation with investor proximity to the firm, are causal in nature. Lastly, as suggested by theory, we find that the determinants of the geographic dispersion of ownership and trading have similarly consistent effects on stock returns. Namely, firms whose locations are furthest from potential investors have higher stock returns than those located closest to potential investors, and this effect is largely tempered by the quantity and quality of public information. Collectively, our evidence lends support to the idea that the geographic distribution of private information and the public information environment jointly determine firms' ownership structure and cost of capital.

References

Aarland, Kristin, James C. Davis, J. Vernon Henderson, and Yukako Ono, 2007, Spatial organization of firms: the decision to split production and administration, *RAND Journal of Economics* 38, 480494.

Baik, Bok, Jun-Koo Kang, and Jin-Mo Kim, 2010, Local institutional investors, information asymmetries, and equity returns, *Journal of Financial Economics* 97, 81–106.

Baiman, Stanley, and Robert E. Verrecchia, 1996, The relation among capital markets, financial disclosure, production efficiency, and insider trading, *Journal of Accounting Research* 34, 1–22.

Barber, Brad M., and Terrance Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785–818.

Bernile, Gennaro, Alok Kumar, and Johan Sulaeman, 2015, Home away from home: Geography of information and local investors, *Review of Financial Studies* , forthcoming.

Bradshaw, Mark T., Brian J. Bushee, and Gregory S. Miller, 2004, Accounting choice, home bias, and U.S. investment in non-U.S. firms, *Journal of Accounting Research* 42, 795–841.

Brown, Nerissa C., J. Han Stice, and Roger White, 2015, Mobile communication and local information flow: Evidence from distracted driving laws, *Journal of Accounting Research*.

Bushee, Brian J., 1998, The influence of institutional investors on myopic R&D investment behavior, *Accounting Review* 73, 305–333.

——— , and Christopher F. Noe, 2000, Corporate disclosure practices, institutional investors, and stock return volatility, *Journal of Accounting Research* 38, 171–202.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chemmanur, Thomas J., and An Yan, 2009, Advertising, attention, and stock returns, Working Paper; Available at SSRN: http://ssrn.com/abstract=1340605.

Corvig, Vicentiu M., Mark L. DeFond, and Mingyi Hung, 2007, Home bias, foreign mutual fund holdings, and the voluntary adoption of international accounting standards, *Journal of Accounting Research* 45, 41–70.

Coval, Joshua D., and Tobias J. Moskowitz, 1999, Home bias at home: Local equity preference in domestic portfolios, *Journal of Finance* 54, 2045–2073.

——— , 2001, The geography of investment: Informed trading and asset prices, *Journal of Political Economy* 109, 811–841.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In search of attention, *Journal of Finance* 66, 1461–1499.

Dai, Lili, Jerry T. Parwada, and Bohui Zhang, 2015, The governance effect of the media's news dissemination role: Evidence from insider trading, *Journal of Accounting Research*, forthcoming.

Daniel, Kent D., Mark Grinblatt, Sheridan Titman, and Russell Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035–1058.

Davis, James C., and J. Vernon Henderson, 2008, The agglomeration of headquarters, *Regional Science and Urban Economics* 38, 445–460.

Dechow, Patricia M., and Ilia D. Dichev, 2002, The quality of accruals and earnings: The role of accrual estimation errors, *The Accounting Review* 77, 35?59.

Derrien, François, and Ambrus Kecskés, 2013, The real effects of financial shocks: Evidence from exogenous changes in analyst coverage, *Journal of Finance* 68, 1407–1440.

Diamond, Douglas W., and Robert E. Verrecchia, 1991, Disclosure, liquidity, and the cost of capital, *Journal of Finance* 46, 1325–1359.

Easley, David, and Maureen O'Hara, 2004, Information and the cost of capital, *Journal of Finance* 59, 1553–1583.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Fang, Lily, and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *Journal of Finance* 64, 20232052.

Garcia, Diego, and Oyvind Norli, 2012, Geographic dispersion and stock returns, *Journal of Financial Economics* forthcoming.

Gervais, Simon, Ron Kaniel, and Dan H. Mingelgrin, 2001, The high-volume return premium, *Journal of Finance* 56, 877–919.

Gillan, Stuart L., and Laura T. Starks, 2003, Corporate governance, corporate ownership, and the role of institutional investors: A global perspective, *Journal of Applied Finance* 13, 4–22.

Giroud, Xavier, 2011, Proximity and investment: Evidence from plant-level data, Working Paper; Available at SSRN: http://ssrn.com/abstract=1958200.

Grullon, Gustavo, George Kanatas, and James P. Weston, 2004, Advertising, breadth of ownership, and liquidity, *Review of Financial Studies* 17, 439–461.

Hong, Harrison, and Marcin Kacperczyk, 2010, Competition and bias, *Quarterly Journal of Economics* 125, 1683?1725.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2008, The only game in town: Stock-price consequences of local bias, *Journal of Financial Economics* 90, 20–37.

Hou, Kewei, Lin Peng, and Wei Xiong, 2009, A tale of two anomalies: The implications of investor attention for price and earnings momentum, Working Paper; Available at SSRN: http://ssrn.com/abstract=976394.

Huberman, Gur, 2001, Familiarity breeds investment, *Review of Financial Studies* 14, 659–680.

Ivković, Zoran, and Scott Weisbenner, 2005, Local does as local is: Information content of the geography of individual investors' common stock investments, *Journal of Finance* 60, 267–306.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *Journal of Finance* 45, 881–898.

Kelly, Bryan, and Alexander Ljungqvist, 2012, Testing asymmetric-information asset pricing models, *Review of Financial Studies* 25, 1366?1413.

Kogan, Shimon, Bryan Routledge, Jacob S. Sagi, and Noah A. Smith, 2011, Information content of public firm disclosures and the Sarbanes-Oxley Act, Working Paper, Available at SSRN: http://ssrn.com/abstract=1584763.

Korniotis, George M., and Alok Kumar, 2010, State-level business cycles and local return predictability, Working Paper (November); Available at http://ssrn.com/abstract=1094560.

Lambert, Richard, Christian Leuz, and Robert E. Verrecchia, 2007, Accounting information, disclosure, and the cost of capital, *Journal of Accounting Research* 45, 385–420.

Leuz, Christian, and Robert E. Verrecchia, 2000, The economic consequences of increased disclosure, *Journal of Accounting Research* 38, 91–124.

Li, Feng, 2006, Do stock market investors understand the risk sentiment of corporate annual reports?, Working Paper.

Li, Feng, 2011, Textual analysis of corporate disclosures: A survey of the literature, *Journal of Accounting Literature* 29, 143?165.

Liberti, Jose M., and Atif R. Mian, 2009, Estimating the effect of hierarchies on information use, *Review of Financial Studies* 22, 4057–4090.

Lou, Dong, 2009, Attracting investor attention through advertising, Working Paper; London School of Economics.

Lustig, Hanno, and Stijn van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia: An empirical perspective, *Journal of Finance* 60, 1167–1219.

———, 2010, How much does household collateral constrain regional risk sharing?, *Review of Economic Dynamics* 13, 265–294.

Malloy, Christopher J., 2005, The geography of equity analysis, *Journal of Finance* 60, 719–755.

Merton, Robert C., 1987, A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483–510.

Miller, Gregory S., and Douglas J. Skinner, 2015, The evolving disclosure landscape: How changes in technology, the media, and capital markets are affecting disclosure, *Journal of Accounting Research*, forthcoming.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite heteroskedasticity and auto-correlation consistent variance-covariance matrix, *Econometrica* 55, 703–708.

Petersen, Mitchell A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

Pirinsky, Christo, and Qinghai Wang, 2006, Does corporate headquarters location matter for stock returns?, *Journal of Finance* 61, 19912015.

Seasholes, Mark S., and Guojun Wu, 2007, Predictable behavior, profits, and attention, *Journal of Empirical Finance* 14, 590–610.

Seasholes, Mark S., and Ning Zhu, 2010, Individual investors and local bias, *Journal of Finance* 65, 1987-2010.

van Nieuwerburgh, Stijn, and Laura Veldkamp, 2009, Information immobility and the home bias puzzle, *Journal of Finance* 64, 1187–1215.

Verrecchia, Robert E., 2001, Essays on disclosure, *Journal of Accounting and Economics* 32, 97?180.

Yuan, Yu, 2011, Attention and trading, Working Paper; Available at SSRN: http://ssrn.com/abstract=1105532.

Appendix A: Model

In this section we present a partially revealing equilibrium model with multiple risky assets and two signal types. Less theoretically inclined readers can skip this section and proceed to section 3 where we provide the empirical predictions of the model.

A.1 Model setup

Building on Easley and O'Hara (2004), we assume a two-period world in which investors make their portfolio choice decisions at time 0 and realize the return on their portfolio at time 1. There are K firms and I investors. There is one risk-free asset, whose price is normalized to 1. The value of stock k at time 1, $\tilde{\vartheta}_k$, is normally distributed with mean $\bar{\vartheta}_k$ and precision ρ_k. The supply of stock k, $\tilde{x}_k$, is random and distributed normally with mean $\bar{x}_k$ and precision η_k. We assume that all risky assets are in positive net supply. The price vector of assets is given by $(1, p_1, ...p_k)$ and the payoff vector that is realized at time 1 is $(1, \tilde{\vartheta}_1, ..., \tilde{\vartheta}_k)$.

Each investor receives two signals about the future value of each stock. The first signal is a public signal observed by all investors. The precision of this signal is determined by the firm. Some firms might prefer to send a more precise signal than others. We call this signal the *disclosure signal.* The second signal is a private signal, which provides more precise information compared to the public signal. We assume that the more familiar a firm is to investors, the more precise the signal is. We call this signal the *familiarity signal. Informed investors*, who are familiar with the firm, receive the private signal with high precision; *uniformed investors*, who are unfamiliar with the firm, do not receive a private signal. The number of (or the ratio of) informed investors can be driven by the location of the firm, e.g., how many investors happen to reside next to the firm's HQ and the geographic distribution of the firm's economic interests. Firms that are geographically dispersed or have higher exposure to potential traders have more informed traders than geographically concentrated firms. [12]

While informed investors receive two signals about the future value of the stock before they make their investment decisions, uninformed investors essentially receive only one. The public signal, s_{k_0}, is normally distributed with mean ϑ_k and precision γ_{k_0}, whereas the private signal, s_k is normally distributed with mean ϑ_k and precision γ_k. By definition, the private signal is more precise than the public one such that $\gamma_k > \gamma_{k_0}$. Firm k has I_k investors who have enough exposure to the firm to receive the private signal. This corresponds to α_k fraction of total investor base I. All of the random variables are independent and investors know their distributions. All signals are independent conditional on ϑ_k. Firms with greater exposure will have a higher α_k, a greater informed investor base. For simplicity, assume that all firms operate in a single location and every location has the same total number of investors. The aim is to obtain comparative statistics between firms with high and low exposure. By varying α_k, we vary the ratio of privately informed investors and by varying γ_{k_0}, we vary the precision of the public signal that the firm chooses to disclose.

[12] Here, for simplicity, we assume that the uninformed investor does not receive the private signal. Another way to model this is allow for each type of investor to receive one private signal that has a different precision than the signal of the other type, depending on their exposure to the firm.

A.2 Investors' problems

Each investor chooses his demands for assets $k = 1, ..., K$ in order to maximize his expected utility, given the initial wealth ω_0^i. We assume that investors have CARA utility, and because all distributions are normal, they solve the following standard mean-variance objective function:

$$\max_{\{z_k^i\}_{k=1}^K} \quad E_i[\tilde{\omega}^i] - \frac{\delta}{2} Var_i[\tilde{\omega}^i]$$

where the next period's wealth is given by:

$$\tilde{\omega}^i = \sum_k (\tilde{\vartheta}_k - p_k) z_k^i + \omega_0^i$$

Here, z_k^i denotes the number of shares of stock k the investor purchases. We assume that the payoffs on stocks are independent. Let $\bar{\vartheta}_k^i$ and $\bar{\rho}_k^i$ denote the expectation and variance of $\tilde{\vartheta}_k$ conditional on all of investor i's information. Then, investor i's demand function for asset k is given by:

$$z_k^i = \frac{\bar{\vartheta}_k^i - p_k}{\delta (\bar{\rho}_k^i)^{-1}}$$

The demand function for asset k depends on the signals that investor i received. Let us first solve for the informed investor's demand. The expected value of the payoff from asset k conditional on the informed investor's information is given by:

$$\bar{\vartheta}_k^i = \frac{\rho_k \bar{\vartheta}_k + \gamma_{k_0} s_{ko} + \gamma_k s_k}{\bar{\rho}_k^i}$$

with precision

$$\bar{\rho}_k^i = \rho_k + \gamma_{k_0} + \gamma_k$$

Then, the demand function for the informed investor is given by:

$$z_k^{i*} = \frac{\rho_k \bar{\vartheta}_k + \gamma_{k_0} s_{ko} + \gamma_k s_k - p_k(\rho_k + \gamma_{k_0} + \gamma_k)}{\delta} \equiv DI_k^*(s_{k_0}, s_k, p_k)$$

Uninformed investors receive a public signal but not a private signal. They also know that the informed traders' demand affects the equilibrium price. Suppose the uninformed investors conjecture the following price function:

$$p_k = a\bar{\vartheta}_k + bs_k + cs_{k_0} - dx_k + e\bar{x}_k$$

where a, b, c, d, and e are coefficients to be determined. Define:

$$\theta_k = \frac{p_k - a\bar{\vartheta} - cs_{k_0} + \bar{x}_k(d-e)}{b} = s_k - \frac{d}{b}(x_k - \bar{x}_k)$$

The random variable θ_k is distributed normally with mean ϑ_k and precision

$$\rho_{\theta_k} = \left[\left(\frac{d}{b}\right)^2 \eta_k^{-1} + \gamma_k^{-1}\right]^{-1}$$

Having the same mean value $\bar{\vartheta}_k$ with the signals, the variable θ_k simplifies our derivations. The conditional mean and variance from the uninformed investor's perspective are given by:

$$\bar{\vartheta}_k^i = \frac{\rho_k\bar{\vartheta}_k + \gamma_{k_0}s_{ko} + \rho_{\theta_k}\theta_k}{\bar{\rho}_k^i}$$

where

$$\bar{\rho}_k^i = \rho_k + \gamma_{k_0} + \rho_{\theta_k}$$

Uninformed investor's demand is given by:

$$z_k^{i*} = \frac{\rho_k\bar{\vartheta}_k + \gamma_{k_0}s_{ko} + \rho_{\theta_k}\theta_k - p_k(\rho_k + \gamma_{k_0} + \rho_{\theta_k})}{\delta} \equiv DU_k^*(s_{k_0}, \theta_k, p_k)$$

A.3 Equilibrium

In equilibrium, for each asset k, supply must equal demand:

$$\alpha_k DI_k^*(s_{k_0}, s_k, p_k) + (1-\alpha_k)DU_k^*(s_{k_0}, \theta_k, p_k) = x_k$$

Solving for p_k yields Proposition 1.

Proposition 1: There exists a partially revealing rational expectations equilibrium in which, for each asset k,

$$p_k = a\bar{\vartheta}_k + bs_k + cs_{k_0} - dx_k + e\bar{x}_k$$

where

$$a = \frac{\rho_k}{C_k}, \quad b = \frac{\alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}}{C_k}, \quad c = \frac{\gamma_{k_0}}{C_k}, \quad d = \frac{\alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}}{\alpha_k\gamma_k C_k}\delta$$

$$e = \frac{(1-\alpha_k)\rho_{\theta_k}}{\alpha_k\gamma_k C_k}\delta, \quad \rho_{\theta_k} = \left[\left(\frac{\delta}{\alpha_k\gamma_k}\right)^2 \frac{1}{\eta_k} + \frac{1}{\gamma_k}\right]^{-1}$$

where

$$C_k = \rho_k + \gamma_{k_0} + \alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}$$

Proof: There exists a partially revealing rational expectations equilibrium in which, for each asset k,

$$\begin{aligned}
\delta x_k &= \rho_k\bar{\vartheta}_k + \gamma_{k_0}s_{ko} - p_k(\rho_k+\gamma_{k_0}) + \alpha_k\gamma_k(s_k - p_k) + (1-\alpha_k)\rho_{\theta_k}(\theta_k - p_k) \\
&= \rho_k\bar{\vartheta}_k + \gamma_{k_0}s_{ko} + \alpha_k\gamma_k s_k + (1-\alpha_k)\rho_{\theta_k}\theta_k - p_k[\rho_k + \gamma_{k_0} + \alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}] \\
&= \rho_k\bar{\vartheta}_k + \gamma_{k_0}s_{ko} + \alpha_k\gamma_k s_k + (1-\alpha_k)\rho_{\theta_k}\left[s_k - \frac{d}{b}(x_k - \bar{x}_k)\right] \\
&\quad - p_k[\rho_k + \gamma_{k_0} + \alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}] \\
&= \rho_k\bar{\vartheta}_k + \gamma_{k_0}s_{ko} + [\alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}]\,s_k - (1-\alpha_k)\rho_{\theta_k}\left[\frac{d}{b}(x_k - \bar{x}_k)\right] \\
&\quad - p_k[\rho_k + \gamma_{k_0} + \alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}]
\end{aligned}$$

Then,

$$\begin{aligned}
\left[\delta + (1-\alpha_k)\rho_{\theta_k}\frac{d}{b}\right]x_k &= \rho_k\bar{\vartheta}_k + \gamma_{k_0}s_{ko} + [\alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}]\,s_k + \ (1-\alpha_k)\rho_{\theta_k}\frac{d}{b}\bar{x}_k \\
&\quad - p_k[\rho_k + \gamma_{k_0} + \alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}]
\end{aligned}$$

Define

$$C_k = \rho_k + \gamma_{k_0} + \alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}$$

$$\begin{aligned}
d &= \frac{\delta + (1-\alpha_k)\rho_{\theta_k}\frac{d}{b}}{C_k} = \frac{\alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}}{\alpha_k\gamma_k C_k}\delta \\
a &= \frac{\rho_k}{C_k} \\
b &= \frac{\alpha_k\gamma_k + (1-\alpha_k)\rho_{\theta_k}}{C_k} \\
c &= \frac{\gamma_{k_0}}{C_k} \\
e &= \frac{(1-\alpha_k)\rho_{\theta_k}\frac{d}{b}}{C_k} = \frac{(1-\alpha_k)\rho_{\theta_k}}{\alpha_k\gamma_k C_k}\delta \\
\rho_{\theta_k} &= \left[\left(\frac{\delta}{\alpha_k\gamma_k}\right)^2\eta_k^{-1} + \gamma_k^{-1}\right]^{-1} = \frac{\alpha_k^2\eta_k\gamma_k^2}{\delta^2 + \alpha_k^2\gamma_k\eta_k}
\end{aligned}$$

Then, substituting and solving for p_k yields

$$p_k = a\bar{\vartheta}_k + bs_k + cs_{k_0} - dx_k + e\bar{x}_k$$

A.4 Cross-section of asset returns

The return on holding asset k is given by $\vartheta_k - p_k$, and the expected return to an investor with the information set I is $E[\vartheta_k|I] - p_k$. The average return that an outside investor could compute is $E[E[\vartheta_k|I] - p_k] = E[\vartheta_k - p_k]$.[13]

Proposition 2: The expected return per share for stock k is given by

$$E[v_k - p_k] = \frac{\bar{x}_k \delta}{C_k}$$

Proof: The expected return per share for stock k is given by

$$\begin{aligned} E[v_k - p_k] &= E\left[\vartheta_k - a\bar{\vartheta}_k - bs_k - cs_{k_0} + dx_k - e\bar{x}_k\right] \\ &= \bar{\vartheta}_k[1 - a - b - c] + \bar{x}_k(d - e) \\ &= \frac{\bar{x}_k \delta}{C_k} = \frac{\bar{x}_k \delta}{\rho_k + \gamma_{k_0} + \alpha_k \gamma_k + (1 - \alpha_k)\rho_{\theta_k}} \end{aligned}$$

Here, there are two ways to incorporate the firm's disclosure decision into the model. First, by increasing the number of public signals, one could allow for a greater amount of public information to be available. Second, which is the case that we follow, is to increase the quality of the public information. The next proposition explains the impact of the precision of public signal on expected stock returns.

Proposition 3.a: For any stock k, an increase in the precision of the public signal γ_{k_0} decreases the equilibrium required return, or

$$\frac{\partial E[v - p_k]}{\partial \gamma_{k_0}} = \frac{-\bar{x}_k \delta}{(C_k)^2} < 0$$

Holding the precision of the private signal constant, the expected stock return is lower when the public signal is more precise. This result follows from the fact that more precise public signal will result in more informative prices, which in turn will increase the uninformed demand for the stock and decrease the cost of capital. In the next proposition, we derive the impact of exposure on expected stock returns.

Proposition 3.b: For any stock k, the equilibrium required return decreases with the exposure parameter

[13] Here, we use the words *payoff* and *return* interchangeably. We do this because it is easier to calculate the expectations of payoffs than that of returns. Throughout our comparative statistics analysis, we will fix the mean value of the firm. This is the sufficient condition that allows us to use payoffs instead of returns. To see this note that $E[\vartheta - p_k] > E[\vartheta - p_j]$ holds if and only if $E[p_j - p_k] > 0$, which in turn implies that $E\left[\frac{\vartheta}{p_k}\right] > E\left[\frac{\vartheta}{p_j}\right]$. In other words, from two assets that yield the same expected value ϑ, the one with lower price will yield a higher expected return and a higher expected payoff. Note that this follows from the formula $E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)}\left[1 + \frac{\sigma_Y}{(E(Y))^2} - \frac{\sigma_{XY}}{E(X)E(Y)}\right]$ and the fact that $\frac{\sqrt{Var(p_k)}}{E(p_k)} = \frac{Cov(p_k, v_k)}{E(v_k)}$.

such that

$$\frac{\partial E[v-p_k]}{\partial \alpha_k} = \frac{-\delta \bar{x}_k \left[\frac{\delta^2\gamma_k}{\delta^2+\alpha_k^2\gamma_k\eta_k} + (1-\alpha_k)\frac{2\alpha_k\eta_k\gamma_k^2\delta^2}{[\delta^2+\alpha_k^2\gamma_k\eta_k]^2}\right]}{(C_k)^2} < 0 \tag{1}$$

Proof: For any stock k, the equilibrium required return decreases with the exposure parameter such that

$$\begin{aligned}
\frac{\partial E[v_k-p_k]}{\partial \alpha_k} &= \frac{-\delta \bar{x}_k \left[\gamma_k - \rho_{\theta_k} + (1-\alpha_k)\frac{\partial \rho_{\theta_k}}{\partial \alpha_k}\right]}{(C_k)^2} \\
&= \frac{-\delta \bar{x}_k \left[\gamma_k - \rho_{\theta_k} + (1-\alpha_k)\frac{2\alpha_k\eta_k\gamma_k^4\delta^2}{[\delta^2\gamma_k+\alpha_k^2\gamma_k^2\eta_k]^2}\right]}{(C_k)^2} \\
&= \frac{-\delta \bar{x}_k \left[\gamma_k - \frac{\alpha_k^2\eta_k\gamma_k^3}{\delta^2\gamma_k+\alpha_k^2\gamma_k^2\eta_k} + (1-\alpha_k)\frac{2\alpha_k\eta_k\gamma_k^4\delta^2}{[\delta^2\gamma_k+\alpha_k^2\gamma_k^2\eta_k]^2}\right]}{(C_k)^2} \\
&= \frac{-\delta \bar{x}_k \left[\frac{\delta^2\gamma_k^2}{\delta^2\gamma_k+\alpha_k^2\gamma_k^2\eta_k} + (1-\alpha_k)\frac{2\alpha_k\eta_k\gamma_k^4\delta^2}{[\delta^2\gamma_k+\alpha_k^2\gamma_k^2\eta_k]^2}\right]}{(C_k)^2} < 0
\end{aligned}$$

The derivative of C_k with respect to α_k is given by

$$\frac{\partial C_k}{\partial \alpha_k} = \frac{\delta^2\gamma_k}{\delta^2+\alpha_k^2\gamma_k\eta_k} + (1-\alpha_k)\frac{2\alpha_k\eta_k\gamma_k^2\delta^2}{[\delta^2+\alpha_k^2\gamma_k\eta_k]^2}$$

There are two channels through which exposure affects the price and the expected return. First, when the firm has more exposure to investors, there are more informed investors, which decreases the return to existing informed investors. However, note that the total informed demand might be higher or lower than the low-exposure case. Second, when there are more informed investors, price is more informative about the private signal, which makes the asset less risky for the uninformed investors, further decreasing the cost of capital.

A.5 Equilibrium portfolios of informed and uninformed investors

Stocks are riskier for uniformed investors than they are for informed investors. This difference would affect their demand for stocks. To see this, we calculate the difference between the demands of informed and uninformed investors:

$$z_k^I - z_k^U = [\gamma_k s_k - \rho_{\theta_k}\theta_k + p_k(\rho_{\theta_k} - \gamma_k)]\delta^{-1}$$

The ex-ante difference in the holdings of informed and uninformed investors is given by:

$$\begin{aligned}E[z_k^I - z_k^U] &= \delta^{-1}E[\tilde{\vartheta}_k(\gamma_k - \rho_{\theta_k}) + p_k(\rho_{\theta_k} - \gamma_k)] \\ &= \left(\frac{\gamma_k - \rho_{\theta_k}}{\delta}\right)E[\tilde{\vartheta}_k - \rho_{\theta_k}] \\ &= (\gamma_k - \rho_{\theta_k})\left(\frac{\bar{x}_k}{C_k}\right) > 0\end{aligned}$$

Note that the result follows from the fact that $\gamma_k > \rho_{\theta_k}$ always holds. On average, informed investors hold more of the risky asset compared to uninformed investors. However, informed investors' holdings will be larger or smaller than the uninformed investor's holding depending upon the private information they receive. The next two propositions report the impact of exposure and precision of the public signal on the ex-ante difference in holdings.

Proposition 4.a: The derivative of the ex-ante difference between the investors' holdings with respect to the public signal precision is given by

$$\frac{\partial E[z_k^I - z_k^U]}{\partial \gamma_{k_0}} = -(\gamma_k - \rho_{\theta_k})\frac{\bar{x}_k}{(C_k)^2} < 0 \tag{2}$$

An increase in the precision of the public signal increases the uninformed investor demand and decreases the average cost of capital.

Proposition 4.b: The derivative of the average difference between the investors' holdings with respect to the exposure parameter is given by

$$\frac{\partial E[z_k^I - z_k^U]}{\partial \alpha_k} = \left(\frac{\gamma_k - \rho_{\theta_k}}{\delta}\right)\frac{\partial E[v - p_k]}{\partial \alpha_k} - \left(\frac{\bar{x}_k}{C_k}\right)\frac{2\alpha_k\eta_k\gamma_k^2\delta^2}{[\delta^2 + \alpha_k^2\gamma_k\eta_k]^2} < 0 \tag{3}$$

which follows from $\frac{\partial E[v-p_k]}{\partial \alpha_k} < 0$.

An increase in the exposure has two impacts. First, it increases the aggregate demand from informed investors since there are more informed investors. Second, the price becomes more informative, which in turn decreases the risk for the uninformed. As the derivative above shows, the second effect is larger than the first one and in expectation, an increase in exposure increases holdings of uninformed investors more than that of informed investors.

A.6 Volatility of informed demand

In this subsection, we derive the implications of the model for the variance of informed investors' holdings. Before that, we calculate the volatility of the equilibrium price:

$$Var(p_k) = \frac{b^2}{\gamma_k} + \frac{c^2}{\gamma_{k_0}} + \frac{d^2}{\eta_k} + \frac{2bc}{\rho_k}$$

Define:

$$y_k = \rho_k + \gamma_{k_0} + \gamma_k$$

The variance of the demand by an individual informed investor is given by:

$$Var(z_k^I) = \frac{1}{\delta^2}\left[\gamma_{k_0} + \gamma_k + y_k^2 Var(p_k) + \frac{2\gamma_{k_0}\gamma_k}{\rho_k} - 2y_k\left(\gamma_{k_0}\left[\frac{c}{\gamma_{k_0}} + \frac{b}{\rho_k}\right] + \gamma_k\left[\frac{c}{\rho_k} + \frac{b}{\gamma_k}\right]\right)\right]$$

Proposition 5.a: The derivative of the variance of demand by an individual informed investor with respect to the precision of the public signal γ_{k_0} is given by:

$$\begin{aligned}\frac{\partial Var(z_k^I)}{\partial \gamma_{k_0}} &= \frac{1}{\delta^2}\left[1 + 2y_k Var(p_k) + y_k^2 \frac{\partial Var(p_k)}{\partial \gamma_{k_0}} + \frac{2\gamma_k}{\rho_k}\right] - \frac{2}{\delta^2}\left(\gamma_{k_0}\left[\frac{c}{\gamma_{k_0}} + \frac{b}{\rho_k}\right] + \gamma_k\left[\frac{c}{\rho_k} + \frac{b}{\gamma_k}\right]\right) \\ &\quad -2\frac{y_k}{\delta^2}\left[\left(1 + \frac{\gamma_{k_0}}{\rho_k}\right)\frac{\partial b}{\partial \gamma_{k_0}} + \left(1 + \frac{\gamma_k}{\rho_k}\right)\frac{\partial c}{\partial \gamma_{k_0}}\right]\end{aligned}$$

Proposition 5.b: The derivative of the variance of demand by an individual informed investor with respect to the exposure parameter α_k is given by:

$$\frac{\partial Var(z_k^I)}{\partial \alpha_k} = \frac{y_k}{\delta^2}\left[y_k \frac{\partial Var(p_k)}{\partial \alpha_k} - 2\frac{\partial b}{\partial \alpha_k}\left(1 + \frac{\gamma_{k_0}}{\rho_k}\right)\right]$$

Table A.1 reports the values of variance for different parameter values. The increase in the precision of the public signal decreases the variance of individual informed investors. Similarly, the exposure parameter has a negative impact on individual demand.

	α_k								
γ_{k_0}	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	3.807	3.394	2.956	2.538	2.164	1.842	1.571	1.344	1.156
0.2	3.564	3.221	2.839	2.463	2.117	1.814	1.555	1.336	1.153
0.3	3.355	3.070	2.736	2.395	2.075	1.789	1.541	1.329	1.150
0.4	3.173	2.937	2.644	2.335	2.037	1.766	1.528	1.323	1.148
0.5	3.014	2.820	2.562	2.280	2.002	1.745	1.516	1.317	1.146
0.6	2.873	2.715	2.488	2.230	1.971	1.726	1.505	1.311	1.144
0.7	2.748	2.620	2.421	2.185	1.942	1.708	1.495	1.306	1.142
0.8	2.636	2.535	2.360	2.144	1.915	1.692	1.486	1.302	1.140
0.9	2.535	2.458	2.304	2.106	1.890	1.677	1.477	1.297	1.138

Table A.1: Variance of individual informed demand

Appendix B: MD&A Informativeness Measure

Financial statements are an important vehicle for public dissemination of information but measuring their quality is difficult. In order to measure this quality, we examine firms' annual reports and focus on "Item 7: Management's Discussion and Analysis" (MD&A) section of those reports. In the MD&A section, managers discuss the firm's past performance as well as its future directions and challenges. Following Kogan, Routledge, Sagi, and Smith (2011), we use machine learning tools to understand what features of the MD&A section's text related to future, long-horizon, volatility. In particular, we capture the informativeness of each MD&A section by measuring the degree to which the text in the section can be used to forecast *out-of-sample* stock return volatility over the next year. We first estimate an econometric model that determines how various terms in the reports contribute to increases or decreases in firm volatility. We then take the estimated weights for each term and apply them out-of-sample to forecast annual volatility for each firm in the subsequent year based on its disclosure. The MD&A informativeness measure is constructed by subtracting the estimation error obtained from the MD&A text-based model to predict next year's volatility from the corresponding estimation error obtained from a predictive statistical model that uses the firm's past annual volatility. Thus, our measure of MD&A informativeness is higher when the text in the MD&A section is more useful for predicting future stock volatility beyond its past volatility.

The data used to construct this measure were obtained by extracting the MD&A section from all available annual reports in the 1996–2006. We then matched the observations for which this procedure resulted in a meaningful size report with CRSP return data. On a rolling basis, we used two years of data to estimate word weights and then applied these weights to forecast volatility for firms in the subsequent year. Thus, our procedure produced informativeness measure for firms in the 1998–2006 period as we exclude the first two years of our sample for training purposes. Because the algorithm we use to estimate MD&A informativeness requires that the MD&A section contains at least 1,000 words, we assign the minimum informativeness measure to missing observations, i.e., -0.6.

In this appendix, we compare this measure to other models. First, we examine their ability to predict future volatility. Table B.1 reports pairwise correlations among realized volatility ("realized"), the prediction from the text-based model above ("text_prediction"), the prediction from a model based on the count of risk-related words in the financial statement as proposed by Li (2006) ("count_prediction"), and the prediction from a model based on accruals as proposed by Dechow and Dichev (2002) ("accrual_prediction"). We find that the text-based model is more strongly correlated with subsequent realized volatility than the other two models. Moreover, while the count- and accruals-based models are strongly correlated, the text-based model is not very correlated with the other two models, suggesting that it captures a distinct aspect of the link between financial statements and firm volatility.

	Realized	Predicted		
		text	*count*	*accrual*
Realized Volatility	1			
Predicted Volatility, based on ...				
... MD&A statement (*text_prediction*)	0.80	1		
... count of risk-related words (*count_prediction*)	0.46	0.54	1	
... accruals (*accrual_prediction*)	0.48	0.59	0.91	1

Table B.1: Correlations of various volatility measures

Table B.2 reports estimates from panel regression models where the dependent variable is realized volatility. While the univariate regressions suggest that all three models have some power in predicting realized volatility, the text-based model has a much higher R^2 than the other models. Moreover, the coefficient on the text-based prediction remains high when the other predictions are included while the latter's coefficients significantly decline when the text-based prediction is included.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.:	Realized volatility					
text_prediction	0.756***			0.692***	0.728***	0.708***
	[0.008]			[0.010]	[0.010]	[0.011]
count_prediction		0.851***		0.179***		0.264***
		[0.011]		[0.016]		[0.032]
accruals_prediction			0.873***		0.118***	-0.133***
			[0.014]		[0.019]	[0.037]
Constant	-0.894***	-0.574***	-0.510***	-0.531***	-0.624***	-0.660***
	[0.026]	[0.039]	[0.047]	[0.041]	[0.050]	[0.050]
Observations	10,867	13,062	10,452	9,881	7,884	7,884
Number of permno	3,319	3,618	2,749	3,081	2,364	2,364
Pseudo R-squared	0.604	0.191	0.208	0.628	0.641	0.641

Table B.2: Multiple regressions of out-of-sample realized volatility

Lastly, we compare how realized volatility and measures of predicted volatilities vary with a number of firm and stock characteristics. Table B.3 reports estimates from panel regression models of realized volatility and each of the three predicted volatility measures on various characteristics. The estimates indicate that the text-based measure and the realized volatility have similar correlations with various characteristics, and more so relative to the other two models. In particular, the text-based measure pick up firm complexity (number of segments and segment concentration), corporate governance (G index), analyst dispersion, and illiquidity in a similar manner to the realized volatility, unlike the latter two measures.

In summary, these results suggests that while the alternative measures of disclosure quality are correlated with the text-based measure, the text-based approach performs much better in predicting future volatility than alternative measures. Moreover, it does so by capturing firm dimensions that cannot be easily captured by simple word counts or accruals.

Dep. Var.:	Realized	Predicted *text*	*count*	*accrual*
Mkt. Cap.	-0.000**	-0.000***	0	0
	[0.000]	[0.000]	[0.000]	[0.000]
H(Sales)	-0.158***	-0.153***	-0.063**	-0.054**
	[0.055]	[0.053]	[0.031]	[0.024]
N. Segs.	-0.069**	-0.072**	-0.006	-0.003
	[0.028]	[0.028]	[0.017]	[0.013]
Book/Market	-0.067***	0.004	0.030**	0.012
	[0.023]	[0.023]	[0.015]	[0.012]
$G-index$	-0.043***	-0.039***	-0.003	-0.004
	[0.007]	[0.007]	[0.004]	[0.003]
N. Analysts	0.001	0	0.004**	0.002
	[0.003]	[0.003]	[0.002]	[0.002]
Analyst Disp.	0.039*	0.049**	0.001	0.013
	[0.023]	[0.025]	[0.017]	[0.014]
Illiquidity	5.386***	5.397***	3.051***	2.433***
	[0.751]	[0.791]	[0.532]	[0.438]
Constant	-3.192***	-3.167***	-3.439***	-3.456***
	[0.108]	[0.104]	[0.061]	[0.047]
Observations	897	897	897	897
Number of permno	340	340	340	340
Pseudo R^2	0.173	0.16	0.039	0.0435

Table B.3: Multiple regressions of various volatility measures

Appendix C: Variable Definitions

Firm-level Distance Variables

- *Potential Investor Distance* (Miles) = Weighted-average of the distance (in miles) between the firm's HQ location and the population-weighted centroids of the states of residence of all 13-F institutional investors (including both shareholders and non-shareholders of the firm), with weights equal to the fraction of the aggregate 13-F equity portfolio located in each state. Same-state distances are set equal to 100 miles (this seemed reasonable since no cross-state distance is 100 miles or less). Each variable is measured at the end of each quarter using holdings reported in 13-F filings. For each firm-fiscal year, we average the relevant four fiscal quarters to obtain firm-year measures.

- *Potential Investor Log-Distance* = Weighted-average of the natural logarithm of potential investors distance above. The distance in miles is divided by 100 before taking the logarithm, so that same state log-distances are equal to 0.

- *Actual Shareholder Distance* (Miles) = Weighted-average of the distance (in miles) between the firm's HQ location and the population-weighted centroids of the firm's 13-F institutional shareholders' states of residence, with weights equal to the fraction of the firm's shares held by 13-F filers located in each state.

- *Actual Shareholder Log-Distance* = Weighted-average of the natural logarithm of actual shareholder distance above. The distance in miles is divided by 100 before taking the logarithm, so that same state log-distances are equal to 0.

- *Potential Investor PW HHI* = Proximity-weighted Herfindahl Index of state-level institutional *capital*, calculated for firm i in quarter t (suppressed for convenience) as follows:

$$Potential\ Investor\ PW\ HHI_i = \frac{\sum_{s=1}^{50} \frac{(State\ Aggr.\ Inst.\ Capital_s)^2}{d_{is}}}{[\sum_{s=1}^{50} (State\ Aggr.\ Inst.\ Capital_s)]^2}$$

  where $s = 1, , 50$, are the 50 US states. The log-distance, d_{is}, and state-level aggregate value of institutional investors' portfolios, $State\ Aggr.\ Inst.\ Capital$, are measured at the end of each quarter using holdings reported in 13-F filings. Note that the *State Aggr. Inst. Capital* does not have a firm i subscript as it is not firm-specific (unlike *State Aggr. Inst. Ownership* below).

  This index increases as state-level institutional holdings' concentration increases, holding constant the distance, OR as the firm HQ location gets closer to larger fractions of the institutional portfolio, holding constant the state-level institutional holdings' concentration.

- *Actual Shareholder PW HHI* = Proximity-weighted Herfindahl Index of state-level institutional *ownership*, calculated for firm i in quarter t (suppressed for convenience) as follows:

$$Actual\ Shareholder\ PW\ HHI_i = \frac{\sum_{s=1}^{50} \frac{(State\ Aggr.\ Inst.\ Ownership_{is})^2}{d_{is}}}{[\sum_{s=1}^{50}(State\ Aggr.\ Inst.\ Ownership_{is})]^2}$$

where $s = 1, , 50$, are the 50 US states. The log-distance, d_{is}, and state-level aggregate ownership of 13-F institutional investors, $State\ Aggr.\ Inst.\ Ownership$, are measured at the end of each quarter using holdings reported in 13-F filings.

Firm-State Level Variables

- $State\ Institutional\ Holdings_{i,s}$ = The fraction of firm i's institutional ownership controlled by institutional investors in state s.
- $State–Firm\ Distance_{i,s}$ = The distance between firm i's HQ state and state s, in 100 miles. Same-state distances are set equal to 100 miles (this seemed reasonable since no cross-state distance is 100 miles or less).

Firm-level Characteristics

- *Log*($Number\ of\ States$) = Natural logarithm of the number of U.S. states cited at least once in the relevant sections of the annual financial statement (see Bernile, Kumar, and Sulaeman (2015) for further details).
- *GEO Tercile* = Annual firm tercile rank (Low=0, High=2) of *Log*($Number\ of\ States$).
- *MDA Info* = MD&A Informativeness measure, which quantifies the usefulness of the textual content in each firm's annual 10-K filing's "Item 7: Managements Discussion and Analysis" (MD&A) section for predicting the firm's risk, as defined in Kogan, Routledge, Sagi, and Smith (2011).
- *MDA Info Tercile* = Annual firm tercile rank (Low=0, High=2) of $MDAInfo$.
- *Analyst Coverage* = Number of analysts covering the firm.
- *Analyst Coverage Tercile* = Annual firm tercile rank (Low=0, High=2) of *Analyst Coverage*.
- *Log (Market Cap.)* = Natural logarithm of equity market cap measured at close of prior fiscal year-end as stock price multiplied by the number of shares outstanding.
- *Leverage* = (Total Long Term Debt + Preferred Equity Liquidation Value)/(Market Value of Equity at Fiscal Year End + Total Long Term Debt + Preferred Equity Liquidation Value).
- *Age* = Number of years since a firm's first appearance in the CRSP database (i.e., since it becomes a publicly traded firm).

- *Young* = Indicator variable for firms with five years or less since their first appearance in the CRSP database.
- *B/M Ratio* = Ratio of book to market equity.
- *Price* = Stock price at the beginning of the institutional ownership measurement quarter.
- *Lottery-type* = Indicator variable for firms that are not in the bottom third of volatility, the bottom third of skewness, or the top third of price.
- *Turnover* = Trading volume divided by shares outstanding.
- *Effective Spread* = The difference between the executed price and the midpoint of the market quotes.

Firm-level Return Variables

- *Raw Return* = Raw monthly stock returns.
- *4-Factor Alpha (Carhart)* = Monthly abnormal stock return estimated using Carhart (1997) four-factor model.
- *Idio. Vol.* = Monthly stock returns volatility calculated from residuals of annual market-model regressions of monthly stock returns.
- *Idio. Skew.* = Monthly stock returns skewness calculated from residuals of annual market-model regressions of monthly stock returns.
- *Char.-Adjusted (DGTW) Return* = Characteristic-adjusted monthly abnormal stock return using the Daniel, Grinblatt, Titman, and Wermers (1997) method.
- *Lag (6-Month Return)* = Raw stock return in the six months leading up to the beginning of the institutional ownership measurement quarter.
- *Lag (1-Month Return)* = Raw stock return in the immediately preceding month.
- *Lag (11-Month Return)* = Raw stock return in the 11-month period leading up to the beginning of the immediately preceding month.

State-level Characteristics

- *Education* = Fraction of college graduates in the firm's HQ state.
- *Population Density* = Firm HQ state's population divided by its land area.
- *C/P Ratio* = Ratio of Catholic adherents to Protestant adherents in the firm's HQ state.
- *Religious Ratio* = Fraction of religious adherents in the firm's HQ state.

- *State Econ. Index* = State's macroeconomic indicator defined in Korniotis and Kumar (2010).
- *Republican* = Percentage of the state's registered voters that voted for the presidential candidate from the Republican party in the last election.

Table 1

Summary Statistics

This table reports summary statistics for the main variables used in the empirical analysis. The sample covers firm-fiscal years ending between 1997 and 2006, conditional on availability of a valid 10-K filing in the SEC's Edgar system and of financial statement and stock return data in the CRSP/Compustat merged database. Panel A reports the descriptive sample statistics for the geographic dispersion of institutional investors and shareholders relative to the firm's HQ location, as defined in detail in Appendix C. Panel B reports the statistics for investor-type classifications of Bushee (1998): Transient (TRA), Quasi-Indexer (QIX), and Dedicated (DED). Panel C reports the descriptive sample statistics for firm-level characteristics and firm/HQ State demographics.

	N	Mean	Median	1^{st} Pctl	25^{th} Pctl	75^{th} Pctl	99^{th} Pctl	Std Dev
Panel A: Ownership Geographic Dispersion								
Actual Shareholder Distance	25,207	1,074.9	937.1	176.2	673.2	1,380.1	2,295.6	537.6
Potential Investor Distance	25,207	1,116.5	842.2	596.4	679.9	1,353.5	2,096.7	546.9
Distance Difference	25,207	-41.6	-34.6	-946.5	-170.7	86.2	849.9	299
Actual Shareholder Log-Distance	25,207	2.867	2.936	1.28	2.35	3.421	4.031	0.667
Potential Investor Log-Distance	25,207	2.969	2.984	2.069	2.32	3.516	3.942	0.62
Log-Distance Difference	25,207	-0.102	-0.054	-1.491	-0.26	0.09	1.019	0.419
Actual Shareholder PWHHI	25,207	0.1311	0.0898	0.0329	0.0581	0.1477	0.7766	0.1331
Potential Investor PWHHI	25,207	0.0661	0.0492	0.0355	0.0431	0.0915	0.1283	0.031
PWHHI Difference	25,207	0.065	0.0254	-0.042	0.0044	0.0723	0.6873	0.126
Panel B: Ownership Geographic Dispersion, by Institutional Types								
TRA Act. Shareholder Log-Dist.	22,100	2.934	3.037	1	2.49	3.456	4.224	0.734
TRA Log-Distance Diff.	22,100	-0.11	-0.046	-1.656	-0.32	0.158	1.21	0.516
QIX Act. Shareholder Log-Dist.	22,655	2.864	2.914	1.238	2.329	3.43	4.095	0.688
QIX Log-Distance Diff.	22,655	-0.089	-0.037	-1.693	-0.259	0.116	1.257	0.478
DED Act. Shareholder Log-Dist.	20,269	2.888	3.087	1	2.128	3.681	4.28	1.006
DED Log-Distance Diff.	20,269	-0.125	-0.099	-2.093	-0.535	0.293	1.486	0.694
Panel C: Firm and HQ State Characteristics								
Log(Number of States)	25,207	2.27	2.3	0.69	1.79	2.71	3.83	0.68
MDA Info	25,207	-0.17	-0.03	-0.6	-0.6	0.02	0.48	0.3
Num Analysts	25,207	7.76	5.00	0.00	1.00	11.00	38.00	8.53
Exog. Analyst Loss	25,207	0.11	0	0	0	0	1	0.31
Log (Market Cap)	25,207	12.92	12.81	9.26	11.58	14.11	17.76	1.85
B/M Ratio	25,207	0.568	0.472	-0.367	0.273	0.732	2.714	0.512
Leverage	25,207	0.17	0.105	0	0.004	0.277	0.857	0.193
Age	25,207	14.7	9	0	4	20	74	15.3
Lag (6-Month Ret.)	25,207	10.45	5.36	-71.61	-12.22	25.28	191.61	42
Price	25,207	21.93	17.78	1.15	8.78	30.06	88.42	17.51
Turnover	25,207	1.34	0.86	0.05	0.43	1.69	7.56	1.41
Idio. Vol.	25,207	9.81	8.01	0	5.1	12.72	36.15	6.65
Idio. Skew.	25,207	0.14	0.14	-1.57	-0.33	0.6	1.91	0.71
Lottery-type	25,207	0.3153	0	0	0	1	1	0.4647
Education	25,207	25.5	26.1	16.9	22.4	27.4	33.2	3.74
Population Density	25,207	300.43	218	26.82	138.37	395.15	1,136.65	272.43
C/P Ratio	25,207	1.42	0.89	0.03	0.4	2.02	4.7	1.26
Religious Ratio	25,207	0.5	0.52	0.33	0.39	0.54	0.73	0.11
State Econ. Index	25,207	0.5	0.3	-0.91	-0.07	1.11	2.45	0.82
Republican	25,207	0.72	1	0	0	1	1	0.45

Table 2

Geographic Distance of Potential and Actual Institutional Equity Holdings

This table reports sample means of institutional investors and shareholders distance or log-distance from the firm's HQ location, as well as their difference. We report the unconditional averages (means and medians) in Panel A. In Panel B, firms are sorted (low to high) into terciles based on Geographic Dispersions of their Economic Interests (lognumstates) annually; while in Panel C, firms are sorted (low to high) into terciles based on MD&A Informativeness (MD&A Quality) annually. The sample covers firm-fiscal years ending between 1997 and 2006. Appendix C provides details about the computation of all variables. Asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, for the relevant sample statistics.

	Distance, in Miles		*Log(Distance/100)*	
	Mean	Median	Mean	Median
Panel A: All Firm-years				
Actual Shareholders	1074.9	937.1	2.867	2.936
Potential Investors	1116.5	842.2	2.969	2.984
Distance Difference	-41.6***	-34.6***	-0.102***	-0.054***
Panel B: Distance Difference, By Geographic Dispersion of Economic Interests				
Low Geographic Dispersion	-67.0***	-57.7***	-0.152***	-0.096***
Medium	-36.9***	-30.9***	-0.090***	-0.051***
High Geographic Dispersion	-18.8***	-22.8***	-0.060***	-0.033***
High − Medium	18.1***	8.1***	0.030***	0.018***
High − Low	48.1***	34.9***	0.092***	0.062***

Table 3

Determinants of Geographic Dispersion of Institutional Ownership

This table reports pooled cross-sectional regression coefficient estimates and corresponding t-statistics, in italics, for the relation between the geographic dispersion of institutional shareholdings relative to the firm's HQ location, conditional on the dispersion of all institutional investors' capital relative to the firm's HQ location, the geographic dispersion of the firm's economic interests, and the informativeness of the MD&A in the firm's 10-K. We include (but do not report the estimates of) the following control variables: time-varying firm characteristics (log market cap, leverage, age, b/m ratio, lagged 6-month returns, stock price, turnover, idiosyncratic volatility, idiosyncratic skewness, and lottery stock indicator), time-varying HQ state characteristics (population density, education level, religiosity, Catholic/Protestant ratio, state economic index, and political affiliation), time-invariant HQ state fixed effects, and year fixed effects. All variables are defined in Appendix C. Following Petersen (2009), the t-statistics reported in parentheses are adjusted for clustering by year and state. The sample period is from 1998 to 2006. Asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Actual Shareholder Log Distance	*Actual Shareholder Distance*	*Actual Shareholder PW HHI*
Potential Investor Dist.	0.7910***	0.7701***	1.3213***
	(36.55)	*(39.94)*	*(14.49)*
GEO Tercile	0.0331***	13.7031***	-0.0075***
	(6.64)	*(3.84)*	*(-5.85)*
MDA Info Tercile	0.0218***	10.6548***	-0.0078***
	(3.75)	*(2.73)*	*(-4.47)*
GEO Tercile * MDA Info Tercile	-0.0069**	-2.1737	0.0024***
	(-2.03)	*(-0.83)*	*(2.62)*
	Time-varying firm and HQ state characteristics are included, but their estimates are suppressed.		
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	25,207	25,207	25,207
R-square	65.00%	73.20%	26.70%

Table 4

Determinants of Geographic Dispersion of Institutional Ownership

This table reports pooled cross-sectional regression coefficient estimates and corresponding t-statistics, in italics, for the relation between the geographic dispersion of institutional shareholdings relative to the firm's HQ location, conditional on the dispersion of all institutional investors' capital relative to the firm's HQ location, the geographic dispersion of the firm's economic interests, and the informativeness of the MD&A in the firm's 10-K. We include (but do not report the estimates of) the following control variables: time-varying firm characteristics (log market cap, leverage, age, b/m ratio, lagged 6-month returns, stock price, turnover, idiosyncratic volatility, idiosyncratic skewness, and lottery stock indicator), time-varying HQ state characteristics (population density, education level, religiosity, Catholic/Protestant ratio, state economic index, and political affiliation), time-invariant HQ state fixed effects, and year fixed effects. All variables are defined in Appendix C. Following Petersen (2009), the t-statistics reported in parentheses are adjusted for clustering by year and state. The sample period is from 1998 to 2006. Asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Actual Shareholder Log Distance	*Actual Shareholder Distance*	*Actual Shareholder PW HHI*
Potential Investor Dist.	0.7898***	0.7695***	1.3091***
	29.84	*34.47*	*13.97*
GEO Tercile	0.0417***	21.6601***	-0.0194***
	3.85	*2.8*	*-6.21*
AN COV. Tercile	0.0239**	11.3306	-0.0322***
	2.48	*1.63*	*-12.20*
GEO Tercile * AN COV Tercile	-0.0156**	-10.1552**	0.0146***
	-2.30	*-2.10*	*7.46*
	Time-varying firm and HQ state characteristics are included, but their estimates are suppressed.		
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	25,207	25,207	25,207
R-square	65.00%	73.20%	27.60%

Table 5

Determinants of Institutional Ownership Levels and Changes

This table reports pooled tobit regression estimates and corresponding t-statistics, in italics, for the relation between firm-state-quarter fraction of institutional ownership or its absolute changes and: a) investors' proximity to the firm's HQ; b) the firm's overall geographic exposure (i.e., dispersion of economic interests); and c) the firm's disclosure quality (i.e., informativeness of MD&A); controlling for state and time fixed effects, and for time-varying firm characteristics. All variables are defined in Appendix C. The t-statistics reported in parentheses are adjusted for clustering at the firm-quarter level. The sample period is from 1998 to 2006. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% probability levels, respectively.

Dependent Variable:	Firm-State-Qtr IO Share					
	Level	|Change|	Level	|Change|	Level	|Change|
Exposure Measure (EXPO):	*HQ State Indicator*		*1/Log(HQ Dist)*		*Contact State Ind.*	
(1) Investors' Exposure	11.597***	3.665***	9.349***	3.134***	5.729***	1.747***
(EXPO)	*(19.91)*	*(24.33)*	*(21.84)*	*(22.79)*	*(22.84)*	*(21.15)*
(2) Firm Geo. Disp.	0.469***	0.303***	1.637***	0.775***	0.569***	0.365***
(GEO)	*(16.61)*	*(15.72)*	*(15.38)*	*(19.15)*	*(15.39)*	*(16.13)*
(3) Firm MD&A Info.	0.166***	0.126***	0.682***	0.222***	0.195***	0.129***
(MDA)	*(6.01)*	*(6.91)*	*(6.43)*	*(5.83)*	*(5.93)*	*(6.72)*
(4) GEO * MDA	-0.096***	-0.076***	-0.434***	-0.208***	-0.158***	-0.106***
	(-5.42)	*(-6.14)*	*(-6.38)*	*(-7.86)*	*(-6.87)*	*(-7.56)*
(5) EXPO * GEO	-4.745***	-1.706***	-3.562***	-1.419***	-2.789***	-0.951***
	(-13.60)	*(-18.21)*	*(-13.82)*	*(-15.48)*	*(-20.49)*	*(-19.57)*
(6) EXPO * MDA	-1.967***	-0.219**	-1.451***	-0.234***	-0.827***	-0.055
	(-5.54)	*(-2.18)*	*(-5.58)*	*(-2.65)*	*(-5.64)*	*(-1.01)*
(7) EXPO*GEO*MDA	1.205***	0.334***	0.960***	0.361***	0.564***	0.123***
	(5.38)	*(5.06)*	*(5.71)*	*(5.72)*	*(6.87)*	*(3.85)*
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Firm-Qtr	Firm-Qtr	Firm-Qtr	Firm-Qtr	Firm-Qtr	Firm-Qtr
Pseudo-R^2	48.70%	31.20%	48.50%	31.10%	48.50%	31.10%
N	4,978,050	4,978,050	4,978,050	4,978,050	4,978,050	4,978,050
(2)+(5)	-4.276***	-1.403***	-1.924***	-0.645***	-2.220***	-0.586**
p-value [(2)+(5)]	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
(3)+(6)	-1.800***	-0.093	-0.768***	-0.012	-0.631***	0.0738
p-value [(3)+(6)]	< 0.01	*0.248*	< 0.01	*0.832*	< 0.01	*0.157*

Table 6

Determinants of Institutional Ownership Levels and Changes

This table reports pooled tobit regression estimates and corresponding t-statistics, in italics, for the relation between firm-state-quarter fraction of institutional ownership or its absolute changes and: a) investors' proximity to the firm's HQ; b) the firm's overall geographic exposure (i.e., dispersion of economic interests); and c) the firm's disclosure quality (i.e., informativeness of MD&A); controlling for state and time fixed effects, and for time-varying firm characteristics. All variables are defined in Appendix C. The t-statistics reported in parentheses are adjusted for clustering at the firm-quarter level. The sample period is from 1998 to 2006. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% probability levels, respectively.

Dependent Variable:	Firm-State-Qtr IO Share					
	Level	|Change|	Level	|Change|	Level	|Change|
Exposure Measure (EXPO):	*HQ State Indicator*		*1/Log(HQ Dist)*		*Contact State Ind.*	
(1) Investors' Exposure	15.12***	5.96***	11.95***	4.91***	7.87***	3.22***
(EXPO)	(22.17)	(33.69)	(21.76)	(29.53)	(27.70)	(33.4)
(2) Firm Geo. Disp.	0.60***	0.34***	1.45***	0.60***	0.42***	0.24***
(GEO)	(15.01)	(13.59)	(9.39)	(12.21)	(8.83)	(9.46)
(3) Analyst Coverage	0.93***	0.48***	2.41***	1.22***	1.03***	0.54***
(ANCOV)	(23.77)	(18.20)	(17.03)	(24.10)	(23.54)	(19.76)
(4) GEO * ANCOV	-0.22***	-0.11***	-0.47***	-0.17***	-0.08***	-0.03*
	(-9.13)	(-6.95)	(-5.03)	(-5.21)	(-2.81)	(-1.85)
(5) EXPO * GEO	-4.03***	-1.41***	-2.66***	-0.85***	-2.87***	-1.08***
	(-8.59)	(-11.57)	(-7.11)	(-7.45)	(-17.74)	(-19.44)
(6) EXPO * ANCOV	-6.38***	-2.92***	-4.46***	-2.23***	-3.14***	-1.61***
	(-14.77)	(-25.12)	(-13.02)	(-20.22)	(-18.05)	(-25.59)
(7) EXPO*GEO*ANCOV	1.53***	0.54***	0.79***	0.21***	0.99***	0.45***
	(5.27)	(6.90)	(3.40)	(2.71)	(9.94)	(12.30)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Firm-Qtr	Firm-Qtr	Firm-Qtr	Firm-Qtr	Firm-Qtr	Firm-Qtr
Pseudo-R^2	48.90%	31.40%	48.70%	31.30%	48.70%	31.30%
N	4,978,050	4,978,050	4,978,050	4,978,050	4,978,050	4,978,050
(2)+(5)	-3.43	-1.10	-1.20	-0.30	-2.50	-0.84
p-value [(2)+(5)]	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
(3)+(6)	-5.45	-2.44	-2.05	-1.01	-2.11	-1.07
p-value [(3)+(6)]	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 7

Determinants of Institutional Ownership Levels and Changes, post Regulation

This table reports pooled tobit regression estimates and corresponding t-statistics, in italics, for the relation between firm-state-quarter fraction of institutional ownership or its absolute changes and: a) investors' proximity to the firm's HQ; b) the firm's overall geographic exposure (i.e., dispersion of economic interests); and c) the improvement in firm's disclosure quality as proxied by the adoptions of regulations in early 2000's; controlling for state and time fixed effects, and for time-varying firm characteristics. All variables are defined in Appendix C. The t-statistics reported in parentheses are adjusted for clustering at the firm-quarter level. The sample period is from 1998 to 2006. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% probability levels, respectively.

Dependent Variable:	Firm-State-Qtr IO Share					
	Level	|Change|	Level	|Change|	Level	|Change|
Exposure Measure (EXPO):	*HQ State Indicator*		*1/Log(HQ Dist)*		*Contact State Ind.*	
(1) Investors' Exposure (EXPO)	10.84***	4.22***	8.72***	3.42***	5.60***	2.19***
	(23.30)	(29.45)	(24.28)	(27.28)	(29.07)	(30.03)
(2) Firm Geographic Disp. (GEO)	0.45***	0.27***	1.35***	0.65***	0.48***	0.29***
	(18.19)	(15.90)	(15.03)	(18.27)	(16.33)	(15.97)
(3) Post MD&A Regulation (POSTREG)	1.07***	0.15***	1.75***	0.56***	1.10***	0.20***
	(17.02)	(2.82)	(12.07)	(7.95)	(16.62)	(3.65)
(4) GEO*POSTREG	-2.89***	-1.83***	-1.99***	-1.26***	-1.60***	-1.11***
	(-6.68)	(-12.98)	(-6.07)	(-9.99)	(-9.06)	(-15.16)
(5) EXPO*GEO	-0.18***	-0.09***	-0.31***	-0.19***	-0.14***	-0.07***
	(-7.15)	(-4.59)	(-3.52)	(-4.98)	(-4.80)	(-3.38)
(6) EXPO*POSTREG	-3.77***	-1.59***	-2.78***	-1.20***	-2.50***	-1.03***
	(-13.34)	(-17.61)	(-12.67)	(-14.38)	(-23.92)	(-24.11)
(7) EXPO*GEO*POSTREG	0.51*	0.46***	0.39*	0.30***	0.60***	0.43***
	(1.79)	(4.94)	(1.75)	(3.41)	(5.97)	(9.83)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Firm-Qtr					
Pseudo-R^2	48.70%	31.20%	48.50%	31.10%	48.50%	31.10%
N	4,978,050					

Table 8

Determinants of Institutional Ownership Levels and Changes, Difference in Difference around Exogenous Reduction of Analyst Coverage

This table reports pooled tobit regression estimates and corresponding t-statistics, in italics, for the relation between firm-state-quarter fraction of institutional ownership or its absolute changes and: a) investors' proximity to the firm's HQ; b) the firm's overall geographic exposure (i.e., dispersion of economic interests); and c) the decline in firm's information environment as proxied by analyst coverage loss due to brokerage houses' merger and acquisition activities; controlling for state and time fixed effects, and for time-varying firm characteristics. All variables are defined in Appendix C. The t-statistics reported in parentheses are adjusted for clustering at the firm-quarter level. The sample period is from 1998 to 2006. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% probability levels, respectively.

Dependent Variable:	Firm-State-Qtr IO Share					
	Level	|Change|	Level	|Change|	Level	|Change|
Exposure Measure (EXPO):	*HQ State Indicator*		*1/Log(HQ Dist)*		*Contact State Ind.*	
(1) Investors' Exposure	15.36***	2.88***	15.56***	2.86***	14.16***	2.80***
(EXPO)	(15.08)	(10.73)	(17.79)	(10.63)	(23.03)	(17.18)
(2) Firm Geographic Disp.	0.10***	0.16***	0.84***	0.22***	0.22***	0.08*
(GEO)	(3.76)	(3.19)	(3.65)	(3.30)	(4.55)	(1.79)
(3) EXPO*GEO	-2.52***	-0.40*	-2.66***	-0.30*	-5.08***	-1.00***
	(-3.64)	(-1.86)	(-4.02)	(-1.79)	(-14.65)	(-10.50)
(4) PostQtr	-0.01	-0.04	-0.08***	-0.06	0.00	-0.04
(POST)	(-0.37)	(-0.80)	(-3.21)	(-1.04)	(0.16)	(-0.74)
(5) Analyst Loss	-0.01	-0.01	-0.02	-0.01	-0.04	-0.02
(LOSS)	(-0.42)	(-0.21)	(-1.01)	(-0.36)	(-1.40)	(-0.43)
(6) POST*LOSS	-0.20*	-0.10*	-0.24*	-0.10*	-0.18*	-0.10*
	(-1.81)	(-1.94)	(-1.85)	(-1.74)	(-1.67)	(-1.73)
(7) POST*LOSS*GEO	0.04	0.08	0.10*	0.10*	0.18*	0.10
	(0.99)	(1.52)	(1.72)	(1.67)	(1.67)	(1.61)
(8) POST*LOSS*EXPO	0.42*	0.54*	0.64*	0.42*	0.76**	0.24*
	(1.68)	(1.76)	(1.76)	(1.77)	(2.25)	(1.75)
(9) POST*LOSS*GEO*EXPO	-0.24	-0.62*	-0.32*	-0.56*	-0.28*	-0.14*
	(1.53)	(-1.81)	(-1.68)	(-1.81)	(-1.81)	(-1.81)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Firm-Qtr					
Pseudo-R^2	3.34%	1.37%	3.32%	1.21%	5.45%	1.95%
N	190,100					

Table 9

Monthly Returns of Portfolios formed Conditional on Potential Investors Distance from the HQ Location

This table reports average monthly raw and excess returns of stocks sorted into terciles of Potential Investor Distance as of the prior fiscal year-end date - Potential Investor Short/Medium/Long Distance. Detailed definitions of Distance Terciles, Raw Return, 4-Factor Alpha (Carhart), and Char.-Adjusted Return (DGTW) are provided in Appendix C. Long minus Short is the mean difference between monthly returns of firms in the Long and Short Potential Investor Distance Portfolios. Panel A reports equal-weighted portfolio returns, Panel B reports value-weighted portfolio returns with weights equal to firms' equity market cap at the beginning of the month. t-statistics, reported in parentheses, are adjusted using Newey and West (1987) correction for heteroskedasticity and serial correlation. The sample period is from 1998 to 2006. Asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Potential Investor Distance	Raw Return	4-Factor Alpha (Carhart)	Char.-Adjusted (DGTW)
	Panel A: Equal-Weighted		
Short	1.0615%	0.0314%	0.1024%
Medium	1.0029%	-0.0354%	0.0663%
Long	1.4618%	0.3105%	0.4762%
Long minus Short	0.4003%**	0.2791%*	0.3738%**
	(2.26)	*(1.72)*	*(2.74)*
	Panel B: Value-Weighted		
Short	0.2459%	-0.2360%	-0.1189%
Medium	0.5174%	-0.0250%	0.0158%
Long	0.9690%	0.4452%	0.2574%
Long minus Short	0.7230%***	0.6810%***	0.3763%**
	(2.75)	*(3.11)*	*(2.53)*

Table 10
Monthly Returns Regression Analysis – MD&A Informativeness

This table reports regression coefficient estimates and corresponding t-statistic, in italics, for the relation between monthly stock returns and the geographic dispersion of all institutional investors relative to the firm's HQ location, conditional on the firm's geographic dispersion of economic interests and informativeness of MD&A as contained in the last available 10-K. The dependent variable is the characteristic-adjusted value-weighted return computed using the Daniel, Grinblatt, Titman, and Wermers (1997) method and expressed in percentages. All variables in the table are defined in Appendix C. The table reports parameter estimates from monthly Fama and MacBeth (1973) style regressions. The t-statistics reported are adjusted using the Newey and West (1987) correction for heteroskedasticity and serial correlation. The sample period is from 1998 to 2006. Asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Log(Distance)			Distance Tercile
Potential Inv. Dist. (PIDist)	0.267***	0.572***	0.685***	0.411***
	(3.69)	(3.68)	(3.67)	(3.35)
MD&A Info. Tercile (MDA)	-0.078	0.148*	0.31*	0.117
	(-0.9)	(1.93)	(1.73)	(1.22)
GEO Tercile (GEO)	-0.216**	0.028	0.208	0.01
	(-2.24)	(0.21)	(0.77)	(0.05)
MDA * GEO			-0.195	-0.107
			(-1.18)	(-0.93)
PIDist * MDA		-0.246*	-0.368**	-0.224*
		(-1.97)	(-2.10)	(-1.68)
PIDist * GEO		-0.194**	-0.337**	-0.217**
		(-2.36)	(-2.38)	(-2.25)
PIDist * MDA * GEO			0.148	0.185
			(1.44)	(1.61)
Log (Market Cap)	-0.086	-0.097	-0.099	-0.085
	(-0.86)	(-0.97)	(-0.99)	(-0.85)
Log (B/M)	-0.113	-0.091	-0.076	-0.105
	(-0.81)	(-0.65)	(-0.52)	(-0.72)
Lag (11-Month Ret.)	0.21	0.208	0.199	0.203
	(0.68)	(0.67)	(0.65)	(0.66)
Lag (1-Month Ret.)	-3.174***	-3.189***	-3.208***	-3.188***
	(-3.27)	(-3.29)	(-3.31)	(-3.29)
Effective Spread	-0.182	-0.175	-0.174	-0.177
	(-1.25)	(-1.20)	(-1.20)	(-1.21)
Idio. Vol.	0.017	0.016	0.016	0.017
	(0.58)	(0.54)	(0.54)	(0.59)
Idio. Skew.	0.015	0.013	0.016	0.018
	(0.25)	(0.22)	(0.28)	(0.31)
Avg N	2625.54	2625.54	2625.54	2625.54
Avg R^2	0.142	0.143	0.145	0.145

Table 11
Monthly Returns Regression Analysis – Analyst Coverage

This table reports regression coefficient estimates and corresponding t-statistic, in italics, for the relation between monthly stock returns and the geographic dispersion of all institutional investors relative to the firm's HQ location, conditional on the firm's geographic dispersion of economic interests and analyst coverage. The dependent variable is the characteristic-adjusted value-weighted return computed using the Daniel, Grinblatt, Titman, and Wermers (1997) method and expressed in percentages. All variables in the table are defined in Appendix C. The table reports parameter estimates from monthly Fama and MacBeth (1973) style regressions. The t-statistics reported are adjusted using the Newey and West (1987) correction for heteroskedasticity and serial correlation. The sample period is from 1998 to 2006. Asterisks ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Log(Distance)			Distance Tercile
Potential Inv. Dist. (PIDist)	0.266***	0.468***	0.600***	0.354***
	(3.62)	(2.94)	(3.09)	(2.64)
Analyst Coverage Tercile (ANCOV)	-0.041	0.021	0.367	0.102
	(-0.43)	(0.28)	(1.48)	(0.85)
GEO Tercile (GEO)	-0.22**	0.03	0.357	0.123
	(-2.26)	(0.25)	(1.6)	(0.8)
ANCOV * GEO			-0.342**	-0.214**
			(-2.38)	(-2.06)
PIDist * ANCOV		-0.092	-0.208*	-0.187*
		(-1.36)	(-1.92)	(-1.83)
PIDist * GEO		-0.202**	-0.392***	-0.246***
		(-2.58)	(-2.85)	(-2.66)
PIDist * ANCOV * GEO			0.195**	0.109*
			(2.1)	(1.76)
Log (Market Cap)	-0.093	-0.103	-0.103	-0.089
	(-0.95)	(-1.05)	(-1.06)	(-0.91)
Log (B/M)	-0.108	-0.099	-0.081	-0.103
	(-0.73)	(-0.67)	(-0.55)	(-0.7)
Lag (11-Month Ret.)	0.218	0.221	0.229	0.231
	(0.74)	(0.75)	(0.79)	(0.79)
Lag (1-Month Ret.)	-3.181***	-3.186***	-3.19***	-3.171***
	(-3.28)	(-3.29)	(-3.29)	(-3.28)
Effective Spread	-0.218	-0.21	-0.194	-0.201
	(-1.36)	(-1.3)	(-1.2)	(-1.24)
Idio. Vol.	0.016	0.015	0.013	0.014
	(0.56)	(0.52)	(0.45)	(0.5)
Idio. Skew.	0.029	0.029	0.03	0.032
	(0.53)	(0.52)	(0.54)	(0.58)
Avg N	2625.54	2625.54	2625.54	2625.54
Avg R^2	0.144	0.145	0.146	0.146