

# Investor Sophistication and the Participation, Home Bias, Diversification, and Employer Stock Puzzles

MILES S. KIMBALL and TYLER SHUMWAY\*  
University of Michigan

September 21, 2006

## Abstract

Using data from the April 2005 Survey of Consumer Attitudes, we develop an index of investor sophistication from a set of 15 quiz-like questions. We correlate our measure of sophistication with stock market participation, percentage of wealth in stocks, holdings of international investments, measures of diversification, and holdings of an employer's stock. We find that each of these variables is correlated with sophistication, with more sophisticated investors consistently behaving the way that financial economists would recommend. We also regress sophistication on some simple measures of financial education, finding evidence that financial education might help investors overcome suboptimal behavior.

---

\*Kimball is at the Department of Economics, University of Michigan. Shumway is at the Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109. Kimball can be reached at 734-764-2375 or kimball@umich.edu. Shumway can be reached at 734-763-4129 or shumway@umich.edu. We thank Matthew Shapiro and Bob Willis for comments. We are particularly grateful to Jim Lepkowski and his students for helping us design our survey questions and collect the data. This research is supported by a grant from the National Institutes of Aging.

Researchers have identified a number of puzzling investing behaviors by households. Many households do not participate in the stock market despite the large equity premium that stockholders have earned in the past. Among households that do participate in the market, many neglect to hold stocks of companies located outside their home country, hold undiversified portfolios, or hold the stock of the company they work for. Researchers have attempted to explain each of these behaviors in various ways. In this paper, we test the hypothesis that financially unsophisticated households are particularly likely to exhibit each of these behaviors.

The puzzles mentioned above have been studied extensively, with a small academic literature discussing each of them. However, very little research has attempted to link these puzzles together. Relatively little research has correlated these behaviors with financial sophistication, in part because sophistication is generally difficult to measure. Instead, researchers have attempted to explain most of these behaviors with information asymmetries, costs, holdings of other (nonfinancial) assets, or nonpecuniary motives to invest. If these puzzling behaviors are correlated with each other and with investor sophistication, it is unlikely that they can be completely explained in the context of fully rational, frictionless markets.

We test our hypotheses with data from the April 2005 Survey of Consumer Attitudes. We were able to suggest questions for this particular survey, and our analysis largely uses questions that we proposed. While there are potential disadvantages to using survey data in research, one advantage of using these data is that we can measure a number of relevant quantities simultaneously.

The stylized fact that many households do not hold stocks has been named the participation puzzle, and has been documented by Mankiw and Zeldes (1991), Haliassos and Bertaut (1995), Campbell (2006) and several other authors. Explanations for this phenomenon include the existence of fixed costs of investing and holdings of nonfinancial risky assets.<sup>1</sup> The fixed cost to investing can be thought of as the cost of becoming sufficiently financially sophisticated to learn how to invest. However, we do not expect investors with nonfinancial assets to be particularly unsophisticated.

The home bias, first documented by French and Poterba (1991), is the stylized fact that despite the substantial diversification gains investors can achieve by holding stocks outside of their home countries, they typically hold only domestic stocks. One potential explanation of this fact is that investors have more information about companies that are close to them geographically than they do about foreign companies. Consistent with this hypothesis, Coval and Moskowitz (1999) show that

---

<sup>1</sup>See, for example, Vissing-Jørgensen (2002).

mutual fund managers are particularly likely to hold stocks of companies with headquarters that are geographically close. Another potential explanation maintains that investors have a preference for holding domestic stocks because they feel a certain affinity for their home country (Morse and Sive, 2006).

Goetzmann and Kumar (2005) and Ivkovich, Sialm and Weisbenner (2005) show that investors are not generally well diversified. Using data from a large discount brokerage firm in the U.S., the median number of stocks held in an account is two. While Goetzmann and Kumar (2005) relate this lack of diversification to some measures of investor sophistication, Ivkovich, Sialm and Weisbenner find that individuals that are less diversified have better average returns. Thus, one potential explanation of a lack of diversification is information trading. An alternative explanation is that transactions costs make diversification prohibitively expensive.

The employer stock puzzle has been documented by Benartzi (2001). The observation that employees are particularly likely to hold their company's stock is sufficiently contrary to what financial economists would prescribe that no reasonable explanations of Benartzi's evidence have been advanced. However, it seems reasonable to think that some sort of company loyalty motivates people to hold their employers' stocks. It may also be the case that employers frequently offer company stock as a primary option for retirement savings, so individuals hold company stock by default.

Considering each of these puzzles, a natural hypothesis is that people sometimes make mistakes in their financial choices. It is also natural to hypothesize that investors that are relatively unsophisticated financially will make more of these mistakes than more sophisticated investors. This is the central hypothesis of this paper.

Another natural hypothesis that we would like to test is that financial education helps investors become sophisticated. While we can measure financial education and correlate it with sophistication, we are somewhat concerned that financial education might be endogenously related to sophistication. Thus, we will report results on the estimated relation between education and sophistication, but we interpret our results with some caution.

The paper proceeds as follows. In the next section, we discuss our data and the methods that we use to make our statistical inferences. In the subsequent section, we discuss our empirical results in detail. Finally, we conclude with some thoughts about the implications of our research and some directions for future work.

## I. Data and Method

To test our hypotheses, we use the responses to a number of relevant questions on the April 2005 Survey of Consumer Attitudes administered by the Survey Research Center at the University of Michigan. The Survey of Consumer Attitudes is a monthly survey that uses a nationally representative sample of approximately 500 people to gauge consumers' attitudes about current economic conditions. The Survey Research Center publishes several Consumer Sentiment Indexes using data collected in this survey. The survey started in the 1940s, and has been conducted monthly since 1977.

We suggested the text of several of the questions that were asked in the April 2005 survey, but we did not have complete autonomy over the text of any of the questions.<sup>2</sup> Several of the survey questions are listed in the Appendix, along with a brief description of how we coded the responses to create our variables of interest. We divide the variables into three categories, including dependent variables, sophistication variables, and controls and instruments. Summary statistics for each of the variables appear in Tables 1 through 3. We code each of our variables so that the “correct” response is associated with a higher variable value.

Our sophistication variables are designed to measure the financial sophistication of respondents. Each of the variables represents the response to a quiz-like question. Some of the quiz-like questions ask for a true or false response, while others ask for a degree of agreement, for example strongly disagree, disagree, neither agree nor disagree, and so forth. To construct an index of investor sophistication, we conduct a principal components analysis on the responses to these questions. We analyze the first two factors that are generated by this analysis, finding that the first component is easily identified as a measure of investor sophistication. We use this sophistication index throughout the rest of the paper.

Our nine dependent variables measure the puzzling behavior that we want to explain. Three variables, *Partic*, *StkFrac* and *Frac>0* are related to the participation puzzle. *Partic* is simply an indicator variable that equals one if the respondent holds any stocks. *StkFrac* measures the fraction of a respondent's portfolio that consists of stock, including respondents that hold no stock. When respondents are asked about their stock holdings, they are given a number of categories to describe their holdings, and *StkFrac* is inferred from these categories.<sup>3</sup> *Frac>0* is the same as

---

<sup>2</sup>Some of our questions were inspired by the previous work of Hilgert, Hogarth and Beverly (2003).

<sup>3</sup>Responses of “less than a tenth” are assigned 5%, “between a tenth and a quarter” get 17.5%, “between a quarter

StkFrac, except that it is missing for investors that have no stocks. In our sample, 68 percent of the respondents participate in the stock market. Among all respondents, the average fraction of total assets in stock is 20 percent, and the fraction is 32 percent among stockholders. We hypothesize that each of these participation measures will increase with investor sophistication.

We have two dependent variables that measure respondents' propensity to invest in the stock of the company that employs them. Our most direct measure, EmplStk, is equal to one for respondents that hold stocks, work for companies with publicly traded stock, and do not hold company stock. Since a respondent must work for a publicly traded firm and hold stock for this measure to be defined, it is only available for 88 respondents. Those 88 respondents seem quite likely to hold company stock, with 62 percent of them reporting that they hold stock in the company that employs them. However, the question that asks about company stock is not perfectly designed (it asks about employer stock in mutual funds, for example) and it is not defined for the majority of respondents. Our second company stock dependent variable, ESAtt, measures respondents' attitude about holding the stock of an employer. Asked whether "employees should have the majority of their retirement funds in their current employers stock," 2% strongly agreed and 17% agreed. Again, ESAtt is coded so that higher values correspond to more correct answers, so disagreeing with the statement generates higher values of ESAtt. Given the way we code our variables, we hypothesize that both EmplStk and ESAtt will increase with sophistication.

We also have two variables that measure respondents' propensity to hold investments in companies located in other countries. Our direct holdings variable, HomeBs, is one for investors that report holding global or international mutual funds. Our attitude measure, HBAtt, measures the degree to which investors agree that "it is best to invest in domestic stock." We hypothesize that both of these variables will covary positively with sophistication.

Finally, we have two dependent variables that measure the diversification of respondents. Our stock diversification variable, StkDivr, measures the number of individual stocks held by an investor and is then calculated as one minus the inverse of this number. This allows the variable to be increasing and concave in the number of stocks held, just as the actual benefit of diversification is increasing and concave in the number of different assets held. Investors that hold stock mutual funds are assumed to be holding an equivalent of 3 different individual stocks, but our results are not sensitive to that particular assumption. We also measure respondents' diversification across  

---

and a half" get 37.5%, "between a half and three quarters" get 62.5% and "more than three quarters" get 80%.

different classes of assets by simply counting the number of different asset types held for investment purposes (stocks, bonds, real estate) in a variable called NumAsst.

Before discussing our regression results, we first describe the correlations between our various dependent variables reported in Table 2. Looking at the table, it is immediately clear that a number of the correlations are significantly positive, but none of them are significantly negative. While some of these correlations are positive almost by construction (for example, StkFrac and Frac>0), it is noteworthy that respondents that hold more stocks are more likely to hold international stocks and less likely to think it is a good idea to hold an employer's stock. Investors that hold more stock are also more likely to be more diversified. Our most direct measure of company stock holdings (EmplStk) is not significantly related to any of the other variables, presumably because it is not measured for a very large sample of respondents. These correlations are our first preliminary evidence that the puzzles we examine are related.

Our controls and instruments are variables that measure the demographic and economic characteristics of our respondents. Notable among our controls and instruments is Fin Educ, which is a very basic measure of financial education. Respondents are asked whether they have ever taken a class, read a book, participated in a workshop, or taken an online course on investing. Fin Educ is then the sum of the different types of financial education investors have received. Both the mean and the median of this variable are 1.

## II. Results

In this section we present and discuss our empirical results. First we discuss the construction of our sophistication index, then we describe regressions relating sophistication to our dependent variables. We also discuss some results that control for the potential endogeneity of sophistication, and we present some results on the relation of financial education to sophistication.

### A. Principal Components Analysis

Summary statistics about each of our 15 quiz-like sophistication questions appear in Table 3. For each question, Column 2 lists the percentage of respondents that gave the correct answer, and Column 3 reports the percentage that either claim to not know the answer or respond with a "neither agree nor disagree" statement. Most respondents seem to understand that diversification is

a good idea (question 8). Most know the definition of a mutual fund (question 1), and that checking accounts don't generally pay high rates of interest. However, almost half of the respondents agree with the statement that "if a stock's value is down, it will eventually come back up," and a majority agree with "investing in stocks is just like gambling." Overall, the responses to our sophistication questions appear close to what we might expect. Using 15 such questions, each with significant numbers of both correct and incorrect responses, gives us some confidence that we will be able to measure sophistication with some accuracy.

In order to construct an index of financial sophistication, we need to somehow aggregate individuals' responses to the sophistication questions. While we could simply sum up each respondents' correct responses to the questions, this simple method ignores any correlation between the responses and assumes that each question provides the same amount of information about respondents' sophistication. Rather than take this approach, we analyze the responses to our questions with a principal components analysis. The first two factors that our principal components analysis identifies are described in Columns 4 and 5 of Table 3.

The first factor that results from the principal components analysis is what we consider to be investor sophistication. This factor gives positive weight to all of the question variables (all of which are coded so that more correct answers correspond to higher variable values) except for the third question. The third question asks survey subjects whether they agree or disagree with the statement "I can usually tell when it is a good time to buy or sell stock." It seems likely that believing in market inefficiency is not highly correlated with being financially sophisticated in other ways (e.g. knowing market institutions, understanding interest rates and compounding). Questions 2, 9, 11, and 15 have the largest coefficients in our sophistication index. Apparently relatively unsophisticated investors like extremely safe investments and perceive banks quite favorably. Our index is scaled to have a mean of zero and a standard deviation of one.

The second factor that results from the analysis appears to be large for respondents that understand institutions but do not believe in market efficiency. For lack of a better label, we call this factor the "stock jock" index. People that load heavily on the stock jock index believe that they can tell when it is a good time to buy or sell stock, that there is an optimal time of year to invest, that stocks that have declined in value will eventually come back, and that investing in one type of stock makes sense. They also believe that if you want to do well in the stock market, you have to buy and sell often and that it is easy to make money in the market if you are smart. While the

stock jock factor is interesting, it does not seem to have much explanatory power for any of our dependent variables.

## B. Relating the Puzzles to Sophistication

Table 4 contains the results of a number of OLS regressions that relate our dependent variables to sophistication. The first panel of the table contains simple regressions of the dependent variables on sophistication alone. As can be seen from Panel A, all of these regressions, except the regression for `EmplStk`, generate significantly positive coefficient estimates on our sophistication variable. Since these are univariate regressions, the economic significance of the coefficients can be determined by looking at the  $R^2$  of each regression, reported in the penultimate row of the table. Sophistication explains almost 24 percent of the participation variable, and about 15 percent of attitudes toward holding an employer's stock. Sophistication explains about 15 percent of international asset holdings, but it has little explanatory power for respondents' attitudes about international assets. It also has economically significant explanatory power for diversification.

The second panel of Table 4 reports the results of regressions that include a large number of control variables. These controls include regional dummy variables, race dummy variables, polynomial terms in age and income, demographic variables such as gender, marital status, and children, pension variables and expectation variables. Even after controlling for each of these variables, our sophistication index is significantly related to all of our dependent variables except `EmplStk` and `NumAsst`. Since `EmplStk` is not significantly related to sophistication in univariate regressions, it is not surprising that it is still not significantly related to sophistication in the regression with controls. The coefficient on sophistication in the `NumAsst` regression is somewhat smaller than the coefficient in the univariate regression, but it is still of the correct sign and of a reasonable magnitude.

The economic significance of the coefficients in Table 4 can be assessed by recalling that our sophistication factor is scaled to have a mean of zero and a standard deviation of one. Thus, altering sophistication by one standard deviation increases participation by 12.3 percent. Since `Partic` is an indicator variable, this means that a standard deviation of sophistication increases the odds of holding stocks by approximately 12.3 percent, which is quite economically significant. Performing similar calculations for each of the other dependent variables, it is clear that statistically significant coefficients in this regression are generally also economically significant.



Looking at the statistical significance of the coefficients on the control variables in Panel B of Table 4, most of the coefficients are not significant in most of the regressions. Some notable exceptions include the variable that measures financial education, which has significant explanatory power for the home bias variable and for market participation, and the income and pension variables, which again have some power for the participation variables.

### C. Correcting for Possible Endogeneity

While the results of Table 4 are quite striking, it is possible that sophistication is endogenously related to some of our dependent variables. Suppose, for example, that some respondents are exogenously endowed with portfolios of stocks (perhaps by inheritance) and that owning stocks causes them to read books about investing and become sophisticated. In such a scenario, our results relating sophistication to participation will not be due to sophistication causing investors to participate, but they will be due to participation causing sophistication. To control for this sort of endogeneity, we estimate regressions of our dependent variables on sophistication and some controls with two-stage least squares. Our results are reported in Table 5.

The instruments that we use in our two-stage least squares estimates must be variables that are correlated with sophistication but uncorrelated with the residuals from regressions like those in Table 4. Intuitively, they must be quantities that cannot be correlated with the dependent variables by an endogeneity story like the story discussed in the previous paragraph. We choose to use education, age, race, region, gender and FEP as instruments. When respondents are asked probabilistic questions, people that are not mathematically sophisticated tend to answer with exact probabilities (e.g. 100 percent). FEP measures the “fraction of exact probability” given by respondents, or the fraction of four survey questions that are answered with exact probabilities of 0, 50 or 100 percent. Each of our instruments seems unlikely to be driven by our dependent variables. In the case of our age, region, race and gender variables, the results in Table 4 confirm that these variables are not generally significantly related to our dependent variables when sophistication is included in the regression. We exclude age, financial education, and the stock jock index from the second stage regressions.

Looking at the first-stage regression results reported in Table 5, it is noteworthy that education and age are very significant predictors of financial sophistication. Our first-stage regression also indicates that race is related to sophistication but FEP and region are not related. Our second-

stage results make it clear that using two-stage least squares decreases the power of our tests somewhat. However, the participation variables, the employer stock attitude variable and the home bias variable are all still significantly related to financial sophistication. Since the endogeneity story discussed above probably makes the most sense for participation, this is strong evidence in favor of our hypothesis.

#### **D. Explaining Sophistication**

Having successfully measured sophistication and related it to several puzzling behaviors, it is natural to ask what drives sophistication? While this is a very interesting question to ask, it is difficult for us to answer unambiguously because of the possibility of endogeneity of many variables that we might use to explain sophistication. Nevertheless, we make an attempt at answering this question with the regression reported in Table 6.

The regression reported in Table 6 explains sophistication with all of the controls and instruments used in the rest of the paper except for Fin Educ. Instead of Fin Educ, the components of Fin Educ are included as separate regressors in the analysis. These components include a dummy variable that indicates whether the respondent has taken an online investment course, one that corresponds to a classroom-based course, one that corresponds to participating in an investment workshop, and one that indicates whether the respondent has read books and/or articles on investing. The regression also includes several variables that control for whether the respondent relies on advice from other people or sources to make investing decisions. The variables labeled “Consult” followed by Media, Professional, Colleague, Friend and Family all indicate whether respondents consult regularly with these particular people or sources. These variables are coded with a value of 1 for respondents that “always” consult with the given source, they are coded with a 3 for “sometimes” consultations, and they are coded with a value of 5 when respondents indicate that they “never” consult the given source.

While we need to interpret these results with some caution, the estimates in Table 6 indicate that respondents with more education tend to be more financially sophisticated. Furthermore, those that read books or articles about investing and those that regularly consult with investment professionals also appear to be more sophisticated. The coefficient on market expectations is significantly positive in this regression, probably indicating that more sophisticated individuals are more optimistic about future returns than others. Overall, we interpret this as evidence that

financial education and investment professionals might help investors to become more financially sophisticated.

### **III. Conclusion**

We present survey evidence that the participation, home bias, diversification and employer stock puzzles are significantly related to each other and to investor sophistication. Using data from the April 2005 Survey of Consumer Attitudes, we construct an index of investor sophistication using responses to 15 quiz-like questions. We relate our sophistication index to investment behavior and attitudes about investment, finding that relatively sophisticated investors consistently behave the way that financial economists would expect them to. Regressing sophistication on several measures of financial education, we find evidence that financial education might help investors to become more sophisticated.

One positive reason to find our results important is that they help us think about potential explanations for the various puzzling behaviors that we observe in individuals' portfolio choices. If all of these puzzles are related to each other, and if relatively unsophisticated investors are particularly likely to exhibit these behaviors, it is unlikely that perfectly rational, frictionless models will ever be able to completely explain them.

One normative way to think about the value of our contribution is to interpret our estimates as measures of the potential benefit of making individuals more financially sophisticated. While a complete analysis of the welfare consequences of each of the puzzles we address is beyond the scope of this paper, our findings indicate that more sophistication will bring investors closer to the economist's ideal portfolio holdings. If it is relatively cheap to make people more financially sophisticated, that appears to be a goal worth pursuing.

## References

- Benartzi, Shlomo, 2001, Excessive extrapolation and the allocation of 401(k) accounts to company stock, *Journal of Finance* 56, 1747-1764.
- Campbell, John Y., 2006, Household finance, *Journal of Finance* 61, 1553-1604.
- Coval, Joshua D. and Tobias Moskowitz, 1999, Home bias at home: Local equity preference in domestic portfolios, *Journal of Finance* 54, 2045-2073.
- French, Kenneth R., and James M. Poterba, 1991, Investor diversification and international equity markets, *American Economic Review* 81, 222-226.
- Goetzmann, William N., and Alok Kumar, 2005, Why do individual investors hold under-diversified portfolios, working paper, Yale University.
- Haliassos, Michael, and Carol Bertaut, 1995, Why do so few hold stocks? *Economic Journal* 105, 1110-1129.
- Hilgert, Marianne A., Jeanne M. Hogarth, and Sondra G. Beverly, 2003, Household financial management: The connection between knowledge and behavior, *Federal Reserve Bulletin* 89, 309-322.
- Ivkovich, Zoran, Clemens Sialm, and Scott J. Weisbenner, 2005, Portfolio Concentration and the Performance of Individual Investors, working paper, University of Michigan.
- Mankiw, N. Gregory, and Stephen P. Zeldes, 1991, The consumption of stockholders and nonstockholders *Journal of Financial Economics* 29, 97-112.
- Morse, Adair, and Sophie Shive, 2006, Patriotism in your portfolio, working paper, University of Notre Dame.
- Vissing-Jørgensen, Annette, 2002, Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures, working paper, Northwestern University.

**Table 1: Summary Statistics**

Table 1 reports summary statistics for all the variables used in the regressions below. The data were collected during the April 2005 Survey of Consumer Attitudes, by telephone interviews conducted by researchers in the Survey Research Center at the University of Michigan. Each of the variables is described in the Appendix. The first nine variables are dependent variables in subsequent regressions, and the rest of the variables are either control variables or instruments.

Variable	Mean	Median	Std Dev	Minimum	Maximum	Obs
Dependent Variables						
Partic	0.681	1.0	0.466	0.0	1.0	490
StkFrac	20.036	5.0	25.557	0.0	80.0	411
Frac>0	32.294	17.5	25.630	0.0	80.0	255
EmplStk	0.375	0.0	0.487	0.0	1.0	88
ESAtt	3.728	4.0	1.029	1.0	5.0	485
HomeBs	0.417	0.0	0.494	0.0	1.0	276
HBAtt	3.041	3.0	0.909	1.0	5.0	482
StkDivr	0.711	0.7	0.234	0.0	1.0	253
NumAsst	1.640	2.0	0.696	1.0	3.0	253
Controls and Instruments						
Black	0.072	0.0	0.259	0.0	1.0	499
Hispanic	0.072	0.0	0.259	0.0	1.0	499
Asian	0.034	0.0	0.182	0.0	1.0	499
Caucasion	0.798	1.0	0.402	0.0	1.0	499
Age	49.768	49.0	16.443	18.0	90.0	499
Male	0.465	0.0	0.499	0.0	1.0	499
Married	0.561	1.0	0.497	0.0	1.0	499
Kids	0.699	0.0	1.120	0.0	9.0	499
Fin Educ	0.995	1.0	0.926	0.0	4.0	499
Education	14.168	14.0	2.832	4.0	18.0	499
Region 1	0.222	0.0	0.416	0.0	1.0	499
Region 2	0.271	0.0	0.445	0.0	1.0	499
Region 3	0.178	0.0	0.383	0.0	1.0	499
Region 4	0.329	0.0	0.470	0.0	1.0	499
Home Owner	0.776	1.0	0.418	0.0	1.0	499
Income	9.149	10.0	3.912	1.0	16.0	464
Pension	0.186	0.0	0.390	0.0	1.0	499
Has 401k	0.475	0.0	0.500	0.0	1.0	499
Market Exp.	50.233	50.0	29.302	0.0	100.0	480
Sentiment	72.945	76.0	39.139	-12.8	135.2	499
FEP	0.311	0.3	0.231	0.0	0.8	499

## Table 2: Correlations

Table 2 reports correlation coefficients for all of the dependent variables used throughout the paper. The variables are defined in the appendix. One asterisk indicates statistical significance at the 5% level and two asterisks indicates significance at the 1% level.

Variable	Partic	StkFrac	Frac>0	EmplStk	ESAtt	HomeBs	HBAtt	StkDivr
StkFrac	61.4**							
EmplStk	-	-6.3	-6.3					
ESAtt	19.4**	22.6**	16.2**	4.8				
HomeBs	-	18.7**	18.7**	-9.6	11.5			
HBAtt	-2.8	10.7*	18.7**	3.5	10.9*	20.4**		
StkDivr	-	26.4**	26.4**	3.5	11.9	20.0**	2.6	
NumAsst	23.6**	13.6*	4.0	-11.6	10.5	3.8	3.8	17.8**

### Table 3: Sophistication Index

Table 3 describes the inputs to the Sophisticate and Stock Jock indexes that we relate to the dependent variables described above. The indexes were calculated by performing principal components analysis on the responses to fifteen quiz-like questions about financial topics. The questions are listed in the Appendix. For each question, the percentage of correct responses is given in the second column, while the number of “don’t know” or “refuse to answer” responses is recorded in the third column. The fourth and fifth columns report the coefficients generated by the principal components analysis for the first two factors recoverable in the data. Throughout the paper we call the first factor “Sophisticate” and we call the second factor “Stock Jock.” Both factors are scaled to have zero mean and unit variance.

Question	Percent Correct	Percent DK or NA	Sophisticate Coefficient	Stock Jock Coefficient
1	85.0	7.2	0.1082	0.2432
2	84.0	4.4	0.1882	0.0348
3	67.3	12.6	-0.0398	-0.1533
4	51.3	22.9	0.0891	-0.3374
5	75.8	8.2	0.1434	0.0745
6	38.1	17.0	0.0679	-0.4367
7	39.1	11.8	0.1020	0.1263
8	87.8	7.0	0.1791	-0.0945
9	46.7	9.6	0.2256	-0.0291
10	73.0	13.8	0.1789	0.1623
11	73.7	14.6	0.2060	-0.0388
12	85.0	6.4	0.0777	0.3017
13	54.3	16.2	0.1276	-0.3465
14	35.1	13.4	0.1643	0.1325
15	64.1	13.6	0.2363	-0.0059

## Table 4: OLS Regressions

Table 4 presents the results of regressing our dependent variables on our sophistication index and several control variables. T-statistics are in parentheses. The regressions reported in Panel B include region dummies, but their coefficients are excluded from the table for space considerations. The sample size and adjusted  $R^2$  of each regression are reported in the last two rows of the table.

<b>Panel A: Simple Regressions</b>									
Independent	Dependent Variable								
Variable	Partic	StkFrac	Frac>0	EmplStk	ESAtt	HomeBs	HBAtt	StkDivr	NumAsst
Intercept	0.681	20.1	27.8	0.393	3.715	0.314	3.038	0.659	1.575
	(37.0)	(18.1)	(15.7)	(7.44)	(86.5)	(10.1)	(73.9)	(38.0)	(30.6)
Sophisticate	0.227	11.9	10.7	-0.087	0.410	0.249	0.113	0.101	0.142
	(12.4)	(10.9)	(5.14)	(-1.53)	(9.47)	(6.95)	(2.74)	(5.05)	(2.35)
Adj. $R^2$	23.8	22.3	9.1	1.5	15.5	14.7	1.3	8.9	1.8
No. Obs.	490	411	244	88	485	276	482	253	253



**Table 4: OLS Regressions (continued)**

Independent Variable	Panel B: Regressions with Controls								
				Dependent Variable					
	Partic	StkFrac	Frac>0	EmplStk	ESAtt	HomeBs	HBAtt	StkDivr	NumAsst
Intercept	0.316 (1.76)	-2.047 (-0.16)	8.617 (0.40)	0.725 (0.79)	3.607 (7.14)	-0.668 (-1.52)	2.696 (5.68)	0.407 (1.85)	0.572 (0.82)
Sophisticate	0.123 (5.87)	5.627 (3.85)	5.428 (2.02)	-0.010 (-0.11)	0.351 (6.05)	0.159 (3.50)	0.133 (2.43)	0.067 (2.78)	0.097 (1.32)
Stock Jock	0.027 (1.60)	0.073 (0.06)	-0.908 (-0.51)	-0.018 (-0.28)	-0.144 (-3.07)	0.007 (0.23)	-0.063 (-1.41)	0.005 (0.28)	0.029 (0.59)
Black	-0.002 (-0.03)	-2.162 (-0.45)	-2.963 (-0.34)	0.160 (0.78)	-0.545 (-2.96)	0.017 (0.13)	0.615 (3.53)	-0.053 (-0.75)	0.158 (0.67)
Hispanic	0.041 (0.64)	-3.029 (-0.69)	-9.242 (-1.40)	0.116 (0.50)	0.027 (0.14)	0.085 (0.76)	0.224 (1.20)	0.073 (1.09)	0.114 (0.56)
Asian	-0.060 (-0.65)	2.006 (0.34)	1.948 (0.22)	0.254 (0.43)	-0.362 (-1.41)	0.097 (0.56)	-0.006 (-0.03)	0.147 (1.70)	0.118 (0.38)
Age	-0.006 (-0.90)	0.142 (0.32)	0.344 (0.45)	-0.008 (-0.30)	-0.003 (-0.18)	0.022 (1.56)	0.021 (1.24)	0.007 (1.04)	0.013 (0.60)
Age <sup>2</sup> /1000	0.046 (0.73)	-0.568 (-0.13)	-1.728 (-0.23)	0.085 (0.28)	0.071 (0.41)	-0.239 (-1.73)	-0.186 (-1.13)	-0.069 (-1.02)	-0.034 (-0.16)
Male	-0.017 (-0.49)	0.461 (0.20)	1.913 (0.58)	-0.165 (-1.37)	0.039 (0.42)	0.126 (2.18)	-0.091 (-1.02)	0.037 (1.22)	0.186 (2.05)
Married	0.006 (0.16)	-0.066 (-0.02)	1.301 (0.34)	0.041 (0.25)	0.189 (1.73)	0.056 (0.84)	0.066 (0.64)	-0.016 (-0.45)	-0.167 (-1.49)
Kids	-0.034 (-2.09)	0.552 (0.50)	0.794 (0.46)	-0.059 (-0.87)	-0.051 (-1.05)	-0.030 (-0.98)	0.018 (0.39)	-0.012 (-0.78)	0.036 (0.78)
Fin Educ	0.038 (1.90)	2.916 (2.09)	2.538 (1.38)	-0.017 (-0.24)	-0.037 (-0.67)	0.089 (2.74)	0.022 (0.43)	0.021 (1.25)	-0.015 (-0.30)
Home Owner	0.076 (1.63)	0.742 (0.23)	-3.927 (-0.72)	0.257 (1.55)	-0.039 (-0.30)	-0.002 (-0.02)	-0.138 (-1.12)	0.026 (0.52)	-0.098 (-0.56)
Income	0.053 (2.77)	-1.501 (-1.17)	-1.996 (-0.78)	-0.003 (-0.03)	0.041 (0.77)	0.050 (1.02)	-0.040 (-0.79)	-0.005 (-0.20)	0.034 (0.44)
Inc <sup>2</sup>	-0.002 (-1.73)	0.169 (2.45)	0.168 (1.36)	-0.000 (-0.05)	-0.002 (-0.72)	-0.001 (-0.28)	0.004 (1.36)	0.001 (0.81)	0.002 (0.61)
Pension	0.157 (3.00)	4.781 (1.30)	0.865 (0.14)	-0.269 (-0.87)	-0.201 (-1.38)	0.078 (0.67)	-0.150 (-1.08)	-0.032 (-0.57)	-0.134 (-0.81)
Has 401K	0.336 (7.46)	9.668 (3.06)	0.835 (0.16)	-0.376 (-1.71)	-0.066 (-0.52)	-0.055 (-0.60)	-0.202 (-1.69)	0.020 (0.42)	-0.013 (-0.09)
Market Exp.	0.000 (0.69)	0.108 (2.66)	0.141 (2.35)	-0.000 (-0.24)	0.001 (0.41)	0.002 (1.63)	0.000 (0.31)	-0.001 (-0.95)	0.000 (0.24)
Sentiment	-0.001 (-1.89)	-0.014 (-0.45)	0.008 (0.19)	0.001 (0.88)	-0.001 (-0.62)	-0.001 (-1.76)	-0.001 (-0.75)	0.000 (0.53)	-0.003 (-2.76)
Adj. R <sup>2</sup>	45.7	32.3	9.62	-0.4	17.7	23.1	4.5	12.2	11.5
No. Obs.	447	381	248	83	441	261	439	241	236

**Table 5: Two-Stage Least Squares Regressions**

Table 5 presents the results of regressing our dependent variables on our sophistication index and several control variables, adjusting for possible endogeneity of sophistication by using two-stage least squares. Our first-stage regression for each of the estimates is:

$$\begin{aligned} \text{Sophisticate} = & -4.278 + 0.140\text{Educ} - 0.090\text{FEP} + 0.095\text{Age} - 0.894\text{Age}^2/1000 \\ & (-11.9) \quad (10.01) \quad (-0.53) \quad (7.37) \quad (-7.12) \\ +0.036\text{Male} & - 0.667\text{Black} - 0.209\text{Hispanic} - 0.401\text{Asian} + \sum_{j=1}^3 \beta_j \text{Region}_j + \epsilon, \\ & (0.48) \quad (-4.51) \quad (-1.37) \quad (-1.90) \quad (-) \end{aligned}$$

where the t-statistics are given in parentheses. The adjusted R<sup>2</sup> from our first-stage regression is 31.6 percent.

Independent Variable	Dependent Variable								
	Partic	StkFrac	Frac>0	EmplStk	ESAtt	HomeBs	HBAtt	StkDivr	NumAsst
Intercept	0.056 (0.69)	2.154 (0.40)	22.901 (1.74)	0.640 (1.12)	3.644 (15.74)	-0.555 (-2.09)	3.156 (14.67)	0.504 (3.71)	1.385 (3.44)
Sophisticate	0.112 (2.86)	6.800 (2.63)	9.274 (1.52)	-0.084 (-0.54)	0.406 (3.66)	0.338 (3.39)	0.119 (1.15)	0.079 (1.20)	0.141 (0.80)
Black	0.005 (0.07)	-0.617 (-0.12)	3.039 (0.30)	0.090 (0.41)	-0.522 (-2.66)	0.236 (1.53)	0.636 (3.48)	-0.035 (-0.47)	0.042 (0.17)
Hispanic	0.040 (0.61)	-2.956 (-0.67)	-7.785 (-1.12)	0.148 (0.61)	0.097 (0.49)	0.171 (1.40)	0.231 (1.26)	0.069 (1.00)	0.017 (0.08)
Asian	-0.041 (-0.46)	2.209 (0.38)	2.680 (0.30)	0.448 (0.79)	-0.299 (-1.18)	0.109 (0.61)	-0.027 (-0.11)	0.114 (1.29)	0.116 (0.36)
Male	-0.018 (-0.53)	0.150 (0.07)	1.280 (0.38)	-0.172 (-1.55)	-0.017 (-0.18)	0.132 (2.24)	-0.103 (-1.17)	0.042 (1.42)	0.183 (2.04)
Married	-0.001 (-0.02)	1.024 (0.39)	1.760 (0.47)	0.066 (0.51)	0.230 (2.16)	0.073 (1.06)	0.114 (1.14)	-0.006 (-0.18)	-0.131 (-1.16)
Kids	-0.033 (-2.18)	-0.047 (-0.05)	0.379 (0.23)	-0.059 (-1.10)	-0.066 (-1.45)	-0.017 (-0.58)	0.003 (0.08)	-0.012 (-0.85)	-0.011 (-0.27)
Home Owner	0.062 (1.40)	1.315 (0.42)	-1.986 (-0.38)	0.228 (1.40)	-0.076 (-0.60)	-0.002 (-0.02)	-0.114 (-0.96)	0.053 (1.10)	-0.048 (-0.28)
Income	0.064 (3.49)	-1.499 (-1.20)	-2.808 (-1.15)	-0.017 (-0.16)	0.022 (0.42)	0.087 (1.73)	-0.047 (-0.96)	-0.005 (-0.19)	0.000 (0.00)
Inc <sup>2</sup>	-0.002 (-1.91)	0.181 (2.66)	0.215 (1.81)	0.001 (0.14)	-0.000 (-0.15)	-0.002 (-0.84)	0.004 (1.65)	0.001 (0.89)	0.004 (1.00)
Pension	0.159 (3.14)	5.475 (1.55)	1.611 (0.27)	-0.264 (-0.93)	-0.172 (-1.22)	0.063 (0.52)	-0.121 (-0.92)	-0.012 (-0.22)	-0.033 (-0.20)
Has 401K	0.361 (8.19)	11.301 (3.66)	-0.48 (-0.10)	-0.420 (-2.02)	-0.123 (-0.99)	-0.036 (-0.39)	-0.158 (-1.36)	0.054 (1.18)	-0.004 (-0.03)
Market Exp.	0.001 (1.55)	0.124 (3.11)	0.151 (2.56)	-0.001 (-0.40)	0.002 (0.96)	0.002 (1.96)	0.001 (0.55)	-0.001 (-1.13)	0.001 (0.35)
Sentiment	-0.000 (-1.09)	-0.000 (-0.01)	0.008 (0.20)	0.002 (1.04)	-0.001 (-0.81)	-0.001 (-1.15)	-0.001 (-0.95)	0.000 (0.80)	-0.003 (-2.75)
Adj. R <sup>2</sup>	43.2	29.8	8.3	4.8	10.7	17.0	2.9	7.9	8.0

**Table 6: Explaining Sophistication**

Table 6 presents the results of regressing our sophistication index on all of our control variables and instruments plus some variables that measure financial education and financial advice. The sample size is 407n and the adjusted  $R^2$  is 35.6 percent. The variables about financial education (Online Course, Class, Workshop, Books/Articles) are indicator variables that are equal to one if the respondent indicates that he or she has learned about investments from such a source. The variables about consulting are equal to 1 if the respondent indicates that he or she always consults with the given source, is equal to 3 for sometimes consults, and is equal to 5 if the respondent indicates that he or she never consults with the given source.

<b>Dependent Variable: Sophisticate</b>			
Variable	Mean	Coefficient	T-statistic
Intercept	1.000	-3.348	-6.54
Black	0.059	-0.716	-4.36
Hispanic	0.059	-0.314	-1.89
Asian	0.029	-0.319	-1.38
Male	0.477	-0.018	-0.22
Married	0.592	0.041	0.45
Kids	0.703	0.011	0.28
Home Owner	0.818	-0.095	-0.84
Income	9.727	-0.030	-0.59
Inc <sup>2</sup>	106.985	0.004	1.40
Pension	0.199	0.032	0.26
Has 401K	0.555	0.062	0.60
Market Exp.	51.405	0.004	2.69
Sentiment	74.618	0.001	1.00
Education	14.604	0.069	3.75
Region 1	0.221	0.045	0.42
Region 2	0.278	0.111	1.12
Region 3	0.182	0.069	0.61
FEP	0.302	-0.059	-0.33
Age	49.39	0.064	4.19
Age <sup>2</sup> /1000	2.685	-0.572	-3.82
Online Course	0.027	-0.349	-1.47
Class	0.197	0.064	0.61
Workshop	0.199	0.191	1.87
Books/Articles	0.690	0.331	3.56
Consult Media	3.752	-0.024	-0.76
Consult Professional	3.133	-0.056	-2.05
Consult Colleague	4.002	0.037	0.84
Consult Friend	3.821	0.058	1.26
Consult Family	3.364	0.044	1.48

## Appendix: Survey Questions

We use the responses to a number of questions from the April 2005 Survey of Consumer Attitudes for our sample. The survey was asked of 499 people at least 18 years of age. The survey's sample is carefully designed to be nationally representative. In this appendix, we list some of the questions asked on the survey and then, in brackets, a brief description of how we coded the variable used in the study.

### A Dependent Variable Questions

**Partic:** The next questions are about investments in the stock market. First, do you (or any member of your family living there) have any investments in the stock market, including any publicly traded stock that is directly owned, stocks in mutual funds, stocks in any of your retirement accounts, including 401(K)s, IRAs, or Keogh accounts?

[yes = 1, no = 0]

**StkFrac:** Roughly speaking, what fraction of your (family's) financial assets, including any savings in retirement plans, is invested in stocks or stock mutual funds - would you say less than a tenth, between a tenth and a quarter, between a quarter and a half, between a half and three quarters, or more than three quarters?

[midpoint of range cited if Partic = 1 and stated value of investments > \$5,000, zero otherwise]

**Frac>0**

[= StkFrac; only defined if Partic = 1]

**EmplStk:** Do you own stock, including stock in 401ks, 403bs, or mutual funds, in the company you work for?

[yes = 0, no = 1; only defined if Partic = 1]

**ESAtt:** Employees should have the majority of their retirement funds in their current employers stock. (Would you say you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree?)

[strongly agree = 1, agree = 2, ..., strongly disagree = 5]

**HomeBs:**(Including whats in your retirement accounts,) do you have global or international mutual funds?

[yes = 1, no = 0; only defined if Partic = 1]

**HBAtt:**It is best to invest in domestic stock. (Would you say you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree?)

[strongly agree = 1, agree = 2, ..., strongly disagree = 5]

**StkDivr:** You said you have money invested in stocks, not including mutual funds. In how many different companies do you hold stocks?

[num-stocks = number reported]

Including what's in your retirement account, do you have stock mutual funds?

[if yes then add 3 to num-stocks; StkDivr = 1 - 1/num-stocks; only defined if Partic = 1]

**NumAsst:**

[from three separate questions, identify if respondents hold stocks, bonds, and real estate (excluding primary residence) - sum number of asset classes held]

## B Sophistication Questions

Survey respondents were read a number of statements, and they were asked to indicate whether the statements were true or false, or whether they agreed with the statements or not. The statements used to measure investor sophistication are listed below, with indicators of how we scored each question. Table 2 reports a principal components analysis of the data generated by these questions.

1. A mutual fund combines the money of many investors to buy a variety of stocks or bonds.  
[true = 1, false = -1, DK or NA = 0]
2. Checking accounts earn a higher rate of return than other types of investments purchased from a bank.  
[true = -1, false = 1, DK or NA = 0]
3. I can usually tell when it is a good time to buy or sell stock.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]
4. There is an ideal time of the year to invest.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]
5. Only brokers make money in the stock market.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]
6. If a stock's value is down, it will eventually come back up.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]
7. Investments with high risk are best for younger individuals.  
[strongly agree or agree = 1, strongly disagree or disagree = -1, DK or neither agree nor disagree = 0]
8. Investing in only one type of stock, like tech stocks, makes sense.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]
9. You should always put your money into the safest investment you can find.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]
10. Choosing to invest in both small and large companies at the same time is wise.  
[strongly agree or agree = 1, strongly disagree or disagree = -1, DK or neither agree nor disagree = 0]
11. Investments offered by a bank are the best investments you can buy.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]
12. The earlier in life that you invest, the better off you will be financially.  
[strongly agree or agree = 1, strongly disagree or disagree = -1, DK or neither agree nor disagree = 0]
13. If you are smart, it is easy to make money in the stock market.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]
14. Investing in stocks is just like gambling.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]
15. To do well in the stock market, you have to buy and sell your stocks often.  
[strongly agree or agree = -1, strongly disagree or disagree = 1, DK or neither agree nor disagree = 0]

## C Control Variables and Instruments

Variable	Definition
Black	1 if respondent identifies self as black
Hispanic	1 if respondent identifies self as hispanic
Asian	1 if respondent identifies self as asian
Caucasian	1 if respondent identifies self as caucasian
Age	age of respondent in years
Male	1 if respondent is male
Married	1 if respondent is married, 0 otherwise
Kids	number of kids under 18 years in household
Fin Educ	number of types of financial education (e.g. books, courses, online)
Education	number of years of school
Region 1	1 if respondent lives in the West
Region 2	1 if respondent lives in the Midwest
Region 3	1 if respondent lives in the Northeast
Region 4	1 if respondent lives in the South
Home Owner	1 if respondent owns a home
Income	SCA coding of reported income into brackets
Pension	1 if respondent has a pension but no 401k plan
Has 401k	1 if respondent has a 401k plan
Market Exp.	respondents probability that market returns are positive in next year
Sentiment	consumer sentiment value - calculated from 5 questions
FEP	fraction of 4 probability questions answered with 0, 50, or 100 percent

Guiso, L., P. Sapienza, and L. Zingales, 2005, Trusting the stock market, NBER Working Paper 11648. Haliassos, M., and C. Bertaut, 1995, Why do so few hold stocks?, *Economic Journal*, 105, 1110-1129. Hilgert, M., and J. Hogarth, 2002, Financial knowledge, experience and learning preferences: Preliminary results from a new survey on financial literacy, *Consumer Interest Annual*, 48. 28

Kimball, M., and T. Shumway, 2006, Investor sophistication and the participation, home bias, diversification, and employer stock puzzles, mimeo, University of Michigan. Lusardi, A., and O. Mitchell, 2006, Financial literacy and planning: Implications for retirement wellbeing, Pension Research Council Working Paper 1, The Wharton School.

The Home bias puzzle is the term given to describe the fact that individuals and institutions in most countries hold only modest amounts of foreign equity. This is puzzling since observed returns on national equity portfolios suggest substantial benefits from international diversification. The home bias in equities was first documented by French and Poterba (1991) and Tesar and Werner (1995). Investor sophistication and the home bias, diversification, and employer stock puzzles. Article. Feb 2007. This paper documents that investors are more likely to hold, buy, and sell the stocks of Finnish firms that are located close to the investor, that communicate in the investor's native tongue, and that have chief executives of the same cultural background. The influence of distance, language, and culture is less prominent among the most investment-savvy institutions than among both households and less savvy institutions. many behavioral anomalies (Investor Sophistication and the Home Bias, Diversification and Employer Stock Puzzles) The alternative has been a separate explanation for each anomaly. Measuring Sophistication Previous work of Hilgert, Hogarth and Beverly (2003), Lusardi and Mitchell (2007) Questions on the April 2005 Survey of Consumers Kimball and Shumway (2007) Fox Run Survey, ALP used for development Cognitive Economics Survey Ask 16 sophistication questions Many outcome variables attitudes & actions Measuring Sophistication We count correct answers to sophistication questions to form an index Correct answers. 2007. Home Sweet Home: Home Bias and International Diversification Among Individual Investors. *Journal of Banking & Finance* 31: 317-333. CrossRefGoogle Scholar. Karolyi, G.A., and R.M. Stulz. 2010. Investor sophistication and the home bias, diversification, and employer stock puzzles. Working Paper. University of Michigan, USA. Google Scholar. 2011. Financial Literacy and Stock Market Participation. *Journal of Financial Economics* 101: 449-472. CrossRefGoogle Scholar. van Rooij, M., A. Lusardi, and R. Alessie.