THE JOURNAL OF FINANCE • VOL. LXV, NO. 5 • OCTOBER 2010

Individual Investors and Local Bias

MARK S. SEASHOLES and NING ZHU*

ABSTRACT

The paper tests whether individuals have value-relevant information about local stocks (where "local" is defined as being headquartered near where an investor lives). Our methodology uses two types of calendar-time portfolios—one based on holdings and one based on transactions. Portfolios of local holdings do not generate abnormal performance (alphas are zero). When studying transactions, purchases of local stocks significantly *underperform* sales of local stocks. The underperformance remains when focusing on stocks with potentially high levels of information asymmetries. We conclude that individuals do not help incorporate information into stock prices. Our conclusions directly contradict existing studies.

THIS PAPER STUDIES THE GEOGRAPHY of individual investors' portfolios. Recent research provides ample evidence that individuals tilt their portfolios towards local stocks. For example, the typical U.S. household has about 30% of its portfolio invested in stocks headquartered within a 250-mile radius of the family's home. On average, only 12% of all firms (the market) are headquartered within the same radius. In Finland, the median non-Helsinki-headquartered firm has 12% greater weight among investors in its municipality than it does among all Finnish investors. And, in mainland China, individuals invest 8% more in firms from their province of residence than a market capitalization portfolio would predict.[1] However, despite consistent and strong evidence of local bias, an unanswered question remains: Why do investors have a preference for local stocks?

The ability to exploit nonpublic information could be one reason individuals tilt their portfolios toward local stocks. Ivkovic and Weisbenner (2005, p. 305) conclude that individual households are "able to process and

*Seasholes is with Hong Kong University of Science and Technology (HKUST), and Zhu is with the Shanghai Advanced Institute of Finance and University of California at Davis. We thank a large discount broker, Brad Barber, and Terry Odean for providing the data used in this study. Terry Hendershott, Chris Hennessy, Gur Huberman, Alok Kumar, Tobias Moskowitz, Terry Odean, and Richard Stanton have given many helpful comments. Cynthia Qi Pan provided excellent research assistance. Zhu acknowledges support from the National Natural Science Foundation of China (70932002). We are especially grateful to Brad Barber, who has been instrumental in the development of this paper.

[1]For U.S. results see Ivkovic and Weisbenner (2005) and Zhu (2003). Related results are in Huberman (2001). For Finnish results see Grinblatt and Keloharju (2001). For results from mainland China see Feng and Seasholes (2004). Throughout this paper we primarily consider a distance-based measure of "local" and "remote." Stocks that are headquartered within a certain radius of a household are considered local for that household. Stocks that are headquartered outside the radius are considered remote. We consider radii of 100 km, 100 miles, and 250 miles.

exploit locally available information to earn excess returns." The authors note that their findings are "particularly strong where information asymmetries are likely to play the most pronounced role—among the non-S&P 500, less widely known stocks." Similarly, Massa and Simonov (2006, p. 682) study individual portfolio holdings in Sweden and find that investors "deliberately tilt their portfolio towards stocks that are most closely related to them," including geographically local stocks. The authors conclude that information drives these investment choices.

Finding that individuals exploit nonpublic information and achieve superior returns on their local investments has a number of implications. First, such a finding may help financial economists better understand how markets aggregate information—a topic that has been central to the field since early rational expectations models.[2] Two obvious empirical questions arise from these theory models: (i) which agents help impound information into prices? and (ii) can we uncover evidence of information being incorporated into prices by examining investors' holdings and trades? If the conclusions reached by Ivkovic and Weisbenner (2005) are correct, individual investors may play a key role in incorporating information into prices.

Second, finding that individuals achieve superior returns on local stocks has implications regarding the advice (implicit or explicit) that financial economists provide to the investing public. The ability to achieve superior returns on local stocks might justify shifting one's portfolio away from both the market portfolio and indexing. In particular, if superior returns on local stocks are possible, individuals might rationally engage in stock picking—provided that the returns of these stocks cover the costs associated with acquiring them.

Our paper tests empirically whether individuals earn superior returns on their local investments. We study households' portfolio holdings and also their transactions. Our basic methodology involves forming portfolios based on the holdings or trades of many investors. Throughout the paper the corresponding aggregations are referred to as "holdings-based calendar-time portfolios" and "transactions-based calendar-time portfolios," respectively. Studying both holdings and trades is one aspect of our paper that sets it apart from existing studies of individual investor geography.

Calendar-time portfolios address four pitfalls that potentially affect studies of individuals' investments. These same four pitfalls affect earlier studies of individual investors and local bias. First, calendar-time portfolios take into account cross-sectional correlation of stock returns. Earlier studies of individuals' portfolio returns assume independence. Second, portfolios dampen the effect of small stocks on returns. Existing studies of individuals' portfolios tend to overweight the effect of small positions and/or small stocks. Third, when studying investor geography specifically, portfolios from a given geographic area can be compared against a passive benchmark from the same geographic area. Fourth, and finally, portfolios allow a financial economist to make full use of a data set's entire time series. Some earlier papers use holdings based on a single point in time.

[2]For examples, see Hellwig (1980), Diamond and Verrecchia (1981), and Admati (1985).

Our paper starts with a standard performance analysis of individuals' local holdings (using holdings-based calendar-time portfolios). Holdings of local stocks outperform the riskless asset by 1.1% per month over our sample period. Regression analysis, with controls for the market's excess return, an investor-specific passive index, and the Fama-French-Carhart factors, paints a different picture: individuals' local portfolios do not generate abnormal performance, portfolio "alphas" are not statistically different from zero, and point estimates of alphas are economically close to zero. Tables IV–VI in this paper, combined with tables in the Internet Appendix, test 45 different regression permutations.[3] None of the alphas is significantly different from zero at the 5% level (nor at the 10% level.) We conclude that individuals do not appear to have value-relevant information about the local stocks they hold.

Next, we use transactions-based calendar-time portfolios to test the following hypothesis: if individuals have value-relevant information, the local stocks they buy should outperform the local stocks they sell. We study 249,555 local stock transactions. We form one portfolio based on purchases of local stocks, and a second portfolio based on sales of local stocks. Stocks are held in our portfolios for 1 year, which is approximately the average holding period for investors in our sample. We then measure the average return of the "Buys-minus-Sells" portfolio and find it to be −1.7% per year. When we confine the analysis to only trades of local non-S&P 500 stocks, the Buys-minus-Sells portfolio has a return of −2.3% per year.[4] In other words, local buys of these stocks *underperform* local sells by 2.3% per year. Individuals again appear to have no value-relevant information about the local stocks they trade. The finding regarding local, non-S&P 500 stocks is especially important as these stocks have potentially high levels of information asymmetries. Our conclusions directly contradict the conclusions reached in earlier papers.

Pitfalls When Studying Individual Investor Behavior

The methodology used in our paper addresses four pitfalls that arise when studying individuals' portfolios.

Pitfall #1. Cross-sectional Correlation of Individuals' Portfolio Returns: The first pitfall comes from the fact that individuals' portfolio returns are cross-sectionally correlated. For instance, although our data set of local holdings-based portfolio returns contains 938,644 monthly observations, we do not have nearly this many independent observations. Investors in our 71-month data set hold 5,779 different stocks in their local portfolios, with 215,512 unique stock-month combinations. The monthly returns of two investors who hold the

[3] An Internet Appendix for this article is available online in the "Supplements and Datasets" section at http://www.afajof.org/supplements.asp.

[4] Non-S&P 500 stocks are generally smaller than, have less institutional trading than, and are not followed by as many analysts as stocks in the index. We refer to these stocks as stocks with higher levels of information asymmetries. Please see Section I for additional discussion of the classification methodology.

same stocks, over the same months, are obviously correlated and dependent. Similar issues arise when studying transactions.

We address issues related to the number of independent observations in two ways. With the holdings-based portfolios, we compute Rogers (1993) standard errors that are robust to heteroskedasticity and contemporaneous correlation (clustered by month). With the transactions-based portfolios, we calculate the returns of a Buys-minus-Sells portfolio, which results in a single time series. The single time-series takes care of cross-sectional correlation. Please see Appendix A and Appendix C for additional notes related to Pitfall #1.

Pitfall #2. Small Stocks and Individual Portfolios: The second pitfall arises when studying investor-level portfolios. As mentioned above, our local holdings data comprise 938,644 household-month observations. The median household has approximately three stocks in its portfolio in a given month. If one-third of a typical portfolio is classified as local, then some individuals have a single stock in their local portfolios. Thus, the monthly return of a single small stock may be counted as an observation in a typical regression analysis.[5]

When conducting standard performance analysis, the monthly return of a wealthy, well-diversified individual is counted as a single observation. The monthly return of a not-so-wealthy, poorly diversified individual is also counted as a single observation. Thus, single small stocks can overly influence results (in much the same way that equal-weighting a portfolio can alter its returns). We address possible biases caused by small stocks by forming portfolios that combine the value-weighted holdings of many investors. We also use transactions-based portfolios to combine all purchases or sales into a single value-weighted portfolio. Weights in the transactions-based portfolios are initially set equal to the value of the underlying transactions. See Appendix B for additional notes related to small stocks and Pitfall #2.

Pitfall #3. Geographic Selection Biases: The third pitfall concerns cross-sectional sample selection biases. In this paper, we study 43,132 households that invest through a large discount broker in the United States. Neither investors, nor firms, nor industries are uniformly distributed across the country. A graphic example of the nonuniform distributions can be seen in Figure 1 of Ivkovic and Weisbenner (2005). In the figure, it is clear that firm headquarters have highest density in the Northeast of the country, around Chicago, and in the San Francisco Bay Area. The figure also shows that investors are densely populated in the Northeast and in the Bay Area.

During our 1991–1996 sample period, some industries experienced high returns (Banks, Finance, Business Services, and Electrical Equipment all had average monthly returns over 2%) while other industries experienced low returns (Utilities, Construction, Tobacco, Gold, and Coal had average monthly returns less than 1%). Since investors in our sample live disproportionately near Silicon Valley and New York City (and all investors tend to tilt their portfolios towards local stocks), we might incorrectly conclude that

[5] Returns of individuals' portfolios are always value-weighted in our paper. Some portfolios, however, may only contain a single local stock.

local stock pickers are able to earn abnormally high returns. The incorrect conclusion would be due to a cross-sectional sampling error. Our holdings-based portfolio analysis addresses geographic biases by including passive local indices as control variables. Transactions-based portfolios address potential geographic biases by testing whether buys of local stocks outperform sales of local stocks.

Pitfall #4. Time-Series Selection Biases: The fourth and final pitfall relates to time-series selection biases. The main sample of 34,517 households studied by Ivkovic and Weisbenner (2005) is based on households with at least $1,000 in stock holdings at the end of 1991. Individuals (as a group) have varying performance between 1991 and 1996. We show that local buys outperform local sells in 1991, break even in 1995, and underperform in 1992, 1993, 1994, and 1996. Our paper addresses potential time-series biases by considering all holdings and transactions over the entire 1991–1996 sample period.

Individual Investors and Performance

Our paper is part of an ongoing investigation into the performance of individual investors. Odean (1998) finds that winning stocks sold by individuals outperform losing stocks held by individuals over a 1-year horizon. Odean (1999) considers all stocks individuals trade (not just recent winners and losers in their portfolios) and finds that buys underperform sells. Our paper differs by testing whether individuals achieve abnormally high performance when trading local stocks. We also consider stocks with potentially lower and higher levels of information asymmetries.

Two recent papers show that individual investors' Buys-minus-Sells predict future positive returns at short horizons in the United States. Kaniel, Saar, and Titman (2008) study trades by individual investors on the New York Stock Exchange (NYSE). They find that stocks most heavily bought outperform stocks most heavily sold over a 1- to 4-week horizon. The authors argue that individuals are being compensated for providing immediacy to institutions: individuals buy (sell) when institutions are selling (buying) and prices are falling (rising), and the subsequent price increase (decrease) compensates the individuals in a manner similar to how market makers are compensated. Barber, Odean, and Zhu (2009) study small trades (those less than $5,000 in value), which they interpret as coming from individual investors. They find that buys outperform sales at horizons of 1 week to 1 month, writing that "stocks heavily bought by individual investors one week earn strong returns contemporaneously in the subsequent week, while stocks heavily sold one week earn poor returns contemporaneously and in the subsequent week." This pattern persists for a total of 3–4 weeks and then reverses for the subsequent several weeks. See also Coval, Hirshleifer, and Shumway (2005) and Hau (2001).

Finally, Kumar (2004), like our paper, examines whether individual investors have superior information about local stocks. Using time-varying, investor-specific benchmarks, Kumar (2004) finds that, at most, one-third of investors possess local information. He further finds that information drives purchases of small-cap, value stocks, and that information effects are particularly strong for investors who reside in remote geographical locations.

Our paper proceeds as follows. Section I describes the data and gives summary statistics. Section II presents our results. Section III concludes.

I. Data

A. Individual Investor Data

Our individual investor data come from a large discount brokerage house and have been extensively studied in the behavioral finance literature.[6] Transaction data start on January 1, 1991 and end on November 30, 1996. Monthly portfolio positions start on January 31, 1991 and end on November 1, 1996. A demographic file with information on a subset of households contains the five-digit zip code and state in which a household is located. Throughout this paper we use the terms "individual" and "household" interchangeably.

B. Stock Price and Return Data

Our holdings-based calendar-time portfolios use monthly returns obtained from Center for Research in Security Prices (CRSP). The transactions-based calendar-time portfolios use daily returns also obtained from CRSP. Returns of the value-weighted market portfolio, the risk-free rate, and the Fama-French-Carhart factors are downloaded from Ken French's website (monthly and daily). We recalculate monthly returns of the value-weighted market portfolio using only stocks with zip code information. Doing so eliminates a potential selection bias but has no effect on our results.

C. Location Data

We translate household zip codes into latitudes and longitudes using the 1990 Census U.S. Gazetteer available at www.census.gov. For each of 8,773 different listed common stocks (CUSIP numbers), state and county information is obtained from Compustat. The U.S. Gazetteer is used to translate state and county information into latitudes and longitudes of the firms' headquarters. Using state and county information (instead of zip codes obtained from Compustat) increases the sample size considerably.

Throughout this paper we divide each household's portfolio into local stocks and remote (nonlocal) stocks using a distance-based measure. Stocks whose

[6] See Barber and Odean (2000, 2001, 2002), Barber et al. (2009), Dhar and Zhu (2006), Goetzmann and Kumar (2008), Graham and Kumar (2006), Ivkovic, Poterba, and Weisbenner (2005), Ivkovic and Weisbenner (2005), Kumar (2004, 2009a, 2009b), Kumar and Lee (2006), Kumar and Lim (2008), and Odean (1998, 1999). A detailed description of the data can be found in Barber and Odean (2000).

headquarters are within a 250-mile radius of where an investor lives are considered local *for that investor*. Stocks whose headquarters are outside the 250 mile radius are considered remote.[7] Thus, the same stock (say Microsoft) is considered to be *local* for investors living in Seattle but is considered to be *remote* for investors living in Atlanta. The distance-based measure allows us to create investor-specific definitions of local and remote.

D. Information Asymmetry/Size Classification

We use S&P 500 index membership to classify stocks into those likely to have lower levels of information asymmetries and those likely to have higher levels of information asymmetries. The S&P 500 index consists of large firms and constitutes approximately 80% of the market capitalization of U.S. equities. Stocks in the index are generally followed by large numbers of institutional investors and by many analysts. Non-S&P 500 stocks are typically smaller firms that constitute the remaining 20% of the market capitalization of U.S. equities (approximately).[8] We obtain the history of firms in the S&P 500 from Barclay, Hendershott, and Jones (2008).

E. Summary Statistics

Table I provides summary statistics of our data. Panel A starts with 77,795 household accounts. Focusing on households with location information reduces the sample size to 54,538 households. We further narrow the sample to households within the continental U.S. and those with at least one holding of common stock between 1991 and 1996. The last filter requires households to hold at least one common stock with location information. The final sample contains 43,132 households. While the average household holds approximately three stocks at a time, many hold only a single stock (which may or may not be classified as local). The sparsity of holdings at the household level is one reason this paper advocates forming portfolios.

Table I, Panel B reports the number of households and value of holdings at the end of each year. There are 32,723 households at the end of 1991 and this

[7]Coval and Moskowitz (2001) provide a distance formula in footnote 3 on p. 815. We present results based on the 250-mile radius in order to facilitate straightforward comparisons with existing papers. Results based on 100 miles and 100 km radii are qualitatively similar and shown in the Internet Appendix. Ideally we would also like to measure the distance from each investor's home to each stock's closest branch office or subsidiary. Although we do not have such data, Massa and Simonov (2006, p. 652) are able to construct this measure using Swedish data. They find that results do not differ materially when using a measure based on firm headquarters compared with a measure based on the closest branch office/subsidiary.

[8]The Internet Appendix shows an alternative measure of information asymmetry based on dollar trading volume, analyst coverage, and S&P 500 inclusion. The alternative measure has an 88–98% overlap with S&P 500 inclusion. The advantages of using only S&P 500 inclusion are that it is straightforward, does not reduce the sample size by requiring stocks to have certain data, allows for comparison with existing papers, and avoids any dependence on stock returns (which are used to quantify investor performance).

Table I

Summary Statistics

This table presents summary statistics for our data. Investor data come from a large discount brokerage. Household location is based on each household's zip code. Firm location information is based on the county and state of the headquarters as reported by Compustat. In Panel B, households consist of those with at least one common stock holding with firm location information from (v) in Panel A. In Panels D and E, local stocks are defined as those headquartered within a 250-mile radius of a household, while remote stocks are defined as those headquartered outside the 250-mile radius. In Panel E, stocks are also classified by whether they are included in the S&P 500 index.

Panel A: Number of Households

(i)	Households	77,795
(ii)	Households from (i) with location information	54,538
(iii)	Households from (ii) in Continental United States	53,443
(iv)	Households from (iii) with at least one common stock holding during our sample period	44,836
(v)	Households from (iv) with at least one common stock holding with firm-location information	43,132

Panel B: Holdings

Portfolio Date	Number of Households	Value of Holdings ($ mil)
December 1991	32,723	1,012.0
December 1992	33,483	1,701.9
December 1993	28,736	1,592.4
December 1994	21,021	1,187.6
December 1995	16,738	1,069.9

Panel C: Transactions per Year

	Number of Transactions			Value ($ million)		
Year	Buys	Sells	Total	Buys	Sells	Total
1991	92,164	68,434	160,598	885	828	1,712
1992	88,962	68,209	157,171	902	864	1,765
1993	84,791	75,177	159,968	913	943	1,855
1994	72,268	62,678	134,946	784	800	1,584
1995	93,268	85,092	178,360	1,221	1,297	2,518
1996	102,892	89,387	192,279	1,426	1,512	2,938
Total	534,345	448,977	983,322	6,130	6,243	12,373

Panel D: Transactions by Location Based on 250-Mile Radius; All Years Together

	Number of Transactions			Value ($ million)		
	Buys	Sells	Total	Buys	Sells	Total
Local stocks	134,766	114,789	249,555	1,585	1,719	3,304
Remote stocks	399,579	334,188	733,767	4,545	4,524	9,069
Total	534,345	448,977	983,322	6,130	6,243	12,373

(*continued*)

Table I—*Continued*

Panel E: Transactions by Location and S&P 500 Inclusion Location Based on 250-Mile Radius; All Years Together

	Number of Transactions			Value ($ million)		
	Buys	Sells	Total	Buys	Sells	Total
Local + S&P 500	47,696	44,624	92,320	655	736	1,390
Local + Non-S&P 500	87,072	70,163	157,235	930	983	1,914
Remote + S&P 500	162,546	141,806	304,352	2,250	2,300	4,550
Remote + Non-S&P 500	237,031	192,384	429,415	2,295	2,224	4,519
Total	534,345	448,977	983,322	6,130	6,243	12,373

number falls to 16,738 households by the end of 1995. The aggregate value of holdings is just over US$ 1 billion at the end of 1991. The value rises to US$ 1.6 billion by the end of 1993 before falling back to about US$ 1 billion by the end of 1995.

Panels C, D, and E provide an overview of our transactions data. There are a total of 983,322 transactions, of which 249,555 involve local stocks. Local buys outnumber local sells (134,766 vs. 114,789) though local sales have higher value. Of the transactions in local stocks, 92,320 are in S&P 500 stocks while 157,235 transactions are not.

Holdings, Location, and Local Bias

We test whether individual investors overweigh local stocks in their portfolios relative to a market capitalization portfolio. Table II presents results based on year-end holdings. We show that investors hold approximately 30% of their portfolios in stocks located within a 250-mile radius of their home. Approximately 12% of the market is headquartered within the same radius.

We calculate three measures of local bias. The difference between local holdings and available local holdings ranges between 17.1% and 18.4% with little variation over the 5 years. The ratio of local holdings to available local holdings ranges between 1.40 and 1.56 each year (after subtracting one). The natural log of the previous ratio ranges between 0.88 and 0.94 each year. All measures generate consistent results that households tilt their portfolios towards local stocks. The Internet Appendix shows the evidence of local bias is robust to using a 100-mile radius, a 100-km radius, and state boundaries.

II. RESULTS

A. Holdings-Based Calendar-Time Portfolios

We test whether individuals' local investments earn superior returns using standard performance analysis. For each individual, we calculate the

Table II

Holdings, Location, and Local Bias

This table shows the degree to which households overweigh local stocks. We report averages across households. In the table below, we calculate the fraction of each household's portfolio invested within a 250-mile radius of the family's home. Distance is measured from the household's zip code to the zip code of the firm's headquarters. For each household, we also calculate the fraction of the market (all stocks) within the same radius. The difference or ratio of columns A and B represents a measure of local bias.

	(A) Average % of Household's Portfolio	(B) Average % of Market	Local Bias Measures		
Portfolio Date	$\leq$250 miles	$\leq$250 miles	Difference A – B	Ratio #1 A/B – 1	Ratio #2 ln(A/B)
December 1991	30.3	12.6	17.7	1.40	0.88
December 1992	29.3	12.1	17.2	1.42	0.88
December 1993	30.2	11.8	18.4	1.56	0.94
December 1994	29.1	12.0	17.1	1.43	0.89
December 1995	30.2	12.0	18.2	1.52	0.92

value-weighted return of his local holdings. Our holdings data are at a monthly frequency and each individual produces a single time series with up to 71 months of local returns. Table III defines variables used in Tables IV–VI.

Table IV, Panel A provides summary statistics for the monthly portfolio returns. Our data comprise 938,644 individual-month observations. We report means and standard deviations across all 938,644 observations.

We regress an individual's excess local return ($R_{local,i} - R_f$) on the market's excess return ($R^*_m - R_f$), the excess return of a passive zip code–level index ($R_z - R_f$), and the Fama-French-Carhart factors (*SMB*, *HML*, and *MOM*). To lessen the effect of small stocks (see Pitfall #2 mentioned in the introduction), we winsorize local portfolio returns at the 0.5% and 99.5% levels. The market's return is denoted R^*_m and is defined as the value-weighted return of all CRSP stocks with zip code data. Each household in our data is located within one of 7,832 zip codes. A single, passive zip code–level index is calculated as the value-weighted return of all stocks headquartered within a 250-mile radius of the given zip code. Individuals who live in different zip codes are associated with different passive zip code–level indices.

Estimation is conducted using pooled ordinary least squares. Since individual *i*'s return in a given month may be correlated with individual *j*'s return in the same month, we compute standard errors that are robust to heteroskedasticity and contemporaneous correlation (clustered by month). Please see Appendix C for additional information about statistical inference and clustered standard errors.

Regression 1 in Table IV, Panel B shows that the average excess return is 105.8 basis points (bp) per month. Regression 2 shows that investors' portfolios of local holdings outperform the market (R^*_m) by only 7.3 bp per month after adjusting for market beta. Regression 3 shows that the local holdings

Table III
Variable Definitions

This table reports variable definitions used Tables IV–VI.

Variable	Definition
R_f	The risk-free return from Ken French's website.
$R_{local,i}$–R_f	The excess return of the local portion of investor *i*'s portfolio. When this is the dependent variable, regressions are run at the investor level.
$R_{local,z}$–R_f	The excess return of a portfolio that aggregates the local holdings of all investors living at a given zip code. When this is the dependent variable, regressions are run at the zip code level.
$R_{local,i}$–$R_{remote,i}$	The return of the local portion of investor *i*'s portfolio minus the remote portion of the investor's portfolio. When this is the dependent variable, regressions are run at the investor level.
R^*_m–R_f	The value-weighted excess market return for all stocks with zip code data. This is a passive index.
R_z–R_f	The excess return of a passive zip code level index. Contains all stocks headquartered within a 250-mile radius of a given zip code.
R_z–$R^*{}_{m\setminus z}$	The difference between two passive indices. The first is a passive zip code level index. The second is the value-weighted return of a passive remote index, and is defined differently for each zip code. The "m \ z" portfolio contains all stocks in the market except those stocks headquartered within a 250-mile radius of a given zip code.
SMB	The small-minus-big portfolio from Ken French's website.
HML	The high-minus-low (book-to-market) portfolio from Ken French's website.
MOM	The momentum (winners-minus-losers) portfolio from Ken French's website.

outperform the respective passive, zip code–level index (R_z) by 6.9 bp per month, with 0.4 *t*-statistic. Note that had we not used clustered standard errors, we would have (erroneously) reported a *t*-statistic of 6.6 for the constant in Regression 3. The last regression in Table IV, Regression 7, contains all control variables and shows an alpha of 6.7 bp per month. Outperformance of 6.7 bp per month translates to an alpha of 0.8% per annum, which is not economically significant.

We aggregate households' returns into value-weighted portfolios at the zip code level to check the robustness of our results. There are at least two advantages of this alternative approach. First, zip code–level analysis helps alleviate potential influence from outliers at the household level (Pitfall #2). Second, compared with individual-level analysis, the zip code–level analysis assigns equal weight to each geographic area. This helps circumvent problems related to geographic selection biases and the distribution of retail investors (Pitfall #3). Table V reports the regression results. After controlling for market level–or zip code–level returns, the alphas in the table are not significantly different from zero, with *t*-statistics that range from 0.11 to 0.76 in Regressions 2 to 7. The alphas are economically small as well.

Our next set of results involves holdings-based portfolios and the difference between an individual's local versus remote portfolio returns. We use this difference portfolio as the dependent variable in our regression analysis. We

Table IV

Holdings-Based Calendar-Time Portfolios

This table reports results from analysis using holdings-based calendar-time portfolios. Panel A reports summary statistics of monthly return variables. Panel B reports pooled regression results with $R_{local,i} - R_f$ as the dependent variable, where $R_{local,i} - R_f$ is the monthly excess return of an individual's local holdings. Local stocks are defined as those headquartered within 250 miles of an investor's home. Variable definitions can be found in Table III. t-statistics are based on Rogers (1993) standard errors (clustered by month) and are robust to heteroskedasticity.

Panel A: Summary Statistics of Monthly Returns

	Mean (%)	*SD* (%)
$R_{local,i}$	1.4042	10.66
R^*_m	1.2511	2.88
R_z	1.3328	3.53
R_f	0.3460	0.09
$R_{local,i}$–R_f	1.0582	10.66
$R_{local,i}$–R^*_m	0.1531	10.19
$R_{local,i}$–R_z	0.0714	10.06

Panel B: Regressions with $R_{local,i}$–R_f as the Dependent Variable

	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7
Alpha (bp)	105.82	7.25	6.87	5.57	6.80	11.40	6.72
(t-stat)	*(2.42)*	*(0.34)*	*(0.40)*	*(0.31)*	*(0.50)*	*(1.00)*	*(0.56)*
R^*_m–R_f		1.0891		0.0642	1.1395		0.1961
(t-stat)		*(12.36)*		*(0.58)*	*(30.02)*		*(4.49)*
R_z–R_f			1.0027	0.9570		1.0122	0.8804
(t-stat)			*(20.54)*	*(18.21)*		*(38.34)*	*(29.00)*
SMB					0.5093	0.4227	0.4338
(t-stat)					*(8.01)*	*(7.90)*	*(8.11)*
HML					0.0863	0.0533	0.0776
(t-stat)					*(1.38)*	*(1.15)*	*(1.55)*
MOM					−0.2317	−0.2034	−0.2208
(t-stat)					*(−5.10)*	*(−5.26)*	*(−5.92)*
# of Obs.	938,644	938,644	938,644	938,644	938,644	938,644	938,644
# of Months	71	71	71	71	71	71	71

do not necessarily advocate forming long-short portfolios based on individual holdings because individuals rarely short stocks. That said, forming such a portfolio allows arguably the closest comparison with results in Table V of the Ivkovic and Weisbenner (2005) paper. Our results using local-minus-remote holdings are shown in Table VI of this paper.

In our Table VI, we regress the local-minus-remote returns ($R_{local,i} - R_{remote,i}$) on the difference between passive local and remote portfolios ($R_z - R^*_{m\setminus z}$). The passive local portfolio is denoted R_z and is formed at the investor's zip code level. The passive, remote portfolio is denoted $R^*_{m\setminus z}$ and contains all stocks not in certain radius of the investor. Our Table VI, Regression 3 shows an alpha of 0.9 with a 0.2 t-statistic. The alpha is neither economically nor statistically

Table V

Holdings-Based Calendar-Time Portfolios: Aggregated at the Zip Code Level

This table reports results from analysis of holdings-based calendar-time portfolios. Panel A reports summary statistics of monthly return variables. Panel B reports pooled regression results with $R_{local,z}$–R_f as the dependent variable. Variable definitions can be found in Table III. *t*-statistics are based on Rogers (1993) standard errors (clustered by month) and are robust to heteroskedasticity.

Panel A: Overview Statistics of Monthly Returns

	Mean (%)	SD (%)
$R_{local,z}$	1.3320	8.94
R^*_m	1.2715	2.87
R_z	1.3075	3.37
R_f	0.3523	0.09
$R_{local,z}$–R_f	0.9797	8.93
$R_{local,z}$–R^*_m	0.0605	8.44
$R_{local,z}$–R_z	0.0245	8.31

Panel B: Regressions with $R_{local,z}$–R_f as the Dependent Variable

	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7
Alpha (bp)	97.97	3.08	4.87	1.51	2.26	7.18	1.25
(t-stat)	(*2.55*)	(*0.21*)	(*0.37*)	(*0.11*)	(*0.22*)	(*0.76*)	(*0.13*)
R^*_m–R_f		1.0324		0.1558	1.0732		0.2371
(t-stat)		(*15.91*)		(*1.98*)	(*37.06*)		(*6.13*)
R_z–R_f			0.9746	0.8599		0.9842	0.8197
(t-stat)			(*24.80*)	(*26.69*)		(*44.88*)	(*33.70*)
SMB					0.3819	0.3278	0.3372
(t-stat)					(*8.30*)	(*7.39*)	(*7.89*)
HML					0.0911	0.0547	0.0858
(t-stat)					(*2.03*)	(*1.39*)	(*2.09*)
MOM					−0.1734	−0.1440	−0.1642
(t-stat)					(*−5.05*)	(*−4.06*)	(*−5.10*)
# of Obs	369,219	369,219	369,219	369,219	369,219	369,219	369,219
# of Months	71	71	71	71	71	71	71

different from zero. Our Table VI provides additional and compelling evidence that an individual's local holdings do not outperform his remote holdings.

The Internet Appendix contains robustness checks using regression analysis and holdings-based calendar-time portfolios. As mentioned earlier, our first three robustness checks define local stocks as those within a 100-mile radius, those within a 100-km radius, and those that are headquartered within the same state as the investors. In a fourth check, we consider only states with a large number of individuals and firms. Our goal is to avoid states with few investors and/or few firms overly influencing the regression results. Appendix D discusses forming holdings-based portfolios at the zip code level and the state level. Fifth, and finally, we estimate a more flexible state-level regression that allows market betas to vary by state. Results of these robustness checks are

Table VI

Holdings-Based Calendar-Time Portfolios: Local–Remote on Left Hand Side

The dependent variable is the monthly return of a long-short holdings-based calendar-time portfolio. The long-short portfolio is the return of locals stocks held by an investor in our sample minus the return of remote stocks held by the same investor ($R_{local,i} - R_{remote,i}$). Local stocks are defined as those headquartered within a 250-mile radius of an investor's home. Remote stocks are defined as those headquartered outside a 250-mile radius of an investor's home. Variable definitions can be found in Table III. t-statistics are based on Rogers (1993) standard errors (clustered by month) and are robust to heteroskedasticity.

Panel A: Summary Statistics of Monthly Returns

	Mean (%)	*SD* (%)
$R_{local,i} - R_{remote,i}$	0.1068	12.91
$R^*_{m\setminus z}$	1.2478	2.84
R_z	1.3206	3.48
R_f	0.3454	0.09

Panel B: Regressions with $R_{local,i} - R_{remote,i}$ as the Dependent Variable

	*Reg*1	*Reg*2	*Reg*3
Alpha (bp)	10.68	5.26	0.85
(t-stat)	(*2.20*)	(*1.21*)	(*0.18*)
R_z–$R^*_{m\setminus z}$		0.7435	0.7431
(t-stat)		(*24.51*)	(*24.64*)
SMB			0.0375
(t-stat)			(*2.20*)
HML			0.0550
(t-stat)			(*3.13*)
MOM			0.0058
(t-stat)			(*0.38*)
# of Obs	647, 899	647, 899	647, 899
# of Months	71	71	71

consistent with the results presented in Tables IV–VI in the main paper.

In sum, Tables IV–VI paint a very clear picture. After controlling for passive state-level indices and/or Fama-French-Carhart factors, individual investors do not generate a significantly positive alpha with their local holdings. In Table IV, Regression 7, the alpha is 6.7 bp per month or 0.8% per annum with a 0.6 t-statistic. In Table V, Regression 7, the alpha is 1.3 bp per month or 0.2% per annum with a 0.1 t-statistic. In Table VI, Regression 3, the alpha is 0.9 bp per month or 0.1% per annum with a 0.2 t-statistic.

B. Transactions-Based Calendar-Time Portfolios

We test whether purchases of local stocks predict future positive returns and whether sales predict future negative returns. Our methodology uses transactions-based calendar-time portfolios to aggregate the trades of many individuals and control for cross-sectional correlation of stock returns.[9] Stocks are held in the portfolios for 1 year, which is close to the average holding period of the investors in our sample. Our data contain hundreds of transactions each day and returns are calculated on a daily basis. Despite using daily returns, our calendar-time portfolios are free from potential microstructure effects. Please see Appendix E for additional descriptions of transactions-based calendar-time portfolios.

The returns of our transactions-based calendar-time portfolios have two natural economic interpretations. First, they represent the returns experienced by an investor who mimics the trades of individuals in our data and holds stocks for a set period of time (i.e., 1 year). Second, they represent the opposite returns of an investor (such as an institution) who trades against individuals in our data and who holds stocks for a year. If individuals lose 2% per year, an investor trading against the individuals stands to make 2% per year. All returns are calculated before transactions costs.

Table VII presents the transactions-based calendar-time portfolio results. In Panel A, the calendar-time "Buys" portfolio has an average return of 6.1 basis points per day. The calendar-time "Sells" portfolio has an average return of 6.9 bp per day. The return difference (Buys-minus-Sells) is −0.8 bp per day, which amounts to −2.0% per annum. This difference is statistically significant with a −2.5 t-statistic.

We calculate abnormal returns (or "Alpha") for the transactions-based calendar-time portfolios by regressing the returns of the Buys-minus-Sells portfolio on a constant and the market's excess returns. Statistical inference is straightforward as there is only a single time series of returns. t-statistics are based on Newey-West standard errors. Please see Appendix C for additional descriptions of statistical inference. In Table VII, Panel A, the alpha is −2.2% per annum with a −2.7 t-statistic.

Table VII, Panel B classifies stocks by location. Local buys minus local sells have an average return of −1.7% per annum, with a −1.8 t-statistic. We conclude that locals buys underperform local sells and the difference is statistically significant at the 7% level.

C. Stocks with Lower/Higher Levels of Information Asymmetries

We test whether trades of individual investors have value-relevant information for stocks with lower/higher levels of information asymmetries. As discussed in the Section I, the S&P 500 index is used to help identify stocks with different levels of information asymmetries. Stocks not in the index are identified as those with higher levels of information asymmetries.

[9]Events-based and transactions-based calendar-time portfolios have been used in a number of applications. Barber and Lyon (1997) and Lyon, Barber, and Tsai (1999) detail statistical properties relating to long-run tests of abnormal stock returns. Brav and Gompers (1997) measure returns to investing in initial public offerings. Jeng, Metrick, and Zeckhauser (2003) use calendar-time portfolios to study the returns to insider trading.

Table VII

Transactions-Based Calendar-Time Portfolios

This table shows average returns of transactions-based calendar-time portfolios. Portfolios are formed by mimicking the trades of all investors in our sample between 1991 and 1996. Stocks are held in a calendar-time portfolio for 1 year. For a given group of stocks, we form one calendar-time portfolio based on stocks bought ("Buys") and another portfolio based on stocks sold ("Sells"). We show the difference of returns between the Buys and Sells portfolios ("Diff") in both basis points per day and annualized in percentages. The "Alpha" reports the annualized constant from a regression of the Buys-minus-Sells portfolio returns on the market's excess returns. Local stocks are defined as being headquartered within a 250-mile radius of an investor's home. In Panels C and D, we consider whether or not a stock is part of the S&P500 Index. t-statistics are based on Newey-West standard errors with five lags and robust to heteroskedasticity and serial correlation of residuals.

	Average Returns (bp/Day)			Annual Diff		Alpha	
	Buys	Sells	Diff	%	t-stat	%	t-stat
Panel A: All Stocks							
All	6.083	6.889	−0.805	−2.01%	*−2.50*	−2.16%	*−2.66*
Panel B: Sorted by Location							
Local	6.725	7.421	−0.696	−1.74%	*−1.81*	−1.66%	*−1.69*
Remote	6.072	6.878	−0.806	−2.01%	*−2.46*	−2.25%	*−2.72*
Panel C: S&P500 Stocks and Location							
Local	7.260	7.638	−0.378	−0.95%	*−0.64*	−0.87%	*−0.58*
Remote	6.324	7.153	−0.829	−2.07%	*−1.67*	−2.29%	*−1.83*
Panel D: Non-S&P500 Stocks and Location							
Local	6.313	7.215	−0.902	−2.25%	*−2.49*	−2.11%	*−2.28*
Remote	5.736	6.603	−0.867	−2.16%	*−2.83*	−2.43%	*−3.20*

While Table VII, Panel C reports results related to S&P 500 stocks, we focus on Panel D, which reports the average Buys-minus-Sells return of local non-S&P 500. The average return is -2.3% per annum with a -2.5 t-statistic. We note that results involving the transactions-based calendar-time portfolios do not suffer from low-powered tests. We show economically and statistically significant underperformance.

The Internet Appendix contains 10 tables that use alternative specifications and robustness checks. In particular, we define local using 100-mile and 100-km radii, we form equally weighted transactions-based calendar-time portfolios, and we consider holding periods of 3 months and 6 months. The results are consistent with those presented in Table VII of the main paper.

As noted in the introduction to this paper, our results involving local non-S&P 500 stocks are among the paper's most important findings. If individuals have any valuable information, we hypothesize that the information is likely to be about local stocks with high levels of information asymmetries. However,

Table VIII

Transactions-Based Calendar-Time Portfolios by Year

This table expands the average daily calendar-time returns shown in Table VI (Panels A and B). We report the average return of the Buys-minus-Sells calendar-time portfolio in basis points per day. Average returns are calculated during each of 6 calendar years. Portfolios are formed based on mimicking the trades of all investors in our sample. Numbers in the column labeled "All Years" match the results shown Table VII, Panels A and B of this paper. The earlier table provides measures of statistical significance.

Buys-minus-Sells	1991	1992	1993	1994	1995	1996	All Years
All stocks	0.663	−1.359	−1.793	−1.400	−0.629	0.165	−0.805
Local stocks	1.379	−0.856	−1.649	−2.258	−0.003	−0.385	−0.696
Remote stocks	0.483	−1.432	−1.757	−1.064	−0.865	0.278	−0.806

when we focus only on these types of stocks, individual buys do not predict future price increases. Individual sells do not predict future price decreases. In fact, we see the opposite.

D. Different Time Periods

We measure time-variation in the transactions-based results. Table VIII shows the average daily return in basis points for the Buys-minus-Sells portfolios over each of the 6 calendar years. Buys of local stocks outperform sells by 1.4 bp per day in 1991. The local buys and local sells portfolios have roughly similar returns in 1995. Over the remaining years, local buys underperform local sells with the right-hand column in Table VIII showing an average −0.7 bp per day underperformance over all years (matching results shown in Table VII, Panels A and B). The strong outperformance of local stocks in 1991 might account for the conclusions in Ivkovic and Weisbenner (2005), as these authors define locality based on 1991 holdings.

E. Individual Trades and Contemporaneous Returns

We end our empirical analysis by checking whether individuals tend to buy (sell) stocks on days when prices are rising or falling. Due to data limitations, the contemporaneous daily relationship between trades and prices represents the highest frequency we can study. If individuals have value-relevant information, we expect prices to go up on days individuals are buying (or prices may go up in the immediate future). Likewise, prices should go down on days individuals are selling (or prices may go down in the immediate future).

We again form two calendar-time portfolios, namely a "Buys" portfolio and a "Sells" portfolio. We measure the value-weighted return on the day individuals trade, that is stocks enter the portfolios for 1 day only, and we specifically do not skip a day after trade dates as we want to measure a contemporaneous effect. As before, we again form a Buys-minus-Sells transactions-based calendar-time portfolio. We find an average daily return of −57.4 bp, which is significantly different from zero at all conventional levels. The negative sign might indicate that individuals actively buy as prices are falling and sell as prices are rising. Alternatively, individuals might have stale limit orders. As prices fall, individuals are "picked off" and thus appear to be selling into a falling market. The result in this section is consistent with findings in Kaniel et al. (2008). We do not see prices going up on days individuals are buying, nor do we see prices falling on days individuals are selling. Rather, we find the opposite. This test thus provides further evidence that individuals do not have value-relevant information.

III. Conclusions

It is well documented that individuals tilt their portfolios towards locally headquartered stocks. Yet, despite strong evidence from United States and international studies, a question remains: Why do investors have a preference for local stocks? Recent papers by Ivkovic and Weisbenner (2005) and Massa and Simonov (2006) suggest that individuals may exploit local information. If this is true, one might rationalize tilting portfolios away from the market portfolio and towards local stocks.

Our paper tests whether individuals earn superior returns on their local investments. Our methodology involves forming calendar-time (aggregate) portfolios based on holdings and transactions. We classify holdings and transactions based on whether the stock is local or remote for a given investor. The use of both holdings-based and transactions-based calendar-time portfolios is one aspect that sets our paper apart from prior work.

The calendar-time portfolios address four potential pitfalls that arise when studying individuals' investments and geography. These pitfalls are: 1) cross-sectional correlation of portfolio returns, 2) small stock effects in individual portfolios, 3) geographic selection biases, and 4) time-series selection biases.

We analyze holdings of local stocks using standard portfolio analysis. After including the market's excess return and the excess return of a passive local index, investors are found not to generate abnormally high returns. The average alpha is 0.8% per annum with a 0.6 t-statistic. The finding that individuals do not generate abnormally high returns differs from extant studies.

We next use transactions-based calendar-time portfolios and find that buys of local stocks do not predict positive returns and sells of local stocks do not predict negative returns. In fact, a local Buys-minus-Sells portfolio earns -1.7% per annum with a -1.8 t-statistic. We hypothesize that if individuals have valuable information about a stock, it is likely that the stock is local and has a high level of information asymmetries (i.e., the stock is not in the S&P 500 index). Even when we constrain our analysis to only these local, non-S&P 500 stocks, buys do not outperform sells. The Buys-minus-Sells portfolio has a return of -2.3% per annum with a -2.5 t-statistic. The finding that local buys underperform local sells directly contradicts conclusions reached in existing papers.

Our results are robust to many alternative specifications. For instance, we divide the time series of transactions-based calendar-time portfolio returns by calendar year; we consider definitions of local and remote based on radii of 100 miles, 100 km, and state boundaries; and we consider alternative measures of information asymmetry that include dollar trading volume and number of analysts following a stock. The supplementary Internet Appendix contains 17 pages of additional tables. The results are qualitatively similar and provide additional support for our conclusions.

Our findings point to indexing as a straightforward solution to the perils faced by individual investors. An investor who indexes can minimize transaction costs and avoid losses associated with trading individual stocks. On average, individuals do not have value-relevant information about the local stocks they hold and trade.

Appendix A: Cross-sectional Correlation of Individuals' Portfolio Returns

When studying returns to individual investors' portfolios, the number of independent observations is often related to the number of stocks in the market and not to the number of investors in the data set, which can cause confusion. This pitfall is labeled Pitfall #1 in the paper's introduction. Since the number of independent observations is a quantity that affects statistical inference, we provide additional insights by way of an example.

Consider the investment decisions of 40,000 individuals in a market with only two stocks. Suppose that at the beginning of the year, each individual independently decides what fraction $\alpha_{A,i}$ of his portfolio to invest in Stock A. The remaining fraction of his portfolio ($\alpha_{B,i} = 1 - \alpha_{A,i}$) is invested in Stock B. Next record each of the 40,000 portfolio returns ($r_{p,i}$) at the end of the year. Assume no rebalancing for simplicity. Since $r_{p,i} = r_B + \alpha_{A,i}(r_A - r_B)$, financial economists will notice a relationship between $r_{p,i}$ and $\alpha_{A,i}$ provided the two stocks have different returns over the year.

How statistically significant is the observed relationship between an individual's portfolio returns ($r_{p,i}$) and his portfolio choice ($\alpha_{A,i}$) in the previous example? Even if all 40,000 individuals make independent portfolio decisions, there are not 40,000 independent observations. In this example, there are only two independent observations and a financial economist has no ability to reliably conclude whether increased investments in one type of stock (e.g., local stocks) enhance an investor's returns. In more realistic examples, the number of independent stock returns increases as the number of different stocks in investors' portfolios increases (though there must be corrections if an underlying factor structure determines returns). Using calendar-time portfolios and clustered standard errors allows us to account for cross-sectional correlation of stock returns.

Appendix B: Small Stocks and Individual Portfolios

When studying investor-level portfolios, small stocks can overly influence results as some portfolios consist of very few stocks, and it is possible that a single stock constitutes the local part of an investor's portfolio.

During our sample period, CRSP reports 192 instances of stock returns over 200% in a given month. Incredibly, one stock reports a single month's return of 1,250%. When looking at the firms associated with these high return events, the average market capitalization at the beginning of the month is less than US$ 7.5 million. Half of the events are associated with stocks that have a market capitalization of less than US$ 2.5 million. Not surprisingly, small stocks can (on occasion) have extreme returns.

To understand how extreme returns can affect large data samples, consider the following example. The local holdings-based portfolios shown in Table IV have an average return of 1.4% per month. Our data consists of 938,644 observations (or household-months). If only 75 of the stock-months had a return of 1,250%, the average would go up 1.5% per month. Note that 75 out of 938,644 represents less than 0.008% of the sample.

As mentioned in the introduction, we address possible small-stock biases by forming portfolios that combine and value-weight the holdings of many investors. We also use transactions-based portfolios that combine all purchases or sales into a single portfolio.

Appendix C: Statistical Inference and Calendar-Time Portfolios

Holdings-Based Calendar-Time Portfolios: We test whether portfolios of local stock holdings have abnormally high returns. The portfolios are formed at the individual level. For a given individual i, we have up to 71 monthly observations. Table IV in the paper, for example, estimates stacked (pooled) regressions using ordinary least squares.

Returns of portfolios formed from investors i and j may be contemporaneously correlated. To account for this cross-sectional dependence, we use Rogers (1993) standard errors that are clustered by month—see Rogers (1993), Petersen (2009), and Cameron, Gelbach, and Miller (2010) for details regarding cluster-robust standard errors. The variance of the estimator is

$$\mathrm{Var}(\hat{\beta}) = (X'X)^{-1}\left[\sum_{g=1}^{G} X'_g \hat{u}_g \hat{u}'_g X_g\right](X'X)^{-1}$$

$$\hat{u}_g = y_g X_g(\hat{\beta}),$$

where G is the number of clusters. In our paper, $G = 71$ since we have 71 months of data.

Transactions-Based Calendar-Time Portfolios: Our transactions-based calendar-time portfolios produce single time series of daily returns. Statistical inference is straightforward. The difference of two calendar-time portfolios is regressed on a constant and the market's excess returns. Serial correlation and heteroskedasticity of residuals are addressed with Newey-West standard errors with five lags. The regression equation is

$$R_{Buy-Sell,t} = \alpha_p + \beta_p(R_{m,t} - R_{f,t}) + \varepsilon_t,$$

where $R_{Buy-Sell,t}$ is the difference between two daily calendar-time portfolios, $R_{m,t}$ is the daily CRSP value-weighted return, and $R_{f,t}$ is the daily return on the risk free asset.

Appendix D: Holdings-Based Calendar-Time Portfolios Formed at the Zip Code Level and State Level

To overcome possible biases caused by positions in small/low-priced stocks (Pitfall #2), investors are aggregated by the zip code or state in which they live. When focusing on zip codes, we create a portfolio that contains the local investments of all investors living in a given zip code. Local investments are defined using a 250-mile radius as described in the paper. When focusing on states, local stocks are defined as those headquartered in the state where an individual lives. The return of an aggregate portfolio is the monthly value-weighted return of stocks in the portfolio. Portfolio weights are determined by the values of holdings at the beginning of each month. We consider all zip codes with investors. The net result is 7,832 time series of local (zip code-level) returns. With states, we start with all 50 states plus the District of Columbia, but exclude Delaware, Maine, Montana, New Mexico, North Dakota, and Wyoming due to insufficient investor and/or firm data. The net result is 45 time series of local (state-level) returns.

Holdings-based calendar-time portfolios have the advantage that an investor's actual holding period is reflected in the time series of portfolio returns. Consider a household in Seattle that buys Microsoft stock on March 29, 1991 and sells the stock on February 28, 1992. Microsoft stock will be in Washington state's local portfolio for the 11 months the investor holds the stock. The disadvantage is that holdings are only sampled once a month. To address possible geographic sampling biases (Pitfall #3), we create passive zip code–specific or state-specific indices (R_z or R_s, respectively) and require investors to outperform these indices if we are to conclude they have value-relevant information.

We estimate pooled regressions using ordinary least squares. When aggregating by zip code the dependent variable has dimensions $369,219 \times 1$. When we use data from 45 states, the stacked dependent variable has dimensions $3,190 \times 1$, indicating an average of 70.9 months of data per state. To account for cross-sectional correlation of zip code *i*'s and zip code *j*'s returns (or state *i*'s and state *j*'s returns), *t*-statistics are based on clustered standard errors (clustered by month). Table V shows results when investors are aggregated by zip code. The alpha associated with local holdings remains insignificantly different from zero. The point estimate is only 1.3 bp per month in Regression 7. The Internet Appendix shows results when investors and firms are classified by state rather than by zip code. The alpha associated with local holdings remains insignificantly different from zero. What's more, the point estimate is only 1.6 bp per month in Regression 7. This works out to 0.2% per annum.

To further control for possible biases caused by small positions in small/low-priced stocks (Pitfall #2), we reestimate regressions using only returns from the 20 most represented states. See the Internet Appendix. As a final robustness check, we allow all coefficients on the market excess return and the passive state portfolios to vary by state. In the Internet Appendix, the alpha is still constrained to be the same across the 45 states. The alpha associated with local holdings remains insignificantly different from zero. The robustness checks support the conclusions reached in this paper.

Appendix E: Transactions-Based Calendar-Time Portfolios

We mimic the buys and sells of investors by forming "Buys" and "Sells" calendar-time portfolios. Each time an investor buys a stock, we place the same number of shares in our calendar-time "Buy" portfolio. Similarly, each time an investor sells a stock, we place the same number of shares in our calendar-time "Sell" portfolio. Shares are held in a portfolio for a pre-determined length of time. Our strategy of mimicking the number of shares traded is called a value-weighted calendar-time portfolio. A value-weighted calendar-time portfolio refers to buying or selling the same number of shares that individual investors buy or sell. In this way, large transactions initially receive more weight than small transactions. An equal-weighted calendar-time portfolio refers to initially buying (selling) \$1 of each stock bought (sold). Buying (selling) \$1 of a stock corresponds to buying (selling) $\$1 \div P_t$ shares of the stock, where P_t is the share price in dollars.

The value of shares held in a portfolio changes as the stock price goes up and down. Thus, both value-weighted and equal-weighted calendar-time portfolios account for changes in stock prices. Both the value-weighted and equal-weighted calendar-time portfolios calculate the weighted average return of stocks in the portfolio each day. The main difference between the two types of portfolios is that a position in the equal-weighted portfolio starts at \$1 worth of shares while a position in the value-weighted portfolio starts at the value of shares actually bought by individuals in our data set.

The tables in the main paper consider a 1-year holding period, which is close to the average holding period for investors in our data. The Internet Appendix gives results for holding periods of 3 months, 6 months, and 1 year. The Internet Appendix also shows results from an equal-weighted calendar-time portfolio with a 1-year holding period.

As investors in our data set buy stocks (for example), the composition of our "calendar-time buys portfolio" changes. The return of the portfolio on any day is the weighted average return of the portfolio on that day. Weights are determined by the beginning of the day values of shares held in the portfolio. Share values are recalculated each day based on close-of-market prices from the previous day. Importantly, we do not change the number of shares (of a single position) over its holding period. We also do not rebalance the portfolio (other than adding stocks and holding them for the pre-determined length of time). We follow a buy-and-hold strategy. A calendar-time portfolio produces a single time series of returns. The total length of the time series is equal to the number of days between the first and last transaction in the sample plus the holding period (i.e., approximately 7 years long in our case.)

Our transactions-based calendar-time portfolios address potential microstructure effects in two ways. First, we place shares in our portfolios at the end of the day after an individual actually buys or sells the shares. In other words, we skip a day between the actual purchase/sale date and when we start recording prices. If buys take place at the ask and sales at the bid, we avoid any mean-reversion induced by the bid-ask spread. Second, our economic results and conclusions are based on the differences between two calendar-time portfolios (buys minus sells.) Since each portfolio consists of hundreds or thousands of stocks at any point in time, our returns are not subject to the bias described in Blume and Stambaugh (1983). That is, bid-ask bounce does not bias returns due to volatility. Our calendar-time portfolios diversify away individual stock bid-ask bounce.

REFERENCES

Admati, Anat R., 1985, A noisy rational expectations equilibrium for multi-asset securities markets, *Econometrica* 5, 629–658.

Barber, Brad M., and John D. Lyon, 1997, Detecting long-run abnormal stock returns: The empirical power and specification of test statistics, *Journal of Financial Economics* 43, 341–372.

Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773–806.

Barber, Brad M., and Terrance Odean, 2001, Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics* 116, 261–292.

Barber, Brad M., and Terrance Odean, 2002, Online investors: Do the slow die first? *Review of Financial Studies* 15, 455–489.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Do retail trades move markets? *Review of Financial Studies* 22, 151–186.

Barclay, Michael J., Terrance Hendershott, and Charles M. Jones, 2008, Order consolidation, price efficiency, and extreme liquidity shocks, *Journal of Financial and Quantitative Analysis* 43, 93–121.

Blume, Marshall E., and Robert F. Stambaugh, 1983, Biases in computed returns: An application to the size effect, *Journal of Financial Economics* 12, 387–404.

Brav, Alon, and Paul A. Gompers, 1997, Myth or reality? The long-run underperformance of initial public offerings, *Journal of Finance* 52, 1791–1821.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, 2010, Robust inference with multiway clustering, *Journal of Business and Economic Statistics* (forthcoming).

Coval, Joshua D., David A. Hirshleifer, and Tyler Shumway, 2005, Can individual investors beat the market? Working paper, Harvard University.

Coval, Joshua D., and Tobias J. Moskowitz, 2001, The geography of investment: Informed trading and asset prices, *Journal of Political Economy* 109, 811–841.

Dhar, Ravi, and Ning Zhu, 2006, Up close and personal: Investor sophistication and the disposition effect, *Management Science* 52, 726–740.

Diamond, Douglas W., and Robert E. Verrecchia, 1981, Information aggregation in a noisy rational expectations equilibrium, *Journal of Financial Economics* 9, 221–235.

Feng, Lei, and Mark S. Seasholes, 2004, Portfolio choice and location of trade, Working paper, U.C. Berkeley.

Goetzmann, William N., and Alok Kumar, 2008, Equity portfolio diversification, *Review of Finance* 12, 433–463.

Graham, John R., and Alok Kumar, 2006, Do dividend clienteles exist? Evidence on dividend preferences of retail investors, *Journal of Finance* 61, 1305–1336.

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

Hau, Harald, 2001, Location matters: An examination of trading profits, *Journal of Finance* 56, 1959–1983.

Hellwig, Martin F., 1980, On the aggregation of information in competitive markets, *Journal of Economic Theory* 22, 477–498.

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

Ivkovic, Zoran, James Poterba, and Scott Weisbenner, 2005, Tax-motivated trading by individual investors, *American Economic Review* 95, 1605–1630.

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

Jeng, Leslie A., Andrew Metrick, and Richard Zeckhauser, 2003, Estimating the returns to insider trading: A performance-evaluation perspective, *Review of Economics and Statistics* 85, 453–471.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2008, Individual investor trading and stock returns, *Journal of Finance* 63, 273–310.

Kumar, Alok, 2004, Is the local bias of individual investors induced by familiarity or information-asymmetry? Working paper, Notre Dame University.

Kumar, Alok, 2009a, Dynamic style preferences of individual investors and stock returns, *Journal of Financial and Quantitative Analysis* 44, 607–640.

Kumar, Alok, 2009b, Hard-to-value stocks, behavioral biases, and informed trading, *Journal of Financial and Quantitative Analysis* 44, 1375–1401.

Kumar, Alok, and Charles M. C. Lee, 2006, Retail investor sentiment and return comovements, *Journal of Finance* 61, 2451–2486.

Kumar, Alok, and Sonya Lim, 2008, How do decision frames influence the stock investment choices of individual investors? *Management Science* 54, 1052–1064.

Lyon, John D., Brad M. Barber, and Chih-Ling Tsai, 1999, Improved methods for tests of long-run abnormal stock returns, *Journal of Finance* 54, 165–201.

Massa, Massimo, and Andrei Simonov, 2006, Hedging, familiarity, and portfolio choice, *Review of Financial Studies* 19, 633–685.

Odean, Terrance, 1998, Are investors reluctant to realize their losses? *Journal of Finance* 53, 1775–1798.

Odean, Terrance, 1999, Do investors trade too much? *American Economic Review* December, 1279–1298.

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

Rogers, William H., 1993, Regression standard errors in clustered samples, *Stata Technical Bulletin* 13, 19–23.

Zhu, Ning, 2003, The local bias of individual investors, Working paper, Yale University.