Three Essays on Stock Markets

Three Essays on Stock Markets

By
PARASTOO OSTAD,

A Thesis Submitted to the School of Graduate Studies
in the Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
in
Finance and Business Economics

McMaster University
Hamilton, Ontario

Doctor of Philosophy (2024)
DeGroote School of Business
McMaster University
Hamilton, Ontario, Canada

TITLE: Three Essays on Stock Markets

AUTHOR:
Parastoo Ostad,

SUPERVISOR:
Dr. Ronald Balvers
Professor, Finance,
McMaster University, ON, Canada

SUPERVISORY COMMITTEE CHAIR:
Dr. Peter Miu,
Professor, Finance,
McMaster University, ON, Canada

SUPERVISORY COMMITTEE MEMBERS:
Dr. Richard Deaves
Professor, Finance,
McMaster University, ON, Canada

Dr. Narat Charupat
Professor, Finance,
McMaster University, ON, Canada

EXTERNAL EXAMINER:
Dr. Philippe d'Astous,
Associate Professor, Finance,
HEC Montreal, Quebec, Canada

NUMBER OF PAGES: vii, 138

Abstract

This thesis studies three important topics regarding the Stock Markets.
The first chapter of my Ph.D. thesis, titled "*The Value Relevance of Corporate Tax Expenses in the Presence of Partisanship: International Evidence*", contributes to the existing literature on the information contents of corporate taxes by investigating whether the political orientation of tax decision-makers affects the informativeness of corporate tax expenses. We confirm that firms bear higher tax expenses under left-leaning governments supporting the partisan theory of political cycles even in the wake of globalization. Furthermore, we introduce a simple model suggesting that the information contents of corporate tax expenses are conditional on the political orientation of the government, driven by investors' perspectives on a firm's future cash flows and cash-flow volatility. By analyzing cross-sectional country-level partisanship differences, we find that corporate tax expenses are only informative about future returns under right-leaning governments. This suggests that corporate tax expenses are arguably a more direct profitability indicator when right-leaning governments are in power.

The second chapter, titled "*The Direct and Indirect Impact of Non-cognitive Skills on Stock Market Participation*" focuses on the Household Finance area. Using data from the Understanding America Study (UAS) survey panel, we explore the roles of cognitive and non-cognitive skills in the empirically observed stock market participation (SMP) puzzle. Specifically, we examine both the direct and indirect effects of the "primitive" factors of intelligence and non-cognitive skills (derived from conscientiousness and emotional stability scores) on stock holdings. This examination extends to their influence through well-observed proximate factors of SMP such as general and task-specific financial literacy, education, income, and trust. We find that non-cognitive skills have both direct and indirect, via the proximate factors, positive impacts on stock market participation. In contrast, intelligence solely affects participation indirectly through proximate factors. Overall, higher levels of intelligence and non-cognitive skills significantly enhance the likelihood of owning stock equities.

The third chapter, titled "*Market Literacy and Stock Market Participation*", delves deeper into the determinants of SMP and investigates the role of specific stock market knowledge in household stock-holding decisions using the UAS, survey panel. I examine whether a demonstration of knowledge about a reasonable distribution for market return increases equity participation. I demonstrate that possessing market return expectations that are in line with historical performance and the current environment, labeled as

market literacy in return, enhances the likelihood of stock holding and this effect persists even after accounting for established factors such as general and task-specific financial knowledge, intelligence, and education. Moreover, overoptimistic households, whose market return expectation is greater than the market literacy range, are less likely to participate since the higher expected outcome is dominated by the perceived literacy cost. I also explore the impact of market literacy in risk and overprecision on SMP. Market literacy in risk refers to providing reasonable volatility estimates for stock returns, while overprecision signifies overestimating market outcomes volatility. The findings suggest that the roles of market literacy in risk and overprecision on equity holding decisions are indirect and fully subsumed by other drivers of SMP.

Acknowledgements

I would like to express my sincerest gratitude to my supervisor, Dr. Ronald Balvers, whose unwavering guidance, support, and mentorship have shaped my academic journey. He has not only been an exceptional supervisor but also a role model whose dedication to research and teaching has inspired me all these years. I am immensely grateful to him for being patient with me and pointing me in the right direction every time I was desperately stuck.

My genuine appreciation to my supervisory committee members, Dr. Richard Deaves, and Dr. Narat Charupat. To Dr. Deaves who introduced the fascinating world of Behavioral and Household Finance to me and trained me along the way. His support and guidance continued even after his retirement. To Dr. Charupat, who believed in me to be a part of this program and supported me throughout my Ph.D. studies. I would not be here if it wasn't for him.

I would also like to thank my external reviewer, Dr. Philippe d'Astous, for his constructive comments and valuable suggestions.

I am sincerely grateful to my co-author and friend, Dr. Adam Stivers for his invaluable help and support, particularly during the job market process. I extend my great appreciation to Anita Blaney and Bani Rafeh for their excellent administrative support.

My heartful gratitude goes to my Mom and Dad, for their unlimited support and unconditional love, and to my brothers, Pouria and Peyman, for always being there for me, even though we have been miles away from each other all these years.

I am extremely thankful to my dearest friends, Lulu and Mike, who have been my family in Hamilton and have stuck with me through thick and thin.

Last but not least, my deepest love to my toy poodle Coconut, who kept me sane and happy, and gave me a reason to keep going every day.

Contents

Declaration of Academic Achievement

I, Parastoo Ostad, declare that this thesis titled, **Three Essays on Stock Markets**, and works presented in it are my own. I confirm that the thesis comprises the following chapters:

- The Value Relevance of Corporate Tax Expenses in the Presence of Partisanship: International Evidence
- The Direct and Indirect Impact of Non-cognitive Skills on Stock Market Participation
- Market Literacy and Stock Market Participation

This thesis is entirely my own original work unless otherwise indicated. Any use of the work of other authors is acknowledged at their point of use

Introduction

The dynamic realm of the capital market where myriad forces converge is shaped by the intricate interplay of economic variables, political dynamics, and individual behavior. This dissertation is a venture into unraveling the complexities that underscore stock market behavior, examining two distinct dimensions through empirical analyses and comprehensive investigations.

The first dimension explores the realm of asset pricing, specifically, the first chapter of this dissertation delves into equity analysis under specific political circumstances. This study investigates the informativeness of corporate tax expenses in the presence of political cycles in an international setting and finds that corporate tax expenses are value-relevant for investors during the administration of right-wing governments.

Previous literature finds that greater tax expenses are good news and can positively explain contemporaneous and future returns, since it signals the core profitability of a firm and is not affected by accounting procedures or managerial manipulation. Nevertheless, there are circumstances that can influence corporate taxation, potentially either enhancing or diminishing its informativeness to investors. For instance, Kerr (2019) documents positive and critical effect of tax enforcement in the informativeness of corporate tax expense.

Corporate taxation is a function of the national tax system administrated by the government. The "partisan theory" suggests that one of the primary sources of differences among governments from different parties comes from taxation and the ideal tax structures. This is supported by previous studies documenting that left-leaning governments tend to impose higher corporate tax expenses (Osterloh and Debus, 2012) through policies such as increasing the statutory corporate tax rate and expanding the tax base, Therefore, the political preferences of the government should play a role in determining the information contents of corporate tax expenses.

For this purpose, we investigate these relationships in a sample of 10 developed countries with a parliamentary system. We find that corporate tax expenses are only informative about future returns during right-leaning governments. This result implies that corporate tax expenses are arguably a more direct profitability indicator and are only

informative about future returns when right-leaning governments are in charge. These results can be substantiated by both potential mechanisms; cash flow and risk-based effects.
This article has been published in the Global Finance Journal. It represents a collaborative work with Javier Mella. However, my contribution to the research is substantial, accounting for approximately 90% of the content.

The second dimension of this dissertation navigates the multifaceted nature of individual decision-making in participating in the stock market. This is the subject of chapter two and chapter three which contribute to the Stock Market Participation (SMP) puzzle literature. Chapter two explores to what extent intelligence and non-cognitive skills, which we label as primitive factors, contribute to individuals' decisions regarding stock market participation. We investigate whether these primitive factors affect individuals' equity-holding decisions. Non-cognitive skills have been largely unexplored in the literature, as opposed to general personality. Therefore, we initiate our study by identifying the most relevant elements of Big Five personality traits for decision-making in equity holding, and eventually, we define Non-cognitive skills as an equal-weighted combination of survey-based Conscientiousness and Emotional Stability scores. Then we explore both the direct impact of non-cognitive skills on SMP and the indirect impact through traditional explanatory factors such as education, income, financial literacy, task-specific financial knowledge, and trust, which we label as proximate factors.

To carry out our study, we utilize survey data from the Understanding America Study panel, which provides a large, generalized sample from the US population with high response rates. We use information on whether respondents hold stocks of any kind, whether directly or indirectly. Thus, our SMP measure is binary. In isolation, we find that the primitive factors affect SMP; higher levels of both intelligence and non-cognitive skills lead to significantly higher SMP. Still, we find that most of the effects are indirect in that higher levels of the primitives lead to higher levels of the proximates, which then lead to higher levels of SMP. This is entirely true for intelligence, which we find to have no residual direct impact on SMP after controlling for the proximates. We do find, however, that our measure of non-cognitive skills retains some direct impact on SMP, even after controlling for the proximates.
This paper has received a "Revise and Resubmit" at Journal of Banking and Finance. It is a collaborative effort with Dr. Richard Deaves, my Ph.D. supervisory committee member, and Dr. Adam Stivers. We all equally contributed to the creation of this paper.

The final chapter digs deeper into the drivers of SMP. The objective of this study

is to investigate whether specific market knowledge, which I label as market literacy in return and market literacy in risk and are distinct from general financial knowledge, is incrementally predictive of SMP.

Moreover, I explore whether behavioral factors, in particular overoptimism and overprecision, as two main forms of overconfidence bias, contribute to stockholding decisions. The conventional equity exposure decision suggests that higher expected return and lower perceived risk should increase investment in stocks. Nevertheless, whether such linear relationships hold when it comes to stock market participation decisions is an empirical question. I show that the probability of SMP does not simply rise with the expected market return; instead, a demonstration of greater awareness about future returns is associated with higher SMP. This can be explained by the perceived costs of obtaining information for equity investment which deter households from in the territory of overoptimism. In terms of perceived risk, although I conjecture that the ability to provide reasonable volatility for stock returns increases the probability of SMP, the impact of market literacy in risk and overprecision factor, are indirect and fully subsumed by other drivers of SMP.

As a precursor to the empirical work, I first develop a simple representative-agent model to illustrate the role of market literacy in SMP decision-making, by incorporating the agent's belief about their knowledge as well as the literacy cost. I discuss that, in the realm of overoptimism, while higher expected returns may encourage participation, this behavioral inclination is outweighed by the rational consideration of the costs associated with obtaining market knowledge, ultimately leading to lower participation rates.

Second, I utilize survey data from the Understanding America Study panel, which provides a large, representative sample from the US population with high response rates and shows that market literacy in return matters for market entry and its effect is incremental to the previously documented financial literacy and absence of stock return ignorance determinants. I also compare the impact of the three forms of financial/market literacy, finding that market literacy in return is an important determinant of SMP, and its impact is roughly on par with the other two literacy factors. To address the reverse casualty between SMP and market literacy, I employ two instrumental variables, namely US states and parents' education, and show that these are valid instruments and that the causal impact of market literacy on SMP continues to hold.

Chapter 1

The value relevance of corporate tax expenses in the presence of partisanship: International evidence.

Contents

Global Finance Journal 57 (2023) 100832

Contents lists available at ScienceDirect

Global Finance Journal

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

The value relevance of corporate tax expenses in the presence of partisanship: International evidence☆

Parastoo Ostad [a], Javier Mella [b,*]

[a] DeGroote School of Business, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4M4, Canada
[b] Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Chile. Monseñor Álvaro del Portillo 12455, Santiago, Chile

ARTICLE INFO

JEL classification:
G12
G15
G18
H25
F65

Keywords:
Asset pricing
Political cycles
Corporate tax expenses
Value relevance
Stock market

ABSTRACT

This study examines the value relevance of corporate tax expenses in an international setting and determines whether partisanship plays a role in its informativeness. Our empirical results indicate the greater value relevance of corporate taxes during the administration of right-leaning governments. Moreover, our cross-sectional analysis suggests that corporate tax expense is value-relevant during the administration of right-leaning governments; however, corporate tax expense does not convey information regarding returns when left-leaning parties are in office. These findings highlight the importance of the political orientation of tax policymakers in determining value-related information on corporate tax expenses.

1. Introduction

This paper investigates the informativeness of corporate tax expense in the presence of political cycles. In general, empirical results in an international setting indicate that corporate tax expenses are value-relevant for investors under right-wing governments and that two potential mechanisms can help describe their effects on stock returns: cash flow and risk-based effects. Previous literature also provides evidence on the information contents of corporate tax expenses, finding that greater tax expenses (or tax surprises) are good news, which can positively explain contemporaneous and future returns (e.g., Hanlon, Kelley Laplante, & Shevlin, 2005; Thomas & Zhang, 2011). Scholars further argue that the explanatory and predictive power of tax expenses is incremental to that of generally accepted accounting principles (GAAP) proxies for profitability, since it signals the core profitability of a firm. Related research suggests that both current and deferred portions of tax expenses are value-relevant. Regarding the current portion, Hanlon et al. (2005) measure taxable income from this portion and study it as an alternative proxy for profitability based on a tax system, rather than an accounting system. The advantages of such a measure of profit are that, first, it is unlikely to be subject to managerial manipulation, and second, it provides a consistent measurement across firms. In terms of the deferred portion of tax expenses, Lev and Thiagarajan (1993) find that it signals earnings persistence.

☆ We are grateful to Ronald Balvers, Narat Charupat, Richard Deaves, the Editor-in-chief Ali M. Fatemi, the Senior Editor Arman Eshraghi, anonymous reviewers and workshop participants at McMaster University, Universidad de los Andes (Chile), and the Southwestern Finance Association 2021 Annual Meeting for their helpful comments and suggestions.

* Corresponding author.
E-mail addresses: ostads@mcmaster.ca (P. Ostad), jamellab@uandes.cl (J. Mella).

https://doi.org/10.1016/j.gfj.2023.100832
Received 14 July 2022; Received in revised form 29 April 2023; Accepted 30 April 2023
Available online 8 May 2023

Previous studies also examine the circumstances that can either strengthen or weaken the value relevance of corporate taxes. The most relevant research to the present study include Kerr (2019), which examines the value relevance of tax expenses within its interaction with tax enforcement in an international setting, and Mella (2021), which reports the effect of corporate tax rates in the United States (U.S.) that are conditional on the party in office.

Furthermore, extensive literature shows that economic policies differ between left- and right-leaning parties. The main results can be found in Alesina, Roubini, and Cohen (1997), Drazen (2000), and Blinder and Watson (2016). The "partisan theory" suggests that governments from different parties implement distinct policies according to their platforms (Hibbs Jr., 1992). However, one of the primary sources of differences comes from taxation (since it is under the government's influence) and the ideal tax structure, which can considerably differ across parties. Since corporate tax expenses are a function of the national tax system, the political preferences of the government should play a role in determining the information contents of corporate tax expenses. Thus, we examine the informativeness of corporate tax expenses and its interaction with the political orientation of tax policymakers on future stock returns.

For this purpose, we investigate these relationships in a sample of 10 developed countries with a parliamentary system. Specifically, the sample consists of G12 countries, excluding the U.S. (which has been previous studied and includes a presidential system), Japan (which does not have the necessary variation in the political variable), and Switzerland (for which the political variable is not available in the database used in this study). Overall, we extend the literature on the value relevance of corporate tax expenses by showing that the political orientation of tax policymakers is relevant for determining the information contents of such expenses in developed financial markets outside of the U.S.

2. Literature review and hypothesis development

2.1. Partisanship and corporate tax expenses

In a political context, one of the primary sources of the ideological differences among candidates for government administration comes from taxation, since their focus area and ideal tax structure are dissimilar. To clarify, the parliamentary system of government is dominant in European national politics, and the most relevant ideological framework is the left-right dichotomy. In this case, left-leaning politicians are inclined to target voters whose socioeconomic status tends to be lower than the average. Thus, they have a stronger tendency to favor policies that redistribute wealth in society and fund public expenditure (Benoit & Laver, 2006), whereas the reverse is true for right-leaning politicians. Corporate taxes are one of the financial resources that left-leaning incumbents rely on to fulfill their obligations of public expenditure. In the U.S., Inclan, Quinn, and Shapiro (2001) show that this increases corporate taxes during Democratic terms but decreases them during Republican periods. Osterloh and Debus (2012) investigate the effect of partisanship on the statutory corporate tax rate and effective tax rate at the country level, employing an unbalanced panel of 32 European countries during 1980–2006; their findings show that left-wing governments tend to increase statutory corporate taxes, while right-wing parties favor lower taxes.

2.2. Discussion regarding globalization

A counterargument that may be valid in this context states that during the globalization of the international economy over the past 30 years, statutory corporate tax rates have been gradually slashed to attract multinational businesses and prevent capital outflows. In addition, corporate income taxes provide an essential contribution to the national tax system in terms of the revenue they generate and their impact on the financial decisions of the affected entities. More specifically, globalization has increased competition in the global economy and influenced the national level of corporate taxation (Osterloh & Debus, 2012). Moreover, increases in firms' mobility have made capital more sensitive to the international differences in corporate taxation. This implies that tax incumbents, regardless of their political preferences, may feel compelled to lower the statutory corporate tax rate to avoid the expected costs from excessive outflows of mobile capital. Consequently, statutory corporate income tax rates have shown a downward trend in Organization for Economic Co-operation and Development (OECD) countries since the 1990s. The Organization for Economic Co-operation and Development (OECD) (2020) reports a significant decline of 9% in average statutory corporate tax rates across OECD member countries, from 32.2% in 2000 to 23.2% in 2020. In this context, politicians with preferences for higher corporate tax rates may be unable to implement their desired tax policies, making partisanship irrelevant. We address this issue by raising the following discussion and conducting a validation test.

The tax burden imposed by a tax system on corporations is not fully captured by statutory tax rates, in the sense that there are other factors that can determine the ultimate taxable income (Government Accountability Office (GAO), 2008, p. 1). Furthermore, corporate tax expenses are not only influenced by statutory tax rates but also the tax base, which is broadly defined as the total amount of corporate income that is subject to taxation by the tax authority. In the process of a substantial but gradual decline in statutory corporate tax rates (due to globalization), the OECD recommended that its member states increase their tax revenue by broadening their tax base (Organization for Economic Co-operation and Development (OECD), 2010) to avoid the risk of capital outflows in the international economy. Examples of mechanisms that governments can use to expand their tax base are the elimination (or decrease) of accelerated depreciation, the elimination of tax credits, the restriction of carrying losses backward and forward, and the elimination of credits for previously paid foreign taxes.[1]

[1] For an exhaustive list, see Kawano and Slemrod (2016).

Bernardi (2011) finds that tax base broadening paired with tax rate cuts has been an apparent goal of European countries since the mid-1980s. In related research, Kawano and Slemrod (2016) study the interaction between corporate tax rates and corporate tax base changes in estimating corporate tax revenue. Based on a sample of 30 OECD countries, they find that changes in the corporate tax base are likely to occur following changes in statutory tax rates. In particular, they conclude that governments tend to adopt policies that broaden the tax base and maintain the desired level of tax revenue when the statutory tax rate drops due to international competitive pressure. Hence, the downward trend in the statutory tax rate has not been accompanied by fluctuations in tax revenue to the same degree. Figs. A-1 and A-2 present the average statutory corporate tax rates and the average corporate tax revenue for OECD countries since 2000.

The international trend toward lowering statutory tax rates has impeded legislatures from implementing their desired tax policies. Statutory tax rates are widely visible and employed to attract new investments from multinational companies. Conversely, changes in the tax base are not widely visible and could be a more convenient tool for tax incumbents to maintain (or increase) their tax revenue. The intuition is that left-leaning governments may have a greater propensity to broaden the tax base. Thus, we assume that (on average) companies incur greater tax expenses, either from a greater tax base or tax rate when left-leaning politicians are in office, relative to the periods of right-leaning leadership. This leads to the following validation exercise.

2.3. Validation

To investigate the validity of our assumption, we determine whether the tax expenses incurred by companies are higher under left-leaning governments. In this case, we use panel data approaches and control for firm fixed effects[2] and time fixed effects, along with other controls for firm characteristics that are known to affect a firm's tax expenses.

$$\begin{aligned}\mathbf{TaxExp}_{it} =& \lambda_0 + \lambda_1 \mathbf{LeftGov}_{t-1} + \lambda_2 \mathbf{BVA}_{it-1} + \lambda_3 \mathbf{LEV}_{it} + \lambda_4 \mathbf{PPE}_{it} + \lambda_5 \mathbf{ROA}_{it} + \lambda_6 \mathbf{Intangible}_{it} \\ &+ \lambda_7 \mathbf{MTB}_{it} + \lambda_8 \mathbf{Inventory}_{it} + \lambda_9 \mathbf{R\&D}_{it} + \lambda_{10} \mathbf{PCM}_{it} + \lambda_{11} \mathbf{FCF}_{it} + \lambda_{12} \mathbf{LossFirm}_{it} \\ &+ \lambda_{13} \mathbf{HHI}_{it} + \lambda_{14} \mathbf{GDPgrowth}_{t} + \lambda_{15} \mathbf{StatutoryTaxRate}_{t} + \eta_i + \tau + \varepsilon_{it}\end{aligned} \tag{1}$$

In Eq. (1), **TaxExp** is the variable for corporate tax expenses, while **LeftGov** is the left-leaning dummy variable representing the political orientation of the head of the government. When changes in a tax policy occur, the new tax law comes into effect on the first of January of the following year (Osterloh & Debus, 2012). Thus, the corporate tax expense at time *t* is ascribable to the decision of the tax incumbent regarding corporate taxes at *t*-1. We also include the coefficient λ_1 to examine the effect of partisanship on corporate tax expenses. A positive λ_1 is expected to support our validation test that corporate tax expenses are higher under left-leaning governments. Consistent with previous studies, we include a wide array of control variables that influence firm level tax expenses. These variables are described in the Appendix, while the definitions of all the variables are found in Table A-3.

Table 1 presents the coefficients and t-statistics estimated from a panel with fixed effect regression. In this case, the coefficient for **LeftGov** is positive and significant, suggesting that (on average) firms incur greater corporate tax expenses when the head of the government belongs to a left-leaning party. The results confirm the existence of a partisan effect on corporate taxation, despite globalization costs. This finding also motivates us to determine whether the partisan effect has a bearing on the information contents of corporate tax expenses.

3. Information contents of corporate tax expenses conditional on partisanship

3.1. A simple model

In this model, we introduce corporate taxes as a function of the political cycle, i.e., the level of taxes differs between left- and right-leaning governments. However, we keep the model simple, since it is sufficient for obtaining the hypothesis and mechanisms for our empirical analysis. Specifically, the earnings before taxes of firm *i* (EBT_i) are described as the sum of the true value (EBT_i^*) and the noise (εi). The noise term represents earnings management, which is of a higher variance when a firm is engaged in the management of book accruals to manipulate earnings. Accordingly, tax expenses are the products of the effective tax rate (ETR) and earnings before taxes, while the effective tax rate consists of a constant (k*i*) and a political multiplier (PM_s). These two states affect the PMs when a left-leaning government is in office (LL) and a right-leaning government is in office (RR), where $0 \leq PM_{RR} < PM_{LL} \leq 1$, indicating that the PMs is higher during left-leaning tax incumbents. Consequently, the variance of tax expenses ($\sigma^2_{\eta is}$) is a function of the PMs and the noisiness of the EBT_i. Eqs. 2 to 7 show these concepts:

$$EBT_i = EBT_i^* + \varepsilon i \qquad \varepsilon i \sim N\left(0, \sigma^2_{\varepsilon i}\right) \tag{2}$$

$$Tax_{is} = ETR_{is} \times EBT_{is} \tag{3}$$

$$ETR_{is} = ki \times PM_s \tag{4}$$

[2] The fixed-effects model specification is supported by the Hausman test.

Table 1
Corporate tax expenses in the presence of partisanship.

	CTE
LeftGov	**0.000667****
	(2.207)
BVA	−0.00571***
	(−13.56)
LEV	−0.00619***
	(−4.081)
PPE	0.00802***
	(4.100)
Inventory	0.0262***
	(7.143)
ROA	0.0270***
	(5.384)
Intangible	0.000528
	(0.304)
MTB	0.000874***
	(4.254)
R & D	−0.000269
	(−0.0548)
PCM	−0.000319***
	(−2.908)
FCF	0.0134***
	(5.294)
LossFirm	−0.0141***
	(−9.492)
HHI	−0.00581***
	(−3.762)
GDPgrowth	0.000375***
	(2.658)
StatutoryTaxRate	0.0136
	(1.651)
Fixed Effects	Firm & Time
R-squared	0.168
Number of Firms	12,049
Number of observations	126,058
Time periods	31

This table presents the fixed effect regression results of corporate tax expenses (CTEs) and the political orientation of the government. All the variables are defined in Table A-3 of the Appendix. Robust standard errors are clustered at the industry level. Significance levels are based on two-sided *t*-tests, provided in parentheses and indicated as follows: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

$$\mathrm{Tax_{is}} = \mathrm{EBT}_i^* \times \mathrm{k}i \times \mathrm{PM_s} + \eta_{\mathrm{is}} \tag{5}$$

$$\eta_{\mathrm{is}} = \varepsilon_i \times \mathrm{k}i \times \mathrm{PM_s} \qquad \eta_{\mathrm{is}} \sim \mathrm{N}\left(0, \sigma^2_{\eta is}\right) \tag{6}$$

$$\sigma^2_{\eta is} = (\mathrm{ki} \times \mathrm{PMs})^2 \sigma^2_{\varepsilon i} \tag{7}$$

In the absence of partisanship, in which PM is constant, consider two firms (*a* and *b*) that report identical earnings before taxes $\mathrm{EBT_a} = \mathrm{EBT_b}$. However, firm *a* discloses higher tax expenses: $\mathrm{Tax_a} > \mathrm{Tax_b}$. This implies that firm *a* carries higher true earnings: $\mathrm{EBT}_a^* > \mathrm{EBT}_b^*$. This signaling effect also suggests that the higher tax expenses of firm *a* signal its greater true earnings, compared to those of firm *b*." Even though the earnings after taxes of firm *a* are lower than those of firm *b*, they are considered as better quality, due to smaller noise. Hence, the market's reward for high-quality earnings outweighs the compensation for lower magnitude, leading to higher returns for firm *a*. This effect becomes stronger for firms situated in the top bracket of the tax system, i.e., firms with higher *k*.

However, in the presence of partisanship, PM is defined by the political orientation of the government, $0 \leq PM_{\mathrm{RR}} < PM_{\mathrm{LL}} \leq 1$. From Eq. (4), firms with greater *k* are more affected by changes in PM. Assume that firm *i* holds an identical true value of EBT during either left-leaning or right-leaning governments ($\mathrm{EBT}_{iLL}^* = \mathrm{EBT}_{iRR}^*$). Since $PM_{\mathrm{RR}} < PM_{\mathrm{LL}}$, firm *i* will end up paying higher taxes under a left-leaning government, even though its true earnings remain unchanged. In fact, the higher tax expenses of firm *i* under such a government in part signals the true value of earnings (the signaling effect) and in part signals greater cash outflow, due to the heavier tax setting (the partisanship effect). We believe that the signaling effect and partisanship effect move in the opposite direction when a left-leaning government is in charge, indicating that the signaling effect of tax expenses is weakened through a higher PM, thus leading to lower returns for firms with greater *k*. The opposite is true when the government is right-leaning. To clarify, the signaling effect and

partisanship effect go in the same direction, due to lower PM, and firms with higher *k* are more affected by reduced PM. Therefore, the higher tax expenses under a right-leaning government are representative of the core profitability of a firm, resulting in higher returns for such firms.

The relationship between a firm's stock returns and corporate tax expenses conditional on the political orientation of the government can be explained by investors' perspectives toward future cash flows, in general, and investors' underreaction (uncertainty) regarding a firm's future cash flows, in particular. While the risk-based explanation is consistent with rational investors, such underreaction requires additional theoretical support, as seen in Daniel, Hirshleifer, and Subrahmanyam (1998) and Hong and Stein (1999).

3.1.1. Cash flow-based explanation

When the government is known for its desire to collect higher tax revenue as a result of its political preferences, increases in corporate tax expenses are in part concerned with a cash outflow that is not informative about the future profitability of a firm. Consequently, if the true value of a firm's earnings before taxes (EBT_i^*) is constant during left- or right-leaning periods, then the cash outflow of the firm is higher under a left-leaning government, thus causing lower returns. Eqs. 8 to 10 demonstrate these relations:

$$\mathrm{Tax}_{i\mathrm{LL}-}\ \mathrm{Tax}_{i\mathrm{RR}} = \left(\mathrm{EBT}_i^* \times \mathrm{k}_i\right)\left[\mathrm{PM}_{\mathrm{LL}} - \mathrm{PM}_{\mathrm{RR}}\right] + \left[\eta_{i\mathrm{LL}} - \eta_{i\mathrm{RR}}\right] \tag{8}$$

$$\mathrm{E}[\mathrm{Tax}_{i\mathrm{LL}-}\ \mathrm{Tax}_{i\mathrm{RR}}] = \mathrm{E}\left[\left(\mathrm{EBT}_i^* \times \mathrm{k}_i\right)\left(\mathrm{PM}_{\mathrm{LL}} - \mathrm{PM}_{\mathrm{RR}}\right)\right] + \mathrm{E}\left[(\varepsilon i \times \mathrm{k}_i)(\mathrm{PM}_{\mathrm{LL}} - \mathrm{PM}_{\mathrm{RR}})\right] \tag{9}$$

Since $\mathrm{E}(\varepsilon_i) = 0$;

$$\mathrm{E}[\mathrm{Tax}_{i\mathrm{LL}-}\ \mathrm{Tax}_{i\mathrm{RR}}] = \left(\mathrm{EBT}_i^* \times \mathrm{ki}\right) \times \left(\mathrm{PM}_{\mathrm{LL}} - \mathrm{PM}_{\mathrm{RR}}\right) \geq 0 \tag{10}$$

3.1.2. Risk-based explanation

The aforementioned theory can also be explained through a risk-based channel. The variable earnings after taxes of firm *i* (EAT_i) represents the difference between the earnings before taxes and tax expenses (Eq. (11)). From Eq. (14), the variance of EAT is a function of the PMs and the variance of EBT. Since $PM_{RR} < PM_{LL}$, the earnings after taxes of firm *i* have lower variance under a left-leaning government. Moreover, Var ($\mathrm{EAT}_{\mathrm{iLL}}$) < Var ($\mathrm{EAT}_{i\mathrm{RR}}$) supports lower expected returns under such a government.

$$\mathrm{EAT}_{is} = E\mathrm{BT}_i - \mathrm{Tax}_{is} \tag{11}$$

$$\begin{aligned} EAT_{is} &= \mathrm{EBT}_i^* + \varepsilon\mathrm{i} - \mathrm{EBT}_i^* \times \mathrm{ki} \times \mathrm{PM_s} - \eta_{\mathrm{is}} \\ EAT_{is} &= (1 - \mathrm{ki} \times \mathrm{PMs}) \times \mathrm{EBT}_i^* + \varepsilon\mathrm{i} - \eta_{\mathrm{is}} \end{aligned} \tag{12}$$

$$\mathrm{Var}\ (EAT_{is}) = \mathrm{Var}\ (\varepsilon\mathrm{i} - \eta_{\mathrm{is}}) \tag{13}$$

$$\begin{aligned} \mathrm{Var}\ (EAT_{is}) &= \mathrm{Var}\ (\varepsilon\mathrm{i} - \varepsilon\mathrm{i} \times \mathrm{ki} \times (\mathrm{PM_s}) \\ \mathrm{Var}\ (EAT_{is}) &= \mathrm{Var}\ (\varepsilon\mathrm{i}\ (1 - \mathrm{ki} \times (\mathrm{PM_s}) \\ \mathrm{Var}\ (EAT_{is}) &= (1 - \mathrm{ki} \times \mathrm{PMs})^2 \sigma_{\varepsilon i}^2 \end{aligned} \tag{14}$$

3.2. Hypothesis development

Previous studies document the value relevance and informativeness of accounting income numbers (e.g., Ball & Brown, 1968; Beaver, 1968). Intuitively, revenues positively relate to value, whereas expenses negatively relate to value. However, some empirical results demonstrate the reverse, in which tax expenses are good news and positively related to a firm's value. For example, Graham, Raedy, and Shackelford (2012) provide comprehensive reviews regarding the information contents of income-tax-related disclosures in financial statements, while Thomas and Zhang (2014) extend the literature on income taxes showing that tax surprises (either the current or deferred component of tax expenses) positively explain contemporaneous returns as well as predict future returns and contain incremental value information to pre-tax GAAP income. Additionally, Hanlon et al. (2005) measure taxable income using the current portion of tax expenses, as an alternative proxy for pre-tax book income, and document that tax expenses are good news since it signals the core profitability of a firm.

Previous research also suggests a relationship between firm performance and partisanship. For instance, Alesina (1987) supports the partisan view of a macroeconomic policy, in which governments from different parties introduce different policies according to their respective platforms, while Santa-Clara and Valkanov (2003) investigate U.S. presidential administrations from 1927 to 1998 and find that excess and real stock returns are economically and statistically higher under Democratic presidencies than under Republican ones. Apergis, Cooray, and Rehman (2022) even find the spillover effects of U.S. partisan conflicts in the United Kingdom (U·K).

Interestingly, information contained in tax expenses becomes more transparent and informative to investors and analysts under some circumstances. At the firm level, Baik, Kim, Morton, and Roh (2016) study the role of pre-tax-income and earnings forecasts on tax surprises and conclude that the mispricing of tax expenses becomes less severe since such forecasts include the persistence of future earnings. At the macro level, Kerr (2019) finds that greater tax enforcement boosts the informativeness of tax surprises. Indeed, the information contents of tax surprises almost entirely remain within its interaction with higher tax enforcement. In related research, Mella (2021) finds that firms with a high level of corporate taxes have higher returns than firms with a low corporate tax level when the

president is from the Republican party. However, the opposite is true during periods with a Democratic president.

As discussed earlier, corporate tax expenses are expected to be higher under left-leaning governments due to their stronger desire for a higher level of public expenditure than right-leaning parties. Thus, the greater corporate tax expenses when the government belongs to left-leaning parties is unlikely to be fully interpreted as a signal for underlying profitability, but rather a higher statutory tax rate or broadened tax base. Thomas and Zhang (2011) argue that the market initially underreacts to the information contents of tax expenses, since the predictable implications of tax-related numbers are confusing, even to sell-side analysts (Dhaliwal, Gleason, & Mills, 2004). However, when tax policymakers are known for their desire to collect higher tax revenue as a result of their political preferences, increases in corporate tax expenses are in part concerned with a cash outflow that is not informative about a firm's future profitability. Consequently, investors should respond differently to firms with higher tax burdens conditional on the political orientation of tax incumbents in the markets outside of the U.S. To examine whether partisanship interacts with the information credibility of tax expenses, we present the following hypotheses:

H1. *Consistent with the literature regarding the positive relationship between stock returns and corporate tax expenses, for a firm, higher corporate tax expenses are related to higher returns under right-leaning administrations. However, corporate tax expenses are less informative under left-leaning governments.*

H2. *Corporate tax expenses explain firm-wide cross-sectional disparities in future returns under right-leaning governments. However, the explanatory power of corporate tax expenses on cross-sectional disparities in future returns is significantly weaker under left-leaning governments.*

Note that the main difference between the first and second hypotheses is that the former refers to the average effect for firms over time, while the latter refers to the cross-sectional variation among firms.

4. Material and methods

To test our hypotheses, we start with our sample of G12[3] countries, which include developed Western countries with parliamentary political systems. Among this group, three members are excluded from our investigation: Japan (which does not include governments from both sides of the political spectrum), Switzerland (which does not have a political variable from our source of political data (ParlGov), and the U.S. (which has been previous studied and includes a presidential system). Thus, the final sample consists of 10 countries: Australia, Belgium, Canada, France, Germany, Italy, the Netherlands, Spain, Sweden, and the U.K.

Table A-1 in the Appendix represents the system of government for the G12 countries in this study. Our sample size covers monthly data from 1990 to 2020. As opposed to the presidential system in the U.S., in which the president is the head of government as well as the head of state, this is not the case for the countries in our sample in which the head of state is usually different from the head of government. To address this issue, we take the political orientation of the head of government (Prime Minister) to define the political variables (Bohl & Gottschalk, 2006). Financial statement data and stock price data are obtained from Compustat Global Daily and Compustat Global Fundamental Annuals,[4] respectively. The former database provides daily prices and dividend adjustment factors to compute total returns at the firm level and convert them into a monthly basis, while the latter database provides historical financial statement data from 1986 to 2006. Our analysis begins in 1990, which is the earliest time available after matching both databases and excluding missing observations. We also exclude finance, real estate, and insurance industries (referring to the Standard Industry Classification (SIC) codes 6000–6799). All the returns and financial data are converted to U.S. dollars using the nominal currency exchange rate from Bloomberg. All the variables are defined in Appendix A-3 and winsorized at their 1st and 99th percentiles. Table 2 reports the descriptive statistics of these variables and ideology, while Table 3 shows the correlation matrix for the returns, corporate tax expenses, and tax expenses interacting with the ideology dummy variable.

4.1. Measure of corporate tax expenses

Most research on corporate taxes employs the ETR, which is the ratio of tax paid or accrued to pre-tax income (excluding loss firms to make the ratio meaningful). Loss firms (firms with negative pre-tax income) are a significant and relevant part of our sample. Moreover, Henry and Sansing (2018) document that the book value of assets is a reasonable scalar for normalizing total taxes. Thus, consistent with previous valuation research, as well as to avoid data truncation bias and capture all taxes borne by companies during a fiscal year, we compute the ratio of total income taxes at time t (Compustat TXT) to total assets at time $t-1$ (Compustat AT) to measure corporate tax expenses.

4.2. Measure of ideology

The ideology of the governments, as measured by party information, is extracted from the Parliament and Government Database

[3] The G12 consists of 13 countries.
[4] Financial statement data for Canada are obtained from Compustat-North America.

Table 2
Descriptive statistics.

Panel A All	N	Mean	SD	P10	P25	P50	P75	P90
Return%	1,438,517	0.94	20.01	−16.92	−7.41	−0.23	7.09	18.13
TaxExp%	1,438,517	1.36	3.74	−0.96	0.00	0.61	2.63	5.17
EBT	1,438,517	−0.06	0.33	−0.43	−0.12	0.03	0.09	0.18
SIZE	1,438,517	4.46	2.42	1.50	2.76	4.28	6.03	7.76
BTM	1,438,517	−0.59	1.24	−2.05	−1.27	−0.56	0.10	0.77
MOM	1,438,517	0.09	0.65	−0.59	−0.22	0.07	0.36	0.73
TaxEnforcement	1,438,517	5.50	1.20	3.89	4.76	5.57	6.46	7.00

Panel B RR (57%)	N	Mean	SD	P10	P25	P50	P75	P90
Return%	818,110	0.92	21.06	−16.54	−7.14	−0.19	6.90	17.52
TaxExp%	818,110	1.21	3.64	−1.08	0.00	0.55	2.44	4.86
EBT	818,110	−0.06	0.32	−0.42	−0.12	0.02	0.09	0.17
SIZE	818,110	4.49	2.50	1.43	2.75	4.32	6.10	7.88
BTM	818,110	−0.57	1.28	−2.03	−1.27	−0.57	0.10	0.80
MOM	818,110	0.09	0.67	−0.56	−0.21	0.07	0.36	0.71
TaxEnforcement	818,110	5.44	1.14	4.05	4.80	5.52	6.23	6.76

Panel C LL (43%)	N	Mean	SD	P10	P25	P50	P75	P90
Return%	620,407	0.97	18.53	−17.41	−7.76	−0.30	7.36	18.93
TaxExp%	620,407	1.56	3.88	−0.82	0.00	0.71	2.91	5.58
EBT	620,407	−0.06	0.34	−0.44	−0.12	0.03	0.10	0.18
SIZE	620,407	4.43	2.31	1.60	2.78	4.23	5.94	7.59
BTM	620,407	−0.61	1.18	−2.07	−1.28	−0.55	0.10	0.75
MOM	620,407	0.09	0.64	−0.62	−0.24	0.07	0.37	0.77
TaxEnforcement	620,407	5.59	1.28	3.89	4.72	5.84	6.58	7.03

Panel D Country		Return%			TaxExp%			Government	
	N	All	RR	LL	All	RR	LL	Right	Left
Australia	294,354	1.41	1.80	0.61	0.75	0.70	0.85	0.67	0.33
Belgium	24,795	1.69	1.62	2.16	1.84	1.89	1.43	0.88	0.12
Canada	278,820	1.19	0.22	2.02	0.96	0.55	1.33	0.46	0.54
France	150,837	0.77	0.81	0.69	1.73	1.68	1.81	0.62	0.38
Germany	170,804	0.63	0.63	0.64	1.86	1.79	2.06	0.73	0.27
Italy	41,878	0.47	0.08	0.91	1.95	2.06	1.82	0.53	0.47
Netherlands	34,757	0.78	1.19	−0.19	2.01	1.54	3.13	0.70	0.30
Spain	31,507	0.83	1.05	0.58	1.48	1.41	1.55	0.53	0.47
Sweden	89,852	0.98	0.55	1.20	1.27	1.45	1.17	0.34	0.66
U.K.	320,913	0.56	0.70	0.43	1.66	1.31	1.99	0.49	0.51

This table presents the summary statistics for all the variables (Panel A), the variables under right-leaning governments (Panel B), the variables under left-leaning governments (Panel C), and the variables by country (Panel D) used in the regression analyses. All the variables are defined in Table A-3 of the Appendix.

Table 3
Correlation matrix.

	Return	*TaxExp*	*TaxExp_LeftGov*	*TaxExp_RightGov*
Return	1	0.0012	0.0005	0.0012
TaxExp	0.0561*	1	0.6638*	0.7053*
TaxExp_LeftGov	0.0283*	0.5726*	1	−0.0620*
TaxExp_RightGov	0.0404*	0.6451*	−0.1827*	1

This table presents the pairwise correlations of returns and corporate tax expenses. The variables are defined in Table A-3 of the Appendix. Pearson's correlation coefficients are above the diagonal, while Spearman's correlation coefficients are below the diagonal. The correlations that are significant at the 5% level (or better) are in asterisks.

(ParlGov),[5] which is a survey-based database that combines parties, elections, and cabinets for all European Union (EU) and most OECD democracies. It also includes the "left-right" variable, which measures the position of each party according to its political orientation on a scale from 0 to 10. In this setting, 0, 5, 10 are indicative of extreme left-leaning, center-leaning, and extreme right-leaning governments, respectively. Consequently, we assign left-leaning (the variable **LeftGov**) to the numbers from 0 to 5 and right-leaning to the remaining numbers of the spectrum (the variable **RightGov**).

5. Empirical models and results

5.1. Panel with fixed effects analysis

To examine the value relevance of corporate tax expenses in the presence of partisanship and test our first hypothesis, we use panel analysis with fixed effects, as depicted in the following equation:

$$\begin{aligned}\mathbf{Return}_{it} =& \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 \mathbf{TaxExp}_{it-1} \\ &+ \boldsymbol{\theta}_2 \mathbf{LeftGov} + \boldsymbol{\theta}_3 \mathbf{LeftGov} \times \mathbf{TaxExp}_{it-1} \\ &+ \boldsymbol{\theta}_4 \mathbf{EBT}_{it-1} + \boldsymbol{\theta}_5 \mathbf{SIZE}_{it-1} + \boldsymbol{\theta}_6 \mathbf{BTM}_{it-1} + \boldsymbol{\theta}_7 \mathbf{MOM}_{it-1} + \boldsymbol{\theta}_8 \mathbf{TaxEnforcement}_{t-1} + \boldsymbol{\eta}_i + \boldsymbol{\tau} + \boldsymbol{\varepsilon}_{it}\end{aligned} \tag{15}$$

In this case, **Return** refers to buy and hold monthly stock returns (calculated in USD) and **LeftGov** is the dummy variable for left-leaning governments. The political dummy variable is known at the start of the return period. In addition, **TaxExp** refers to corporate tax expenses, while the control variables include **EBT** (which represents pre-tax income scaled by total assets) and **SIZE** (which is the natural logarithm of the firm's market value in USD). **BTM** is the natural logarithm of the book-to-market ratio and **MOM** indicates firm momentum, which is the return from months −2 to −12. Following Kerr (2019), we control the tax evasion index (**TaxEnforcement**), which is a proxy for tax enforcement obtained from the IMD World Competitiveness Online Database. The first column (R1) of Table 4 represents the coefficients and t-statistics estimated from the fixed effects regression depicted in Eq. (15). For the sake of clarity, we replace **LeftGov** and **LeftGov** × $\mathbf{TaxExp}_{it-1}$ with **RightGov** and **RightGov** × $\mathbf{TaxExp}_{it-1}$, respectively, and rerun the regression. The results are provided in the second column (R2).

Consistent with the existing literature on the value relevance of tax expenses, the coefficient for **TaxExp** is significant and positive, while the coefficient for the interaction term between tax expenses and the left-leaning government dummy (**LeftGov** × **TaxExp**) is significant and negative, as presented in column R1 (the corresponding coefficient (**RightGov** × **TaxExp**) under a right-leaning government is shown in column R2). These results are consistent with the first hypothesis that higher corporate taxes are associated with higher future returns under right-leaning governments. However, its impact is relatively lower under left-leaning administrations. Thus, the signaling effect (information contents) of corporate tax expenses is influenced by the party of the government. Moreover, the inclusion of the tax evasion index (**TaxEnforcement**) shows that our results hold after controlling for tax enforcement and differ from those of Kerr (2019).

5.2. Cross-sectional analysis

To test our second hypothesis, we estimate Fama and MacBeth (1973) cross-sectional regression coefficients of tax expenses and their interaction with the government's ideology, as presented in Eq. (16). We also include firm characteristics that are potentially correlated with a firm's tax burden and are known to explain the cross-section of returns.

$$\begin{aligned}\mathbf{Return}_{it} =& \boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_1 \mathbf{RightGov} \times \mathbf{TaxExp}_{it-1} + \boldsymbol{\lambda}_2 \mathbf{LeftGov} \times \mathbf{TaxExp}_{it-1} + \boldsymbol{\lambda}_3 \mathbf{EBT}_{it-1} + \boldsymbol{\lambda}_4 \mathbf{SIZE}_{it-1} + \boldsymbol{\lambda}_5 \mathbf{BTM}_{it-1} \\ &+ \boldsymbol{\lambda}_6 \mathbf{MOM}_{it-1} + \boldsymbol{\lambda}_7 \mathbf{TaxEnforcement}_{t-1} + \boldsymbol{\varepsilon}_{it}\end{aligned} \tag{16}$$

In this case, **Return** refers to buy and hold monthly stock returns (calculated in USD), while **RightGov** and **LeftGov** are the dummy variables for right- and left-leaning governments, respectively. Again, the political dummy variable is known at the start of the return period. Additionally, **TaxExp** refers to corporate tax expenses, while the control variables include **EBT** (which represents pre-tax income scaled by total assets) and **SIZE** (which is the natural logarithm of a firm's market value in USD). As earlier, **BTM** is the natural logarithm of the book-to-market ratio and **MOM** indicates firm momentum, which is the return from months −2 to −12. Following Kerr (2019), we also control the tax evasion index (**TaxEnforcement**). Overall, we are interested in the sign and magnitude of λ_1 and λ_2. Our second hypothesis is supported if λ_1 is significantly greater than λ_2.

Columns 3 and 4 of Table 4 display the coefficients for regressing returns on corporate tax expenses, the political dummy interaction with corporate tax expenses, and the control variables. The results in column 3 show that corporate taxes have predictive power for future returns in our sample. Up to this point, the results are consistent with Kerr (2019), but differ from Mella (2021), since they are unconditionally relevant to the party in office. Column 4 represents the interaction between the partisan effect and corporate tax expenses. Consistent with the literature regarding value-related information on corporate taxes, the coefficient for the right-leaning government indicator is positive and significant. However, tax expenses do not seem to b\eqalign{e informative about the underlying profitability of a firm under left-leaning governments. The results support our second hypothesis: Firms with higher tax expenses

[5] The ParlGov database (Döring & Manow, 2020) includes information on parties, elections, and cabinets in modern democracies. Development version: http://www.parlgov.org/#documentation.

Table 4
Value relevance of corporate tax expenses in the presence of partisanship.

	R1	R2	R3	R4
TaxExp	**0.044***** **(3.212)**	**0.021**** **(2.159)**	**0.046***** **(5.264)**	
LeftGov	0.00200 (1.470)			
RightGov		−0.00200 (−1.470)		
LeftGov × TaxExp	**−0.024**** **(−2.031)**			**−0.138** **(−0.0732)**
RightGov × TaxExp		**0.024**** **(2.031)**		**0.0408***** **(3.356)**
EBT	0.005** (2.425)	0.005** (2.425)	−0.000 (−0.135)	−0.000 (−0.223)
SIZE	−0.019*** (−8.619)	−0.019*** (−8.619)	−0.001*** (−3.235)	−0.001*** (−3.256)
BTM	0.001* (1.886)	0.001* (1.886)	0.004*** (6.776)	0.004*** (6.769)
MOM	0.005*** (2.856)	0.005*** (2.856)	0.014*** (8.777)	0.014*** (8.776)
TaxEnforcement	−0.002 (−1.632)	−0.002 (−1.632)	0.002** (2.019)	0.002* (1.953)
Constant	0.014 (0.647)	0.016 (0.698)	0.000483 (0.0704)	0.000726 (0.104)
Fixed effects	Firm & Time	Firm & Time	No	No
R-squared	0.105	0.105	0.038	0.039
Number of Firms	12,349	12,349		
Number of observations	1,438,517	1,438,517	1,438,517	1,438,517
Time periods	372	372	372	372

This table presents the fixed effect regression (R1 and R2) and Fama–MacBeth regression (R3 and R4) results of returns, TaxExp, and the interaction of TaxExp with the political orientation of the government. All the variables are defined in Table A-3 of the Appendix. Robust standard errors are clustered at the industry level. Significance levels are based on two-sided t-tests, provided in parentheses and indicated as follows: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

have higher returns under right-leaning governments, compared with firms that bear lower tax expenses. However, tax expenses do not explain the cross-sectional differences in future returns under left-leaning administrations. This extends the findings of Kerr (2019) by adding the political orientation of tax policymakers into corporate tax expenses and showing that this effect is incremental to that of tax enforcement.

The lack of informativeness regarding tax expenses and future returns under left-leaning governments can be explained by investors' underreaction. Thomas and Zhang (2011) discuss that the opacity of tax disclosure in financial reporting causes underreaction to future implications of tax variables, which justifies the predictive power of tax surprises. Conversely, when an administration is known for its desire to increase tax revenue, the underreaction to the news received from the corporate tax level most likely decreases because it is no longer considered a signal of the underlying profitability of a firm. Unlike accounting rules, tax rules are not set to create value, but to target other considerations such as welfare. Although the profit measured by tax rules contains information that is not conveyed by accounting profit (Lev & Nissim, 2004), the tax amount is an expense representing cash outflow, thus negatively affecting a firm's value. In related research, Mella (2021) implies that tax expenses are good news during Republican administrations and bad news under Democratic governments. In our sample, it appears that the good news of tax expenses (signal of core profitability) is offset by the bad news of being an expense (signal of cash outflow) under left-leaning governments.

6. Robustness tests

In this section, we perform a robustness test to investigate the impact of potential changes in tax policies caused by the 2008 financial crisis on our results. The aftermath of this financial crisis threatened the welfare state of most developed countries, while the economic issues within each country were worsened by global forces. Thus, the development of the welfare state of economies had to be mainly financed by setting higher taxes (Kowalski & Shachmurove, 2014). Following this financial crisis, OECD countries faced pressure to enhance tax revenues to stabilize public budgets and debts, while sustaining economic recovery through tax reductions. This dilemma may have enhanced the incentives for fiscal policymakers to enlarge the tax base in the following years, given the continuing downward trend in corporate tax rates.

The Organization for Economic Co-operation and Development (OECD) (2020) reports that average corporate tax revenues scaled by the GDP peaked in 2008, but dramatically declined in 2009 and 2010 (see Fig. A-2), mostly due to lower corporate profitability. Interestingly, we observe that the increase in corporate tax revenue following 2010 is not accompanied by the same trend in the average corporate statutory tax rate, since it remains relatively flat for approximately half a decade. This could imply the implementation of tax base expansion policies by governments. As discussed in Section 2.2, left-leaning governments are more inclined to implement tax base broadening policies, compared to their right-leaning counterparts. To address this financial matter, we split our

Table 5
Robustness check.

	1990–2010		2011–2020	
	R1	R2	R3	R4
TaxExp	0.0504***		0.0357***	
	(4.515)		(2.706)	
LeftGov × TaxExp		**−0.219**		**0.0321**
		(−0.0786)		**(1.582)**
RightGov × TaxExp		**0.0409****		**0.0406****
		(2.504)		**(2.580)**
EBT	−0.00273	−0.00306	0.00454**	0.00446**
	(−0.672)	(−0.752)	(2.168)	(2.126)
SIZE	−0.000675*	−0.000681*	−0.00167***	−0.00167***
	(−1.674)	(−1.691)	(−3.838)	(−3.853)
BTM	0.00433***	0.00431***	0.00262***	0.00264***
	(5.612)	(5.590)	(4.424)	(4.463)
MOM	0.0170***	0.0170***	0.00830***	0.00830***
	(7.627)	(7.629)	(4.923)	(4.917)
TaxEnforcement	0.00342**	0.00337**	−0.000	−0.000
	(2.168)	(2.105)	(−0.0239)	(−0.0480)
R-squared	0.046	0.048	0.021	0.021
Number of observations	808,639	808,639	629,878	629,878
Time periods	252	252	120	120

This table presents the Fama–MacBeth regression results of returns, CTEs, and the interaction of CTEs with the political orientation of the government after dividing the dataset into two time periods. All the variables are defined in Table A-3 of the Appendix. Significance levels are based on two-sided t-tests, provided in parentheses and indicated as follows: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

dataset into the 1990–2010 vs. 2011–2020 periods to statistically investigate whether these global trends cause any difference in our results.[6] We start with the second period in 2011 to ensure that the potential changes have been implemented. Table 5 presents the results for the split sample. Consistent with our main results, the coefficient for corporate tax expenses (without controlling for the political variable) is positive and significant. However, the informativeness of corporate tax expenses is contained within its interaction with right-leaning governments in both time periods.

7. Conclusions

This study adds to the existing literature on the information contents of corporate taxes by investigating whether the political orientation of tax decision-makers contributes to the informativeness of corporate tax expenses. The information contents of corporate taxes have been extensively studied, providing evidence that such tax expenses contain information regarding a firm's future profitability, which can explain contemporaneous and future returns (Graham et al., 2012; Hanlon et al., 2005; Thomas & Zhang, 2011). Using tax expenses as a more informative proxy for the future profitability of a firm arises from the possible manipulation of pre-tax accounting income through earnings management and tax planning strategies that can lower the quality of book earnings (Thomas & Zhang, 2014). Nevertheless, do corporate tax expenses generally reflect a firm's future profitability and explain stock returns? The circumstances that can affect tax informativeness have been largely unexplored in the literature, especially at an international level. Kerr (2019) utilizes an international setting and shows that tax expenses are more value-relevant in the presence of greater tax enforcement, and their informativeness is almost fully contained when interacting with such enforcement.

There are other circumstances that are influential on corporate taxation. Previous studies investigate the relevance of partisanship on corporate taxation and document that left-leaning governments tend to impose higher corporate tax expenses (Osterloh & Debus, 2012) through policies such as increasing the statutory corporate tax rate and expanding the tax base (Kawano & Slemrod, 2016). We explore the impact of ideological preferences of the party in office for corporate taxation on the informativeness of corporate tax expenses about future returns.

To carry out our study, we introduce a simple model suggesting that the information contents of corporate tax expenses are conditional on the political orientation of the government, which can be explained by investors' perspectives toward a firm's future cash flows and cash-flow volatility. Furthermore, we build our international sample by utilizing G12 countries, excluding the U.S., Japan, and Switzerland, and covering the period from 1990 to 2020. We first validate that firms bear greater tax expenses under left-leaning governments. This contributes to the political economy literature and supports the partisan theory of political cycles by showing that the political affiliation of the government on tax policies still matters, even in the wake of globalization. Then, after providing evidence on the value relevance of corporate taxes in our international setting, this study presents the results supporting the critical impact of the political orientation of the government on the informativeness of corporate tax expenses about future returns. Specifically, using cross-sectional differences in country level partisanship, we find that corporate tax expenses are only informative about future returns under right-leaning governments. This result implies that corporate tax expenses are arguably a more direct

[6] We would like to thank our anonymous reviewers for this suggestion.

profitability indicator and an important determinant of the information contents of corporate tax expenses when right-leaning governments are in charge. Moreover, following Kerr (2019), who documents the positive and critical effect of tax enforcement in the literature, our results show that the effect of partisanship on the information contents of corporate taxes is present after controlling for tax enforcement. However, it would be worthwhile for future research to explore the robustness of our results in developing markets and in countries with presidential systems. In addition, other measures of political orientation can be investigated, such as the policy preferences of the party, through content analysis.

CRediT authorship contribution statement

Parastoo Ostad: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Javier Mella:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

All authors declare that they have no conflicts of interest.

Data availability

The authors do not have permission to share data.

Appendix A. Appendix

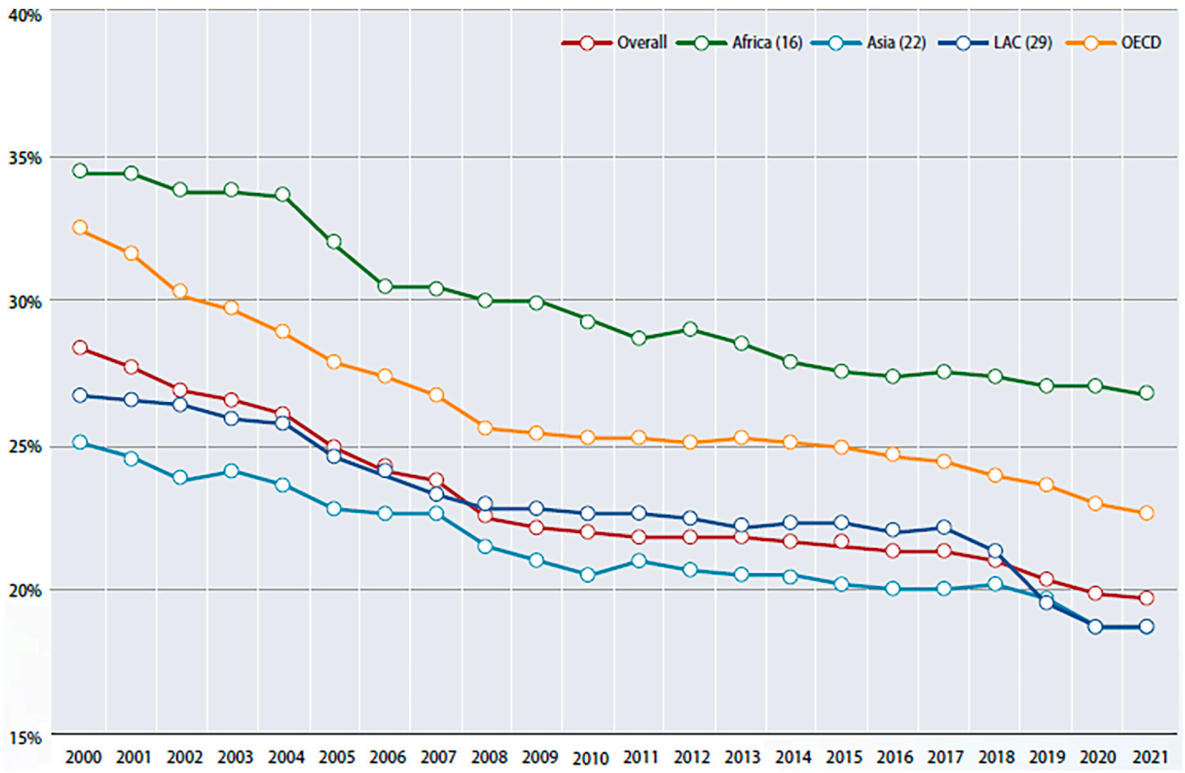


Fig. A-1. Average Statutory Corporate Tax Rates by Region.
Source: Organization for Economic Co-operation and Development (OECD) (2021), Corporate Tax Statistics.

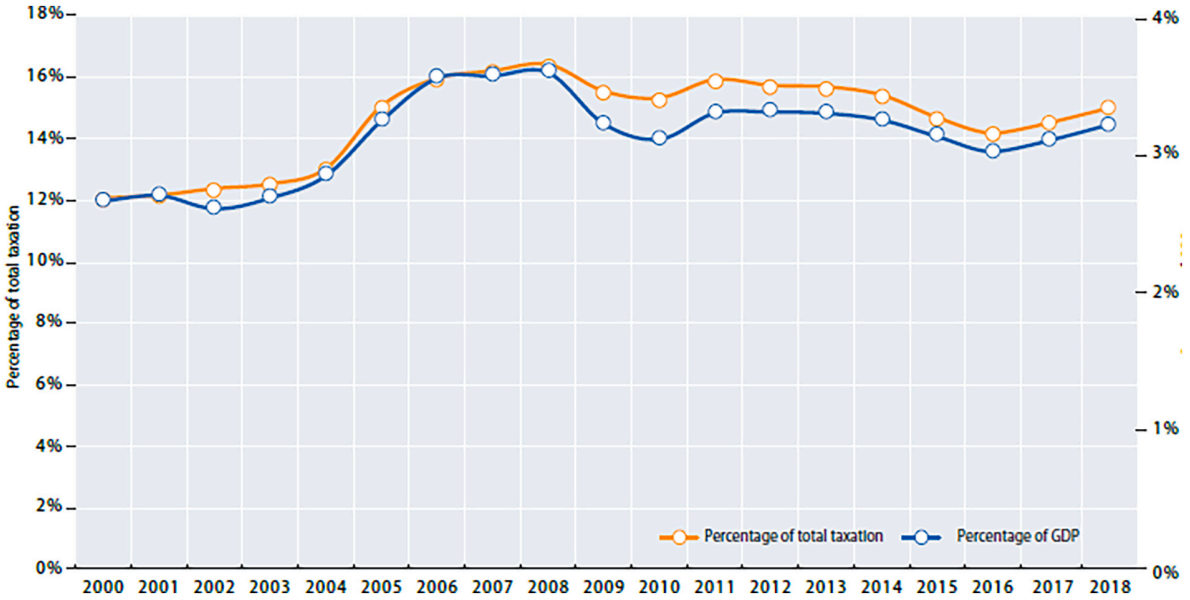


Fig. A-2. Average Corporate Tax Revenues as a Percentage of Total Tax and a Percentage of the GDP.
Source: Organization for Economic Co-operation and Development (OECD) (2021), Corporate Tax Statistics.

Table A-1
System of Government for the Members of the G12.

Country	Government	CM/R	Head of State	Head of Government
Australia	Parliamentary democracy	CM	Queen	Prime Minister
Belgium	Parliamentary democracy	CM	King	Prime Minister
Canada	Parliamentary democracy	CM	Queen	Prime Minister
France	Semi-presidential	R	President	Prime Minister
Germany	Parliamentary	R	President	Federal Chancellor
Italy	Parliamentary	R	President	President
Japan	Parliamentary	CM	Emperor	Prime Minister
Netherlands	Parliamentary	CM	King	Prime Minister
Spain	Parliamentary	CM	King	Prime Minister
Sweden	Parliamentary	CM	King	Prime Minister
Switzerland	Semi-direct & assembly	R	Federal Council	Federal Council
U.K.	Parliamentary democracy	CM	Queen	Prime Minister
U.S.	Presidential	R	President	President

This table shows the system of government for the members of the G12. CM and R stand for Constitutional Monarchy and Republic, respectively. We take the political orientation of the head of government to define the political variable.

A.1. Determinants of Corporate Tax Expenses

In this section, we describe the control variables specified in Eq. (1). Specifically, we describe the determinants of corporate tax expenses at the firm level. Firm size has been widely used in the literature as one of the main determinants of the ETR. However, the results are inconclusive, since firm size can positively (Zimmerman, 1983), negatively (Porcano, 1986), or insignificantly (Gupta & Newberry, 1997) affect corporate taxes. In this regard, we include the book value of assets (**BVA**) to control for firm size.

Tangible fixed assets provide tax benefits to firms since the depreciation expense of property, plant, and equipment is tax-deductible in all tax regimes. Most empirical evidence favors a negative relationship between the ETR and tangible fixed assets (Gupta & Newberry, 1997; Stickney & McGee, 1982). Conversely, Plesko (2003) shows a positive relationship between the two variables, whereas no significant association is observed in the study by Rodríguez (2004). Hence, we add net property, plant, and equipment (**PPE**) to control for depreciation, while the variable **Intangible** controls for the impact of intangible assets on corporate tax expenses, since they are arguably used for profit shifting to countries with lower tax rates (Dischinger & Riedel, 2011).

Profitability undoubtedly explains the variation in effective tax rates since larger profits lead to higher tax expenses for companies (Stickney & McGee, 1982). In this regard, we include return on assets (**ROA**) to control for a firm's profitability. We also include a dummy variable (**LossFirm)** as an indicator of firms with a net operating loss. In addition, the asset mix varies across activity sectors. Some industries, such as manufacturing, heavily depend on capital expenditures more than other sectors, such as wholesale and retail trades, which are inventory intensive. Consequently, inventories carry significant weight on the asset mix of the latter sectors and are a substitute for capital investment. Nevertheless, unlike capital-intensive companies, inventory intensive firms cannot benefit from the

tax-deductibility of depreciation, resulting in higher ETRs. Gupta and Newberry (1997) show a direct relationship between the ETR and inventory intensity. Thus, the variable **Inventory** in the model represents inventory.

To control for growth opportunities at the firm level, we use leverage (**LEV**), research and development expenses (**R&D**), the market to book ratio (**MTB**), and free cash flows (**FCF**). The impact of capital structure on tax expenses is extensively analyzed in the literature, with mixed results. For example, Gupta and Newberry (1997) suggest a negative relationship, while Harris and Feeny (2003) find a positive relationship. Conversely, Kim and Limpaphayom (1998) show no association between capital structure and a firm's tax expenses. The tax shield nature of R&D expenses, as well as the availability of tax credits on incremental R&D expenses, has turned them into beneficial tools for managing tax expenses and tax revenue by firms and tax policymakers, respectively. Gupta and Newberry (1997), and Richardson and Lanis (2007) find a reverse relationship between the ETR and R&D expenses. Regarding FCF, Dhaliwal, Huang, Moser, and Pereira (2011) suggest a direct relationship between cash holding and effective tax rates.

Furthermore, to control for growth opportunities at the industry level, we use the product market power (**PCM**) and the Herfindahl–Hirschman Index (**HHI**). Regarding the former, Kubick, Lynch, Mayberry, and Omer (2015) find that a firm's tax policy is affected by its ability to be the market leader. As for the latter, Hou and Robinson (2006) suggest the HHI to control for market competition within an industry.

It is also argued that competing nations tend to implement tax-lowering strategies to overcome their counterparts over capital mobility. Cassette and Paty (2008) find tax interdependence patterns for Western European countries (the former EU 15).[7] They further explain that the tax policymaking of such countries is highly affected by the tax policies of the region's leaders in terms of GDP per capita. Hence, we include the weighted average of the corporate statutory tax rate of the countries in our sample (**StatutoryTaxRate**), based on their GDP per capita. To avoid endogeneity issues, we instrument this variable using the residuals of STR over the other control variables.

Finally, we include GDP growth (**GDPgrowth**) to control for the business cycle. Moreover, Markussen (2008) finds that the policy sentiments are inclined to the left when the economy is booming, while Pástor and Veronesi (2020) find that when risk aversion is high, agents tend to vote Democrat and returns and growth are higher. Therefore, including lagged GDP growth can help avoid the possible endogeneity issue of ideology. In this case, $\boldsymbol{\eta}_i$ represents firm fixed effects, while $\boldsymbol{\tau}$ represents time fixed effects. Table A-2 demonstrates the descriptive statistics of the variables in Eq. (1).

Table A-2
Summary statistics of the determinants of corporate tax expenses.

Variable	N	mean	SD	min	p10	p25	p50	p50	p75	p90	max
TaxExp	126,058	0.016	0.039	−0.425	−0.009	0.000	0.009	0.009	0.029	0.055	0.635
LeftGov	56,832	0.451		0							1
RightGov	69,226	0.549		0							1
BVA	126,058	4.664	2.387	−2.191	1.667	2.983	4.533	4.533	6.214	7.919	12.163
LEV	126,058	0.142	0.252	0.000	0.000	0.000	0.074	0.074	0.216	0.374	56.223
PPE	126,058	0.340	0.416	0.000	0.018	0.068	0.226	0.226	0.482	0.801	20.626
Inventory	126,058	0.120	0.169	−0.346	0.000	0.002	0.057	0.057	0.190	0.320	24.239
ROA	126,058	−0.049	0.390	−8.178	−0.362	−0.075	0.037	0.037	0.100	0.183	2.961
Intangible	126,058	0.180	0.340	−0.002	0.000	0.000	0.049	0.049	0.247	0.513	31.261
MTB	126,058	2.261	2.500	0.000	0.000	0.673	1.497	1.497	2.905	5.606	15.868
R & D	126,058	0.024	0.078	−0.000	0.000	0.000	0.000	0.000	0.003	0.068	2.273
PCM	126,058	−1.050	3.928	−36.078	−1.676	−0.187	−0.033	−0.033	0.031	0.133	0.267
FCF	126,058	−0.049	0.200	−0.867	−0.326	−0.089	0.000	0.000	0.064	0.134	0.253
LossFirm	126,058	0.325	0.468	0							1.000
HHI	126,058	0.303	0.233	0.031	0.067	0.137	0.241	0.241	0.397	0.640	1.000
GDPgrowth	126,058	2.186	1.825	−5.694	0.255	1.487	2.374	2.374	3.421	3.970	6.869
StatutoryTaxRate	126,058	0.349	0.061	0.218	0.274	0.299	0.354	0.354	0.384	0.429	0.525

This table presents the summary statistics for the variables used in the validation test. All the variables are defined in Table A-3.

Table A-3
Variable definitions.

Variable	Definition	Source
Corporate Tax Expenses		
TaxExp	Total tax (TXT_{it}) divided by lagged total assets (AT_{it-1})	CGF*
Ideology		
LeftGov	ParlGov database (Döring & Manow, 2020), "left-right" variable. LeftGov = 1 if "left-right" ranges from 0 to 5, with LeftGov = 0 for the remaining numbers of the spectrum.	ParlGov
RightGov	ParlGov database (Döring & Manow, 2020), "left-right" variable. RightGov = 1 if "left-right" ranges from 5 to 10, with RightGov = 0 for the remaining numbers of the spectrum.	ParlGov

(*continued on next page*)

[7] Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, and the U.K.

Table A-3 (*continued*)

Variable	Definition	Source
Dependent Variable		
Return	$Return_i = (P_{monthlyt} - P_{monthlyt-1}/P_{monthlyt-1})$ P = (prccd * trf d/ajexdi), converted to a monthly basis and converted to USD = $P_{monthly.}$ Exchange Rate: Nominal rate from Bloomberg.	CGD Bloomberg
Control Variables		
EBT	$Profitability_{it-1}$ = pre-tax $income_{it-1}$ divided by lagged total assets (AT_{it-2}).	CGD
SIZE	$SIZE_{it-1}$ = Log [number of outstanding $shares_{t-1}$ (CSHOC)* closing $price_{t-1}$ (PRCCD)]; converted to USD.	CGD
BTM	Book-to-Market Ratio $_{it-1}$ = Log (Book Value of $Equity_{it-1}$)–$Size_{it-1.}$ Book Value of Equity = Equity total + deferred Tax + Deferred Taxes and Investment credit–preferred stock The equity total is equal to SEQ, if available. If SEQ is missing, then it is set equal to common equity plus the carrying value of preferred stock (CEQ + UPSTK), if available. If missing, then it is set to be equal to total assets (AT) less total liabilities (LT). Note that book value changes annually, while market value changes monthly.	CGD CGF
MOM	$Momentum_{-2,\ -12} = r-2 + r-3 + \ldots + r-12$	CGD
TaxEnforcement	Tax Evasion Index taken from the IMD World Competitiveness Online Database.	IMD
Validation Test Control Variables		
BVA	Log of lagged book value of assets = Log of (AT_{it-1})	CGD**
LEV	Leverage = Total Long-term Debt ($DLTT_{it}$) divided by lagged total assets (AT_{it-1}). If DLTT is missing, then we set it equal to 0.	CGF
PPE	Plant, Property and Equipment = Net property, plant, and equipment ($PPENT_{it}$) divided by lagged total assets (AT_{it-1}). If the PPENT is missing, then we set it equal to 0.	CGF
ROA	Return on Assets = Pre-tax income (PI_{it}) divided by lagged total assets (AT_{it-1}).	CGF

Variable	Definition	Source
Intangible	Intangible Assets = Intangibles ($INTAN_{it}$) divided by lagged total assets (AT_{it-1}). If the INTAN is missing, then we set it equal to 0.	CGF
MTB	Market To Book Ratio = lagged Market Value of Equity (MVA_{it-1}) to lagged Book Value of Equity (CEQ_{it-1}). Lagged Market Value of Equity = MVA_{it-1} = (prccd * cshoc) $_{t-1}$.	CGF CGD
Inventory	Inventories ($INVT_{it}$) divided by lagged total assets (AT_{it-1}). If INVT is missing, then we set it equal to 0.	CGF
R & D	Research and Development = research and development expense (XRD_{it}) scaled by lagged total assets (AT_{it-1}). If XRD is missing, then we set it equal to 0.	CGF
PCM	Price to Cost $Margin_i$ = [($SALE_{it}$ − $COGS_{it}$ − $EXGA_{it}$) /$SALE_{it}$] − [value weighted (based on sales) industry average (based on two-digit SIC classification)]. If COGS or EXGA are missing, then we set them equal to the operating profit after depreciation (OIADP).	CGF
FCF	Free Cash Flow = operating cash flows ($OANCF_{it}$) less Capital Expenditure ($CAPX_{it}$) divided by lagged total assets (AT_{it-1}). If OANCF or CAPX are missing, then we set them equal to 0.	CGF
LossFirm	Net Operating Loss = dummy variable that takes the value 1 if the operating profit after depreciation is negative (OIADP <0), or 0 otherwise.	CGF
HHI	Herfindahl–Hirschman Index = the sum of squares of ($SALE_i$/$SALE_{industry}$). Industry is defined by the two-digit SIC classification.	CGF
GDPgrowth	The annual percentage growth rate of the GDP at market prices based on constant local currency. World Bank and World Development Indicators (WDI).	WDI
StatutoryTaxRate	Weighted Average of Corporate Statutory Tax Rate: $W_{ij} = \frac{\left(\left\lvert \text{GDPperCapita}_i - \text{GDPperCapita}_j \right\rvert\right)^{-1}}{\sum_j \left(\left\lvert \text{GDPperCapita}_i - \text{GDPperCapita}_j \right\rvert\right)^{-1}}$. $\text{CompTax} = \sum_j W_{ij} \times T_j$. GDP per capita: (constant 2015 US$), World Bank. T: Corporate statutory Tax rate, TaxFoundation.org	WDI

Note that the financial statement data are assumed to be available three months after the end of a firm's fiscal year.
All the amounts are converted to USD using the nominal exchange rate from Bloomberg.
* Compustat Global Fundamental Annuals.
** Compustat Global Daily.

References

Alesina, A. (1987). Macroeconomic policy in a two-party system as a repeated game. *Quarterly Journal of Economics, 102*(3), 651–678. https://doi.org/10.2307/1884222

Alesina, A., Roubini, N., & Cohen, G. (1997). *Political cycles and the macroeconomy*. MIT Press.

Apergis, N., Cooray, A., & Rehman, M. U. (2022). US partisan conflict uncertainty and households' access to bank credit: Evidence from UK survey data. *Global Finance Journal, 100744*. https://doi.org/10.1016/j.gfj.2022.100744

Baik, B., Kim, K., Morton, R., & Roh, Y. (2016). Analysts' pre-tax income forecasts and the tax expense anomaly. *Review of Accounting Studies, 21*(2), 559–595. https://doi.org/10.1007/s11142-016-9349-z

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research, 6*(2), 159–178. https://doi.org/10.2307/2490232

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research, 6*, 67–92. https://doi.org/10.2307/2490070

Benoit, K., & Laver, M. (2006). *Party policy in modern democracies*. Cambridge University Press.

Bernardi, L. (2011). *Economic crisis and taxation in Europe [Working paper 652]*. Società Italiana di Economia Pubblica.

Blinder, A. S., & Watson, M. W. (2016). Presidents and the US economy: *An econometric exploration. American Economic Review, 106*(4), 1015–1045. https://doi.org/10.1257/aer.20140913
Bohl, M. T., & Gottschalk, K. (2006). International evidence on the democrat premium and the presidential cycle effect. *North American Journal of Economics and Finance, 17*(2), 107–120. https://doi.org/10.1016/j.najef.2005.10.001
Cassette, A., & Paty, S. (2008). Tax competition among eastern and Western European countries: With whom do countries compete? *Economic Systems, 32*(4), 307–325. https://doi.org/10.1016/j.ecosys.2008.06.003
Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance, 53*(6), 1839–1885. https://doi.org/10.1111/0022-1082.00077
Dhaliwal, D., Huang, S., Moser, W., & Pereira, R. (2011). *Corporate tax avoidance and the level and valuation of firm cash holdings [Working paper]*. The University of Arizona.
Dhaliwal, D. S., Gleason, C. A., & Mills, L. F. (2004). Last-chance earnings management: Using the tax expense to meet analysts' forecasts. *Contemporary Accounting Research, 21*(2), 431–459. https://doi.org/10.1506/TFVV-UYT1-NNYT-1YFH
Dischinger, M., & Riedel, N. (2011). Corporate taxes and the location of intangible assets within multinational firms. *Journal of Public Economics, 95*(7–8), 691–707. https://doi.org/10.1016/j.jpubeco.2010.12.002
Döring, H., & Manow, P. (2020). Data from: Parliaments and governments database (ParlGov): Information on parties, elections and cabinets in modern democracies. Development version [Dataset]. http://www.parlgov.org/.
Drazen, A. (2000). The political business cycle after 25 years. In *, 15. NBER Macroeconomics Annual* (pp. 75–117). MIT Press. https://doi.org/10.1086/654407.
Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy, 81*(3), 607–636. https://doi.org/10.1086/260061
Government Accountability Office (GAO). (2008). *Effective tax rates are correlated with where income is reported [Report to the committee on finance U.S. Senate, GAO-08-950]*. GAO. August.
Graham, J. R., Raedy, J. S., & Shackelford, D. A. (2012). Research in accounting for income taxes. *Journal of Accounting and Economics, 53*(1–2), 412–434. https://doi.org/10.1016/j.jacceco.2011.11.006
Gupta, S., & Newberry, K. (1997). Determinants of the variability in corporate effective tax rates: Evidence from longitudinal data. *Journal of Accounting and Public Policy, 16*(1), 1–34. https://doi.org/10.1016/S0278-4254(96)00055-5
Hanlon, M., Kelley Laplante, K., & Shevlin, T. (2005). Evidence for the possible information loss of conforming book income and taxable income. *Journal of Law and Economics, 48*(2), 407–442. https://doi.org/10.1086/497525
Harris, M. N., & Feeny, S. (2003). Habit persistence in effective tax rates. *Applied Economics, 35*(8), 951–958. https://doi.org/10.1080/0003684032000050577
Henry, E., & Sansing, R. (2018). Corporate tax avoidance: Data truncation and loss firms. *Review of Accounting Studies, 23*(3), 1042–1070. https://doi.org/10.1007/s11142-018-9448-0
Hibbs, D. A., Jr. (1992). Partisan theory after fifteen years. *European Journal of Political Economy, 8*(3), 361–373. https://doi.org/10.1016/0176-2680(92)90001-W
Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance, 54*(6), 2143–2184. https://doi.org/10.1111/0022-1082.00184
Hou, K., & Robinson, D. T. (2006). Industry concentration and average stock returns. *Journal of Finance, 61*(4), 1927–1956. https://doi.org/10.1111/j.1540-6261.2006.00893.x
Inclan, C., Quinn, D., & Shapiro, R. (2001). Origins and consequences of changes in U.S. corporate taxation, 1981–1998. *American Journal of Political Science, 45*(1), 179–201.
Kawano, L., & Slemrod, J. (2016). How do corporate tax bases change when corporate tax rates change? With implications for the tax rate elasticity of corporate tax revenues. *International Tax and Public Finance, 23*(3), 401–433. https://doi.org/10.1007/s10797-015-9375-y
Kerr, J. N. (2019). The value relevance of taxes: International evidence on the proxy for profitability role of tax surprise. *Journal of Accounting and Economics, 67*(2–3), 297–305. https://doi.org/10.1016/j.jacceco.2018.10.001
Kim, K. A., & Limpaphayom, P. (1998). Taxes and firm size in Pacific-Basin emerging economies. *Journal of International Accounting, Auditing and Taxation, 7*(1), 47–68. https://doi.org/10.1016/S1061-9518(98)90005-2
Kowalski, T., & Shachmurove, Y. (2014). The reaction of the U.S. and the European monetary union to recent global financial crises. *Global Finance Journal, 25*(1), 27–47. https://doi.org/10.1016/j.gfj.2014.03.002
Kubick, T. R., Lynch, D. P., Mayberry, M. A., & Omer, T. C. (2015). Product market power and tax avoidance: Market leaders, mimicking strategies, and stock returns. *Accounting Review, 90*(2), 675–702. https://doi.org/10.2308/accr-50883
Lev, B., & Nissim, D. (2004). Taxable income, future earnings, and equity values. *Accounting Review, 79*(4), 1039–1074. https://doi.org/10.2308/accr.2004.79.4.1039
Lev, B., & Thiagarajan, S. R. (1993). Fundamental information analysis. *Journal of Accounting Research, 31*(2), 190–215. https://doi.org/10.2307/2491270
Markussen, S. (2008). How the left prospers from prosperity. *European Journal of Political Economy, 24*(2), 329–342. https://doi.org/10.1016/j.ejpoleco.2007.07.003
Mella, J. (). *Corporate taxes, partisan politics, and stock returns [Working paper]*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3634417. Universidad de los Andes.
Organization for Economic Co-operation and Development (OECD). (2010). *Choosing a broad base-low-rate approach to taxation. OECD Tax Policy Studies 19*. OECD.
Organization for Economic Co-operation and Development (OECD). (2020). *Corporate tax statistics* (2nd ed.). OECD https://oe.cd/corporate-tax-stats.
Organization for Economic Co-operation and Development (OECD). (2021). *Corporate tax statistics* (3nd ed.). OECD https://oe.cd/corporate-tax-stats.
Osterloh, S., & Debus, M. (2012). Partisan politics in corporate taxation. *European Journal of Political Economy, 28*(2), 192–207. https://doi.org/10.1016/j.ejpoleco.2011.11.002
Pástor, Ľ., & Veronesi, P. (2020). Political cycles and stock returns. *Journal of Political Economy, 128*(11), 4011–4045. https://doi.org/10.1086/710532
Plesko, G. A. (2003). An evaluation of alternative measures of corporate tax rates. *Journal of Accounting and Economics, 35*(2), 201–226. https://doi.org/10.1016/S0165-4101(03)00019-3
Porcano, T. (1986). Corporate tax rates: Progressive, proportional, or regressive. *Journal of American Taxation Association, 7*(2), 17–31.
Richardson, G., & Lanis, R. (2007). Determinants of the variability in corporate effective tax rates and tax reform: Evidence from Australia. *Journal of Accounting and Public Policy, 26*(6), 689–704. https://doi.org/10.1016/j.jaccpubpol.2007.10.003
Rodríguez, E. F. (2004). Los factores condicionantes de la presión fiscal empresarial española a partir de la información contable. Especial mención a las decisiones financieras [the determining factors of the Spanish business tax burden through the accounting information. The case of financing decisions]. *Spanish Journal of Finance and Accounting / Revista Española de Financiación y Contabilidad, 33*(120), 125–159. https://doi.org/10.1080/02102412.2004.10779516
Santa-Clara, P., & Valkanov, R. (2003). The presidential puzzle: Political cycles and the stock market. *Journal of Finance, 58*(5), 1841–1872. https://doi.org/10.1111/1540-6261.00590
Stickney, C. P., & McGee, V. E. (1982). Effective corporate tax rates: The effect of size, capital intensity, leverage, and other factors. *Journal of Accounting and Public Policy, 1*(2), 125–152. https://doi.org/10.1016/S0278-4254(82)80004-5
Thomas, J., & Zhang, F. X. (2011). Tax expense momentum. *Journal of Accounting Research, 49*(3), 791–821. https://doi.org/10.1111/j.1475-679X.2011.00409.x
Thomas, J., & Zhang, F. X. (2014). Valuation of tax expense. *Review of Accounting Studies, 19*, 1436–1467. https://doi.org/10.1007/s11142-013-9274-3
Zimmerman, J. L. (1983). Taxes and firm size. *Journal of Accounting and Economics, 5*(2), 119–149. https://doi.org/10.1016/0165-4101(83)90008-3

Chapter 2

The Direct and Indirect Impact of Non-cognitive Skills on Stock Market Participation

Contents

The Direct and Indirect Impact of Non cognitive Skills on Stock Market Participation

Richard Deaves[a], Parastoo Ostad[b,*], Adam Stivers[c]

[a]McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada
[b]McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada
[c]University of Wisconsin, 1725 State St, La Crosse, WI 54601, United States

September 2023

Abstract: Data from the *Understanding America Study* survey panel are used to investigate the role that both cognitive and non-cognitive skills play in the empirically-observed stock market participation puzzle. Specifically, we examine both the direct and indirect impacts of the "primitive" factors of survey-measured intelligence and non-cognitive skills (found as an equal-weighted combination of survey-based Conscientiousness and Emotional Stability instruments) on stock market participation levels (i.e., whether one has direct or indirect stock holdings). We examine to what extent these primitive factors operate through well-observed proximate factors of stock market participation such as financial literacy, education, income, and being trusting. We find evidence that non-cognitive skills both directly and indirectly (via the proximate factors) encourage stock market participation. We also show that cognitive skills affect participation rates, but the impact operates entirely in an indirect fashion through proximate factors. Overall, we show that higher levels of both cognitive and non-cognitive skills significantly increase the likelihood of owning stocks.

Keywords: stock market participation; cognitive skills; intelligence; non-cognitive skills; financial literacy; stock return ignorance

JEL classification: D91, G11, J24, D14, G41, G53

* Corresponding author
E-mail addresses: deavesr@mcmaster.ca (R. Deaves), ostads@mcmaster.ca (P. Ostad), astivers@uwlax.edu (A. Stivers)

1. Introduction

One of the most intensively-researched puzzles in household finance is the stock market participation ('SMP') puzzle: despite the well-known theoretical result that even at very high levels of risk aversion some equity exposure is optimal (Merton, 1969, 1971), many (as much as 60% according to Haliassos & Bertaut, 1995) do not invest in the stock market at all. An inexplicably high level of risk aversion is required to eliminate all equity demand (Heaton & Lucas, 2000). Ruling out extreme risk aversion, for an individual to abstain from the equity market, the most obvious reasons are: 1) they do not have enough money or basic knowledge; 2) they do not trust the market and those involved; and 3) they do not believe they have enough *specific* knowledge for the task at hand. Indeed, the empirical evidence supports that these are important reasons for reluctance. Income (or wealth), providing the means, and general education, providing the knowledge base, both help surmount any participation costs (e.g., Haliassos & Bertaut, 1995). A feeling that "most people can be trusted" induces the holding of equity (Guiso, Sapienza & Zingales, 2008). And task specific-knowledge, whether financial literacy (Van Rooij, Lusardi & Alessie, 2011) or the (more-specific) absence of stock return ignorance (Merkoulova & Veld, 2022), have all been shown to be associated with higher SMP.

While these readily observable factors have been found to help explain the SMP puzzle, we argue that additional forces may be at play. Sometimes much of what is going on is obscured: a focus on what is in plain sight can miss important details. The five determinants (and other such variables) we stressed previously can be viewed as *proximate* as opposed to *primitive* determinants of SMP. Proximate determinants (or 'proximates') tend to be in front of the curtain; primitive determinants (or 'primitives') often behind. By "primitive" determinants we mean what you are born with, the skills and other attributes that (for better or worse) influence your behavior (either

directly or indirectly through proximate determinants). The main innate skill groupings are cognitive skills (loosely categorized as intelligence) and non-cognitive skills, commonly organized as beneficial personality trait groups.

While Sias, Starks & Turtle (2020) have recently drilled down to genetic primitives to explain SMP, most research (ours included) using primitives works with measured-in-adulthood measures of (in large part) innate characteristics, which are thus arguably not quite fully primitive. Nonetheless, the evidence is strong that cognitive skills and non-cognitive skills – even if both are somewhat influenced by environment – are quite stable once someone enters adulthood.[1]

To clarify the distinction between primitives and proximates, consider the (primitive) intelligence and the (proximate) financial literacy. Grinblatt, Keloharju & Linnainmaa (2011) find that for Finland a high IQ was strongly predictive of SMP. And yet later work finds that in multivariate regressions which include an array of proximate causes, intelligence may be rendered statistically insignificant (e.g., Merkoulova & Veld, 2022). This mini-puzzle is easily explained (as we later illustrate): intelligence is highly correlated with (and likely causative of) various proximate determinants of SMP. While important, its impact is almost entirely subsumed by proximate causes.

The other salient set of primitives is personality traits. While there are various ways to organize personality traits, it is now most commonly done using the Five-Factor Model (or the 'Big 5'), whose five trait groups are Openness-to-Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism. Note that it will normally be convenient for ease of interpretation

[1] See, for example, Donnellan & Lucas (2008) and Roberts & DelVecchio (2000). An example of a proximate influencing a measured primitive is the finding that education enhances measured intelligence (e.g., Rindermann, 2008).

to work with the opposites (i.e., negatives) of the latter two, Disagreeableness and Emotional Stability. It has been shown in numerous studies that these Big 5 traits do a good job capturing much of an individual's core personality.[2] When traits encourage *positive* financial behaviors, it is now common to refer to them as non-cognitive skills or abilities (e.g., Kuhnen & Melzer, 2018; Choi & Laschever, 2018; and Parise & Peijnenburg, 2019). Two of the Big 5 are often highlighted in this regard, namely Emotional Stability and Conscientiousness (e.g., Parise & Peijnenburg, 2019). These two traits have consistently been associated with positive economic outcomes such as effective credit management (Parise & Peijnenburg, 2019; and Choi & Laschever, 2018). Besides, Emotional Stability and Conscientiousness can be viewed not merely as personality traits but as skills. In contrast, the other three Big Five traits are not typically considered 'skills' because their effects can be negative, positive, or neutral depending on the context and dataset. Therefore, we treat Emotional Stability and Conscientiousness as indicators of non-cognitive abilities, while using the remaining three traits as control variables to avoid their problematic interpretation.

To our knowledge, little work has been undertaken on the relationship between SMP and various components of personality, exceptions being Brown & Taylor (2014), Conlin *et al.* (2015), Sias, Starks & Turtle (2020) and Salamanca, De Grip, Fouarge & Montizaan (2020), who (despite multiple divergences) show that various traits are positively or negatively associated with SMP. While we review this research more fully in section 2, none of this work focuses on non-cognitive skills as we do. Indeed, there is a lack of evidence on the impact, either direct or indirect through proximate factors, that non-cognitive skills may have on SMP. One of our main goals is to remedy this deficiency. We focus on a binary SMP measure: whether one has any stock holdings, be it

[2] It is conventional to capitalize when referring to the Big 5 traits. Using the first letter of each trait the commonly used mnemonic OCEAN results. See McCrae (2009) for an overview of the Big 5. While these are more properly viewed as trait groups (i.e., groups of closely related traits), we will usually abbreviate by calling them "traits."

direct, indirect, or within retirement savings accounts. While studies of equity holding levels and retirement savings versus non-retirement savings have their merits, our focus is on shedding further light on why some individuals fully reject the equity market. Thus, our discussion of SMP throughout should be viewed through that lens. Another related goal is to explore the relative roles of primitive and proximate factors in explaining the stock market participation puzzle. The major questions we address are: 1) ignoring proximate determinants, to what extent are primitive determinants, namely cognitive skills and non-cognitive skills, predictive of SMP? 2) To what extent are proximate determinants of SMP driven by primitive determinants? 3) Do primitive determinants operate only through proximate determinants, or do they also directly affect SMP?

To the extent that SMP is driven by primitive factors, policy and interventions such as financial literacy programs and education are likely to be rendered ineffective. However, if these primitive factors are mostly driving SMP indirectly via proximate factors, then interventions aimed at the proximate factors (such as increasing financial literacy, education, and addressing income inequality) can be effective. It is also helpful to understand the relative role or contribution level of a given proximate factor after common primitive factors have been controlled for. For example, we know that all else equal, higher education leads to higher SMP. However, if it can be shown that all or most (or at least all statistical significance) of education's impact on SMP is due to primitive factors, this would indicate that education is not a viable intervention mechanism to increase SMP. Thus, it is important from both a policy-oriented and educational perspective to understand the contributions (both direct and indirect/relative) of both primitive and proximate factors in explaining SMP.

To preview, our main finding is that non-cognitive skills are incrementally predictive of higher SMP. While we examine the impact of individual beneficial Big 5 traits separately, we believe that

the growing recognition in the literature that several non-cognitive facets of personality induce positive financial behaviors justifies the use of a composite measure constructed from these individual beneficial Big 5 traits. Indeed, one of our goals is to begin exploring what such a composite measure might look like. Non-cognitive skills, as we measure via an equal-weighted combination of survey-based Conscientiousness and Emotional Stability instruments, operate both indirectly through their impact on such proximate variables as income, education and financial literacy and also via a direct channel of influence not operating through such intermediaries. On the other hand, cognitive skills' entire impact is indirect. Further, we find that the common proximate factors from the literature have a significant contribution due to both primitive and other, unobserved, external factors (with the latter typically driving a majority of the contribution to SMP). Graph 1 outlines the study's framework by categorizing the primitive and proximate factors and illustrating how primitive factors can influence stock market participation (SMP) both directly and indirectly through proximate factors.

(Graph 1 HERE)

Thus, the main contribution of our study can be summarized as follows. Our results suggest that education, financial literacy, and policy interventions aimed at increasing entry into equity markets are likely to be effective, although a sizable portion of SMP is due to innate primitive factors. We also contribute to the stock market participation debate in showing that non-cognitive skills are another important determinant. Our third contribution is that we begin to explore what elements of personality should and should not be incorporated into a workable definition of non-cognitive skills as it relates to financial decision making.

In section 2, we review the relevant literature on the determinants of stock market participation. The next section describes the data and variables. The main empirical findings are presented in

section 4, with extensions and robustness in the penultimate section. Finally, section 6 concludes.

2. Literature Review

2.1. Proximate determinants

Vissing-Jørgensen (2002, 2003) stress the role of fixed market-participation costs in explaining non-participation in equity markets.[3] These costs, both monetary and non-monetary, include the effort and any charges incurred setting up an investment account and various information costs (such as learning about investments), and could rationally deter those with insufficient means from investing in risky securities.[4] High enough income or wealth levels should eliminate this impediment. Indeed, it is empirically clear that income and wealth are predictive of SMP. Virtually all SMP research controls for income or wealth universally finding it to be positively associated with this behavior. On this theme, "windfall research" illustrates the importance of means, while also showing it cannot be the whole story. A natural experiment based on the behavior of lottery winners reinforces the fact that participation costs are only a piece of the puzzle: only 12% of non-participating people winning a $150,000 prize (which should safely obviate the low-wealth threshold) begin to participate (Briggs, Cesarini, Lindqvist & Östling, 2021). Similar boosts in SMP occur after unexpected inheritances (Andersen & Nielsen, 2011) and unexpected housing market appreciation (Kong, Cheng & Liu, 2021). General education levels are also associated with greater SMP, an example being Cole, Paulson & Shastry (2014) who find that an additional year of compulsory schooling increases SMP by 4%. Almost all SMP studies control for general education, finding it to be predictive of participation.

[3] Alan (2006) estimates participation costs. Paiella (2007) finds supportive evidence for their existence.
[4] Attanasio & Paiella (2011) argue that participation costs are unlikely to be the full story.

Information sharing can serve to reduce participation costs. Those with close family recently entering the market are more likely to enter themselves (Li, 2014). And those actively using the internet are more likely to invest in equities (Bogan, 2008).[5] Hurdles are also more easily mounted when the subjective expected market return is high, perceived (forward-looking) market risk is low, or risk aversion is low. Along these lines, Hurd, Van Rooij & Winter (2011) find that an individual's subjective expected market return and perceived market risk are both predictive of whether or not stocks are purchased. Psychology can play a role: those high in dispositional optimism are more likely to buy stocks (Puri & Robinson, 2007), while those more susceptible to the pessimism-inducing negativity bias (which is the tendency to stress negative events) are less likely to do so (Sias, Starks & Turtle, 2023).[6] Many researchers have controlled for risk aversion and have found this proximate variable too has the expected impact (Dimmock & Kouwenberg, 2010; Dimmock & Kouwenberg, Mitchell & Peijnenburg, 2016; Guiso, Sapienza & Zingales, 2008; Haliassos & Bertaut, 1995; Hong & Kubik, 2004; and Shum & Faig, 2006).

Beyond participation costs, a lack of trustfulness can be impactful: when one does not trust someone or the environment (such as the stock market or those involved in it), it is natural to stay away. Guiso, Sapienza & Zingales (2008) survey about 2,000 Dutch households, asking whether "most people can be trusted or... (do) …you have to be very careful in dealing with people." Being trustful significantly predicts stock ownership: those who are trusting are 50% more likely to invest in common stock. On the other hand, Georgarakos & Pasini (2011) find that "sociability" can somewhat offset low community-level trust leading to greater participation.

[5] For contrary evidence, see Liang & Guo (2015).
[6] For expected return, also see Das, Kuhnen & Nagel (2020) and Guiso, Sapienza & Zingales (2008).

In support of this trust explanation, other researchers find that when highly publicized corporate fraud occurred in a US state, this served to lower market participation in that state going forward (Giannetti & Wang, 2016). Community and social interaction likely reflect and facilitate trust (along with information sharing, thereby reducing participation costs). Social households (where being "social" manifests itself in such behaviors as volunteering, joining clubs, going to church, etc.) are substantially more likely to invest in the market (Hong, Kubik & Stein, 2004). The impact of socialization may also operate through peer effects. In Finland, witnessing peer market success encourages stock ownership (Kaustia & Knüpfer, 2012). A causal relation between someone investing in the markets and average stock market participation in their community also speaks to peer effects (Brown, Ivkovic, Smith & Weisbenner, 2008). Further, risk aversion can be influenced by peer effects (Ahern, Duchin & Shumway, 2014).

Someone with means, general knowledge and trustfulness might still hold back if they do not understand the specific environment of investing in equities. Christiansen, Joensen & Rangvid (2008) find that completing an economics degree helps further. Higher scores obtained from explicit instruments to measure financial literacy also boost SMP (Van Rooij, Lusardi & Alessie, 2011).[7] Also, the ability to provide point estimates and confidence intervals for future market returns (even if the answers are unreasonable) – what Merkoulova & Veld (2022) call the absence of "stock return ignorance" – is also positively associated with SMP. Specifically, survey respondents were asked for their 10-year average market return forecasts along with 10% and 90% distributional percentiles. Merkoulova & Veld (2022) characterize about 70% of their sample as "stock return-ignorant," because they 1) could not answer the question though prompted (the authors label this "expected return-ignorance"); 2) could not come up with a plausible distribution

[7] Also, see Lusardi & Mitchell (2014) for the evolution of financial literacy testing.

("risk-ignorance"); or 3) their return forecasts, at 30% per year or more for 10 years, were improbably bullish ("overoptimism"). The absence of the first two behaviors is shown to increase SMP, while the absence of the last behavior (affecting only about 5% of their sample) decrease it. Along the lines of ignorance of the market, even being merely "aware" of the market helps (Guiso & Jappelli, 2005).

All factors reviewed up to now have been arguably rational. But, behavioral factors, especially non-standard preferences, also have been shown impact market participation.[8] If investors are ambiguity-averse as well as risk-averse, a low percentage of people invested in the market can be explained (Antoniou, Harris & Zhang, 2015; Dimmock, Kouwenberg, Mitchell & Peijnenburg, 2016; Peijnenburg, 2018; and Dangl & Weissensteiner, 2020). And loss aversion, a key component of prospect theory, also predicts low SMP (Dimmock & Kouwenberg, 2010).[9] Another non-standard alternative model is presented in Ang, Bekaert & Liu (2005), who show that a model incorporating "disappointment aversion" can explain low participation.

2.2. Primitive determinants

Turning to primitives, there are many possible innate factors. Some, perhaps surprisingly, are associated with SMP. For example, Addoum, Korniotis & Kumar (2017) document an impact from body shape. However, our stress will be on cognitive skills and non-cognitive skills.

Cognitive skills (or ability) for our purposes will be viewed as the ability to perform well on standard intelligence tests. Three papers published around the same time using three different datasets conclude that cognitive ability is an important positive determinant of SMP, even when

[8] Deaves (2023) reviews the SMP literature in terms of both rational and behavioral factors.
[9] Also see Benartzi & Thaler (1995, 1999), Barberis & Huang (2007) and Barberis, Huang & Thaler (2006).

general education is included in estimation (Grinblatt, Keloharju & Linnainmaa, 2011; Christelis, Jappelli & Padula, 2010; and Georgarakos & Pasini, 2011). Nevertheless, as mentioned earlier, intelligence may cease to be statistically significant when task-specific knowledge is included as *an explanatory variable.*

2.3. Personality and SMP

The other major category of primitives is personality traits. As mentioned in the introduction, Brown & Taylor (2014), Conlin *et al.* (2015), Sias, Starks & Turtle (2020) and Salamanca, De Grip, Fouarge & Montizaan (2020) investigate the impact of personality on SMP (as well as sometimes other behaviors). Nevertheless, none of this research focuses on Big 5-based non-cognitive skills (as we do), nor do any of them include the same range of proximate determinants, especially those relating to task-based knowledge (as we do). Beginning with Brown & Taylor (2014), they use the *British Household Panel Survey*, finding that Disagreeableness is positively predictive of SMP, while Emotional Stability and Conscientiousness play no significant role (unlike our findings reported below).[10] They control for education and income, but not for trustfulness or any form of task-specific knowledge.

Conlin *et al.* (2015), in investigating SMP in Finland, employ a different taxonomy of traits, the so-called Temperament and Character Inventory of Cloninger, Svrakic, & Przybeck (1993), whose four traits are Novelty-Seeking (related to Openness-to-Experience from the Big 5), Harm-Avoidance (related to Neuroticism), Reward-Dependence (related to Extraversion) and Persistence (related to Conscientiousness). They show that Harm-Avoidance is negatively related to SMP at

[10] These findings are for single people. Not strictly comparable to our work, when they look at couples (and use the average trait value of the two people in a couple) they find that Extraversion/Openness-to-Experience is negatively/positively related to SMP.

5% significance when controlling for income and education. Like Brown & Taylor (2014), they do not control for trust and various forms of task-specific knowledge.

Sias, Starks & Turtle's (2020) study is novel in the SMP literature in that they examine whether genetic endowments related to cognition, personality, health, and body shape (known as "genotypes") are predictive of SMP. They use eight such genetic endowments; each genetic endowment is captured by a "polygenic score." For example, the polygenic score related to neuroticism is a weighted average of genomes that form one's potential characteristic of being neurotic. They document that these genotypes can predict SMP directly or through realized characteristic channels (known as "phenotypes"). Very loosely, our primitives can (for the most part) be viewed as intermediate between their genotypes and phenotypes. One such genotype is associated with a manifestation of Neuroticism, and it turns out to be negatively associated with SMP, similar to what we find for Emotional Stability. However, there is an important distinction to be made between the proxy we employ for Neuroticism and their analogous genotype. In their work, Neuroticism is a genetic endowment, while in our study a Neuroticism score is extracted from a self-reported Big 5 personality trait questionnaire (see section 3) and is a realized characteristic resulting from the interaction of genetic endowment with the environment. Therefore, our proxy for Neuroticism is the realization of one's genetic potential. Two other differences between Sias, Starks & Turtle (2020) and our work is they do not study the impact of Conscientiousness on SMP (which is one of the components of our proxy for non-cognitive skills), and financial literacy (likely because of data limitations) does not enter their group of phenotypes.

Finally, Salamanca, De Grip, Fouarge & Montizaan (2020) employ as their personality driver (internal) locus-of-control (Rotter, 1990), defined as the extent to which one believes that life outcomes are due to personal efforts, finding that it is predictive of SMP. They consider various

other variables as possible intermediaries. For example, using the data from Hurd, Van Rooij & Winter (2011) who (it will be recalled) show that those with higher return expectations and lower perceived forward-looking volatility are more likely to be in the market, the introduction of these explanatory variables into a regression of SMP on locus-of-control (plus various controls) does not eliminate the significance of locus-of-control. Their reasonable interpretation is that locus-of-control is not entirely operating through these two variables. They perform similar analyses for financial literacy, overconfidence, optimism, trustfulness and the Big 5 traits, in each case finding that locus-of-control retains its impact. It is also worth noting that when they control for the Big 5, while Emotional Stability is insignificant in most specifications, Conscientiousness has an unexpected sign. While these patterns seem anomalous, given the evidence that locus-of-control is positively correlated with non-cognitive skills and intelligence (Hattrup, O'Connell & Labrador, 2005), it is not surprising that the impact of the former is largely subsumed by locus-of-control.

3. Data

Our main dataset is survey data collected in 2019 from the *Understanding America Study* ('UAS'), a probability-based household internet panel representing the US adult population, to study the proximate and primitive determinants of stock market participation.[11] Administered by the University of Southern California, the UAS panel includes over 500 surveys, distributed to the same (as well as new) participants, on assorted topics such as health and retirement, economic well-being, cognitive and non-cognitive skills and other psychological constructs. Conveniently, each survey (denoted by survey number) can be linked to data collected from previous UAS

[11] Since its inception in 2014 up until Sep 2023, 334 published academic papers have used data collected in the UAS. See https://uasdata.usc.edu/index.php. Alattar, Messel & Rogofsky (2018) provide further background on the UAS.

surveys. Participants are provided anonymity and are paid USD 1.50 per minute spent answering survey questionnaires. The surveys we employ went to the field during 2019 (or shortly before). Note that our dataset is an augmented version of the dataset in Merkoulova & Veld (2022). Details on the definitions of variables used appear in Table A.1 of the Empirical Appendix. Table A.2 of the Empirical Appendix presents characteristics of the sample which consists of 3,134 panel respondents. Importantly, the survey response rates are excellent, and the panel resembles the general population in most respects.

The binary variable *SMPart* (we use italics for variable names) is an indicator for stock market participation, here defined as holding equity either directly or indirectly in retirement accounts.[12] Note that 48.5% of the respondents declared they had stock in their financial investment portfolios, which is in line with other survey results. Regarding the primitive factors, *CS* (measuring intelligence or cognitive skills) is calculated by summing up respondents' scores on questions from the Woodcock-Johnson Tests of Cognitive Abilities (Mather & Jaffe, 2016) regarding number series (obtained from UAS 83), picture vocabulary (UAS 84) and verbal analogies (UAS 85), and then standardizing to create a zero-mean unit-variance variable for ease of interpretation. *NCS2*, our two-trait principal proxy for non-cognitive skills, is the standardized average of *Conscientious* (the instrument score for Conscientiousness) and *EmotStab* (the instrument score for Emotional Stability). These and other Big 5 trait scores are based on a 44-question survey module from UAS 121. The mean of *Conscientious* demonstrates that on average respondents had a propensity to deliberate, plan, control impulses, or in a word, be conscientious. The distribution of *EmotStab*

[12] UAS survey 184 provides data on SMP. Stocks are defined as the stocks of individual firms and investments in equity mutual funds.

scores is skewed to the right, corresponding to a higher level of Emotional Stability vs. Neuroticism in our sample.

Beyond these primitive variables we also focus on proximate factors (which we also standardize) that are known to help explain SMP. Income is measured in 16 categories ranging from less than $5,000 to more than $150,000. *Inc* is its continuous representation, with the lowest category assigned '1'and the highest '16.' About 50% of the sample earned $60,000 or more. Education is measured in five categories, allowing for a similar continuous representation *Educ*: not having a high school diploma (which takes zero); a high school diploma or college below bachelor's ('1'); a bachelor's degree ('2'); a master's degree or professional school degree ('3'); and PhD ('4'). The majority of respondents in the sample (57%) held either a high school diploma or attended college, with bachelor's degrees (22%) and master's degrees (14%) placing second and third. Financial literacy (*FinLit),* between 0 and 14, is the respondent's score on a 14-question financial literacy test obtained from UAS 121, with 13 of the 14 questions used coming from the questions of Van Rooij, Lusardi & Alessie (2011). The mean of *FinLit* is closer to the upper bound of the bracket showing that respondents who correctly answered more than 50% of the questions regarding financial literacy outnumbered the respondents who were not able to do so. *Trust,* the variable measuring the intensity of being trusting, is calculated in the same manner as Guiso, Sapienza & Zingales (2008), who employ a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted." *RT* is a risk tolerance measure ranging from 1 (lowest risk tolerance) to 5 (highest risk tolerance). *AbStkRetIg* (i.e., the absence of stock return ignorance) takes one if respondents can come up with answers when asked for their personal expectation of the average market return over the next 10 years in the US (where respondent answers form the variable *ExRet*) *and* an associated 80% confidence interval (if the

latter is feasible). Note that 70% of the respondents are stock return-ignorant based on this definition of *AbStkRetIg*, a definition slightly different from that of Merkoulova & Veld (2022), because those they characterize as being overoptimistic are also defined as stock return-ignorant, even though this behavior leads to *less* (not more) SMP.

Various appropriate control variables are also included in estimation. These include age (measured as *AgeYrs*), the other three Big 5 traits, namely *Disagreeable* (for Disagreeableness), *Openness* (for Openness-to-Experience) and *Extravert* (for Extraversion), and an indicator for gender (*Gender* = 1 if the respondent is male), another for marital status (*Married* = 1 if the respondent is married) and a third for birth country (*BornUS* = 1 if the respondent was born in the US). Finally, a series of indicators for labor status (six categories) and race (six categories) are used as controls.

4. Empirical Results

Table 1 provides a correlation matrix of *SMPart* and all potentially causative variables at our disposal. In the first column, we see that simple correlations of *SMPart* with potentially causative variables (i.e., the primitives *Conscientious*, *EmotStab*, *NCS2* and *CS*; and the proximates *Inc*, *Educ*, *FinLit*, *AbStkRetIg* and *Trust*) are in the expected direction and significant at 1%. The exceptions are *RT* and *ExRet*. As will be described, the former is measured with abundant noise, and the latter exhibits substantial non-linearity. Note as well that *Gender* and *AgeYrs* are also strongly positively correlated with *SMPart*.

(TABLE 1 HERE)

As for the relationship between the primitive and proximate variables, importantly, *Inc*, *Educ*,

FinLit, *AbStkRetIg* and *Trust* (hereafter the 'core proximate determinants') are *all* positively correlated with *both CS* and *NCS2*. These observations together point in the direction of much (if not all) of the impact of the primitives being subsumed by proximate determinants. The extent to which this may be true is empirically addressed below.

4.1. Proximate determinants of SMP

We begin exploring the stock market participation puzzle by focusing on proximate determinants, for now leaving aside primitive determinants. In Table 2, probit estimation of *SMPart* is conducted using the aforementioned seven proximate determinants as explanatory variables as well as a series of control variables. The displayed controls are *Gender* and *AgeYrs*. Abundant evidence documents higher risk tolerance among males (e.g., Falk *et al.*, 2018). Given previous evidence of rising participation throughout the life cycle with a decline as one approaches and moves into retirement (e.g., Fagereng, Gottlieb & Guiso, 2017), a quadratic for *AgeYrs*, where *Age^2* is its square, is investigated and found valid. Additional controls include a married (vs. not) indicator, a US-born (vs. foreign-born) indicator, and a series of labor market and race indicators.[13]

(TABLE 2 HERE)

Regressions R1 (with *Educ*) and R2 (with education represented by four indicators) exclude *ExRet* and *RT*. R3 and R4 introduce *RT*, with a resultant sample drop of slightly more than 1,000 because the risk tolerance question is taken from a later survey. R5 and R6 drop *RT* but now include *ExRet* (R5) or a third-order *ExRet* polynomial (R6) where *ExRet^2* and *ExRet^3* are squared and cubed *ExRet*. Notice when *ExRet* is included, the sample size drops by 670 observations. This is

[13] The coefficients and significance levels of these additional controls are not shown for brevity (as is also true for other tables); they are however available on request from the authors.

because those people who were unable to answer the question which asked what they believed the market's return was likely to be in the future are by necessity dropped. This further implies that the definition of *AbStkRetIg* is somewhat different depending on whether *ExRet* is included or not: when it is included, those who are expected return-ignorant disappear from the sample.

Regressions R1 and R2 show that, as expected, *Inc*, *Educ* (or its categorical indicators), *FinLit*, *AbStkRetIg* and *Trust* are all positively associated with *SMPart* at the 1% significance level (the exception being categorical education where high school completion is significant at only 5%). Moreover, *Gender* (recall *Gender* = 1 for males) and both coefficients for *AgeYrs* and *Age^2*, are significant at 5% or better in the expected direction (implying an inverted-U relationship). In R3, when *RT* is included as a continuous variable, it is insignificant. When a set of indicators for *RT* is employed (R4), several are significant, with the overall (anomalous) implication being that *SMPart* is more probable at medium levels of *RT* vs. extremely low levels (which is sensible) or very high levels (which is counterintuitive).[14] Viewing R5, consistent with Hurd, Van Rooij and Winter (2011), *ExRet*'s marginal impact is also positive and significant at 1%. Regression R6 however shows that the relationship is actually non-linear, a matter that is pursued in more detail in Ostad (2023). Given these complications as well as the necessitated sample contraction, for the rest of this section we drop *ExRet* and *RT* from consideration.

4.2. Primitive determinants of SMP

Table 3 explores to what extent primitive determinants have explanatory power for SMP absent proximate factors. The individual regressions differ depending on whether we look at the

[14] It could be that these counterintuitive *RT* results are due to 1) the variable being taken in the later (wave 2) UAS 295 survey; 2) the fact the latter survey was conducted during the COVID-19 pandemic; or 3) the large sample reduction involved when merging wave 1 and wave 2 participants.

primitives one by one or group them appropriately, and whether or not controls are included.

(TABLE 3 HERE)

Focusing on R7 from Panel A where *NCS2* and *CS* are regressors and controls are excluded, these two primitives are strongly associated with stock market participation. It can be shown that a one-standard-deviation increase in *CS* increases the likelihood of stock market participation by 15.1 percentage points, while a one-standard-deviation increase in *NCS2* leads to a more modest but still meaningful boost of 6.1 percentage points.[15] The impact of *CS* mirrors the results of Grinblatt, Keloharju & Linnainmaa (2011). The second result, that a measure of non-cognitive skills based on the Big 5 traits is explanatory, is novel. In sum, for now ignoring proximate factors, both cognitive and non-cognitive skills are important influences in inducing individuals to participate in equity markets.

4.3. Impact of primitives on proximates

Next, we consider the interaction between primitives and proximates. Table 4 shows OLS and probit regressions of the core proximate factors on primitive determinants. While *Inc*, *Educ*, *FinLit* and *Trust* are categorical, there are sufficient choices to justify a continuous variable approximation, so we use OLS for these. Alternatively, as *AbStkRetIg* is binary, we use probit for this variable. In Panels A (without controls) and B (with controls), *CS*, *EmotStab* and *Conscientious* are regressors; in Panels C (without controls) and D (with controls), we use *CS* and

[15] The constant (unshown) is -0.049; the average marginal effects of CS and *NCS2* are calculated as:

$$\frac{\Delta\text{Prob}(SMPart = 1)}{\Delta CS} = \varphi(-0.049 + 0.437 \times CS + 0.169 \times NCS2) \times 0.437 = 15.1\%$$

$$\frac{\Delta\text{Prob}(SMPart = 1)}{\Delta NCS2} = \varphi(-0.049 + 0.437 \times CS + 0.169 \times NCS2) \times 0.169 = 6.1\%$$

where φ is the probability density function of standard normal.

NCS2.

(TABLE 4 HERE)

Focusing on Panel C (with *NCS2* and *CS*, and absent controls), largely mirroring what we observed in the correlation matrix, we find that in all ten cases cognitive and non-cognitive skills are positively associated with the five core proximate determinants. In nine out of ten cases (with the one exception being *AbStkRetIg* on *NCS2*), statistical significance is at the 1% level or better. Additionally, in Table A.3 of the Empirical Appendix, we provide some non-parametric tests for the interaction between primitives and proximates. Here we merely note that our results are also supported by the non-parametric tests for these interactions. The clear conclusion is that primitive determinants impact *SMPart* indirectly via proximate determinants. The extent to which direct impact remains is next addressed.

4.4. Relative strength of indirect and direct impact of primitives

How much of the impact of cognitive and non-cognitive skills is indirectly occurring through the proximate determinants? Tables 5 and 6 offer insight. Referring to Panel A of Table 5 which examines *CS* vs. the core proximate variables, following Sias, Starks and Turtle (2020), we first regress *SMPart* and each of the core proximate factors on the control variables to remove the variation related to control variables.[16] The residuals from these regressions are extracted and denoted control-excluded *SMPart* and control-excluded proximates (where it is to be noted that control-excluded proximates are standardized to have unit variance and zero mean). Then we regress each of the control-excluded proximates on *CS* to partition them into the portion predicted

[16] These controls are *Gender*, *AgeYrs* (and its square), *Openness, Extravert, Disagreeable, BornUS, Married* and the labor status and race indicators.

by *CS* (i.e., the fitted value) and the portion attributed to other factors (i.e., the residuals). Again, fitted values and residuals are rescaled to unit variance and zero mean. Finally, we regress control-excluded *SMPart* on the fitted values of each proximate (the indirect impact of *CS* through control-excluded proximates) as well as the associated residuals (the direct impact of control-excluded proximates). Finally, to ascertain the indirect impact of *CS* on *SMPart* through proximates relative to the total variation, we take the ratio of the squared coefficient for the fitted value and divide it by the sum of the two coefficients squared.

For instance, referring to the first column of Panel A of Table 5 which examines the *CS* vs. *FinLit* pairing, the relative impact of *CS* on *SMPart* through *FinLit* is 31%, meaning that 31% of the relation between control-excluded *SMPart* and control-excluded *FinLit* arises from the portion of *FinLit* predicted by *CS* and 69% of the relation arises from the residuals (which reflects the direct impact of *FinLit*). The rest of Panel A examines *CS* vs. the other core proximates. The indirect impact of *CS* on *SMPart* ranges from 27% (through *Inc*) to 67% (through *Trust*). Panel B works in the same way for *NCS2* vs. the core proximates, and Panel C for both *NCS2* and *CS* together. According to Panel B, similar patterns hold for *NCS2*, with a reduced role for indirect impacts (vs. direct). Now the indirect impact of *NCS2* on *SMPart* ranges from 5.8% (through *FinLit*) to 24% (through *Trust*). From Panel C, it is apparent that when *CS* and *NCS2* are taken together, the indirect impact of cognitive and non-cognitive skills on *SMPart* ranges from 30% (through *Inc*) to 69% (through *Trust*). To summarize, these results demonstrate that a meaningful portion of the relation between stock market participation and each of the proximate determinants arises from the primitive factors. However, the other portion (the direct impact of the proximates not explained by primitives) is generally greater than 50%, suggesting that these factors may represent viable intervention points.

(TABLE 5 HERE)

Table 6 provides another perspective. Fairlie-Oaxaca-Blinder decompositions of the direct vs. indirect impact of *CS* and *NCS2* on *SMPart* are shown (Oaxaca, 1973; Blinder, 1973; and Fairlie, 1999). The analysis is based on interdecile comparisons. In the first column, we can calculate that the difference in *SMPart* between the lowest and highest *CS* deciles is 50.55%. The procedure attributes indirect impact via the five core proximates. We see that 11.1% is explained by *Inc*, 16.9% by *Educ*, 12.9% by *FinLit*, 8.5% by *AbStkRetIg* and 3% by *Trust*. Together these more than explain the entire impact of *CS* on *SMPart*. As for *NCS2*, we can calculate that the difference in *SMPart* between the lowest and highest *NCS2* deciles is 21.78%, which is about 40% of the *CS* interdecile spread but still substantial. Further, 7.2% is explained by *Inc* and 6.1% by *FinLit*; the other contributions are much more modest. Overall, 5.2% remains unexplained by the core proximates. In sum, much of the impact of both cognitive skills and non-cognitive skills on stock market participation is accounted for by proximate variables, with perhaps some direct role remaining for non-cognitive skills.

(TABLE 6 HERE)

4.5. Proximate and primitive determinants of SMP

Table 7 shows probit regressions of *SMPart* on *both* proximate and primitive determinants for the full sample. These regressions mirror those of Table 2, the principal difference being the present inclusion of all relevant determinants at our disposal.[17] For brevity, we focus on R7. Consistent with Table 2, all five core proximates continue to be highly significant positive factors

[17] Survey analyses such as ours can be criticized on the basis of omitted variables bias. When relevant variables are omitted, the coefficients and standard errors of included ones may be biased. That said, since our focus is on primitives (with a high claim to exogeneity, as discussed in section 5) we suggest that any omitted variable bias is unimportant.

for encouraging people to enter markets. *AgeYrs* once again appears significantly in quadratic form (with, again, its impact on market participation first rising and then falling). Figure A.1 in the Empirical Appendix (based on mean values for all other variables) shows the implications of the age coefficients for *SMPart* over the life cycle. Participation rises to about 60 years of age, after which it begins to decline as the average person enters or approaches retirement.

(TABLE 7 HERE)

Importantly, *CS* is now rendered insignificant. In other words, in these data, the entirety of the impact of *CS* on stock market participation is an indirect one operating through *Inc*, *Educ*, *FinLit*, *AbStkRetIg* and *Trust* (and potentially other controls). Given the evidence of Tables 5 and 6 this is perhaps not surprising. Nonetheless, some (albeit modest in terms of statistical significance) impact persists for *NCS2*. Non-cognitive skills, though *largely* operating indirectly through core proximate determinants, do retain some direct role as well.

In the Table 7 regressions the other Big 5 traits, *Disagreeable*, *Openness* and *Extravert*, are also included as explanatory variables. *Disagreeable* has a significantly positive impact on *SMPart* at 5% or better in all regressions. While this result may be sample-specific and illusory, it is also consistent with previous research. Specifically, risk taking requires resilience (Klein & Kunda, 1994), and being tough-mined, even disagreeable, can reflect such resilience (Nicholson, Soane, Fenton-O'Creevy & Willman, 2005; and West & Hall, 1997). In other areas of household finance, it pays to be disagreeable: Brown & Taylor (2014) find that agreeable individuals (and couples) held more debt. And Parise & Peijnenburg (2019) also find being agreeable was not helpful in the context of debt decisions.

Collectively, the independent variables improve the predictability of market participation. Since

48% of the sample participate, implying that non-participation would be the best guess without knowing any explanatory variable values, and 75% of the time the correct prediction is made conditioning on all explanatory variables (appearing in R7), errors are made about half as often (48% vs. 25%). R8, the last regression of Table 7, restores *ExRet* and *RT;* the results are unchanged (albeit with a smaller sample).

Comparing these results to those of Salamanca, De Grip, Fouarge & Montizaan (2020), who seek to show a residual impact for locus-of-control (which is their non-cognitive ability variable), there is a crucial difference between their findings and our Table 7. We show that in our data non-cognitive skills retain a direct influence after incorporating *all* proximate variables into the model. On the other hand, they do not provide comparable evidence, so we have no way of knowing whether locus-of-control retains direct influence or just operates indirectly through intermediaries.

5. Robustness and Extensions

In this section, we perform several robustness checks and extensions. First, we consider whether non-cognitive skills are predictive of SMP or merely associated with it. To avoid exogeneity concerns, it is crucial to consider the stability of non-cognitive skills. As previously mentioned, the evidence suggests a high degree of trait stability once adulthood is reached. Moreover, previous researchers in this area have performed stability and exogeneity tests, reporting reasonable stability. For example, Parise & Peijnenburg (2019) use instrumental variables, instrumenting Emotional Stability measured in adulthood on childhood trauma. Salamanca, De Grip, Fouarge & Montizaan (2020) perform various analyses using time-varying traits. Kuhnen & Melzer (2018) use averaged-over-time trait values. And Choi & Laschever (2018) exclude younger adults from their sample. In all these cases, results are robust.

Here we provide two kinds of evidence in support. First, most of those in the main sample repeated the personality questionnaire (which was conducted in three waves across UAS 1, UAS 121, and UAS 237). Those doing so exhibited high correlations for all the Big 5 traits. The stability assumption of personality traits is supported in our data by the high correlation in personality traits across three waves of surveys. More specifically, the average correlation of *Conscientious*/*EmotStab* scores across the three surveys was 72%/77%. Second, we show that the results are unchanged if younger adults with likely less stable personality traits are excluded from the sample. Support appears in Table A.4 of the Empirical Appendix, with results essentially unchanged when those younger than 25/45 are removed (in R1/R2).

A second issue is whether using a simple normalized average of *Conscientious* and *EmotStab* (i.e., *NCS2*) as a proxy for non-cognitive skills is appropriate. This proxy can be questioned on the basis of it being: 1) confounded with intelligence; 2) too inclusive; 3) insufficiently inclusive; and 4) mis-weighted. As for the first issue, note the significant correlation between *NCS2* and *CS* in Table 1. In R3 of Table A.4 we show that if we orthogonalize *NCS2* vs. *CS* and then use the residuals instead of raw *NCS* in estimation, the results are essentially unchanged. The second issue speaks to whether it is appropriate to combine *Conscientious* and *EmotStab* into a composite measure of non-cognitive skills. While we believe doing so is consistent with the direction of the literature (e.g., Parise & Peijnenburg, 2019), for those that prefer that the individual Big 5 traits be kept separate we note that R6 of Table 7 shows that *EmotStab* alone is weakly (at 10%) associated with *SMPart* when all proximate and control variables are included (while *Conscientious* alone is rendered insignificant).[18] As for the third issue, while recalling the discussion on Disagreeableness

[18] While we have focused mostly on a composite measure of non-cognitive skills, we routinely examine both Conscientiousness and Emotional Stability separately as well. For example, in Table 3 we find that both are positive and significant predictors of SMP when proximates are not included. Further, in Table 4 we see that both are significantly related to the proximate factors.

in the previous section, as an alternative to our favored proxy for non-cognitive skills, we define *NCS3* as a normalized simple average of *Conscientious, EmotStab* and *Disagreeable*. R4 and R5 from Table A.4 uses this broader proxy. While the impact of non-cognitive skills so-defined is little different, it does become significant at 1%. Finally, turning to the fourth issue, the potential mis-weighting of the components of non-cognitive skills, we address this through an alternative approach using partial least squares ('PLS') weighting. Using weights from PLS to form *NCSPLS*, as shown in R6 of Table A.4, the coefficient of this proxy for non-cognitive skills is virtually unchanged and retains comparable significance.

A third issue speaks to participation intensity. While participation is beneficial for most, participating at an appropriate level is also crucial. Our study has only focused on the former, primarily because the data at our disposal are weak when it comes to participation intensity. The closest we have is dollars invested (with interval categories). In Table A.5 of the Empirical Appendix, we repeat the estimation of Table 7 for dollars invested instead of *SMPart*, finding that when only people in the market are considered, both *NCS2* and *CS* are insignificant (or have a counter-intuitive sign). Notably, *Trust* is also insignificant, a possible interpretation being that the tendency to be trusting is needed for market entry but once someone is a participant it no longer has influence. The three knowledge determinants remain highly significant.

A final issue is whether the main results hold up in further UAS waves. The main analysis is based on the "first wave" of stock market participation and personality-trait data. More specifically, the first wave (UAS 184) was conducted between May 23rd and June 22nd of 2019. A second wave (UAS 295) was in the field from June 3rd to June 30th of 2020. And a third wave (UAS 387) took place from July 15th to August 15th of 2021. These SMP surveys were combined with the latest waves of personality trait surveys available in the UAS before the respective SMP

survey dates, denoted as UAS 121 and UAS 237. Table A.6 performs panel analysis for the first two waves (R1) and for all three waves (R2). These regressions are analogous to R7 from Table 7. It is apparent that the key results hold up. Specifically, *NCS2* remains a significant (at 5%) determinant of stock market participation.

6. Concluding Remarks

In this paper, we contribute to the stock market participation puzzle by investigating to what extent intelligence and non-cognitive skills, which we label as primitive factors, contribute to equity market participation. These primitive factors are likely to be fully developed and stable in adulthood. Thus, our contribution to the SMP puzzle literature is to examine how these primitive factors, which should remain relatively constant over an individual's adulthood, affect the decision whether to hold equities. Non-cognitive skills (as opposed to general personality) have been largely unexplored in the literature. We explore both the direct impact of non-cognitive skills (which we define as an equal-weighted combination of survey-based Conscientiousness and Emotional Stability scores) on SMP and the indirect impact through traditional explanatory factors such as education, income, financial literacy, task-specific financial knowledge, and trust (which we label as proximate factors). While these proximate factors vary over one's lifetime, we show that both variation due to primitive factors and other more time-varying factors (which are unobserved) can lead to variation in SMP.

To carry out our study, we utilize survey data from the *Understanding America Study* panel, which provides a large, generalized sample from the US population with high response rates. We use information on whether respondents hold stocks of any kind, whether directly or indirectly. Thus, our SMP measure is binary. In isolation, we find that the primitive factors affect SMP: higher

levels of both intelligence and non-cognitive skills lead to significantly higher SMP. Still, we find that most of the effects are indirect in that higher levels of the primitives lead to higher levels of the proximates, which then lead to higher levels of SMP. This is entirely true for intelligence and the Conscientiousness sub-component of non-cognitive skills, which we find to have no residual direct impact on SMP after controlling for the proximates. We do find, however, that a composite measure of non-cognitive skills retains some direct impact on SMP, even after controlling for the proximates. In our data then this direct impact seems to come more from Emotional Stability than Conscientiousness, but their combined direct impact (via the composite measure) is more impactful than either skill in isolation.

While others such as Brown & Taylor (2014), Conlin et al. (2015), Sias, Starks & Turtle (2020) and Salamanca, De Grip, Fouarge & Montizaan (2020) also find a relationship between personality and SMP (as discussed in section 2), we utilize a unique dataset of non-cognitive skills, proximate factors, and task-specific knowledge that has yet to be examined in the SMP puzzle literature. We also utilize the robust UAS panel data, which, to the best of our knowledge, has yet to be explored as it pertains to non-cognitive skills and SMP. Our novel result is that non-cognitive skills retain a direct impact on SMP even after controlling for the commonly-used proximates. This result suggests that, when comparing individuals with differing levels of non-cognitive skills, it is likely that these individuals will have differing levels of SMP, even if they have similar levels of income, education, financial literacy, trustfulness, and task-specific knowledge. Alternatively, there remains a direct impact of the proximates on SMP even after controlling for the primitives. This suggests that policy and financial literacy interventions can be effective. Our results suggest that irrespective of the level of intelligence or beneficial personality traits, efforts aimed at improving both financial literacy and understanding of the stock market itself can improve stock market

participation, thereby boosting future retirement savings through appropriate risk taking. Finally, our findings are consistent with the behavioral finance perspective shedding light on why some households completely avoid stock market participation despite the rational factors, grounded in traditional economic theory, advocating for stock holding.

There are a number of issues that future research could explore, especially with new surveys designed to address them. It would be useful to see how results are sensitive to the decomposition of equity holdings into direct and indirect (e.g., through retirement accounts). Higher-quality data would also be useful for careful examination of the determinants of participation intensity. Additionally, the robustness of our results could be investigated by utilizing different instruments to measure cognitive and non-cognitive skills. It would also be worthwhile for researchers to explore whether the results hold for different samples, such as those from other countries, both emerging and developed. While for the most part our study uses variables measured at roughly the same point in time, a long-term longitudinal study would also be of interest so that changes in SMP over an individual's life cycle could be studied. For example, do intelligence and beneficial personality traits affect the decision to enter and exit the equity market over a 10–year or even a 20-year period? The SMP literature has mostly focused on static measures of SMP, but the dynamics of SMP merit further exploration. This type of study may help lead to a well-defined standard model of SMP's explanatory variables, which at this point consists of a large and growing set of variables.

References

Addoum, J. M., G. Korniotis & A. Kumar (2017). Stature, obesity and portfolio choice. *Management Science,* 63: 3147-3529.

Ahern, K. R., R. Duchin & T. Shumway (2014). Peer effects in risk aversion and trust. *Review of Financial Studies,* 27: 3213-3240.

Alan, S. (2006). Entry costs and stock market participation over the life cycle. *Review of Economic Dynamics,* 9: 588-611.

Alattar, L., M. Messel & D. Rogofsky (2018). An introduction to the Understanding America Study internet panel. *Social Security Bulletin,* 78(2): 13-28.

Andersen, S. & K. M. Nielsen (2011). Participation constraints in the stock market: Evidence from unexpected inheritance due to sudden death. *Review of Financial Studies,* 24: 1667-1697.

Ang, A., G. Bekaert & J. Liu (2005). Why stocks may disappoint. *Journal of Financial Economics,* 76: 471-508.

Antoniou, C., R. D. F. Harris & R. Zhang (2015). Ambiguity aversion and stock market participation: An empirical analysis. *Journal of Banking & Finance,* 58: 57-70.

Attanasio, O. P. & M. Paiella (2011). Intertemporal consumption choices, transaction costs and limited participation to financial markets, reconciling data and theory. *Journal of Applied Econometrics,* 26: 322–343.

Barberis, N. & M. Huang (2007). The loss aversion/narrow framing approach to the equity premium puzzle. *Handbook of Investments: Equity Premium*. R. Mehra, Elsevier North-Holland.

Barberis, N., M. Huang & R. H. Thaler (2006). Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American Economic Review,* 96: 1069-1090.

Benartzi, S. & R. Thaler (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics,* 110: 73-92.

Benartzi, S. & R. H. Thaler (1999). Risk aversion or myopia? Choices in repeated gambles and retirement investments. *Management Science,* 45: 364-381.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural variables. *Journal of Human Resources,* 8: 436-455.

Bogan, V. (2008). Stock market participation and the Internet. *Journal of Financial and Quantitative Analysis,* 43: 191-212.

Briggs, J. S., D. Cesarini, E. Lindqvist & R. Östling (2021). Windfall gains and stock market participation. *Journal of Financial Economics,* 139: 57-83.

Brown, J., Z. Ivkovic, P. Smith & S. Weisbenner (2008). Neighbors matter: Causal community effects and stock market participation. *Journal of Finance,* 63: 1509–1531.

Brown, S. & K. Taylor (2014). Household finances and the 'Big Five' personality traits. *Journal of Economic Psychology,* 45: 197–212.

Choi, H.-S. & R. A. Laschever (2018). The credit card debt puzzle and noncognitive ability. *Review of*

Finance, 22: 2109–2137.

Christelis, D., T. Jappelli & M. Padula (2010). Cognitive abilities and portfolio choice. *European Economic Review,* 54: 18–38.

Christiansen, C., J. S. Joensen & J. Rangvid (2008). Are economists more likely to hold stocks?. *Review of Finance,* 12: 465–496.

Cloninger, C. R., D. M. Svrakic & T. R. Przybeck (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry,* 50: 975-990.

Cole, S., A. Paulson & G. K. Shastry (2014). Smart money? The Effect of education on financial outcomes. *Review of Financial Studies,* 27: 2022-2051.

Conlin, A., P. Kyröläinen, M. Kaakinen, M.-R. Järvelin, J. Perttunen & R. Svento (2015). Personality traits and stock market participation. *Journal of Empirical Finance,* 33: 34-50.

Dangl, T. & A. Weissensteiner (2020). Optimal portfolios under time-varying investment opportunities, parameter uncertainty, and ambiguity aversion. *Journal of Financial and Quantitative Analysis,* 55: 1163-1198.

Das, S., C. M. Kuhnen & S. Nagel (2020). Socioeconomic status and macroeconomic expectations. *Review of Financial Studies,* 33: 395–432.

Deaves, R. (2023). *Household Finance: An Introduction to Individual Financial Behavior,* Forthcoming, Oxford University Press.

Dimmock, S. G. & R. Kouwenberg (2010). Loss-aversion and household portfolio choice. *Journal of Empirical Finance,* 17: 441–459.

Dimmock, S. G., R. Kouwenberg, O. S. Mitchell & K. Peijnenburg (2016). Ambiguity aversion and household portfolio choice puzzles: Empirical evidence. *Journal of Financial Economics,* 119: 559-577.

Donnellan, M. & R. Lucas (2008). Age differences in the big five across the life span: Evidence from two natural samples. *Psychological Aging,* 23: 558-566.

Fagereng, A., C. Gottlieb & L. Guiso (2017). Asset market participation and portfolio choice over the life-cycle. *Journal of Finance,* 72: 705-750.

Fairlie, R. W. (1999). The absence of African American owned business: An analysis of the dynamics of self-employment. *Journal of Labor Economics,* 17: 80-108.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman & U. Sunde (2018). Global evidence on economic preferences. *Quarterly Journal of Economics,* 133: 1645–1692.

Georgarakos, D. & G. Pasini (2011). Trust, sociability, and stock market participation. *Review of Finance,* 15: 693–725.

Giannetti, M. & T. Y. Wang (2016). Corporate scandals and household stock market participation. *Journal of Finance,* 71: 2591-2636.

Grinblatt, M., M. Keloharju & J. Linnainmaa (2011). IQ and stock market participation. *Journal of Finance,* 66: 2121-2164.

Guiso, L. & T. Jappelli (2005). Awareness and stock market participation. *Review of Finance,* 9: 537–567.

Guiso, L., P. Sapienza & L. Zingales (2008). Trusting the stock market. *Journal of Finance,* 63: 2557-2601.

Haliassos, M. & C. C. Bertaut (1995). Why do so few hold stocks? *Economic Journal,* 105: 1110–1129.

Hattrup, K., M. S. O'Connell & J. R. Labrador (2005). Incremental validity of locus of control after controlling for cognitive ability and conscientiousness. *Journal of Business & Psychology,* 19: 461-481.

Heaton, J. & D. Lucas (2000). Portfolio choice and asset prices: The importance of entrepreneurial risk. *Journal of Finance,* 55: 1163-1198.

Hong, H., J. D. Kubik & J. Stein (2004). Social interaction and stock-market participation. *Journal of Finance,* 59: 137–163.

Hurd, M. D., M. Van Rooij & J. Winter (2011). Stock market expectations of Dutch households. *Journal of Applied Econometrics,* 26: 416–436.

Kaustia, M. & S. Knüpfer (2012). Peer performance and stock market entry. *Journal of Financial Economics,* 104: 321–338.

Kong, D., Y. Cheng & S. Liu (2021). Unexpected housing wealth appreciation and stock market participation. *Journal of Housing Economics,* 52: 101768.

Kuhnen, C. M. & B. T. Melzer (2018). Noncognitive abilities and financial delinquency: The role of self-efficacy in avoiding financial distress. *Journal of Finance,* 73: 2837-2869.

Li, G. (2014). Information sharing and stock market participation: Evidence from extended families. *Review of Economics and Statistics,* 96: 151-160.

Liang, P. & S. Guo (2015). Social interaction, Internet access and stock market participation—An empirical study in China. *Journal of Comparative Economics,* 43: 883-901.

Lusardi, A. & O. S. Mitchell (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature,* 52(1): 5-44.

Klein, W. M., & Kunda, Z. (1994). Exaggerated self-assessments and the preference for controllable risk. *Organizational Behavior and Human Decision Processes,* 59: 410-417.

Mather, N. & Jaffe, L.E. (2016). Woodcock–Johnson IV: Reports, Recommendations, and Strategies. Jossey-Bass, Hoboken, NJ.

McCrae, R. R. (2009). The Five-Factor-Model of personality traits: Consensus and controversy. *The Cambridge Handbook of Personality Psychology*. P. J. Corr and G. Matthews, Cambridge University Press.

Merkoulova, Y. & C. Veld (2022). Stock return ignorance. *Journal of Financial Economics,* 144(3): 864-884.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics,* 51: 247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory,* 3: 373–413.

Nicholson, N., E. Soane, M. Fenton-O'Creevy & P. Willman (2005). Personality and domain-specific risk taking. *Journal of Risk Research,* 8: 157-176.

Oaxaca, R. (1973). Male-female differentials in urban labor markets. *International Economic Review*, 14: 693-709.

Ostad, P. (2023). Market literacy and stock market participation. Working paper (Available at SSRN: http://dx.doi.org/10.2139/ssrn.4566280).

Paiella, M. (2007). The forgone gains of incomplete portfolios. *Review of Financial Studies*, 20: 1623–1646.

Parise, G. & K. Peijnenburg (2019). Noncognitive abilities and financial distress: Evidence from a representative household panel. *Review of Financial Studies*, 32: 3884–3919.

Peijnenburg, K. (2018). Life-cycle asset allocation with ambiguity aversion and learning. *Journal of Financial and Quantitative Analysis*, 53: 1963-1994.

Puri, M. & D. T. Robinson (2007). Optimism and economic choice. *Journal of Financial Economics*, 86: 71–99.

Rindermann, H. (2008). Relevance of education and intelligence for the political development of nations: Democracy, rule of law and political liberty. *Intelligence*, 36: 306-322.

Roberts, B. W. & W. F. DelVecchio (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 226: 3-25.

Rotter, J. B. (1990). Internal vs. external control of reinforcement: A case history of a variable. *American Psychologist*, 45: 489-493.

Salamanca, N., A. De Grip, D. Fouarge & R. Montizaan (2020). Locus of control and investment in risky assets. *Journal of Economic Behavior and Organization*, 177: 548-568.

Shum, P. & M. Faig (2006). What explains household stock holdings? *Journal of Banking and Finance*, 30: 2579-2597.

Sias, R., L. Starks & H. J. Turtle (2020). Molecular genetics, risk aversion, return perceptions, and stock market participation. *Working paper*.

Sias, R., L. Starks & H. J. Turtle (2023). The negativity bias and perceived return distributions: Evidence from a pandemic, *Journal of Financial Economics*, 147: 627-657.

Van Rooij, M., A. Lusardi & R. Alessie (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 101: 449-472.

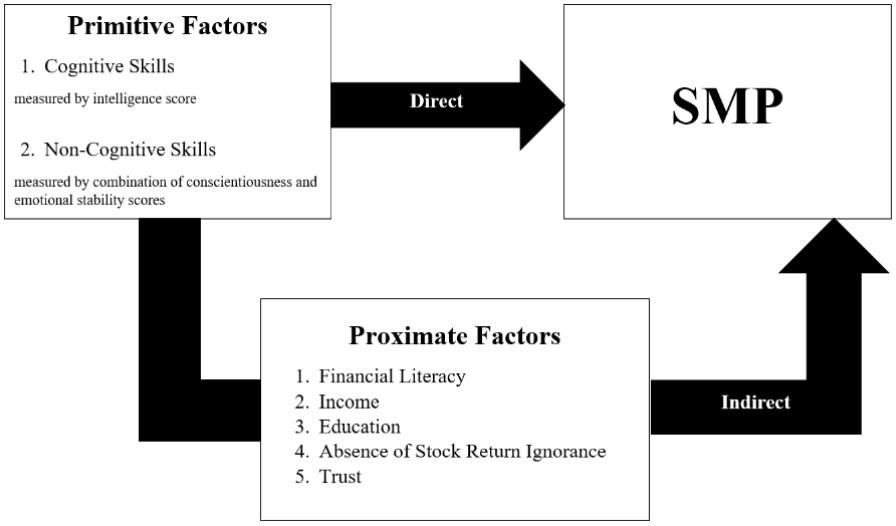
Vissing-Jørgensen, A. (2002). Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures. *Working paper*.

Vissing-Jørgensen, A. (2003). Perspectives on behavioral finance: Does irrationality disappear with wealth? Evidence from expectations and actions. *NBER Macroeconomics Annual*, 2003: 139–194.

West, R. & J. Hall (1997). The role of personality and attitudes in traffic accident risk. *Applied Psychology: An International Review*, 46: 253-264.

Tables and Graphs

Graph 1
The Direct and Indirect Impacts of Primitive Factors on Stock Market Participation



Graph 1 depicts that the impact of primitive factors on stock market participation can be direct or indirect through the proximate factors. We examine the extent to which these primitive factors affect stock market participation through well-observed proximate factors and whether a direct impact remains.

Table 1
Correlation Matrix

Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1. *SMPart*	1.000													
2. *Conscientious*	0.097***	1.000												
3. *EmotStab*	0.150***	0.422***	1.000											
4. *NCS2*	0.149***	0.811***	0.873***	1.000										
5. *CS*	0.323***	0.025	0.115***	0.088***	1.000									
6. *RT*	0.023	0.013	0.080***	0.060***	0.053**	1.000								
7. *Inc*	0.460***	0.147***	0.172***	0.190***	0.412***	-0.002	1.000							
8. *Educ*	0.332***	0.094***	0.126***	0.132***	0.384***	0.080***	0.404***	1.000						
9. *FinLit*	0.442***	0.128***	0.208***	0.203***	0.561***	0.094***	0.442***	0.388***	1.000					
10. *AbStkRetIg*	0.294***	0.019	0.057***	0.047***	0.351***	0.075***	0.255***	0.244***	0.356***	1.000				
11. *Trust*	0.203***	0.021	0.177***	0.126***	0.198***	0.016	0.208***	0.203***	0.225***	0.169***	1.000			
12. *ExRet*	-0.032	0.025	-0.005	0.010	-0.172***	-0.010	-0.139***	-0.068***	-0.127***	-0.073***	-0.019	1.000		
13. *Gender*	0.177***	-0.028	0.160***	0.088***	0.183***	0.076***	0.173***	0.072***	0.253***	0.160***	0.045**	-0.058***	1.000	
14. *AgeYrs*	0.106***	0.144***	0.199***	0.206***	0.108***	0.037*	0.009	0.027	0.265***	-0.009	0.091***	-0.074***	0.138***	1.000

This table presents correlations between variables. *SMPart* is equal to one if an individual is a stock market investor and zero otherwise. *Conscientious* and Neuroticism come from scores obtained from UAS 121; *EmotStab* is the negative of Neuroticism. *NCS2* is the average of *Conscientious* and *EmotStab*. *CS* is measured using number series, picture vocabulary and verbal analogies. *RT* is risk tolerance ranging from 1 to 5. *Inc* is measured in 16 categories from less than $5,000 to $150,000 or more. *Educ* is measured in five categories: not having a high school diploma takes zero; high school diploma or college below bachelor's takes 1; bachelor's degree takes 2; master's degree or professional school degree takes 3; and PhD takes 4 (note that not having a high school diploma is the omitted category). *FinLit* is the respondent's score on the financial literacy index, between 0 and 14. *AbStkRetIg* takes a value of one if respondents can answer the questions about expected return, high return (r_{90}) and low return (r_{10}) and consistently order the three returns ($r_{10}<r_{50}<r_{90}$). *Trust* is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted". *ExRet* is the participant's answer to the question about average annual return over the next 10 years. *Gender* is equal to one if the respondent is male. *AgeYrs* is age in years *** p<0.01, ** p<0.05, and * p<0.1.

Table 2
Proximate Determinants of Stock Market Participation

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**
ExRet					0.00880***	0.0443***
					(3.457)	(3.777)
ExRet^2						-0.00107***
						(-2.744)
ExRet^3						0.000007**
						(2.355)
AbStkRetIg	0.344***	0.348***	0.273***	0.290***	0.260***	0.238***
	(5.756)	(5.802)	(3.987)	(4.171)	(4.114)	(3.742)
Trust	0.0945***	0.0947***	0.107***	0.0955***	0.0960***	0.0951***
	(3.518)	(3.519)	(3.210)	(2.823)	(3.154)	(3.118)
Inc	0.427***	0.426***	0.490***	0.471***	0.483***	0.480***
	(11.64)	(11.59)	(10.57)	(10.01)	(11.48)	(11.40)
FinLit	0.328***	0.319***	0.311***	0.281***	0.324***	0.315***
	(9.741)	(9.376)	(7.090)	(6.258)	(8.098)	(7.853)
Gender	0.122**	0.124**	0.147**	0.152**	0.182***	0.186***
	(2.241)	(2.259)	(2.218)	(2.243)	(3.013)	(3.065)
AgeYrs	0.0474***	0.0481***	0.0318**	0.0366**	0.0425***	0.0432***
	(3.905)	(3.953)	(2.031)	(2.274)	(3.031)	(3.073)
Age^2	-0.000427***	-0.000433***	-0.000297**	-0.000333**	-0.000380***	-0.000385***
	(-3.639)	(-3.681)	(-1.976)	(-2.148)	(-2.823)	(-2.854)
Educ	0.161***		0.175***		0.145***	0.150***
	(5.364)		(4.805)		(4.397)	(4.529)
High school		0.389**		0.378*		
		(2.303)		(1.668)		
Bachelor's		0.646***		0.665***		
		(3.639)		(2.808)		
Master's		0.652***		0.599**		
		(3.544)		(2.461)		
PhD		1.141***		1.451***		
		(4.187)		(3.784)		
RT			-0.0341			
			(-1.098)			
RT low				0.506***		
				(3.397)		
RT moderate				0.528***		
				(3.850)		
RT high				0.724***		
				(5.206)		
RT very high				0.146		
				(1.087)		
Control Vars.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,134	3,134	2113	2113	2,463	2,463
Pseudo R2	0.272	0.274	0.267	0.291	0.260	0.263

This table presents the results of probit regressions. The dependent variable (*SMPart*) is equal to one if an individual is a stock market investor. The independent variables are defined as follows: *ExRet* is the participant's answer to the question about average annual return over the next 10 years. *ExRet^2* is the square of *ExRet. ExRet^3* is the cube of *ExRet. AbStkRetIg* takes a value of one if respondents can answer the questions about expected return, high return (r_{90}) and low return (r_{10}), and consistently order the three returns ($r_{10}<r_{50}<r_{90}$). *Trust* is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted." *Inc* is measured in 16 categories from less than $5,000 to $150,000 or more. *FinLit* is the respondent's score on the financial literacy index, between 0 and 14. *Gender* is equal to one if the respondent is male. *Educ* is measured in five categories: not having a high school diploma takes zero; high school diploma or college below bachelor's takes 1; bachelor's degree takes 2; master's degree or professional school degree takes 3; and PhD takes 4 (note that not having a high school diploma is the omitted category). *AgeYrs* is age in years. *Age^2* is the square of *AgeYrs*. *RT* is risk tolerance ranging from 1 to 5 (when dummy variables for *RT* are used very low risk tolerance is the omitted category). *Trust*, *Income*, *FinLit*, *Educ* and *RT* are standardized. Control variables not shown are *Married* (1 if married), *BornUS* (1 if born in US) and a series of indicators for race and labor status. The z-statistics are provided in parentheses; *** p<0.01, ** p<0.05, and * p<0.1.

Table 3
Primitive Determinants of Stock Market Participation

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
CS	0.445***					0.435***	0.437***
	(17.89)					(17.33)	(17.44)
Conscientious		0.123***		0.0525**		0.0708***	
		(5.440)		(2.108)		(2.755)	
EmotStab			0.191***	0.169***		0.128***	
			(8.368)	(6.724)		(4.930)	
NCS2					0.190***		0.169***
					(8.329)		(7.171)
Control Vars.	No	No	No	No	No	No	No
Observations	3,134	3,134	3,134	3,134	3,134	3,134	3,134
Pseudo-R2	0.0799	0.00686	0.0164	0.0174	0.0162	0.0921	0.0919

Panel B	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
CS	0.334***					0.325***	0.325***
	(11.64)					(11.25)	(11.30)
Conscientious		0.0931***		0.0626**		0.0701**	
		(3.469)		(2.251)		(2.495)	
EmotStab			0.145***	0.128***		0.0944***	
			(5.078)	(4.343)		(3.148)	
NCS2					0.161***		0.139***
					(5.447)		(4.631)
Control Vars.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,134	3,134	3,134	3,134	3,134	3,134	3,134
Pseudo R2	0.150	0.121	0.124	0.125	0.125	0.155	0.155

This table presents the results of probit regressions. The dependent variable is equal to one if an individual is a stock market investor. *CS* is measured using number series, picture vocabulary and verbal analogies. *Conscientious* and Neuroticism come from scores obtained from UAS 121. *EmotStab* is the opposite of Neuroticism. *NCS2* is the average of *Conscientious* and *EmotStab*. *CS, Conscientious*, *EmotStab* and *NCS2* are standardized. Control variables not shown are *AgeYrs*, *Age^2*, *Gender*, *Openness*, *Extravert*, *Disagreeable*, *Married* (1 if married), *BornUS* (1 if born in US) and a series of indicators for race and labor status. The z-statistics are provided in parentheses; *** p<0.01, ** p<0.05, and * p<0.1.

Table 4
Impact of Primitive Determinants on Proximate Determinants

	Panel A					**Panel B**				
	Panel A-1				**Panel A-2**	**Panel B-1**				**Panel B-2**
	FinLit	***Inc***	***Trust***	***Educ***	***AbStkRetIg***	***FinLit***	***Inc***	***Trust***	***Educ***	***AbStkRetIg***
CS	0.546***	0.400***	0.178***	0.376***	0.557***	0.458***	0.278***	0.157***	0.343***	0.507***
	(37.30)	(24.76)	(10.25)	(22.75)	(19.13)	(28.40)	(17.54)	(7.879)	(18.46)	(15.55)
Conscientious	0.0653***	0.102***	-0.0603***	0.0602***	0.0146	0.0512***	0.0741***	-0.0984***	0.0512***	0.0392
	(4.071)	(5.763)	(-3.167)	(3.326)	(0.531)	(3.165)	(4.656)	(-4.923)	(2.745)	(1.342)
EmotStab	0.117***	0.0831***	0.182***	0.0571***	0.0191	0.0587***	0.0302*	0.134***	0.0358*	0.0232
	(7.267)	(4.667)	(9.505)	(3.132)	(0.696)	(3.445)	(1.801)	(6.346)	(1.818)	(0.745)
Control Vars.	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	3,134	3,134	3,134	3,134	3,134	3,134	3,134	3,134	3,134	3,134
R-squared (Pseudo R2)	0.339	0.194	0.066	0.157	(0.1135)	0.408	0.427	0.096	0.213	(0.1338)

	Panel C					**Panel D**				
	Panel C-1				**Panel C-2**	**Panel D-1**				**Panel D-2**
	FinLit	***Inc***	***Trust***	***Educ***	***AbStkRetIg***	***FinLit***	***Inc***	***Trust***	***Educ***	***AbStkRetIg***
CS	0.547***	0.399***	0.188***	0.375***	0.557***	0.457***	0.276***	0.168***	0.342***	0.506***
	(37.52)	(24.74)	(10.77)	(22.79)	(19.21)	(28.48)	(17.42)	(8.382)	(18.46)	(15.57)
NCS2	0.155***	0.155***	0.109***	0.0987***	0.00282	0.0928***	0.0876***	0.0320	0.0732***	0.0527*
	(10.63)	(9.648)	(6.254)	(5.991)	(1.136)	(5.423)	(5.198)	(1.502)	(3.709)	(1.696)
Control Vars.	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	3,134	3,134	3,134	3,134	3,134	3,134	3,134	3,134	3,134	3,134
R-squared (Pseudo R2)	0.339	0.194	0.051	0.157	(0.1135)	0.408	0.426	0.081	0.213	(0.1337)

This table presents the results of OLS (A-1, B-1, C-1 and D-1) and probit (A-2, B-2, C-2 and D-2) regressions. The dependent variables are one of the proximate determinants of stock market participation. *FinLit* is the respondent's score on the financial literacy index, between 0 and 14. *Inc* is measured in 16 categories from less than $5,000 to $150,000 or more. *Trust* is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted." *Educ* is measured in five categories: not having a high school diploma takes zero; high school diploma or college below bachelor's takes 1; bachelor's degree takes 2; master's degree or professional school degree takes 3; and PhD takes 4. *AbStkRetIg* takes one if respondents can answer the questions about expected return, high return (r_{90}) and low return (r_{10}), and consistently order the three returns ($r_{10}<r_{50}<r_{90}$.) *CS* is measured using number series, picture vocabulary and verbal analogies. *Conscientious* and Neuroticism scores are obtained from UAS 121; *EmotStab* is the opposite of Neuroticism. *NCS2* is the average of *Conscientious* and *EmotStab*. *Trust, Inc, FinLit, Educ, CS, Conscientious, EmotStab* and *NCS2* are standardized. Control variables are *AgeYrs, Age^2, Gender, Openness, Extravert, Disagreeable, Married* (1 if married), *BornUS* (1 if born in US) and a series of indicators for race and labor status. The z-statistics are provided in parentheses; *** p<0.01, ** p<0.05, and * p<0.1.

Table 5
Relative Impact of Proximate Variables on *SMPart* through Primitive Variables

Panel A: ***CS***	***FinLit***	***Inc***	***Trust***	***Educ***	***AbStkRetIg***
Fitted value	0.086***	0.086***	0.086***	0.086***	0.086***
	(11.09)	(11.11)	(10.66)	(10.86)	(10.80)
Residuals	0.128***	0.141***	0.061***	0.106***	0.095***
	(17.83)	(18.27)	(7.50)	(13.41)	(11.98)
Relative Impact	**30.98%**	**27.01%**	**66.90%**	**39.62%**	**44.84%**
Panel B: ***NCS2***	***FinLit***	***Inc***	***Trust***	***Educ***	***AbStkRetIg***
Fitted value	0.039***	0.039***	0.039***	0.039***	0.039***
	(5.06)	(5.05)	(4.81)	(4.93)	(8.43)
Residuals	0.158***	0.156***	0.070***	0.123***	0.112***
	(20.37)	(20.10)	(8.56)	(15.49)	(14.00)
Relative Impact	**5.82%**	**5.94%**	**23.98%**	**9.22%**	**10.93%**
Panel C: ***CS & NCS2***	***FinLit***	***Inc***	***Trust***	***Educ***	***AbStkRetIg***
Fitted value	0.090***	0.091***	0.090***	0.090***	0.089***
	(11.62)	(11.81)	(11.11)	(11.39)	(11.16)
Residuals	0.136***	0.140***	0.060***	0.105***	0.095***
	(17.59)	(18.04)	(7.44)	(13.25)	(11.89)
Relative Impact	**30.41%**	**29.98%**	**69.03%**	**42.52%**	**46.81%**

Panel A reports coefficients on proximate variables from separate regressions of *SMPart* on the portion of proximates predicted by *CS* (Fitted value), and the portion that cannot be explained by *CS* (Residuals). *SMPart* and all proximate variables are first orthogonalized to remove the variation of control variables (*Openness, Extravert, Disagreeable, Gender, AgeYrs, Married, BornUS* and dummies for race and labor status). All independent variables are rescaled to zero mean and unit variance. **Panel B** repeats Panel A by substituting *NCS2* for *CS*. **Panel C** incorporates both *CS* and *NCS2*. The relative impact is the ratio of the squared coefficient for the fitted value divided by the sum of the two coefficients squared. The z-statistics are provided in parentheses; *** p<0.01, ** p<0.05, and * p<0.1.

Table 6
Fairlie-Blinder-Oaxaca Decomposition of the Effect of Primitive Variables on *SMPart*

	CS%	***NCS2 %***
AbStkRetIg	-8.5**	-0.84***
	(-2.31)	(-2.58)
Trust	-3.00*	-1.30
	(1.92)	(-1.57)
Inc	-11.06***	-7.22***
	(-2.76)	(-5.63)
FinLit	-12.89**	-6.09***
	(-2.18)	(-3.69)
Educ	-16.90***	-1.19
	(-3.43)	(0.75)
Bottom 10 participation rate	20.06	33.12
Top 10 participation rate	70.61	54.90
Explained difference in participation rates	-52.37	-16.61
Unexplained difference in participation rates	1.82	-5.16
Number of observations	627	597

This table reports on a Fairlie-Blinder-Oaxaca decomposition. This analysis measures how much of the difference in high and low *CS* and *NCS2* deciles of individuals' stock market participation rates can be explained by differences in proximate variables such as *AbStkRetIg*, *Trust*, *Educ*, *Inc*, and *FinLit*. *CS* is measured using number series, picture vocabulary and verbal analogies. *NCS2* is the average of *Conscientious* and *EmotStab*. *AbStkRetIg* takes one if respondents can answer the questions about expected return, high return (r_{90}) and low return (r_{10}) and consistently order the three returns ($r_{10}<r_{50}<r_{90}$). *Trust* is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted". *Inc* is measured in 16 categories from less than 5,000 to 150,000 or more. *FinLit* is the respondent's score on the financial literacy index, between 0 and 14. *Educ* is measured in five categories; where not having a high school diploma takes zero, high school diploma or college below bachelor's takes 1, bachelor's degree takes 2 master's degree or professional school degree takes 3 and PhD takes 4. *Trust, Inc, FinLit,* and *Educ* are standardized. The z-statistics are provided in parentheses; *** p<0.01, ** p<0.05, and * p<0.1.

Table 7
Proximate and Primitive Determinants of Stock Market Participation

	R1	R2	R3	R4	R5	R6	R7	R8
CS	-0.0475					-0.0490	-0.0483	-0.0234
								(-0.516)
Conscientious		0.0467		0.0342		0.0323		
		(1.590)		(1.128)		(1.063)		
EmotStab			0.0618**	0.0528*		0.0552*		
			(1.994)	(1.650)		(1.721)		
NCS2					0.0736**		0.0740**	0.0950**
					(2.280)		(2.292)	(2.355)
AbStkRetIg	0.358***	0.346***	0.345***	0.345***	0.345***	0.358***	0.358***	0.303***
	(5.883)	(5.748)	(5.745)	(5.745)	(5.745)	(5.886)	(5.883)	(4.257)
FinLit	0.348***	0.330***	0.327***	0.327***	0.327***	0.343***	0.343***	0.303***
	(9.662)	(9.729)	(9.616)	(9.604)	(9.616)	(9.517)	(9.521)	(6.381)
Inc	0.430***	0.420***	0.422***	0.419***	0.419***	0.425***	0.425***	0.482***
	(11.57)	(11.36)	(11.44)	(11.34)	(11.35)	(11.42)	(11.42)	(10.06)
Trust	0.103***	0.106***	0.0977***	0.101***	0.102***	0.102***	0.103***	0.0972***
	(3.800)	(3.874)	(3.591)	(3.693)	(3.755)	(3.735)	(3.806)	(2.837)
Educ	0.171***	0.164***	0.164***	0.164***	0.164***	0.170***	0.170***	0.157***
	(5.566)	(5.391)	(5.411)	(5.383)	(5.381)	(5.523)	(5.517)	(4.111)
Gender	0.105*	0.104*	0.0822	0.0858	0.0877	0.0861	0.0888	0.114
	(1.890)	(1.880)	(1.454)	(1.514)	(1.566)	(1.519)	(1.585)	(1.642)
AgeYrs	0.0508***	0.0478***	0.0488***	0.0477***	0.0476***	0.0491***	0.0489***	0.0359**
	(4.159)	(3.924)	(4.017)	(3.915)	(3.909)	(4.008)	(3.998)	(2.234)
Age^2	-0.00045***	-0.0004***	-0.00044***	-0.00043***	-0.00043***	-0.00044***	-0.00044***	-0.000324**
	(-3.844)	(-3.634)	(-3.729)	(-3.643)	(-3.637)	(-3.737)	(-3.727)	(-2.094)
ExRet								0.00956***
								(3.260)
RT = low								0.486***
								(3.252)
RT = moderate								0.527***
								(3.822)
RT = high								0.714***
								(5.114)
RT = very high								0.136
								(1.008)
Openness	-0.0237	-0.0321	-0.0327	-0.0340	-0.0341	-0.0279	-0.0280	0.0183
	(-0.837)	(-1.144)	(-1.166)	(-1.212)	(-1.213)	(-0.982)	(-0.986)	(0.524)
Extravert	0.0208	0.0190	0.0159	0.0120	0.0121	0.00620	0.00634	-0.0259
	(0.752)	(0.686)	(0.570)	(0.429)	(0.431)	(0.218)	(0.224)	(-0.748)
Disagreeable	0.0672**	0.0805***	0.0872***	0.0954***	0.0953***	0.0980***	0.0978***	0.0873**
	(2.440)	(2.759)	(2.939)	(3.122)	(3.119)	(3.199)	(3.194)	(2.257)
Control Vars.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,134	3,134	3,134	3,134	3,134	3,134	3,134	2,113
Pseudo-R2	0.274	0.274	0.275	0.275	0.275	0.276	0.276	0.293

This table presents the results of the probit regression. The dependent variable (*SMPart*) is equal to one if an individual is a stock market investor. The independent variables are defined as follows; *CS* is measured using number series, picture vocabulary and verbal analogies. *NCS2* is the average of *Conscientious* and *EmotStab*. *AbStkRetIg* takes one if respondents can answer the questions about expected return, high return (r_{90}) and low return (r_{10}) and consistently order the three returns ($r_{10}<r_{50}<r_{90}$). *FinLit* is the respondent's score on the financial literacy index, between 0 and 14. *Inc* is measured in 16 categories from less than 5,000 to 150,000 or more. *Trust* is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted". *Educ* is measured in five categories; where not having a high school diploma takes zero, high school diploma or college below bachelor's takes 1, bachelor's degree takes 2 master's degree or professional school degree takes 3 and PhD takes 4. *Gender* is equal to one if the respondent is male. *AgeYrs* is age in years. *Age^2* is the square of *AgeYrs*. *ExRet* is the participant's answer to the question about average annual return over the next 10 years. *RT* is risk tolerance ranging from 1 to 5 (very low risk tolerance is the omitted category). *Openness*, *Extravert*, *Conscientious*, Agreeableness and Neuroticism scores are obtained from UAS 121. *EmotStab* and *Disagreeable* are the opposite of Neuroticism and Agreeableness, respectively. Control variables not shown are *Married* (1 if married), *BornUS* (1 if born in US) and a series of indicators for race and labor status. *Trust, Inc, FinLit, Educ, CS, Openness, Extravert, Disagreeable, Conscientious, EmotStab, RT* and *NCS2* are standardized. The z-statistics are provided in parentheses; *** p<0.01, ** p<0.05, and * p<0.1.

Appendix

Table A.1
Variable Names, Sources and Definitions

Variable	Survey	Definition/Source/Calculation
Stock Market Participation (*SMPart*)	UAS 184	Var. "as_003a" *SMPart* = 1 indicates investment in equity, *SMPart* = 0, otherwise
Proximate Determinants of Stock Market Participation		
Income (*Inc*)	UAS 121	Var. "hhincome" Categorical variable for annual income ranging from 1 (less than $5,000) to 16 ($150,000 or more)
Education (*Educ*)	UAS 121	Var. "education" Categorical variable ranging from 0 to 4 where below high school=0, high school diploma or college below bachelor's =1, bachelor's degree=2, master's degree, or professional school degree =3, PhD=4
Financial Literacy (*FinLit*)	UAS 121	Var. "finlitscore" Scores ranging from 0 to 14
Absence of Stock Return Ignorance (*AbStkRetIg*)	UAS 184	Var. "as_004a, as_004b, as_004c" *AbStkRetIg* =1, if as_004a, as_004b, as_004c have been answered and consistently ordered (as_004b < as_004a < as_004c), *AbStkRetIg* =0, otherwise
Being Trusting *(Trust)*	UAS 184	Var. "as_005" It is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted".
Expected Return (*ExRet*)	UAS 184	Var. "as_004a" ranging from -9 to 100%, upper bound truncated at 100%
Risk Tolerance (*RT*)	UAS 295	Var. "as_006", ranging from 1 to 5 where *RT*=1 represents very low risk tolerance and *RT*=5 represents very high-risk tolerance
Primitive Determinants of Stock Market Participation		
Conscientiousness (*Conscientious*)	UAS 121	Var." conscientiousness" ranging from 9 to 45 with higher value corresponding to higher level of conscientiousness.
Emotional Stability (*EmotStab*)	UAS 121	Var." neuroticism" *-1 Ranging from -40 to 8 with higher value corresponding to higher level of Emotional Stability.
Non-Cognitive Skills (*NCS2*)	UAS 121	*NCS2*= (*Conscientious* + *EmotStab)*/2
Cognitive Skills (*CS*)	UAS 83, 84, 85	Measured by Intelligence scores*: CS*= Var. "uas83cog" (number series) + Var. "uas84cog" (picture vocabulary) + Var. "uas85cog**"** (verbal analogies)

Variable[19]	Survey	Definition/Source/Calculation
Control Variables		
Gender *(Gender)*	UAS 121	Var. "gender" Gender=1, if male, Gender =0 if female
Age (*AgeYrs*)	UAS 121	Var. "age"
Openness to experience *(Openness)*	UAS 121	Var." openness" ranging from 10 to 50 with higher value corresponding to higher level of openness.
Extraversion (*Extravert*)	UAS 121	Var." extraversion" ranging from 8 to 40 with higher value corresponding to higher level of extraversion.
Disagreeableness (*Disagreeable*)	UAS 121	Var." agreeableness" *-1 ranging from -45 to -9 with higher value corresponding to higher level of disagreeableness.
Marital Status (*Married*)	UAS 121	Var. "maritalstatus" *Married* = 1 if married, *Married* = 1 otherwise
Born in US (*BornUS*)	UAS 121	Var. "bornus" *BornUS* = 1 if the respondent was born in the US, *BornUS* = 0, otherwise
Race	UAS 121	Var. "race", ranging from 1 to 6
Labor Status	UAS 121	Var. "laborstatus", ranging from 1 to 6

[19] Our *SMPart* variable extracted from **UAS 184** was in the field from May 23 to June 22, 2019. It contains questions on the approximate total value of stocks in respondents' current financial portfolios, and questions on expected returns and their distribution. The size of selected sample is 3754 which addresses all active respondents who completed that provide scores for intelligence (UAS 83,84,85), and UAS 121 which contains financial literacy score. The response rate is 85.66%, which is in line with previous studies that have used the panel (Merkoulova & Veld, 2022a).
UAS 121 asks questions on personality traits and financial literacy. The survey took place between January 15, 2018, to August 25, 2020. The Bigfive section consist of 44 questions, such as, I am a person who…' and ask respondents to rate given attributes using a 5-ponit scale; 1 for Disaggree strongly, 2 for Disagree a little, 3 for Neither agree or disagree, 4 for Agree a little, and 5 for Agree strongly. these questions are grouped to generate scores for each of the five major personality traits.
UAS 83, **84**, and **85** measure respondent's intelligence (cognitive skills) scores. The tests are designed to measure the respondent's quantitative reasoning (number series- UAS 83) and lexical knowledge (picture vocabulary-UAS 84 and verbal analogies-UAS 85). Each measure consists of 15 items, which are scored dichotomously as correctly or incorrectly solved.

Table A.2
Sample Characteristics

Panel A: Distribution of Categorical Variables

	SMPart	*Married*	*AbStkRetIg*	*BornUS*	*Gender*	
Yes	1521	1885	953	2962	Female	1765
No	1613	1249	2181	172	Male	1369
Total	**3134**	**3134**	**3134**	**3134**		**3134**

Race		***Edu***		**Labor Status**	
White	2631	No high school diploma	146	Working	1739
African American	264	High School and some college or associate degree	1786	Retired	649
American Indian or Alaska	30			Unemployed or on leave	128
Asian	74	Bachelor's degree	692	Disabled	238
Hawaiian/ Pacific Islander	7	Master's degree or professional degree	439	Mixed	224
Mixed	128			Other	156
		Doctorate	71		
Total	**3134**	**Total**	**3134**	**Total**	**3134**

Inc						***RT*** [20]	
Less than 5000	84	20,000 to 24,999	137	60,000 to 74,999	348	1: very low	138
5000 to 7499	34	25,000 to 29,999	176	75,000 to 99,999	424	2: low	254
7500 to 9999	49	30,000 to 34,999	166	100,000 to 149,999	485	3: medium	503
10,000 to 12,499	80	35,000 to 39,000	131	150,000 or more	306	4: high	564
12,500 to 14,999	76	40,000 to 49,999	262			5: very high	654
15,000 to 19,999	108	50,000 to 59,999	268	**Total**	**3134**	**Total**	**2113**

[20] Risk tolerance (*RT*) is a self-reported measure based on respondents answer to the following question from UAS 295: How important is the following factor in determining the percentage of your investable assets that is currently invested in stocks? The possibility of even small losses on my stock investments makes me worry: 1 Not important at all; 2 A little important; 3 Moderately important; 4 Very important; 5 Extremely important. The answers ranges from Extremely important (very low *RT*) to Not important at all (very high *RT*).

Panel B: Summary Statistics

	N	mean	sd	min	max
Openness	3,134	35.53	6.252	14	50
Conscientious	3,134	36.12	5.443	11	45
Extrovert	3,134	25.73	6.263	8	40
Disagreeable	3,134	-35.84	5.380	-45	-9
EmotStab	3,134	-21.66	6.526	-40	-8
NCS2	3,134	7.23	5.05	-14.5	18.5
FinLit	3,134	9.015	3.001	0	14
CS	3,134	158.9	21.68	58.30	209.4
Trust	3,134	4.103	2.706	0	10
ExpRet	2,463	8.329	11.61	-9	100
AgeYrs	3,134	53.30	15.04	19	100

Figure A.1

Predicted Probability of Stock Market Participation versus Age

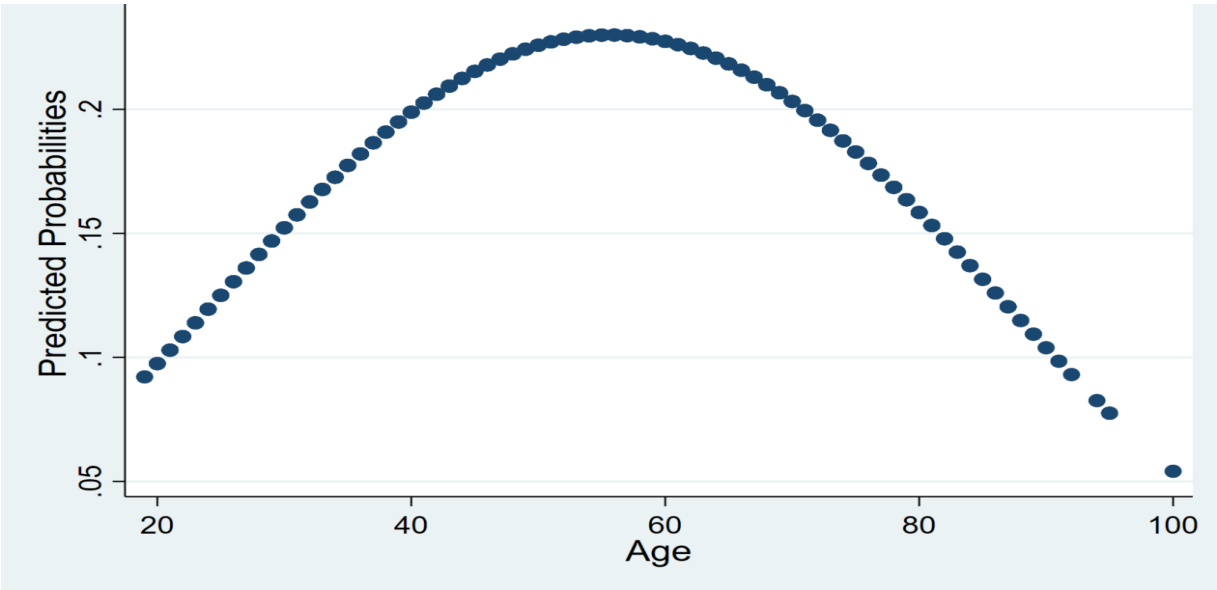


Table A.3
Non-Parametric Tests

Panel A: *CS*	***SMPart***	***FinLit***	***Inc***	***Trust***	***Educ***	***AbStkRetIg***
	Mean	Mean	Mean	Mean	Mean	Mean
Q1	0.25	6.57	8.86	3.22	1.09	0.09
Q2	0.44	8.63	11.38	3.93	1.38	0.22
Q3	0.56	9.91	12.28	4.58	1.65	0.39
Q4	0.69	10.94	12.85	4.68	1.96	0.52
Test Type	proportions	t-test	M-W	M-W	M-W	proportions
Q4-Q1	0.44***	4.37***	435***	240***	406***	0.43***
	(17.44)	(34.85)	(19.114)	(10.621)	(19.587)	(18.22)

Panel B: *NCS2*	***SMPart***	***FinLit***	***Inc***	***Trust***	***Educ***	***AbStkRetIg***
	Mean	Mean	Mean	Mean	Mean	Mean
Q1	0.34	7.85	9.86	3.53	1.27	0.24
Q2	0.49	8.98	11.44	4.07	1.56	0.32
Q3	0.55	9.55	11.96	4.31	1.60	0.35
Q4	0.55	9.56	11.94	4.45	1.61	0.30
Test Type	proportions	t-test	M-W	M-W	M-W	proportions
Q4-Q1	0.21***	1.71***	200***	141***	127***	0.06**
	(7.98)	(10.67)	(9.15)	(6.504)	(6.629)	(2.51)

Panel C: **Double Sorts:** Proximate Factor & *CS*	***FinLit***	***Inc***	***Trust***	***Educ***	***AbStkRetIg***
	SMPart	*SMPart*	*SMPart*	*SMPart*	*SMPart*
$Q1_{prox}Q1_{CS}$	0.10	0.13	0.14	0.16	0.17
$Q4_{prox}Q4_{CS}$	0.80	0.93	0.79	0.83	0.80
diff	0.70***	0.80***	0.65***	0.67***	0.63***
Z proportions test	(12.83)	(12.30)	(11.80)	(13.30)	(12.40)

Panel D: Double Sorts: Proximate Factor & *NCS2*	***FinLit***	***Inc***	***Trust***	***Educ***	***AbStkRetIg***
	SMPart	*SMPart*	*SMPart*	*SMPart*	*SMPart*
$Q1_{prox}Q1_{NCS2}$	0.13	0.12	0.17	0.24	0.25
$Q4_{prox}Q4_{NCS2}$	0.77	0.85	0.66	0.74	0.77
diff	0.64***	0.73***	0.49***	0.50***	0.52***
Z (proportions test)	(10.99)	(10.75)	(8.62)	(9.67)	(9.78)

Panel E: Double Sorts: Proximate Factors & Cognitive Skills (proportions test)

Financial Literacy		**$Q1_{FL}Q4_{CS}$**	**$Q2_{FL}Q4_{CS}$**	**$Q3_{FL}Q4_{CS}$**	**$Q4_{FL}Q4_{CS}$**
	SMPart	**0.20**	**0.42**	**0.68**	**0.80**
$Q1_{FL}Q1_{CS}$	**0.10**	0.10**	0.32***	0.58***	0.70***
		(2.46)	(6.74)	(11.10)	(12.83)
$Q2_{FL}Q1_{CS}$	**0.28**	-0.8*	0.14***	0.40***	0.52***
		(-1.65)	(3.17)	(8.39)	(10.42)
$Q3_{FL}Q1_{CS}$	**0.60**	-0.40***	-0.18***	0.08	0.20***
		(-7.74)	(-3.83)	(1.67)	(4.16)
$Q4_{FL}Q1_{CS}$	**0.65**	-0.44***	-0.23***	0.03	0.15***
		(-8.29)	(-4.58)	(0.62)	(3.16)

Income		$Q1_{Inc}Q4_{CS}$	$Q2_{Inc}Q4_{CS}$	$Q3_{Inc}Q4_{CS}$	$Q4_{Inc}Q4_{CS}$
	SMPart →	**0.22**	**0.51**	**0.73**	**0.93**
$Q1_{Inc}Q1_{CS}$	**0.13**	0.09**	0.38***	0.60***	0.80***
		(2.32)	(7.49)	(13.40)	(12.30)
$Q2_{Inc}Q1_{CS}$	**0.27**	-0.05	0.25***	0.46***	0.66***
		(-0.96)	(4.18)	(9.51)	(9.37)
$Q3_{Inc}Q1_{CS}$	**0.44**	-0.22***	0.07	0.29***	0.49***
		(-5.17)	(1.36)	(7.94)	(7.82)
$Q4_{Inc}Q1_{CS}$	**0.76**	-0.54***	-0.25***	-0.03	0.17***
		(-8.22)	(-3.61)	(-0.65)	(2.94)

Trust		$Q1_{Trust}Q4_{CS}$	$Q2_{Trust}Q4_{CS}$	$Q3_{Trust}Q4_{CS}$	$Q4_{Trust}Q4_{CS}$
	SMPart →	**0.48**	**0.65**	**0.71**	**0.79**
$Q1_{Trust}Q1_{CS}$	**0.14**	0.34***	0.51***	0.57***	0.65***
		(6.79)	(10.08)	(11.36)	(11.80)
$Q2_{Trust}Q1_{CS}$	**0.25**	0.23***	0.40***	0.46***	0.54***
		(4.81)	(8.47)	(9.89)	(10.39)
$Q3_{Trust}Q1_{CS}$	**0.33**	0.15***	0.32***	0.38***	0.46***
		(3.20)	(6.92)	(8.38)	(9.01)
$Q4_{Trust}Q1_{CS}$	**0.38**	0.10*	0.27***	0.33***	0.41***
		(1.88)	(5.25)	(6.61)	(7.41)

Education		$Q1_{Educ}Q4_{CS}$	$Q2_{Educ}Q4_{CS}$	$Q3_{Educ}Q4_{CS}$	$Q4_{Educ}Q4_{CS}$
	SMPart →	**0.48**	**0.56**	**0.66**	**0.83**
$Q1_{Educ}Q1_{CS}$	**0.16**	0.32***	0.40***	0.50***	0.67***
		(6.74)	(8.33)	(10.13)	(13.30)
$Q2_{Educ}Q1_{CS}$	**0.26**	0.22***	0.30***	0.40***	0.57***
		(4.41)	(6.07)	(7.79)	(11.32)
$Q3_{Educ}Q1_{CS}$	**0.32**	0.16***	0.24***	0.34***	0.51***
		(3.25)	(4.94)	(6.87)	(10.32)
$Q4_{Educ}Q1_{CS}$	**0.55**	-0.07	0.01	0.11**	0.28***
		(1.88)	(-1.36)	(0.36)	(2.38)

Panel F: Double Sorts: Proximate Factors & Non- Cognitive Skills (proportions test)

Financial Literacy		$Q1_{FL}Q4_{NCS2}$	$Q2_{FL}Q4_{NCS2}$	$Q3_{FL}Q4_{NCS2}$	$Q4_{FL}Q4_{NCS2}$
	SMPart →	**0.20**	**0.38**	**0.69**	**0.77**
$Q1_{FL}Q1_{NCS2}$	**0.13**	0.07	0.25***	0.56***	0.64***
		(1.56)	(4.89)	(9.94)	(10.99)
$Q2_{FL}Q1_{NCS2}$	**0.24**	-0.04	0.14***	0.45***	0.53***
		(-0.87)	(3.10)	(9.20)	(10.37)
$Q3_{FL}Q1_{NCS2}$	**0.61**	-0.41***	-0.23***	0.08*	0.16***
		(-7.73)	(-4.59)	(1.76)	(3.38)
$Q4_{FL}Q1_{NCS2}$	**0.67**	-0.47***	-0.29***	0.02	0.10**
		(-8.54)	(-5.59)	(0.41)	(2.04)

Income		$Q1_{Inc}Q4_{NCS2}$	$Q2_{Inc}Q4_{NCS2}$	$Q3_{Inc}Q4_{NCS2}$	$Q4_{Inc}Q4_{NCS2}$
	SMPart →	**0.19**	**0.42**	**0.68**	**0.85**
$Q1_{Inc}Q1_{NCS2}$	**0.12**	0.07*	0.30***	0.56***	0.73
		(1.85)	(6.00)	(11.82)	(10.75)
$Q2_{Inc}Q1_{NCS2}$	**0.29**	-0.10**	0.13**	0.39***	0.56***
		(-2.09)	(2.13)	(7.53)	(7.38)
$Q3_{Inc}Q1_{NCS2}$	**0.54**	-0.35***	-0.12**	0.14**	0.31***
		(-7.61)	(-2.43)	(3.70)	(4.70)
$Q4_{Inc}Q1_{NCS2}$	**0.75**	-0.56***	-0.33***	-0.07	0.10
		(-8.14)	(-4.49)	(-1.25)	(1.36)

Trust		**$Q1_{Trust}Q4_{NCS2}$**	**$Q2_{Trust}Q4_{NCS2}$**	**$Q3_{Trust}Q4_{NCS2}$**	**$Q4_{Trust}Q4_{NCS2}$**
	SMPart →	**0.38**	**0.50**	**0.59**	**0.66**
$Q1_{Trust}Q1_{NCS2}$	**0.17** ↓	0.21***	0.33***	0.42***	0.49***
		(4.20)	(6.48)	(8.06)	(8.62)
$Q2_{Trust}Q1_{NCS2}$	**0.36**	0.02	0.14***	0.23***	0.30***
		(0.603)	(3.02)	(4.84)	(5.66)
$Q3_{Trust}Q1_{NCS2}$	**0.41**	-0.03	0.09*	0.18***	0.25***
		(-0.52)	(1.93)	(3.73)	(4.67)
$Q4_{Trust}Q1_{NCS2}$	**0.49**	-0.11*	0.01	0.10*	0.17***
		(-1.78)	(0.35)	(1.96)	(3.01)

Education		**$Q1_{Educ}Q4_{NCS2}$**	**$Q2_{Educ}Q4_{NCS2}$**	**$Q3_{Educ}Q4_{NCS2}$**	**$Q4_{Educ}Q4_{NCS2}$**
	SMPart	**0.40**	**0.44**	**0.59**	**0.74**
$Q1_{Educ}Q1_{NCS2}$	**0.24**	0.16***	0.20***	0.35***	0.50***
		(3.42)	(4.05)	(6.92)	(9.67)
$Q2_{Educ}Q1_{NCS2}$	**0.28**	0.18**	0.16***	0.31***	0.46***
		(2.40)	(3.03)	(5.95)	(8.73)
$Q3_{Educ}Q1_{NCS2}$	**0.37**	0.03	0.07	0.59***	0.40***
		(0.57)	(1.20)	(4.17)	(7.02)
$Q4_{Educ}Q1_{NCS2}$	**0.64**	-0.24***	-0.20***	-0.05	0.10**
		(-4.53)	(-3.91)	(-0.84)	(2.12)

Panel A reports *cognitive skills* (CS) quartile analysis. *SMPart* shows the average SMP for each quartile of *CS* followed by a proportions test. Other columns show the average of proximate factors for each quartile. For categorical variables, a Mann-Whitney test is used to compare the mean rank of categories. Following the same logic, **Panel B** reports non-cognitive skills (*NCS2*) quartile analysis. **Panels C** and **D** represent the average of *SMPart* based on a double sort on each proximate factor in the first sort and *CS* (Panel C) and *NCS2* (Panel D) in the second sort. **Panels E** and **F** report a proportions test for the average of *SMPart* for each quartile of the double sorts on a given proximate factor in the first sort followed by *CS* (**Panel E**) and *NCS2* (**Panel F**) in the second sort. The z-statistics are provided in parentheses; *** p<0.01, ** p<0.05, and * p<0.1.

Table A.3 provides some non-parametric tests for the interaction between primitives and proximates. To start, we sort into quartiles based on *CS* (Panel A) and *NCS2* (Panel B) and report both average *SMPart* and the average for each of the proximate factors. We report a monotonically increasing relationship between *SMPart* and both *CS* and *NCS2*. Further, the difference between *SMPart* of the highest quartile and the lowest quartile of each primitive factor is significantly positive at the 1% level based on a proportions test. We also find significantly different levels of the proximate factors for the highest and lowest quartiles of the primitive factors in all cases.[21] As was suggested by correlations, the relationship between primitives and proximates is clearly

[21] With the exception of *AbStkRetIg* which requires a proportions test because it is binary, and *FinLit* where a t-test is used, for the other three core proximates Mann-Whitney U tests are conducted.

evident.

In Panels C and D of Table 5 we perform a double sort on each proximate factor in the first sort and *CS* (Panel C) and *NCS2* (Panel D) in the second sort, finding that differences in *SMPart* for the lowest quartile of each sort versus the highest quartile of each sort are significant and in the expected direction, in line with our earlier results. In Panel E we perform a proportions test for each quartile of the double sorts on a given proximate factor in the first sort followed by *CS* in the second sort.[22] For brevity, focusing on the diagonals of each proximate factor sort, we find that the *SMPart* of the lowest quartile of *CS* is significantly lower (at the 1% level based on a proportions test) vs. the highest quartile of *CS* in 15 out of 16 cases. Thus, once the proximate factors are controlled in the first sort, we find that higher *CS* results in significantly higher *SMPart* for almost all quartiles of all proximate factors. In Panel F we repeat the analysis in Panel E now using *NCS*2 in place of *CS*. While the results are not quite as strong as those in Panel E – significant differences are observed in 14 of 16 cases – the pattern is similar in that higher levels of *NCS2* typically result in significantly higher *SMPart* even after controlling for the proximate factors in the first sort. In sum, these non-parametric results show that *CS* and *NCS2* affect *SMPart* both directly and indirectly, typically in a statistically significant manner.

[22] Because *AbStkRetIg* is binary, we cannot analyze it in Panels E and F.

Table A4
Robustness Tests

	R1	**R2**	**R3**	**R4**	**R5**	**R6**
	***AgeYrs*>=25**	***AgeYrs* >=45**	***NCS2* orthogonalized**	***NCS3***	***NCS3* orthogonalized**	***NCS2PLS***
CS	-0.0380	-0.0639	-0.0418	-0.0438	-0.0343	-0.0491
	(-1.057)	(-1.437)	(-1.178)	(-1.241)	(-0.968)	(-1.387)
Non-cognitive Skills Proxy	**0.0795****	**0.0891****	**0.0737****	**0.0741*****	**0.0735*****	**0.0747****
	(2.439)	**(2.223)**	**(2.292)**	**(2.604)**	**(2.604)**	**(2.312)**
AbStkRetIg	0.351***	0.331***	0.358***	0.357***	0.357***	0.358***
	(5.704)	(4.393)	(5.883)	(5.861)	(5.861)	(5.886)
FinLit	0.344***	0.404***	0.343***	0.342***	0.342***	0.343***
	(9.427)	(8.434)	(9.521)	(9.500)	(9.500)	(9.517)
Inc	0.431***	0.490***	0.425***	0.425***	0.425***	0.426***
	(11.35)	(10.55)	(11.42)	(11.43)	(11.43)	(11.44)
Trust	0.0945***	0.0866***	0.103***	0.0967***	0.0967***	0.102***
	(3.440)	(2.637)	(3.806)	(3.587)	(3.587)	(3.758)
Educ	0.167***	0.186***	0.170***	0.170***	0.170***	0.170***
	(5.376)	(4.819)	(5.517)	(5.541)	(5.541)	(5.526)
Gender	0.0745	0.0581	0.0888	0.107*	0.107*	0.0856
	(1.314)	(0.840)	(1.585)	(1.935)	(1.935)	(1.524)
AgeYrs	0.0538***	0.00278	0.0489***	0.0473***	0.0473***	0.0492***
	(4.021)	(0.0769)	(3.998)	(3.882)	(3.882)	(4.021)
Age^2	-0.0004***	-0.00011	-0.000440***	-0.000431***	-0.000431***	-0.000442***
	(-3.794)	(-0.391)	(-3.727)	(-3.656)	(-3.656)	(-3.749)
Openness	-0.0283	-0.0179	-0.0280	-0.0357	-0.0357	-0.0279
	(-0.986)	(-0.510)	(-0.986)	(-1.266)	(-1.266)	(-0.981)
Extravert	0.0104	0.0136	0.00634	-0.00305	-0.00305	0.00623
	(0.362)	(0.385)	(0.224)	(-0.109)	(-0.109)	(0.220)
Disagreeable	0.0985***	0.125***	0.0978***			0.0979***
	(3.180)	(3.236)	(3.194)			(3.200)
Control Vars.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,060	2,100	3,134	3,134	3,134	3,134
Pseudo-R2	0.273	0.295	0.276	0.274	0.274	0.276

This table presents the results of probit regressions. In all regressions, the dependent variable is *SMPart* and is equal to one if an individual is a stock market investor. *CS* is measured using number series, picture vocabulary and verbal analogies. *NCS2* is the average of *Conscientious* and *EmotStab*. *NCS2*orthogonalized is the residuals of regressing *NCS2* on *CS*. *NCS3* is obtained by the average of *Conscientious*, *EmotStab* and *Disagreeable*. *NCS3*orthogonalized is the residuals of regressing *NCS3* on *CS. NCS2PLS* is the weighted average of *Conscientious* (40%) and *EmotStab* (60%). *AbStkRetIg* takes one if respondents can answer the questions about expected return, high return (r_{90}) and low return (r_{10}), and consistently order the three returns ($r_{10}<r_{50}<r_{90}$). *FinLit* is the respondent's score on the financial literacy index, between 0 and 14. *Inc* is measured in 16 categories from less than $5,000 to $150,000 or more. *Trust* is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted". *Educ* is measured in five categories; where not having a high school diploma takes zero, high school diploma or college below bachelor's takes 1, bachelor's degree takes 2 master's degree or professional school degree takes 3 and PhD takes 4. *Gender* is equal to one if the respondent is male. *AgeYrs* is age in years. *Age^2* is the square of *AgeYrs*. *Openness*, *Extravert*, *Conscientious*, Agreeableness and Neuroticism scores are obtained from UAS 121; *EmotStab* and *Disagreeable* are the opposite of Neuroticism and Agreeableness, respectively. *Trust, Inc, FinLit, Educ, CS, Openness, Extravert, Disagreeable* and all non-cognitive skills proxies are standardized. The control variables are *Married* (1 if married), *BornUS* (1 if born in US) and a series of indicators for race and labor status. The z-statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table A.5
Proximate and Primitive Determinants of Equity Investment

	SMPart	**Equity Investment**		**Equity Exposure**	
	R1	R2: Full Sample	R3: *SMPart*=1	R4: Full Sample	R5: *SMPart*=1
CS	-0.0483	-0.0434**	-0.0713*	-0.0514	-0.0684
	(-1.365)	(-2.128)	(-1.909)	(-1.291)	(-0.901)
NCS2	**0.0740****	**0.0317***	**0.0252**	**0.0440**	**0.00751**
	(2.292)	**(1.688)**	**(0.763)**	**(1.201)**	**(0.112)**
AbStkRetIg	0.358***	0.262***	0.221***	0.309***	0.260**
	(5.883)	(7.213)	(3.950)	(4.354)	(2.291)
FinLit	0.343***	0.144***	0.177***	0.290***	0.404***
	(9.521)	(6.995)	(4.353)	(7.200)	(4.888)
Inc	0.425***	0.220***	0.423***	-0.107***	-1.003***
	(11.42)	(10.52)	(9.445)	(-2.610)	(-11.04)
Trust	0.103***	0.0406**	0.0397	0.0651**	0.0192
	(3.806)	(2.552)	(1.406)	(2.096)	(0.336)
Educ	0.170***	0.161***	0.138***	0.158***	0.144**
	(5.517)	(9.018)	(4.996)	(4.532)	(2.561)
Gender	0.0888	0.160***	0.197***	0.235***	0.334***
	(1.585)	(4.847)	(3.556)	(3.649)	(2.962)
AgeYrs	0.0489***	0.0300***	0.0674***	0.0578***	0.112***
	(3.998)	(4.303)	(5.198)	(4.254)	(4.269)
Age^2	-0.000440***	-0.000222***	-0.000467***	-0.000474***	-0.000837***
	(-3.727)	(-3.268)	(-3.830)	(-3.582)	(-3.379)
Control Vars.	Yes	Yes	Yes	Yes	Yes
Observations	3,134	3,134	1,521	3,134	1,521
Pseudo R2	0.276	0.300	0.286	0.138	0.211

This table presents the results of the probit (column 1) and OLS (columns 2 to 5) regressions. In column 1, the dependent variable is *SMPart* which is equal to one if an individual is a stock market investor. In columns 2 and 3 the dependent variable is the dollar amount (in thousands of dollars) invested in equity and is measured as the midpoint of equity investment intervals. In columns 4 and 5, the dependent variable is a ratio of midpoint of equity investment intervals to midpoint of income intervals. The independent variables are defined as follows; *NCS2* is obtained by sum of *Conscientious* and *EmotStab*. *CS* is measured using number series, picture vocabulary and verbal analogies. *AbStkRetIg* takes one if respondents can answer the questions about expected return, high return (r_{90}) and low return (r_{10}) and consistently order the three returns ($r_{10}<r_{50}<r_{90}$). *FinLit* is the respondent's score on the financial literacy index, between 0 and 14. *Inc* is measured in 16 categories from less than 5,000 to 150,000 or more. *Trust* is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted". *Educ* is measured in five categories; where not having a high school diploma takes zero, high school diploma or college below bachelor's takes 1, bachelor's degree takes 2 master's degree or professional school degree takes 3 and PhD takes 4. *Gender* is equal to one if the respondent is male. *AgeYrs* is age in years. *Age^2* is the square of *AgeYrs*. Equity Investment, *Trust*, *Inc*, *FinLit*, *Educ*, *CS* and *NCS2* are standardized. Other control variables are *Openness*, *Extrovert*, *Disagreeable*, *Married* (1 if married), *BornUS* (1 if born in US) and a series of indicators for race and labor status. The z-statistics in parentheses and *** p<0.01, ** p<0.05, * p<0.1

Table A.6
Proximate and Primitive Determinants of Stock Market Participation: A Panel Analysis

	Panel With the First Two Waves	**Panel With All Three Waves**
CS	0.0218	0.0430
	(0.421)	(0.929)
NCS2	**0.107****	**0.0857****
	(2.204)	**(1.964)**
AbStkRetIg	0.529***	0.475***
	(6.607)	(7.026)
FinLit	0.504***	0.448***
	(9.617)	(9.932)
Inc	0.734***	0.675***
	(11.54)	(12.61)
Trust	0.201***	0.232***
	(4.521)	(5.542)
Educ	0.348***	0.370***
	(6.669)	(7.516)
Gender	0.234**	0.280***
	(2.563)	(3.243)
AgeYrs	0.0749***	0.0778***
	(3.695)	(4.074)
Age^2	-0.000658***	-0.000695***
	(-3.400)	(-3.839)
Openness	-0.0607	-0.0731*
	(-1.393)	(-1.846)
Extrovert	0.0311	0.0381
	(0.717)	(0.956)
Disagreeable	0.0911**	0.0632
	(2.031)	(1.562)
Controls	Yes	Yes
Observations	5,280	6,858

This table presents the results of the probit regressions. The dependent variable (*SMPart*) is equal to one if an individual is a stock market investor. The independent variables are defined as follows; *CS* is measured using number series, picture vocabulary and verbal analogies. *NCS2* is the average of *Conscientious* and *EmotStab*. *AbStkRetIg* takes one if respondents can answer the questions about expected return, high return (r_{90}) and low return (r_{10}) and consistently order the three returns ($r_{10}<r_{50}<r_{90}$). *FinLit* is the respondent's score on the financial literacy index, between 0 and 14. *Inc* is measured in 16 categories from less than 5,000 to 150,000 or more. *Trust* is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted". *Educ* is measured in five categories; where not having a high school diploma takes zero, high school diploma or college below bachelor's takes 1, bachelor's degree takes 2 master's degree or professional school degree takes 3 and PhD takes 4. *Gender* is equal to one if the respondent is male. *AgeYrs* is age in years. *Age^2* is the square of *AgeYrs*. *Openness*, *Extravert*, *Conscientious*, Agreeableness and Neuroticism scores are obtained from UAS 121. *EmotStab* and *Disagreeable* are the opposite of Neuroticism and Agreeableness, respectively. Control variables not shown are time, *Married* (1 if married), *BornUS* (1 if born in US) and a series of indicators for race and labor status. *Trust, Inc, FinLit, Educ, CS, Openness, Extravert, Disagreeable, Conscientious, EmotStab* and *NCS2* are standardized. The z-statistics are provided in parentheses; *** p<0.01, ** p<0.05, and * p<0.1.

Chapter 3

Market Literacy and Stock Market Participation

Contents

Market Literacy and Stock Market Participation

Parastoo Ostad

April 2024

Abstract: This paper employs data from the Understanding America Study (UAS) panel survey to examine whether a demonstration of knowledge about a reasonable distribution for market return increases stock market participation. My analysis reveals that the ability to formulate market return expectations aligned with the historical performance of the stock market and the current environment, denoted as "market literacy in return", significantly enhances the likelihood of stockholding. This effect remains robust after controlling for established factors such as general and task-specific financial knowledge, intelligence, and education. Interestingly, overoptimistic households, whose market return expectation exceeds the boundaries of market literacy, exhibit reduced participation. This phenomenon can be attributed to the perceived literacy cost, which outweighs the higher expected outcomes. In conclusion, literacy in return expectations plays a pivotal role in influencing stock market participation, warranting attention in policy and investor education initiatives.

Keywords: stock market participation; financial literacy; return distribution; overconfidence, overprecision, overoptimism

JEL Classification: D91, G11, D83

1. Introduction

The objective of this paper is to investigate whether specific market knowledge, as distinct from general financial knowledge, is incrementally predictive of Stock Market Participation, hereafter *SMP*. The earliest research on the *SMP* puzzle focuses on factors such as general education, income, and wealth, indicating that individuals with higher levels of these factors tend to participate more frequently in the stock market (Haliassos & Bertaut, 1995).[1] This perspective suggests that these attributes matter because individuals with sufficient resources and basic knowledge can more easily overcome the obstacles posed by fixed market participation costs (Vissing-Jørgensen, 2002 & 2003). Such costs, monetary and non-monetary, include the effort and any charges incurred setting up an investment account and various information costs, such as learning about investments, and can rationally deter those lacking means from investing in risky securities Other contributing factors, as previously documented, are trust (Guiso, Sapienza & Zingales, 2008) and intelligence (Grinblatt, Keloharju & Linnainmaa, 2011), with higher levels of trusting and cognitive skills correlating with a higher likelihood of investing in the stock market.[2]

Then researchers turn to task-specific knowledge rather than general expertise. Christiansen, Joensen & Rangvid (2008) show that economists are more likely to hold stocks. Van Rooij, Lusardi & Alessie (2011; hereafter 'VLA') identify that financial literacy, gauged by assessing subjects' comprehension of concepts such as interest rates, compounding, inflation, the stock market, and diversification, is a strong predictor of *SMP*.[3] More recently, Merkoulova & Veld (2022; hereafter 'MV') collect average market return forecasts over the next 10 years along with 80% confidence intervals for these predictions from participants. They find that the mere ability to generate plausible answers (which is the absence of what they call "stock return ignorance") is predictive of participation.

However, no study, to my knowledge, has delved beyond these investigations to explore whether a genuine understanding of a rational distribution for market returns—termed 'market literacy'—holds predictive power for *SMP*. Specifically, I consider whether the ability to come up with market return expectations that are in line with history and the current environment, referred

[1] For income/ wealth also see Shum & Faig (2006); for general education also see Cole, Paulson & Shastry, 2014; Vaarmets, Liivamagi, & Talpsepp, 2019).
[2] For cognitive skills, also see Christelis, Jappelli & Padula (2010); for trust, also see Georgarakos & Pasini (2011) and Giannetti & Wang (2016), and for non-cognitive skills see (Deaves, Ostad & Stivers, 2022).
[3] Their question set built on Lusardi & Mitchell (2014).

to as ‘market literacy in return’, is predictive of participation. I designate market return expectations falling below/above those harmonizing with market literacy in return as underoptimistic/overoptimistic.

Two important antecedents of this study, described in greater detail in section 2, are MV and Hurd, Van Rooij, and Winter (2011; hereafter ‘HVR’). While MV explicitly solicit market return expectations and confidence intervals, HVR indirectly elicit return forecasts and perceived volatility. They find that *SMP* rises with expected market return and falls with perceived volatility. This correlation is logically anticipated, as a suitably high return expectation is essential for overcoming both risk-free investments and participation costs. MV's findings demonstrate that individuals who are extremely overoptimistic (with a 10-year average expected return of 30% or more, herein referred to as 'superoptimistic') tend to invest more frequently in the market than their counterparts within the sample.[4] Notably, these two studies yield empirically consistent results: a positive linear relationship found by HVW is perfectly aligned with the MV’s observation that superoptimistic households exhibit higher *SMP* compared to the sample complement. However, I show that the situation is rather more complicated.

I conjecture that a demonstration of market knowledge in the domain of market return expectations is associated with higher *SMP*, paralleling findings for financial literacy and the absence of stock return ignorance. Therefore I anticipate observing SMP increases with expected returns within low to reasonable (market literacy) expected return ranges. However, the intricate interplay of diverse factors becomes evident when we move beyond the literacy range toward high expected returns. On the one hand, an overoptimistic belief in a high market return should induce more market entry. On the other hand, an overoptimistic view implies a lack of sufficient literacy regarding returns. Overoptimistic households should be aware of the limitations in their knowledge and the fact that becoming literate is costly. The question then arises: would the higher expected return justify the perceived cost of becoming literate? Therefore, it is an empirical question as to which force dominates.

Subsequently, I introduce the concept of 'market literacy in risk', encompassing the ability to provide a sensible market return confidence interval (or measure of volatility) that aligns with

[4] MV use the term “overoptimistic” for those with expected returns of 30% or higher. However, I have previously used this term for those whose expectations exceed the market-literate range, necessitating the need for a different term (“superoptimistic”).

market history and current conditions. I further conjecture that this demonstration of market knowledge should likewise offer predictive insight into *SMP*. As previously stated, HVW report that, as perceived volatility declines, increased *SMP* is observed. As in the case of expected return, this is consistent with the need to surpass the hurdle of participation costs. However, what occurs when perceived volatility reaches very low levels, such as when confidence intervals are notably narrow? Could this potentially impact *SMP* if these intervals are conspicuously too confined?[5] This occurrence, recognized as overprecision, introduces the possibility of influencing *SMP*. Evidence referenced in section 2 documents that high overprecision leads to more security trading. It is a plausible extension to suggest that overprecision might also be associated with more market entry. However, once more, there are two forces working at cross purposes. A correct confidence interval, rather than overprecision, is reflective of financial and market knowledge, and should also lead to higher *SMP*. Once again, it is an empirical question as to whether rational or behavioral forces dominate.

This paper utilizes data from the *Understanding America Study* (UAS) panel to address whether market literacy impacts stock market participation.[6] Additionally, it delves into whether the behavioral forces of overprecision and overoptimism impact market entry. In 2019, more than 3,000 UAS panel participants were asked to provide their subjective estimates of the average return on the US stock market over the next 10 years (*ExRet*) along with corresponding 80% confidence intervals (*Ret10* to *Ret90*). For instance, one participant provided figures of *ExRet*=12%, *Ret10*=2%, *Ret90*=15%; while another submitted *ExRet*=2%, *Ret10*=1%, *Ret90*=3%. To interpret, the subjective 10-year expected (average annual) return of the first (second) participant was 12% (2%), and she believes there was an 80% probability of the actual return falling between 2% (1%) and 15% (3%). These distinct viewpoints undoubtedly offer disparate perspectives. However, the pivotal question remains: What insights, if any, can be gleaned from these views regarding the propensity to invest in equities?

My main findings document a direct positive correlation between market literacy in return and *SMP*; a subject who displays greater awareness of what future returns are likely to resemble exhibits an increased likelihood of engaging in stock market participation. This effect is incremental to the impact coming from two related previously documented, task-specific

[5] As will be described in section 2, overprecision is a variant of the behavioral flaw of overconfidence.
[6] The UAS data are publicly available. My dataset is an augmented version of the MV dataset.

knowledge determinants, namely financial literacy, and the absence of stock return ignorance. Moreover, higher levels of *ExRet* in the territory of overoptimism, exceeding the domain of market literacy in return result in reduced *SMP*. This is contrary to MV and HVW which find a positive linear relationship between SMP and Expected returns. Regarding market literacy in risk and overprecision, those positioned within these ranges have higher *SMP* compared to people subject to underprecision. This suggests that overconfidence seems to be a factor in the *SMP* debate. However, it is important to note that the impact of market literacy in risk operates indirectly, as it is entirely absorbed by controlling factors.

I contribute to the literature in the following ways. First, I empirically show that greater knowledge about future returns of risky securities increases the probability of participating in the stock market, and its predictive power is incremental to those reflecting general and task-specific knowledge about capital markets. In addition, I explore the role of overconfidence in the *SMP* debate and find that an overoptimistic view regarding the expected return of risky assets does not necessarily induce stock holding. This is contrary to previous studies documenting that higher return expectations increase the probability of stock market entry. Moreover, I present a simple model to illustrate that high perceived literacy costs can prevent households from stock holding in the territory of overoptimistic belief in expected returns, meaning that rational force prevails over behavioral force. Finally, I introduce a new instrumental variable for market literacy to address the known issue of reverse causality between *SMP* and financial and market literacy.

In section 2, I provide the pertinent background on related *SMP* research and put overoptimism and overprecision in perspective. An illustrative representative agent model and the role of perceived literacy cost in deterring *SMP* are presented in section 3. In section 4 the survey and data are described. Section 5 details the hypotheses. The main empirical findings are presented in section 6. In the penultimate section, I cover extensions and robustness checks. The final section presents some discussion and concludes.

2. Background

2.1. Market return forecasts

As mentioned previously, HVW and MV investigate the relationship between a survey respondent's market return forecast and *SMP*. Salient differences between the two papers in their

methodologies and forecast time frames are worthy of note. HVW did not ask for return forecasts directly but rather used indirect elicitation, asking subjects for eight probabilities of staggered market gains and losses occurring over the next year. One example is: What is the probability that the market will earn 20% or better over the next 12 months? Based on the assumption of normality, they used nonlinear least squares to estimate a subject-specific expected market return (and volatility). MV, on the other hand, came right out and asked for expectations. About 20% of MV's sample were not able to arrive at any kind of answer, even when prompted. Comparing HVW and MV, the elicitation methodology of the former has the disadvantage of noisy estimation, while MV's methodology has the disadvantage of substantial sample erosion.

Another key difference between MV and HVW is the time interval of forecasts, with HVW's elicitation based on market evolution over the next 12 months, whereas MV asked for average returns over the next 10 years. While a rational forecaster might consider temporary market undervaluation leading to optimistic forecasts or overvaluation leading to pessimistic forecasts, a long-term projection should predominantly reflect the forecaster's belief in the market's overall performance, accounting for the balancing of ups and downs. This point will be further discussed, highlighting that a rational forecaster's long-term projection should align closely with the sum of the equity premium, anticipated inflation, and the real interest rate, as outlined in section 3. Consequently, when evaluating the potential impact of market literacy on *SMP*, a longer-term forecast interval akin to MV's approach becomes more suitable.

As previously indicated, the empirical conclusions drawn from both of these papers highlight a positive linear correlation between *SMP* (HVW) and the subset displaying superoptimistic outlooks, which holds even when compared to the remaining portion of the sample (MV). Nevertheless, neither study includes assessments of potential non-linear relationships, nor do they address whether those with market literacy exhibit higher *SMP* values than individuals classified as superoptimistic.

2.2. Overconfidence

Overconfidence constitutes a category of behavioral flaws. The standard taxonomy of overconfidence involves various biases in self-perception. This classification of overconfidence, as articulated by Moore & Healy (2008), comprises three main components: overplacement,

overestimation, and overprecision. I will elaborate on the two latter components below, leaving overplacement to readers as it is not relevant in this context.

2.2.1. Overoptimism

The second aspect of overconfidence refers to "the overestimation of one's true ability, performance, level of control, or chance of success". While the first part of this definition implies a degree of control and personal stake in the outcome, such as the performance of a personally managed portfolio, the latter part, "the overestimation of one's chance of success" suggests a reliance on luck rather than skills. When there is a personal stake involved (for instance, inheriting a portfolio of securities without having assembled it yourself), and luck plays a significant role, this form of overconfidence is frequently labeled as excessive optimism or overoptimism.[7]

Does this overoptimism align with my usage? Is it equivalent to what MV refers to as overoptimism, which I have labeled as superoptimism? The answer may not necessarily be so. As for my usage, I employ it in a well-specified context, specifically, to denote an expectation of returns that exceeds what is consistent with market literacy. As for MV, their respondents are asked to forecast the stock market over the next 10 years. However, how does the tendency to provide a high expected market return align with the prevailing interpretation of overoptimism in the overconfidence literature? In situations where a respondent is already invested in the market and anticipates a high market return this indeed represents the classic case of overoptimism (involving personal stake). However, this scenario is patently different from forecasting the market when an individual is not invested in equities and therefore lacks a personal stake. Importantly, a personal stake introduces the potential for wishful thinking or even a form of cognitive dissonance.[8] I will maintain awareness of this distinction as I proceed with interpreting the results below.

[7] For example, someone might believe that she has a 10% of dying of cancer during her life when the real probability *ceteris paribus* is (say) 20%. One also hears "the optimism bias" (e.g., Sharot, 2011) or mere "optimism" (e.g., Marshall et al, 1992) as opposed to pessimism (with the name of the interval between the optimism and pessimism being left vague). By using the term overoptimism one can call pessimism "underoptimism," with the intermediate territory called something like "properly balanced optimism."

[8] the tendency to ignore reality if it serves self interest.

2.2.2. Perceived volatility and overprecision

The next variant of overconfidence is overprecision, the belief that one's knowledge about some currently knowable fact or the accuracy of their prediction about a forthcoming event exceeds what is justified by reality.[9] This can be clarified through an example: I could present individuals with a series of questions, preferably in a multiple-choice format, and subsequently inquire how many questions they believe they answered correctly. If someone believes they answered around 75% correctly but in actuality only answered 30% accurately, this discrepancy illustrates overprecision.

Often confidence intervals are used to measure overprecision. For instance, if a 90% confidence interval is specified and 50 questions are asked, then a properly calibrated individual (i.e., someone with appropriately balanced precision or, equivalently, devoid of miscalibration) would likely answer about 45 questions correctly. As previously mentioned, MV, in line with Deaves, Lüders & Schröder (2010) and Ben-David, Graham & Harvey (2013), directly ask for 80% confidence intervals for the 10-year forward-looking average market return. In the case of these two related papers, many participants were subject to overprecision in the sense that x% confidence intervals contained eventual realizations less than x% of the time. Though a point of interest, this is not my concern here. In fact, because the MV survey was conducted in 2019, assessing forecast accuracy won't be feasible until 2029. Nonetheless, I can assess whether confidence intervals seem overly narrow (suggesting overprecision), excessively wide (indicating underprecision), or appropriately balanced (reflecting market literacy in risk) based on historical market performance.

Importantly, MV's survey is also different from these two aforementioned papers in that these two papers survey sophisticated parties (namely, market forecasters and CFOs), resulting in nearly everyone being capable of devising logical confidence intervals. However, more than half of MV's remaining sample were not able to provide a confidence interval that was even theoretically plausible.[10] This disparity raises questions: Are MV's confidence intervals subject-specific proxies for perceived volatility or do they reflect overprecision? Formally, there is no difference. Yet, for a group of respondents lacking sophistication, many of whom may possess a limited understanding of what a confidence interval is, the confidence interval frame might not be sufficient to induce

[9] The term "miscalibration" is sometimes used for overprecision, but technically speaking an underprecise person is miscalibrated as well, so the increasingly common usage over-/underprecision is preferable. Miscalibration however is useful in describing underprecision *or* overprecision.

[10] An example is a lower bound of 5% and an upper bound of 4%. MV show that those unable to come up with possible confidence intervals are less likely to invest in stocks.

them to truly think of outcome probabilities. Instead, they may be thinking merely how certain they are in their answer, perhaps not fully remembering that their answer essentially represents a guess of a mean around which random outcomes may be observed.

Suppose the overprecision interpretation holds for many cases. Several suboptimal financial behaviors exhibited by households, such as the excessive trading of securities (e.g., Barber & Odean, 2000; and Deaves, Luo & Lüders, 2009) have been attributed to overprecision.[11] In line with this body of evidence, it is reasonable to anticipate that this brand of overconfidence would lead to more *SMP*.

2.3. Financial literacy and market literacy

In my exploration of market literacy, I draw upon the research of VLA and MV. VAL study surveyed approximately 2,000 Dutch respondents asking a series of 16 questions, with five forming a "basic" test and 11 forming an "advanced" test of financial literacy. The first test asked simple questions assessing knowledge related to interest rates, compounding, and inflation, while the second asked questions centered around topics such as stocks, bonds, mutual funds, and diversification.[12] Respondents performed relatively well on the first test but often displayed weak knowledge on the second more advanced test. VLA used factor analysis to develop indices, both basic and advanced, of financial literacy. Subsequently, they document that the advanced index was highly predictive of stock market participation.

Recall MV asked their subjects for 10-year-forward-looking average market return predictions along with 80% confidence intervals. Their purpose, like VLA, was to investigate the impact on *SMP* of various measures of financial knowledge (two of which I will focus on), the absence of which they term "stock return ignorance." By inverting their indicators to literacy metrics, they find that both an absence of "expected return ignorance" (i.e., the mere ability to make any type of market forecast, even if it is inaccurate) and an absence of "risk ignorance" (i.e., the ability to

[11] Another example of suboptimal household investment behavior attributed to overconfidence is portfolio underdiversification (e.g., Goetzmann & Kumar, 2008; and Von Gaudecker, 2015).

[12] Here are two examples from the latter group:
Question 14: Stocks are normally riskier than bonds. True or false?
Question 15: Buying a company stock usually provides a safer return than a stock mutual fund. True or false?

generate a plausible confidence interval for future market returns) are associated with *SMP*.[13] In this paper, I elevate the level of market knowledge by a degree or two. I aim to explore whether possessing knowledge about both market return and risk is predictive of *SMP*.

3. An illustrative model

A prominent factor contributing to the prevalent preference for risk-free assets over stocks among households is the perceived cost associated with acquiring information. It might be believed that the cost of acquiring information necessary for equity investment is too high relative to the expected outcomes. Haliassos & Bertaut (1995) document that actual or perceived information is costly and restrains agents from stock market participation.

Entering the equity market and effectively managing risky portfolios are costly and time-consuming. Haliassos & Bertaut (1995) introduce the costs of stock market participation as expenses such as transaction costs, information costs, expenses of purchasing investment guides, investment magazines subscription fees, broker advice and the opportunity cost of time spent in tracking stock performance, internet surfing to obtain information, attending investment seminars to learn about investment strategies and so on. Obtaining knowledge about the stock market can reduce the cost of participation and managing risky assets, thereby fostering a desire for equity market entry. However, if the literacy cost is perceived as too expensive to justify the anticipated benefit, the preference often shifts towards investing in risk-free assets. In fact, Lusardi, Michaud & Mitchell (2017) explain that some individuals might reasonably choose not to invest in financial knowledge, as it can be costly to obtain, and not everyone may derive benefits from increased financial sophistication.

Households who expect a very high future return for stocks surpassing a reasonable expectation, essentially those who are overoptimistic, presumably would have a low level of stock market literacy. In the model below I explain why these individuals are more likely to demonstrate lower participation in equity even though the expected return on stocks is high.

I assume a static one-period model in which households are risk averse with Constant Absolute Risk Aversion (CARA) preferences. Each household maximizes its expected utility subject to its

[13] As will become clear, even to get into my sample subjects cannot be expected return-ignorant since my methodology requires expected returns. Then it will be seen that my sample divides by whether or not people are risk ignorant. They also slotted their "overoptimistic" group (as described earlier in the section) into the stock return-ignorant camp.

end of period wealth. In this model, only two financial assets are available; a risk-free asset and a risky asset. The risk-free rate and the risk premium are denoted as r_f and $\mathrm{r_p}$, respectively. Additionally, there is an entry cost associated with participating the stock market. If households choose to acquire more information, they will incur a literacy cost. The utility of final wealth is represented by:

$$U(w) = -e^{-\gamma w} \quad \text{(Eq.1)}$$

Where $\gamma > 0$ and w indicate risk aversion and the final wealth, respectively. Risk aversion is assumed to be the same for literate and non-literate households. The budget constraint is given as:

$$w_0 = F + S \quad \text{(Eq.2)}$$

Where w_0 indicates household's initial wealth and F and S represent investment in the risk-free asset and risky asset, respectively. Final wealth is fully consumed by each household, which is given by:

$$w = (w_0 - S)\big(1 + r_f\big) + S\big(1 + r_f + r_p\big) - a - bn \quad \text{(Eq.3)}$$

Where a is the cost of stock market entry, b is the literacy cost and n signifies the number of rounds of obtaining knowledge. $\mathrm{r_p}$ is random and assumed to be normally distributed with mean of μ and variance of σ^2, $r_p \sim N(\mu, \sigma^2)$. All households maximize their expected utility. Given the normality assumption and the CARA utility function the expected utility can be expressed as:

$$E(-e^{-\gamma w}) = -e^{-\gamma[E(w) - \frac{\gamma}{2}\sigma_w^2]} \quad \text{(Eq.4)}$$

A monotonic transformation of the equation above yields a simpler objective for households:

$$\max_S E(w) - \frac{\gamma}{2}\sigma_w^2 \quad \text{(Eq.5)}$$

Where:

$$w = w_0(1 + r_f) + Sr_p - a - bn$$

$$E(w) = w_0(1 + r_f) + S\mu - a - bn$$

$$\sigma_w^2 = -\frac{\gamma}{2}S^2\sigma^2$$

The true value of mean is unknown to households, necessitating the sampling of r_p to obtain an estimate beyond the prior. Let $X_1 \ldots, X_n$ be a random sample from a normal distribution with an unknown value of the mean and a specified value of the precision indicated by r ($r > 0$), which indicates the inverse of the variation from the true mean. With each sampling of the true mean, households gain more information and update their prior. Therefore, subsequent samples should provide more informative, higher precision estimates of the true mean.

Suppose that households' prior distribution of the true mean is a normal distribution with mean μ_0 and precision κ (κ > 0). Then the posterior distribution of μ is a normal distribution with mean μ_n and precision $\kappa + nr$ where: $\mu_n = \frac{\kappa\mu_0 + nr\overline{x}}{\kappa + nr}$ and $\frac{1}{\sigma^2} = \kappa + nr$. Here, $\overline{x}$ represents the sample mean which is an estimate of μ formed from the sample and nr is the precision of the conditional distribution of the sample mean for any given value of μ. Given this setup and from Eq.5, all households maximize their expected utility:

$$\max_S [w_0(1 + r_f) + S\mu_n - \frac{\gamma}{2}S^2\sigma^2 - a - bn] \quad \text{(Eq.6)}$$

Subject to $\mu_n = \frac{\kappa\mu_0 + nr\overline{x}}{\kappa + nr}$

$$\frac{1}{\sigma^2} = \kappa + nr$$

The first order condition of the expected utility with respect to S is:

$$S^* = \frac{\mu_n}{\gamma\sigma^2} \quad \text{(Eq.7)}$$

Assuming the prior distribution, $N(\mu_0, \frac{1}{\kappa})$, and considering no extra education, $n = 0$ $b = 0$, $\mu_n = \mu_0$ and $\frac{1}{\sigma^2} = \kappa$, Eq. 7 is rewritten and substituted into Eq.6 to obtain:

$$E(w) = w_0(1 + r_f) + \frac{1}{2\gamma}\mu_0{}^2\kappa - a \quad \text{(Eq.8)}$$

Three cases can be discussed:

1. If a household has an underoptimistic view of the risk premium, it means that their prior distribution of true mean is already too low to even justify the entry cost, or their expected stock return might be perceived as lower than the risk-free rate. Therefore, individuals may

neglect to seek out information or update their beliefs about the risk premium due to the lack of strong incentives, resulting in a reluctance to invest in stocks.

2. If a household is market literate, their expected risk premium is close to the true mean, indicating a high precision. If they believe that the risk premium exceeds the entry cost, $\frac{1}{2\gamma}{\mu_0}^2\kappa > a$, they are more likely to invest in the stock market.
3. Overoptimistic households face a unique dilemma. Their high priors regarding the risk premium provide a strong incentive to invest in stocks, as they perceive the potential returns to be substantial. However, this optimism is tempered by the low precision of their estimates, indicating a high degree of risk. While this combination does increase the incentive to learn more about the risk premium, the trade-off is less straightforward. Increased learning raises precision but at a literacy cost and also should lower the perceived risk premium, leading to a potential reduction in the overall incentive to invest in stocks. This nuanced interplay between optimism, risk perception, and the incentive to invest highlights the complex decision-making processes involved in household investment behavior.

   First, consider an overoptimistic household that chooses not to seek additional information about the stock market. In this case, their end-of-period wealth conditional on stock market participation will be represented by Eq.8. Then assume that the overoptimistic household chooses to obtain more knowledge about stock returns to raise precision by undergoing the first round of market literacy at the cost of b. Given $n = 1$, $\mu_1 = \frac{\kappa\mu_0 + r\bar{x}}{\kappa + r}$ and $\frac{1}{\sigma^2} = \kappa + \mathrm{r}$, Eq. 7 is rewritten and substituted into Eq.6 to obtain:

$$E(w) = w_0(1 + r_f) + \frac{1}{2\gamma}\frac{(\kappa\mu_0 + r\bar{x})^2}{\kappa + r} - a - b \qquad \text{(Eq.9)}$$

Eq.9 illustrates the household's end-of-period wealth with one round of education conditional on stock market participation. To simplify the comparison of choices for an overoptimistic household, I retain the relevant parameters from Eq.8 and Eq.9:

$$E^{OverOpt}(w^{n=0}) = \mu_0^2\kappa \qquad \text{(Eq.10)}$$

$$E^{OverOpt}(w^{n=1}) = \mu_1^2(\kappa + \mathrm{r}) - b$$

For overoptimistic households, the prior expectation of the risk premium is substantially higher than the true mean: $\mu_0 > \mu$. Thus, the posterior expectation should be lower than the prior after they have educated themselves by paying b ($\mu_1 < \mu_0$). As a result, their expected risk premium should decrease, albeit with higher precision ($\kappa + r > \kappa$). In fact, obtaining more information incurs a cost of literacy, and they may discover that the expected risk premium is actually lower than initially expected. They may choose to undergo more rounds of education to increase precision, where n=2,3,…, updating their posterior. Nevertheless, this implies higher costs and the realization of a lower expected risk premium. Consequently, overoptimistic households would often not invest in equity despite their optimism.

In conclusion, market literate households who have an expected premium close to the true mean are more likely to invest in equities. Conversely, both underoptimistic and overoptimistic households are less likely to invest in equities due to either low precision in their expectations, too low on expected return, or being too high literacy costs. Furthermore, precision is expected to be high for those who have an expected return close to the mean (market literate households), and low for those who are under/over optimistic (market illiterate households).

4. Survey questions and data

I utilize various surveys from the *Understanding America Study* (UAS) survey panel. A total of 3,134 participated in relevant surveys were pooled to obtain my dataset. UAS 184 contains questions regarding the distribution of market returns. The three survey questions under special scrutiny (along with the preamble that introduces them) are provided as follows:[14]

[Preamble] *Please answer the next questions based on your best guess on how the stock market will perform over the* ***next 10 years****. Here is an example of what I mean by average annual (yearly) return: It is a simple average of each year's return. So if in the first year, the*

[14] These come from UAS 184 (in the field during 2019). A question whether someone participates in the stock market also comes from UAS 184, as does a question on how trusting they are. Measures of intelligence, non-cognitive skills, financial literacy come from other UAS surveys. Intelligence is the sum of scores from tests of number series (UAS 83), picture vocabulary (UAS 84) and verbal analogies (UAS 85). Non-cognitive skills and other elements of the Big 5 personality traits and financial literacy come from UAS 121. I also use UAS 184 for data on income, education, and demographic information (such as gender, age, family status, place of birth, labor market status and race). Unique participant identifiers allow for the easy merging of survey modules.

stock market goes up by 2% and in the second year the market goes up by 4%, the average annual return over two years is 3% - I want your own best guess, so please do not look anything up.

1. *I expect the average annual return over the next 10 years will be*_____. It is labeled as *ExRet*.
2. *I believe that there is a small (1-in-10, or 10 percent) chance the actual return over 10 years will be less than:* _____. It is labeled as *Ret10.*
3. *I believe that there is a small (1-in-10 or 10 percent) chance the actual return over 10 years will be greater than*_____. It is labeled as *Ret90*.

4.1.Market-literate responses

Before describing the pattern of responses, it is essential to outline the expected answers that a market-literate respondent would likely offer in response to the aforementioned questions. A market-literate respondent is an individual who possesses a certain level of familiarity with both the historical performance of the stock market and the potential range of future returns.

First, in addressing the problem of coming up with a sensible expected return, one can focus on the return itself or decompose a nominal expected equity return as follows:

$$ExRet = Real\ interest\ rate + Exp.inflation + Equity\ premium$$

Various methodologies are available for estimating expected returns, including historical averages, forecasts provided by experts, or model-based techniques.[15] However, there is often a hesitation to rely solely on historical averages because the current circumstances may not closely mirror the past. For instance, the present economic conditions might significantly diverge from historical trends, such as a transition to a low-inflation environment following years of high inflation. Indeed, inflation has exhibited substantial variation over time.[16]

[15] See Damodaran (2019) for a review of various approaches.
[16] The 10-year average inflation rate varied from 2.94% to 1.62% over a 20-year period ending in 2018. This variability maps on to the *ex-post* one-year real rate of interest which over a 20-year period ending in 2018 ranged from -2.91% to 2.43%.

A roughly time-matched survey of close to 1,200 finance and economics professors, analysts, and managers of companies conducted by Fernandez, Martinez & Acin (2019) which solicited both the equity premium and the riskfree rate used to calculate the required return to equity, came up with averages of 5.6% and 2.7% respectively (summing to 8.3%).[17] On the other hand, a model-based approach that backs out an estimate of the equity premium from market prices produced a value of 5.96% in 2018 (Damodaran, 2019). To this latter figure must be added estimates for future inflation and the real interest rate. To arrive at the total expected return, one must account for projections of future inflation and the real interest rate. In 2018, the ex-post real interest rate was nearly zero, while inflation hovered around 2%, with a 10-year average of 1.80%. When combined, these figures imply a nominal risk-free interest rate of approximately 2%. In sum, utilizing this method yields an overall expected return of about 8%, which closely aligns with the findings of the earlier survey. Given this evidence, I use 8% as my best rough guess. Applying a somewhat arbitrary range of plus or minus 2%, a market-literate estimation for *ExRet* emerges within the 6-10% spectrum.[18]

Confidence interval width, *CIWidth*, is easily calculated as:

$$CIWidth = Ret90 - Ret10$$

Regarding the construction of a confidence interval from a market-literate perspective, considering the standard deviation of the annual S&P 500 return up to the year of 2018 over a 40-year period, which stood at 16.08%, serves as a starting point. Given an 80% confidence interval based on the assumption of a normal distribution, spanning 2.56 standard deviations, and considering the objective of assessing 10-year volatility (requiring division by the square root of 10), the following calculation is applied:

$$\frac{16.08\% \times 1.282 \times 2}{\sqrt{10}} = 13.02\%$$

Again, applying a somewhat arbitrary range of plus or minus 3% leads to a 10-16% market-

[17] Graham & Harvey have also periodically disseminated results from an ongoing survey of CFOs. Their last available report (2018), taken at the end of 2017 for the 10-year average S&P 500 return, produced a consensus of 6.79%.
[18] To be clear, this is not a confidence interval.

literate range for *CIWidth*.[19]

4.2. ExRet responses

I first focus on provided subjective expected returns. Beginning with the usable 3,253 respondents who participated in the survey soliciting their subjective expected returns along with corresponding confidence intervals, 657 were unable, or possibly, unwilling to answer question #1 even when prompted.[20] Due to the inability to evaluate either their degree of optimism or their level of precision, these individuals have been excluded from the sample, resulting in a remaining sample size of 2,477 subjects.[21]

(TABLE 1 HERE)

Table 1 displays relevant *ExRet* ranges including some integer ranges, along with the number of respondents and the average *SMP* for each range.[22] For now, the focus will be solely on the number of respondents within each range, postponing the discussion of *SMP* within each range till later. Among the respondents, there are 799 individuals categorized as market-literate in return; 1,331 people as underoptimistic; 217 as overoptimistic, and 130 as superoptimistic. The latter figure will be further clarified below.

Looking at the pattern of *ExRet* responses while considering question 1, at least four errors or biases are likely present in the *ExRet* data, and it is possible for an individual subject to be affected by more than one of these. Firstly, a few people are evidently providing *ExRet* in decimal form. This is the most plausible explanation for forecasted 10-year average returns such as 0.03, 0.05,

[19] At present I don't require symmetry or near-symmetry, but I will return to this issue in the empirical appendix.
[20] More precisely, 3,253 respondents participated in UAS 184, (41 of them started but didn't complete the survey; 3212 completed the survey). A further 119 respondents were lost in merging with other modules, ending up with 3,134 respondents. As for unwilling, see Bucher-Koenen, Lusardi & Van Rooij (2021), whose experiment shows that women tend to disproportionately respond "do not know" to questions measuring financial knowledge, but when this response option is unavailable, they often choose the correct answer. Note that 70% of respondents categorized as expected return-ignorant are women.
[21] The lowest *ExRet* was -50%; 13 people gave answers improbably over 100%. Interestingly, the four highest answers, all market participants, are 3,000%, 5,000%, and 1,000,000% (twice)."
[22] Ranges with no observations, of which there are many, are omitted.

0.08, and 0.10. Fortunately, this issue is confined to only 12 observations.[23]

Secondly, I conjecture (though not proven) that some people might be considering either the real (inflation-adjusted) market return or the equity premium (which is inherently real) in their responses. As discussed earlier, either of these considerations would generate an *ExRet* prediction of about 6% (rather than around 8%). If this figure is further rounded down by certain respondents to 5% (as discussed in the following paragraph), it might elucidate why *ExRet* = 5% emerges as the overwhelmingly preferred answer. The following two issues are more consequential and potentially actionable.

(FIGURE 1 HERE)

Thirdly, the influence of the round-number bias (as evidenced in studies such as Fraser-Mackenzie, Sung & Johnson, 2015) is quite substantial and likely to play a prominent role in shaping the responses. This bias pertains to people's inclination to provide round or focal point numbers instead of more subtle answers. This influence can be better understood by referring to Figure 1, which illustrates the frequency distribution of *ExRet* integer values.[24] The fact that 5% and 10% are round numbers but 6%, 7%, 8%, and 9% are not explains the spikes at these two focal points. Those strongly influenced by this bias and "thinking" 6-7% will perhaps input 5%, which potentially accounts for the popularity of 5%. Similarly, those "thinking" 8-9% might have entered 10%, explaining its fairly popularity as an answer.

Despite my lack of concrete proof regarding the presence or magnitude of this bias, I refrain from making arbitrary or disputable adjustments. However, it is crucial to bear in mind this bias as I proceed with the analysis.

Fourth, and most importantly, it is natural to be skeptical of a forecast of 30% or higher, which is triple the top of the market-literate range at 10%. Instead of attributing these responses to extreme overoptimism, it seems more plausible that many of these individuals might be demonstrating a form of commonplace carelessness. It is reasonable to assume that while reading, or perhaps skimming over, or even disregarding, the preamble, they might have overlooked the

[23] While I keep these participants in my main dataset (and did not adjust their data), one of my numerous robustness checks is to drop them. Not surprisingly, there is essentially no impact.
[24] A very high percentage (97%) of *ExRet* responses are integers.

clarification that the inquiry pertained to an annual average of returns over 10 years, not a 10-year cumulative return. This oversight could account for forecasts such as 100% or even higher (with at least 21 respondents providing such responses.[25] If this interpretation is correct, a forecast of 100% would be equivalent to a far more reasonable 10% per year. Moreover, it is worth noting that more respondents provided answers of 50% (33 individuals) than those who answered 9% (32 individuals). This phenomenon is probably a result of the previously described carelessness, combined with the round number bias. Respondents choosing 50% likely have the round number focal point of 5% per year in their minds, reflecting a broader belief range of 3-7%.

Going forward it seems sensible to center my primary estimation on a sample in which *ExRet* falls within the range of *ExRet* > -50% and *ExRet* < 30%. Doing so excludes one observation at the bottom and 130 at the high end. The respondents omitted from this sample are more likely to be careless rather than truly egregiously overoptimistic. I will however perform robustness checks where some or all of these observations are restored.

4.3. Ret10 and Ret90 responses

I now turn to the *Ret10* and *Ret90* responses. The majority of the 2,477 people who provided *ExRet* responses (1,537) are "risk-ignorant" as defined by MV. These individuals were not able to provide a plausible confidence interval,[26] which means that only their degree of optimism (but not precision) can be evaluated. However, for the remaining 940 subjects, I can assess both their levels of optimism and precision.

(<u>FIGURE 2 HERE</u>)

[25] There are eight respondents who gave an answer of *exactly* 100%.

[26] Respondents are not considered risk-ignorant if they fulfill all three of the following conditions: they can 1) come up with any expected return (i.e., they are not expected return-ignorant); 2) answer both questions regarding *Ret10* and *Ret90* and; 3) consistently order *Ret10, ExRet and Ret90* (*Ret10<ExRet<Ret90*). If respondents cannot fulfill all of these conditions, they are categorised as risk ignorant. Out of the 3,134 usable participants, 657 were expected return-ignorant, 56 could not come up with *Ret10* or *Ret90* (or both) and 1,468 could not consistently order the three figures, so I end up with 953 respondents. Therefore, the number of respondents cut by the MV risk-ignorant filter is 1,524 (i.e., 56 + 1468). I cut a further 13 with egregiously narrow (less than 1%) or egregiously wide (75% or more) confidence intervals. If I multiply *CIWidth* by $\frac{\sqrt{10}}{2.56}$ I get an estimate of subject-specific perceived volatility (Davidson & Cooper, 1976).

Referring to Figure 2, it becomes evident how the respondents distribute themselves across different bins of Confidence Interval Width, *CIWidth*. A substantial number of subjects provided too narrow, hence overprecise, confidence intervals and are thus present in the 1 to < 6 and 6 to <10 bins.[27] In contrast, there is a notably smaller representation within the 10 to 16 bin, corresponding to the market-literate range, and in the two higher range bins, reflecting underprecision. This pattern aligns well with abundant research that shows that most people are overconfident in the overprecision sense as evidenced by their confidence intervals being too narrow [28] (Soll & Klayman, 2004). For instance, notable research indicates that even individuals with heightened market awareness, such as CFOs, often offer 80% confidence levels that only encompass eventual market returns less than 40% of the time (as demonstrated by Ben-David, Graham & Harvey, 2013).

5. Hypotheses

In this section, I outline my principal hypotheses. I start with expected return (*ExRet*) and compare overoptimism, underoptimism, and market literacy concerning return. Similar to the empirical findings that those who are more financially literate are more likely to participate in the stock market, I conjecture that those who possess greater market literacy, especially in terms of return assessment will also *ceteris paribus* participate in the market more:

Hypothesis 1: Those with higher levels of market literacy concerning expected return participate in the stock market more.

Tentatively using 6-10% as the *ExRet* market-literate range, I would anticipate higher levels of *SMP* within this range, but progressively lower levels as one moves outside this range in either direction.

The second hypothesis relates to confidence interval width, and compares underprecision, overprecision, and market literacy concerning risk:

[27] The 1 to <6 range predominantly includes values such as 1, 2, 3, 4, and 5. Additionally, the 6 to <10 range has a notable concentration (mostly comprising values 6, 7, 8, and 9).
[28] This pattern is observed in novice as well as expert judgments (Clemen, 2001; Henrion & Fischhoff, 1986; Juslin, Winman & Hansson, 2007; McKenzie, Liersch, & Yaniv, 2008; Morgan & Keith, 2008).

Hypothesis 2: Those with higher levels of market literacy concerning risk participate in the stock market more.

Tentatively using 10-16% as the *CIWidth* market-literate range, I would anticipate higher levels of *SMP* within this range, but progressively lower levels as one moves outside this range in either direction. In both cases, one should picture an inverted V.

Now, I examine how market participation costs and behavioral forces might introduce complexity. As for the participation cost, it is insightful to begin with considering the utility score associated with an investment in stocks such as the market portfolio against that of a risk-free alternative. The conventional utility score formula takes the form:

$$(U^P) = E(r_p) - 0.5A\sigma^2$$
$$U^f = r_f$$

where U^P is the utility score of equity investment, $E(r_p)$ is the expected return of the market portfolio, A is risk aversion, σ^2 is market return variance, U^f is the utility score of risk-free investment and r_f is the risk-free rate. Focusing on perceived risk and return, while holding *A* and r_f constant, it becomes apparent that higher expected returns and lower perceived risk should increase the inclination for equity holding. Therefore, given the historical performance of the stock market and the absence of entry costs, all financial decision-makers regardless of perceived return, (which is provided to be greater than the risk-free rate), and perceived risk (and risk aversion) should participate in the market ($U^P > U^f$). Nevertheless, these factors could matter for some investors with constrained financial resources, characterized by lower wealth or income. For such investors, threshold levels of these variables are needed to justify making equity investments. thus, ignoring market literacy and participation costs, the expectation is to observe *SMP* rising with *ExRet* and falling with *CIWidth (*as a proxy of σ^2).

Next, I consider potential factors that might affect the influence of *ExRet* on *SMP*. One such factor is the irrational exuberance of overoptimism, which could become more significant, particularly within the realm of moderate to high *ExRet* values. Moreover, it is crucial to address

the counteracting behavioral force that might impact the connection between *CIWidth* and *SMP*, which is the phenomenon of overprecision. As one moves outside of the market-literate territory into the realm of overoptimism and overprecision, these behavioral forces could provide a full or partial offset to market literacy. It is crucial to note that, at low levels of *ExRet* and high levels of *CIWidth*, overconfidence and a given level of market literacy are not able to be differentiated as they point in the same direction. However, at high levels of *ExRet* (overoptimism) and low levels of *CIWidth* (overprecision), there is potential for differentiation between overconfidence and market literacy, as they point in opposing directions. Specifically, if I observe *SMP* increasing when moving further into the realm of overoptimism, the suggestion is that this form of overconfidence is dominating participation and literacy costs implying that the behavioral force prevails. However, if I observe *SMP* decreasing when moving further into the range of overoptimism this suggests that this form of overconfidence is being dominated by participation and literacy costs suggesting that the rational force prevails.[29] Similarly, if I observe *SMP* increasing moving further into the realm of overprecision, the suggestion is that this form of overconfidence is surmounting participation and literacy costs implying the dominance of a behavioral force. However, if I observe *SMP* decreasing moving further into the range of overprecision this suggests that this form of overconfidence is being overcome by participation and literacy costs suggesting the dominance of a rational force.

6. Empirical results

6.1. Market literacy and overoptimism: a first look

Before proceeding to a comprehensive multivariate analysis, I examine Figure 3, which displays the average *SMP* for all integer *ExRet* values with at least 25 respondents.[30] It is apparent that *SMP* reaches its highest levels within the 6-10% market-literate range, consistent with a knowledge-based rational story. Moving away from this interval in either direction, either towards ever-greater underoptimism or towards ever-greater overoptimism, leads to a reduction in *SMP*.

[29] Figure A1 in the Appendix shows these possibilities using heuristic graphs.
[30] From Table 1 it is clear that that there are few observations in each non-integer range so an *SMP* average would be quite noisy.

(FIGURE 3 HERE)

At first glance, consistent with Hypothesis 1, those who are market-literate in return are more likely to buy stocks. In line with the concept that overoptimism does not foster increased participation, when controls are absent, the decline in participation while transitioning into the overoptimistic range appears to be roughly equivalent to the decline observed during the transition into the underoptimistic range.

(TABLE 2 HERE)

In Table 2, I perform probit estimation for *SMP* on various *ExRet* ranges primarily focusing on a sample with -50% < *ExRet* < 30%. Thus, I mostly remove from consideration those who are superoptimistic.[31] Consequently, I mainly exclude those classified as superoptimistic. Panel A encompasses models without any control variables, while Panel B incorporates a comprehensive set of control variables. Beginning with Panel A, R1 has only market literacy in return (*MktLitRet*) as a regressor, while R2 has both *MktLitRet* and overoptimis (*OverOpt).* In the absence of control variables, *SMP* is higher for those who are market-literate in return compared to both the sample complement (R1) and those who are underoptimistic or overoptimistic separately (R2). Surprisingly contrary to expectations based on MV, the coefficient for *OverOpt* is negative and statistically indistinguishable from zero in the absence of control variables as reported in R2. For now, I defer a discussion of regressions R3 to R7. For future reference, note that R3 exclusively employs *ExRet* as a regressor, and R4 is a piecewise estimation over the three optimism ranges.

6.2. Market literacy and overoptimism: multivariate analysis

Referring to Table 3, I analyze the average values of the main control variables for individuals categorized as overoptimistic, underoptimistic, and market-literate in return, as well as group differences. I also present the values and differences for the superoptimistic group, however, for the time being, I postpone the discussion of this extreme segment of the overoptimistic category, concentrating instead on the larger, less extreme overoptimistic segment. A notable observation is

[31] Thus the 130 superoptimistic subjects were removed along with one supplying an expected return of less than -50%.

that households who exhibit market literacy in return possess the highest levels of intelligence, non-cognitive skills, financial literacy, income, trust, education, and absence of risk ignorance. These variables all positively predict *SMP*. Conversely, overoptimistic households tend to have the lowest levels of these driving factors within the three groups except for non-cognitive skills.

(TABLE 3 HERE)

Given these observed patterns, I can infer that a significant factor contributing to the lower *SMP* among overoptimistic people compared to those with market literacy is the influence of their control variable average levels on behavior. Therefore, it becomes necessary to conduct a multivariate analysis to disentangle these influences. This analysis is carried out in Panel B of Table 2. I am now revisiting the estimation from Panel A while incorporating as controls the driving variables that have been documented elsewhere as predictors of a positive *SMP*.[32]

R1 and R2 show that *MktLitRet* remains highly significant, although its coefficient experiences a slight reduction in magnitude. In R2, when controls are taken into account, the coefficient for *OverOpt* becomes positive but remains insignificant. Comparing these two estimations with their counterparts in Panel A, it is apparent that the control-adjusted coefficient on *MktLitRet* declines, while the corresponding coefficient on the *OverOpt* indicator rises and becomes positive. This narrowing gap is a result of those who demonstrate market literacy in return being more educated and trusting, having higher income and financial literacy, and being less prone to risk ignorance than the overoptimistic group.

R3 essentially replicates the findings of HVW: when *ExRet* is included linearly, it shows a highly significant positive association at the 1% significance level, both in the absence of controls (Panel A) and when controls are accounted for (Panel B). However, as revealed in R4, when a piecewise regression is performed, allowing for different slopes and starting points for underoptimistic, market-literate, and overoptimistic *ExRet* ranges, the assumption of linearity is clearly rejected. Indeed, the positive relationship between *SMP* and *ExRet* derive entirely from the

[32] Consistent with most previous research, income, education, trust, financial literacy, the absence of risk ignorance, non-cognitive skills, gender, and age are positively associated with *SMP*. Also, align with MV and Deaves, Ostad & Stivers (2022), the impact of intelligence is subsumed by other determinants.

underoptimistic segment as indicated by R4.[33]

6.3. Superoptimism

R5 and R6 in Panels A and B include *all ExRet* observations inclusive of the superoptimistic group. In these regressions, I revisit the findings of MV that the egregiously overoptimistic, whom I have labeled as superoptimistic, have higher *SMP*. The impact of control variables is significant here. When control variables are not considered, this pattern is evidently not observed. However, I find that the inclusion of controls alters the scenario completely. Specifically, R5 in Panel B essentially reproduces MV's finding that, with the incorporation of control variables, the superoptimistics invest more than the sample complement. Yet, as I infer from R6, this does not diminish the importance of market literacy in return. Individuals literate in this context continue to display a greater propensity for equity investment compared to the sample complement, except for those classified as egregiously overoptimistic. It is noteworthy that the disparity in SMP between individuals who are literate in return and the superoptimistic group cannot be statistically differentiated from zero.

Now the question is what distinguishes superoptimistic groups. The evidence points in two directions. First, as earlier argued, with *ExRet* levels at 30% or more it becomes challenging to distinguish respondents who are truly egregiously overoptimistic from those who are merely careless. A clue in this regard comes from Table 3. Being careless is likely to be correlated with lower intelligence among other factors. Indeed, even in comparison to those who are more moderately overoptimistic (who also face challenges in this aspect), the superoptimistic group exhibits significantly lower intelligence, limited financial literacy, lower income levels, and an increased tendency toward risk ignorance. It is the presence of these deficiencies in control variables that adds an element of surprise to any level of participation from the superoptimistic group.

One way to address this issue is to examine those who are still substantially overoptimistic but not egregiously. This can be achieved by focusing on individuals with *ExRet* values in the range of $20 \leq ExRet < 30$. At the low end of this range (20%), an investment worth $100,000 grows to

[33] Figure A2 in the Empirical Appendix plots the fitted *SMP* value versus *ExRet.* Both the decreased boost (for no controls vs. controls) in the market-literate range and the increased boost in the overoptimistic range (for no controls vs. controls) are visible.

about $619,200 in a span of ten years, which is an overoptimistic projection; at the high end, the figure climbs to $1,378,600, which is even more bullish. In my assessment, there is a greater probability of genuine overoptimism within this range rather than mere carelessness. Referring to R7 in both Panels A and B of Table 2, where $20 \leq ExRet < 30$ forms the new upper level of optimism, as observations with $ExRet \geq 30$ observations are omitted, it is apparent that the quite overoptimistic group participates less than the underoptimistic and market-literate in return who still participate in the stock market more than all other groups with or without incorporating control variables.

Another distinguishing aspect can be attributed to at least a portion of the $ExRet \geq 30$ group. A considerable number of individuals within this category (29 out of 130) provide responses ranging from 60% to 100%, an unquestionably overly optimistic range. However, when these values are divided by 10, these respondents conveniently fall within the market-literate in return range. The average level of participation for this subgroup stands at 55%, compared to 35% for individuals with $30\% \leq ExRet < 60\%$. This disparity provides a clear indication that what sets this subgroup apart is their simultaneous blend of confusion and market literacy in return.

To summarize, it is reasonable to conclude that individuals who exhibit market literacy in return are more likely to participate in equity markets, aligning with Hypothesis 1. Those who demonstrate overoptimism participate slightly more than those categorized as underoptimistic but still less than individuals with market literacy (when control variables are taken into account). Regarding the superoptimistic group, after incorporating control variables, their participation is on par with the market-literate group and exceeds that of the remaining sample. It appears to be a peculiar finding based on their lower control variable scores and the likelihood that a significant number of them are primarily confused rather than genuinely overoptimistic.

6.4. Market literacy in risk and overprecision: a first look

This part shifts focus to the interplay between overprecision and market literacy. Recalling that 10-16% is the market-literate range for *CIWidth*, with narrower intervals implying overprecision and broader ones suggesting underprecision, Figure 2 displays that *SMP* peaks at mild level of overprecision. At first glance, it seems that while task-specific knowledge is one force encouraging participation, overconfidence via overprecision may be a contributing factor as well.

Table 4 explores the potential impact of market literacy in risk and overprecision on *SMP*.[34] Panel A is presented without any control variables, while Panel B integrates control variables. Beginning with Panel A, market literacy in risk (R1) and overprecision (R2) taken individually have no explanatory power, however, when examined jointly (R3) they both are predictive of *SMP* with statistical significance at the 10% level. This observation is in line with Figure 3 illustrating that these two groups tend to participate more compared to the underprecise group. R4 is a simple probit estimation with *CIWidth* as a regressor. Not surprisingly, considering R3, the coefficient is negative. R5 presents piecewise estimation coefficients. Although R4 displays a negative linear relationship between *SMP* and perceived risk, which is consistent with findings from previous studies, I observe a deviation from linearity upon introducing the interaction between CIWidth and overprecision (R5). To elaborate further, as the width of the confidence interval expands (i.e., as the level of perceived risk rises), the overprecise group exhibits a higher tendency to participate in the stock market compared to the underprecise group, which contradicts previous findings.

(TABLE 4 HERE)

6.5. Multivariate analysis of market literacy in risk and overprecision

Table 5, analogous to Table 3, displays average values of the control variables for the market-literate, underprecise, and overprecise groups. Importantly, the underprecise group exhibits lower levels of intelligence, non-cognitive skills, financial literacy, income, education, and trust.[35] Therefore, it should not be surprising that when the Panel A regressions are repeated in Panel B while incorporating all control variables, the coefficients for market literacy in risk and overprecision decline and become indistinguishable from zero[36]. The interpretation here is that the impacts of market literacy in risk and overprecision on *SMP* are subsumed by the presence of other drivers of *SMP*.[36]

(TABLE 5 HERE)

[34] Note in these regressions I keep all superoptimistic respondents. However, most of them are risk-ignorant (all but 22 of the 130). Consequently, there are a few superoptimistic subjects included. For robustness I repeat these estimations exclusive of superoptimistic group. The results remain essentially unchanged.
[35] This is consistent with Bhandari & Deaves (2006) who find that those with higher education are more overconfident.
[36] Figure A3 in the Empirical Appendix plots the fitted *SMP* value versus *CIwidth*.

6.6. Both forms of market literacy, precision and optimism

In this section, I consider both varieties of market literacy simultaneously. I begin the analysis with double sorts, as shown in Table 6. Each matrix cell shows the average *SMP* for the relevant optimism/precision combination with the number of observations in parentheses beside it. Additionally, numbers surrounding the average *SMP* show t-statistics and significance of differences relative to *SMP* values in adjacent cells. This encompasses cells situated above, below, to the left, to the right, or diagonally in any direction.

Notice that the average *SMP* for households who are market-literate in both expected return and risk stands at 0.83, which is the highest among all cells containing 20 or more observations. This value is significantly higher than the *OverOpt&MktLitRisk*, *UnderOpt&MktLitRisk*, *UnderOpt&OverPre*, and *OverOpt&UnderPre* combinations. The significance is absent only in instances where the bins have fewer than 20 observations or when one of the conditions is *MktLitRet*. Also, note that the lowest average *SMP* is for those who are simultaneously underprecise and overoptimistic. This is consistent with findings from Figure 2 and Figure 3.

(TABLE 6 HERE)

While this double-sorting approach is intriguing, It is important to reiterate the same caveats stated earlier concerning controls. R1 to R4 in Table 7 present probit estimation with indicators for both forms of market literacy, along with overprecision and overoptimism. For the time being, I defer the discussion of R5 to R7 to section 7. Panel A is executed without control variables while Panel B includes a comprehensive set of control variables. Directing attention to R4 in Panel A, which excludes control variables and employs all these indicators, it becomes apparent that *MktLitRet*'s coefficient is positive and statistically significant at 1%. This finding, in fact, serves as a robustness check, as the sample size is considerably reduced compared to Table 2. However, the coefficients for the other three indicators, *MktLitRisk*, *OverPre*, and *OverOpt* are statistically indistinguishable from zero. As expected, in Panel B which accounts for control variables, the coefficient and level of significance of *MktLitRet* diminishes albeit that statistical significance at the 5% level persists.

(TABLE 7 HERE)

In summary, the evidence strongly suggests that market literacy in return serves as a predictive factor for *SMP*. This holds particularly when controls are not introduced, and its significance diminishes to some extent when accounting for all available factors. A notable caveat is that individuals labeled as superoptimistic, contrasted with those who are moderately optimistic, exhibit SMP levels similar to those who are market-literate in return. However, this similarity appears to arise from confusion stemming from the questions rather than genuine overoptimism. Conversely, the evidence concerning market literacy in risk is notably weaker, with any potential impact being overshadowed by other influencing factors.

7. Extensions, robustness and causality

In this section, I begin by extending my analysis by broadening the characterization of market literacy. Specifically, I investigate the impact of a potential third marker of market literacy, namely distributional symmetry. Then I consider whether my principal findings concerning market literacy in return are robust to changes in the necessarily arbitrary range utilized previously. After this, I broach the important issue of causality. Finally, I compare the impact on *SMP* of the three forms of financial/market literacy.

7.1. Distributional Symmetry

Although the distribution of market returns deviates from multivariate normality, and there is evidence of both positive skewness and fat tails (eg., Richardson & Smith,1993; Fama[37], 1976), a prevailing assumption in the finance literature asserts that market returns have a multivariate normal distribution. In fact, for standard deviation to be an sufficient measure of risk, the normality assumption of market returns distribution is inevitable. As stated earlier, HVW estimates subject-specific return volatility based on the assumption of normality. Following this assumption, I argue that a third requirement of full market literacy is to understand that the forward-looking

[37] Fama (1976) finds that the normal distribution assumption of 14 out of 30 DJIA companies is rejected. This is sufficient to reject multivariate normality assumption for DJIA portfolio returns, since if the distribution of a random variables is not univariate normal, it cannot come from a multivariate normal distribution. Besides, excess skewness and kurtosis are evident in the work of Fama (1976) and Richardson & Smith,1993

distribution of market returns should be close to symmetrical. To measure market literacy in symmetry a threshold of 1% in absolute value is established. To illustrate, consider a confidence interval spanning from 5% to 15%, indicating a midpoint of 10%. In this scenario, a projected return ranging from 9% to 11% would be categorized as approximately symmetric.

I conduct a triple sort of the data classifying the respondents' *SMP* values based on three criteria: optimism, and precision ranges, as well as literacy in symmetry. Interestingly, the *SMP* of all those who are fully market-literate (i.e., market-literate in return, risk, and symmetry) is 0.83. This category includes 35 subjects. In contrast, those who are entirely market-illiterate, totaling 108 subjects, hold an average *SMP* score of 0.57. The difference between these averages is statistically significant at 1%.[38]

Referring back to Table 7, we incorporate the market literacy in symmetry (*MktLitSym*) indicator into regressions R5, R6, and R7. In Panel A, the indicator is notably positive and statistically significant at a 10% significance level or better across all three cases, without the inclusion of controls. In regression R6, which incorporates indicators for all three aspects of market literacy, the indicator achieves significance at a 5% level. However, it is worth noting that while market literacy in return holds a stronger influence and maintains statistical significance at a 1% level, market literacy in risk does not demonstrate a significant impact. Consistent with our prior findings, when moving to Panel B and introducing controls, only market literacy in return maintains its statistical significance.

7.2. Varying the market literacy in return range

I perform a series of robustness checks for the definitions of market literacy types.[39] Of particular importance is the range chosen for market literacy in return, 6%-10%, which I acknowledge may seem somewhat arbitrary. I explore whether my principal findings are robust to changes in this range. Firstly, I narrowed the interval on the left side, moving to a 7%-10% range. Secondly, I narrowed the interval on the right side, shifting to a 6%-9% range. Remarkably, in both scenarios, even when incorporating a comprehensive set of controls, the coefficient for

[38] These findings are provided in Table A2 of the empirical appendix.

[39] As well as the robustness checks based on the interval for market literacy in return (described in the text), I performed robustness checks for the market literacy in risk interval and separately for the symmetry benchmark percentage. Broadly speaking results are similar.

MktLitRet is positive and significant. The outcomes of these two scenarios are provided in Table A3 and Table A4 of the empirical appendix. Finally, I consider a case where the range is contracted on both ends, resulting in a narrower 7%-9% range for market literacy in return. The outcomes of this adjustment reveal that *MktLitRet* is even more influential. Notably, by including control variables, its coefficient surpasses that of the baseline estimations for R1, R2, R6, and R7 of Table 2. Furthermore, in contrast to the baseline results provided in Table 2, the coefficient for *MktLitRet* in R6 is now greater than that of the superoptimistic range. This finding suggests a heightened significance of market literacy in return within this more focused range. Table A5 of the empirical appendix illustrates the outcomes of this adjustment.

7.3. Causality

It is argued that the relationship between market literacy in return, and perhaps financial/market literacy in general, with *SMP* is tantamount to reverse causality. To elaborate, if person A possesses market literacy while person B does not, my findings would suggest that person A is more likely to invest in equities. Nevertheless, it is crucial to recognize that this observation does not establish a causal link. In reality, it is conceivable that person B holds an equity portfolio and subsequently chose to gain market knowledge to better manage the acquired securities. This situation exemplifies the possibility of reverse causation. I begin by assuming that in all cases of task-specific knowledge, such as financial literacy, the absence of stock return ignorance, or, in this instance, market literacy, the presence of both causality and reverse causality is undoubtedly reasonable. However, the intriguing question does not revolve around this duality but rather centers on the extent to which reverse causality outweighs direct causality. While MV did not address exogeneity, VLA addressed the issue by using instrumental variables. Specifically, they use the financial situation of the oldest sibling and parents of respondents as instruments. To justify their choice of instruments, they argue that respondents cannot control the financial situation of those around them, but they can learn from them to improve their literacy. Therefore, the financial situation of siblings and parents can explain the financial literacy of respondents while exogenous with respect to *SMP*.

Similar to the approach taken by VLA, I implement instrumental variables, using two inherently exogenous variables in relation to stock market participation: 1) US states and 2) parents'

education. Regarding US states, the geography of financial literacy and its disparities among different regions has been extensively studied.[40] Peng *et al.* (2018) and Bumcrot *et al.* (2013) study the distribution of financial literacy in the US and find considerable geographic variation across the states. Therefore, I argue that the state in which respondents are residing can serve as an instrument for my market literacy proxy. I utilize UAS 121[41] to assign financial literacy scores to each US state based on the average number of correct answers provided by the participants residing in that state. The average of financial literacy across US states is 9. Accordingly, I define a dummy variable that takes a value of 1 if the financial literacy of a state is greater than the mean and zero otherwise.

The other instrument that I utilize is the education level of a respondent's parents. Grohmann *et al* (2015) argue that childhood characteristics such as family can impact adults' financial knowledge and consequently predict their financial behavior such as whether they invest in risky assets or not. Behrman *et al* (2010) use the father's and mother's level of education as an instrumental variable for their respondents' financial literacy. Exogeneity seems clear as one's parents' education is not under the control of offspring but can affect the learning curve of children. Utilizing UAS 76, I consider the education of fathers and mothers and merge them with my sample. The variable takes the value of 1 if either a respondent's mother or father has a college degree or higher and zero otherwise.

Table 8 shows the correlation matrix for market literacy in returns, *SMP*, and the instrumental variables. The first column indicates that both instruments are positively correlated with market literacy in returns at 1% or better supporting the relevance of the instruments.

(TABLE 8 HERE)

Panel A of Table 9 reports the first stage of the regressions of market literacy in return on US states and parents' education level. The coefficient for US states in R1 is positive and significant showing that respondents residing in states with greater financial literacy are more knowledgeable

[40] For instance, Fornero & Monticone (2011) study disparities of financial literacy in Italy, Klapper & Panos (2011) in Russia, Bucher-Koenen & Lusardi (2011) in Germany, Beckmann (2013) in Romania and Boisclair *et a.* (2017) find evidence for Canada.

[41] UAS 121 contains financial literacy scores. The survey took place between January 15, 2018, to August 25, 2020. The size of sample is 10945 and response rate is 85.66%,

about the likely market expected return. This finding is consistent with the evidence of Bumcrot *et al* (2013). R2 demonstrates the impact of parents' education on the financial literacy of their children. The results indicate that having at least one well-educated parent encourages the acquisition of market knowledge. Moreover, the Cragg-Donald statistic p-value at less than 5% in all specifications rules out the weak-instrument concern.

Panel B of Table 9 shows the second-stage IV results using GMM.[42] I instrument *MrktLitRet* using each IV separately (R1 and R2) and jointly (R3). The relationship between market literacy and *SMP* in all three specifications remains positive and significant, confirming my previous result. Moreover, the Hansen J-statistic from R3 shows that the null hypothesis that the instrument is uncorrelated with the error term cannot be rejected. The p-value of the endogeneity test shows that the original regressor was endogenous, confirming the validity of using IV estimation.

(TABLE 9 HERE)

It is worthwhile using US states and parents' education to instrument *AbRiskIg* and *FinLit*, in the first case because MV did not explore exogeneity for *AbRiskIg* and in the second case as an exogeneity robustness check. The results are available in Table A6 and Table A7 of the Empirical Appendix. To be brief, *FinLit* passes all the IV tests and instrumented *FinLit* remains positive and significant in *SMP* probit regressions. On the other hand, US states and parents' education are rejected as relevant instruments for *AbRiskIg.*

7.4. Comparing the impact of the three forms of financial and market literacy

In this section, I conduct an analysis to compare the effect of financial and market literary proxies on *SMP*. Because *FinLit* is based on 14 questions and is thus close to continuous, whereas both *AbRiskIg* and *MktLitRet* are dichotomous, to properly compare the relative impacts of these variables, it is appropriate to convert *FinLit* to an indicator. To accomplish this *FinLit* takes a value

[42] The white test suggests GMM estimator due to existence of heteroskedasticity since its more efficient than standard IV estimator when the variance of error term is not constant.

of 1 if a subject's score is greater than the sample median and zero otherwise[43]. Table 10 reports correlations between the three task-specific literacy indicators, as well as vs. *SMP* and other important controls such as education, income, and trust. Not surprisingly, all three are significantly positively correlated with *SMP* and with each other. It is also worth noting that while they are all positively related to income, education, and trust, *FinLit* has the strongest association with these other determinants, and *MktLitRet is* the weakest.

(TABLE 10 HERE)

Table 11 presents the results of probit estimations when these literacy indicators are used as independent variables either individually or in combination. Panel A excludes control variables while Panel B incorporates all control variables. Focusing on R4 of Panel A, *FinLit* has the largest impact, followed by *MktLitRet and AbRiskIg* in second and third positions, respectively. However, when controls are introduced in Panel B, all three coefficients experience a decline due to the influence of other pertinent factors like education and income, which absorb a portion of their effects. Notably, the *FinLit* coefficient experiences the most decrease reaching a level of similarity with the coefficients of the other two task-specific indicators. At this point, the differences between these coefficients are statistically insignificant. The reason for the greater decrease in the coefficient of *FinLit* can be attributed to its higher correlation with the other important determinants, leading to a more indirect influence through these intermediary variables.

(TABLE 11 HERE)

To summarize, market literacy in return stands out as an important determinant of *SMP*, with its impact not too far from that of financial literacy. This is perhaps surprising considering that *MktLitRet* comes from a single question, while *FinLit* is derived from multiple questions. When the probability of participation for an individual who is literate in all three senses is compared with the likelihood of participation for another individual who is illiterate across all three aspects, Table

[43] I use another assumption for converting continuous *FinLit to a* dichotomous variable which is *FinLit* = 1 if a subject's score is greater than or equal to the sample median and zero otherwise and repeat my analysis. The results remain essentially unchanged and are provided in the Table A8 of the empirical appendix.

11 reveals a remarkable boost of 53.4% in the probability of equity participation when control variables are not taken into account.

8. Discussion and conclusion

This study contributes to the stock market participation puzzle debate by investigating whether specific knowledge about likely future stock market returns, which I label as market literacy in return and market literacy in risk, encourages equity market participation. Moreover, I explore whether behavioral factors, in particular overoptimism and overprecision, as two main forms of overconfidence bias, contribute to stockholding decisions. The conventional equity exposure decision suggests that higher expected return and lower perceived risk should increase investment in stocks. Nevertheless, whether such linear relationships hold when it comes to stock market participation decisions is an empirical question. I show that the probability of *SMP* does not simply rise with the expected market return; instead, a demonstration of greater awareness about future returns is associated with higher *SMP*. The costs of obtaining information for equity investment (literacy cost), such as resources and time spent in learning about investment strategies, tracking stock performance, managing the risky portfolio, and so on deter households from stockholding at higher levels of expected return in the territory of overoptimism, with expected returns being greater than those reflecting market literacy in returns. In terms of perceived risk, Hurd, Van Rooij, and Winter (2011) find that *SMP* falls with perceived volatility. Although I conjecture that the ability to provide reasonable volatility for stock returns increases the probability of *SMP*, the impact of market literacy in risk and the overprecision factor, which is determined as confidence intervals being narrower than those reflecting market literacy in risk, are indirect and fully subsumed by other drivers of *SMP*.

As a precursor to the empirical work, I first develop a simple representative-agent model to illustrate how rational and behavioral forces interact in *SMP* decision-making. I incorporate investors' prior beliefs on risk premium coupled with the perceived cost of becoming literate, I illustrate that in the territory of overoptimism, the behavioral force of higher expected returns encouraging participation could potentially be outweighed by the rational consideration of the perceived literacy cost which discourages participation. This interplay can result in a lower participation rate in the territory of an overoptimist view regarding the expected return.

Second, I utilize survey data from the *Understanding America Study* panel, which provides a large, representative sample from the US population with high response rates and shows that market literacy in return matters for market entry and its effect is incremental to the previously documented financial literacy and absence of stock return ignorance determinants. I also compare the impact of the three forms of financial/market literacy, finding that market literacy in return is an important determinant of *SMP*, and its impact is roughly on par with the other two literacy factors. To address the reverse casualty between *SMP* and market literacy, I employ two instrumental variables, namely US states and parents' education, and show that these are valid instruments and that the causal impact of market literacy on *SMP* continues to hold.

My findings indicate that when individuals become knowledgeable about expected stock returns, they can make more informed and rational investment decisions. This can lead to increased stock market participation and reduced behavioral biases, such as overconfidence. Thus, interventions focusing on financial literacy and education should explicitly include instructions on understanding expected returns and risks as well as how to balance risk and reward more effectively.

Future research could explore the robustness of my results by exploring whether the results hold for different samples, such as those from other countries, both emerging and developed. While for the most part, my study uses variables measured at roughly the same point in time[44], a long-term longitudinal study would also be of interest so that changes in SMP over an individual's life cycle could be studied. The SMP literature has mostly focused on static measures of SMP, but the dynamics of SMP merit further exploration. This type of study may help lead to a well-defined standard model of SMP's explanatory variables, which at this point consists of a large and growing set of variables. Moreover, other samples may contain variables that would allow utilizing other instrumental variables to tackle the endogeneity issue.

[44] Recently, new waves of stock market participation (UAS295 and UAS387) and financial literacy (UAS 237 and UAS 458) surveys have been available within the UAS database. To enhance my sample, I integrated these new waves and established a panel set with three time periods After reperforming my analysis, I can affirm that my findings have remain essentially unchanged.

References

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

Beckmann, E. (2013). Financial literacy and household savings in Romania. *Numeracy*, *6*(2), 9.

Ben-David, I., J. R. Graham & C. R. Harvey (2013). Managerial miscalibration. *Quarterly Journal of Economics*, 128, 1547–1584.

Bhandari, G. & R. Deaves (2006). The demographics of overconfidence. *Journal of Behavioral Finance*, 7(1): 5-11.

Behrman, J. R., Mitchell, O. S., Soo, C. K., & Bravo, D. (2010). Financial literacy, schooling, and wealth accumulation. NBER Working Paper 16452.

Boisclair, D., Lusardi, A., & Michaud, P. C. (2017). Financial literacy and retirement planning in Canada. *Journal of pension economics and finance*, *16*(3), 277-296.

Bucher-Koenen, T., & Lusardi, A. (2011). Financial literacy and retirement planning in Germany. *Journal of Pension Economics and Finance*, *10*(4), 565-584.

Bucher-Koenen, T., Alessie, R. J., Lusardi, A., & Van Rooij, M. (2021). *Fearless woman: Financial literacy and stock market participation* (No. w28723). National Bureau of Economic Research.

Lusardi, A., Bumcrot, C. B., & Lin, J. (2013). The Geography of Financial Literacy. *Scholar Commons*.

Christelis, D., T. Jappelli & M. Padula (2010). Cognitive abilities and portfolio choice. *European Economic Review*, 54, 18–38.

Christiansen, C., J. S. Joensen & J. Rangvid (2008). Are economists more likely to hold stocks? *Review of Finance*, 12, 465–496.

Clemen, R. T. (2001). Assessing 10-50-90s: A surprise. *Decision Analysis Newsletter*, *20*(1),2.

Cole, S., A. Paulson & G. K. Shastry (2014). Smart money? The Effect of education on financial outcomes. *Review of Financial Studies*, 27, 2022-2051.

Damodaran, A. (2019). Equity Risk Premiums (ERP): Determinants, Estimation and Implications–The 2019 Edition. *NYU Stern School of Business*.

Davidson, L. B. & D. O. Cooper (1976). A simple way of developing a probability distribution of present value. *Journal of Petroleum Technology*, 28(09), 1069–1078.

Deaves, R., E. Lueders & G. Y. Luo (2009). An experimental test of the impact of overconfidence and gender on trading activity. *Review of Finance*, 13, 555-575.

Deaves, R., E. Lueders & M. Schroeder (2010). The dynamics of overconfidence: Evidence from stock

market forecasters. *Journal of Economic Behavior and Organization*, 75: 402-412.

Deaves, R., Ostad, P., & Stivers, A. (2022). Look behind the Curtain: The Direct and Indirect Impact of Non-cognitive Skills on Stock Market Participation. *Available at SSRN 4244525*.

Fama, E. F. (1976). Foundations of Finance Basic Books: New York.

Fernandez, P., Martinez, M., & Fernández Acín, I. (2019). Market risk premium and risk-free rate used for 69 countries in 2019: A survey. *Available at SSRN 3358901*.

Fornero, E., & Monticone, C. (2011). Financial literacy and pension plan participation in Italy. *Journal of Pension Economics and Finance*, *10*(4), 547-564.

Fraser-Mackenzie, P., M. Sung & J. E. V. Johnson (2015). The prospect of a perfect ending: Loss aversion and the round-number bias. *Organizational Behavior and Human Decision Processes*, 131, 67–80.

Georgarakos, D. & G. Pasini (2011). Trust, sociability, and stock market participation. *Review of Finance*, 15, 693–725.

Giannetti, M. & T. Y. Wang (2016). Corporate scandals and household stock market participation. *Journal of Finance*, 71, 2591-2636.

Goetzmann, W. N. & A. Kumar (2008). Equity portfolio diversification. *Review of Finance* 12, 433-463.

Graham, J. R. & C. R. Harvey (2018). The equity risk premium in 2018. *Available at SSRN 3151162*.

Grinblatt, M., M. Keloharju & J. Linnainmaa. (2011). IQ and stock market participation. *Journal of Finance*, 66, 2121-2164.

Grohmann, A., Kouwenberg, R., & Menkhoff, L. (2015). Childhood roots of financial literacy. *Journal of Economic Psychology*, *51*, 114-133.

Guiso, L., P. Sapienza & L. Zingales (2008). Trusting the stock market. *Journal of Finance*, 63, 2557-2601.

Haliassos, M. & C. C. Bertaut (1995). Why do so few hold stocks?. *Economic Journal*, 105, 1110–1129.

Henrion, M., & Fischhoff, B. (1986). Assessing uncertainty in physical constants. *American Journal of Physics*, 54, 791–798.

Hurd, M. D., M. Van Rooij & J. Winter (2011). Stock market expectations of Dutch households. *Journal of Applied Econometrics*, 26, 416–436.

Juslin, P., Winman, A., & Hansson, P. (2007). The Naïve Intuitive Statistician: A Naïve Sampling Model of Intuitive Confidence Intervals. *Psychological Review*, 114, 678–703.

Klapper, L. F., & Panos, G. A. (2011). Financial literacy and retirement planning in view of a growing youth demographic: The Russian case.

Lusardi, A. & O. S. Mitchell (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1), 5-44.

Lusardi, A., Michaud, P. C., & Mitchell, O. S. (2017). Optimal financial knowledge and wealth inequality. *Journal of Political Economy*, *125*(2), 431-477.

McKenzie, C., Liersch, M., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes*, 107, 179–191.

Marshall, G. N., Wortman, C. B., Kusulas, J. W., Hervig, L. K., & Vickers Jr, R. R. (1992). Distinguishing optimism from pessimism: Relations to fundamental dimensions of mood and personality. *Journal of personality and social psychology*, *62*(6), 1067.

Merkoulova, Y., & Veld, C. (2022). Does it pay to invest? The personal equity risk premium and stock market participation. *Journal of Banking and Finance*, *136*, 106220.

Merkoulova, Y., & Veld, C. (2022). Stock return ignorance. *Journal of Financial Economics*, *144*(3), 864-884.

Merkoulova, Y., & Veld, C. (2022). Why do individuals not participate in the stock market?. *International Review of Financial Analysis*, *83*, 102292.

Moore, D. M. & P. J. Healy (2008). The trouble with overconfidence. *Psychological Review*, 115, 502-517.

Morgan, M. G., & Keith, D. W. (2008). Improving the way we think about projecting future energy use and emissions of carbon dioxide. *Climatic Change*, 90, 189–215.

Peng, G., Liu, F., Lu, W., Liao, K., Tang, C., & Zhu, L. (2018). A spatial-temporal analysis of financial literacy in United States of America. *Finance Research Letters*, *26*, 56-62.

Richardson, M., & Smith, T. (1993). A test for multivariate normality in stock returns. *Journal of Business*, 295-321.

Sharot, T. (2011). The optimism bias: A tour of the irrationally positive brain. New York, NY, US: Pantheon.

Shum, P. & M. Faig (2006). What explains household stock holdings? *Journal of Banking and Finance*, 30: 2579-2597.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299.

Van Rooij, M., A. Lusardi & R. Alessie (2011). Financial literacy and stock market participation. *Journal of Financial Economics* 101: 449-472.

Vaarmets, T., Liivamägi, K., & Talpsepp, T. (2019). From academic abilities to occupation: What drives stock market participation?. *Emerging Markets Review*, *39*, 83-100.

Vissing-Jørgensen, A. (2002). Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures. *Working paper.*

Vissing-Jørgensen, A. (2003). Perspectives on behavioral finance: Does "irrationality" disappear with wealth? Evidence from expectations and actions. *NBER Macroeconomics Annual*, 2003, 139–194.

Von Gaudecker, H.-M. (2015). How does household portfolio diversification vary with financial literacy and financial advice? *Journal of Finance*, 70, 489-507.

Tables and figures

Table 1
Number of Respondents and Average *SMP* for *ExRet* Ranges

Value	Obs.	*SMP*	Value	Obs.	*SMP*
Negative	5	0.60	13	6	0.17
0	**130**	**0.10**	14	2	1.00
0<*ExRet*<1	12	0.33	**15**	**57**	**0.42**
1	**30**	**0.37**	16	1	0.00
1<*ExRet*<2	5	0.40	17	2	1.00
2	**159**	**0.42**	18	8	0.50
2<*ExRet*<3	20	0.70	19	3	0.00
3	**232**	**0.56**	**20**	**57**	**0.30**
3<*ExRet*<4	20	0.65	22	1	0.00
4	**235**	**0.56**	23	1	0.00
4<*ExRet*<5	8	0.63	**25**	**26**	**0.46**
5	**474**	**0.57**	29	1	1.00
5<*ExRet*<6	1	1.00	**30**	**27**	**0.41**
6	**192**	**0.69**	32	1	1.00
6<*ExRet*<7	14	0.86	33	3	0.00
7	**147**	**0.79**	34	1	0.00
7<*ExRet*<8	5	1.00	35	5	0.40
8	**172**	**0.82**	40	8	0.25
8<*ExRet*<9	2	0.50	42	1	1.00
9	**32**	**0.56**	45	3	0.33
9<*ExRet*<10	1	0.00	46	1	1.00
10	**234**	**0.59**	47	1	0.00
11	6	0.83	**50**	**33**	**0.33**
11<*ExRet*<12	1	0.00	Above 50	46	0.59
12	**45**	**0.67**	**Total**	**2477**	**0.56**

Figure 1

Frequency Distribution for *ExRet* Integer Values

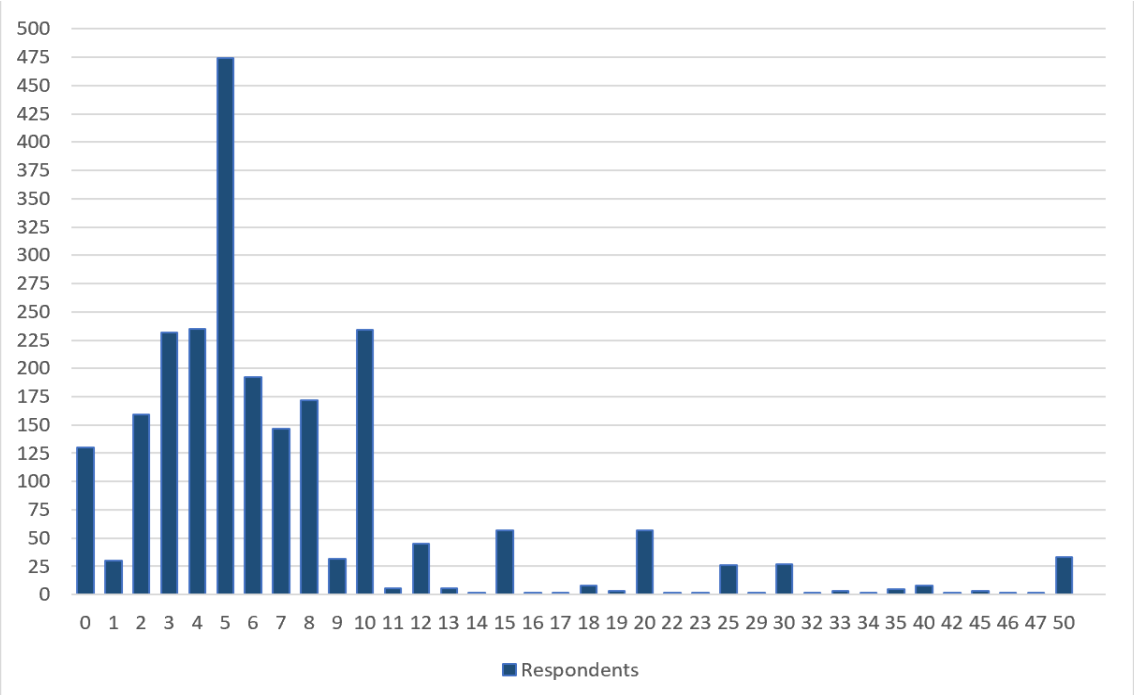


Figure 2
Number of Respondents and Average *SMP* for *CIWidth* Ranges

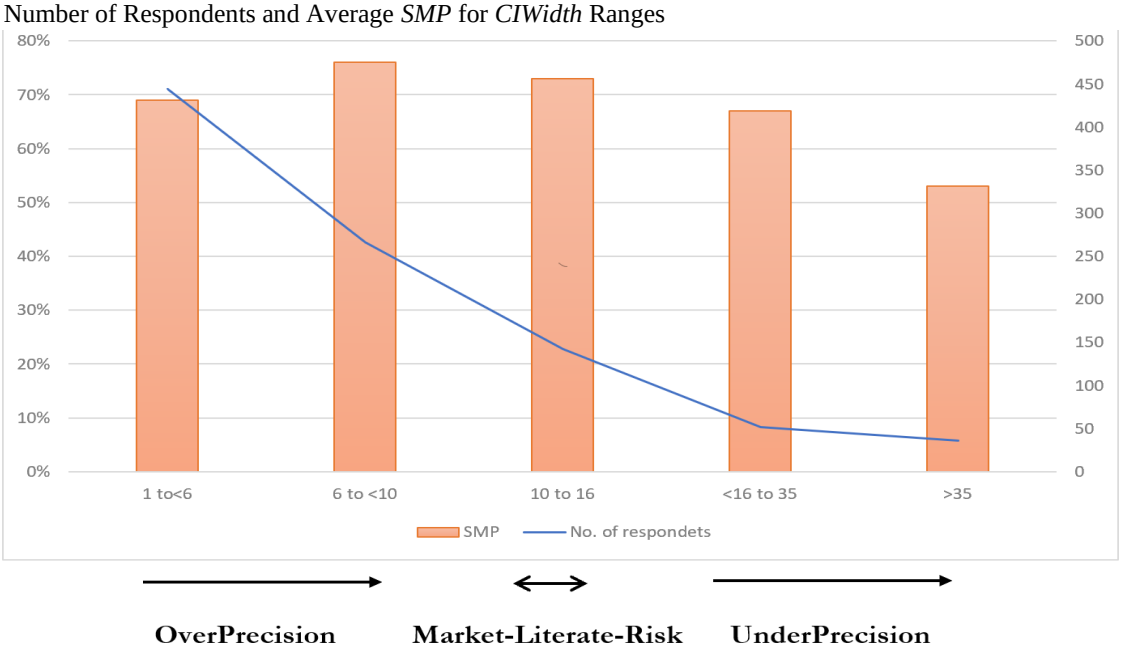


Figure 3
Average *SMP* for *ExRet* Integer Values with at least 25 Respondents

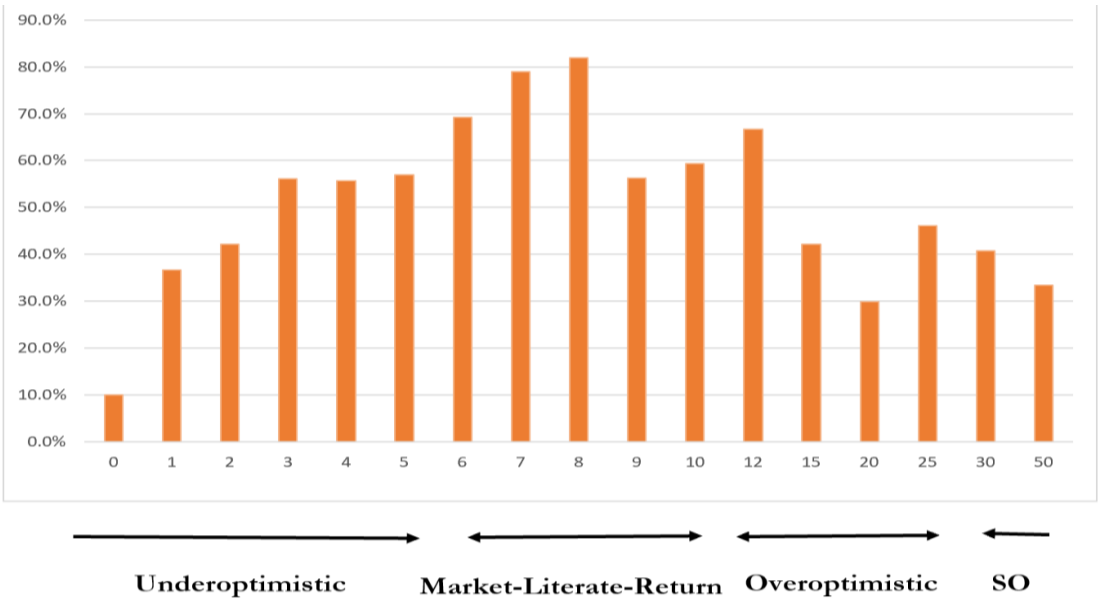


Table 2
Probit Estimation of *SMP* on *ExRet* Ranges

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
ExRet			0.015***	0.181***			
			(2.680)	(8.473)			
MktLitRet	0.565***	0.549***		1.992***		0.564***	0.544***
	(9.981)	(9.450)		(7.700)		(9.968)	(9.521)
OverOpt		-0.118		1.308***			
		(-1.281)		(3.680)			
ExRet × *MktLitRet*				-0.284***			
				(-7.710)			
ExRet × *OverOpt*				-0.229***			
				(-7.857)			
SuperOpt					-0.320***	-0.135	
					(-2.817)	(-1.179)	
ExRet_20_to_30							-0.389***
							(-2.728)
Controls	No	No	No	No	No	No	No
Observations	2,346	2,346	2,346	2,346	2,477	2,477	2,346
Pseudo-R2	0.0317	0.0322	0.0022	0.0608	0.00235	0.0322	0.0340

Panel B	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
ExRet			0.0191***	0.0954***			
			(2.906)	(3.875)			
MktLitRet	0.313***	0.326***		0.794***		0.314***	0.309***
	(4.837)	(4.920)		(2.685)		(4.865)	(4.737)
OverOpt		0.0978		0.731*			
		(0.920)		(1.762)			
ExRet × *MktLitRet*				-0.112***			
				(-2.641)			
ExRet × *OverOpt*				-0.113***			
				(-3.349)			
SuperOpt					0.355***	0.435***	
					(2.629)	(3.207)	
ExRet_20_to_30							-0.0772
							(-0.480)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,477	2,477	2,346
Pseudo-R2	0.258	0.258	0.253	0.263	0.255	0.262	0.258

Panel A presents the results of probit regressions. The dependent variable is equal to one if an individual is a stock market investor. The independent variables are defined as follows: *ExRet* is the participants' answer to the question about average annual return over the next 10 years. *MktLitRet* takes a value of one if 6≤ *ExRet* ≤10 and zero otherwise. *OverOpt* takes a value of one if 10<*ExRet* <30 and zero otherwise. *SuperOpt* takes a value of one if *ExRet* ≥30 and zero otherwise. *ExRet_20_to_30* takes a value of one if 20≤ *ExRet* <30, and zero otherwise. Panel B repeats probit regressions described in Panel A with control variables. Control variables are *Age*, *Age*^2, *Gender*, *Intelligence*, *NCS*, *Income*, *Education*, *Trust*, *FinLit*, *AbRiskIg*, *Openness*, *Extroversion*, *Agreeableness*, *BornUS* (born in the US=1), *MaritalStatus* (married=1), a series of indicators for race (where white is omitted) and a series of indicators for labor status (where employed is omitted). All variables are defined in Table A1 of the empirical appendix. The z-statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table 3
Average *SMP* and Control Variables with Non-parametric Tests for Optimism Ranges

***ExRet* Intervals**	**Obs**	***SMP***	***Intell.***	***NCS***	***FinLit***	***Income***	***Trust***	***Educ.***	***Age***	***Gender* (male=1)**	***AbRiskIg***
SuperOpt (SO)	130	0.44	144.32	14.56	7.56	9.56	3.90	1.35	51	0.34	0.17
OverOpt (OO)	217	0.45	155.13	15.10	8.96	10.78	4.02	1.29	51	0.41	0.36
MktLitRet (ML)	799	0.71	166.54	16.17	10.42	12.76	4.57	1.76	54	0.54	0.51
UnderOpt (UO)	1331	0.50	160.99	14.43	9.39	11.69	4.13	1.58	56	0.46	0.33
Total	**2477**										
Test Type		p-test	t-test	t-test	t-test	M-W	M-W	M-W	t-test	p-test	p-test
SO – OO		0.01	-11.19***	-0.54	-1.40***	-27.41**	-3.44	1.15	0	-0.7	-0.19***
test statistics		(-0.24)	(-4.60)	(-0.48)	(-4.17)	(-2.47)	(0.31)	(0.12)	(0)	(-1.42)	(-3.87)
SO – ML		-0.27***	-22.22***	-1.61*	-2.86***	-209.76***	-63.36**	-124.93***	-3***	-0.2***	-0.34***
test statistics		(-6.04)	(-11.96)	(-1.75)	(-12.06)	(-8.35)	(-2.52)	(-5.29)	(-2.66)	(-4.17)	(-7.24)
SO – UO		-0.06	-16.67***	0.13	-1.83***	-209.88***	-32.12	-105.89***	-5***	-0.12***	-0.16***
test statistics		(-1.31)	(-8.54)	(0.14)	(-7.24)	(-5.45)	(-0.40)	(-3.05)	(-3.73)	(-2.61)	(-3.84)
OO – ML		-0.26***	-11.41***	-1.07	-1.46***	-157.58***	-62.32***	-148.70***	-3***	-0.13***	-0.15***
test statistics		(-7.01)	(-7.81)	(-1.45)	(-7.55)	(-7.07)	(-2.80)	(-7.18)	(-2.73)	(-3.16)	(-3.83)
OO – UO		-0.05	-5.86***	0.67	-0.43**	-102.81***	-18.69	-124.19***	-5***	-0.5	0.03
test statistics		(-1.29)	(-3.85)	(0.92)	(-2.11)	(-3.61)	(-0.58)	(-4.25)	(-4.12)	(-1.17)	(0.88)
ML – UO		0.21***	5.55***	1.74***	1.03***	206.82***	105.85***	131.75***	-2.0**	0.08***	0.18***
test statistics		(9.42)	(6.11)	(3.97)	(8.95)	(7.68)	(3.90)	(5.22)	(-2.22)	(3.49)	(8.07)

This table reports the results of non-parametric tests for the average of *SMP* and various control variables across four intervals of expected returns. *MktLitRet* takes a value of one if 6≤ *ExRet* ≤10 and zero otherwise. *OverOpt* takes a value of one if 10<*ExRet* <30 and zero otherwise. *SuperOpt* takes a value of one if *ExRet* ≥30 and zero otherwise. *UnderOpt* takes a value of one if *ExRet* <6 and zero otherwise. The tests used to compare averages across various levels of expected return are p-tests (proportions tests), t-tests, and Mann-Whitney (M-W) tests. Control variables are defined in Table A1 of the empirical appendix. The test statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table 4
Probit Estimation of *SMP* on *CIWidth* Ranges

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**
MktLitRisk	0.0694		0.310*		0.459
	(0.570)		(1.760)		(0.631)
OverPre		0.0862	0.272*		-0.553
		(0.866)	(1.885)		(-1.522)
CIWidth				-0.00795*	-0.0146
				(-1.879)	(-1.605)
CIWidth × *MktLitRisk*					-0.0387
					(-0.733)
CIWidth × *OverPre*					0.0808***
					(3.130)
Controls	No	No	No	No	No
Observations	940	940	940	940	940
Pseudo-R2	0.000287	0.000657	0.00337	0.00309	0.0132
Panel B	**R1**	**R2**	**R3**	**R4**	**R5**
MktLitRisk	0.0373		-0.0826		-0.387
	(0.270)		(-0.412)		(-0.474)
OverPre		-0.0880	-0.137		-0.851**
		(-0.767)	(-0.827)		(-2.090)
CIWidth				0.000605	-0.0165
				(0.126)	(-1.631)
CIWidth × *MktLitRisk*					-0.00456
					(-0.0771)
CIWidth × *OverPre*					0.0476*
					(1.648)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	940	940	940	940	940
Pseudo-R2	0.214	0.215	0.215	0.214	0.218

Panel A presents the results of probit regressions. The dependent variable is equal to one if an individual is a stock market investor. The independent variables are defined as follows. *CIWidth* is defined for those who can answer the questions about the expected return, high return (*Ret90*), and low return (*Ret10*) and consistently order the three returns (*Ret10*<*ExRet*< *Ret90*) as the difference between *Ret*90 and *Ret*10. *OverPre* takes a value of 1 if 1≤ *CIWidth* <10 and zero otherwise. *MktLitRisk* takes a value of 1 if 10≤ *CIWidth* ≤16 and zero otherwise. Panel B repeats probit regressions presented in Panel A using the same controls as in Table 2. The z-statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table 5
Average *SMP* and Control Variables with Non-parametric Tests for Precision Ranges

***CIWidth* Intervals**	**Obs.**	***SMP***	***Intell.***	***NCS***	***FinLit***	***Income***	***Trust***	***Educ.***	***Age***	***Gender (male=1)***
OverPre (OP)	710	0.71	171.28	15.38	10.67	13.03	4.88	1.88	54	0.56
MktLitRisk (ML)	142	0.73	171.14	15.59	10.99	12.78	4.83	1.91	52	0.57
UnderPre (UP)	88	0.61	162.94	13.25	9.76	11.63	4.09	1.45	48	0.49
Total	**940**									
Test Type		ptest	t-test	t-test	t-test	M-W	M-W	M-W	t-test	ptest
OP – ML		-0.02	0.14	-0.21	-0.32	-5.47	9.02	-7.98	2*	-0.01
test statistics		(-0.31)	(0.09)	(-0.24)	(-1.62)	(-0.425)	(0.40)	(-0.373)	(1.76)	(-0.15)
OP – UP		0.1*	8.34***	2.13**	0.91***	93.69***	66.14**	101.51***	6***	0.07
test statistics		(1.92)	(4.42)	(2.01)	(3.70)	(3.63)	(2.56)	(4.14)	(3.85)	(1.33)
ML – UP		0.12*	8.20***	2.34*	1.23***	25.95***	15.39*	31.65***	4**	0.08
test statistics		(1.77)	(3.18)	(1.80)	(3.96)	(2.90)	(1.79)	(3.76)	(2.07)	(1.20)

Panel A reports the results of non-parametric tests for the average of *SMP* and various control variables across three confidence intervals. *CIWidth* is defined for those who can answer the questions about the expected return, high return (*Ret90*), and low return (*Ret10*) and consistently order the three returns (*Ret10*<*ExRet*< *Ret90*) as the difference between *Ret*90 and *Ret*10. *OverPre* takes a value of 1 if 1≤ *CIWidth* <10 and zero otherwise. *MktLitRisk* takes a value of 1 if 10≤ *CIWidth* ≤16 and zero otherwise. *UnderPre* takes a value of one if 16<*CIWidth* <75 and zero otherwise. The tests used to compare averages across various levels of expected return are p-tests (i.e., proportions tests), t-tests, and Mann-Whitney (M-W) tests. Control variables are defined in Table A1 of the empirical appendix. The test statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table 6
Double Sort for *SMP* vs. Optimism and Precision Ranges

	OverPre	***MktLitRisk***		***UnderPre***	Obs.
OverOpt* + *SuperOpt	0.85 (14)	(2.21***) (3.31***)	0.52 (33)	(0.23) (1.69*)	0.49 (45)	92
	(0.70)	(0.24)	(3.51***)	(4.08***)	(1.99**)	
MktLitRet	0.78 (297) (3.41***)	(1.06) (3.08***)	0.83 (83) (1.94*)	(1.13) (0.24)	0.73 (26) (0.14)	406
		(1.44)		(0.60)		
UnderOpt	0.66 (399)	(0.06)	0.65 (26)	(0.65)	0.75 (17)	442
Obs.	710		142		88	940

This table reports the results of non-parametric tests for the average of *SMP* based on a double sort on three intervals of *ExRet* in the first sort and the three intervals of *CIWidth* in the second sort. *OverOpt* interval is defined when 10< *ExRet* <30. *SuperOpt* interval is defined when 30≤ *ExRet* <100. *MktLitRet* in defined when 6≤ *ExRet* ≤10. *UnderOpt* interval is defined when *ExRet* <6. *OverPre* is defined when 1≤ *CIWidth* <10. *MktLitRisk* in defined when 10≤ CI ≤16. UnderPre is defined when 16 < CI <75. A proportions test is used to compare the average across various levels of *ExRet* and *CIWidth*. The test statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table 7
Probit Estimation of *SMP* on All Forms of Market Literacy

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
MktLitRet	0.422***	0.412***	0.396***	0.401***		0.426***	0.410***
	(4.693)	(4.575)	(4.212)	(4.261)		(4.726)	(4.346)
MktLitRisk	0.000659	0.202	0.0364	0.197		0.0901	0.148
	(0.00531)	(1.126)	(0.280)	(1.095)		(0.690)	(0.809)
OverPre		0.226		0.202			0.0574
		(1.551)		(1.285)			(0.324)
OverOpt			-0.157	-0.0743			-0.0757
			(-0.964)	(-0.422)			(-0.429)
MktLitSym					0.181*	0.227**	0.200*
					(1.943)	(2.293)	(1.764)
Controls	No	No	No	No	No	No	No
Observations	940	940	940	940	940	940	940
Pseudo-R2	0.0199	0.0220	0.0207	0.0222	0.0033	0.0245	0.0249
Panel B	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
MktLitRet	0.236**	0.241**	0.264**	0.260**		0.238**	0.267**
	(2.343)	(2.381)	(2.512)	(2.465)		(2.356)	(2.530)
MktLitRisk	-0.00564	-0.142	-0.0440	-0.133		0.0335	-0.183
	(-0.0402)	(-0.699)	(-0.302)	(-0.652)		(0.228)	(-0.888)
OverPre		-0.155		-0.112			-0.258
		(-0.930)		(-0.627)			(-1.277)
OverOpt			0.177	0.130			0.134
			(0.938)	(0.639)			(0.658)
MktLitSym					0.0788	0.100	0.199
					(0.753)	(0.907)	(1.575)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	940	940	940	940	940	940	940
Pseudo-R2	0.219	0.220	0.220	0.220	0.215	0.220	0.222

Panel A presents the results of probit regressions. The dependent variable is equal to one if an individual is a stock market investor. The independent variables are: *MktLitRet* is equal if 6<= *ExRet* <=10 and zero otherwise; *OverPre* takes 1 if 1=<CI <10 and zero otherwise; *MktLitRisk* takes 1 if 10<= CI <=16 and zero otherwise; *OverOpt* takes one if 10<*ExRet* <30 and zero otherwise. *MktLitSym* takes 1 if the midpoint of the respondent's confidence interval is -1 to 1 different from their *ExRet* and zero otherwise. Panel B repeats probit regressions presented in Panel A by adding other control variables. The z-statistics are in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table 8
Correlation Matrix of Instrumental Variables

	MktLitRet	***Statedum***	***Parentsduc***	***SMP***
MktLitRet	1.000			
USstates	0.209***	1.000		
Parentsduc	0.110***	0.189***	1.000	
SMP	0.206***	0.393***	0.165***	1.000

This table presents correlations between variables. *SMP* is equal to one if an individual is a stock market investor and zero otherwise. *MktLitRet* takes a value of one if 6≤ *ExRet* ≤10 and zero otherwise. *USstates*=1 if the financial literacy of a state is greater than the average of financial literacy scores across states and 0, otherwise. *Parents Educ*=1 if either mother or father has a college degree or higher and 0, otherwise. *** p<0.01, ** p<0.05, * p<0.1

Table 9
Exogeneity

Panel A **First-Stage Regressions OLS**	**US States** **R1**	**Parents, Education** **R2**	**US States & Parents' Education** **R3**
USstates	**0.129***** **(3.762)**		**0.116***** **(2.813)**
Parents Educ		**0.0647**** **(2.406)**	**0.0595**** **(2.211)**
Controls	Yes	Yes	Yes
Observations	2,346	1,541	1,541
R-squared	0.073	0.079	0.084
Cragg and Donald statistic	15.416	5.595	7.0554
(p-value)	(0.000)	(0.018)	(0.000)

Panel B **Second-Stage Regressions GMM**	**US States** **R1**	**Parents' Education** **R2**	**US States& Parents' Education** **R3**
MrktLitRet	**0.843***** **(2.676)**	**0.910*** **(1.787)**	**0.780**** **(2.524)**
Controls	Yes	Yes	Yes
Observations	2,346	1,541	1,541
R-squared (uncentered)	-0.167 (0.49)	-0.244 (0.423)	-0.082(0.49)
Hansen J statistic (p-value)	N/A[45]	N/A	0.130 (0.718)
Endogeneity test Chi2	10.024	4.840	8.163
(p-value)	(0.001)	(0.027)	(0.004)

Panel A shows the estimates of the first-stage OLS regressions of market literacy in return on the set of controls and dummy variables indicating literacy of US states and parents' education. Panel B reports GMM estimates of the effect of market literacy in return on *SMP*. *MktLitRet* has been instrumented using two dummy variables indicating literacy of US states and parents' education. *MktLitRet* takes a value of one if 6≤ *ExRet* ≤10 and zero otherwise. *USstates*=1 if the financial literacy of a state is greater than the average of financial literacy scores across states and 0, otherwise. *Parents Educ*=1 if either mother or father has a college degree or higher and 0, otherwise. Controls are *Age*, *Age*^2, *Gender*, *Intelligence*, *NCS*, *Income*, *Education*, *Trust*, *FinLit*, *AbRiskIg*, *Openness*, *Extroversion*, *Agreeableness*, *BornUS* (born in the US=1), *MaritalStatus* (married=1), a series of indicators for race (where white is omitted) and a series of indicators for labor status (where employed is omitted). All variables are defined in Table A1 of the empirical appendix. The z-statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

[45] Hansen test is not applicable when when the model is exactly specified.

Table 10
Correlation Matrix of Literacy variables

Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. SMP	1.000										
2. FinLit	**0.338*** **	1.000									
3. AbRiskIg	**0.236*** **	**0.284*** **	1.000								
4. MktLitRet	**0.207*** **	**0.193*** **	**0.178*** **	1.000							
5. MktLitRisk	0.019	0.086***		0.130***	1.000						
6. MktLitSym	0.064*	0.002		-0.052	-0.303***	1.000					
7. Education	0.313***	0.340***	0.216***	0.125***	0.029	0.045	1.000				
8. Income	0.467***	0.332***	0.220***	0.172***	-0.010	0.063*	0.384***	1.000			
9. Trust	0.205***	0.185***	0.161***	0.081***	0.005	0.038	0.207***	0.218***	1.000		
10. Age	0.059***	0.142***	-0.083***	-0.016	-0.040	0.053*	0.008	-0.050**	0.077***	1.000	
11. Gender	0.182***	0.230***	0.134***	0.087***	0.011	0.047	0.082***	0.177***	0.037*	0.111***	1.000

This table presents correlations between variables. *SMP* is equal to one if an individual is a stock market investor and zero otherwise. *FinLit* is the respondent's score on the financial literacy index, between 0 and 14. *AbStkRetIg* takes a value of one if respondents can answer the questions about the expected return, high return (r_{90}), and low return (r_{10}) and consistently order the three returns ($r_{10}<r_{50}<r_{90}$). *MktLitRet* takes a value of one if $6\leq$ *ExRet* ≤ 10 and zero otherwise. *MktLitRisk* in defined when $10\leq$ *CIWidth* ≤ 16. *MktLitSym* takes a value of one if the difference between the midpoint of *CIWidth* and the expected return is greater than or equal to -1 and less than or equal to 1, and zero otherwise. *Education* is measured in five categories: not having a high school diploma takes zero; high school diploma or college below bachelor's takes 1; bachelor's degree takes 2; master's degree or professional school degree takes 3; and PhD takes 4 (note that not having a high school diploma is the omitted category). *Income* is measured in 16 categories from less than $5,000 to $150,000 or more. *Trust* is defined on a Likert scale from zero to 10, where zero corresponds to "you can never be too careful" and 10 to "most people can be trusted". *Gender* is equal to one if the respondent is male. *** p<0.01, ** p<0.05, and * p<0.1.

Table 11
SMP Probits with Task-Specific Indicators

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**
FinLit	0.893***			0.745***	0.683***
	(16.78)			(13.36)	(7.525)
AbRiskIg		0.631***		0.393***	
		(11.78)		(6.873)	
MktLitRet			0.576***	0.403***	0.350***
			(10.29)	(6.815)	(3.770)
MktLitRisk					0.0106
					(0.0799)
MktLitSym					0.215**
					(2.132)
Controls	No	No	No	No	No
Observations	2,477	2,477	2,477	2,477	940
Pseudo-R2	0.0857	0.0417	0.0319	0.117	0.0747

Panel B	**R1**	**R2**	**R3**	**R4**	**R5**
FinLit	0.378***			0.319***	0.243**
	(5.794)			(4.802)	(2.181)
AbRiskIg		0.313***		0.245***	
		(4.969)		(3.811)	
MktLitRet			0.357***	0.309***	0.244**
			(5.701)	(4.852)	(2.432)
MktLitRisk					0.0703
					(0.481)
MktLitSym					0.126
					(1.148)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	2,477	2,477	2,477	2,477	940
Pseudo-R2	0.237	0.235	0.237	0.249	0.208

Panel A presents the results of the probit regressions. The dependent variable is equal to one if an individual is a stock market investor. The independent variables are *FinLit* which takes 1 if the respondent's score is greater than the median and zero otherwise. *AbRiskIg* which is 1 if a respondent consistently orders *Ret10*<*ExRet*<*Ret90* and zero otherwise. *MktLitRet* which equals 1 if 6≤ *ExRet* ≤10 and zero otherwise; *MktLitRisk* which takes 1 if 10≤ *CIWidth* ≤16 and zero otherwise; and *MktLitSym* which takes 1 if the midpoint of a respondent's confidence interval is -1 to 1 different from their *ExRet* and zero otherwise. Panel B repeats probit regressions presented in panel 1 by adding other control variables as shown in Table 2. All variables are defined in Table A1 of the empirical appendix. The z-statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Appendix

Table A1
Variable Names, Sources and Definitions

Variable	Survey	Definition/Source/Calculation
Stock Market Participation (*SMP*)	UAS 184	Var. "as_003a" *SMP* = 1 indicates investment in equity and 0, otherwise
Literacy Variables		
Expected Return (*ExRet*)	UAS 184	Var. "as_004a" ranging from -9 to 100%, upper bound truncated at 100%
Market Literacy in Return (*MktLitRet*)		*MktLitRet* = 1 if 6≤ *ExRet* ≤10 and 0, otherwise
Overoptimistiv (*OverOpt*)		*OverOpt* =1 if 10<*ExRet* <30 and 0, otherwise
Superoptimistic (*SuperOpt*)		*SuperOpt* =1 if *ExRet* ≥30 and 0, otherwise
ExRet_20_to_30		*ExRet_20_to_30* =1 if 20≤ *ExRet* <30, and 0, otherwise
Confidence Interval Width (*CIWidth*)	UAS 184	Var. "as_004a, as_004b, as_004c" ,if as_004a, as_004b, and as_004c have been answered and consistently ordered (as_004b < as_004a < as_004c), *CIWidth*= as_004c - as_004b
Market Literacy in Risk (*MktLitRisk*)		*MktLitRisk* = 1 if 10≤ *CIWidth* ≤16 and 0, otherwise
OverPrecision (*OverPre*)		*OverPre* = 1 if 1≤ *CIWidth* <10 and 0, otherwise
Market Literacy in Symmetry (*MktLitSym*)		*MktLitSym* =1 if the difference between the midpoint of *CIWidth* and expected return is greater than or equal to -1 and less than or equal to 1, and 0, therwise.
Other Literacy Variables		
Financial Literacy (*FinLit*)	UAS 121	Var. "finlitscore" Scores ranging from 0 to 14
Absence of Stock Return Ignorance (*AbStkRetIg*)	UAS 184	Var. "as_004a, as_004b, as_004c" *AbStkRetIg* =1, if as_004a, as_004b, and as_004c have been answered and consistently ordered (as_004b < as_004a < as_004c) and 0, otherwise
Instrumental Variables		
US states (*USstates*)	UAS 184	Var. "finlitscore" The financial literacy of each state is calculated using the mean value of correct responses provided by the residents of that state. *USstates*=1 if the financial literacy of a state is greater than the average of financial literacy scores across states and 0, otherwise.
Parents' Education (*Parents Educ*)	UAS 76	*Parents Educ*=1 if either mother or father has a college degree or higher and 0, otherwise

Table A1 Continued: Variable	**Survey**	**Definition/Source/Calculation**
Control Variables		
Income	UAS 184	Var. “hhincome” Categorical variable for annual income ranging from 1 (less than $5,000) to 16 ($150,000 or more)
Education	UAS 184	Var. “education” Categorical variable ranging from 0 to 4 where below high school=0, high school diploma or college below bachelor’s =1, bachelor’s degree=2, master’s degree, or professional school degree =3, PhD=4
Trust	UAS 184	Var. “as_005” It is defined on a Likert scale from zero to 10, where zero corresponds to “you can never be too careful” and 10 to “most people can be trusted”.
Intelligence (*Intell*)	UAS 83, 84, 85	*Intelligence*= Var. “uas83cog” + Var. “uas84cog” + Var. “uas85cog”
Gender	UAS 184	Var. “gender” Gender=1, if male and 0 if female
Age	UAS 184	Var. “age”
Openness	UAS 121	Var.” openness” ranging from 10 to 50 with higher value corresponding to higher level of openness.
Extraversion	UAS 121	Var.” extraversion” ranging from 8 to 40 with higher value corresponding to higher level of extraversion.
Disagreeableness	UAS 121	Var.” agreeableness” *-1 ranging from -45 to -9 with higher value corresponding to higher level of disagreeableness.
Conscientiousness	UAS 121	Var.” conscientiousness” ranging from 9 to 45 with higher value corresponding to higher level of conscientiousness.
Emotional Stability (*EmotStab*)	UAS 121	Var.” neuroticism” *-1 Ranging from -40 to 8 with higher value corresponding to higher level of Emotional Stability.
Non-Cognitive Skills (*NCS*)	UAS 121	*NCS*= (*Conscientiousness* + *EmotStab)*/2
Marital Status (Married)	UAS 184	Var. “maritalstatus” *Married* = 1 if married, and 0, otherwise
Born in US (*BornUS*)	UAS 184	Var. “bornus” *BornUS* = 1 if the respondent was born in the US and 0, otherwise
Race	UAS 184	Var. “race”, ranging from 1 to 6
Labor Status	UAS 184	Var. “laborstatus”, ranging from 1 to 6

Table A2
Triple Sort for *SMP* vs. Symmetry plus Optimism and Precision Ranges

	OverPre	***MktLitRisk***	***UnderPre***	**Total**	**Obs.**
OverOpt* + *SuperOpt	s: 0.85 (13) a: 1 (1) (-)	s: 0.55 (11) a: 0.50 (22) (0.25)	s: 0.63 (8) a: 0.46 (37) (0.85)	**s: 0.69 (32)** **a: 0.48 (60)** **(1.88*)**	92
MktLitRet	s: 0.78 (237) a: 0.75 (60) (0.58)	s: 0.83 (35) a: 0.83 (48) (0.06)	s: 1 (2) a: 0.71 (24) (-)	s: 0.79 (274) a: 0.77 (132) (0.44)	406
UnderOpt	s: 0.67 (346) a: 0.58 (53) (1.21)	s: 0.86 (7) a: 0.58 (19) (1.32)	s: -(0) a: 0.75 (17) (-)	s: 0.67 (353) a: 0.61 (89) (1.08)	442
Total	s: 0.72 (596) a: 0.68 (114) (0.96)	s: 0.77 (53) a: 0.70 (89) (0.99)	s: 0.70 (10) a: 0.60 (77) (0.63)	**s: 0.72 (659)** **a: 0.66 (281)** **(1.95*)**	
Obs.	710	142	88		**940**

This table reports the results of non-parametric tests for the average of *SMP* based on a triple sort on three intervals of *ExRet* in the first sort, the three intervals of *CIWidth* in the second sort, and *MktLitSym* =1 (s) and *MktLitSym* =0 (a) in the third sort. *OverOpt* interval is defined when 10< *ExRet* <30. *SuperOpt* interval is defined when 30≤*ExRet* <100. *MktLitRet* is defined when 6≤ *ExRet* ≤10. *UnderOpt* interval is defined when *ExRet* <6. *OverPre* is defined when 1≤ *CIWidth* <10. *MktLitRisk* in defined when 10≤ *CIWidth* ≤16. UnderPre is defined when 16<*CIWidth* <75. *MktLitSym* takes a value of one if the difference between the midpoint of *CIWidth* and the expected return is greater than or equal to -1 and less than or equal to 1, and zero otherwise. A proportions test is used to compare the average across various levels of *ExRet* and *CIWidth*. The test statistics are provided in parentheses, and *** p<0.01, ** p<0.05, * p<0.1.

In Table A2, I conduct a triple sort of data based on optimism, precision and symmetry. This division results in 18 distinct categories (3 x 3 x 2). While none of the within-cell differences are statistically indistinguishable from zero, in all cases where there are at least 10 observations in a three-way bin, the *SMP* of the symmetric bin was greater than or equal to that of the asymmetric bin. Additionally, I did marginal and total tests of significant differences. For example, 710 subjects are overprecise; of these 114 are market-illiterate in symmetry and the rest market-literate in this sense. The symmetry vs. asymmetry *SMP*s are 0.72 and 0.68. In all six marginal cases the *SMP* for symmetry was higher than that for asymmetry, though the difference was only significant at 10% for overoptimism, while holding precision constant. Finally, in the bottom right, I found for the entire sample that those who are market-literate/-illiterate in symmetry had an *SMP* of 0.72/0.66, and the difference was statistically significant at 10%.

Table A3
Robustness Check when *MktLitRet* Range is 7-10%

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
ExRet			0.0153***	0.188***			
			(2.678)	(10.07)			
MktLitRet	0.506***	0.483***		3.231***		0.505***	0.484***
	(8.153)	(7.655)		(7.989)		(8.142)	(7.755)
OverOpt		-0.187**		1.325***			
		(-2.051)		(3.736)			
ExRet × *MktLitRet*				-0.423***			
				(-8.687)			
ExRet × *OverOpt*				-0.235***			
				(-8.664)			
SuperOpt					-0.320***	-0.198*	
					(-2.817)	(-1.729)	
ExRet_20_to_30							-0.452***
							(-3.180)
Controls	No	No	No	No	No	No	No
Observations	2,346	2,346	2,346	2,346	2,477	2,477	2,346
Pseudo-R2	0.0212	0.0225	0.00223	0.0659	0.00235	0.0223	0.0244

Panel B	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
ExRet			0.0191***	0.0943***			
			(2.906)	(4.360)			
MktLitRet	0.329***	0.338***		1.626***		0.331***	0.325***
	(4.666)	(4.713)		(3.567)		(4.686)	(4.579)
OverOpt		0.0723		0.734*			
		(0.685)		(1.775)			
ExRet × *MktLitRet*				-0.201***			
				(-3.625)			
ExRet × *OverOpt*				-0.113***			
				(-3.548)			
SuperOpt					0.352***	0.417***	
					(2.598)	(3.071)	
ExRet_20_to_30							-0.0980
							(-0.611)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,477	2,477	2,346
Pseudo-R2	0.258	0.258	0.253	0.265	0.255	0.261	0.258

Panel A presents the results of probit regressions. The dependent variable is equal to one if an individual is a stock market investor. The independent variables are defined as follows: *ExRet* is the participant's answer to the question about average annual return over the next 10 years. *MktLitRet* takes a value of one if 7≤ *ExRet* ≤10 and zero otherwise. *OverOpt* takes a value of one if 10<*ExRet* <30 and zero otherwise. *SuperOpt* takes a value of one if *ExRet* ≥30 and zero otherwise. *ExRet_20_to_30* takes a value of one if 20≤ *ExRet* <30, and zero otherwise. Panel B repeats probit regressions described in Panel A with control variables. Control variables are *Age*, *Age*^2, *Gender*, *Intelligence*, *NCS*, *Income*, *Education*, *Trust*, *FinLit*, *AbRiskIg*, *Openness*, *Extroversion*, *Agreeableness*, *BornUS* (born in the US=1), *MaritalStatus* (married=1), a series of indicators for race (where white is omitted) and a series of indicators for labor status (where employed is omitted). All variables are defined in Table A1 of the empirical appendix. The z-statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table A4
Robustness Check when *MktLitRet* Range is 6-9%

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
ExRet			0.0153***	0.181***			
			(2.678)	(8.473)			
MktLitRet	0.679***	0.695***		0.878**		0.678***	0.659***
	(10.47)	(10.36)		(2.003)		(10.46)	(10.11)
OverOpt		0.0648		1.342***			
		(0.949)		(6.713)			
ExRet × *MktLitRet*				-0.119*			
				(-1.859)			
ExRet × *OverOpt*				-0.230***			
				(-9.218)			
SuperOpt					-0.320***	-0.168	
					(-2.817)	(-1.471)	
ExRet_20_to_30							-0.421***
							(-2.960)
Controls	No	No	No	No	No	No	No
Observations	2,346	2,346	2,346	2,346	2,477	2,477	2,346
Pseudo-R2	0.0356	0.0359	0.00223	0.0639	0.00235	0.0359	0.0384

Panel B	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
ExRet			0.0191***	0.0957***			
			(2.906)	(3.891)			
MktLitRet	0.347***	0.386***		-0.217		0.350***	0.343***
	(4.653)	(5.012)		(-0.434)		(4.693)	(4.572)
OverOpt		0.157**		0.711***			
		(2.025)		(3.132)			
ExRet × *MktLitRet*				0.0384			
				(0.521)			
ExRet × *OverOpt*				-0.113***			
				(-3.914)			
SuperOpt					0.352***	0.403***	
					(2.598)	(2.975)	
ExRet_20_to_30							-0.105
							(-0.654)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,477	2,477	2,346
Pseudo-R2	0.258	0.259	0.253	0.265	0.255	0.261	0.258

Panel A presents the results of probit regressions. The dependent variable is equal to one if an individual is a stock market investor. The independent variables are defined as follows: *ExRet* is the participants' answer to the question about average annual return over the next 10 years. *MktLitRet* takes a value of one if 6≤ *ExRet* ≤9 and zero otherwise. *OverOpt* takes a value of one if 9<*ExRet* <30 and zero otherwise. *SuperOpt* takes a value of one if *ExRet* ≥30 and zero otherwise. *ExRet_20_to_30* takes a value of one if 20≤ *ExRet* <30, and zero otherwise. Panel B repeats probit regressions described in Panel A with control variables. Control variables are *Age*, *Age*^2, *Gender*, *Intelligence*, *NCS*, *Income*, *Education*, *Trust*, *FinLit*, *AbRiskIg*, *Openness*, *Extroversion*, *Agreeableness*, *BornUS* (born in the US=1), *MaritalStatus* (married=1), a series of indicators for race (where white is omitted) and a series of indicators for labor status (where employed is omitted). All variables are defined in Table A1 of the empirical appendix. The z-statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table A5
Robustness Check when *MktLitRet* Range is 7-9%

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
ExRet			0.0153***	0.188***			
			(2.678)	(10.07)			
MktLitRet	0.725***	0.724***		2.929***		0.724***	0.704***
	(9.120)	(8.944)		(3.280)		(9.112)	(8.841)
OverOpt		-0.00428		1.359***			
		(-0.0638)		(6.851)			
ExRet × *MktLitRet*				-0.383***			
				(-3.288)			
ExRet × *OverOpt*				-0.237***			
				(-10.44)			
SuperOpt					-0.320***	-0.220*	
					(-2.817)	(-1.929)	
ExRet_20_to_30							-0.473***
							(-3.333)
Controls	No	No	No	No	No	No	No
Observations	2,346	2,346	2,346	2,346	2,477	2,477	2,346
Pseudo-R2	0.0276	0.0276	0.00223	0.0660	0.00235	0.0284	0.0312

Panel B	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
ExRet			0.0191***	0.0943***			
			(2.906)	(4.360)			
MktLitRet	0.436***	0.465***		0.665		0.440***	0.432***
	(4.797)	(5.029)		(0.651)		(4.841)	(4.736)
OverOpt		0.132*		0.709***			
		(1.719)		(3.149)			
ExRet × *MktLitRet*				-0.0724			
				(-0.540)			
ExRet × *OverOpt*				-0.111***			
				(-4.236)			
SuperOpt					0.352***	0.391***	
					(2.598)	(2.892)	
ExRet_20_to_30							-0.121
							(-0.759)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,346	2,346	2,346	2,346	2,477	2,477	2,346
Pseudo-R2	0.258	0.259	0.253	0.266	0.255	0.262	0.258

Panel A presents the results of probit regressions. The dependent variable is equal to one if an individual is a stock market investor. The independent variables are defined as follows: *ExRet* is the participants' answer to the question about average annual return over the next 10 years. *MktLitRet* takes a value of one if 7≤ *ExRet* ≤19 and zero otherwise. *OverOpt* takes a value of one if 9<*ExRet* <30 and zero otherwise. *SuperOpt* takes a value of one if *ExRet* ≥30 and zero otherwise. *ExRet_20_to_30* takes a value of one if 20≤ *ExRet* <30, and zero otherwise. Panel B repeats probit regressions described in Panel A with control variables. Control variables are *Age*, *Age*^2, *Gender*, *Intelligence*, *NCS*, *Income*, *Education*, *Trust*, *FinLit*, *AbRiskIg*, *Openness*, *Extroversion*, *Agreeableness*, *BornUS* (born in the US=1), *MaritalStatus* (married=1), a series of indicators for race (where white is omitted) and a series of indicators for labor status (where employed is omitted). All variables are defined in Table A1 of the empirical appendix. The z-statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Table A6
Addressing Exogeneity for Financial Literacy

Panel A **Fist-Stage Regressions** **OLS**	**US States** **R1: *FinLit***	**Parents' Education** **R2: *FinLit***	**US States & Parents' Education** **R3: *FinLit***
USstates	**0.0754****		**0.0700***
	(2.367)		**(1.775)**
ParentsEduc		**0.113*****	**0.115*****
		(2.590)	**(2.636)**
Controls	Yes	Yes	Yes
Observations	2,346	1,541	1,541
R-squared	0.432	0.430	0.431
Cragg and Donald statistic	5.82	6.692	4.919
(p-value)	(0.016)	(0.009)	(0.007)

Panel B **Second-Stage Regressions** **GMM**	**US States** **R1: *SMP***	**Parents' Education** **R2: *SMP***	**US States& Parents' Education** **R3: *SMP***
FinLit	**0.640***	**0.570****	**0.513****
	(1.942)	**(2.020)**	**(2.314)**
Controls	Yes	Yes	Yes
Observations	2,346	1,541	1,541
R-squared (uncentered)	-0.352 (0.412)	-0.167 (0.459)	-0.057 (0.51)
Hansen J statistic	N/A[46]	N/A	0.141
(p-value)			(0.706)
Endogeneity test Chi2	5.496	4.840	5.511
(p-value)	(0.019)	(0.027)	(0.018)

[46] Hansen test is not applicable when when the model is exactly specified.

Table A7
Addressing Exogeneity for Absence of Risk Ignorance

Panel A **Fist-Stage Regressions** **OLS**	**US States** **R1: *AbRiskIg***	**Parents' Education** **R2: *AbRiskIg***	**US States & Parents' Education** **R3: *AbRiskIg***
USstates	**-0.0258**		**-0.0163**
	(-1.346)		**(-0.683)**
ParentsEduc		**-0.0313**	**-0.0318**
		(-1.185)	**(-1.203)**
Controls	Yes	Yes	Yes
Observations	2,346	1,541	1,541
R-squared	0.185	0.193	0.194
Cragg and Donald statistic (p-value)	1.841 (0.175)	1.355 (0.244)	0.895 (0.409)

Panel B **Second-Stage Regressions** **GMM**	**US States** **R1: SMP**	**Parents' Education** **R2: SMP**	**US States& Parents' Education** **R3: SMP**
AbStkRetIg	**-1.535**	**-1.612**	**-1.502**
	(-1.117)	**(-0.985)**	**(-1.114)**
Controls	Yes	Yes	Yes
Observations	2,346		
R-squared (uncentered)	-1.746	-1.996	-1.709
Hansen J statistic (p-value)	N/A[47]	N/A	0.020 (0.887)
Endogeneity test Chi2 (p-value)	5.496 (0.019)	4.840 (0.028)	5.650 (0.017)

[47] Hansen test is not applicable when when the model is exactly specified.

Table A8

SMP Probits with Task-Specific Indicators

Panel A	**R1**	**R2**	**R3**	**R4**	**R5**
FinLit	1.051***			0.907***	0.956***
	(19.57)			(16.16)	(9.550)
AbRiskIg		0.631***		0.373***	
		(11.78)		(6.463)	
MktLitRet			0.576***	0.348***	0.310***
			(10.29)	(5.802)	(3.289)
MktLitRisk					0.0143
					(0.106)
MktLitSym					0.203**
					(1.981)
Controls	No	No	No	No	No
Observations	2,477	2,477	2,477	2,477	940
Pseudo-R2	0.117	0.0417	0.0319	0.142	0.106

Panel B	**R1**	**R2**	**R3**	**R4**	**R5**
FinLit	0.533***			0.470***	0.471***
	(7.953)			(6.917)	(3.821)
AbRiskIg		0.313***		0.240***	
		(4.969)		(3.729)	
MktLitRet			0.357***	0.284***	0.226**
			(5.701)	(4.434)	(2.240)
MktLitRisk					0.0598
					(0.410)
MktLitSym					0.123
					(1.120)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	2,477	2,477	2,477	2,477	940
Pseudo-R2	0.246	0.235	0.237	0.257	0.217

Panel A presents the results of the probit regressions. The dependent variable is equal to one if an individual is a stock market investor. The independent variables are *FinLit* which takes 1 if the respondent's score is greater than or equal to the median and zero otherwise. *AbRiskIg* which is 1 if a respondent consistently orders *Ret10<ExRet<Ret90* and zero otherwise. *MktLitRet* which equals 1 if 6≤ *ExRet* ≤10 and zero otherwise; *MktLitRisk* which takes 1 if 10≤ *CIWidth* ≤16 and zero otherwise; and *MktLitSym* which takes 1 if the midpoint of a respondent's confidence interval is -1 to 1 different from their *ExRet* and zero otherwise. Panel B repeats probit regressions presented in panel 1 by adding other control variables as shown in Table 2. All variables are defined in Table A1 of the empirical appendix. The z-statistics are provided in parentheses and *** p<0.01, ** p<0.05, * p<0.1.

Figure A1
Rational Vs. Behavioral Forces

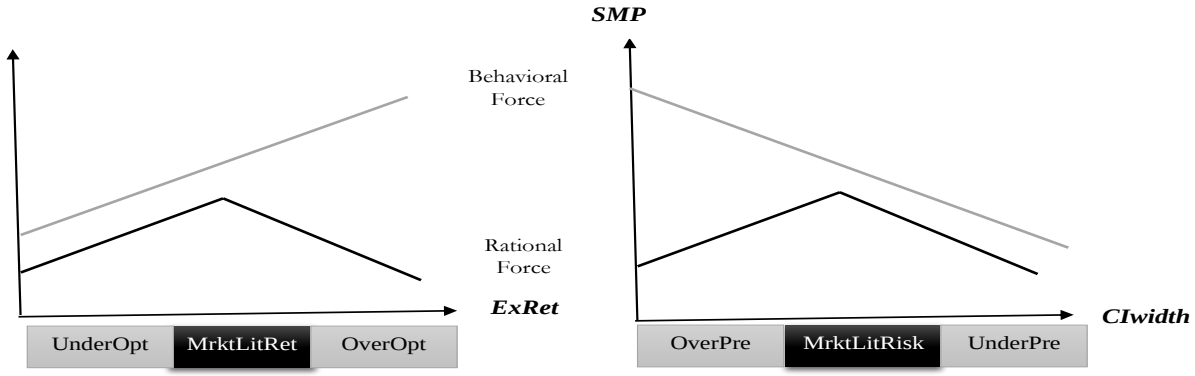


Figure A2
Predicted SMP and Number of Respondents for Different *ExRet* Ranges

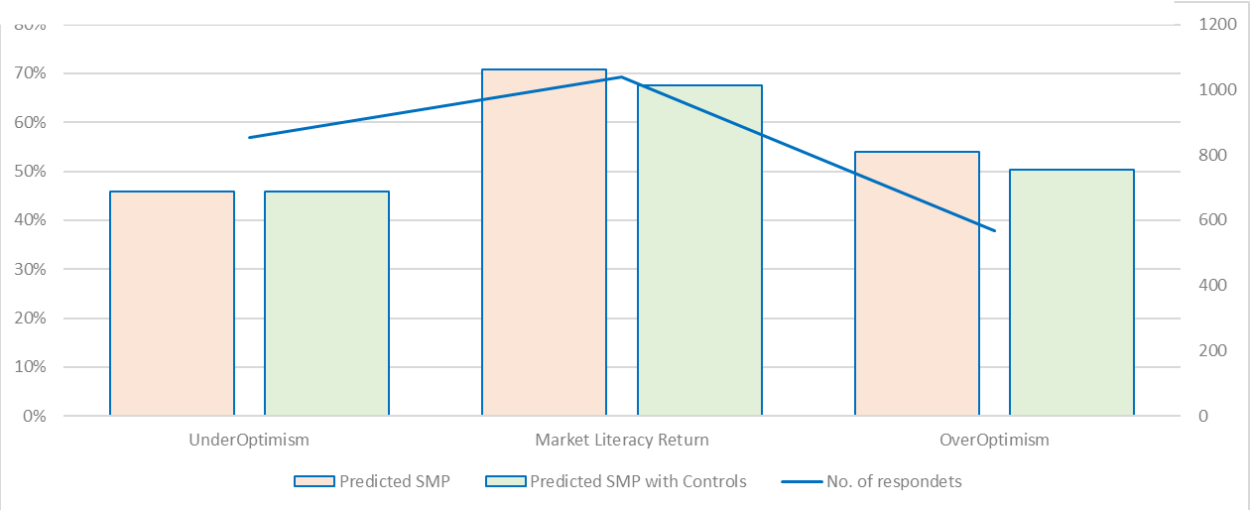
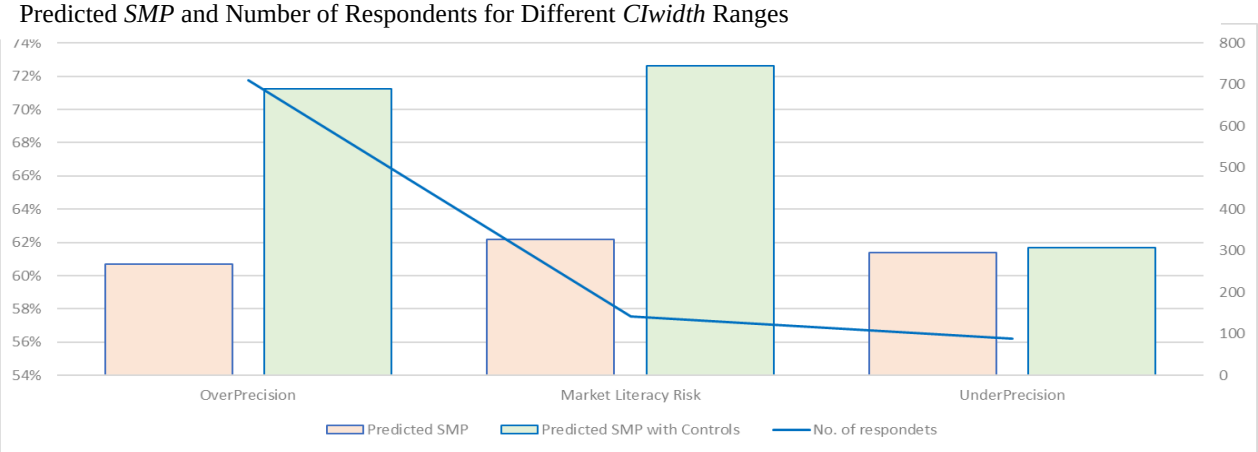


Figure A3
Predicted *SMP* and Number of Respondents for Different *CIwidth* Ranges



Conclusion

In this doctoral dissertation I present three distinct papers, the first paper contributes to the asset pricing literature and the second and third papers are related to household and behavioral Finance.

The first paper contributes to the literature on the information contents of corporate taxes in the presence of political cycles in an international setting. It studies whether the political orientation of tax decision-makers affects the informativeness of corporate tax expenses. Utilizing a sample of 10 developed countries with parliamentary system, we confirm that firms bear higher tax expenses under left-leaning governments supporting the partisan theory of political cycles even in the wake of globalization. We find that corporate tax expenses are only informative about future returns under right-leaning governments. This suggests that corporate tax expenses are arguably a more direct profitability indicator when right-leaning governments are in power.

The second chapter focuses on the realm of household finance. Using data from the Understanding America Study survey panel, we explore the direct and indirect effects of the "primitive" factors of intelligence and non-cognitive skills, derived from conscientiousness and emotional stability scores, the empirically observed stock market participation (SMP) puzzle. We find that non-cognitive skills both directly and indirectly via the previously observed proximate factors defined as general and task-specific financial literacy, education, income, and trust, have positive impacts on stock market participation. However, intelligence solely affects participation indirectly through proximate factors.

The third chapter is related to household finance and Behavioral Finance. Using data from the UAS, I find that a demonstration of knowledge about a reasonable expectation for future market return , labeled as market literacy in return, enhances the likelihood of stock holding and this effect persists even after accounting for established factors such as general and task-specific financial knowledge, intelligence, and education. Moreover, overoptimistic households are less likely to

participate since the higher expected outcome is dominated by the perceived literacy cost. I also explore the impact of market literacy in risk and overprecision on SMP. My findings suggest that the roles of market literacy in risk and overprecision on equity holding decisions are indirect and fully subsumed by other drivers of SMP.