

## Participation Puzzle Revisited: Panel Evidence from England

Ashok Thomas<sup>†</sup> Luca Spataro<sup>‡</sup>

### Abstract

We analyse the determinants of the decision to enter the stock market in England through a panel analysis on data drawn from the English Longitudinal Survey of Ageing dataset, for years 2002-2012. For doing this we use several methodologies including a probit model controlling for both unobserved heterogeneity and serial correlation through Correlated Random Effects, Generalized Estimating Equations and Generalized Linear Models. Additionally, the endogeneity of financial literacy is controlled for by using the Control Function approach. Risky asset participation is found to be positively and significantly related to financial, with an average partial effect of 5.8%. The education quality (proxied by student-teacher ratios) and the financial incentives observed at early ages (captured by the sharpe-ratios observed by individuals at early adult life) play a significant role as well. As for other individual variables, both financial resources and social interaction affect positively the probability to join the stock market. Evidence on the fact that individual non-cognitive ability e measured in terms of trust also matters for participation, further exist.

**JEL classification codes:** C23, D14, G11, I25

**Keywords:** stock market participation, financial literacy, education quality, social interaction, panel analysis.

---

<sup>†</sup> Assistant Professor, Indian Institute of Management, Kozhikode, India Email: [ashok.thomas@iimk.ac.in](mailto:ashok.thomas@iimk.ac.in) , tel 0495-2809128

<sup>‡</sup> Professor in Public Economics, Dipartimento di Economia e Management, University of Pisa, Via Cosimo Ridolfi, Italy email: [luca.spataro@unipi.it](mailto:luca.spataro@unipi.it), Tel : 050-2216224

We thank the participants of SSES annual conference on Household Finance at St; Gallen, Pencon 2018 at Lodz, the seminar participants at Department of Economics and Management at University of Pisa for their valuable comments and in particular Thomas Renström, Davide Fiaschi, Alessandro Belmonte, Nebojsa Dimic for helpful insights and suggestions on a previous version of this paper.

## 1. Introduction

The present work analyses the determinants of stock market participation in England in the first decade of the present century.

Although the issue of stock market holding has been largely analysed, there is still much concern about the low rates of participation throughout the world, even in the face of new financial services and products that should have facilitated a higher degree of portfolio diversification among households. In fact several contributions have pointed out that there can be considerable welfare loss in non-participation of individuals, in the form of reduced returns to household saving and lesser asset accumulation. Higher participation rates could also favor a greater breadth and depth of capital markets, which are important determinants of the equity premium and of the stock market volatility (Vissing-Jørgensen, 2002; Brav et al., 2002). Moreover, reforms of pension systems are increasingly shifting the responsibility for retirement saving from governments to individuals. Hence, unveiling the determinants of stock market participation has relevant policy implications, since it can help removing the barriers to efficient portfolio diversification.

By building on some consolidated findings of previous literature, the contribution of our work is twofold.

First of all, we provide new empirical evidence on the determinants of stock market participation in England drawing the data from the panel component of the English Longitudinal Survey of Ageing (hereafter ELSA) for the years 2000-2012. Apart from Brown et al (2016), using a Bayesian framework to explore the relationship between social interaction and household liabilities & assets using British Household Panel Survey, panel studies exploring household financial decisions are none. Therefore, our study using the ELSA panel data to determinant of risky asset participation is expected to contribute to this arena. In fact, earlier studies are either cross-sectional analyses at country level (Van Rooij et al, 2011; Yoong, 2011) or cross-country analyses (Thomas and Spataro, 2018; Christelis et al, 2010). By extending the analysis to a panel framework we can disentangle age and cohort effects on portfolio choice behavior, control for the time-invariant unobserved heterogeneity and model a dynamic relationship among the variables.

The second contribution of our work is to apply to portfolio choices the methodology pioneered by Papke and Wooldridge (2008) for estimating fractional response models for panel data with a large cross-section and few time periods. More precisely, by this approach we can take into account possible endogeneity of financial literacy and allow for time-constant unobserved effects to be correlated with explanatory variables.

In fact, several studies have argued that financial literacy<sup>§</sup>, while being a significant variable influencing the decision to join the stock market, is endogenous in nature. As pointed out, among others, by Kimball and Shumway (2006), Christelis et al. (2010), Van Rooij et al. (2011), on one hand, financial literacy helps alerting individuals about the excess returns on stocks/bonds, which in turn induces them to invest in risky activities; on the other hand, investing in advanced financial products could provide some kind of financial literacy training. Additionally, this positive correlation may reflect the fact that financial literacy is not distributed randomly in the population and those who possess high levels of literacy are likely to have

---

<sup>§</sup> For a detailed review on the role of financial literacy on individual financial decision-making, see Lusardi and Mitchell (2014).

certain characteristics, often unobservable, such as talent, ability, or patience that may lead also to “better” financial decisions.

Following these lines we find that endogeneity of financial literacy causes a negative bias of the associated average partial effect, which, after controlling for endogeneity, increases from 1.3% to 5.8%. Moreover, we find that the observed hump shape in the age profile of participation rates and discussed in previous works (Poterba and Samwick, 2001; Gomes and Michaelides, 2005; Alan, 2006) is in fact the result of a variety of effects. Once controlling for these factors, the participation in financial markets turns out to be an increasing function of age. Cohort effects captured by the quality of education and sharpe-ratios observed at young ages play a role in explaining the attitude towards stock market of different cohorts. As for other individual characteristics, financial resources affect positively the probability of joining the stock market, pointing to the presence of entry costs, while marital status and gender are not significant in explaining stock ownership in England. Finally, the presence of social interaction, trust and self-satisfaction increase the probability to own stocks.

The paper is organized as follows: in Section 2 we review the existing literature on the topic of stock market participation and in Section 3 we present the data. In Section 4 we lay out the empirical strategy and in Section 5 we present the results. Section 6 concludes.

## 2. Review of the literature

Existing studies have highlighted several determinants of stock market participation. Empirical evidence of industrialized countries provided by Guiso et al. (2003) documents a relevant positive correlation between stock market participation and household financial wealth, supporting the entry costs thesis (see also Alan, 2006). Other studies have suggested that participation depends on a variety of factors, including age<sup>\*\*</sup> and education (Bertraut, 1998), risk aversion (Campbell and Cochrane, 2000), social interaction (Hong et al., 2004), home ownership (Vestman, 2013), and social capital (Guiso et al., 2004).

Other works have shown that education and financial literacy play a role in stock market participation. For example, Guiso and Jappelli (2005), Kimball and Shumway (2006) Van Rooij et al. (2011, 2012) find that lack of awareness of stocks is a primary reason for the limited participation.

As for the role of general education, several authors have shown that colleges educated are more likely to own stocks than less educated individuals (Haliassos and Bertaut 1995; Campbell, 2006; Lusardi and de Bassa Scheresberg 2013). Cole and Shastry (2008) argue that one year of schooling increases the probability of financial market participation by 7-8%. On the same lines some empirical studies on stock holding have shown that including control for educational attainment does enhance the significance of the variable financial literacy (Van Rooij et al. 2011, Behrman et al. 2012) underlying the fact that general knowledge (education) and specialized knowledge (financial literacy) both contribute for financial decision making, both in Netherlands and United States.

One of the main drivers of the risky asset market participation is the influence of financial literacy, a specialized cognitive skill. As far as the questions concerning the channels

---

<sup>\*\*</sup> The hump shape in the age profile of people who hold risky assets has been much discussed (see Alan, 2006; Ameriks and Zeldes, 2004; Poterba and Samwick, 2001; Gomes and Michaelides, 2005).

through which financial literacy influence financial decision like risky asset participation is concerned, two main channels are discovered by existing literature. Calvet (2014) and Grinblatt (2011) adhere to non-preference driven explanations, i.e., quality of financial decisions improve with higher cognitive ability (financial literacy). On the other hand, scholars including Frederick (2006), Dohmen et al., (2010), Beauchamp et al., (2017) point towards the effect on risk preferences, reflecting the fact that higher financial literacy is related to lower risk aversion ((and the likelihood for participation increasing).

We take the latter channel and assume that with higher financial literacy increases the quality in information processing (in line with Grinblatt et al., 2011 and Calvet et al., 2007) leading to more precise estimates of subjective expected returns and risks which thereby increase the likelihood for participation.

Among other individual characteristics, Arrondel et al., (2012) point out that stock ownership significantly correlates with both expectations and realizations of stock market returns. In fact, the recent research by Giuliano and Spilimbergo (2014) indicates that generations who grew up in economic recessions have systematically different socio-economic beliefs compared to generations who grew up during boom periods. On the same lines Malmendier and Nagel (2011) and Thomas and Spataro (2018) argue that households experiencing higher stock returns early in life are more likely to participate in stock market.

Another stream of literature has explored the gender bias in stock ownership. The pioneering work by Haliassos and Bertaut, (1995) provided the empirical evidence of limited participation of female workers, while others have gone one step ahead in explaining the reasons behind the phenomenon, resorting to women's higher risk aversion (see, among others, Croson and Gneezy 2009 or Bertrand 2011 for reviews) and to differences in the production processes for financial literacy across genders (Fonseca et al. 2012), due to the household specialization.

The lack of social interaction and reduced general intensity of participation are also found to affect individuals' decision to enter risky markets (Hong et al., 2004; Duflo and Saez, 2003; Brown and Taylor, 2007; Brown et al., 2008; Liu et al., 2014). For example, Hong et al. (2004) from HRS survey in U.S. find evidence that a "social" investor finds the market more attractive when more of his/her peers participate. Brown et al. (2008) find a positive link between individual's decision to participate and the average level of stock market participation present in the individual's social group/community.

Finally, there is a recent literature discussing the role of health status on the portfolio decisions of respondents. Rosen and Wu (2004) analyze the role of health status on household portfolio decisions using self -perceived health status data from the Health and Retirement Study (HRS) in the U.S. and find a positive relationship between better health status and portfolio choice. Edwards (2008) shows that retired individuals view their health status to be risky and try to hedge against it by decreasing their exposure to financial risk.

To sum up, it is a well-established fact that not all households participate in risky asset markets. The empirical studies mentioned above are cross-sectional analyses at either country or cross-country level. However, the respondents of different age vary in unobservables, which are correlated with age, and therefore the estimated age pattern may reflect only cohort effects. Hence, by using a panel framework, in the present work we aim to disentangle age and cohort effects. Moreover, we also take into account the endogeneity of an explanatory variable (financial literacy) and allow for time-constant unobserved effects to be correlated with explanatory variables.

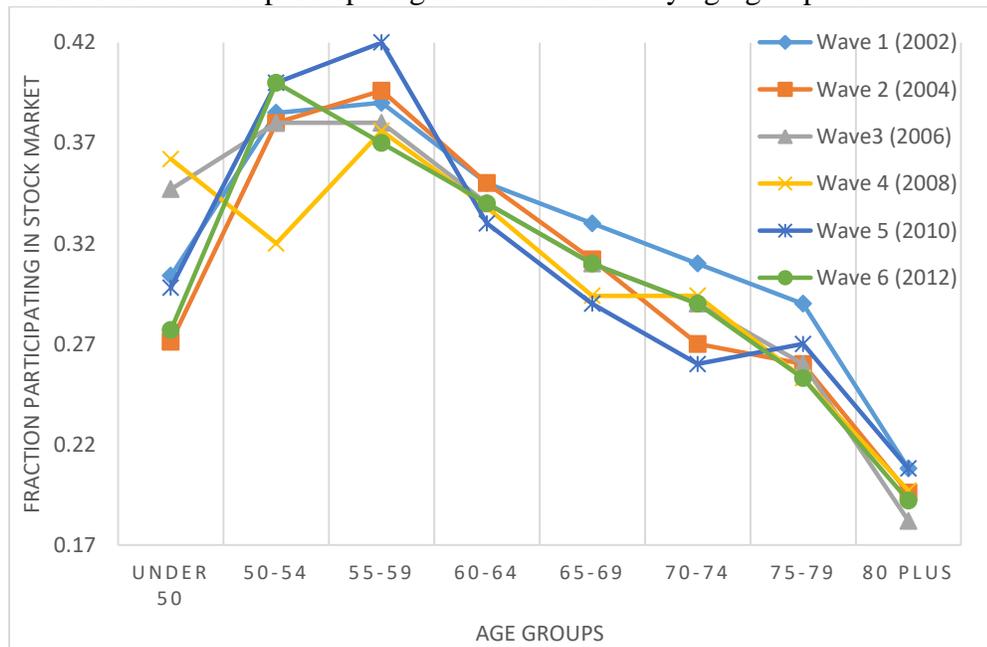
### 3. Data

The data for the present analysis are drawn from the English longitudinal Survey of Ageing (hereafter ELSA) from year 2002 to 2012. This is a longitudinal survey on a large representative sample of men and women living in England, designed to understand the implications of ageing and containing information on demographic factors, economic circumstances, social and psychological variables, health, cognitive function and biology. The study began in 2002 and the sample was re-examined every two years. Our sample consists of 5064 individuals who were interviewed in all waves (for the sake of simplicity in estimation we focus on a balanced panel). Individuals who have exited the survey because of death or migration and new entrees are not included. Overall, the sample contains 30506 individual/year observations (all variables are summarized in Table 1 and Appendix 2).

As for the dependent variable, *stockpart* is a binary variable which takes value 1 if the individual participates in stock market and 0 otherwise.

Figure 1 plots the age profile of the share of individuals participating in the stock market for five-year age groups and waves. Two features emerge: first, the age profile displays an inverse U shape, with peaks associated with the 50-54 and 55-59 age groups, at around 40%. Second, there has been a general drop in participation rates, after 2002, especially at the tails of the age-distribution. We also split the age group into working and retirees to separate the effect of retirement on risky asset participation.

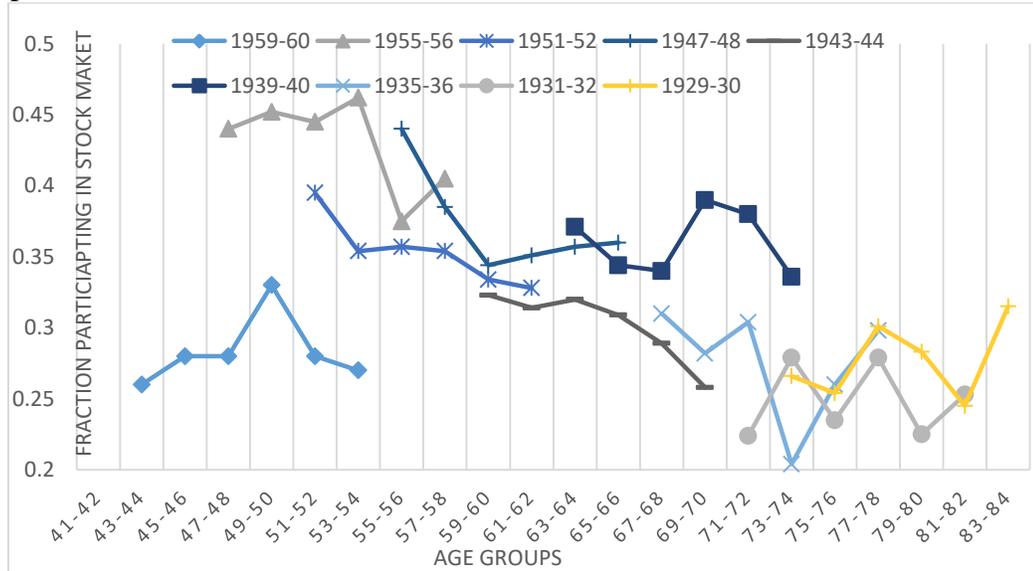
**Figure 1:** Fraction of individuals participating in stock markets by age-groups and waves



Source: Authors' calculation

Figure 2 presents the time path of the share of individuals participating in the stock market for selected two-year cohort groups and it confirms the general negative time-trend, with a recovery in year 2008. Moreover, negative cohort effects are particularly relevant for younger cohorts (born after 1959) and for individuals that were born before the II World War.

**Figure 2:** Time path of fraction of individuals participating in stock markets by cohort-groups

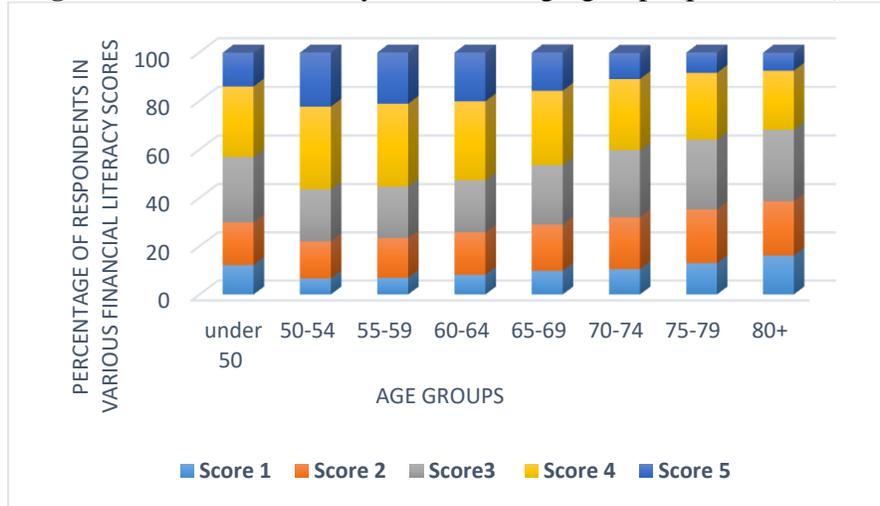


We try to capture such cohort effects through two variables. The evidence provided in previous works suggests that variations in experienced stock market returns can effectively capture the cohort effects. Following these lines, we use average sharpe-ratios observed between ages 18-25 as a proxy for cohort effects (i.e. five-year cohort groups) by using the data from historical stock returns of United Kingdom.

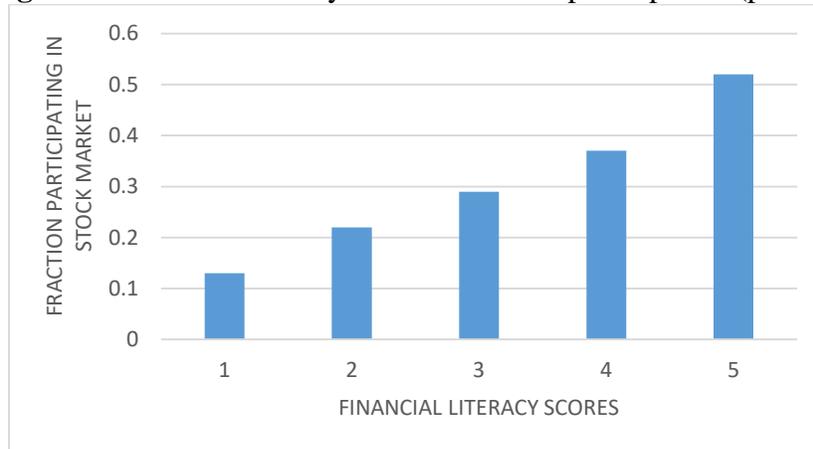
Moreover, following the insights contained in Thomas and Spataro (2018), we include a variable capturing the effectiveness of education (Education quality effect), proxied by the student-teacher ratio that a respondent experienced during her childhood (6-15). Indeed, as education quality increases in a country, both individual and social capital improve; given the positive effect of human capital on participation, there are higher possibilities for an agent to find individuals in the same cohort group that are engaged in stock markets (peer effect). Hence, from the International Historical Statistics on Education, the 10-year average student-teacher-ratio is calculated for each individual belonging to a specific five-year-cohort group.

Among other explanatory variables, financial literacy is of primary interest. By following Jappelli and Padula (2013) we use the index provided by ELSA, whereby each individual is presented with four or five financial and numerical questions and the answers are collected to obtain a value ranging from 0 to 5. Details of the actual questions and the construction of this indicator are given in the Appendix 4 is a slightly modified version of Lusardi and Mitchell (2011).

**Figure 3: Financial literacy scores and age-groups (pooled data)**



**Figure 4: Financial literacy scores and stock participation (pooled data)**



Our main measure of financial literacy is based on the answers to a series of questions related to (basic) principles of household finance. The measure follows the spirit of the methodology proposed by Lusardi and Mitchell (2011). The questions from which we calculate financial literacy is asked in wave 1, wave 4. The cognition module used in wave 2, wave 3, wave 5, wave 6 are very similar to the ones used in wave 1, wave 4 (though the specific questions on numeracy and prospective memory questions are dropped). This method of dropping the numeracy questions is aimed to ward off any possible similarity with the protocols used in the earlier waves which might create bias in the measurement arising from recalling the nature of questions by the respondents. Additionally we find across waves, there is only slight change in the financial literacy scores with the summary statistics across wave 1, 3 and 5 (see appendix 1). Therefore using the financial literacy scores of Wave 1 for wave 2 and wave 3 and using the financial literacy score of wave 4 for wave 5, wave 6 is attempted in our analysis.

Interestingly, the age profiles of financial literacy and stock market participation show a similar pattern. From Figs. 3 and 4 it emerges that both participation to stock market and

financial literacy scores peak before the period of retirement and then fall with age. Hence, a higher financial literacy score also reflects a higher participation to stock market.

We also take into account financial variables like income and wealth quintiles, given that the latter are considered as good predictors of stock market participation (Van Rooij et al, 2011, Thomas and Spataro, 2018), also due to the presence of entry costs.

The social interaction variables are also likely to have significant bearing over the decision to enter the stock market. Hence, we include a variable depicting the lack of social interaction (i.e a dummy variable taking value 1 if the respondent is not taking part to any social, religious or organisational groups and 0 otherwise).

Participation in stock markets is also influenced by non-cognitive ability, through the non-preference channel is loosely related to work of Guiso et al (2008) who assumes that trust affects beliefs in data used to form subjective expectations. Borrowing from, Gyllenram et al. (2013), Olga Goldfayn-Frank (2018) on the influence on non-cognitive ability on stock participation we incorporate the personality dimension of agreeableness by proxing, “trust”, which is a variable ranked from 1 to 7 depending on one’s perception of trust on others (1 means almost none in this area can be trusted). A higher trust among respondents is expected to have a positive effect on the decision to join the stock market. Guiso et al. (2008) explore the relationship between trust and stock market participation and find that less trusting individuals are less likely to purchase stocks in a setting controlling for risk and ambiguity aversion. More recently, Christelis et al. (2010); (Georgarakos and Pasini, 2011) found that higher trust individuals are more likely to own shares.

**Table 1:** Sample statistics of all variables from ELSA balanced panel

<b>VARIABLES</b>	<b>Obs</b>	<b>Mean</b>	<b>Standard deviation</b>	<b>Min</b>	<b>Max</b>
Stock ownership (dependent variable)	30,372	0.316	0.465	0	1
Financial literacy scores	30,372	3.199	1.253	0	5
Age	30,368	66.50	9.842	20	99
Age <sup>2</sup>	30,368	4,519	1,356	400	9,801
Dummy for married	30,372	0.602	0.489	0	1
Dummy for female	30,372	0.578	0.494	0	1
Income quintiles	28,343	3.104	1.396	1	5
Wealth quintiles	28,343	3.269	1.375	1	5
Dummy for no social interaction	30,372	0.338	0.473	0	1
Trust	29,136	2.658	1.656	1	7
Self-perceived social status	27,960	57.65	18.54	0	100
Self-perceived health status	27,458	2.626	1.075	1	5
Education quality	30,038	26.02	1.121	18.71	29.52
Average sharpe-ratios at early adult life	30,017	0.198	0.0933	-0.0259	0.385

See Appendix 2 for details

The measures of sociability (participation in social activities and trust) could reflect measurement problem. In fact, these variables may not only capture information on the degree of social interaction but also other personality traits associated with the propensity to invest in the

stock market. For example, more socially interacting respondents may have traits like boldness, risk –taking and optimism, qualities which are likely to enhance financial market participation (see also the discussion in Hong et al 2004). Consequently, in order to pinpoint the effect of social interaction we also include a proxy for personality/psychological traits, namely self-reported life satisfaction (“self-perceived social status”). This variable can be linked to optimism, which has been studied in prior research<sup>††</sup>. The link between these two characteristics is relatively obvious since optimistic persons are most likely to be more satisfied with their lives and more likely to take active steps to improve their current or future situation. Consequently, individuals reporting higher levels of life-satisfaction are expected to have a greater tendency to choose risky options (Weinstein, 1980, 1984). Therefore, we include this variable which ranks from 0 to 100 in our analysis (100 is the highest level of reported life-satisfaction).

In line with previous literature, we also include self-perceived health status in our model as a potential predictor of stock market participation, with negative expected sign.

We are also interested in considering the effect of demographic variables like marital status and gender. The latter two variables are specified by dummies that equal to 1 if the person is married and if the person is a female, respectively. Finally time dummies are also included. Sample statistics of all variables are summarized in Table 1 (time dummies are omitted).

## 4. Empirical methodology

### 4.1. Overview

In this Section we discuss the choice of the estimation strategy that we adopted to investigate the determinants of portfolio choice outcomes of English individuals belonging to the panel component.

Panel data models usually are affected by such issues as unobserved heterogeneity and omitted time-varying variables, which cause biased estimations. Traditionally, unobserved heterogeneity is treated as parameters to estimate, when T (time periods) is large: in fact, under fairly weak assumptions, one could obtain consistent asymptotically normal estimators of average structural functions, provided suitable instruments are found. However, when T is small, such a methodology can lead to the incidental parameter problem (that is, lack of convergence of estimators) and additionally the bias could be higher when weak dependence or even independence is assumed across the time dimension (for an insight into these issues see Hardin and Hilbe 2007; Wooldridge 2002).

To overcome these issues in this work we follow the methodology pioneered by Wooldridge (2002) which clarifies how to specify and estimate fractional-binary/response models for panel data with a large cross-section and few time periods (the so called correlated random effect<sup>‡‡</sup> model using the Chamberlin-Mundlak device) under strict exogeneity and serial

---

<sup>††</sup> Life satisfaction is a relatively new characteristic to household finance, and as expected, it operates similarly as moderate optimism, which has been found to result in sensible economic decision making and tendency to search information on risky assets (Felton et al., 2003; Puri and Robinson, 2007).

<sup>‡‡</sup> In the exogenous case, among different methodologies like Fixed effects (FE) or Conditional Maximum Likelihood Estimation (CMLE) approaches, we prefer the correlated Random Effect approach (CRE). In fact the CRE approach, puts a restriction on the conditional distribution of heterogeneity compared with FE or CMLE

independence. We further take insight from the works by Papke and Wooldridge (2008) and Wooldridge (2005), which provide a methodology, again under exogeneity assumption, to identify the average partial effects without the conditional serial independence assumption, by using the Bernoulli- quasi MLE and generalised estimating equation. Finally, we relax the assumption of strict exogeneity and we resort to a Control Function approach employing the two stage probit QMLE proposed by Papke and Wooldridge (2008).

#### 4.2. Estimation methods under strict exogeneity

We start with a standard specification of a static unobserved effects probit model for panel data, which can be written as:

$$E(y_{it} | x_{it}, c_i) = \Phi(x_{it}\beta + c_i), t = 1, \dots, T \dots \dots \dots (1)$$

where  $y_{it}$  is the binary variable (stock market participation in our case) and  $x_{it}$  is the vector of explanatory variables,  $\Phi$  is the standard normal cumulative distribution function and  $c_i$  are the unobserved effects. Note that the magnitude of partial effects not only depend on the value of covariates  $x_{it}$ , but also on unobserved heterogeneity  $c_i$ . Thus to identify  $\beta$  or the average partial effects (APEs) in presence of unobserved effects, we require some further assumptions.

The first assumption is the exogeneity of  $x_{it}$ : conditional upon  $c_i$ , we assume that  $x_{it}$  remains exogenous, so that we can write the following:

$$E(y_{it} | x_i, c_i) = E(y_{it} | x_{it}, c_i), t = 1 \dots \dots T \dots \dots \dots (2)$$

The second assumption concerns the distribution of  $c_i$  given  $x_i$ . We follow the Chamberlain (1980)-Mundlak (1978) approach by assuming conditional normality of unobserved effects as follows:

$$c_i = \psi + \bar{x}_i \zeta + a_i, a_i | x_i \sim \text{Normal}(0, \sigma_a^2) \dots \dots \dots (3)$$

where  $\bar{x}_i$  is the vector of time averages and  $\sigma_a^2 = \text{Var}(c_i | x_i)$  is the conditional variance of  $c_i$ . Notice that assumptions in equations (1), (2) and (3) do not impose any additional restrictions on the distribution  $D(y_{it} | x_i, c)$  nor on the serial dependence in  $y_{it}$  and allow to identify the average partial effects (APE).

This setting is called a ‘‘Correlated Random Effects (CRE) probit’’ model, which we adopt in the first part of our empirical study. Under the assumption of exogeneity of explanatory variables, as a robustness check, we use other two models in which APEs are identified without the conditional serial independence assumption: first of all, the pooled Bernoulli quasi MLE model (or Probit QMLE model); second, also to possibly enhance efficiency, a generalised estimating equation approach (GEE). As for the latter approach, we report the results obtained by using an exchangeable working correlations matrix<sup>§§</sup>, given that the ones obtaining from independent correlation matrix are not significantly different.

---

approach. However in the balanced panel case, the CRE approach requires few other assumptions for estimating average partial effects, and the restrictions needed on the conditional heterogeneity distribution can be fairly weak making it the most preferable choice in this framework. For a discussion see e Altonji and Matzkin (2003).

<sup>§§</sup> Briefly, GEE applied to panel data is essentially weighted multivariate non-linear least squares with explicit recognition that the weighting matrix might not be the inverse of the conditional variance matrix. The ‘‘working’’ correlation matrix is usually specified as ‘‘exchangeable’’. See Zeger and Liang (1986), Ballinger (2004), Hardin (2005).

4.3. Estimation methods with endogenous explanatory variables

We now briefly show how we treat endogeneity in the presence of unobserved heterogeneity and omitted time-varying variables. The Control Function methodology adopted by Papke and Wooldridge (2008) for the case of fractional response model is recast here under a binary response variable. As noted above, the results obtained from this approach are robust and comparable to the models shown in Section 4.2, where endogeneity is ignored.

Suppose  $y_{it2}$  is an endogenous explanatory variable; provided we have sufficient instruments, we can express the conditional mean model as:

$$E(y_{it1} | y_{it2}, z_i, c_{i1}, v_{it1}) = E(y_{it1} | y_{it2}, z_{it1}, c_{i1}, v_{it1}) = \Phi(\alpha_1 y_{it2} + z_{it1} \delta_1 + c_{i1} + v_{it1}) \dots \dots \dots (4)$$

where  $y_{it1}$  is binary,  $c_{i1}$  is the time-constant unobserved effect and  $v_{it1}$  is a time-varying omitted factor that can be correlated with  $y_{it2}$ , the potentially endogenous variable<sup>\*\*\*</sup> (financial literacy in our case),  $z_{it}$  is the vector of exogenous variables. The traditional instrumental variable method which could provide results by eliminating  $c_i$  cannot be attempted here (as these common estimation methods eliminates  $c_i$  along with any time-constant explanatory variables) and therefore the CRE approach of modelling the distribution of unobserved heterogeneity,  $D(c_{i1}|z_i)$ , is again attempted. Additionally, one has to model how  $y_{it2}$  is related to  $v_{it1}$ . Control Function approach allows to deal with both issues. More precisely, we model the unobserved heterogeneity as a linear function of all exogenous variables, allowing the instruments to be correlated with time-constant omitted factors. Hence, assuming  $a_{i1}$  to be independent of  $z_i$ , we can write:

$$c_{i1} = \psi_1 + \bar{z}_i \zeta_1 + a_{i1}, a_{i1} | z_i \sim \text{Normal}(0, \sigma_{a1}^2) \dots \dots \dots (5)$$

Plugging equation (5) into (4) we get:

$$E(y_{it1} | y_{it2}, z_i, a_{i1}, v_{it1}) = \Phi(\alpha_1 y_{it2} + z_{it1} \delta_1 + \psi_1 + \bar{z}_i \zeta_1 + a_{i1} + v_{it1}) = \Phi(\alpha_1 y_{it2} + z_{it1} \delta_1 + \psi_1 + \bar{z}_i \zeta_1 + r_{it1}) \dots \dots (6)$$

Assuming a linear reduced form of the suspected endogenous variable  $y_{it2}$  we get:

$$y_{it2} = \psi_2 + z_{it} \delta_2 + \bar{z}_i \zeta_2 + v_{it2}, t = 1, \dots, T \dots \dots \dots (7)$$

with  $r_{it1} = a_{i1} + v_{it1}$ . The addition of time averages of strictly exogenous variables  $\bar{z}_i$  in eq. (7) follows a Mundlak (1978) device. As for the source of endogeneity of  $y_{it2}$ , it stems from the relationship between  $v_{it2}$ , the reduced-form error term and the new term  $r_{it1}$  in eq. (6). Thus  $y_{it2}$  is allowed to be correlated with unobserved heterogeneity and time-varying omitted factors.

Compared to the estimation method where every explanatory variable is considered exogenous as in equation (1), we explicitly allow contemporaneous endogeneity in equation (7), while also allowing for possible feedback from unobserved idiosyncratic changes in  $y_{it1}$  as captured by  $v_{it1}$ . Finally, since we do not assume strict exogeneity of the endogenous variable, the GEE estimation is inconsistent and thus we employ the Pooled Probit method in the second - stage estimation. To sum up, the two-step procedure employed is as follow:

- (1) Estimate the reduced form of  $y_{it2}$  (pooled across t) and obtain the residuals  $\hat{v}_{it2}$

---

<sup>\*\*\*</sup> In a cross-sectional context this Control Function approach has been employed by River and Voung (1988) for a binary response variable.

(2) Use the probit QMLE of  $y_{it1}$  on  $y_{it2}$ ,  $z_{it1}$ ,  $\bar{z}$ ,  $\hat{v}_{it2}$ .

#### 4.4 Application of methodology to dataset

Analogous to equations (1) and (4) we build the econometric model

$$\text{Stockpart} = \alpha_1 + \beta_1 \text{FL} + \beta_2 \text{Age} + \beta_3 \text{Agesq} + \beta_4 \text{MS} + \beta_5 \text{FE} + \beta_6 \text{WE} + \beta_7 \text{IN} + \beta_8 \text{SI} + \beta_9 \text{TR} + \beta_{10} \text{SS} + \beta_{11} \text{HE} + \beta_{12} \text{EQ} + \beta_{13} \text{SH} + \beta_{14} \text{WE}^{\text{BAR}} + \beta_{15} \text{IN}^{\text{BAR}} + \beta_{16} \text{TR}^{\text{BAR}} + \beta_{17} \text{SS}^{\text{BAR}} + \beta_{18} \text{HE}^{\text{BAR}} + \beta_{19} \text{DUM}_{2002} \dots + \beta_{22} \text{DUM}_{2010} + v_i \dots \dots \dots (8)$$

where FL is financial literacy scores and the usual demographic variables like age, age squared, marital status (dummy for married, MS), FE (dummy for female) and the self-perceived health status (HE) are included. The income and wealth quintiles variables IN and WL are also kept in the regression, together with self-perceived social status (SS), meant to capture psychological traits of individuals (such as optimism). The social interaction variables include trust (TR) and a dummy for the lack of social interaction (SI). EQ is education quality proxied by the average student/teacher-ratio at cohort level (five-year-cohort groups), when individuals were within their 6-15 age interval. SH is the average sharpe-ratio observed by respondents, grouped into five-year cohort groups, when they were between 18-25 years of age. The time averages of the time-varying variables (with subscript BAR in eq. 8): income, wealth, self-perceived social status, self-perceived health status and trust are allowed to be correlated with the individual unobserved effect. Finally, the time dummies for years 2002 to 2010 are added (2012 is the omitted dummy).

As anticipated, we aim to take care of endogeneity of financial literacy and of the unobserved heterogeneity, on one hand, and to compare such results with those emerging in the case of assumed exogeneity of financial literacy, on the other hand. Thus we augment the traditional instrumental variable approach by including the time averages of the time varying variables, allowing them to be correlated with the individual-level unobserved heterogeneity.

The reduced form of the financial literacy is:

$$\text{FL} = \eta_i + \pi_1 \text{books} + \pi_2 \text{rooms} + \pi_3 \text{diMAedu} + \pi_4 \text{Age} + \pi_5 \text{Agesq} + \pi_6 \text{MS} + \pi_7 \text{FE} + \pi_8 \text{WE} + \pi_9 \text{IN} + \pi_{10} \text{SI} + \pi_{11} \text{TR} + \pi_{12} \text{SS} + \pi_{13} \text{HE} + \pi_{14} \text{EQ} + \pi_{15} \text{WE}^{\text{BAR}} + \pi_{16} \text{IN}^{\text{BAR}} + \pi_{17} \text{TR}^{\text{BAR}} + \pi_{18} \text{SS}^{\text{BAR}} + \pi_{19} \text{HE}^{\text{BAR}} + \pi_{20} \text{DUM}_{2002} \dots + \pi_{24} \text{DUM}_{2010} + v_i \dots \dots \dots (9)$$

where books, rooms, diMAedu are number of books in the shelf at age 10, number of rooms of the house individual lived in at age 10 and mother's education when the individual was 10 year old, respectively<sup>†††</sup> (see Table 2 for summary statistics of the instruments). In using these instruments, which are indexes of family background and level of intergenerational cognitive ability, our identification assumption is that stock market participation depends on unobserved heterogeneity in a smooth fashion and the relationship between heterogeneity and the instruments is smooth (see Section 5.2 for details). Given the strength of the instruments, we then estimate equation (9) by instrumental variable and Probit QMLE approach.

---

<sup>†††</sup> The specification reported here is most efficient one. Other instruments which we checked for robustness are self-reported health at age 10, occupation of the main breadwinner at age 14. Upon estimation, we find that all instruments are exogenous in nature and their strength is fair.

**Table 2:** Sample statistics of instruments

INSTRUMENTS	Obs	Mean	Standard error	Min	Max
Mother's education at age 10	30,372	2.309	1.159	1	7
Number of rooms at age 10	27,834	2.915	0.935	1	15
Number of books at age 10	27,834	2.490	1.210	1	5

Note: for details see Appendix 2 and 4.

## 5. Results and discussion

### 5.1. Results and discussion of the empirical model with financial literacy is exogenous in nature

In this Section we present and discuss the results under the assumption of exogeneity of all explanatory variables. Table 3 contains the estimated coefficients and average partial effects (APE) of the correlated random effects model (CRE), the pooled QMLE and the GEE estimation models respectively. The three sets of estimates tell a consistent story: with the exception of three coefficients (i.e. dummy for female, the self-perceived social status and self-reported health), all other variables are significant with expected signs.

According to the CRE model, financial literacy has a positive and significant effect on the portfolio decision choice. One standard deviation of financial literacy increases the probability to invest in stock market by 1.3%. The financial variables like wealth and income quintiles have positive influence on the decision to participate in risky markets across the waves (marginal effects are 5.2% and 2.0% respectively). The cohort effect proxied by the sharpe-ratio observed between ages of 18 to 25 shows a positive influence on stock participation, suggesting that individuals who observed a bullish market during their earlier years have higher probability to own stocks (marginal effect is 18.3%). The school effect by which we measure the education effectiveness at cohort level (student-teacher ratio) has the expected negative sign and the coefficient of -2.1% implies that deterioration in education quality provides less impetus for individuals to participate in stock market, either directly or through peers effect. The social interaction variable, dummy for lack of interaction are significant and with the expected sign, with marginal effects of -0.3% respectively. The non-cognitive ability is proxied by trust is with an expected sign and across waves, lower trust reduces stock market participation by 0.5%.

Among the demographic variables, dummy for working age displays a positive sign, while dummy for retired representatives shows an expected negative sign. The hump-shape of participation rates depicted in Figure 1 is a combination of different effects, while dummy for married shows a positive estimated coefficient.

Results from the Pooled QMLE model reveal that the APE for financial literacy is 1.4% and is statistically significant (fully robust t statistic= 5.20). The estimate of the GEE approach is close to the one stemming from the CRE model (APE around 1.3%). Interestingly, as for the financial literacy coefficient, the fully robust standard error for the probit QMLE estimate (0.002) is lower than the fully robust standard error for the GEE estimation (0.001, column 5). Hence, in this work, using exchangeable working correlation matrix in multivariate weighted

non-linear least squares estimation appear to enhance efficiency. The socio-demographic variables like age and marital status display a similar effect as in the CRE model and the estimated coefficients for female, self-perceived social status and self-perceived health status remain insignificant.

Wealth and income quintiles for the last two models also show a positive effect on stock market participation. The APEs of the income quintiles are 1.5% and 1.7% in the probit QMLE and GEE models, respectively, with a negligible change in the robust standard errors. The cohort effect proxied by sharpe-ratio observed between 18 to 25 years of age shows the same pattern with a lower estimated coefficient but with a lower standard errors too. The school effect (lower quality of schooling) also shows a negative effect (APEs around -1.6%).

All the social interaction variables (trust and dummy for no social interaction) display the expected sign, while self-perceived social status and self-perceived health status remain insignificant in these models.

## 5.2. Results and discussion of the empirical model with financial literacy is endogenous in nature

Table 4 provides the estimated coefficients obtained from a Two-Stage-Least-Squares model (Column 1) and a Control Function approach (“Probit pooled QMLE”, Column 2, and Average Partial Effects in Column 3), which is directly comparable with the cases in which all variables were treated as exogenous in nature (Column 4 in Table 3). Notice that the number of observation is lower due to missing values of the instruments used in this model.

The first stage results using equation (9) are reported in Appendix 6. As for the choice of the instruments for correcting endogeneity, we pick up the idea that childhood experiences may be a good predictor for financial literacy (Grohmann et al 2015). As these experiences clearly happened in the long past, their direct effect on financial decisions today should be of little concern. Thus we compile a set of possible (instrumental) variables, most of them suggested in the literature, and examine which ones may be important in explaining financial literacy. Among the various possible instruments we select number of rooms and number of books in the shelf at age 10, commonly used in the literature (Jappelli and Padula, 2013; Thomas and Spataro 2018). Also education of mother at age 10 may be seen as proxy for positive early childhood experiences, which are important for favorable later outcomes (Carneiro and Heckman, 2007 and Heckman, 2006; Carneiro et al 2013). The results of the first stage regression show that these instruments exert a positive effect on financial literacy acquisition and are significant at 1% level. As for the relevance of the instruments (signifying the fact that they influence the suspected endogenous regressor) we observe the F-statistics are high and above the value recommended to avoid the weak instrument problem (Staiger and Stock 1997) as reported in the first stage regression. Given that our instruments are strong and overcome the exclusion restriction (Hansen J statistic) we estimate the model using equation (9) in the first step and the estimated coefficients from the second stage regression (eq. 8) are reported in Column (1) of the Table 4. We also report the  $\chi^2$  to obtain the Hausman (1978) test for endogeneity; its fully robust t statistic is -1.92, providing evidence that financial literacy is endogenous in nature. Comparing the estimates in which financial literacy was considered exogenous, the financial literacy coefficient improves from 0.7% to 7.9%. e and serial correlation)

**Table 3:** Estimates assuming Financial Literacy is conditionally strictly exogenous. Dependent (binary) variable: stock market participation. (Standard errors in parentheses).

VARIABLE	Correlated Random effects Probit model coefficient	APE	Probit Pooled QMLE Coefficient	APE	Probit GEE Coefficient	APE
Financial Literacy	0.0770*** (0.021)	0.0153*** (0.003)	0.0528*** (0.008)	0.0147*** (0.001)	0.0486** (0.012)	0.0136*** (0.002)
Dummy for Working age	0.0633* (0.067)	0.0126* (0.004)	0.0641* (0.015)	0.0179*** (0.011)	0.0437* (0.039)	0.0064* (0.009)
Dummy for Retired Age	-0.0478* (0.067)	-0.0095* (0.010)	-0.0475*** (0.000)	-0.0133*** (0.015)	-0.0228* (0.039)	-0.0064* (0.009)
Dummy for married	0.2125** (0.045)	0.0042*** (0.007)	0.389* (0.021)	0.0108** (0.004)	0.1209** (0.033)	0.0345** (0.006)
Dummy for female	-0.0476 (0.0552)	-0.095 (0.008)	-0.0325* (0.019)	-0.0093** (0.004)	-0.0276* (0.032)	-0.007* (0.007)
Wealth quintiles	0.2660*** (0.016)	0.0529*** (0.002)	0.1750*** (0.012)	0.0487*** (0.002)	0.160*** (0.011)	0.0449*** (0.002)
Income quintiles	0.1010*** (0.014)	0.0202*** (0.002)	0.0549*** (0.010)	0.0153*** (0.002)	0.0606*** (0.009)	0.0170*** (0.002)
Dummy for no social interaction	-0.186*** (0.041)	-0.037*** (0.006)	-0.1820*** (0.022)	-0.0506*** (0.004)	-0.1120*** (0.031)	-0.0314*** (0.005)
Trust	-0.0268** (0.016)	-0.0053** (0.002)	-0.0153** (0.005)	-0.0042** (0.001)	0.0151* (0.009)	0.0042** (0.002)
Self-perceived social statu	0.0013 (0.001)	0.0002 (0.000)	0.0007 (0.000)	0.00002 (0.000)	0.0008 (0.001)	0.0002 (0.000)
Self-perceived health status	0.0072 (0.018)	0.0014 (0.003)	0.0069 (0.014)	0.0019 (0.003)	0.0040 (0.011)	0.0001 (0.002)
Education quality	-0.110** (0.034)	-0.0219*** (0.005)	-0.0480*** (0.021)	-0.0134*** (0.004)	-0.0586** (0.019)	-0.0164** (0.004)
Sharpe-Ratio	0.920***	0.183***	0.540***	0.1510***	0.5860**	0.1596***

Constant	(0.313) 0.697*** (0.0457)	(0.051)	(0.110)	(0.025)	(0.019)	(0.041)
Observations	24762		24762		24762	
Number of persons	4127		4127		4127	

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

All models contain year dummies for 2002-2012. (b) The pooled probit QMLE estimation includes time averages of the time-varying explanatory variables (c) The standard errors for coefficients in parenthesis are robust to general second moment misspecification (conditional variance and serial correlation)

**Table 4:** Estimates allowing financial literacy to be endogenous. Dependent (binary) variable: stock market participation (Standard errors in parentheses).

Variables	Linear Instrumental Variables Coefficient	Probit Pooled QMLE Coefficient	APE
Financial Literacy	0.0792** (0.019)	0.2041** (0.003)	0.0580*** (0.018)
Age	-0.0126** (0.031)	-0.0467** (0.021)	-0.0133** (0.004)
Age <sup>2</sup>	0.00008* (0.000)	0.0032** (0.000)	0.00009*** (0.000)
Dummy for married	0.0128 (0.045)	0.0429** (0.007)	0.0197** (0.022)
Dummy for female	-0.0344 (0.0575)	-0.0732 (0.064)	-0.0208 (0.013)
Wealth quintiles	0.0474*** (0.003)	0.1790*** (0.013)	0.0509*** (0.002)
Income quintiles	0.0126*** (0.003)	0.0457*** (0.011)	0.0130*** (0.002)
Dummy for no social interaction	-0.0273*** (0.011)	-0.1656*** (0.033)	-0.0470*** (0.040)
Trust	0.0050* (0.002)	0.0187** (0.002)	0.0053** (0.002)
Self-perceived social status	0.00001 (0.000)	0.0002 (0.000)	0.00003 (0.000)
Self-perceived health status	0.0006 (0.003)	0.0020 (0.003)	0.00005 (0.002)
Education quality	-0.0187* (0.011)	-0.0708** (0.009)	-0.0202** (0.007)
Sharpe-ratio	0.1519*** (0.060)	0.5839*** (0.051)	0.1669*** (0.040)
$v_2$	0.0671 (0.089)	-0.1588 (0.093)	
Observations	21440	21440	
Number of persons	4264	4264	

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Finally, we estimate the effect of financial literacy using the Control Function approach described in Section 4.3 (results in Columns 2 and 3 of Table 4). We add  $v_{i2}$ , obtained from first step linear regression to the pooled probit model, along with other explanatory variables. Again, we find evidence against the null hypothesis that financial literacy is exogenous in nature.

The APE of financial literacy is 0.058, which is almost as 8 times higher as the case in which financial literacy was treated as exogenous. Hence, this finding suggests that the previous estimates were biased. With the exception of the dummy for female, self-perceived social status and self-perceived health status, all other estimated coefficients display expected sign and are significant.

The estimated APEs of income and wealth quintiles reveal that one standard deviation of these variables is associated with 1.3% and 5% increase in the probability to own stocks, respectively. These results are in line with previous findings on the relevant role of financial resources and points to the presence of entry and management costs for investors in financial markets.

Turning to the socio-demographic variables, results are similar to those stemming from the baseline model (Table 3, Column 4) and the changes are concerned with the magnitude of the partial effects of some variables, not even the sign. Married respondents have a higher probability to stock market participation and the estimated coefficient is around 1.9% (significance level 5%). Finally, age and age-square are both significant, showing that stock holding, when purged out from cohort effects, increases with age among English respondents. Self-perceived health status remains insignificant in these models too.

On the other hand, all specifications show that the intensity of social network proxied by the lack of social activities does negatively affects the decision of participating in financial markets, with a partial effect of 4.7%. Also the variable proxing social capital i.e. trust, is found to play a significant and differentiated role: interestingly enough, individuals who believe that more people in his/her area can be trusted are more prone to join the stock market (partial effect of 0.5%).

As far as the level of optimism is concerned, proxied by self-perceived social status results are similar to those shown in Section 5.1: in fact the estimated coefficient remains insignificant, although with the expected sign.

The cohort effect proxied by sharpe-ratios shows that a one unit change in sharpe-ratio observed at young ages is associated with 16% increase in the probability to own risky financial assets. Finally, as for the school effect proxied by student-teacher ratio (meant to capture the effect that a better education system exerts on participation in stock markets through externalities at cohort group level), it has a negative sign, as expected, and its average partial effect is 2% (Table 4, Column 3).

## **6. Conclusion**

In this work we analyse the determinants of the decision to enter the stock market in England through a panel analysis. Data are drawn from the English Longitudinal Survey of Ageing dataset for years 2002-2012.

For doing this we use several methodologies including a probit model controlling for both unobserved heterogeneity and serial correlation through Correlated Random Effects, Generalized Estimating Equations and Generalized Linear Models. We find that financial literacy is a significant determinant of the decision to enter the stock market, with a partial effect of 5.8%.

Not controlling for the endogeneity of the latter variable leads to a dramatic negative bias of the estimate (with a partial effect of 1.3%).

Among individual characteristics, financial resources affect positively the probability of joining the stock market, pointing to the presence of entry costs; interestingly enough, marital status and gender are not significant for explaining participation in stock markets in England. Finally, the presence of social interaction and higher level of trust increase the probability to own stocks.

The hump shape in the age profile of participation rates turns out to be the composite effect of several factors. Once purged out from the latter, age exerts a positive effect on participation. In particular, the cohort effects captured by the quality of education at young age (proxied by student-teacher ratios) and by financial incentives observed at early stages of adult life (sharpe-ratios) are found to play a significant role.

As for policy implications, our findings suggest that the enhancement of financial literacy is crucial for favoring higher participation in capital markets. Moreover, given that financial education is strongly affected by starting conditions, policies should be designed to restore equality in opportunities among young individuals. This goal could be addressed through specific education courses, possibly at compulsory school level.

Finally, in order to promote efficient portfolio diversification, much effort should be put in improving institutional factors such as the effectiveness of the education system and those affecting the performance of the financial markets (for example, by favoring the presence of institutional investors such as pension funds) and to reduce entry costs.

## References

- Alan, S., 2006. Entry costs and stock market participation over the life cycle. *Review of Economic Dynamics*, 9(4), 588-611.
- Altonji, J.G., Matzkin, R.L., 2005. Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors. *Econometrica* 73, 1053-1102.
- Ameriks, J. and Zeldes, S.P., 2004. *How do household portfolio shares vary with age?* Working paper, Columbia University.
- Arrondel, L., Calvo-Pardo, H., Derya T., 2012. Subjective return expectations, information and stock market participation: Evidence from France. Discussion Paper on Economics and Econometrics, University of Southampton.
- Ballinger, G. A., 2004. Using generalized estimating equations for longitudinal data analysis. *Organizational research methods*, 7(2), 127-150.
- Beauchamp, J., Cesarini, D., Johannesson, M. 2017 “The psychometric properties of measures of economic risk preferences *Journal of Risk Uncertainty* 54:203–237
- Behrman, J. R., Mitchell, O. S., Soo, C. K., Bravo, D., 2012. How financial literacy affects household wealth accumulation. *The American Economic Review* 102(3), 300-304.
- Bertaut, C. C., 1998. Stockholding behavior of US households: Evidence from the 1983-1989 survey of consumer finances. *Review of Economics and Statistics*, 80(2), 263-275.
- Bertrand, M., 2011. *New perspectives on gender*, in Ashenfelter, O., Card, D eds., *Handbook of Labor Economics*, Vol. 4B, 1543-1590.

- Brav, A., Constantinides, G.M., Geczy, C.C., 2002. Asset Participation with Heterogeneous Consumers and Limited Participation: Empirical Evidence. *Journal of Political Economy* 110(4), 793-824.
- Brown, J. R., Ivković, Z., Smith, P.A., Weisbenner, S., 2008. Neighbours matter: Causal community effects and stock market participation. *The Journal of Finance* 63(3), 1509-1531.
- Brown, S., Ghosh, P., Taylor, K., 2016 Household Finances and Social Interaction: Bayesian Analysis of Household Panel Data. *Review of Income and Wealth* 62, 467-488
- Brown, S., Taylor, K., 2007. Religion and Education: Evidence from the National Child Development Study. *Journal of Economic Behaviour and Organisation* 63, 439-60.
- Calvet, L.E. and Sodini, P. (2014) Twin Picks: Disentangling the Determinants of Risk Taking in Household Portfolios. *The Journal of Finance*, vol. 69(2), 867-906.
- Calvet, L.E., Campbell, J.Y. and Sodini, P. 2007. Down or Out: Assessing the Welfare Costs of Household Investment Mistakes. *Journal of Political Economy* 115, 707-747.
- Campbell, J. Y., Cochrane, J. H., 2000. Explaining the poor performance of consumption-based asset pricing models. *The Journal of Finance*, 55(6), 2863-2878.
- Campbell, J., 2006. Household finance. *The Journal of Finance* 61(4), 1553-1604.
- Carneiro, P., Heckman, J., 2007. Human Capital Policy, *NBER Working Paper* 9495.
- Carneiro, P., Meghir, C. and Patey, M., 2013. Maternal education, home environments, and the development of children and adolescents. *Journal of the European Economic Association*, 11(s1), pp.123-160.
- Chamberlain, G., 1980. Analysis of variance with qualitative data. *Review of Economic Studies* 47, 225-238.
- Christelis, D., Jappelli, T., Padula, M., 2010. Cognitive abilities and portfolio choice. *European Economic Review* 54(1), 18-38.
- Cole, S., Shastry, G., 2008. If you are so smart, why aren't you rich? The effects of education, financial literacy and cognitive ability on financial market participation. *Working Paper# 09-071. Harvard Business School, Wellesley College (unpublished), Cambridge MA.*
- Croson, R., Gneezy, U., 2009. Gender differences in preferences. *Journal of Economic Literature* 47(2), 1-27.
- Dohmen, T.J., Falk, A., Huffman, D. and Sunde, U. (2010). Are Risk Aversion and Impatience Related to Cognitive Ability? *American Economic Review* 100(3), 1238- 1260.
- Duflo, E., Saez, E., 2003. The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics* 118 (3), 815-842.
- Edwards, R.D., 2008. Health risk and portfolio choice. *Journal of Business & Economic Statistics*, 26(4), pp.472-485.
- Felton, J., Gibson, B., Sanbonmatsu, D., 2003. Preference for risk in investing as a function of trait optimism and gender. *The Journal of Behavioral Finance* 4, 1, 33-40.
- Fonseca, R., Mullen, K.J., Zamarro, G., Zissimopoulos, J., 2012. What explains the gender gap in financial literacy? The role of household decision making. *Journal of Consumer Affairs* 46(1), 90-106.
- Georgarakos, D., Pasini, G., 2011. Trust, sociability, and stock market participation. *Review of Finance*, 15(4), 693-725.

- Giuliano, P., Spilimbergo, A., 2014. Growing up in a Recession. *The Review of Economic Studies* 81, no. 2 (2014): 787-817.
- Gomes, F. and Michaelides, A., 2005. Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence. *The Journal of Finance*, 60(2), pp.869-904.
- Grinblatt, M., Keloharju, M. and Linnainmaa, J. (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., Haliassos, M. and Jappelli, T. 2003. Household Stockholding in Europe: where do we stand, and where do we go? *Economic Policy* 36, 123-70.
- Guiso, L., Jappelli, T., 2005. Awareness and stock market participation. *Review of Finance* 9(4), 537-567.
- Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *Journal of Finance*, 63(6), 2557–2600
- Guiso, L., Sapienza, P., Zingales, L., 2004. The role of social capital in financial development. *The American Economic Review*, 94(3), 526-556.
- Gyllenram, A., Hanes, N And Hellström, J (2013) The Influence of Non-Cognitive and Cognitive Ability on Individuals' Stock Market Participation, Unpublished Manuscript, Umeå School of Business and Economics, Umeå University, 90187 Umeå, Sweden.
- Haliassos, M., Bertaut, C., 1995. Why do so few hold stocks? *The Economic Journal* 105(432), 1110–1129.
- Hardin JW. (2005). Generalized estimating equations (GEE). In *Encyclopedia of Statistics in Behavioral Science*. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa250/compoundindex>.
- Hardin, J.W., Hilbe, J.M., 2007. *Generalized linear models and extensions*. Stata press, College Station, Texas
- Hausman J.A., 1978. Specification tests in econometrics. *Econometrica* 46, 1251-1271.
- Heckman, J, J., 2006., Skill Formation and the Economics of Investing in Disadvantaged Children, *Science*, 312 (No.5782), 1900-1902.
- Hong, H., Kubik, J., Stein, J., 2004. Social interaction and stock-market participation. *The Journal of Finance* 59(1), 137–163.
- J Beauchamp, D Cesarini, M Johannesson (2017) The Psychometric and Empirical Properties of Measures of Risk Preferences *Journal of Risk and Uncertainty* 54 (3), 203–237
- Jappelli, T., Padula, M., 2013. Investment in financial literacy and saving decisions. *Journal of Banking and Finance* 37 (8), 2779-2792.
- Kimball, M., Shumway, T., 2006. Investor sophistication, and the participation, home-bias, diversification, and employer stock puzzles. *Mimeo, University of Michigan*
- Liu, Y. J., Meng, J., You, W., Zhao, L., 2014. Word-of-mouth Communication, Observational learning, and Stock Market Participation. <http://dx.doi.org/10.2139/ssrn.2251570>.
- Lusardi, A. and Mitchell, O.S, (2011) Financial literacy around the world: an overview *Journal of Pension Economics And Finance* 10(04), 497 - 508.
- Lusardi, A., Mitchell, O.S. 2014. The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1), 5-44.

- Lusardi, A., Scheresberg, C.D.B., 2013. Financial literacy and high-cost borrowing in the United States. *National Bureau of Economic Research, No. w18969*.
- Malmendier, U., Nagel, S., 2011. Depression babies: Do macroeconomic experiences affect risk-taking? *The Quarterly Journal of Economics* 126 (1), 373-416.
- Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica* 46, 69–85.
- Olga Goldfayn-Frank (2018) Personality Traits and Financial Decisions of the Households. Unpublished Working paper, <http://dx.doi.org/10.2139/ssrn.2764175>
- Papke, L.E., Wooldridge, J.M., 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1), pp.121-133.
- Poterba, J.M. and Samwick, A.A., 2001. *Household Portfolio Allocation over the Life Cycle* in Ogura S, Tachibanaki, T and W, D. A, edited Aging Issues in the United States and Japan, pp.65-104., University of Chicago Press
- Puri, M., Robinson, D., 2007. Optimism and economic choice. *Journal of Financial Economics* 86, 1, 71-99.
- Rivers, D., Vuong, Q.H., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 39, 347–366.
- Rosen, H.S. and Wu, S., 2004. Portfolio choice and health status. *Journal of Financial Economics*, 72(3), pp.457-484.
- Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.
- Thomas, A, Spataro, L., 2018. Financial Literacy, Human Capital and Stock Market Participation in Europe. *Journal of Family and Economic Issues*, 39, (4), 532–550
- Van Rooij, M. C., Lusardi, A., Alessie, R. J. 2012. Financial literacy, retirement planning and household wealth. *The Economic Journal*, 122(560), 449-478.
- Van Rooij, M., Lusardi, A., Alessie, R., 2011. Financial literacy and stock market participation, *Journal of Financial Economics* 101(2), 449–472.
- Vestman, R., 2010. Limited stock market participation among renters: Explanations and welfare consequences. *Working paper. The Institute for Financial Research (SIFR), Stockholm*.
- Vissing-Jorgensen A., 2002. Limited Stock Market Participation and the Elasticity of Intertemporal Substitution, *Journal of Political Economy* 110 (4), 825–53.
- Weinstein, N., 1980. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology* 39, 5, 806-820.
- Weinstein, N., 1984. Why it won't happen to me: Perceptions of risk factors and susceptibility. *Health Psychology* 3, 5, 431-457.
- Wooldridge, J.M., 2002. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts
- Wooldridge, J.M., 2005. *Unobserved heterogeneity and estimation of average partial effects*. In: Andrews, D.W.K., Stock, J.H. (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press, Cambridge, pp. 27–55.
- Yoong, J., 2011. Financial illiteracy and stock market participation: Evidence from the RAND American Life Panel. *Financial literacy: Implications for retirement security and the financial marketplace* 76-97. *Oxford University Press, New York*.

Zeger, S.L. and Liang, K.Y., 1986. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, pp.121-130.

**Appendix 1:** Financial literacy across waves

Number of numeracy questions correct	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
0	74	74	74	89	89	89
1	420	420	420	435	435	435
2	936	936	936	970	970	970
3	1278	1278	1278	1258	1258	1258
4	1580	1580	1580	1525	1525	1525
5	774	774	774	785	785	785
Total	5062	5062	5062	5062	5062	5062

**Appendix 2:** Description of variables used in the study and their source

Variable	Description	Source
<b>STOCKPART</b>	Stock Ownership (binary)	ELSA
<b>FL</b>	Financial literacy scores (1 to 5)	ELSA
<b>Age</b>	Age at the time of the interview	ELSA
<b>Agesq</b>	Age squared	ELSA
<b>MS</b>	Married or not (Dummy)	ELSA
<b>FE</b>	Female or not (Dummy)	ELSA
<b>WE</b>	Wealth quintiles	ELSA
<b>IN</b>	Income quintiles	ELSA
<b>SI</b>	Lack of social interaction (dummy takes value 1 if he/she does not participate in any social activities)	ELSA
<b>TR</b>	Ranking of trust on others (1-7: 1 stands for minimum level of trust)	ELSA
<b>SS</b>	Self-perceived social status (0-100: 0 stands for worst level of perception of self-status relative to others)	ELSA
<b>HE</b>	Self-perceived health condition (1-6: 1 stands for excellent health)	ELSA
<b>EQ</b>	Average student-teacher ratio observed between 6-15 cohort level (five-years cohort groups)	International Historical Statistics Individual/country level

<b>SH</b>	Average sharpe-ratio observed at 18-25 years of age (five-years cohort groups)	Global Financial Statistics
<b>books</b>	Number of books at shelf when aged 10	ELSA
<b>rooms</b>	Number of rooms in residence when aged 10	ELSA
<b>diMAedu</b>	Mother's years of education when respondent was aged 10 (codes: 1-7) (see Appendix 5)	ELSA

### Appendix 3: Financial literacy groups

Level	Number	Per Cent
Group 1(Lowest)	3145	10.3
Group II	5829	19.9
Group III	7445	24.5
Group IV	9252	30.4
Group V	4701	15.4
ALL	30372	100

### Appendix 4: Financial literacy in ELSA

The questions used to construct the financial literacy indicator are set out below. Possible answers are shown on cards displayed by the interviewer who is instructed not to read them out to respondents:

*The entry level question is the following* A) In a sale, a shop is selling all items at half price. Before the sale, a sofa costs £300. How much will it cost in the sale? Possible answers £150, £100, £200, £250, £600, Other answer and Doesn't know answer

*Then the respondent is asked*

B) If the chance of getting a disease is 10 percent, how many people out of 1,000 (one thousand) would be expected to get the disease?

The possible answers are 100, 10, 90, 900, Other answer and Doesn't know answer

*The third question is*

C) A second hand car dealer is selling a car for 6000 pound. This is two-thirds of what it costs new. How much did the car cost new? The possible answers are 9000, 4000, 8000, 12,000, 18,000, Other answer and Doesn't know answer.

*Now if IF respondent responded incorrectly on all three previous maths questions he is asked the easiest question*

D) If you buy a drink for 85 pence and pay with a one pound coin, how much change should you get back? Possible answers are 15 pence, 25 pence, Other answer and Doesn't know answer

*On the other hand, if he IF respondent replied correctly to any of the first three maths questions: A= 1 OR B = 1 OR C = 1, then he is asked*

*The following question*

E) If 5 people all have the winning numbers in the lottery and the prize is £2 million, how much will each of them get? The possible answers are £400,000 £200,000 £250,000 £500,000, Other answer and Doesn't know answer

Finally, if the respondent is able to correctly tell Question B, Question C and Question E, then he is asked the toughest question.

F) Let's say you have 2000 pound in a savings account. The account earns 10 per cent interest each year. How much would you have in the account at the end of the second year? The possible answers are 2420, 2020, 2040, 2100, 2200 and 2400, Other answer and Doesn't know answer.

Answering (A, B, C, E ,F) correctly results in a score of 5, answering (A, B, C,E) correctly and E incorrectly results in score of 4 and answering correctly any of A, B, C, or atleast two of A,B,C and then correctly E results in score of 3, 2 respectively. If he answers only D correctly he has a score of one and none correctly he collects a score of 0.

**Appendix 5:** Education of mother when respondent was 10 year old (source: ELSA life history wave).

Education of mother when respondent is aged 10	Code
Never went to school	1
Stopped at 14 or under	2
Stopped at 15	3
Stopped at 16	4
Stopped at 17	5
Stopped at 18	6
Stopped at 19 or over	7

**Appendix 6:** First stage regressions of the Endogenous variable *Financial Literacy*. Robust standard errors are reported in parentheses

Regressors	Coefficients
Age	-0.0692*** (0.011)
Age	0.00037*** (0.000)
Dummy for married	-0.1446*** (0.0167)
Dummy for female	-0.6305*** (0.014)
Wealth quintiles	0.0083 (0.010)
Income quintiles	0.0216** (0.008)
Dummy for no social interaction	-0.2022** (0.0168)

Trust	0.0091** (0.004)
Self-perceived social status	0.00063 (0.000)
Self-perceived health status	-0.0066 (0.011)
Education quality	-.0193 (0.0163)
Sharpe-ratio	.05521 (0.088)
n. of rooms	0.1378*** (0.006)
n. of books	0.0326*** (0.008)
Education of Mother	0.0297*** (0.006)
Observations	21,440
R-squared	0.207
F statistic	129.74
Hansen J statistic	3.07
P Value	0.38

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1