Received: 12 November 2019 | Accepted: 30 November 2019
DOI: 10.1002/rfe.1093

ORIGINAL ARTICLE

REVIEW of FINANCIAL ECONOMICS WILEY

Investor sophistication and asset prices

George M. Korniotis[1] | **Alok Kumar**[1] | **Jeremy K. Page**[2]

[1]Department of Finance, University of Miami, Coral Gables, FL, USA
[2]Adobe Inc, Provo, UT, USA

Correspondence
Alok Kumar, Department of Finance, University of Miami Herbert Business School, 514 Jenkins Building, Coral Gables, FL 33124, USA.
Email: akumar@miami.edu

Abstract
We show that geographical variation in the level of investor sophistication influences local asset prices. Investors in less sophisticated regions exhibit stronger trading correlations, and correspondingly, the returns of firms headquartered in less sophisticated areas are more strongly correlated. Furthermore, we show that local economic conditions have a greater ability to predict local stock returns in the U.S. states with less sophisticated retail investors. These asset pricing results are driven by the sophistication of actual local investors, and not by the characteristics of the broader local population.

KEYWORDS
investor sophistication, local return predictability, return comovement, state portfolios, trading correlation

JEL CLASSIFICATION
G11; G12

1 | INTRODUCTION

Investor sophistication plays a central role in behavioral asset pricing models.[1] These models posit that demand shocks generated by the systematic behavior of less sophisticated investors, the noise traders, can move prices away from their intrinsic values. If arbitrage costs are high, arbitrage forces may not be very effective and prices may not move back to their fundamental values quickly. Consequently, stock prices may exhibit excess comovement and higher degree of predictability (e.g., Alti & Tetlock, 2014). Although these theoretical models predict that investor sophistication would generate asset mispricing, there is little empirical evidence to support this idea because it is difficult to measure investor sophistication directly. In this paper, we use an investor-level dataset and construct a direct measure of investor sophistication, which we subsequently relate to mispricing among stocks. Specifically, we examine if the degree of mispricing is stronger in the U.S. states with less sophisticated investors.

The potential relation between local mispricing and local sophistication is motivated by prior evidence on the geographical segmentation in the U.S. capital markets and the cross-sectional variation in investor sophistication across regions. In particular, Korniotis (2008) provides evidence of market segmentation. He demonstrates that since income shocks across the U.S. states are not diversifiable (Asdrubali, Sorensen, & Yosha, 1996; Athanasoulis & van Wincoop, 2001), market returns can be better explained using a model that treats the U.S. economy as a collection of 50 state-level investors instead of one national representative investor. Furthermore, Becker (2007) finds that the market for bank loans and deposits is segmented, while Becker, Ivkovich, and Weisbenner (2010) demonstrate that location of investors influence the demand for dividend-paying stocks. More recently, Korniotis and Kumar (2013a) demonstrate that ownership and trading of local stocks has a significant local component.

Not only the U.S. capital markets are segmented, there is evidence that investor sophistication and behavioral biases vary geographically. For example, Kumar (2009) and Kumar, Page, and Spalt (2011) find that investors' preference for volatile, "lottery-like" stocks differs across the U.S. states. In addition, Korniotis and Kumar (2011) show that investors' ability to smooth income shocks through their financial decision varies geographically.

Motivated by these prior studies, we conjecture that the degree of mispricing would vary geographically since the U.S. capital markets are segmented and investor sophistication varies across regions. To examine this local-mispricing local-sophistication hypothesis, we examine whether local stocks in less sophisticated areas *comove* more with each other and if they are more *predictable*.[2] Comovement among local stocks can arise since the local economy is more salient to local investors than the national economy. Because local risk aversion is affected by local economic conditions (Korniotis & Kumar, 2013a), local economic conditions can become a coordinating mechanism for the trading behavior of less sophisticated investors. The coordinated trading would generate a systematic component to local trading, which would in turn lead to stronger excess return comovement among the stocks of local firms.

Local return predictability can also arise because state-level economic conditions, a source of coordinated local demand shocks, are persistent. Such persistence could create a slow-moving component in the investment mistakes of local investors. For example, prolonged periods of good local economic conditions can lead to persistent local investor optimism and persistent upward demand pressures. If arbitrage forces are limited, the persistent local demand shocks can lead to return predictability (e.g., Alti & Tetlock, 2014). Therefore, in the U.S. states with less sophisticated investors, local economic conditions should be strong predictors for the returns of local stocks.

To test the comovement and predictability hypotheses, we measure investor sophistication for each U.S. state. The sophistication measure is based on investment decisions of retail investors at a U.S. brokerage house. Using the sophistication levels of *actual* investors rather than the sophistication measures of the *overall* local population is a key innovation of our work. Population demographic variables such as education, income, wealth, and so on, are not a good proxy for the sophistication level of local investors because individuals who choose to participate in the market may be quite different from the overall local population. Our measure is able to account for this self-selection among local investors. In fact, we show that all our results become considerably weaker and insignificant when we use population-wide sophistication measures.

To construct the sophistication index, we focus on retail investors, rather than institutional investors, for two reasons. First, institutional investors are broadly considered sophisticated relative to retail investors (e.g., Barber & Odean, 2008). Second, and relatedly, prior research suggests that correlated retail investor demand can have an impact on stock prices and return comovement, while correlated institutional trading seems to attenuate excess return comovement among stocks (Kumar & Lee, 2006; Kumar, Page, & Spalt, 2013).

Using the sophistication measure, we find supporting evidence for our comovement hypothesis. Specifically, we identify the U.S. state where a firm is headquartered (HQ state). Then, we estimate firm-level annual regressions of daily returns on a return index of all firms in the HQ state controlling for various market factors. The beta on the HQ state index is our measure of comovement. We find that these HQ state betas are higher for states with less sophisticated investors. Moreover, supporting our hypothesis that local economic conditions matter for the comovement of local stocks, we find that the HQ state betas are the highest in periods when the local economic conditions are extremely good or bad.

As a validation test, we examine whether the trading of local stocks is more correlated when local investors are less sophisticated. This is an important test because return comovement should result from correlated local trading. Our measure of local trading correlations is the partial correlation coefficient (controlling for the market return) between the buy–sell imbalance of retail trades of a firm with the state-level buy–sell imbalance of retail trades of other firms headquartered in the same state. We find that the local retail trading correlations (RTCs) are higher for firms in states with less sophisticated investors. Similar to our comovement results, we find that the local trading correlations are the strongest when the local economic conditions are extremely good or bad. Overall, our evidence suggests that investor sophistication enhances comovement in returns and trading among local firms.

Next, we test our return predictability hypothesis, which posits that local economic indicators should predict stock returns more strongly in areas where investors are less sophisticated. Specifically, we estimate return predictability regressions similar to those in Korniotis and Kumar (2013a). The dependent variables are the quarterly characteristic-adjusted state portfolio returns. The independent variables are U.S.-level and state-level macroeconomic variables (income growth, relative unemployment, housing collateral ratio; Lustig & Van Nieuwerburgh, 2005) as well as other U.S.-level macroeconomic variables (e.g., the paper-bill spread, term spread, default spread, the dividend-to-price ratio, and the *cay* measure of Lettau & Ludvigson, 2001a).

We estimate the predictability models separately for high (above median) and low (below median) sophistication states. We find that return predictability is stronger among the U.S. states with less sophisticated investors. Consistent with our hypothesis, we also find that local economic predictors are stronger predictors of local returns in less sophisticated states.

Next, we focus on low sophistication states and test whether return predictability is stronger within states where investors *also* exhibit stronger local bias. We expect that local mispricing to be stronger in those states where investors are less sophisticated *and* primarily hold local stocks. Consistent with our conjecture, return predictability is stronger in these low-sophistication high-local bias states.

To examine the economic significance of the return predictability differences between high and low sophistication states, we define Long–Short trading portfolios where the Long (Short) position is in firms located in states with poor (good) economic conditions. Consistent with our main conjecture, we find that the performance of the Long–Short portfolio is stronger among the subset of states where investor sophistication levels are low. In contrast, the evidence of local return predictability is non-existent in the subset of high investor sophistication states.

For example, during the 1978–2009 period, the Long–Short portfolio has a monthly 4-factor alpha of 0.423% (t statistic = 2.07) when we consider low sophistication states. However, for the subsample of high sophistication states, this portfolio has an insignificant alpha (estimate = −0.144%, t statistic = −1.08). These estimates remain qualitatively similar when we consider various other performance measures, including industry-adjusted returns, alphas from extended factor models that account for liquidity risk and reversals, and alphas from conditional factor models that account for time-varying exposures of state portfolios to U.S. systematic risks. Taken together, these results suggest that in regions where investors are less sophisticated, predictability of local stock returns is considerably stronger.

Our results are robust to alternative explanations. To begin with, our results are not driven by geographical clustering of certain industries that may gravitate toward states with a certain level of sophistication. Specifically, in our comovement regression tests, we control for industry clustering and use industry fixed effects. Furthermore, the alpha estimates of our Long-Short portfolios are significant in factor models that include industry factors.

Furthermore, our results are not influenced by potential endogeneity biases. Endogeneity is a concern because the distribution of sophisticated investors (and sophisticated individuals in general) is not random. This distribution is likely to be driven by the location of natural resources across the U.S. states that can affect current production activity and the demographic characteristics of the state population (e.g., Glaeser, Kerr, & Kerr, 2012).

To examine if endogeneity is driving the differences between the high and low sophistication states, we use a different sophistication measure. Specifically, we measure the sophistication of the *overall local population*, which includes investors and other individuals who may not participate in the stock market. Any endogeneity biases that might affect our investor sophistication measure are probably also affecting the population-wide measure.

When we sort the U.S. states based on the overall sophistication of individuals in a state, we no longer find that local return comovement or predictability is stronger in regions with less sophisticated individuals. This evidence suggests that our measures of investor sophistication obtained using the attributes of actual retail investors contain valuable information that gets diluted when we define population-wide sophistication measures. Furthermore, these results with the population-wide measure suggest that endogeneity biases are unlikely to affect our results.

These findings contribute to several strands of the literature. First, we provide evidence of geographical segmentation of the U.S. capital markets that does not originate from frictions in the capital markets but from the behavior and sophistication of local investors. We also show that the sophistication that matters for asset prices is the sophistication of the actual local investors and not the overall local population. This finding is consistent with the limited market participation literature that focuses on the decisions of market participants to better explain asset prices (e.g., Malloy, Moskowitz, & Vissing-Jørgensen, 2009).

Our finding that local comovement and predictability are stronger in low sophistication states also contributes to the predictability literature and highlights that if return predictability is a market anomaly, this anomaly potentially originates from the suboptimal decisions of investors. In broader terms, we contribute to the behavioral asset pricing literature and provide evidence consistent with one of its central premises, which posits that when there are limits to arbitrage, the biased decisions of less sophisticated investors would influence asset prices.

The rest of the paper is organized as follows. Section 2 describes the data sources and the key variables used in the empirical analysis. Sections 3 and 4 present our empirical results. We conclude in Section 5 with a brief discussion.

2 | DATA SOURCES AND KEY VARIABLES

In this section, we describe our measure of investor sophistication and our comovement and trading correlation measures. We also discuss the data we use in our predictability regressions.

2.1 | Investor sophistication measure

In general, it is difficult to directly measure the sophistication levels of actual investors. Most previous studies assume that retail investors are on average less sophisticated than institutional investors and ignore the heterogeneity in sophistication within the group of retail investors. Some studies use population-level demographic variables such as education, income, etc. to proxy for investor sophistication. However, individuals who choose to participate in the market may be considerably more sophisticated than the general population. In this paper therefore, we measure the sophistication levels of *actual retail investors*. Specifically, we use a large sample of individual investors at a large U.S. discount brokerage house. The investor data set contains the portfolio holdings and trades of investors for the 1991–1996 time period. There are 77,995 households in the database and there is significant coverage across all the U.S. states in the dataset. Additional details on the individual investor database are available in Odean (1999), Barber and Odean (2000) and Barber and Odean (2001).

To define the state-level investor sophistication index, we first obtain several measures of behavioral biases for each household. They include portfolio turnover, preference for local stocks, percentage of foreign stocks held, portfolio concentration (measured by the ratio of the portfolio variance to the average variance of stocks in the portfolio), and preference for lottery-type stocks. In addition, we use the demographic characteristics of investors to define a proxy for cognitive abilities that can capture overall investor skill.

Then, we standardize the cognitive abilities-skill proxy and the behavioral bias proxies (mean is set to zero and the standard deviation is one) and add the standardized variables that reflect good investment decisions (skill proxy, local preference, and percentage in foreign stocks) and subtract the standardized variables for measures that are likely to reflect stronger behavioral biases (portfolio turnover, portfolio concentration, and lottery preference). Last, we compute the equal-weighted average of the six investor-level components to obtain the sophistication index. Our approach is similar to Korniotis and Kumar (2011).

In some of our tests, we use a “concise” sophistication index. The concise index uses only the behavioral bias measures that separate low and high sophistication states more successfully. Specifically, we subtract the standardized values of portfolio concentration and lottery preference from the standardized value of percentage in foreign stocks, and divide by three. See Section 4.6 and Table 9 for additional details about the concise sophistication index.

2.2 | Is the investor sophistication measure appropriate?

We recognize that our retail brokerage dataset is not perfect but it is the only dataset available that allows us to measure investor sophistication at the U.S. state level. Its most important weakness is that it covers a relative short period from 1991 to 1996. However, Kumar (2009) provides some evidence that the behavioral biases which form the basis for our sophistication measure are persistent over time. Therefore, we assume that the relative levels of investor sophistication, captured by our 6 years of data, are unlikely to change over time. Even if the actual sophistication can change, the rankings of states are less likely to vary over time.

Despite being measured over a 6-year period, the investor sophistication measure has various advantages that make it appropriate for our study. As mentioned above, it is the only available dataset that has enough information at the state-level to allow us to meaningfully capture investor sophistication differences across the U.S. states. Moreover, existing evidence suggests that the sophistication measure is representative of the investors located in each U.S. state. For example, Korniotis and Kumar (2011) find that the distribution of households across states in the retail brokerage data set is similar to what is reported by the Census Bureau. Also, important demographic variables (like age, income, marital status) are positively correlated between the brokerage data set and the state data reported by the Census Bureau.

As a validation test, we also use data from the 1990 Census to assess the *overall* population-wide sophistication level of the U.S. states. The population-based sophistication index is based on education and occupation data from the 1990 Census and an IQ proxy. Specifically, using the Census data, we identify the percentage of state inhabitants with Bachelor's degree or higher and the percentage of state inhabitants who are professionals (i.e., they have white-collar jobs). The IQ proxy is based on statewide SAT scores from Kanazawa (2006). To compute the index, we standardize the education, occupation, and IQ proxies. The statewide sophistication index is the equal weighted average of these standardized variables.

2.3 | Attributes of low and high sophistication states

In Table 1, we report the values of the investor sophistication index and the statewide index based on the Census data. The values of the two sophistication indices in Table 1 indicate that the two indices rank the U.S. states differently. For example, Kansas has high investor sophistication, but the statewide sophistication is low. In contrast, some states like Connecticut are classified as sophisticated but investors in those states are not very sophisticated. Overall, the two indices are positively correlated but the correlation estimate (= 0.338) is not very high (see Table 2, Panel B).

In Table 2, we report the summary statistics for investor decisions and various demographic variables for low and high sophistication states. In Panel A, we report the univariate statistics, and in Panel B, we report the correlation coefficient estimates. As expected, the average investor sophistication level for low (below median) sophistication states (= −0.266) is substantially

TABLE 1 State-level sophistication estimates

Panel A: Investor sophistication				Panel B: Population-wide sophistication			
State	Index	State	Index	State	Index	State	Index
DC	1.00	*AK*	0.05	*DC*	2.90	*KS*	−0.10
KS	0.70	*ND*	0.05	*MA*	1.79	*OH*	−0.18
DE	0.62	*RI*	0.03	*MD*	1.31	*AZ*	−0.19
CT	0.50	*MD*	0.01	*CT*	1.29	*NE*	−0.24
MO	0.41	*NJ*	0.00	*NJ*	1.15	*IN*	−0.26
WA	0.38	*OK*	−0.03	*VA*	1.04	*SC*	−0.34
WY	0.32	*TX*	−0.10	*NY*	0.98	*ND*	−0.40
GA	0.32	*IA*	−0.13	*CO*	0.83	*ID*	−0.46
UT	0.31	*NH*	−0.16	*NH*	0.78	*MI*	−0.49
ID	0.27	*CO*	−0.17	*VT*	0.78	*SD*	−0.51
NC	0.24	*NY*	−0.17	*DE*	0.62	*WI*	−0.51
PA	0.23	*LA*	−0.20	*WA*	0.56	*MO*	−0.56
MT	0.19	*CA*	−0.31	*MN*	0.46	*IA*	−0.56
IL	0.18	*SD*	−0.31	*RI*	0.44	*NM*	−0.58
AL	0.18	*WV*	−0.32	*PA*	0.37	*UT*	−0.63
TN	0.18	*AZ*	−0.40	*CA*	0.34	*TN*	−0.65
WI	0.14	*FL*	−0.44	*GA*	0.29	*WY*	−0.65
OR	0.13	*ME*	−0.45	*OR*	0.25	*AL*	−0.94
OH	0.13	*KY*	−0.45	*IL*	0.14	*KY*	−0.96
SC	0.11	*MA*	−0.46	*HI*	0.13	*OK*	−0.96
HI	0.11	*AR*	−0.50	*NC*	0.10	*NV*	−1.00
MN	0.08	*NV*	−0.52	*MT*	0.09	*LA*	−1.07
VA	0.08	*MS*	−0.61	*AK*	0.02	*WV*	−1.08
MI	0.06	*VT*	−0.68	*ME*	−0.03	*AR*	−1.41
NM	0.06	*NE*	−0.71	*TX*	−0.05	*MS*	−1.73
IN	0.05			*FL*	−0.09		

Note: The table reports sophistication indices for the U.S. states. In Panel A, we report the investor sophistication index of Korniotis and Kumar (2011). First, they add the standardized measures that are positively related to risk sharing (cognitive ability, local preference, and percentage in foreign stocks) and subtract the standardized variables for measures that are negatively related to risk sharing (portfolio turnover, portfolio concentration, and lottery preference). The resulting total is divided by six and it is the investor sophistication index. In Panel B, we report a population-wide sophistication index based on the proportion of residents in a state with a Bachelor's or higher educational degree, the percentage of management and professional employees, and a proxy for state-level IQ. To compute the index, we take the equal-weighted average of the standardized values of the three measures.

lower than the average investor sophistication for high (above median) sophistication states (= 0.277) and the difference is statistically significant (t statistic = 8.35).

In general, the values of almost all the investor-based measures are different between the low and high investor sophistication measures and the differences become even stronger when we compare the most sophisticated states (top quartile) with the least sophisticated states (bottom quartile). We also find that there are minor differences between the statewide demographic characteristics of high and low sophistication states. However, the differences in statewide demographics are not statistically significant.

2.4 | Local return comovement and trading correlation measures

Our primary measure of return comovement is a local beta, $\beta_{i,HQ}$, that measures how closely the return of a firm in state i moves with the returns of other firms in the same state. We obtain the local beta from the following factor regression:

$$r_{it} - r_f = \alpha + \beta_{i,\mathrm{HQ}}\mathrm{HQState} + \beta_{i,\mathrm{EC}}\mathrm{ECState} + \beta_{i,M}\mathrm{MKTRF} + \beta_{i,S}\mathrm{SMB} + \beta_{i,H}\mathrm{HML} + \varepsilon_{it}.$$

TABLE 2 Demographic characteristics of U.S. states and portfolio decisions

Panel A: Characteristics of high and low sophistication states	Low sophistication (below median)	High sophistication (above median)	High − Low	*t* stat	Low sophistication (below 25th pctl)	High sophistication (above 75 pctl)	High − Low	*t* stat
Investor sophistication index	−0.266	0.277	0.544	8.35	−0.473	0.441	0.914	12.08
Cognitive abilities	−0.142	−0.008	0.134	2.17	−0.223	0.099	0.322	3.13
Portfolio turnover	6.845	5.737	−1.109	−3.22	7.360	5.430	−1.930	−3.75
Local bias	0.154	0.167	0.013	0.52	0.152	0.178	0.026	0.60
Percentage in foreign stocks	13.354	15.502	2.148	2.37	11.964	15.185	3.221	2.11
Portfolio concentration	0.546	0.532	−0.013	−1.38	0.543	0.538	−0.004	−0.32
Lottery preferences	12.191	10.874	−1.317	−2.62	12.474	10.086	−2.388	−3.77
Concise sophistication index	−0.278	0.289	0.568	5.34	−0.435	0.341	0.776	4.76
Population-wide attributes								
Education	26.227	28.064	1.837	1.06	24.454	29.567	5.113	1.78
Percentage unemployed	5.581	5.736	0.155	0.42	5.569	5.800	0.231	0.39
Percentage professionals	32.938	33.576	0.638	0.53	31.700	34.742	3.042	1.49
Average age	50.363	50.299	−0.064	−0.12	50.797	49.667	−1.130	−1.30
Stock market participation rates	24.492	25.328	0.836	0.63	23.254	25.742	2.488	1.37
Percentage urban	67.529	70.135	2.605	0.61	62.998	71.527	8.529	1.19
Percentage married	53.455	53.387	−0.068	−0.04	53.538	52.663	−0.876	−0.31
IQ	93.631	94.352	0.721	0.20	91.292	96.425	5.133	0.93
Population-wide soph. index	−0.082	0.085	0.167	0.70	−0.334	0.310	0.645	1.64

(Continues)

TABLE 2 (Continued)

Panel B: Correlations coefficient estimates

	Soph. index	Cognitive abilities	Port. turnover	Local bias	% in foreign stk	Port. concentr.	Lottery pref.	Concise soph. index
Investor sophistication index	1							
Cognitive abilities	0.607	1	1					
Portfolio turnover	−0.489	−0.150						
Local bias	0.103	−0.134	0.240	1				
Percentage in foreign stocks	0.432	0.207	0.033	−0.030	1			
Portfolio concentration	−0.108	0.005	0.056	0.505	0.038	1		
Lottery preferences	−0.412	−0.087	0.120	−0.132	0.178	−0.275	1	
Concise sophistication index	0.670	0.203	−0.100	−0.284	0.552	−0.483	−0.385	1
Population-wide attributes								
Education	0.371	0.675	−0.268	−0.332	0.161	−0.156	0.130	0.132
Percentage unemployed	0.209	0.130	0.001	0.118	0.160	0.030	−0.071	0.142
Percentage professionals	0.357	0.650	−0.236	−0.336	0.143	−0.156	0.082	0.152
Average age	−0.269	−0.282	0.264	0.048	0.160	−0.035	0.275	−0.056
Stock market participation rate	0.196	0.406	−0.230	−0.393	0.210	−0.212	0.243	0.126
Percentage urban	0.282	0.486	0.129	0.040	0.594	0.226	0.160	0.147
Percentage married	−0.161	−0.437	0.068	0.149	−0.069	0.031	−0.110	0.006
IQ	0.139	0.208	0.014	−0.336	0.167	−0.389	0.114	0.311
Population-wide soph. index	0.338	0.597	−0.191	−0.391	0.184	−0.273	0.127	0.232

Note: The table presents summary statistics of portfolio decisions and demographic variables for the U.S. states. Panel A reports averages for high and low sophistication subsamples. Panel B reports cross-sectional correlations. The portfolio decisions are obtained by aggregating the choices of a sample of brokerage investors. They include the portfolio turnover, a measure of local preference, the portfolio weight in foreign stocks, the portfolio concentration, and the proportion of all trades that are in stocks with lottery-type features. We also use a cognitive abilities proxy (see Korniotis & Kumar, 2011). Based on these six measures, we define the investor sophistication index as in Korniotis and Kumar (2011) (see Table 2 for the definition). We also define a concise index that adds the standardized value of foreign ownership, subtracts the standardized values of portfolio concentration and lottery preferences, and divides by 3. The state-level education is from the 1990 Census, and it is the proportion of state inhabitants with Bachelor's or higher educational degree. The average age is the average age of the inhabitants of each state according to the 1990 Census. Similarly, the percentage unemployed, urban, married in a state are also from the 1990 Census. The stock market participation proxy is the percentage of tax returns in each state that reported dividend income over the 1998–2005 period (Brown, Ivkovic, Smith, & Weisbenner, 2008). The state IQ data based on SAT scores are from Kanazawa (2006). The population-wide sophistication measure is an equal weighted average of the standardized values of the IQ proxy, education, and the percentage of professionals.

In the above regression, the local headquarters (HQ) state index (HQState) is the excess return on the portfolio of stocks of firms headquartered in the same state as a given firm. In the factor regression, we control for the Fama and French (1993) MKTRF, SMB, and HML factors. Moreover, we include in the regression an Economic Center index (ECState) that corresponds to the state (besides the HQ state) where the firm has the strongest economic presence.[3] In the present of the EC index, the local beta captures regional effects that are only related to the headquarter state and not to other states that the firm is connected to. Finally, we estimate the regression annually for each firm, using daily stock returns, and require at least 100 daily returns for the year in order for the $\beta_{i,\mathrm{HQ}}$ to be included in the analysis.

In addition to stock return comovement, we also directly examine trading correlations. We obtain trading data from the Institute for the Study of Security Markets (ISSM) and the Trade and Quote (TAQ) databases. We use small-sized trades (trade size $\leq$ \$5,000) to proxy for retail trades, and obtain monthly retail buying and selling volume for each stock. Like Barber, Odean, and Zhu (2009), we use the ISSM/TAQ data only until 2000 because the assumption that small trades proxy for retail trading is less likely to be valid after 2000. We measure the degree of correlation in the trading activities of retail and institutional investors by estimating the partial correlation of stock-level buy–sell-imbalance (BSI) with the portfolio BSI of other stocks in the same HQ state, controlling for correlation with the market.

To compute the portfolio BSI, we first measure period-t BSI for each stock i using

$$\mathrm{BSI}_{it}=\frac{\mathrm{VB}_{it}-\mathrm{VS}_{it}}{\mathrm{VB}_{it}+\mathrm{VS}_{it}},$$

where VB_{it} and VS_{it} are the period-t dollar buy and sell trading volumes of stock i, respectively.

The period-t portfolio BSI measure BSI_{pt} is the equal-weighted average of the period-t buy–sell imbalance of all other stocks headquartered in the same state as firm i.

The partial correlation is defined as the correlation between the residuals ε_{it} and η_{pt} from the following two regressions:

$$\mathrm{BSI}_{it}=\gamma_{10}+\gamma_{11}\mathrm{MKTRF}_t+\varepsilon_{it}, \mathrm{BSI}_{pt}+\gamma_{20}+\gamma_{21}\mathrm{MKTRF}_t+\eta_{pt}.$$

where BSI_{it} is the period-t buy–sell imbalance of stock i, BSI_{pt} is the equal-weighted buy–sell imbalance of the HQ state portfolio in period t, and MKTRF_t is the period-t market return in excess of the risk-free rate. We use partial correlation measures to isolate the correlation in retail investor demand that is not driven by common correlation with the market. We compute annual estimates of this partial correlation using monthly observations of retail BSI and refer to this measure as RTC.

Table 3 reports summary statistics for the local comovement and trading correlation measures, as well as the explanatory variables used in our regression analysis. The HQ state beta is on average positive, with a mean of 0.117 and standard deviation of 0.827. The HQ state RTC measure is also positive on average, with a mean value of 0.127 and standard deviation of 0.351. There is also substantial variation across stocks in the degree to which their returns comove with their local state index, and in the degree to which retail investor demand is correlated with demand for other local stocks. For instance, the interquartile range for the state betas is −0.229 to 0.425 and for RTC is −0.118 to 0.390.

Panel B reports correlation coefficients among the comovement and trading correlations measures, the different versions of the investor sophistication index, and the absolute value of the state-level macroeconomic index which we use in some of our tests. This state-level economic index is constructed by averaging state income growth, the state housing collateral ratio, and (the negative of) relative state unemployment. Before we construct the index, we standardize its components to have a mean of 0 and standard deviation of 1. We find that the investor sophistication index and the more concise version of the index are closely correlated. Both, however, have a slight *negative* correlation with the alternate version of the sophistication index based on statewide Census data. Consistent with our main conjecture, HQ state betas and trading correlations are both negatively correlated with measures of investor sophistication.

2.5 | Return measures and return predictors

In our predictability regressions, the dependent variable is the quarterly value-weighted portfolio return of all firms headquartered in a state. We use characteristic-adjusted returns of the state portfolios (Daniel, Grinblatt, Titman, & Wermers, 1997). Our return predictors use the U.S. and state income data from the Bureau of Economic Analysis (BEA), U.S. and state unemployment data from the Bureau of Labor Statistics (BLS), and interest rate spreads from the Federal Reserve. We use data from Center for Research on Security Prices (CRSP) to compute the dividend–price ratio of the state portfolios. We also use the

TABLE 3 Summary statistics: Variables for comovement and trading correlation analysis

Panel A: Summary statistics

	Mean	SD	25th pctl	Median	75th pctl	N
HQ state beta	0.117	0.827	−0.229	0.064	0.425	64,615
HQ state retail trading correlation (RTC)	0.127	0.351	−0.118	0.148	0.390	18,820
Investor soph. index	−0.060	0.276	−0.308	−0.103	0.144	64,615
Concise soph. index	−0.138	0.375	−0.530	−0.192	0.165	64,615
Population-wide soph. index	0.362	0.681	−0.047	0.338	0.984	64,615
Low sophistication dummy	0.586	0.492	0	1	1	64,615
|State econ. index|	0.622	0.582	0.177	0.433	0.866	55,002
Momentum	0.196	0.926	−0.231	0.068	0.388	64,615
Turnover	0.140	0.199	0.041	0.087	0.178	64,615
Ln(Firm age)	4.770	0.990	4.060	4.844	5.533	64,615
Ln(Size)	5.549	2.093	4.018	5.464	6.942	64,615
Ln(Market-to-Book)	0.686	0.932	0.114	0.606	1.175	64,615
Industry cluster	0.598	0.490	0	1	1	64,615
Total population (1,000s)	1,489	1,856	509	884	1,537	64,615
Education	26.111	9.582	19	25	32	64,615
Male–female ratio	0.949	0.043	1	1	1	64,615
Married	0.546	0.112	0	1	1	64,615
Minority	0.229	0.152	0	0	0	64,615
Age	32.480	3.445	30	32	35	64,615
Urban	0.917	0.131	1	1	1	64,615

Panel B: Correlation coefficient estimates

	HQ State Beta	HQ State RTC	Soph. Index	Concise Soph. Index	Population Soph. Index	Low Soph. Dummy	State Econ.
HQ state beta	1						
HQ state RTC	0.0083[a]	1[a]					
Investor soph. index	−0.055	−0.0451[a]	1				
Concise soph. index	−0.051	−0.044[a]	0.745	1			
Population-wide soph. index	−0.020	0.0162[a]	−0.078	−0.184	1		
Low sophistication dummy	0.073	0.0375[a]	−0.831	−0.727	0.181	1	
|State econ. index|	0.070	0.0398[a]	−0.404	−0.412	−0.096	0.462	1

Note: Panel A reports summary statistics for variables used in the analysis of local return comovements and trading correlations. The HQ state beta is the beta estimate from a stock-level time series regression of stock returns on the return of an index of other firms headquartered in the same state, as well as MKTRF, SMB, and HML factors. The HQ State Retail Trading Correlation is the partial correlation (controlling for the market return) between the buy–sell imbalance of retail trades of a firm's stock with the aggregate buy–sell imbalance of retail trades in other firms with the same headquarters state. The investor sophistication index and concise sophistication index are based on data on retail investors in the state, while the population-wide sophistication index is based on statewide census data. The low sophistication dummy equals 1 if the investor sophistication index is below median. The state-level economic index is the average of the standardized values of state income growth, relative state unemployment, and the state housing collateral ratio. The industry cluster dummy that equals one of more than 10% of firms in the MSA are from a single industry and more than 10% of firms in that industry are located in the MSA (by market capitalization). The data are from 1994 to 2011. In Panel B, we report correlations among the dependent variables and main explanatory variables.

[a]Correlations computed from the more limited sample for which retail trading data is available. All other correlations are estimated from the full sample.

U.S. and state housing collateral ratio (Lustig & Van Nieuwerburgh, 2005) and the U.S. *cay* residual of Lettau and Ludvigson (2001a).

Most of the state-level data are available from 1980 to 2008. However, Korniotis and Kumar (2013a) use a series of approximations and extend the state data backwards to 1975. We also follow their approach and we use data starting in 1975 when estimating the predictability model and forming our trading strategies. In our trading strategies, we also extend the sample to include year 2009 using the state performance rankings obtained at the end of 2008. In addition, for the U.S. and state-level income growth, relative unemployment, and housing collateral, we follow the convention used in Korniotis and Kumar (2013a) who note that these variables are reported with a lag of two quarters. Thus, when we predict the state portfolio returns in a quarter, we use information on these variables reported two quarters back. In Table 4, we present the summary statistics for the quarterly state returns and the various return predictors. We organize the summary statistics into two panels. In Panel A, we report the univariate statistics and in Panel B, we report the correlation coefficients. The univariate statistics in Panel A show that quarterly state portfolio returns are on average negative and not serially correlated. We also observe that the state-level macroeconomic predictors are more volatile and less serially correlated compared to their U.S. counterparts.

The correlation estimates in Panel B indicate that state-level macroeconomic predictors are moderately correlated with the corresponding U.S.-level variables and with other predictors. To ensure that the state predictors only capture state-specific economic conditions, we include all U.S.-level variables in our empirical analysis.

3 | EMPIRICAL RESULTS: RETURN COMOVEMENT AND TRADING CORRELATIONS

In this section, we report the first set of our main empirical tests. Our conjecture is that local stock returns would comove more with one another in regions where investors are less sophisticated and potentially make more investment mistakes. Specifically, if relatively less sophisticated investors are more sensitive to changes in local economic conditions, their demand for financial assets should be more sensitive to local economic conditions. Moreover, if local investors are more likely to serve as marginal investors for local stocks due to market segmentation, the cross-sectional variation in investor sophistication across regions will create a geographical component to mispricing.

3.1 | Investor sophistication and local comovement

First, we investigate if the stock returns of firms headquartered in states with lower investor sophistication exhibit greater local return comovement. Specifically, we estimate multivariate regressions in which the dependent variable is the firm's HQ state beta. Our main variable of interest is the level of investor sophistication where the firm is headquartered. Because investor sophistication varies at the state level, we cluster the standard errors by state.

In the comovement regressions, we also control for an array of firm-level characteristics which includes momentum, defined as the firm's stock return over the previous 12 months, the natural logarithm of the firm's age in months, the natural log of the firm's market value in millions of dollars, and the natural log of the firm's market-to-book ratio. Furthermore, at the MSA-level, we control for demographic characteristics including total population, the fraction of the population with a Bachelor's degree or higher, the ratio of males to females, the proportion of households with a married couple, the proportion of minority residents, the median age, and the fraction of the population living in urban areas.

A potential concern is that certain industries are clustered in certain states, and that less sophisticated states happen to be those with more concentrated industries. Furthermore, certain industries may naturally gravitate toward areas where the potential workforce is less or more sophisticated. To control for these industry-related concerns, we include an industry cluster dummy in the regressions that equals one if more than 10% of firms in the MSA are from a single industry and more than 10% of firms in that industry are located in the MSA (by market capitalization). We also include industry fixed effects (as well as year fixed effects), so that the effect of investor sophistication we identify is between firms in the same industry that are located in low versus high sophistication states.

Table 5 reports the results of the local return comovement regressions. We find negative and significant coefficient estimates for the investor sophistication index. In particular, when we include the full set of controls we find a coefficient estimate of −0.136 (t statistic = −4.94) for the sophistication index. This estimate implies that going from the 25th percentile (−0.308) to the 75th percentile (0.144) of the investor sophistication index would correspond to an HQ state beta that is −0.06 lower, a meaningful change relative to the mean value of 0.117. We find comparable results when we use the concise version of the investor sophistication index.

Overall, the regression results confirm the expected negative relation between local comovement and investor sophistication. Next, we test an auxiliary hypothesis related to the importance of local economic conditions. We conjecture that if the relation between investor sophistication and local comovement is driven by sensitivity to local economic conditions, then investor sophistication should be particularly important when the state economy is unusually bad or good (as opposed to “normal” economic conditions).

TABLE 4 Summary statistics of variables in return predictability regressions

Panel A: Average univariate statistics across states				
State returns	**Abbreviation**	**Mean**	***SD***	**Autocorrelation**
State-level excess returns	$R^{loc} - R_f$	−0.129	0.045	0.006
State-level predictors				
Income growth	*Inc Gr*	4.872	2.662	0.862
Relative unemployment	*Rel Un*	1.002	0.241	0.837
Housing collateral ratio	*hy*	0.022	0.102	0.897
U.S. macro predictors				
Income growth	*Inc Gr*	1.803	0.014	0.656
Relative unemployment	*Rel Un*	0.998	0.192	0.866
Housing collateral ratio	*hy*	0.024	0.089	0.988
Other predictors				
Dividend–price ratio	*log(1 + D/P)*	0.027	0.013	0.954
U.S. cay residual	*U.S. cay*	0.001	0.016	0.876
30-Day commer. paper minus 30-day Treasury bill	*Paper-Bill Spread*	0.048	0.026	0.955
10-Year minus 1-year government bond	*Term Spread*	0.011	0.011	0.907
Baa corporate bond minus 1-year government bond	*Default Spread*	0.021	0.006	0.843

Panel B: Correlations among state returns and return predictors							
	$R^{loc} - R_f$	*St Inc Gr*	*St Rel Un*	*St. hy*	*US Inc Gr*	*US Rel Un*	*US hy*
$R^{loc} - R_f$	1	−0.007	0.032	−0.014	−0.015	0.040	0.035
State-level predictors							
Inc Gr		1	−0.379	0.308	0.112	−0.348	0.365
Rel Un			1	0.030	−0.147	0.815	−0.155
hy				1	−0.068	−0.024	0.297
U.S. macro predictors							
Inc Gr					1	−0.166	−0.300
Rel Un						1	−0.191
hy							1
U.S. macro predictors							
Inc Gr						−0.166	−0.300
Rel Un						1	−0.191
hy							1
Other predictors							
log(1 + D/P)	0.019	−0.552	0.164	−0.213	−0.117	0.186	−0.233
U.S. cay	−0.004	−0.428	0.076	−0.376	−0.127	0.121	−0.138
Paper-bill spread	−0.010	−0.291	−0.073	−0.170	−0.206	−0.128	−0.084
Term spread	0.033	−0.174	0.342	−0.147	0.127	0.456	0.021
Default spread	0.021	−0.041	0.406	0.017	−0.062	0.482	0.104

Note: The table reports univariate statistics (Panel A) and the correlation matrix (Panel B) for state portfolio returns and return predictors. The sample period is from 1980 to 2008. State portfolios with fewer than 10 firms are excluded from the sample. The main return variable is the state portfolio return over the risk-free rate (Rlocal), where the state portfolio return is the value-weighted state portfolio return of firms headquartered in the state. The risk-free rate is the rate of return of 30-day Treasury bills obtained from Ibbotson Associates. The returns are divided by one plus the inflation rate computed using the U.S. consumer price index from the Bureau of Labor Statistics (BLS). The return predictors include the relative unemployment rate for the U.S. and individual states (US Rel Un, State Rel Un), the U.S. and state labor income growth rates (US Inc Gr, State Inc Gr), the U.S. and state housing collateral ratio (US *hy*, State *hy*), the paper-bill spread (30-day commercial paper minus 30-day Treasury bill return), the term spread (10-year government bond yield minus one-year government bond yield), default spread (Baa-rated corporate bond yield minus ten-year government bond yield), the US cay residual of Lettau and Ludvigson (2001a, b), and the state dividend-price ratio (log-value of (1 + *D*/*P*)). D is the sum of the past four quarterly dividends and *P* is the stock price at the end of the most recent quarter. The state housing collateral ratio is computed using the Lustig and van Nieuwerburgh (2005) method. The unemployment rates are from the BLS. The relative unemployment rate is the ratio of current unemployment rate to the moving average of the unemployment rates from the previous 16 quarters. Labor income is from the Bureau of Economic Analysis (BEA). The U.S. cay and U.S. *hy* are downloaded from Sydney Ludvigson's and Stijn van Nieuwerburgh's web sites, respectively. The three spreads use quarterly data obtained from the Board of Governors of the Federal Reserve System web site.

TABLE 5 Local investor sophistication and stock return comovement

	1	2	3	4	5
Investor soph. index	−0.126	−0.136			
	(−4.57)	(−4.94)			
Concise soph. index			−0.068		
			(−2.63)		
Low sophistication dummy				0.040	
				(1.71)	
Low soph. × |state econ. index|				0.081	
				(2.02)	
|State econ. index|				−0.056	
				(−1.50)	
Population-wide soph. index					−0.014
					(−0.98)
Momentum	0.007	0.009	0.009	0.008	0.009
	(1.28)	(1.47)	(1.50)	(1.60)	(1.52)
Turnover	0.202	0.186	0.190	0.296	0.197
	(2.38)	(2.34)	(2.39)	(1.70)	(2.53)
Ln(Firm Age)	−0.016	−0.011	−0.012	−0.007	−0.013
	(−2.28)	(−2.02)	(−2.24)	(−1.66)	(−2.34)
Ln(Size)	0.045	0.043	0.043	0.041	0.042
	(5.40)	(5.75)	(5.82)	(5.40)	(5.61)
Ln(Market-to-book)	−0.011	−0.011	−0.011	−0.013	−0.010
	(−1.92)	(−1.64)	(−1.71)	(−1.64)	(−1.58)
Industry cluster		0.052	0.052	0.053	0.069
		(3.21)	(2.72)	(3.16)	(4.77)
Total population		0.000	0.000	0.000	0.000
		(−1.41)	(−1.12)	(−1.63)	(−0.61)
Education		0.001	0.000	0.001	0.001
		(1.77)	(0.48)	(0.93)	(0.67)
Male–female ratio		0.335	0.426	0.298	0.484
		(2.10)	(2.14)	(1.60)	(2.63)
Married		0.449	0.443	0.411	0.391
		(2.68)	(2.44)	(2.54)	(1.91)
Minority		0.402	0.384	0.348	0.343
		(5.80)	(4.47)	(4.11)	(3.37)
Age		0.003	0.004	0.003	0.004
		(0.94)	(1.50)	(0.81)	(1.55)
Urban		0.039	0.057	0.059	0.072
		(0.81)	(1.06)	(1.16)	(1.35)
Industry effects	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes

Note: This table reports estimates of the relation between investor sophistication and local comovement of stock returns. We obtain estimates of local stock return comovement by estimating annual regressions of daily returns on state-level stock indices as well as MKTRF, SMB, and HML factors. The state indices are the value-weighted returns (in excess of the risk-free rate) to the portfolio of firms that are headquartered in, or operate in, the same state as a given stock. The HQ index includes stocks of firms with the same headquarters state. The betas are estimates via stock-by-stock time-series regressions in each year using daily returns. We report multivariate regressions where the dependent variable is the firm's HQ state beta, and the independent variables include local investor sophistication, the concise sophistication index, a low-sophistication (below mean) dummy, and the absolute value of a state-level economic index. The state economic index is an average of the standardized values of state income growth, relative state unemployment, and the state housing collateral ratio. The industry cluster dummy that equals one of more than 10% of firms in the MSA are from a single industry and more than 10% of firms in that industry are located in the MSA (by market capitalization). The *t* statistics in are based on standard errors clustered at the state level. The data are from 1994 to 2011.

For this test, we create an interaction term between a low sophistication dummy variable and an index of abnormal state-level economic conditions. The low sophistication dummy variable equals 1 for states with a below-mean value of the sophistication index. To determine abnormal economic conditions, we construct an index of state-level economic conditions. Similar to Korniotis and Kumar (2013a), the index is the sum of state income growth and the state housing collateral ratio minus relative state unemployment divided by 3. Before we construct the index, we standardized the three measures so that they have zero mean and standard deviation of 1. By construction, this state economic index has a mean of zero, and we thus capture deviations from "normal" economic conditions by using the absolute value of the index.

In column (4), we find a positive and significant coefficient on the interaction term between the low sophistication dummy and the absolute value of the state economic index. This evidence indicates that the relation between sophistication and local return comovement is particularly strong during abnormal local economic conditions.

3.2 | Population-wide sophistication and local comovement

In these baseline results, our sophistication index uses the brokerage data and combines the investment skill (i.e., cognitive ability proxy) and five state-level portfolio measures of individuals who participate in the stock market. For comparison, we consider an alternative population-wide sophistication index, which is the average value of the standardized values of state-level IQ, education, and percentage of white-collar workers.

In contrast to the baseline results, we find no relation between the population-wide sophistication measure and local return comovement (see column (5) of Table 5). Therefore, our investor-based sophistication index captures behavioral biases better than a general sophistication index that does not exclusively consider the behavior of stock market participants.

The insignificance of the population-wide sophistication results is important because the distribution of sophisticated investors (and sophisticated individuals in general) may not be random. Households that are more or less sophisticated may be naturally drawn to certain areas, and different types of firms may gravitate towards, or tend to be founded in, areas with more or less sophisticated populations. To the extent that there is some geographic matching or selection occurring between population and firms, it would likely be driven by the characteristics of the general population rather than by the characteristics of the subset of the population which participates in the stock market. It is therefore encouraging that our results are driven by the sophistication level of actual investors, rather than the broader population.

3.3 | Investor sophistication and trading correlations

Next, we provide further evidence that the relation between investor sophistication and local return comovement is related to the investment decisions of local investors. In particular, we examine if the trading activity of local stocks by local investors is correlated. For this analysis, we use the monthly RTC measure described in Section 2.4.

Table 6 reports the results from regressions of RTC on investor sophistication and other firm and local controls. As with the return comovement results, we find a significant negative correlation between investor sophistication and local retail trading correlation, which holds for both the main investor sophistication index and the concise index. The coefficient estimate of −0.039 (t statistic = −3.74) in column (2) of Table 6 implies that going from the 25th percentile (−0.308) to the 75th percentile (0.144) of the investor sophistication index would correspond to an trading correlation measure that is lower by 0.02 (relative to its mean of 0.127 and standard deviation of 0.351). In addition, as with the comovement results in Table 5, we find in column (4) of Table 6 that the relation between sophistication and correlated trading is strongest during abnormal economic conditions. We also find that there is no relation between the degree of trading correlation and the population-wide sophistication measure.

Overall, our evidence suggests that less sophisticated areas trade in a correlated fashion, which can rationalize why the returns of firms headquartered in less sophisticated areas are more strongly correlated. Moreover, local return comovement and local trading correlations are especially strong when local economic conditions are unusually bad or good.

4 | EMPIRICAL RESULTS: RETURN PREDICTABILITY

In this section, we examine whether mispricing effects are strong enough to create predictability in the returns of local firms. Return predictability can arise because local economic conditions are persistent. Because investors' mistakes might be affected by the local economy, their investment mistakes might also have a slow-moving component, which can result in return

TABLE 6 Local investor sophistication and trading correlation

	1	2	3	4	5
Investor soph. index	−0.035	−0.039			
	(−3.63)	(−3.74)			
Concise soph. index			−0.024		
			(−2.56)		
Low soph. dummy				−0.006	
				(−0.44)	
Low Soph. × |State econ. index|				0.049	
				(1.76)	
|State econ. index|				−0.023	
				(−0.96)	
Population-wide soph. index					0.006
					(0.95)
Momentum	0.003	0.004	0.004	0.004	0.004
	(1.22)	(1.61)	(1.60)	(1.60)	(1.62)
Turnover	0.156	0.154	0.155	0.152	0.160
	(4.67)	(4.62)	(4.64)	(4.72)	(4.57)
Ln(Firm age)	−0.009	−0.009	−0.009	−0.009	−0.009
	(−2.81)	(−2.81)	(−2.86)	(−2.81)	(−2.90)
Ln(Size)	−0.011	−0.011	−0.011	−0.011	−0.011
	(−2.98)	(−2.79)	(−2.79)	(−2.77)	(−2.77)
Ln(Market-to-book)	0.007	0.006	0.006	0.006	0.007
	(2.78)	(2.38)	(2.37)	(2.35)	(2.41)
Industry cluster		−0.009	−0.011	−0.007	−0.008
		(−1.29)	(−1.52)	(−0.94)	(−1.22)
Total population		0.000	0.000	0.000	0.000
		(−1.09)	(−0.96)	(−1.32)	(−0.84)
Education		0.000	0.000	0.000	0.000
		(0.02)	(−0.38)	(−0.05)	(−0.84)
Male–female ratio		0.104	0.120	0.030	0.188
		(1.51)	(1.78)	(0.44)	(2.31)
Married		0.063	0.067	0.032	0.044
		(1.56)	(1.56)	(0.75)	(1.03)
Minority		0.013	0.012	−0.027	0.005
		(0.49)	(0.42)	(−0.98)	(0.16)
Age		−0.001	0.000	−0.001	0.000
		(−0.69)	(−0.14)	(−0.56)	(−0.25)
Urban		0.045	0.046	0.044	0.055
		(1.24)	(1.25)	(1.26)	(1.43)
Industry effects	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes

Note: This table reports estimates from a regression of measures of local trading correlation on state investor sophistication and other firm and state-level characteristics. The dependent variable is the retail trading correlation, computed as the partial correlation (controlling for the market return) between the buy–sell imbalance of retail trades of a firm's stock with the aggregate buy–sell imbalance of retail trades in other firms with the same headquarters state. The explanatory variables include local investor sophistication, the concise sophistication index, a low-sophistication (below mean) dummy, and the absolute value of a state-level economic index. The state economic index is an average of the standardized values of state income growth, relative state unemployment, and the state housing collateral ratio. The industry cluster dummy that equals one of more than 10% of firms in the MSA are from a single industry and more than 10% of firms in that industry are located in the MSA (by market capitalization). The t statistics, reported in parentheses below the coefficient estimates, are based on standard errors clustered by state. The sample period is from 1994 to 2000.

predictability (e.g., Alti and Tetlcok). Consequently, local economic conditions might have a greater ability to predict local stock returns in areas with less sophisticated retail investors.

4.1 | Return predictability model

In the predictability tests, we follow Korniotis and Kumar (2013a) and estimate return predictability panel regressions. The dependent variables are the quarterly characteristic adjusted returns of all firm headquartered in a state. The main independent variables are the U.S. and state-level versions of relative unemployment rate, labor income growth rate, and housing collateral ratio (Lustig & Van Nieuwerburgh, 2005). We also include the paper-bill spread (30-day commercial paper minus 30-day Treasury bill return), the term spread (10-year government bond yield minus 1-year government bond yield), default spread (Baa-rated corporate bond yield minus 10-year government bond yield), the U.S. *cay* residual of Lettau and Ludvigson (2001a), and the state dividend–price ratio. For robustness, we include state fixed effects in our panel regressions.

We estimate the panel regressions with OLS using a recursive methodology. Specifically, at the end of each quarter t, we estimate the predictability regression using data that are available at the end of quarter $t-1$. For each recursive regression, we compute t statistics using the Driscoll and Kraay (1998) standard errors that can accommodate cross-sectional and serial correlations. In each recursive step, we collect the estimates and the t statistics. We compute the average values of the recursive estimates. We also compute the percentage of times that an estimate is significant and has the same sign as the average estimate. The sample period for this analysis is from 1980 to 2008.

4.2 | Predictability regression estimates

We estimate the predictability model for all states and separately for states with high (above median) and low (below median) levels of investor sophistication. The regression estimates are reported in Table 7. In column (1) where we use all the U.S. states regardless of investor sophistication, consistent with the evidence in Korniotis and Kumar (2013a), we find that relative unemployment and the housing collateral ratio are the most significant state-level predictors. Specifically, the estimates indicate that when the local economy performs poorly (i.e., unemployment rate is high and housing collateral is low), future state returns are higher.

In columns (2) and (3), we estimate two new specifications. Column (2) focuses only on states with low sophistication investors and column (3) considers states with high levels of investor sophistication. Column (4) reports the difference between the low and high sophistication subsample estimates. The estimation results are very different between the two subsamples. The evidence of return predictability is considerably stronger for the low sophistication subsample. The average adjusted R^2 in the low sophistication sample is about 5% whereas in the high sophistication sample it is less than 1%.

Moreover, the predictive power of state-level unemployment and state-level *hy* is stronger in the low sophistication subsample. For example, the coefficient estimate of relative state unemployment is 0.021 in the low sophistication subsample and it is only 0.002 in the high sophistication subsample (difference in estimates = 0.019; t statistic for the difference = 30.94). Similarly, the coefficient estimate of state-level *hy* is −0.059 in the low sophistication subsample and it is −0.044 in the high sophistication subsample (difference in estimates = −0.015; t statistic for the difference = −9.05). Overall, our evidence suggests that return predictability is strong in states with low investor sophistication and weak in states with high investor sophistication.

4.3 | Local ownership and sophistication-induced predictability

Next, we focus on the low sophistication subsample and test whether our evidence of return predictability is stronger within states where investors exhibit stronger local bias. If local investors hold a greater proportion of local firms (i.e., local bias is high), the impact of investor sophistication on asset prices is likely to be stronger.

To test this prediction, we divide the sample of low sophistication states into low (below median) and high (above median) local bias states. The state-level local bias measure is the difference between the percentage invested locally (within the state) by local investors and the percentage market capitalization of local in-state firms (i.e., expected ownership). The estimates for the low sophistication-low local bias and low sophistication-high local bias subsamples are reported in columns (5) and (6) of Table 7, respectively. In column (7), we report the difference in the estimates between the high and low local bias subsamples.

TABLE 7 Predictability regression estimates

	1	2	3	4	5	6	7
	Average of recursive regressions						
	Baseline: Char. Adj.	*Low Soph*	*High Soph*	*Low − High Soph 2 − 3*	*Low Soph Low LB*	*Low Soph High LB*	*High – Low LB 6 − 7*
State-level predictors							
Inc Gr (× 100)	−0.013	−0.003	−0.017	0.014	−0.009	−0.089	−0.080
	39%	10%	46%	1.22	15%	8%	−3.01
Rel Un	0.011	0.021	0.002	0.019	−0.011	0.041	0.052
	59%	98%	26%	30.94	27%	100%	17.26
hy	−0.051	−0.059	−0.044	−0.015	−0.037	−0.092	−0.055
	98%	100%	95%	−9.05	37%	78%	−10.28
U.S. macro predictors							
Inc Gr	0.041	−0.086	0.162	−0.248	−0.237	−0.320	−0.084
	20%	30%	23%	−3.89	20%	44%	−3.32
Rel Un	0.003	−0.004	0.009	−0.012	0.002	−0.020	−0.022
	34%	28%	21%	−3.87	6%	34%	−7.53
hy	0.028	0.029	0.032	−0.003	−0.027	0.004	0.031
	65%	59%	64%	−0.21	22%	46%	2.01
Other predictors							
log(1 + D/P)	0.317	0.646	−0.291	0.937	0.973	0.735	−0.237
	84%	98%	42%	33.78	62%	100%	−4.53
U.S. cay	−0.132	−0.297	−0.020	−0.277	−0.147	−0.540	−0.392
	65%	87%	7%	−2.45	13%	56%	−5.27
Paper-bill spread	−0.150	−0.337	0.190	−0.527	−0.380	−0.427	−0.047
	74%	96%	56%	−19.31	50%	87%	−0.77
Term spread	−0.226	−0.300	0.011	−0.311	−0.059	−0.371	−0.313
	54%	75%	35%	−4.95	22%	73%	−3.69
Default spread	−0.517	−0.860	−0.003	−0.857	−0.231	−1.916	−1.685
	73%	83%	0%	−13.03	18%	72%	−18.06
Adj R^2	0.026	0.049	0.008		0.048	0.083	
Obs (Ns × Quarters)	4,950	2,420	2,530		880	990	

Note: This table reports the results from one-quarter ahead OLS panel predictive regressions. We predict the quarterly state portfolio return in quarter t using lagged macroeconomic variables measured in quarter $t-1$ or $t-2$. The dependent variable is the characteristic-adjusted return computed using the Daniel et al. (1997) method. In column (1), we report the full sample estimates. In (2) and (3), we report estimates from subsamples with only the low sophistication states (below median), and only the high sophistication states (above median), respectively. In (4), we report the difference between the low and high sophistication subsample estimates. In (5) and (6), we report estimates from subsamples with only the low sophistication and low local bias states (below median), and only the low sophistication and high local bias states (above median), respectively. In (7), we report the difference between the low and high local bias subsample estimates. We report average estimates from recursive regressions where data until time $t-1$ are used to predict returns at time t. Except regressions (`4) and (7), the number below the estimates is the percentage of times that the recursive regression estimates are statistically significant at the 5% level. The t statistics use serial and cross-sectional correlation adjusted Driscoll and Kraay (1998) standard errors. In regressions (4) and (7), the number below the estimates is the t statistics for the differences across the regression estimates. The Adj. R^2 is the average over all the recursive regressions. The estimation period is from 1980 to 2008.

The subsample estimates show that the predictability is stronger in the low sophistication-high local bias subsample. We find that the average adjusted R^2 in the high local bias sample is about 8% whereas in the low local bias sample it is about 5%. More importantly, the predictive power of the state-level unemployment and state-level *hy* is significantly stronger in the low local bias sample: The coefficient estimate of state-level unemployment is 0.041 and the coefficient estimate of state-level *hy* is −0.092. These estimates are the highest across all specifications in Table 7 and they are statistically higher than the estimates in the low local bias subsample. Overall, the local ownership-based subsample indicate that our evidence of return predictability is stronger in states with high local bias.[4]

4.4 | Performance of trading strategies: Baseline estimates

Next, we use the predictability models from the low and high sophistication samples to formulate a set of trading strategies.[5] These trading strategies allow us to better quantify the economic significance of the estimated predictability models. Our trading strategies have no look-ahead bias because the predictability models are re-estimated every quarter using the available data at the end of that quarter. For each subsample, we obtain the performance estimates for the Long–Short portfolio, which is the difference between the Long and Short portfolio returns. The Long (Short) portfolio return is the value-weighted average of the return indices of the five states predicted to have the highest (lowest) characteristic-adjusted returns in the next quarter. We also compute the performance differential between the returns of low-sophistication and high-sophistication Long–Short portfolios.

For the trading strategies, the evaluation period is from 1978(Q3) to 2009(Q4), representing the largest period for which the state macrovariables are available. The evaluation period starts in 1978(Q3) because we require a 3-year pre-estimation period to obtain the first sets of state rankings. The state-level data from 1975 to 1979 are based on approximations. Moreover, for 2009, we form the trading portfolios with the performance rankings for 2008.

We report the monthly performance estimates of our trading strategy in Table 8. The performance estimates are consistent with our conjecture. Specifically, the characteristic adjusted returns of the Long–Short portfolios for low investor sophistication states is substantially higher from those in high investor sophistication states.

In particular, as we see in Table 8, the average characteristic-adjusted return of the low sophistication portfolio (estimate = 0.327, *t* statistic = 2.44) is higher than that of the high sophistication portfolio (average = −0.186, *t* statistic = −1.70). When we compute the performance differences between the low and high sophistication Long–Short portfolios, it is positive and economically meaningful in most months. The average value of the return differential is 0.513 with a *t* statistic of 2.94.

Next, we evaluate the performance of Long–Short portfolios using various factor models. They include the (i) the CAPM, (ii) the 4-factor model with the market (RMRF), size (SMB), value (HML), and momentum (UMD)[6] factors, (iii) the 7-factor model, containing the four factors in (ii) and three industry factors (IND), (iv) the 9-factor model with the seven factors in (iii) and short-term (STR) and long term reversal factors (LTR), and (v) the 10-factor model with the nine factors in (iv) and the liquidity factor (LIQ) of Pastor and Stambaugh (2003).[7]

We report the alpha estimates from the factor models in Table 8. Similar to the characteristic-adjusted returns, we find that the low sophistication portfolio alpha estimates are positive and those of the high sophistication subsample are negative. For example, the 9-factor monthly alpha estimate for the low and high sophistication portfolios are −0.487 (*t* statistic = 2.64) and −0.122 (*t* statistic = −0.94), respectively. Moreover, the difference between these two alpha estimates (= 0.605) is statistically significant (*t* statistic = 2.66).

In contrast to the baseline results using the investor sophistication index, the performance estimates using the population-wide sophistication measures are weak and always statistically insignificant. For example, the 9-factor monthly alpha for the low and high sophistication Long–Short portfolios are 0.066 (*t* statistic = 0.33) and 0.167 (*t* statistic = 1.21), respectively. The difference between the two alpha estimates is −0.101, which is not statistically significant (*t* statistic = −0.45). This evidence confirms that our investor-based sophistication index captures behavioral biases better than a general sophistication index that does not exclusively consider the behavior of stock market participants.

4.5 | Conditional performance estimates

The unconditional factor models show that the alpha estimates for the low sophistication subsample are high and statistically significant. One shortcoming of these factor models is that they assume that exposures to risk factors are constant. In this section, we use conditional factor models to account for the potential time-varying exposures to U.S. systematic risks.

Specifically, we obtain alpha estimates for the Long–Short portfolios using two conditional asset pricing models, which allow portfolio exposures to U.S. risk factors to vary with the U.S. business cycle. The first model is from Lettau and Ludvigson (2001a, b), and it includes eight factors: the three Fama and French factors (i.e., RMRF, SMB, and HML), the momentum factor (UMD), and the interactions of these factors with the mean-free lagged value of the U.S. *cay* residual. The *cay* is the difference between consumption (c) and its long-term value as predicted by the level of assets (a) and income (y).

TABLE 8 Trading strategy performance estimates

Ns = 5	*R*	*R* − *Rm*	*Char Adj*	*Ind Adj*	*α*CAPM	*α*4FF	*α*7Fact	*α*9Fact	*α*10Fact	*α*LL	*α*(4FE × NBER)	*α*(4FE × NBER × cay)
Full sample	0.387	0.378	0.301	0.182	0.432	0.344	0.370	0.389	0.430	0.276	0.303	0.231
	1.91	1.86	2.19	1.10	2.12	1.74	2.01	2.11	2.33	1.39	1.49	1.13
Investor sophistication												
Low	0.435	0.426	0.327	0.241	0.488	0.423	0.454	0.487	0.529	0.333	0.338	0.241
	2.05	2.01	2.44	1.48	2.29	2.07	2.44	2.64	2.87	1.64	1.62	1.16
High	−0.169	−0.168	−0.186	−0.148	−0.220	−0.144	−0.158	−0.118	−0.122	−0.107	−0.128	−0.101
	−1.25	−1.25	−1.69	−1.47	−1.64	−1.08	−1.22	−0.91	−0.94	−0.83	−0.94	−0.77
Low − *High*	0.604	0.594	0.513	0.388	0.708	0.568	0.612	0.605	0.652	0.439	0.467	0.342
	2.34	2.30	2.94	2.15	2.77	2.26	2.70	2.66	2.86	1.82	1.82	1.38
Population-wide sophistication												
Low	−0.036	−0.039	0.082	0.012	−0.007	0.050	0.050	0.066	0.093	0.043	0.127	0.105
	−0.17	−0.19	0.51	0.08	−0.04	0.25	0.25	0.33	0.46	0.21	0.62	0.51
High	0.244	0.238	0.068	0.090	0.244	0.153	0.170	0.167	0.185	0.103	0.122	0.081
	1.63	1.59	0.68	0.78	1.61	1.02	1.23	1.21	1.34	0.69	0.83	0.55
Low − *High*	−0.281	−0.276	0.015	−0.078	−0.252	−0.103	−0.120	−0.101	−0.092	−0.060	0.005	0.024
	−1.26	−1.24	0.08	−0.45	−1.12	−0.45	−0.54	−0.45	−0.41	−0.27	0.02	0.11
Professionals (Population-wide)												
Low	−0.047	−0.051	0.095	−0.047	−0.027	−0.006	0.002	0.061	0.096	−0.040	0.072	0.022
	−0.23	−0.25	0.59	−0.31	−0.13	−0.03	0.01	0.31	0.48	−0.20	0.35	0.11
High	0.260	0.251	0.031	0.188	0.304	0.220	0.235	0.247	0.268	0.182	0.181	0.147
	1.45	1.40	0.26	1.31	1.69	1.26	1.41	1.48	1.60	1.04	1.03	0.83
Low − *High*	−0.307	−0.301	0.064	−0.235	−0.331	−0.226	−0.233	−0.186	−0.172	−0.223	−0.109	−0.125
	−1.35	−1.32	0.33	−1.34	−1.44	−0.96	−0.99	−0.79	−0.72	−0.94	−0.47	−0.53

Note: This table reports the performance estimates of trading strategies defined using the return prediction model. We report the performance estimates of “Long” minus “Short” portfolios. The long portfolio is a value-weighted portfolio of the state portfolios for the U.S. states predicted to have the highest five characteristic-adjusted returns in the next quarter. The “Short” portfolio is a value-weighted portfolio of the state portfolios for the U.S. states predicted to have the lowest five characteristic-adjusted returns in the next quarter. State portfolios with fewer than ten firms are excluded from the analysis. We compute two long-minus-short portfolios: one for low sophistication states and one for high sophistication states. We use three indices of sophistication: investor sophistication based on measures from the retail investor data set, population-wide sophistication based on demographic characteristics from the Census, and portion of state population in white collar jobs (professionals) from the Census. In the table, we report performance estimates for the long-minus-short portfolios from the low and high sophistication subsamples as well as the difference between the low and high sophistication performance estimates. We report the raw, market-adjusted, characteristic-adjusted, and industry-adjusted performance estimates. The *t* statistics for the mean estimates are reported in parentheses below the estimates. We also report alpha estimates from various unconditional and conditional factor models. The unconditional factor models contain some combination of the following factors: the market factor (RMRF), the size factor (SMB), the value factor (HML), the momentum factor (UMD), three industry factors (IND1, IND2, and IND3), two reversal factors (short-term reversal (STR), long-term reversal (LTR)), and the liquidity (LIQ) factor. The conditional factor models contain some combination of the RMRF, SMB, HML, and UMD factors, and interactions between these factors and two U.S. economic indicators. Specifically, the LL model includes the RMRF, SMB, HML, and UMD factors, and the interactions between these four factors and the mean-free, lagged cay residual of Lettau and Ludvigson (2001a, b). The NBER model includes the RMRF, SMB, HML, and UMD factors, and the interactions between these four factors and the U.S. recession dummy variable NBER. The NBER variable is set to one for quarters in which the U.S. economy experienced a contraction according to the National Bureau of Economic Research (NBER). The evaluation period is from 1978(Q3) to 2009(Q4). Specifically, to determine the portfolio held during 1978(Q3), we use data that where available from 1975(Q3) to 1978(Q2), for the 1978(Q4) portfolio we use data from 1975(Q3) to 1978(Q3), and so forth. Due to data restrictions, we follow Korniotis and Kumar (2013b) and approximate the state-level data for the period 1975 to 1979. For the portfolio in 2009, we use the state rankings predicted with data as of 2008(Q4).

TABLE 9 Trading strategy performance conditional on the components of the investor sophistication index

	R	*R − Rm*	*Char Adj*	*Ind Adj*	*αCAPM*	*α4FF*	*α7Fact*	*α9Fact*	*α10Fact*	*αLL*	*α(4FF × NBER)*	*α(4FF × NBER × cay)*
Original investor sophistication index	0.604	0.594	0.513	0.388	0.708	0.568	0.612	0.605	0.652	0.439	0.467	0.342
	2.34	2.30	2.94	2.15	2.77	2.26	2.70	2.66	2.86	1.82	1.82	1.38
Cognitive ability (investors)	−0.214	−0.214	−0.025	−0.122	−0.200	−0.224	−0.218	−0.177	−0.152	−0.261	−0.232	−0.278
	−0.95	−0.95	−0.13	−0.70	−0.88	−0.96	−0.95	−0.77	−0.66	−1.17	−0.98	−1.22
Portfolio turnover	0.098	0.098	−0.106	0.097	0.117	0.087	0.092	0.066	0.045	0.091	0.089	0.109
	0.43	0.43	−0.61	0.55	0.51	0.40	0.42	0.30	0.20	0.41	0.40	0.49
Local bias	0.014	0.008	0.055	−0.062	0.062	−0.182	−0.160	−0.191	−0.159	−0.242	−0.284	−0.344
	0.06	0.03	0.32	−0.35	0.25	−0.84	−0.77	−0.91	−0.75	−1.13	−1.30	−1.57
Percentage in foreign stocks	0.459	0.445	0.377	0.335	0.567	0.575	0.590	0.605	0.606	0.571	0.544	0.546
	2.23	2.16	2.22	1.99	2.81	2.75	3.06	3.11	3.10	2.69	2.54	2.51
Portfolio concentration (normalized variance)	−0.753	−0.748	−0.629	−0.504	−0.913	−0.754	−0.800	−0.811	−0.829	−0.605	−0.695	−0.534
	−3.04	−3.02	−3.65	−2.81	−3.81	−3.14	−3.76	−3.81	−3.88	−2.62	−2.81	−2.24
Lottery preferences	−0.611	−0.607	−0.524	−0.307	−0.681	−0.504	−0.537	−0.564	−0.620	−0.385	−0.496	−0.364
	−2.78	−2.75	−3.16	−1.95	−3.10	−2.26	−2.56	−2.75	−3.03	−1.80	−2.18	−1.65
New (concise) investor sophistication index	0.547	0.542	0.424	0.483	0.680	0.622	0.656	0.647	0.666	0.510	0.565	0.464
	2.12	2.10	2.14	2.50	2.68	2.43	2.84	2.79	2.85	2.03	2.16	1.80

Note: This table reports the performance differences between "Long" minus "Short" portfolios from low sophistication U.S. states and high sophistication U.S. states. We define sophistication based on the original investor sophistication index, its six components (cognitive abilities, portfolio turnover, local bias, foreign ownership, portfolio concentration, and lottery preferences), and a new (concise) sophistication index that adds the standardized value of foreign ownership, subtracts the standardized values of portfolio concentration and lottery preferences, and divides by 3. The evaluation period is from 1978(Q3) to 2009(Q4). For more details, see Table 8.

The second conditional model is based on eight factors: RMRF, SMB, HML, UMD, and the interactions of these factors with a recession indicator REC. REC is a dummy variable that takes the value of one for quarters that are identified as recession quarters by the National Bureau of Economic Research (NBER).

We report the conditional alpha estimates in the last two columns of Table 8. We find that the alpha estimates weaken when we use conditional factor models to account for portfolio risk. For example, the alpha estimates for the low sophistication portfolio, when we use the Lettau and Ludvigson model (αLL), and the conditional model with NBER recession interactions ($\alpha(4FF \times NBER)$) are about 0.33. These estimates are lower than the unconditional factor model alpha estimates that range from 0.423 to 0.529. Furthermore, the statistical significance of the alpha estimates is weaker when we use conditional factor models to account for risk. However, the differential in the conditional alpha estimates between the low and high sophistication portfolios is high (about 0.45) and statistically significant at the 10% level (*t* statistics are about 1.80).

4.6 | Performance estimates using the components of sophistication index

The investor sophistication index combines six measures obtained from the retail investor data set. Next, we focus on each measure separately. We define high and low sophistication states with each measure, estimate predictability regressions for the two subsamples, and form Long-Short portfolios. In Table 9, we report performance differentials between the Long-Short portfolio in the low sophistication subsample and the high sophistication subsample.

The evidence in Table 9 shows that three components of the sophistication index can successfully separate low and high sophistication states. They are the percentage in foreign stocks, portfolio concentration, and lottery preferences. To focus on these three components only, we define a new concise index of investor sophistication, where we add the standardized value of the percentage in foreign stocks, subtract the standardized values of portfolio concentration and lottery preference, and divide the sum by three. We report the performance differentials between the new Long–Short portfolios from the low and high sophistication subsamples in the last row of Table 9. These differentials are similar to those from the original investor sophistication index. Therefore, the original index is not significantly affected by the fact that some of its components cannot separate low and high sophistication states effectively.

4.7 | Subperiod estimates

For robustness, in unreported results, we obtain the trading strategy performance for subperiods to better align the measurement periods of the sophistication index and the state portfolio returns. We first compute, performance estimates for the 1978(Q1) to 1994(Q4) and 1995(Q1) to 2009(Q4) periods. We find that in both subperiods, the evidence of predictability is stronger in states with low sophistication levels. For example, when the evaluation period is from 1995(Q1) to 2009(Q4), the 9-factor monthly alpha differential between the Long–Short portfolios for the low and high sophistication subsamples is 1.072% (*t* statistic = 2.78) when we use the original investor sophistication index. When we use the concise investor sophistication index, the alpha estimate is 1.098% (*t* statistic = 2.78).

In our subperiod analysis, we also consider the 1991 to 2006 period, which excludes the recent years of the financial crisis and starts in 1991, which is the beginning of the retail investor dataset. Again, the performance differentials between the low and high sophistication portfolios are economically large. For example, when we use the original sophistication index, the difference in the average monthly characteristic-adjusted returns between the low and high sophistication portfolios is 0.559 (*t* statistic = 2.32). Similarly, regardless of the sophistication index employed, the differences in the alpha estimates are high.

Overall, the performance of the trading strategy indicates that investor sophistication affects local return predictability. Specifically, we find that the returns of firms located in less sophisticated states exhibit greater predictability because changes in local macroeconomic conditions have a stronger influence on the portfolio decisions of local investors.

5 | SUMMARY AND CONCLUSION

The degree of investor sophistication is an important building block of behavioral models of asset pricing. Theory predicts that less sophisticated investors are likely to exhibit stronger behavioral biases, and if those biases are systematic, they could affect asset prices. While this is an appealing conjecture, it has been difficult to test this directly as it is often difficult to quantify the sophistication level of the shareholders of individual firms.

In this paper, we examine two aspects of asset pricing where investor sophistication can potentially be important: (a) return comovement and (b) return predictability. Our evidence indicates that investors in less sophisticated areas exhibit stronger trading correlations, and correspondingly, the returns of firms headquartered in less sophisticated areas are more strongly correlated, especially when local economic conditions are unusually bad or good. Furthermore, we show that local economic conditions have a greater ability to predict local stock returns in the U.S. states with less sophisticated retail investors. During the 1978–2009 period, a trading strategy that exploits the predictive ability of local economic indicators earns 5%–7% higher risk-adjusted returns in the U.S. regions with less sophisticated investors. These asset pricing results are driven by the sophistication of actual local investors, and not by the characteristics of the broader local population. Our results disappear when we use population-wide measures of sophistication.

Taken together, these results show that geographical variation in the level of investor sophistication influences local asset prices. In future work, it would be interesting to examine the evidence of local return predictability using better measures of local investor sophistication. It would also be useful to examine whether investor sophistication influences corporate policies through their potentially more effective monitoring activities.

ACKNOWLEDGMENTS

We thank Jawad Addoum, Vidhi Chhaochharia, Stefanos Delikouras, Tarun Mukherjee (the Editor), and seminar participants at the University of Miami for helpful comments and suggestions. We are responsible for all remaining errors and omissions.

ORCID

George M. Korniotis https://orcid.org/0000-0002-6706-3212
Alok Kumar https://orcid.org/0000-0002-7350-1397

ENDNOTES

[1] For example, DeLong et al. (1990), Odean (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), Hong and Stein (1999), Barberis, Shleifer, and Wurgler (2005), Coval and Shumway (2005), Grinblatt and Han (2005), and Statman, Thorley, and Vorkink (2006).

[2] Prior research that examines stock return comovement related to investor trading behavior includes Kumar and Lee (2006), Kumar, Page, and Spalt (2013, 2016), Greenwood and Thesmar (2011). Studies that provide evidence of return predictability resulting from investor behavior include, among others, Cohen and Frazzini (2008) and Korniotis and Kumar (2013a).

[3] To compute the EC index, we follow Bernile, Kumar, and Sulaeman (2015). Specifically, we proxy for economic presence using a text-based measure which counts the frequency with which a state is mentioned in the business description of a firm's 10-K report. For each state mentioned in the firm's 10-K, we compute a citation share as the fraction of state references comprised by that state. For each firm, we identify an EC state, which is the non-HQ state with the highest citation share. The corresponding EC state index is a value-weighted excess stock return of the portfolio of stocks where the firm has an economic presence in that state, weighted by the firm's market value times its citation share for that state. We exclude the firm's own stock return when computing its corresponding HQ and EC state indices.

[4] In untabulated results, we find that return comovement is stronger in regions with less sophisticated investors and higher local ownership. However, these results are not statistically significant when controlling for local demographic characteristics.

[5] For the trading strategy analysis, we do not use the subsamples based on low sophistication *and* high and low local bias because these samples are comprised by a small number states that makes it prohibitive to form meaningful long and short portfolios. Specifically, the low sophistication-low local bias sample includes only eight states whereas the low sophistication-high local bias sample includes nine states.

[6] The UMD factor was proposed by Jegadeesh and Titman (1993) and Carhart (1997).

[7] The three industry factors are calculated using the Pastor and Stambaugh (2002) method and are designed to capture industry momentum (Grinblatt & Moskowitz, 1999; Hong, Torous, & Valkanov, 2007). The liquidity factor is obtained from Robert Stambaugh's web site. All other factors are from Kenneth French's data library.

REFERENCES

Alti, A., & Tetlock, P. C. (2014). Biased beliefs, asset prices, and investment: A structural approach. *Journal of Finance*, *69*, 325–361. https://doi.org/10.1111/jofi.12089

Asdrubali, P., Sorensen, B. E., & Yosha, O. (1996). Channels of interstate risk sharing: United States 1963–1990. *Quarterly Journal of Economics*, *111*, 1081–1110. https://doi.org/10.2307/2946708

Athanasoulis, S. G., & van Wincoop, E. (2001). Risk sharing within the United States: What do financial markets and fiscal federalism accomplish? *The Review of Economics and Statistics*, *83*, 688–698. https://doi.org/10.1162/003465301753237768

Barber, B., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, *55*, 773–806. https://doi.org/10.1111/0022-1082.00226

Barber, B., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, *116*, 261–292. https://doi.org/10.1162/003355301556400

Barber, B., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, *21*, 785–818. https://doi.org/10.1093/rfs/hhm079

Barber, B., Odean, T., & Zhu, N. (2009). Do retail trades move markets? *Review of Financial Studies*, *22*, 151–186. https://doi.org/10.1093/rfs/hhn035

Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, *75*, 283–317. https://doi.org/10.1016/j.jfineco.2004.04.003

Becker, B. O. (2007). Geographical segmentation of U.S. capital markets. *Journal of Financial Economics*, *85*, 151–178. https://doi.org/10.1016/j.jfineco.2006.07.001

Becker, B. O., Ivkovich, Z., & Weisbenner, S. (2010). Local dividend clienteles. *Journal of Finance*, *66*, 655–683. https://doi.org/10.1111/j.1540-6261.2010.01645.x

Bernile, G., Kumar, A., & Sulaeman, J. (2015). Home away from home: Geography of information and local investors. *Review of Financial Studies*, *28*, 2009–2049. https://doi.org/10.1093/rfs/hhv004

Brown, J. R., Ivkovic, Z., Smith, P. A., & Weisbenner, S. (2008). Neighbors matter: Causal community effects and stock market participation. *Journal of Finance*, *63*, 1509–1531. https://doi.org/10.1111/j.1540-6261.2008.01364.x

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*, 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Cohen, L., & Frazzini, A. (2008). Economic links and predictable returns. *Journal of Finance*, *63*, 1977–2011. https://doi.org/10.1111/j.1540-6261.2008.01379.x

Coval, J., & Shumway, T. (2005). Do behavioral biases affect prices? *Journal of Finance*, *60*, 1–34. https://doi.org/10.1111/j.1540-6261.2005.00723.x

Daniel, K. D., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance*, *52*, 1035–1058. https://doi.org/10.1111/j.1540-6261.1997.tb02724.x

Daniel, K. D., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance*, *53*, 1839–1886. https://doi.org/10.1111/0022-1082.00077

DeLong, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*, 703–738. https://doi.org/10.1086/261703

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, *80*, 549–560. https://doi.org/10.1162/003465398557825

Fama, E. F., & French, K. R. (1993). Common risk factors in returns on stocks and bonds. *Journal of Financial Economics*, *33*, 3–56. https://doi.org/10.1016/0304-405x(93)90023-5

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. (2012). *Entrepreneurship and urban growth: An empirical assessment with historical mines*. NBER Working Paper No. 18333.

Greenwood, R., & Thesmar, D. (2011). Stock price fragility. *Journal of Financial Economics*, *102*, 471–490. https://doi.org/10.1016/j.jfineco.2011.06.003

Grinblatt, M., & Han, B. (2005). Prospect theory, mental accounting, and momentum. *Journal of Financial Economics*, *78*, 311–339. https://doi.org/10.1016/j.jfineco.2004.10.006

Grinblatt, M., & Moskowitz, T. J. (1999). Do industries explain momentum? *Journal of Finance*, *54*, 1249–1290. https://doi.org/10.1111/0022-1082.00146

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, *54*, 2143–2184. https://doi.org/10.1111/0022-1082.00184

Hong, H., Torous, W., & Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics*, *83*, 367–396. https://doi.org/10.1016/j.jfineco.2005.09.010

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for market efficiency. *Journal of Finance*, *48*, 65–91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Kanazawa, S. (2006). IQ and the wealth of states. *Intelligence*, *34*, 593–600. https://doi.org/10.1016/j.intell.2006.04.003

Korniotis, G. M. (2008). Habit formation, incomplete markets, and the significance of regional risk for expected returns. *Review of Financial Studies*, *21*, 2139–2172. https://doi.org/10.1093/rfs/hhn074

Korniotis, G. M., & Kumar, A. (2011). Do behavioral biases adversely affect the macro-economy? *Review of Financial Studies*, *24*(5), 1513–1559. https://doi.org/10.1093/rfs/hhq110

Korniotis, G. M., & Kumar, A. (2013a). State-level business cycles and local return predictability. *Journal of Finance*, *68*(3), 1037–1069. https://doi.org/10.1111/jofi.12017

Korniotis, G. M., & Kumar, A. (2013b). Do portfolio distortions reflect superior information or psychological biases? *Journal of Financial and Quantitative Analysis*, *48*, 1–45. https://doi.org/10.1017/S0022109012000610

Kumar, A. (2009). Hard-to-value stocks, behavioral biases, and informed trading. *Journal of Financial and Quantitative Analysis*, *44*(6), 1375–1401. https://doi.org/10.1017/S0022109009990342

Kumar, A., & Lee, C. M. C. (2006). Retail investor sentiment and return comovements. *Journal of Finance*, *61*(5), 2451–2486. https://doi.org/10.1111/j.1540-6261.2006.01063.x

Kumar, A., Page, J., & Spalt, O. (2011). Religious beliefs, gambling attitudes, and financial market outcomes. *Journal of Financial Economics*, *102*, 671–708. https://doi.org/10.1016/j.jfineco.2011.07.001

Kumar, A., Page, J., & Spalt, O. (2013). Investor sentiment and return comovements: Evidence from stock splits and headquarters changes. *Review of Finance*, *17*(3), 921–953. https://doi.org/10.1093/rof/rfs010

Kumar, A., Page, J., & Spalt, O. (2016). Gambling and comovement. *Journal of Financial and Quantitative Analysis*, *51*(1), 85–111. https://doi.org/10.1017/S0022109016000089

Lettau, M., & Ludvigson, S. (2001a). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, *21*, 815–849. https://doi.org/10.1111/0022-1082.00347

Lettau, M., & Ludvigson, S. (2001b). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, *109*, 1238–1286. https://doi.org/10.1086/323282

Lustig, H., & van Nieuwerburgh, S. (2005). Housing collateral, consumption insurance, and risk premia: An empirical perspective. *Journal of Finance*, *60*, 1167–1219. https://doi.org/10.1111/j.1540-6261.2005.00759.x

Malloy, C. J., Moskowitz, T. J., & Vissing-Jørgensen, A. (2009). Long-run stockholder consumption risk and asset returns. *Journal of Finance*, *64*, 2427–2479. https://doi.org/10.1111/j.1540-6261.2009.01507.x

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *Journal of Finance*, *53*, 1887–1934. https://doi.org/10.1111/0022-1082.00078

Odean, T. (1999). Do investors trade too much? *American Economic Review*, *89*, 1279–1298. https://doi.org/10.1257/aer.89.5.1279

Pastor, L., & Stambaugh, R. F. (2002). Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, *63*, 315–349. https://doi.org/10.1016/S0304-405X(02)00064-8

Pastor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*, 642–685. https://doi.org/10.1086/374184

Statman, M., Thorley, S., & Vorkink, K. (2006). Investor overconfidence and trading volume. *Review of Financial Studies*, *19*, 1531–1565. https://doi.org/10.1093/rfs/hhj032

How to cite this article: Korniotis GM, Kumar A, Page JK. Investor sophistication and asset prices. *Rev Financ Econ*. 2019;00:1–23. https://doi.org/10.1002/rfe.1093