Measuring the Financial Sophistication of Households

By LAURENT E. CALVET, JOHN Y. CAMPBELL, AND PAOLO SODINI*

Many households invest in ways that are hard to reconcile with standard financial theory and that have been labelled as investment mistakes (Campbell, 2006; Calvet, Campbell and Sodini, henceforth "CCS", 2007). There is increasing interest among household finance researchers in the concept of financial sophistication, defined as the ability of a household to avoid making such mistakes. A growing empirical literature documents a cross-sectional correlation between household characteristics and investment mistakes. Richer, better educated households tend to be better diversified (Blume and Friend, 1975; CCS, 2007; Goetzmann and Kumar, 2008; Vissing-Jorgensen, 2003), display less inertia (Agnew, Balduzzi, and Sundén, 2003; Bilias, Georgarakos and Haliassos, 2008; Campbell, 2006; CCS, 2009; Vissing-Jorgensen, 2002), and have a weaker disposition to hold losing and sell winning stocks (CCS, 2009; Dhar and Zhu, 2006) than other households. One feature of these earlier papers is that mistakes are investigated one at a time, often on a non-representative sample of households.

In this paper, we jointly analyze several investment mistakes in a comprehensive, high-quality panel of household finances. Because Swedish residents pay taxes on both income and wealth, Statistics Sweden has a parliamentary mandate to collect highly detailed information on the finances of every household in the country. We compiled the data supplied by Statistics Sweden into a panel of the entire population (about 4.8 million households) covering four years (1999-2002). We observe detailed demographic and income information, and, most notably, the worldwide assets owned by each resident on December 31 of each year, including bank accounts, mutual funds and stocks. The information is provided for each individual account or each security referenced by its International Security Identification Number (ISIN). We refer the reader to CCS (2007, 2009) for a detailed presentation of this dataset.

We use the Swedish panel to simultaneously investigate three types of investment mistakes: under-diversification, inertia in risk taking, and the disposition effect in direct stockholdings. Consistent with earlier research, financial wealth, family size and education are found to have a negative impact on the level of all three mistakes. These findings motivate the construction of an index of financial sophistication, which is obtained by regressing the negative of the mistake vector on a single combination of household characteristics. The index of financial sophistication increases strongly with log financial wealth and household size, and to a lesser extent with education and proxies for financial experience. We briefly discuss how sophistication can be estimated in less detailed datasets. An Appendix available online further presents the dataset and the estimation methodology.

I. Measuring Investment Mistakes

A. Definitions

Following CCS (2007, 2009), we consider three classes of liquid financial assets, excluding illiquid assets from consideration. Cash consists of bank account balances and money market funds. Mutual funds refer to all other funds. Stocks refer to direct holdings only. We measure a household's *financial wealth* as the sum of its holdings in these asset classes. This definition focuses on gross wealth and does not subtract mortgage or other household debt.

We define the following variables for each household h. The *risky portfolio* contains stocks and mutual funds but excludes cash. The *risky share* $w_{h,t}$ at date t

* Calvet: Department of Finance, HEC Paris, 1 avenue de la Libération, 78351 Jouy-en-Josas, France; and NBER, calvet@hec.fr. Campbell: Department of Economics, Littauer Center, Harvard University, Cambridge, MA 02138, USA, and NBER, john_campbell@harvard.edu. Sodini: Department of Finance, Stockholm School of Economics, Sveavägen 65, Box 6501, SE-113 83 Stockholm, Sweden, Paolo.Sodini@hhs.se. We thank Statistics Sweden for providing the data. This material is based upon work supported by the Agence Nationale de la Recherche under a Chaire d'Excellence to Calvet, BFI under a Research Grant to Sodini, the HEC Foundation, the National Science Foundation under Grant No. 0214061 to Campbell, Riksbank, and the Wallander and Hedelius Foundation.

is the weight of the risky portfolio in financial wealth.

B. Investment Mistakes

For every household h, we denote by $y_{h,t} = (y_{h,t,1}; y_{h,t,2}; y_{h,t,3})'$ a vector of investment mistakes at date t. The first component $y_{h,t,1}$ measures underdiversification, the second component $y_{h,t,2}$ risky share inertia, and the third component $y_{h,t,3}$ the disposition effect. The definition of these variables is now explained.

Since Sweden is a small and open economy, we assess the diversification of household portfolios relative to a global equity portfolio, the MSCI World Index. As in CCS (2007), we assume that assets are priced on world markets in an international currency according to a global version of the CAPM. From the perspective of a Swedish investor, the pricing model induces a domestic CAPM in which the currency-hedged world index is mean-variance efficient. Because currency-hedging is typically unavailable to most retail investors, except perhaps the richest, we view the unhedged version of the index as a more attainable benchmark. We therefore measure underdiversification in household h's risky portfolio by the relative Sharpe ratio loss

$$y_{h,t,1} = 1 - \frac{S_{h,t}}{S_m},$$

where $S_{h,t}$ and S_m respectively denote the Sharpe ratio of the risky portfolio and unhedged index under the CAPM.

In CCS (2009) we have developed a structural model of portfolio rebalancing, in which inertia can be measured by the instrument variable regression of risky share changes on household characteristics. We now construct a proxy that can be readily computed from individual household data. A useful starting point is provided by the absolute value of risky share changes, $|\omega_{h,t+1} - \omega_{h,t}|$, which Annette Vissing-Jorgensen (2002) uses as a measure of inertia. We have found in CCS (2009) that boundary effects are typically more pronounced in levels than in logs. For this reason, we proxy inertia by:

$$y_{h,t,2} = |\ln(\omega_{h,t+1}) - \ln(\omega_{h,t})|,$$

that is by the absolute value of risky share changes in logs

As in Odean (1998) and Dhar and Zhu (2006), our analysis of the disposition effect builds on the proportion of stock gains realized during the year, $PGR_{h,t}$, and the proportion of stock losses realized, $PLR_{h,t}$. A gain in a particular stock is counted as being realized if the investor sells some (but not necessarily all) of its holdings of the stock. The household's proportion of gains realized, $PGR_{h,t}$, is then defined as the number of winning stocks with realized gains divided by the total number of winning stocks. $PLR_{h,t}$ is defined analogously. The disposition effect in direct stockholdings is then measured by the difference:

$$y_{h,t,3} = PGR_{h,t} - PLR_{h,t}.$$

We depart in two ways from Odean (1998) and Dhar and Zhu (2006). First, because the purchase price is unavailable in our dataset, we classify a stock as a winner if it has a higher return than the unhedged world index during the year, and as a loser if it underperforms the index.

Second, the earlier literature focuses on the set of households that have experienced both gains and losses in their stock portfolios. We are concerned that this restriction might bias the analysis towards households with large stock portfolios, so we look at a broader set of households that own stocks at the end of a given year t and still hold risky assets at the end of the following year. We extend the definition of $PGR_{h,t}$ and $PLR_{h,t}$ to this broader set of investors. If the household does not experience a gain during the year, we set $PGR_{h,t}$ equal to the cross-sectional mean for households with gains. Similarly if the household does not experience a loss during the year, we set $PLR_{h,t}$ equal to the cross-sectional mean for households with losses.

II. Empirical Results

A. Unrestricted Regressions

In Table 1, we report the results of the pooled regressions of each investment mistake on household characteristics:

$$y_{h,t,j} = \beta_j' x_{h,t} + \varepsilon_{h,t,j}, \quad 1 \le j \le 3,$$

where all left-hand side and right-hand side variables are demeaned. The vector $x_{h,t}$ contains both financial and demographic characteristics. The first category includes disposable income, contributions to private pension plans as a fraction of a three-year average of disposable income, log financial wealth, log

real estate wealth, log of total liabilities, and dummies for households that are retired, unemployed, self-employed ("entrepreneurs"), and students. The second category includes age, household size, and dummies for households that have high-school education, post-high-school education, or missing education data (most common among older and immigrant households) or are immigrants.

Financial wealth has a strikingly negative impact on all three mistakes. Larger households with higher education make smaller mistakes, while entrepreneurs are more prone to all mistakes. Other variables, such as disposable income and real estate wealth, have a less stable effect, but this appears to result from the collinearity of the characteristics $x_{h,t}$. In the Appendix, we compute the simple correlation between these regressors and investment mistakes, and find that income and real estate wealth are negatively correlated with all three mistakes.

Investment mistakes themselves are only weakly correlated across households. The correlation between underdiversification and risky share inertia is 15.5%, the correlation between underdiversification and the disposition effect measure is -10.7%, and the correlation between risky share inertia and the disposition effect measure is 5.1%. When we consider instead the fitted values of the mistakes from Table 1, the correlations are substantially higher, respectively 76.8%, 53.4%, and 80.9%. These findings suggest that a single combination of household characteristics can be used to explain suboptimal investment behavior.

B. Index of Financial Sophistication

We construct an index of financial sophistication by regressing the vector of financial mistakes on a single linear combination of household characteristics:

$$
\begin{aligned}
-y_{h,t,1} &= (\beta' x_{h,t}) + \varepsilon_{h,t,1}, \\
-y_{h,t,2} &= \gamma_2 (\beta' x_{h,t}) + \varepsilon_{h,t,2}, \\
-y_{h,t,3} &= \gamma_3 (\beta' x_{h,t}) + \varepsilon_{h,t,3}.
\end{aligned} \tag{1}
$$

We interpret $(\beta' x_{h,t})$ as an index of financial sophistication. Note that we have multiplied the mistake vector by -1 on the left-hand side, so that households with a higher index tend to make lower mistakes. The index is multiplied by proportionality constants γ_2 and γ_3 in the last two equations. The proportionality constant is normalized to unity in the first equation.

In Table 2, panel A, we report the results of the nonlinear least squares estimation of β in (1). Households with high financial wealth, education and family size achieve a high index of sophistication. In Table 2, panel B, we also report the proportionality coefficients γ_2 and γ_3. They are both positive, which confirms that the index is associated with a lower level of all three mistakes. We observe that the proportionality restriction causes only a slight loss in explanatory power for underdiversification and inertia, but a more serious loss for the disposition effect compared to the unrestricted regressions reported in Table 1.

The correlation between the sophistication index and the risky share is equal to 0.35. This result confirms the finding in CCS (2007) that more sophisticated agents tend to invest more aggressively and make smaller mistakes.

C. Robustness Checks

In the online Appendix, we have verified the robustness of our results to alternative assumptions about the household sample and the measurement of financial mistakes. First, we obtain similar results in a smaller subsample containing stockholders with both gains and losses in their stock portfolios, as in Dhar and Zhu (2006) and Odean (1998).

Second, we have considered several alternative measure of inertia. Risky share changes yield broadly similar, if slightly weaker, results in levels than in logs. General equilibrium considerations imply that changes in the target risky share are potentially important (CCS 2009). We have considered several proxies for the target, and have found that our main results are remarkably robust to these alternative measures.

Third, in the computation of the disposition effect, we have classified winners and losers according to their absolute performance during the year, rather than their performance relative to the world index. Since absolute gains are relatively rare during the severe bear market of our sample period, we confine attention to stockholders with both absolute gains and losses in their stock portfolios, and obtain similar results. Our results are also robust to counting a gain as realized only if the household fully disposes of the corresponding stock during the year.

Finally, the household-level Sharpe ratios used in Tables 1 and 2 are computed on the highly disaggregated asset-level data provided by Statistics Sweden. In other countries, however, researchers often have access to more limited information on household finances, and must typically rely on statistics such as

the number of stocks, the number of funds, and the share of funds in the risky portfolio. In the Appendix, we have investigated how these measures relate to the Sharpe ratio. The share of funds in the risky portfolio appears to be a reasonable diversification proxy, with a 0.49 cross-sectional correlation with the Sharpe ratio. Furthermore, when we use this proxy in the regression of financial mistakes on characteristics, we obtain results that are broadly consistent with the results obtained with the Sharpe ratio.[1] This is encouraging since the share of funds in the risky portfolio is readily available in a variety of datasets.

III. Summary and Conclusions

In this paper, we have confirmed earlier evidence that richer, educated households of larger size are less prone to making financial mistakes than other households. These results have motivated the construction of a single index of financial sophistication that best explains a set of three investment mistakes. The index of financial sophistication increases strongly with financial wealth and household size, and to a lesser extent with education and proxies for financial experience.

We have also reported a strong positive correlation between a household's sophistication index and its share of risky assets. This correlation is consistent with the intuition developed in CCS (2007) that a household is willing to take financial risk when it is confident in its understanding of financial markets and the basic precepts of investing. In a recent and related contribution, Guiso, Sapienza and Zingales (2007) emphasize the role of trust, both in oneself and in others, as a key determinant of participation and risk-taking. The detailed analysis of these closely related views of risk-taking is left open for further research.

[1] Variables such as the number of stocks or the number of funds, however, are poor diversification proxies, as evidenced by their small or even slightly negative correlation with the risky portfolio's Sharpe ratio. We have also considered a more elaborate imputation method based on the household's number of stocks and funds, the share of funds in the risky portfolio, as well as the average return, standard deviation, and correlation of stocks and funds. This method performs well but is only a very modest improvement over the share of funds.

REFERENCES

Agnew, J., P. Balduzzi and A. Sunden, 2003, Portfolio choice and trading in a large 401(k) plan, *American Economic Review* 93(1), 193–215.

Bilias, Yannis, Dimitris Georgarakos and Michael Haliassos, 2008, Portfolio inertia and stock market fluctuations, working paper, Goethe University Frankfurt.

Blume, Marshall, and Irwin Friend, 1975, The asset structure of individual portfolios and some implications for utility functions, *Journal of Finance* 30, 585–603.

Calvet, Laurent E., John Y. Campbell and Paolo Sodini, 2007, Down or out: Assessing the welfare costs of household investment mistakes, *Journal of Political Economy* 115, 707-747.

Calvet, Laurent E., John Y. Campbell and Paolo Sodini, 2009, Fight or flight? Portfolio rebalancing by individual investors, forthcoming *Quarterly Journal of Economics*.

Campbell, John Y., 2006, Household finance, *Journal of Finance* 61, 1553–1604.

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

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

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2007, Trusting the stock market, forthcoming *Journal of Finance*.

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

Vissing-Jorgensen, Annette, 2002, Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures, NBER Working Paper 8884.

Vissing-Jorgensen, Annette, 2003, Perspectives on behavioral finance: does "irrationality" disappear with wealth? Evidence from expectations and actions, in Mark Gertler and Kenneth Rogoff eds. *NBER Macroeconomics Annual 2003* (MIT Press, Cambridge, MA).

TABLE 1. REGRESSION OF INVESTMENT MISTAKES ON HOUSEHOLD CHARACTERISTICS

	Underdiversification		Risky Share Inertia		Disposition Effect	
	Estimate	**t-stat**	**Estimate**	**t-stat**	**Estimate**	**t-stat**
Financial Characteristics						
Disposable income	0.841	14.50	2.329	13.00	-0.626	-4.27
Private pension premia/income	-0.541	-9.45	-0.387	-2.18	0.076	0.52
Log financial wealth	-3.814	-59.70	-11.510	-58.10	-7.179	-44.40
Log real estate wealth	-0.696	-11.00	1.597	8.14	0.632	3.94
Log total liability	-0.156	-2.17	-1.205	-5.42	-1.196	-6.59
Retirement dummy	-0.401	-1.86	-1.710	-2.56	1.065	1.95
Unemployment dummy	0.768	3.35	-0.390	-0.55	2.340	4.04
Entrepreneur dummy	1.297	5.12	10.835	13.80	6.481	10.10
Student dummy	1.067	2.32	-4.288	-3.01	-1.919	-1.65
Demographic Characteristics						
Age	0.037	5.95	-0.070	-3.61	0.016	1.04
Household size	-1.420	-28.10	-0.991	-6.32	2.022	15.80
High school dummy	-0.654	-4.02	-1.166	-2.31	-2.705	-6.58
Post-high school dummy	0.246	1.85	-0.089	-0.22	-3.834	-11.40
Dummy for unavailable education data	2.930	11.70	0.113	0.15	-3.969	-6.28
Immigration dummy	3.447	19.00	4.289	7.62	-5.216	-11.40
Adjusted R^2	6.96%		4.27%		3.13%	
Number of observations	102,731		102,731		102,731	

Notes: This table reports the pooled regressions of investment mistakes on household characteristics. Underdiversification is measured by the Sharpe ratio loss relative to the unhedged world index under the CAPM. Risky share inertia is proxied by changes in the log risky share. The disposition effect measure is the difference between the proportion of gains realized and the proportion of losses realized during the year. An asset is classified as a gain if it outperforms the unhedged world index during the year, and as a loss otherwise. The estimation is based on participants at t and t+1 with direct stockholdings at t for which the immigration dummy is available. All characteristics are demeaned year by year, and continuous characteristics are standardized year by year.

TABLE 2. RESTRICTED REGRESSION OF INVESTMENT MISTAKES ON HOUSEHOLD CHARACTERISTICS

A. Sophistication Index

	Estimate	t-stat	Correlation
Financial Characteristics			
Disposable income	-0.673	-15.80	0.137
Private pension premia/income	0.322	7.70	0.184
Log financial wealth	4.335	72.40	0.923
Log real estate wealth	0.073	1.58	0.304
Log total liability	0.379	7.24	-0.009
Retirement dummy	0.313	1.99	0.010
Unemployment dummy	-0.614	-3.67	-0.114
Entrepreneur dummy	-2.865	-15.40	-0.095
Student dummy	0.243	0.72	-0.062
Demographic Characteristics			
Age	-0.012	-2.58	0.071
Household size	0.632	17.00	0.277
High school dummy	0.805	6.78	0.164
Post-high school dummy	0.327	3.36	0.212
Dummy for unavailable education data	-1.070	-5.86	-0.070
Immigration dummy	-1.751	-13.10	-0.136
Number of observations	102,731		

B. Proportionality Coefficients and Adjusted R^2

	Proportionality Coefficient		Adjusted R^2
Underdiversification	-	-	6.02%
Risky share inertia	2.414	49.80	3.76%
Disposition effect	1.397	39.20	1.91%

Notes: This table reports the pooled restricted regressions of investment mistakes on household characteristics. In Panel A, we compute the coefficients of the sophistication index, their *t*-statistics, as well as the correlation of the index with each characteristic. In Panel B, we report the proportionality coefficient of risky share inertia and the disposition effect measure, and the adjusted R^2 of all three mistakes. The proportionality coefficient of underdiversification is by definition equal to unity and is not reported. The measure of each mistake is computed as in Table 1. The estimation is based on participants at *t* and *t*+1 with direct stockholdings at *t* for which the immigration dummy is available. All characteristics are demeaned year by year, and continuous characteristics are standardized year by year.