Journal of Financial Economics 106 (2012) 547–565

Contents lists available at SciVerse ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Geographic dispersion and stock returns ☆

Diego García [a,*], Øyvind Norli [b]

[a] University of North Carolina at Chapel Hill, Chapel Hill, 4409 McColl, CB#3490, NC 27599-3490, United States
[b] BI Norwegian Business School, Norway

ARTICLE INFO

Article history:
Received 3 September 2010
Received in revised form
26 September 2011
Accepted 5 December 2011
Available online 28 June 2012

JEL classification:
G11
G12
G14

Keywords:
Geography
Geographic dispersion
Location
Local
Stock returns

ABSTRACT

This paper shows that stocks of truly local firms have returns that exceed the return on stocks of geographically dispersed firms by 70 basis points per month. By extracting state name counts from annual reports filed with the Securities and Exchange Commission (SEC) on Form 10-K, we distinguish firms with business operations in only a few states from firms with operations in multiple states. Our findings are consistent with the view that lower investor recognition for local firms results in higher stock returns to compensate investors for insufficient diversification.

Published by Elsevier B.V.

1. Introduction

It is well documented that both professional investment managers and individual investors display a strong preference for investing in local firms.[1] This finding is unexpected from the point of view of standard portfolio theory, and it has spurred a large literature on the causes and consequences of this local bias. Somewhat surprisingly, the asset pricing implications of the local bias have received relatively little attention.[2] A possible reason for this omission is that the existing literature defines an investment as local if the investor is located geographically close to the firm's headquarters. According to this definition, all firms are local to some investors, and there is no cross-sectional variation in the degree of "localness" among firms. This paper constructs a novel measure that allows us to characterize firms, rather than investments, as local. By distinguishing between truly local firms and firms that are geographically dispersed, we are able to shed light on the asset pricing implications of the local bias.

We define a firm as local if its business activities are concentrated in a small geographic area. To measure the degree of geographic concentration, we extract state name counts from annual reports filed with the SEC on Form 10-K. Based on the state name counts, firms are classified as geographically dispersed if a large number of

☆ We thank Øyvind Bøhren, Jacob Boudoukh, Lauren Cohen, Joey Engelberg, Michael Halling, Tim Loughran, Richard Priestley, Merih Sevilir, Harry J. Turtle and seminar participants at the 1st Tel-Aviv Finance Conference, SFS Cavalcade, UNC at Chapel Hill, Norwegian Business School, Arison School of Business at IDC, Northern Finance Association meetings, and American Finance Association meetings for helpful comments. All remaining errors are our own.

* Corresponding author. Tel.: +1 919 962 8404; fax: +1 919 962 2068.
E-mail addresses: diego_garcia@unc.edu (D. García), oyvind.norli@bi.no (Ø. Norli).

[1] Coval and Moskowitz (1999) were the first to show the presence of a local bias.

[2] Exceptions are Pirinsky and Wang (2006), Hong, Kubik, and Stein (2008), and Gómez, Priestley, and Zapatero (2008).

0304-405X/$ - see front matter Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.jfineco.2012.06.007

states are mentioned in the annual report, and as local if only one or two states are mentioned. Using a large sample of U.S. publicly listed firms from the period 1994 through 2008, we show that the stock returns of truly local firms far exceed the stock returns of geographically dispersed firms.

To study the relation between stock returns and the degree of geographic dispersion, we use both portfolio sorts and Fama–MacBeth cross-sectional regressions. Firms are sorted into portfolios of local firms and geographically dispersed firms using our state count measure. The portfolio of local firms has a Jensen's alpha of 48 basis points per month relative to a factor model that controls for risks related to the market, size, book-to-market ratio, momentum, and liquidity. The portfolio of geographically dispersed firms has a corresponding alpha of -22 basis points. This implies a 70 basis point difference in monthly risk-adjusted returns between local firms and geographically dispersed firms. On an annual basis, this corresponds to a return difference of 8.4%. Using Fama–MacBeth cross-sectional regressions, we find an effect of geographic dispersion that is similar both in terms of economic magnitude and statistical significance. The variation in average returns associated with firms' geographical dispersion is particularly pronounced for smaller firms, less liquid firms, and firms with high idiosyncratic volatility. But the effect of geographic dispersion is not subsumed by these firm characteristics.

Our paper is closely related to a large and rapidly growing literature on how economic decision making is influenced by firms' geographic location. Coval and Moskowitz (1999) show that U.S. money managers are significantly more likely to invest in firms headquartered in the same city as the manager than in other firms. Numerous subsequent studies have confirmed the strong preference for local firms and have suggested explanations that include informational advantages, familiarity, and social interactions.[3] A more recent branch of the literature looks at the effects of geography from the firm's perspective and finds that geographic location also matters for corporate decision making.[4]

Given the strong evidence in favor of a relation between geographic location and both investor decisions and corporate decisions, it seems natural to investigate how geography affects asset prices. Pirinsky and Wang (2006) show how the stock returns of firms headquartered in the same geographic area strongly co-move with each other, and interpret their evidence as favoring the view that the trading pattern of local investors influences stock returns. Hong, Kubik, and Stein (2008) show that the local bias depresses the stock price through an "only game in town" effect.[5] Our paper contributes to this literature by providing evidence on the existence of a link between the geographical scope of a firm and its average stock returns.

Geographic dispersion has been shown to be important for a number of questions in economics.[6] However, we are the first to create a proxy for the geographical dispersion of a firm's operations that is available for virtually the whole cross-section of publicly traded U.S. firms. Most other studies base their measures of dispersion on international data, small proprietary databases, or on information reported in Exhibit 21 of the 10-K statements, where firms break down financial variables by business segments (which sometimes are geographic segments). Although these sources provide data with less noise than our state counts, it can only be collected for a small subsection of listed U.S. corporations. Moreover, local firms are unlikely to be included in these data sets, precisely because they are local.

Theoretically, there are good reasons to expect the local bias to have implications for asset prices. Merton (1987) characterizes equilibrium stock returns when investors are not aware of all securities. Stocks with lower investor recognition have higher expected returns to compensate investors that hold the stock for insufficient diversification. It is reasonable to expect that stocks of local firms will have a smaller investor base, and hence lower investor recognition, than stocks of geographically dispersed firms. It follows that local firms should have higher stock returns than geographically dispersed firms, consistent with our main finding.[7]

More recent theories have tried to explain the anomalies related to geographic location through an informational channel. VanNieuwerburgh and Veldkamp (2009) develop a model where a slight informational advantage on "local assets" makes agents buy more information on those assets and over-weight them in their portfolios. García and Strobl (2011) show that relative wealth concern generates herding in informational choices, and as a consequence, in holdings. These models have different

[3] Coval and Moskowitz (2001), Hau (2001), Ivković and Weisbenner (2005), Ivković, Sialm, and Weisbenner (2008), and Teo (2009) conclude that local investors have an informational advantage. However, see Seasholes and Zhu (2010) for contradicting evidence. Huberman (2001) show that people tend to invest in the familiar. Social interaction among investors is shown to be important for investment decisions by Grinblatt and Keloharju (2001), Hong, Kubik, and Stein (2004, 2005), Ivković and Weisbenner (2007), and Brown, Ivković, Weisbenner, and Smith (2008).

[4] See Kang and Kim (2008), Landier, Nair, and Wulf (2009), Becker, Ivković, and Weisbenner (2011), and Almazan, Motta, Titman, and Uysal (2010).

[5] Gómez, Priestley, and Zapatero (2008) show that a local risk factor has negative risk premium. This evidence is consistent with investors hedging local risk from relative wealth concerns. See Feng and Seasholes (2004), Loughran and Schultz (2005), Loughran (2007), Dorn, Huberman, and Sengmueller (2008), and Bodnaruk (2009) for other related work.

[6] Landier, Nair, and Wulf (2009) show that the geographic dispersion of a firm affects its decision making. Gao, Ng, and Wang (2008) show that geographic dispersion affects firm value. There is also a large literature in economics that studies why Silicon Valley-style geographic agglomeration exists. See, for example, Ellison and Glaeser (1997) and references therein. The international finance literature is also related; see, for example, Doukas and Travlos (1988) and Uysal, Kedia, and Panchapagesan (2008) for studies of mergers and acquisitions in an international context.

[7] As investor recognition is not directly observable, the existing empirical literature has used proxies that include cross listings by non-U.S. firms (Foerster and Karolyi, 1999), trading volume (Gervais, Kaniel, and Mingelgrin, 2001; Kaniel, Ozoguz, and Starks, 2012), media attention (Fang and Peress, 2009), and a measure of the shadow cost of incomplete information (Bodnaruk and Östberg, 2008).

implications for unconditional stock returns than the model of Merton (1987). In particular, models that generate excess information acquisition will typically generate more informative prices, which lowers the equilibrium ex ante equity premium. Thus, our main empirical finding supports the mechanism in Merton (1987) rather than an informational channel.[8]

To further explore predictions of the investor recognition hypothesis, we investigate how returns on stocks of local firms are related to the imbalance between the amount of capital available locally and local investment opportunities. Following the evidence in Hong, Kubik, and Stein (2008), we conjecture that investors will be aware of most firms around them in areas where the amount of investable capital is large relative to the size of local investment opportunities. On the other hand, in areas where the opposite is true, local firms will have a hard time showing up on investors' radar screens, and stock returns should reflect this.

We measure the capital imbalance in two different ways. First, we investigate how returns on stocks of local firms are related to the number of listed firms per capita in the state where the firms are headquartered. We find that local firms from states with a low firm population density generate returns that are significantly lower than the return on local firms from states with high firm population density. Controlling for potential differences in risk between firms from high density states and firms from low density states, the return on a portfolio of local firms from high density states exceeds the return on an equally weighted portfolio of local firms from low density states by 58 basis points. Second, we measure the capital imbalance using the difference between mutual fund capital and listed firm market capitalization in a 100 km diameter circle around the headquarters of the local firm. With this measure, the return on an equally weighted portfolio long in local stocks with low recognition and short in local stocks with high recognition is 31 basis points. For both approaches, the point estimates for value-weighted portfolios are of a similar magnitude but not statistically significant.

In a final test of the investor recognition hypothesis, we look at changes in geographic dispersion. In particular, we study firms that go from being local – and unrecognized by investors – to geographically dispersed and recognized.[9] We find that the realized return on stocks that become local is no different than the return on stocks that were already local. A similar statement applies to stocks that become geographically dispersed. In other words, firms changing their geographic dispersion behave more like the firms they become similar to than the firms they used to be, consistent with investor recognition being priced into asset prices within a year.

Overall, we present several findings that are consistent with the investor recognition hypothesis. However, the size of the difference in monthly risk-adjusted returns between local firms and geographically dispersed firms leads us to conclude that investor recognition most likely is not the only explanation for our findings.

The rest of the paper is organized as follows. Section 2 describes our data selection procedure and explains how we construct our measure of geographic dispersion. Section 3 presents the main findings. In Section 4 we provide possible explanations for the high returns on local firms as well as robustness tests. Section 5 concludes the paper.

2. Data

We use a sample of firms listed on the New York Stock Exchange (NYSE), the American Stock Exchange (Amex), and Nasdaq. The data used to construct our measure of geographic dispersion are downloaded from the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) of the U.S. Securities and Exchange Commission (SEC). Stock returns, stock prices, and data on volume traded are from the Center for Research in Security Prices (CRSP). Accounting variables are from Compustat. The following sections describe our data selection procedure, explain how we construct our measure of geographic dispersion, and report summary statistics on both geographic dispersion and sample firms.

2.1. Geographic dispersion

The degree of geographic dispersion of a firm's business operations is measured using data from 10-K filings. Form 10-K is an annual report required by the SEC that gives a comprehensive summary of a public company's performance and operations. Firms must file such a report with the SEC within 90 days of the end of their fiscal year. In addition to financial data, the annual report typically includes information on the evolution of the firm's operations during that year, details on its organizational structure, executive compensation, competition, and regulatory issues. The 10-K statement also gives information on the firm's properties, such as factories, warehouses, and sales offices. For example, firms may include sales at stores in different states, and/or list the manufacturing facilities they operate together with the city and state where they are located.

Computerized parsing of all 10-Ks filed with the SEC during the period 1994–2008 allows a count of the number of times each 10-K mentions a U.S. state name. The structure of a 10-K filing is standardized, and the vast majority of 10-Ks are subdivided into the same set of sections. We count the occurrence of state names in sections "Item 1: Business," "Item 2: Properties," "Item 6: Consolidated Financial Data," and "Item 7: Management's Discussion and Analysis." In most of the analysis that follows, we simply measure geographic dispersion as the number of different states mentioned in these four sections. Firms that do not mention any U.S. state names in their 10-K are excluded from the analysis. Thus,

[8] García and Strobl (2011) explicitly show how the equilibrium risk premium of an asset varies with the intensity of relative wealth concerns. Only when agents strongly herd on their information acquisition choices does the model predict higher expected returns for local assets (as information does not aggregate via prices). VanNieuwerburgh and Veldkamp (2009) do not study unconditional expected returns explicitly.

[9] We thank the referee for making this suggestion.

Table 1
Summary statistics on geographic dispersion.

Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed by U.S. publicly listed firms on Form 10-K with the SEC. Geographic dispersion for year t is the count from the last annual report filed prior to July of year t. Using the column labeled "Med" (for median) as an example, summary statistics in Panel A are computed as follows. First, the median is computed for each July cross-section in the sample period 1994–2008. This gives a time series of annual medians. Second, using the time series of medians, the rows in Panel A report the average, the median, the minimum, and the maximum. Panel B breaks down the 4,509 observations from the first row in Panel A by market capitalization (firm size). The sample period is 1994–2008.

		Geographic dispersion								
	Number of firms	Mean	Std.	Min	Max	20%	40%	Med	60%	80%
Panel A: Summary statistics on geographic dispersion for all firms										
Average	4,509	7.9	7.7	1	50	2.6	4.3	5.5	6.8	11.3
Median	4,557	7.8	7.7	1	50	3	4	5	7	11
Minimum	934	7.1	6.9	1	50	2	4	5	6	10
Maximum	6,293	9.6	8.4	1	50	3	6	8	9	14
Panel B: Average geographic dispersion by firm size										
Small	1,503	5.9	5.1	1	48	2.1	3.2	4.3	5.3	8.5
Medium	1,503	7.3	6.9	1	49	2.5	4.4	5.4	6.5	10.5
Big	1,503	10.5	9.4	1	50	3.5	6.0	7.6	9.3	15.3

geographic dispersion for firm i based on the 10-K for fiscal year t is an integer in $\{1,2,\ldots,50\}$.

The vast majority of firms file their annual report using SEC Form 10-K. If a firm has not filed the 10-K within a fiscal year, or we cannot identify the right sections, we check if the firm has made an amended filing on form 10-K/A, and we use this filing to count states. If neither of these two forms are filed or contain the sections we are interested in, we repeat the procedure using Forms 10-K405, 10-KSB, 10-KT, 10KSB, 10KSB40, 10KT405, and the amendments to these forms.[10] We only count states in one form in a given fiscal year. Overall, we read and attempt to count states in 118,242 forms.

Firms that file with the SEC using EDGAR are uniquely identified by the Central Index Key (CIK). The CIK is matched with data from CRSP and Compustat using the linkfile from the CRSP/Compustat Merged Database. We are able to match 91,460 forms with data from CRSP using this linkfile. Restricting firms to be listed on NYSE, Amex, or Nasdaq with common equity (sharecodes 10 and 11) and only counting firms with a December return on CRSP, leaves us with a sample of 66,628 firm-years for the sample period 1994–2008. The number of firms that satisfies the above sampling criteria fluctuates between a low of 934 in 1994 (when EDGAR filings were optional) to a high of 6,293 in 1998.

The state names most frequently mentioned in the 10-Ks are: California, Texas, New York, Florida, and Illinois (in that order). The least common state names are Rhode Island, South Dakota, and North Dakota. Delaware and Washington, particularly the former, are outliers in terms of number of counts per population of the state, as many companies are incorporated in Delaware, and Washington is also the name of the United States' capital. We present our results using the counts of the states without any adjustments, but we remark that all of our results are robust to the exclusion of Delaware and Washington as state names.

A prime example of a firm that is clearly geographically dispersed is Sears Holdings Corporation. It has a state count of 50 for all years in our sample period. In its 10-Ks, Sears always breaks down the number of Sears and Kmart stores by state. Other firms that, by our measure, operate in all 50 states are Darden restaurants, the world's largest company owned and operated restaurant company, GameStop Corp., a videogame retailer, and Genworth Financial, Inc., a large retail financial firm. Well-known firms with an average state count that exceeds 45 include: Barnes & Noble, Applebee's, OfficeMax, Zurich Reinsurance, Jo-Ann stores, United Rentals, Regions Financial Corp., and Integrated Health Services.

2.2. Summary statistics on geographic dispersion

Table 1 presents sample summary statistics for our measure of geographic dispersion. These results have interest on their own, as they are the first large sample evidence on the geographical scope of U.S. publicly traded firms. Panel A presents summary statistics for all firms in the sample. Focusing on the first row, the average number of U.S. state names mentioned in the annual report filed on form 10-K is 7.9. This average is computed using the time series of July cross-sectional averages. In this 1994–2008 time series, the minimum average is 7.1 states and the maximum average is 9.6 states. Based on average state counts, geographic dispersion seems to be stable over our sample period. The stability is confirmed by the graph in Panel A of Fig. 1. This graph shows the monthly cross-sectional average geographic dispersion starting in May 1994 and ending in December 2008. At the start of the sample period, the average number of states is relatively high. This reflects the fact that prior to May 1996, filing via the EDGAR system was voluntary, and the firms that chose electronic filing were mostly large firms.

[10] These forms are essentially 10-K statements for either (i) small firms, who are not required to give as many details as large firms (forms ending in SB), (ii) firms that, prior to 2003, had failed to file a Form 4, regarding insider trading activities, in time (forms ending in 405), or (iii) firms in transition (forms ending in KT).

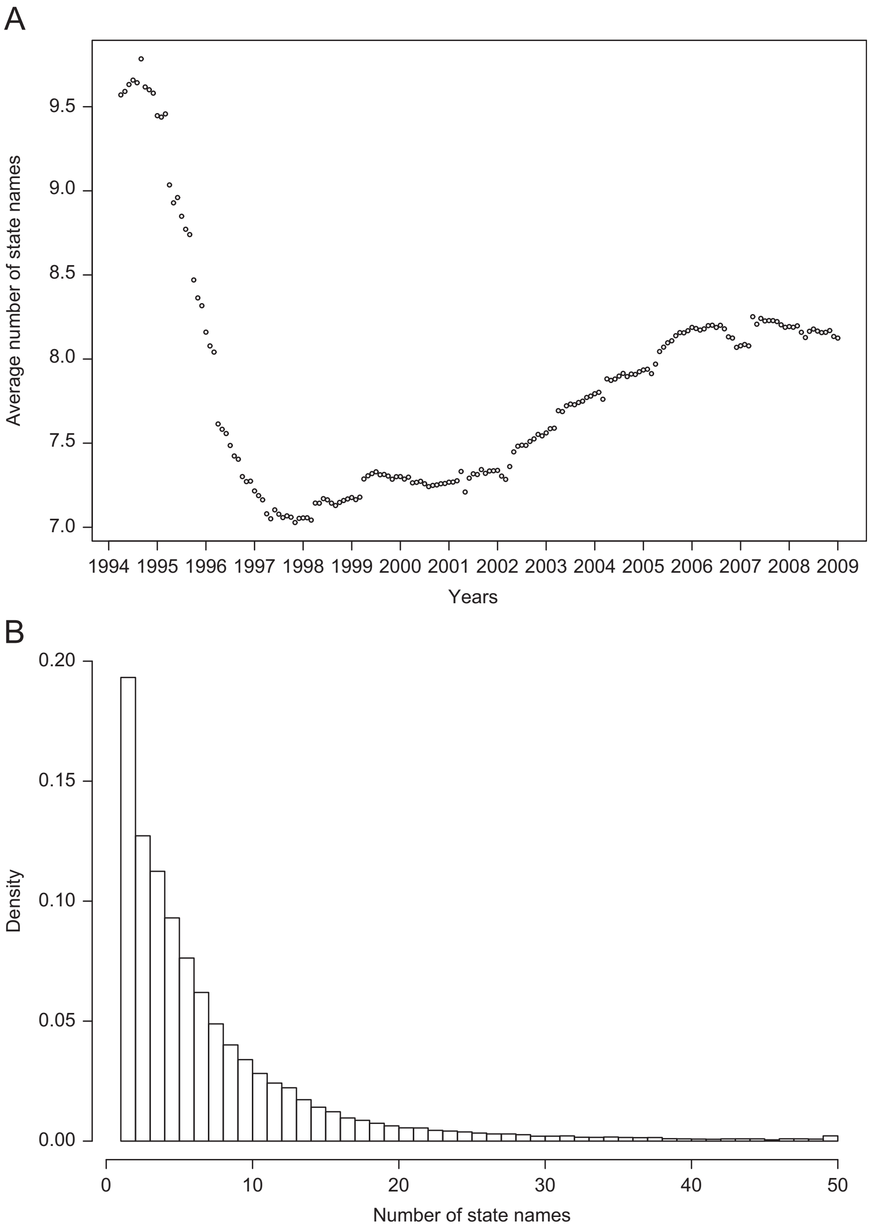

Fig. 1. Geographic dispersion. Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. Panel A plots cross-sectional average geographic dispersion for each month in the period May 1994 through December 2008. The number of firms included in the analysis fluctuates between a low of 934 in 1994 (when EDGAR filings were optional) to a high of 6,293 in 1998. Panel B shows a histogram of geographic dispersion across all firm-years. Panel A: Mean geographical dispersion, Panel B: Histogram of geographical dispersion.

Since 1997, when EDGAR filing was mandatory for all U.S. publicly traded firms, the average number of states mentioned in 10-Ks has increased steadily from around seven states to around eight states. Next we turn to the row labeled Median in Panel A of Table 1. Using the time series of cross-sectional medians, the median firm in the median year mentions five states in its 10-K, indicating a distribution of state counts that is skewed to the right.

More importantly for our purposes, Table 1 shows that there is a significant variation in our measure of geographic dispersion. In particular, the cross-sectional standard deviation of the number of states is 7.7. Moreover, this cross-sectional variation does not change much over time. The minimum standard deviation is 6.9 while the maximum is 8.4. Focusing on the column labeled 20%, we observe that as many as 20% of the firms in our sample do business in three states or less. In the following, we will refer to firms below the 20th percentile as being "local." The last column of Panel A shows that for a typical year in our sample period, 80% of all firms do business in

11 states or less. We will refer to firms that do business in more states than the firm at the 80th percentile as being "geographically dispersed." Looking at the rows labeled Minimum and Maximum, we see that the 20th percentile varies between two and three states over the sample period while the 80th percentile varies between 10 and 14 states. Panel B of Fig. 1 contains the full histogram of our geographical dispersion. As expected, it is heavily skewed to the right, with most firms clustered on single-digit state counts, but with a significant number of companies that operate in multiple states.

Panel B of Table 1 breaks down the averages from the first row of Panel A by the size of firms. As one would expect, big firms are more geographically dispersed, having almost twice as many state names mentioned in their 10-K statement as small firms. The difference is economically large: The average number of state names for small firms is 5.9 while the corresponding average for big firms is 10.5. To study how stock returns vary by geographic dispersion, we require cross-sectional variation in dispersion that is independent of other firm characteristics known to be related to returns. Panel B shows, that even within size terciles, there is a significant amount of variation in geographic dispersion. For small firms, the average 20th percentile is 2.1 states while the average 80th percentile is 8.5 states. The corresponding numbers of states for big firms are 3.5 and 15.3. For all three size groups, the lowest number of states mentioned in a 10-K is one state. The corresponding maximum number of states varies from an average of 48 states for small firms to an average of 50 states for large firms.

In sum, Table 1 shows significant cross-sectional variation in geographic dispersion. This geographic dispersion is stable over time and remains large even when breaking down the cross-section by size. Next, we further explore how geographic dispersion relates to firm size and other firm characteristics such as book-to-market ratio, liquidity, volatility, and stock return momentum.

2.3. Geographic dispersion and other firm characteristics

Previous research has found that, in the cross-section of firms, stock returns are related to a number of firm characteristics. We expect that our measure of geographic dispersion will be related to many of the same firm characteristics. For example, it seems likely that local firms will tend to be smaller and less liquid than geographically dispersed firms. Table 2 investigates this conjecture. Panel A shows how the averages of size (ME) measured using stock market capitalization, the book-to-market ratio (BEME), liquidity (AMI) measured as in Amihud (2002), liquidity measured using the proportional quoted bid–ask spread (SPR), and idiosyncratic volatility (VOL) vary between quintiles of geographic dispersion.

The first row in Panel A shows that the average 10-K state count for firms classified as local is 1.9. The corresponding average state count for firms classified as geographically dispersed is close to 20. As expected, local firms are smaller than dispersed firms. Moving from the first quintile of geographic dispersion (local firms) to the fifth quintile (dispersed firms), the average size (ME) more than doubles. As average stock returns are negatively related to size, the size effect would tend to cause higher returns for local firms. The book-to-market ratio is monotonically increasing as geographic dispersion increases. Although the difference in book-to-market ratios between

Table 2
Geographic dispersion and other firm characteristics.

Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. Geographic dispersion for year t is the number of U.S. states mentioned in the last annual report filed prior to July of year t. Size (the market value of common equity) and the Book-to-market ratio are computed as described in Fama and French (1993). Amihud illiquidity is the price impact liquidity measure of Amihud (2002). Bid–ask spread is the proportional quoted spread measured as: $100(P_A - P_B)/(0.5P_A + 0.5P_B)$, where P_A is the ask price and P_B is the bid price. Volatility is computed as the standard deviation of the error term from a regression using the three-factor model of Fama and French (1993) on one month worth of daily data. Momentum is the buy and hold return for months $t-12$ through $t-2$. In Panel A, all variables are measured as of July each year and the panel reports time series averages of cross-sectional averages. Panel B reports results from a pooled time series cross-sectional regression. All variables in the regression are measured in natural logs. For the momentum variable, the natural log is computed from 1+Momentum. YEARS indicates the presence of dummy variables for each year. DIVS indicates the presence of dummy variables for each of nine U.S. Census divisions. INDS indicates the presence of dummy variables for each of 12 industries from Ken French's Web site. The column labeled R^2 reports adjusted R-squared. The column labeled N reports the number of firm-months used in the regression. Parentheses contain t-statistics computed from the heteroskedasticity-consistent standard errors of White (1980). The sample period is July 1994 through December 2008.

Panel A: Averages by geographic dispersion quintiles

Variable	Local	2	3	4	Dispersed
Geographic dispersion	1.9	3.8	5.7	8.8	19.9
Size (ME)	1,732	1,640	1,862	2,645	3,963
Book-to-market ratio (BEME)	0.689	0.688	0.707	0.760	0.767
Amihud illiquidity (AMI)	0.028	0.021	0.012	0.014	0.006
Bid–ask spread (SPR)	0.030	0.028	0.026	0.023	0.018
Volatility (VOL)	0.031	0.032	0.031	0.028	0.025
Momentum (MOM)	0.150	0.119	0.108	0.107	0.110

Panel B: Regression with geographic dispersion as the dependent variable

ME	BEME	AMI	SPR	VOL	MOM	YEARS	DIVS	INDS	R^2	N
1.03	0.90	−0.31	0.43	0.33	−0.91	Yes	Yes	Yes	0.17	51,902
(25.03)	(24.01)	(−10.72)	(6.15)	(5.99)	(−16.59)					

local firms and dispersed firms is not large, holding other firm characteristics constant, the difference would tend to result in lower returns for local firms.

We study the relation between liquidity and geographic dispersion using both the price impact measure of Amihud (2002) and the proportional quoted bid–ask spread. We set the Amihud illiquidity measure to missing for firm i in month m if the number of days the stock of firm i has traded in month m is below or equal to five. If the dollar volume traded for stock i is high during a month, but the price has moved only very little, the Amihud measure will be small and stock i is said to be liquid. A potential disadvantage of the Amihud measure is that it may be difficult to distinguish liquidity from volatility. We therefore use the bid–ask spread as an alternative measure of liquidity. Proportional quoted spread is computed as $100(P_A - P_B)/(0.5P_A + 0.5P_B)$, where P_A is the ask price and P_B is the bid price. Monthly firm-specific bid–ask spreads are computed as the average daily bid–ask spreads within the month. The fourth and fifth rows in Panel A show that the average liquidity of local firms is lower, using both price impact and bid–ask spread, than the average liquidity of dispersed firms. To the extent that liquidity is priced and illiquid firms are more sensitive to priced liquidity risk than liquid firms, the low liquidity of local firms would cause local firms to have higher average returns than geographically dispersed firms.

Ang, Hodrick, Xing, and Zhang (2006) show that volatility can explain the cross-sectional variation in stock returns. We follow these authors and measure volatility as the standard deviation of the error term from a Fama and French (1993) time series regression using daily data for one month. Ang, Hodrick, Xing, and Zhang (2006) find that firms with high volatility in month $t-1$ tend to experience low stock returns in the following months. Looking at the last row of Panel A of Table 2, local firms tend to be more volatile than dispersed firms. In isolation, this would tend to cause local firms to have lower average returns than dispersed firms. The last row of Panel A shows how average stock return momentum varies by geographic dispersion quintiles. We follow Fama and French (2008) and measure momentum as the cumulative return from month $t-12$ to $t-2$. Even though average past returns are higher for local firms than for dispersed firms, neither group of firms displays stock return momentum that is unusually high, on average.

In Panel B of Table 2, we run a regression with geographic dispersion as the dependent variable and other firm characteristics, year dummies, industry dummies, and U.S. Census division dummies as independent variables. All firm characteristic measures are transformed using the natural logarithm. Each firm is allocated to one of 12 industries using Ken French's industry classification and Standard Industrial Classification (SIC) codes from CRSP. Each firm is also allocated to one of nine U.S. Census divisions based on the location of the firm's headquarters. The headquarters location is from Compustat. Controlling for year, census division, and industry effects, the results from Panel B confirm that geographic dispersion is positively related to size and book-to-market ratio and negatively related to Amihud illiquidity and momentum. However, when controlling for other firm characteristics, the marginal effect of the bid–ask spread and volatility are positive.

3. Results

The analysis presented in the previous section shows that geographic dispersion varies with firm characteristics known to explain some of the cross-sectional variation in stock returns. In this section, where we present results on the relation between geographic dispersion and stock returns, it therefore becomes important to control for the potentially confounding effect of other firm characteristics. We follow two approaches commonly used in the literature to investigate the relation between returns and firm characteristics. First, we sort firms and form portfolios based on geographic dispersion. Second, we perform cross-sectional regressions along the lines of Fama and MacBeth (1973).

3.1. Portfolios sorted on geographic dispersion

To investigate how stock returns are related to the degree of geographic dispersion, we start by forming five portfolios based on our state count measure. A firm that files a 10-K form on or before June of year t is eligible for inclusion in a portfolio starting in July of year t. The firm carries its state count up to and including June of the next year. A firm gets added to the portfolio of local firms if its state count is below the 20th percentile in the June cross-section of state counts. Correspondingly, a firm gets added to the portfolio of dispersed firms if the state count exceeds the 80th percentile. Three more portfolios are formed using the 40th and the 60th percentiles as breakpoints. To ensure that portfolios include a sufficient number of firms, portfolio formation starts in July 1994. The sample period ends in December 2008.

In this section we follow Fama and French (2008) and report results using both equally weighted and value-weighted portfolio returns. The advantage of equally weighted returns is that results will not be driven by a few very large stocks. However, when forming portfolios using geographic dispersion, which is negatively correlated with market capitalization, the portfolio of local firms may be unduly influenced by microcaps (defined by Fama and French, 2008 as firms with market cap below the 20th NYSE percentile). Since microcaps only account for about 3% of the aggregate market cap, equally weighted returns may produce results that are unrepresentative of the market. Reporting results using both value weights and equal weights improves our understanding of the pervasiveness of the relation between stock returns and geographic dispersion.

Table 3 shows equally weighted (EW) and value-weighted (VW) monthly return on the portfolios sorted on geographic dispersion. Focusing on the equally weighted portfolio returns, local firms experienced an average monthly return of 1.18% per month during the sample period. Starting with local firms and moving from left to right along the first row in Table 3, the average returns are monotonically decreasing as firms get more

Table 3
Average return on portfolios sorted by geographic dispersion.

The table reports average portfolio returns in percent. Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. Five portfolios are formed based on geographic dispersion. A firm that files a 10-K form on or before June of year t is eligible for inclusion in a portfolio starting in July of year t. The firm carries its state count up to and including June of the next year. A firm gets added to the portfolio of local firms if its state count is below the 20th percentile in the cross-section of state counts. Correspondingly, a firm gets added to the portfolio of dispersed firms if the state count exceeds the 80th percentile. Three more portfolios are formed using the 40th and the 60th percentiles as breakpoints. The sample period is July 1994 through December 2008.

Variable	Local	2	3	4	Dispersed	Local-dispersed
EW returns	1.18	0.97	0.87	0.83	0.62	0.56 (2.73)
VW returns	0.89	0.78	0.74	0.58	0.49	0.40 (2.06)
Average number of firms	1,084	830	784	757	818	

and more geographically dispersed. The average equally weighted monthly return for the quintile of the most dispersed firms is only 0.62% per month. The 56 basis point difference in average monthly equally weighted returns between local and dispersed firms is economically large and statistically significant at conventional levels.

The second row shows a similar pattern for value-weighted returns. The return difference between the local portfolio and the dispersed portfolio is a statistically significant 40 basis points. The difference in return between the equally weighted and value-weighted portfolios indicates that small local firms have higher returns than large local firms, but the effect of geographic dispersion is clearly also present for large firms. Notice that not only are the point estimates for the top and bottom quintiles statistically different, but they are monotonic along the five quantiles, both for equally and value-weighted portfolios. The last row of the table shows that the average number of firms in each of the quintile portfolios varies between 757 and 1,084. The reason why the five portfolios do not contain the same number of firms is related to the fact that the quintile breakpoints are integers. Many firms are operating in two states—all of which get included in the portfolio of local firms.

The return difference between local firms and dispersed firms is related to size. Earlier we documented a relation between geographic dispersion and other firm characteristics. This raises the question of whether the return spread is compensation for exposure to other risk factors. To take this concern into account, we estimate the following regression model:

$$r_{pt} = \alpha_p + \beta_1(\text{Mkt-Rf})_t + \beta_2 \text{SMB}_t + \beta_3 \text{HML}_t + \beta_4 \text{MOM}_t + \beta_5 \text{LIQ}_t + e_t, \quad (1)$$

where r_{pt} is either the monthly return on a given portfolio, or the monthly return on a zero investment portfolio long local firms and short geographically dispersed firms. The market portfolio proxy Mkt-Rf, the size factor SMB, the book-to-market factor HML, and the momentum factor MOM are all available from Ken French's Web site. The liquidity factor LIQ is the "traded" liquidity factor of Pástor and Stambaugh (2003), available from Wharton Research Data Services (WRDS) as a time series updated to December 2008.

Panel A in Table 4 reports factor loadings and Jensen's alpha for equally weighted portfolios formed using quintiles of geographic dispersion. Focusing on the first row of the table, the portfolio of local firms shows a large and statistically significant Jensen's alpha, 48 basis points with a heteroskedasticity-consistent t-statistic of 2.66, relative to the five-factor model. The return on the local portfolio is closely related to the return on the size factor, reinforcing the earlier finding that local firms tend to be smaller firms. But, since the portfolio has a large alpha, the high return on local firms is not driven by the size effect. Moving down in the column labeled "Alpha," the abnormal returns are monotonically decreasing as portfolios contain more geographically dispersed firms, mimicking the change in raw returns shown in Table 3. For the quintile portfolio with the most dispersed firms, the alpha is a statistically significant -22 basis points (t-statistic of -2.06). This portfolio is less sensitive to the size factor, but it shows much stronger sensitivities to the book-to-market factor and the liquidity factor.

The first row of Panel B in Table 4 reports the result from a regression with the equally weighted zero investment portfolio long local firms and short dispersed firms as the dependent variable. The monthly alpha on this portfolio is 70 basis points—corresponding to an annual abnormal return of 8.4%. The associated t-statistic is 4.45, implying an abnormal return statistically significant at all conventional levels. The return on the long-short portfolio is positively related to the size factor and negatively related to the other four factors. However, the factor loadings are unable to explain the large difference in returns between the portfolio of local firms and the portfolio of dispersed firms. The last row in Panel B constructs the long-short portfolio using value weights. The monthly alpha on this portfolio is 50 basis points, with a t-statistic of 2.81. The smaller alpha on the value-weighted portfolio reinforces our previous finding that small local firms have larger abnormal returns than large local firms.

To investigate the effect of small firms further, Panel C of Table 4 reports results after dropping microcaps from all portfolios. This reduces the overall number of firms by approximately 60%. The reduction is largest in the portfolio of local firms where the average number of firms per month drops from 1,084 to 298. The original portfolio of dispersed firms contains only 300 microcaps—removing these results in a new portfolio containing 518 firms on average. As expected, dropping the smallest firms reduces the abnormal performance of the equally weighted long-short portfolio. The alpha drops from 70 basis points using

Table 4
Jensen's alpha for portfolios sorted on geographic dispersion.

Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. Five portfolios are formed based on geographic dispersion. A firm that files a 10-K form on or before June of year t is eligible for inclusion in a portfolio starting in July of year t. The firm carries its state count up to and including June of the next year. A firm gets added to the portfolio of local firms if its state count is below the 20th percentile in the cross-section of state counts. Correspondingly, a firm gets added to the portfolio of dispersed firms if the state count exceeds the 80th percentile. Three more portfolios are formed using the 40th and the 60th percentiles as breakpoints. The regression model is

$$r_{pt} = \alpha_p + \beta_1(\text{Mkt-Rf})_t + \beta_2\text{SMB}_t + \beta_3\text{HML}_t + \beta_4\text{MOM}_t + \beta_5\text{LIQ}_t + e_t,$$

where r_{pt} is either a portfolio excess return (Panel A) or the return on a zero investment portfolio long local firms and short geographically dispersed firms (Panels B and C). The market portfolio Mkt-Rf, the size factor SMB, the book-to-market factor HML, and the momentum factor MOM are downloaded from Ken French's Web site. The liquidity factor LIQ is the "traded" liquidity factor of Pástor and Stambaugh (2003). The coefficients are estimated using Ordinary Least Squares (OLS). The column labeled T reports the number of monthly observations. The column labeled R^2 contains the adjusted R-squared. The numbers in parentheses are t-statistics computed from the heteroskedasticity-consistent standard errors of White (1980). Portfolio returns are measured in percent. The sample period is July 1994 through December 2008.

Portfolio	Alpha	Mkt-Rf	SMB	HML	MOM	LIQ	T	R^2
Panel A: Portfolios formed based on geographic dispersion quintiles (EW returns)								
Local	0.48	0.85	0.94	0.08	−0.10	0.00	174	0.88
	(2.66)	(16.55)	(12.19)	(1.06)	(−1.32)	(0.10)		
2	0.25	0.94	1.00	0.05	−0.17	0.05	174	0.88
	(1.16)	(16.12)	(10.84)	(0.65)	(−1.76)	(0.99)		
3	0.09	0.95	0.97	0.15	−0.11	0.07	174	0.90
	(0.51)	(18.95)	(13.31)	(2.37)	(−1.54)	(1.68)		
4	0.01	0.99	0.86	0.30	−0.05	0.06	174	0.93
	(0.06)	(24.68)	(16.51)	(6.10)	(−0.99)	(1.88)		
Dispersed	−0.22	0.94	0.67	0.51	0.03	0.07	174	0.94
	(−2.06)	(27.58)	(14.78)	(10.14)	(0.61)	(2.12)		
Panel B: Portfolios long in local firms and short in dispersed firms using all firms								
EW returns	0.70	−0.09	0.27	−0.44	−0.13	−0.06	174	0.50
	(4.45)	(−2.02)	(3.80)	(−6.56)	(−2.36)	(−1.49)		
VW returns	0.50	0.05	0.02	−0.35	−0.13	−0.02	174	0.24
	(2.81)	(1.01)	(0.38)	(−5.19)	(−2.48)	(−0.34)		
Panel C: Portfolios long in local firms and short in dispersed firms dropping microcaps								
EW returns	0.32	0.05	0.22	−0.43	−0.17	−0.02	174	0.58
	(2.40)	(1.17)	(3.68)	(−7.68)	(−4.27)	(−0.59)		
VW returns	0.51	0.06	−0.01	−0.36	−0.13	−0.02	174	0.23
	(2.82)	(1.02)	(−0.12)	(−5.18)	(−2.58)	(−0.29)		

all firms to 32 basis points when excluding microcaps. With an associated t-statistic of 2.4, the abnormal performance remains statistically significant at conventional levels. Moving to the last row of the table, we observe that the alpha for the value-weighted long-short portfolio is practically unaffected by the microcaps. The alpha is 51 basis points with a t-statistic of 2.82.

The results reported in Table 4 show that local firms outperform geographically dispersed firms. The abnormal performance cannot be explained using standard characteristics-based risk factors. As an alternative to the above time series analysis, the next section investigates to what extent geographic dispersion can explain the cross-sectional variation in stock returns while controlling for other firm characteristics known to explain returns.

3.2. Cross-sectional regressions

The analysis of this section is based on cross-sectional regressions similar to Fama and MacBeth (1973). In particular, for each month in the sample period, we run the following cross-sectional regression:

$$R_{it} - R_{ft} = c_0 + \sum_{m=1}^{M} c_{im} Z_{mit} + e_{it},$$

where R_{it} is return on stock i in month t, R_{ft} is the monthly yield on 30-day Treasury bills, Z_{mit} is one of the following M firm characteristics: geographic dispersion, the natural logarithm of our state name count (from last June); size, the natural logarithm of the market capitalization in month $t-2$; book-to-market ratio, the natural logarithm of the firm's book-to-market ratio measured as of last June; Amihud illiquidity, the natural logarithm of the stock's Amihud (2002) illiquidity measure computed using daily returns and volume from month $t-2$; bid–ask spread, the natural logarithm of $(P_A - P_B)/(0.5P_A + 0.5P_B)$ where P_A is the ask price and P_B is the bid price, both measured in month $t-2$; idiosyncratic volatility, the natural logarithm of the standard deviation of the error term from a regression using the three-factor model of Fama and French (1993) and one month worth of daily data; momentum, the buy and hold return for months $t-12$ through $t-2$; and the one-month lagged return.

Table 5 presents the time series averages and associated t-statistics of the cross-sectional regression coefficients from the above model. Focusing first on the column labeled All firms, we see that there is a strong negative relation between geographic dispersion and future one-month stock returns. The average cross-sectional coefficient associated with the natural logarithm of geographic dispersion is −0.22. To compare this estimate with the findings in

Table 5
Time series averages of cross-sectional regression coefficients.
The table reports time series averages of cross-sectional regression coefficients from the following model:

$$R_{it}-R_{ft}=c_0+\sum_{m=1}^{M}c_{im}Z_{mit}+e_{it},$$

where R_{it} is return on stock i in month t, R_{ft} is the monthly yield on 30-day Treasury bills, Z_{mit} is one of M firm characteristics: Geographic dispersion, the natural logarithm of the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. Each monthly cross-sectional regression uses the state count from last June. Size, the natural logarithm of the market capitalization in month $t-2$. Book-to-market ratio, the natural logarithm of the book-to-market ratio measured as of last June. Amihud illiquidity, the natural logarithm of the Amihud (2002) illiquidity measure computed using daily returns and volume from month $t-2$. Bid-ask spread, the natural logarithm of $(P_A-P_B)/(0.5P_A+0.5P_B)$, where P_A is the ask price and P_B is the bid price, both measured in month $t-2$. Volatility, the natural logarithm of the standard deviation of the error term from a regression using the three-factor model of Fama and French (1993) on one month worth of daily data. Momentum, the buy and hold return for months $t-12$ through $t-2$. One-month lagged return, the return for month $t-1$. Each coefficient time series average is multiplied with 100. Microcaps are all firms below the NYSE 20th size decile. Small firms are larger than the firm at the 20th NYSE decile and smaller than or equally sized to the firm at 50th NYSE decile. Big firms are all firms larger than the firm at the 50th NYSE decile. Parentheses contain t-statistics computed from the standard errors of the time series. The sample period is July 1994 through December 2008.

		Cross-sections grouped by size		
Independent variable	All firms	Microcaps	Small	Big
Geographic dispersion	−0.22 (−4.05)	−0.32 (−4.21)	−0.07 (−1.07)	−0.11 (−2.45)
Size	−0.52 (−3.33)	−1.21 (−4.55)	−0.33 (−1.68)	−0.21 (−1.57)
Book-to-market ratio	0.32 (3.38)	0.30 (2.03)	0.23 (2.16)	0.18 (1.91)
Amihud illiquidity	−0.32 (−2.98)	−0.41 (−3.45)	−0.05 (−0.42)	−0.16 (−1.54)
Bid-ask spread	0.11 (0.90)	0.03 (0.17)	−0.26 (−1.76)	−0.01 (−0.13)
Volatility	−0.16 (−0.63)	−0.13 (−0.46)	−0.30 (−1.14)	−0.32 (−1.32)
Momentum	0.43 (2.14)	0.71 (3.38)	0.21 (0.95)	0.49 (1.51)
One-month lagged return	−4.02 (−5.57)	−4.58 (−5.56)	−1.58 (−1.99)	−1.56 (−1.70)
Intercept	3.45 (2.05)	5.22 (2.48)	1.00 (0.50)	−0.27 (−0.14)
Avg. cross-sectional obs.	3,671	2,111	765	795
Number of cross-	174	174	174	174
Avg. R^2	0.05	0.05	0.06	0.09

Tables 3 and 4, notice from the first row of Table 2 that the average state count in the portfolio of local firms is 1.9 while the average state count in the portfolio of geographically dispersed firms is 19.9. Taking the natural logarithm of these numbers, computing the difference, and multiplying with −0.22 shows that predicted monthly returns of geographically dispersed firms is about 52 basis points lower than monthly predicted returns for local firms, holding fixed other firm characteristics. Thus, the effect of geographic dispersion estimated via these Fama and MacBeth (1973) regressions has a similar magnitude as in our time series analysis of Section 3.1.

The last three columns of Table 5 break down the cross-sections by market capitalization. We follow Fama and French (2008) and divide firms into microcaps, small firms, and large firms based on NYSE market capitalization breakpoints. As in previous sections, microcaps are defined as firms below the 20th NYSE size percentile. Small firms are firms between the 20th and the 50th percentile, while big firms are all firms above the 50th percentile. Consistent with the results from Table 4, we find that the effect of geographic dispersion is stronger for microcaps than for larger firms. The effect is weaker and not statistically significant for small firms, but for big firms it is both economically and statistically significant.[11]

Taken together, the results presented in Tables 4 and 5 provide strong evidence in favor of concluding that local firms earn higher returns than geographically dispersed firms. The effect is robust to controlling for characteristics-based risk factors in time series regressions as well as to firm characteristics in cross-sectional regressions. The next section investigates potential explanations for the large return difference between local firms and geographically dispersed firms.

4. Explaining the large returns on local stocks

This section investigates investor recognition and limits to arbitrage, in the form of transaction costs, as potential causes for the return differential between local firms and dispersed firms. The section concludes with several robustness checks of our main finding.

4.1. Investor recognition

Merton (1987) characterizes equilibrium stock returns when investors are not aware of all securities. In such informationally incomplete markets, stocks with lower investor recognition offer higher expected returns to compensate investors that hold the stock for insufficient

[11] This "U-shaped" cross-sectional effect of geographic dispersion is also evident from the alphas in Table 4. In Panel B of Table 4, the alpha from the equally weighted zero investment portfolio exceeds the alpha for the corresponding value-weighted portfolio. However, when dropping microcaps from the portfolios (Panel C in the same table), the alpha for equally weighted portfolios is smaller than the alpha for the value-weighted portfolios.

diversification. To the extent that local stocks have lower investor recognition, the high average return of local firms shown in the previous section is consistent with the investor recognition hypothesis. We provide two sets of tests that further investigate this hypothesis. First, we compare the returns on portfolios of local firms from geographic areas where there is a good chance of being recognized by investors with returns on portfolios of stocks from areas with a smaller chance of being recognized. Under the investor recognition hypothesis, the returns on local stocks should be high in areas where it is hard to become recognized by investors. Second, we study changes in geographic dispersion. As firms expand geographically, they should become more recognized, and stock returns observed after the expansion should reflect this. Similarly, firms that focus their business and become geographically concentrated should experience higher returns as investors expect these firms to become under-recognized in the future.

We begin to investigate the investor recognition hypothesis by focusing on the amount of capital available to recognize the pricing difference between local firms and geographically dispersed firms. Following the discussion and findings in Hong, Kubik, and Stein (2008), we conjecture that firms are trading at a discount if they are located in areas where the competition for investor attention is fierce. Table 6 shows the returns on portfolios of local stocks that are sorted based on three different measures of investor recognition.

Our first measure of investor recognition is computed at the state level. For each state, we compute the ratio of the number of listed firms to the population of the state, which we loosely refer to as the state's firm density. We group states into low, medium, and high firm density states. The group of states with low density is composed of all states with below-median firm density. The remaining states are divided between medium density states and high density states to ensure that the number of listed firms in both state groups are as close as possible. With this approach, the high density states are Massachusetts, Connecticut, Colorado, New Jersey, Minnesota, and California. Using stocks headquartered in high density states, we form quintile portfolios based on geographic dispersion as before. Similar portfolios are created using stocks headquartered in medium density and low density states.

Local stocks headquartered in states with low firm density should have higher investor recognition than local stocks headquartered in states with high firm density. According to the investor recognition hypothesis, the latter group of local stocks should have higher returns than the former group. Panel A of Table 6 investigates this conjecture. When returns are equally weighted, the portfolio of local firms from states with low firm density (high recognition) has an alpha of 46 basis points with a

Table 6
Jensen's alpha for portfolios of local firms sorted on three different proxies for investor recognition.

This table sorts local firms by three different measures that act as proxies for investor recognition. Local firms are associated with each of the measures then ranked to form three portfolios of local firms. A firm is local if it is among the 20% least geographically dispersed firms. Firm density is computed at the state level and is the ratio of the number of listed firms to the population. Local firms are linked to state density through the state of the headquarters. In Panel A, High recognition local firms are firms located in states with low firm density. In Panel B local firms are characterized by local capital imbalance. The imbalance is computed as the difference between mutual fund capital and listed firm capital in a circle of 100 km around the local firm. High recognition local firms are firms headquartered in areas where there is a lot of mutual fund capital relative to market capitalization of all listed firms in the same area. In Panel C, institutional investor ownership in local firm i is the proportion of equity held by investors that have reported ownership in firm i through 13F filings with the SEC. The regression model is

$$r_{pt} = \alpha_p + \beta_1(\text{Mkt-Rf})_t + \beta_2\text{SMB}_t + \beta_3\text{HML}_t + \beta_4\text{MOM}_t + \beta_5\text{LIQ}_t + e_t,$$

where r_{pt} is a portfolio of local firms. The market portfolio Mkt-Rf, the size factor SMB, the book-to-market factor HML, and the momentum factor MOM are downloaded from Ken French's Web site. The liquidity factor LIQ is the "traded" liquidity factor of Pástor and Stambaugh (2003). The coefficients are estimated using OLS. The numbers in parentheses are t-statistics computed from the heteroskedasticity-consistent standard errors of White (1980). The columns labeled R^2 contain the adjusted R-squared. The sample period is July 1994 through December 2008, but months where the geographic dispersion portfolio contains less than 15 firms are dropped from the time series.

	Equally weighted returns			Value-weighted returns		
Portfolio	Alpha	T	R^2	Alpha	T	R^2
Panel A: Firm density						
High recognition	0.46 (3.04)	174	0.85	0.50 (2.03)	174	0.51
Medium	0.78 (4.25)	174	0.86	0.63 (3.55)	174	0.77
Low recognition	1.04 (4.19)	174	0.85	0.88 (3.72)	174	0.82
Low–high	0.58 (2.64)	174	0.63	0.37 (1.18)	174	0.51
Panel B: Mutual fund capital less market capitalization of all firms						
High recognition	0.61 (3.73)	174	0.85	0.35 (1.48)	174	0.65
Medium	0.86 (3.78)	174	0.83	0.61 (2.51)	174	0.68
Low recognition	0.93 (4.57)	174	0.87	0.74 (3.93)	174	0.82
Low–high	0.31 (1.80)	174	0.47	0.39 (1.31)	174	0.15
Panel C: Institutional ownership from 13-F filings						
High recognition	0.24 (3.15)	173	0.97	0.29 (5.53)	173	0.98
Medium	0.54 (3.29)	173	0.91	0.38 (2.75)	173	0.88
Low recognition	0.70 (2.00)	173	0.70	−0.04 (−0.19)	173	0.82
Low–high	0.45 (1.26)	173	0.22	−0.33 (−1.44)	173	0.58

t-statistic of 3.04. Moving to the next row, the portfolio of local firms from medium density states has an alpha of 78 basis points. For high density states (low recognition), the local firm portfolio has an alpha of 1.04 (*t*-statistic of 4.19). The difference in abnormal returns between local firms in low density states and local firms in high density states is a statistically significant 58 basis points. For value-weighted portfolios, the alpha for the portfolio long in local firms from low recognition states and short in high recognitions states is 37 basis points. However, this alpha is not statistically significant at conventional levels.

Taken together, the evidence in Panel A indicates that local firms from states where there are many other listed local firms show average returns that are higher than average returns for local firms in states where competition for attention is not as strong. The lack of significance for the value-weighted portfolio implies that the effect is most prominent among smaller stocks. The larger effect among smaller stocks seems entirely reasonable. Everything else equal, smaller firms probably have a harder time being recognized by investors than larger firms. In other words, a large local firm would probably suffer less in terms of recognition in a state with high competition for attention than a small local firm.

Our second measure of investor recognition is computed at the zip code level. For each zip code where there is at least one firm classified as local in a given year, we draw a 100 km circle around the zip code. Next, we use the Thomson Reuters Mutual Funds data (s12) and the CRSP Mutual Fund data to locate all mutual funds within this circle, using the zip code of each mutual fund, and add up the amount of capital these mutual funds have invested in stocks of listed firms.[12] To arrive at our second measure of investor recognition, we compute the difference between the amount of mutual fund capital and the market capitalization of all listed firms geographically close to these mutual funds. To be specific, we identify all listed firms located closer than 100 km to at least one of the mutual funds identified in the first step. Then we add up the market capitalization of these listed firms and subtract this from the amount of mutual fund capital. This gives us a measure, at the zip code level, of the imbalance between capital available to invest locally and investment opportunities available locally. When this imbalance is large, the likelihood of being recognized by (mutual fund) investors should be larger. To form portfolios, each local firm is associated with capital imbalance using the zip code of the firm's headquarters. At the end of June, all local firms are ranked based on the capital imbalance. Three portfolios of local firms are formed using the 33rd and 67th percentiles of the capital imbalance ranking. Firms are held in the portfolio for one year, at which point the selection procedure is repeated.

The results using the mutual funds measure of investor recognition is reported in Panel B of Table 6. We see a pattern very similar to the one shown in Panel A. In areas where the amount of capital invested by mutual funds is small relative to the market capitalization of all firms, the alpha on a portfolio of local firms is larger than the alpha on a portfolio of local firms from areas with high investor recognition. For the equally weighted portfolio, the five-factor alpha of the long-short portfolio is 31 basis points with a *t*-statistic of 1.8. The point estimate for the value-weighted long-short portfolio is slightly higher, but with a *t*-statistic of only 1.3. Thus, again it seems that the investor recognition story may contribute in explaining the large alphas on portfolios of local firms. However, the tests we are using seem to have limited power. The alphas for the long-short portfolio in Panel B are relatively large, but we have a hard time making a statistically strong case for a difference that is related to our measure of investor recognition.

In the final panel of Table 6, we take a slightly different approach to measure investor recognition. Panel B has focused on the effect of being recognized by mutual fund investors. Another approach to measure the extent to which institutional investors have recognized a local firm is to measure and rank local firms on the actual ownership of institutional investors. To this end, we measure institutional investor ownership in local firm i as the proportion of equity held by investors that have reported ownership in firm i through 13F filings with the SEC.[13] In Panel C, the high investor recognition portfolio contains the one-third of local firms with the largest institutional ownership. The low investor recognition portfolio contains the one-third of local firms with low institutional ownership. The evidence is mixed. For the equally weighted portfolios, local firms with high institutional ownership have lower returns than local firms with low institutional ownership. To the extent that investor recognition is positively correlated with institutional ownership, this is consistent with the investor recognition hypothesis. However, for the value-weighted long-short portfolio, the alpha is negative. Moreover, neither the equally weighted nor the value-weighted long-short portfolio have an alpha that is statistically significant. Thus, forming portfolios of local firms based on institutional ownership does not lend convincing support to the investor recognition hypothesis.

Overall, the evidence presented in Table 6 provides support for the investor recognition hypothesis. When sorting local firms based on firm density and mutual fund capital less market capitalization of listed firms, there is evidence that local firms with high investor recognition had lower returns than local firms with low investor recognition. We find weaker results when sorting on institutional ownership. It is clear that the effects of investor recognition, with our proxies, are more pronounced for small firms than for large firms. We interpret this as evidence in favor of the view that large firms suffer less than small firms in terms of recognition in a state with high competition for attention.

Next, we further investigate the investor recognition hypothesis by looking at changes in investor recognition. Table 7 shows Jensen's alpha on portfolios formed using

[12] We use the approach described in Coval and Moskowitz (2001) to determine the distance between two zip codes.

[13] We rely on the Thomson Reuters Institutional Investor data (s34).

Table 7
Jensen's alpha for portfolios sorted on change in geographic dispersion.

Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. Change in geographic dispersion is measured over 12 months. A firm is local if it is among the 20% least geographically dispersed firms. Firms that change into the quintile of the least geographically dispersed firms during a fiscal year are said to "become local". Having identified changes in geographic dispersion, we form equally weighted and value-weighted portfolios, including firms only after the change is observable through the filing of 10-Ks. Firms are kept in the "become local" portfolio for 12 months. Unless the firm experiences a new change in geographic dispersion, the firm is moved into the "Already local" portfolio after 12 months. The portfolios "Become dispersed" and "Already dispersed" are formed in a similar fashion. The regression model is

$$r_{pt} = \alpha_p + \beta_1(\text{Mkt-Rf})_t + \beta_2\text{SMB}_t + \beta_3\text{HML}_t + \beta_4\text{MOM}_t + \beta_5\text{LIQ}_t + e_t,$$

where r_{pt} is a zero investment portfolio long local firms and short geographically dispersed firms. The market portfolio Mkt-Rf, the size factor SMB, the book-to-market factor HML, and the momentum factor MOM are downloaded from Ken French's Web site. The liquidity factor LIQ is the "traded" liquidity factor of Pástor and Stambaugh (2003). The coefficients are estimated using OLS. The numbers in parentheses are t-statistics computed from the heteroskedasticity-consistent standard errors of White (1980). The sample period is July 1994 through December 2008, but months where the geographic dispersion portfolio contains less than 15 firms are dropped from the time series.

	Equally weighted returns			Value-weighted returns		
Portfolio	Alpha	T	R^2	Alpha	T	R^2
Become local	0.98 (3.86)	162	0.84	0.66 (2.21)	162	0.63
Already local	0.79 (4.27)	162	0.87	0.68 (4.31)	162	0.85
Difference	0.20 (1.28)	162	0.35	−0.01 (−0.04)	162	0.01
Become dispersed	−0.06 (−0.40)	162	0.90	0.24 (1.07)	162	0.72
Already dispersed	0.14 (1.22)	162	0.94	0.14 (1.26)	162	0.89
Difference	−0.20 (−1.38)	162	0.19	0.10 (0.40)	162	0.15

changes in investor recognition. Under the investor recognition hypothesis, a local firm has high returns because investors demand a premium to hold under-recognized stocks. As the firm expands geographically, it should become more recognized, and the stock returns observed after the expansion should reflect this. Firms that focus their business and become geographically concentrated should experience higher returns as investors expect these firms to become under-recognized in the future.

We measure the change in geographic dispersion over 12 months. Those firms that change into the quintile of the least geographically dispersed firms during a fiscal year are said to "Become local." Most companies that become local naturally move from the geographic dispersion quintile closest to local firms. However, there are firms that "Become local" from all the other quintiles. Having identified changes in geographic dispersion, we form equally weighted and value-weighted portfolios, including firms only after the change is observable through the filing of 10-Ks. Firms are kept in the "Become local" portfolio for 12 months. Unless the firm experiences a new change in geographic dispersion, the firm is moved into the "Already local" portfolio after 12 months. We form portfolios "Become dispersed" and "Already dispersed" in a similar fashion.

Table 7 shows alphas on portfolios formed as described in the previous paragraph. Using equally weighted returns, the portfolio of firms that becomes local has a five-factor alpha of 98 basis points. The corresponding alpha for firms that were already classified as local is 79 basis points.[14] The difference is only 20 basis points and is not statistically significant at conventional levels. For the value-weighted portfolios, there is no difference in alphas between firms that are local and firms that become local. A similar conclusion applies when comparing firms that become geographically dispersed and firms that are already dispersed.

A natural conjecture, in the context of the investor recognition hypothesis, is that investors learn about firms that become geographically proximate more quickly than they forget about firms that leave their geographic proximity. As a consequence, one should expect firms that become local to have more modest price reactions than those that become dispersed. The point estimates in Table 7 do not lend support to this conjecture. The year after the change, both firms becoming local and firms becoming geographically dispersed have returns that are similar to comparable firms that already are local and dispersed, respectively. Table 7 shows that firms changing their geographic dispersion behave more like the firms they become similar to than the firms they used to be, consistent with investor recognition being priced into asset prices within a year.

4.2. Liquidity and volatility

Fang and Peress (2009) find that firms with little media coverage have higher returns than comparable firms with high media coverage. They point out that some investors may recognize all securities, but that limits to arbitrage prevent them from taking advantage of the apparent mispricing between stocks. Consistent with this view, they show that the media effect is stronger for low liquidity and high volatility firms. This section investigates if the effect of geographic dispersion is related to liquidity and volatility.

To investigate the importance of liquidity, we first sort firms into three portfolios based on the Amihud illiquidity

[14] The reason why both the reported alphas are larger than the alpha reported in the first row of Table 4 is related to sample composition. To measure a change in geographic dispersion, we require portfolio firms to have data on geographic dispersion in two consecutive years before being included in the portfolio.

measure. Then, within each liquidity portfolio, we sort firms into quintile portfolios based on their geographic dispersion. The same procedure is followed replacing Amihud illiquidity with bid–ask spread and volatility. Table 8 presents alphas for portfolios, within sorts on liquidity and volatility, that are long local firms and short geographically dispersed firms. Using all available firms to form portfolios, the first vertical segment of Panel A shows that the alpha for the long-short portfolio formed using the most liquid firms is 32 basis points. This alpha increases to 71 basis points for firms with medium liquidity and to 93 basis points for the least liquid firms. The difference in alphas for liquid and illiquid firms is 62 basis points with an associated t-statistic of 2.69. The fact that the alpha is monotonically increasing with reduced liquidity, and the economically and statistically significant difference in the alphas, seem to support the conclusions of Fang and Peress (2009) on the importance of liquidity for firms with low investor recognition.

However, when investigating the effect of Amihud liquidity within microcaps, and within the group of firms that are not microcaps, the effect of liquidity is dramatically reduced. First, when only using microcaps to form portfolios, the alphas on the portfolios long local firms and short geographically dispersed firms are economically and statistically significant. However, the alphas for liquid and illiquid firms are not statistically distinguishable from each other. The same conclusion applies to firms that are not microcaps. The implication of this finding is that most of the alpha difference found when using all firms to form portfolios is driven by the difference in liquidity between microcaps and other firms. That is, we cannot separate the liquidity effect from the size effect that we have documented in earlier tables.

Panel B of Table 8 reports the results from a similar analysis using the bid–ask spread as a liquidity measure. For microcaps, there is an effect of liquidity. The alpha for the most liquid firms is 53 basis points smaller than the alpha for the least liquid firms. Using a one-sided test, the difference is statistically significant at below the 5% level. However, moving to the group of firms that are not microcaps, there is no effect of the bid–ask spread, as the alpha of the long-short portfolio is 36 basis points for highly liquid firms, but only 33 basis points for stocks with high bid–ask spreads.

Panel C of Table 8 investigates the effect of volatility on the alphas on the portfolios long local firms and short geographically dispersed firms. The Merton (1987) model implies that investors require compensation for taking on the idiosyncratic risk that follows from holding less than perfectly diversified portfolios. Thus, local firms with high idiosyncratic risk should command higher expected returns than local firms with less idiosyncratic risk. We investigate this prediction by studying the alphas on long-short portfolios when portfolios are formed within groups of firms sorted based on idiosyncratic risk. Using all available firms to form portfolios, the difference in alphas for low volatility firms and high volatility firms is a statistically significant 87 basis points. This difference remains strong among microcaps (68 basis points with a t-statistic of 1.87) but is much weaker among non-microcaps (28 basis points with a t-statistic of 1.26).

Table 8
Jensen's alpha for equally weighted double sorted portfolios long in local firms and short in geographically dispersed firms.

Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. A firm is local if it is among the 20% least geographically dispersed firms. A firm is dispersed if it is among the 20% most geographically dispersed firms. The regression model is

$$r_{pt} = \alpha_p + \beta_1 (\text{Mkt-Rf})_t + \beta_2 \text{SMB}_t + \beta_3 \text{HML}_t + \beta_4 \text{MOM}_t + \beta_5 \text{LIQ}_t + e_t,$$

where r_{pt} is a zero investment portfolio long local firms and short geographically dispersed firms. Portfolio returns are equally weighted. The market portfolio Mkt-Rf, the size factor SMB, the book-to-market factor HML, and the momentum factor MOM are downloaded from Ken French's Web site. The liquidity factor LIQ is the "traded" liquidity factor of Pástor and Stambaugh (2003). The coefficients are estimated using OLS. The numbers in parentheses are t-statistics computed from the heteroskedasticity-consistent standard errors of White (1980). The columns labeled R^2 contain the adjusted R-squared. The sample period is July 1994 through December 2008, but months where the geographic dispersion portfolio contains less than 15 firms are dropped from the time series.

	All firms			Microcaps			All but microcaps		
Portfolio	Alpha	T	R^2	Alpha	T	R^2	Alpha	T	R^2
Panel A: Amihud illiquidity									
Liquid	0.32 (1.80)	174	0.62	0.86 (3.56)	163	0.38	0.50 (2.35)	174	0.49
Medium	0.71 (3.96)	174	0.39	0.98 (4.99)	162	0.20	0.22 (1.04)	174	0.52
Illiquid	0.93 (5.41)	162	0.28	1.04 (4.07)	162	0.20	0.29 (1.78)	174	0.32
Illiquid–liquid	0.62 (2.69)	162	0.46	0.18 (0.52)	162	0.08	−0.21 (−0.98)	174	0.29
Panel B: Bid–ask spread									
Small	0.33 (1.96)	174	0.62	0.59 (3.24)	173	0.23	0.36 (1.81)	174	0.58
Medium	0.71 (3.25)	174	0.21	0.93 (4.19)	162	0.24	0.22 (1.19)	174	0.50
Large	1.13 (5.68)	162	0.40	1.23 (4.84)	162	0.28	0.33 (1.57)	174	0.16
Large–small	0.83 (3.62)	162	0.25	0.53 (1.73)	162	0.05	−0.03 (−0.10)	174	0.31
Panel C: Volatility									
Low	0.40 (3.46)	174	0.35	0.41 (2.62)	174	0.32	0.27 (2.57)	174	0.20
Medium	0.63 (3.80)	174	0.41	1.11 (5.32)	162	0.35	0.29 (1.65)	174	0.24
High	1.25 (5.25)	162	0.40	1.22 (3.68)	151	0.30	0.55 (2.37)	174	0.47
High–low	0.87 (3.54)	162	0.31	0.68 (1.87)	151	0.31	0.28 (1.26)	174	0.38

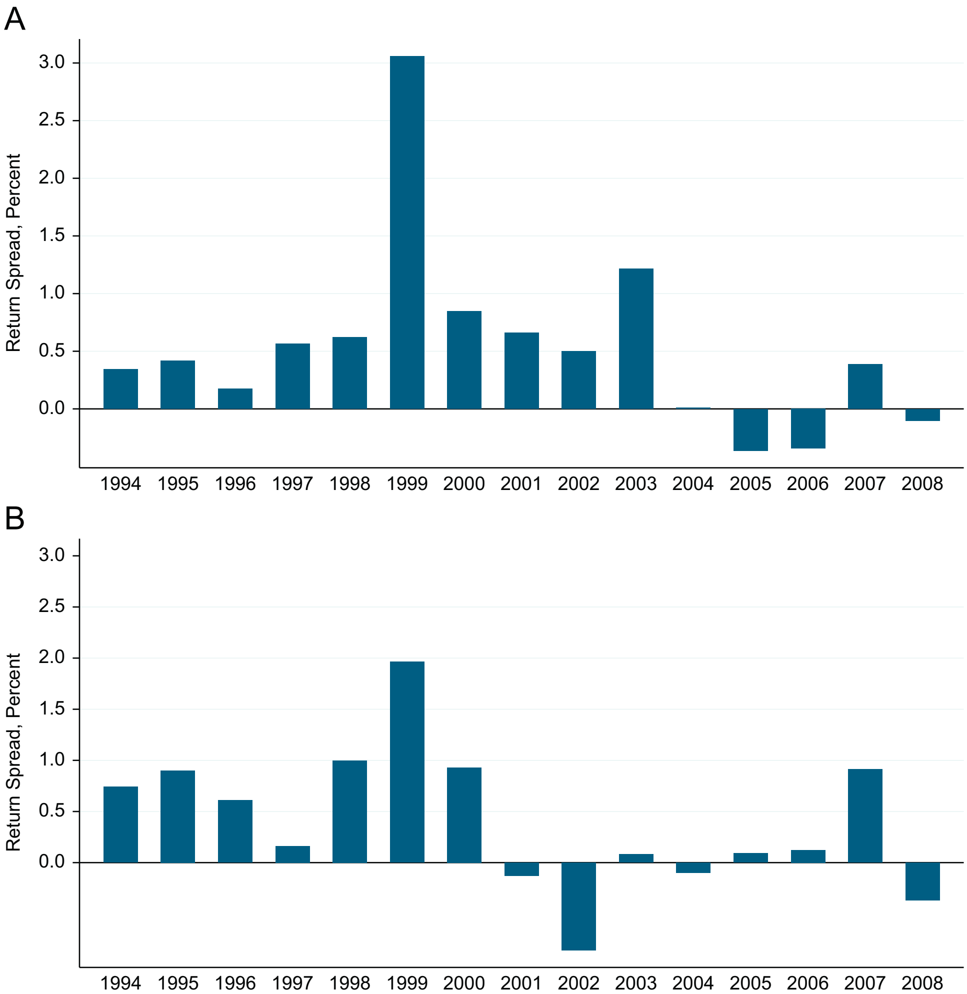

Fig. 2. Returns on portfolios long local firms and short geographically dispersed firms. Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. Two portfolios are formed based on geographic dispersion. A firm that files a 10-K form on or before June of year *t* is eligible for inclusion in a portfolio starting in July of year *t*. The firm carries its state count up to and including June of the next year. A firm gets added to the portfolio of local firms if its state count is below the 20th percentile in the cross-section of state counts. Correspondingly, a firm gets added to the portfolio of dispersed firms if the state count exceeds the 80th percentile. The figure reports the return on a portfolio long in local firms and short in geographically dispersed firms. The sample includes 66,628 firm-years for the sample period 1994–2008. Panel A: EW returns, Panel B: VW returns.

Overall, Table 8 shows a strong effect of liquidity and volatility when using all firms to form portfolios. However, these effects are hard to distinguish from size effects. This does not imply that the effect of geographic dispersion is not stronger for illiquid firms with high volatility, but rather that the effect is hard to disentangle from the size effect. It seems reasonable to conclude that size, liquidity, and volatility together influence the effect of geographic dispersion in much the same way as these variables modify the media effect studied by Fang and Peress (2009). That is, arbitrageurs would find it harder to profit on the mispricing documented in this paper when firms are small, have low liquidity, and are volatile.

We conclude this section by reporting how our results changed throughout our sample period. While EDGAR was put in place in the mid-1990s, the possibility of obtaining a time series with enough observations to make reasonable statements about the effect of geographic dispersion was not available until the latter part of our sample period. Moreover, geography did not take a central role in the finance research community until the early 2000s.[15] If arbitrageurs spotted the pricing anomalies we show, we would expect to see the mispricing diminish throughout our sample. Fig. 2 plots the average returns, for each year in our sample, of the long-short portfolio constructed as in Table 3. The returns from such a trading strategy paid off handsomely during the first 10 years of our sample—both the equally and the value-weighted returns are positive in all 10 years. On the other hand, the effect disappears in the 2004–2008 subsample. Indeed,

[15] See Coval and Moskowitz (1999, 2001) for the first academic studies on the topic.

Table 9

Jensen's alpha for portfolios long in local firms and short in geographically dispersed firms by industries and using EW returns.

Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. A firm is local if it is among the 20% least geographically dispersed firms. A firm is dispersed if it is among the 20% most geographically dispersed firms. The regression model is

$$r_{pt} = \alpha_p + \beta_1(\text{Mkt-Rf})_t + \beta_2\text{SMB}_t + \beta_3\text{HML}_t + \beta_4\text{MOM}_t + \beta_5\text{LIQ}_t + e_t,$$

where r_{pt} is a zero investment portfolio long local firms and short geographically dispersed firms. The market portfolio Mkt-Rf, the size factor SMB, the book-to-market factor HML, and the momentum factor MOM are downloaded from Ken French's Web site. The liquidity factor LIQ is the "traded" liquidity factor of Pástor and Stambaugh (2003). The coefficients are estimated using OLS. The numbers in parentheses are t-statistics computed from the heteroskedasticity-consistent standard errors of White (1980). The column labeled R^2 contains the adjusted R-squared. The sample period is July 1994 through December 2008, but months where the geographic dispersion portfolio contains less than 15 firms are dropped from the time series.

Portfolio	Alpha	T	R^2
Panel A: Double sorts using eight industries and quintiles of geographic dispersion			
Consumer durables and non-durables	−0.09 (−0.36)	174	0.07
Manufacturing	0.55 (2.07)	174	0.23
Energy and chemicals	0.54 (1.67)	174	0.03
Business equipment, telecom, and utilities	0.66 (2.53)	174	0.52
Wholesale and retail	0.82 (3.30)	174	0.09
Health	0.86 (2.21)	173	0.34
Finance	0.62 (3.61)	174	0.27
Other	0.85 (3.22)	174	0.30
F-test of equal alphas	1.68 [0.11]	8×173	
Panel B: Individual stock returns replaced with industry portfolio returns			
Local	0.15 (1.00)	174	0.92
Dispersed	−0.02 (−0.15)	174	0.92
Local-dispersed	0.17 (3.55)	174	0.43

the popularity of text analysis in academic circles started around the 2004 date (Tetlock, 2007). Thus, it is natural to conjecture that the investment community spotted our pricing anomaly, and corrected it by the end of our sample period.

4.3. Industry and other measures of dispersion

Hou and Robinson (2006) conclude that industry concentration affects equilibrium stock returns. If industry membership is correlated with geographical dispersion, our findings could possibly be caused by industry membership rather than geographic dispersion. This section addresses this concern. We also investigate the robustness of our findings using alternative measures of geographic dispersion.

Table 9 investigates the role of industry by creating portfolios within broad industry classifications.[16] In particular, for firms within each of the eight industries listed in Panel A of Table 9, we estimate five-factor alphas for the equally weighted portfolio long in local firms and short in geographically dispersed firms. The estimated alpha from these eight time series regressions is presented in the second column. The numbers are positive and large for all but one industry, for which the point estimate is virtually zero. Furthermore, a formal F-test of the equality of the eight alphas has a p-value of 11%. Thus, we cannot reject the hypothesis that the effect is the same for all industries. We conclude that the effect of geographic dispersion is not driven by one particular industry group.

In Panel B of Table 9 we conduct a further test as to whether our results are driven by industry effects. We repeat the portfolio formation in Section 3.1 with the difference that we replace a firm's stock return by that of its industry, using the Fama and French 38 industries classification to define industry membership. The local portfolio's alpha is 15 basis points, whereas that of the dispersed portfolio is −2 basis points. Comparing this to the estimates from Table 4, 48 and −22 basis points, we see that industry itself cannot explain our findings. A similar conclusion emerges from the long-short portfolio. The last row of Table 9 shows that the alpha of the portfolio obtained substituting a firm's return for that of its industry is 17 basis points, which is less than one-fourth of the point estimate of 70 basis points from Table 4.

Finally, we run Fama and MacBeth (1973) regressions similar to Table 5, but with added industry dummies created based on the Fama–French 38 industries classifications. Results are not reported, but the conclusion remains the same—industry fixed effects do not explain the significance of geographical dispersion as a determinant of stock returns.

Table 10 presents our results using alternative measures of geographical dispersion. The seven rows in this table report the alphas of long-short portfolios, similar to Panel B of Table 4. We change the measure of geographical dispersion in each of the rows. In the first row we use the Herfindahl index to measure geographic dispersion.[17] An argument that favors such a measure is that it is

[16] To have a sufficient number of firms per month, we use eight broadly defined industries, closely aligned to Fama and French 12 industries classification.

[17] We construct the Herfindahl index as follows. We create a vector $x \in \mathbb{R}^{50}$ that has as entry x_i the proportion of all state names mentioned in the 10-K statement that are associated with state i. The Herfindahl index is then defined as usual as $H = \sum_{i=1}^{50} x_i^2$.

Table 10
Jensen's alpha for portfolios long in local firms and short in geographically dispersed firms using alternative measures of geographic dispersion.

Geographic dispersion is measured as the number of U.S. states mentioned in the annual report filed on Form 10-K with the SEC. A firm is local if it is among the 20% least geographically dispersed firms. A firm is dispersed if it is among the 20% most geographically dispersed firms. The regression model is

$$r_{pt} = \alpha_p + \beta_1(\text{Mkt-Rf})_t + \beta_2 \text{SMB}_t + \beta_3 \text{HML}_t + \beta_4 \text{MOM}_t + \beta_5 \text{LIQ}_t + e_t,$$

where r_{pt} is a zero investment portfolio long local firms and short geographically dispersed firms. The market portfolio Mkt-Rf, the size factor SMB, the book-to-market factor HML, and the momentum factor MOM are downloaded from Ken French's Web site. The liquidity factor LIQ is the "traded" liquidity factor of Pástor and Stambaugh (2003). The coefficients are estimated using OLS. The numbers in parentheses are t-statistics computed from the heteroskedasticity-consistent standard errors of White (1980). The columns labeled R^2 contain the adjusted R-squared. The sample period is July 1994 through December 2008, but months where the geographic dispersion portfolio contains less than 15 firms are dropped from the time series.

	EW returns			VW returns		
Measure of geographic dispersion	Alpha	T	R^2	Alpha	T	R^2
1 − Herfindahl	0.50 (3.75)	174	0.39	0.49 (2.87)	174	0.31
U.S. divisions	0.52 (5.71)	174	0.35	0.55 (3.60)	174	0.09
Three-year moving average	0.53 (3.58)	154	0.55	0.34 (1.98)	154	0.14
Max/min	0.54 (3.75)	154	0.52	0.28 (1.72)	154	0.08
Dropping firms when present in						
One or more countries	0.51 (2.07)	144	0.14	0.84 (2.28)	144	−0.02
Five or more countries	0.65 (4.11)	174	0.42	0.44 (1.83)	174	0.34
Country names	−0.12 (−0.80)	174	0.57	−0.25 (−1.27)	174	0.19

continuous and it is widely used to summarize multi-dimensional information such as the state counts that are the core of our analysis. The Herfindahl index is more likely to classify a firm as local even when many states are mentioned in the 10-K but there is one state that receives a large number of counts. The results using the Herfindahl measure parallel those from Table 4. The alpha of the equally weighted portfolio is 50 basis points, whereas that of the value-weighted portfolio is 49 basis points, both highly significant by standard confidence levels.

Another interesting alternative measure of geographical dispersion uses the nine U.S. Census divisions as the measure of location, rather than the 50 U.S. states. A firm that operates in Pennsylvania and California is arguably more geographically dispersed than one that operates in Virginia and North Carolina (both part of the South Atlantic division). Our next alternative geographical dispersion metric is constructed by associating firms with census divisions using the state names mentioned in their 10-K statements. In other words, if a firm mentions Pennsylvania and California in their 10-K statement, the firm is said to do business in both the Mid-Atlantic division and the Pacific division. The second row of Table 10 reports the alpha estimates for the long-short portfolios using both equal and value weights. The conclusion from our previous analysis carries through. The long-short portfolios have alphas above 50 basis points per month, both of which are economically large and statistically different from zero.

The next two rows include the analysis using two alternative metrics of geographic dispersion. The first simply computes a three-year moving average of the state counts from the 10-K statements. Firms are classified as local or geographically dispersed if they are in the bottom or top quintiles of this metric. The second, labeled max/min, classifies a firm as geographically dispersed if it is in the top quintile of geographic dispersion, when geographic dispersion is the maximum number of different states mentioned in a 10-K statement over the last three years. This second metric classifies a firm as local if it is in the bottom quintile of geographic dispersion, where geographic dispersion is the minimum number of different states mentioned in a 10-K statement over the last three years.

These two metrics serve as a conservative anchor, as they should eliminate some of the noise that could stem from our text parsing algorithm. Our previous conclusions are reinforced. Jensen's alphas on the portfolios are on the order of 50 basis points for the equally weighted portfolios and 30 basis points for the value-weighted portfolios.

Finally, we investigate if firms' international presence affects our results. If international presence expands the investor base, it should lead to lower expected returns for international firms. However, the strong home bias of investors (French and Poterba, 1991) suggests that international presence will only have a small effect on the investor base. In other words, a company with operations in China and California may not reach more investors than a similar firm with operations in California only. Nonetheless, it is possible that firms with global operations fail to list U.S. states in their 10-K because it is obvious that they are present in most states. Thus, international presence may cloud our results due to measurement error.

We check if our conclusions are robust to international presence by dropping firms that may have operations outside the U.S. In the third row of Table 10 we drop all firms that mention one or more country names in their 10-K statement. Using our state count measure of geographic dispersion in the sample of non-international firms, the alpha on the equally weighted portfolio is 51 basis points, whereas the value-weighted portfolio has an alpha of 84 basis points. The number of firms remaining in the sample, after dropping firms with some international presence, is significantly lower than for the full sample. Nonetheless, our results are still large in economic terms, and statistically different from zero. To retain more firms, the fourth row reports alphas on the long-short portfolio formed using firms that mention less than five countries in their 10-K statements. The alphas remain large in this subsample as well.

In our final robustness test, we check whether the international dispersion of a firm can have an effect similar to the effect of domestic dispersion. In the last row of Table 10, we report alphas when geographic dispersion is measured using country name counts rather than state name counts. In particular, we redo our previous analysis using the counts of 200 different countries instead of the earlier state name counts. Both the equally weighted and the value-weighted portfolios have alphas that are not distinguishable from zero at conventional levels of statistical significance. Based on the last three rows of Table 10, we conclude that our results are not driven by firms with international presence.

Overall, Tables 9 and 10 show that our main conclusion is not driven by any particular industry, that it is robust to using alternative measures of geographic dispersion, and that it is not driven by firms with international presence.

5. Conclusion

This paper presents the first large sample study of the geographical dispersion of U.S. publicly traded firms. We find a pattern in stock returns that sheds new light on the pricing of local assets. We show that the geographical dispersion of a firm's business activities, measured by the number of states mentioned in a company's annual report, is related to average returns. Local firms, those that operate in two states or less, have average returns that are 70 basis points higher than firms whose operations transcend more than 20 states.

We interpret our evidence as consistent with the predictions of the investor recognition hypothesis of Merton (1987). In Merton's informationally incomplete markets, stocks with lower investor recognition offer higher expected returns to compensate investors for insufficient diversification. To the extent that local stocks have lower investor recognition, the high average return of local firms is consistent with this prediction. The paper also shows that stocks of local firms headquartered in areas where the competition for investor attention is fierce experience higher returns than local stocks headquartered in areas where fewer firms compete for attention.

Our study shows how one can obtain an economically meaningful cross-sectional characterization of firms, in our case the geographical dispersion of operations, from the filings of 10-K forms on EDGAR.[18] The use of textual analysis of business-related information is a promising area for future research. Our study of geographic dispersion and stock returns is only one of many potential questions that can be addressed using these type of data in general—and using our geographic dispersion measure in particular.

[18] See Hoberg and Phillips (2010) for another example of how to use textual information to capture cross-sectional characteristics of firms.

References

Almazan, A., Motta, A.D., Titman, S., Uysal, V., 2010. Financial structure, acquisition opportunities, and firm locations. Journal of Finance 65, 529–563.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. Journal of Finance 61, 259–299.

Becker, B., Ivković, Z., Weisbenner, S., 2011. Local dividend clienteles. Journal of Finance 66, 655–683.

Bodnaruk, A., 2009. Proximity always matters: local bias when the set of local companies changes. Review of Finance 13, 629–656.

Bodnaruk, A., Östberg, P., 2008. Does investor recognition predict returns? Journal of Financial Economics 91, 208–226.

Brown, J.R., Ivković, Z., Weisbenner, S., Smith, P.A., 2008. Neighbors matter: causal community effects and stock market participation. Journal of Finance 63, 1509–1531.

Coval, J.D., Moskowitz, T.J., 1999. Home bias at home: local equity preference in domestic portfolios. Journal of Finance 54, 2045–2073.

Coval, J.D., Moskowitz, T.J., 2001. The geography of investment: informed trading and asset prices. Journal of Political Economy 109, 811–841.

Dorn, D., Huberman, G., Sengmueller, P., 2008. Correlated trading and returns. Journal of Finance 63, 885–920.

Doukas, J., Travlos, N.G., 1988. The effect of corporate multinationalism on shareholders' wealth: evidence from international acquisitions. Journal of Finance 43, 1161–1175.

Ellison, G., Glaeser, E.L., 1997. Geographic concentration in U.S. manufacturing industries: a dartboard approach. Journal of Political Economy 105, 889–927.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 43, 3–56.

Fama, E.F., French, K.R., 2008. Dissecting anomalies. Journal of Finance 63, 1653–1678.

Fama, E.F., MacBeth, J., 1973. Risk, return and equilibrium: empirical tests. Journal of Political Economy 81, 607–636.

Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. Journal of Finance 64, 2023–2052.

Feng, L., Seasholes, M.S., 2004. Correlated trading and location. Journal of Finance 59, 2117–2144.

Foerster, S.R., Karolyi, G.A., 1999. The effects of market segmentation and investor recognition on asset prices: evidence from foreign stocks listing in the United States. Journal of Finance 54, 981–1013.

French, K., Poterba, J., 1991. International diversification and international equity markets. American Economic Review 81, 222–226.

Gao, W., Ng, L., Wang, Q., 2008. Does geographic dispersion affect firm valuation? Journal of Corporate Finance 14, 674–687.

García, D., Strobl, G., 2011. Relative wealth concerns and complementarities in information acquisition. Review of Financial Studies 24, 169–207.

Gervais, S., Kaniel, R., Mingelgrin, R.R., 2001. The high-volume return premium. Journal of Finance 56, 877–919.

Gómez, J.-P., Priestley, R., Zapatero, F., 2008. The effect of relative wealth concerns on the cross-section of stock returns. Unpublished working paper. IE Business School, Norwegian Business School and University of Southern California.

Grinblatt, M., Keloharju, M., 2001. How distance, language, and culture influence stockholdings and trades. Journal of Finance 56, 1053–1073.

Hau, H., 2001. Location matters: an examination of trading profits. Journal of Finance 56, 1959–1983.

Hoberg, J., Phillips, G., 2010. Product market synergies and competition in mergers and acquisitions: a text-based analysis. Review of Financial Studies 23, 3773–3811.

Hong, H., Kubik, J.D., Stein, J.C., 2004. Social interaction and stock-market participation. Journal of Finance 59, 137–163.

Hong, H., Kubik, J.D., Stein, J.C., 2005. Thy neighbor's portfolio: word-of-mouth effects in the holdings and trades of money managers. Journal of Finance 60, 2801–2824.

Hong, H., Kubik, J.D., Stein, J.C., 2008. The only game in town: stock-price consequences of local bias. Journal of Financial Economics 90, 20–37.

Hou, K., Robinson, D.T., 2006. Industry concentration and average stock returns. Journal of Finance 61, 1927–1956.

Huberman, G., 2001. Familiarity breeds investment. Review of Financial Studies 14, 659–680.
Ivković, Z., Sialm, C., Weisbenner, S., 2008. Portfolio concentration and the performance of individual investors. Journal of Financial and Quantitative Analysis 43, 613–656.
Ivković, Z., Weisbenner, S., 2005. Local does as local is: information content of the geography of individual investors' common stock investments. Journal of Finance 60, 267–306.
Ivković, Z., Weisbenner, S., 2007. Information diffusion effects in individual investors' common stock purchases: covet thy neighbors' investment choices. Review of Financial Studies 20, 1327–1357.
Kang, J.-K., Kim, J.-M., 2008. The geography of block acquisitions. Journal of Finance 63, 2817–2858.
Kaniel, R., Ozoguz, A., Starks, L.T., 2012. The high volume return premium: cross-country evidence. Journal of Financial Economics 103, 255–279.
Landier, A., Nair, V.B., Wulf, J., 2009. Trade-offs in staying close: corporate decision making and geographic dispersion. Review of Financial Studies 22, 1119–1148.
Loughran, T., 2007. Geographic dissemination of information. Journal of Corporate Finance 13, 675–694.
Loughran, T., Schultz, P., 2005. Liquidity: urban versus rural firms. Journal of Financial Economics 78, 341–374.
Merton, R.C., 1987. A simple model of capital market equilibrium with incomplete information. Journal of Finance 42, 483–510.
Pástor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.
Pirinsky, C.A., Wang, Q., 2006. Does corporate headquarters location matter for stock returns? Journal of Finance 61, 1991–2015.
Seasholes, M.S., Zhu, N., 2010. Individual investors and local bias. Journal of Finance 65, 1987–2010.
Teo, M., 2009. The geography of hedge funds. Review of Financial Studies 22, 3531–3561.
Tetlock, P.C., 2007. Giving content to investor sentiment: the role of media in the stock market. Journal of Finance 62, 1139–1168.
Uysal, V.B., Kedia, S., Panchapagesan, V., 2008. Geography and acquirer returns. Journal of Financial Intermediation 17, 256–275.
VanNieuwerburgh, S., Veldkamp, L., 2009. Information immobility and the home bias puzzle. Journal of Finance 64, 1187–1215.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.