WHO OWNS STOCKS IN ENGLAND: A PANEL ANALYSIS

Ashok Thomas
Indian Institute of Management
ashok.thomas@iimk.ac.in

Luca Spataro
University of Pisa
luca.spataro@unipi.it

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. Financial literacy is found to be a significant determinant of the decision to enter the stock market, 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 individual variables, both financial resources and social interaction affect positively the probability to join the stock market.
Keywords: financial literacy, panel analysis, stock market participation.
JEL codes: C23, C26, G11.

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. To the best of our knowledge, this has never be done so far. In fact, earlier studies are either cross-sectional analyses at country level (Van Rooij et al. 2011; Yoong 2010) or cross-country analyses (Thomas and Spataro 2015; 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, 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. (2009), 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 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 and education (Bertraut 1998), risk aversion (Campbell and Cochrane 2000), trust in financial institutions (Georgarakos and Pasini 2011), 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) 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 college 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; Lusardi and de Bassa Scheresberg 2013) underlying the fact that general knowledge (education) and specialized knowledge (financial literacy) both contribute for financial decision making, both in Netherlands and United States.

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 Spatharo (2015) 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;
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
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 entrées are not included. Overall, the sample contains
30506 individual/year observations (all variables are summarized in Table 1).

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-years 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.

**Fig. 1.** Fraction of individuals participating in stock markets by age-groups and waves
Source: authors’ calculations.

**Fig. 2.** Time path of fraction of individuals participating in stock markets by cohort-groups
Source: authors’ calculations.
Figure 2 presents the time path of the share of individuals participating in the stock market for selected two-years 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.

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-years cohort groups) by using the data from historical stock returns of United Kingdom.

Moreover, following the insights contained in Thomas and Spataro (2015), 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.

**Fig. 3.** Financial literacy scores and stock participation (pooled data)
Source: authors’ calculations.

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 financial and numerical questions
and the answers are imputed to obtain a value ranging from 1 to 5. Details of the actual questions and the construction of this indicator are discussed in Christelis et al. (2010).

Interestingly, the age profiles of financial literacy and stock market participation show a similar pattern. From Fig.3 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 Spatharo 2015), 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). We also include as a proxy for social capital, “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.

Table 1. Sample statistics of all variables from ELSA balanced panel

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Obs</th>
<th>Mean</th>
<th>Std dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock ownership (dependent variable)</td>
<td>30,372</td>
<td>0.316</td>
<td>0.465</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Financial literacy scores</td>
<td>30,372</td>
<td>3.199</td>
<td>1.253</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Age</td>
<td>30,368</td>
<td>66.50</td>
<td>9.842</td>
<td>20</td>
<td>99</td>
</tr>
<tr>
<td>Age^2</td>
<td>30,368</td>
<td>4.519</td>
<td>1.356</td>
<td>400</td>
<td>9,80</td>
</tr>
<tr>
<td>Dummy for married</td>
<td>30,372</td>
<td>0.602</td>
<td>0.489</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dummy for female</td>
<td>30,372</td>
<td>0.578</td>
<td>0.494</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Income quintiles</td>
<td>28,343</td>
<td>3.104</td>
<td>1.396</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Wealth quintiles</td>
<td>28,343</td>
<td>3.269</td>
<td>1.375</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Dummy for no social interaction</td>
<td>30,372</td>
<td>0.338</td>
<td>0.473</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Trust</td>
<td>29,136</td>
<td>2.658</td>
<td>1.656</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Self-perceived social status</td>
<td>27,960</td>
<td>57.65</td>
<td>18.54</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Self-perceived health status</td>
<td>27,458</td>
<td>2.626</td>
<td>1.075</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Education quality</td>
<td>30,038</td>
<td>26.02</td>
<td>1.121</td>
<td>18.71</td>
<td>29.5</td>
</tr>
<tr>
<td>Average sharpe-ratios at early adult life</td>
<td>30,017</td>
<td>0.198</td>
<td>0.0933</td>
<td>-0.0259</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Source: authors’ calculations.

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. 2008). 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. 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 et al. 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 model using the Chamberlin-Mundlak device) under strict exogeneity and serial 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:

\[ E(y_{it} | x_{it}, c_i) = \Phi(x_{it}\beta + c_i), \quad t=1,\ldots,T \]  

(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:

\[ E(y_{it} | x_{it}, c_i) = E(y_{it} | x_{it}, c_i), \quad t=1,\ldots,T \]  

(2)

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

\[ c_i = \psi + \bar{x}_i + a_i, \quad a_i|x_i \sim \text{Normal}(0, \sigma^2 a) \]  

(3)

where \( \bar{x}_i \) is the vector of time averages and \( \sigma^2 a = \text{Var}(c_i|x_i) \) 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 \( D(y_{it}|x_{it}, c) \) 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, given that the ones obtaining from independent correlation matrix are not significantly different.

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 \( yit2 \) is an endogenous explanatory variable; provided we have sufficient instruments, we can express the conditional mean model as:

\[
E(yit1 | yit2, Zi, c1, vit) = E(yit1 | yit2, zit1, c1, vit) = \Phi(\alpha | yit2 + zit1 + c1 + vit1)
\]  

(4)

where \( yit1 \) is binary, \( c1 \) is the time-constant unobserved effect and \( vit1 \) is a time-varying omitted factor that can be correlated with \( yit2 \), the potentially endogenous variable (financial literacy in our case), \( zit \) is the vector of exogenous variables (See detailed methodology in Appendix 1).

### 4.4 Application of methodology to dataset

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

\[
\text{Stockpart} = \alpha + \beta_1 \text{FL} + \beta_1 \text{Age} + \beta_3 \text{Agesq} + \beta_4 \text{MS} + \beta_5 \text{FE} + \beta_6 \text{WE} + \beta_7 \text{IN} + \beta_8 \text{SI} + \beta_9 \text{TR} + \beta_{10} \text{SS} + \beta_{11} \text{HE} + \beta_{12} \text{EQ} + \beta_{13} \text{SH} + \beta_{14} \text{WEBAR} + \beta_{15} \text{INBAR} + \beta_{15} \text{TRBAR} + \beta_{16} \text{SSBAR} + \beta_{17} \text{HEBAR} + \beta_{18} \text{DUM}2002 + \ldots + \beta_{22} \text{DUM}2010 + \nu
\]  

(5)

where \( \text{FL} \) 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). \( \text{EQ} \) 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. \( \text{SI} \) is the average sharp-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. 5): 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:

\[
\text{FL} = \eta + \pi \text{books} + \pi \text{rooms} + \pi \text{diMAedu} + \pi \text{Age} + \pi \text{Agesq} + \pi \text{MS} + \pi \text{FE} + \pi \text{WE} + \pi \text{IN} + \pi \text{SI} + \pi \text{TR} + \pi \text{SS} + \pi \text{HE} + \pi \text{EQ} + \pi \text{SH} + \pi \text{WEBAR} + \pi \text{INBAR} + \pi \text{TRBAR} + \pi \text{SSBAR} + \pi \text{HEBAR} + \pi \text{DUM}2002 + \ldots + \pi \text{DUM}2010 + \nu
\]  

(6)

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 (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 (5) by instrumental variable and Probit QMLE approach.

Table 2. Sample statistics of instruments

<table>
<thead>
<tr>
<th>INSTRUMENTS</th>
<th>Obs</th>
<th>Mean</th>
<th>Standard error</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother’s education at age 10</td>
<td>30,372</td>
<td>2.309</td>
<td>1.159</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Number of rooms at age 10</td>
<td>27,834</td>
<td>2.915</td>
<td>0.935</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Number of books at age 10</td>
<td>27,834</td>
<td>2.490</td>
<td>1.210</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

Source: authors’ calculations.

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 0.7%. 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 5.6% and 2.1% 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%). 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 -2.6% 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 variables, trust and dummy for lack of interaction are significant and with the expected sign, with marginal effects of 0.5% and -0.3% respectively.
<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Correlated Random effects</th>
<th>APE</th>
<th>Probit Pooled QMLR Coefficient</th>
<th>APE</th>
<th>Probit GLM Coefficient</th>
<th>APE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Literacy</td>
<td>0.0365**</td>
<td>0.071**</td>
<td>0.0475***</td>
<td>0.0147***</td>
<td>0.0249**</td>
<td>0.068***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.002)</td>
<td>(0.009)</td>
<td>(0.001)</td>
<td>(0.014)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0510**</td>
<td>-0.0503**</td>
<td>-0.0596***</td>
<td>-0.0168***</td>
<td>-0.0115**</td>
<td>-0.0089**</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Age 2</td>
<td>0.0405**</td>
<td>0.0000**</td>
<td>0.0415**</td>
<td>0.0058***</td>
<td>0.0004**</td>
<td>0.0006**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Dummy for married</td>
<td>0.2175**</td>
<td>0.2420**</td>
<td>0.1143**</td>
<td>0.0324**</td>
<td>0.1200**</td>
<td>0.0545**</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.007)</td>
<td>(0.022)</td>
<td>(0.009)</td>
<td>(0.033)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Dummy for female</td>
<td>-0.0667</td>
<td>-0.0155</td>
<td>-0.0252</td>
<td>-0.0195**</td>
<td>-0.0077</td>
<td>-0.0107</td>
</tr>
<tr>
<td></td>
<td>(0.0575)</td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Wealth quantiles</td>
<td>0.2868***</td>
<td>0.03647***</td>
<td>0.1871***</td>
<td>0.0827***</td>
<td>0.1873***</td>
<td>0.0471***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.002)</td>
<td>(0.015)</td>
<td>(0.002)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Income quantiles</td>
<td>0.1071***</td>
<td>0.0216***</td>
<td>0.1044**</td>
<td>0.0414**</td>
<td>0.0174**</td>
<td>0.0100**</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Dummy for no social</td>
<td>-0.1677***</td>
<td>-0.0373***</td>
<td>-0.1847***</td>
<td>-0.0529***</td>
<td>-0.0718***</td>
<td>-0.0279***</td>
</tr>
<tr>
<td>interaction</td>
<td>(0.041)</td>
<td>(0.001)</td>
<td>(0.023)</td>
<td>(0.004)</td>
<td>(0.031)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Trust</td>
<td>0.0282**</td>
<td>0.0057**</td>
<td>0.0170**</td>
<td>0.0049**</td>
<td>0.0068**</td>
<td>0.0050**</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.001)</td>
<td>(0.036)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Self-perceived social</td>
<td>0.091**</td>
<td>0.0156**</td>
<td>0.0862**</td>
<td>0.0402**</td>
<td>0.0016**</td>
<td>0.0020**</td>
</tr>
<tr>
<td>status</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Self-perceived health</td>
<td>0.002</td>
<td>0.0055</td>
<td>0.0111**</td>
<td>0.0058**</td>
<td>0.0055**</td>
<td>0.0009**</td>
</tr>
<tr>
<td>status</td>
<td>(0.019)</td>
<td>(0.003)</td>
<td>(0.014)</td>
<td>(0.003)</td>
<td>(0.011)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Education quality</td>
<td>-0.1276**</td>
<td>-0.0269**</td>
<td>-0.0675**</td>
<td>-0.0192**</td>
<td>-0.1260**</td>
<td>-0.0219**</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.002)</td>
<td>(0.021)</td>
<td>(0.002)</td>
<td>(0.02)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Sharpe-Ratio</td>
<td>0.2235***</td>
<td>0.1815***</td>
<td>0.2538***</td>
<td>0.1929***</td>
<td>0.2538***</td>
<td>0.1545***</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.05)</td>
<td>(0.117)</td>
<td>(0.023)</td>
<td>(0.034)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.235</td>
<td>1.529</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>23.008</td>
<td>23.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: authors’ calculations.
Among the demographic variables, age displays a positive sign, showing that 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 1.3% 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 0.7%). Interestingly, as for the financial literacy coefficient, the fully robust standard error for the probit QMLE estimate (0.001) is lower than the fully robust standard error for the GEE estimation (0.002, column 5). Hence, in this work, using exchangeable working correlation matrix in multivariate weighted non-linear least squares estimation does not appear to enhance efficiency. The socio-demographic variables like age and civil 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 1.7% and 1.8% 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 -2%).

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. 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).

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 where 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.

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 2014). 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 2015). 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; 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 u2 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%.

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

<table>
<thead>
<tr>
<th>Variables</th>
<th>Linear Instrumental Variables Coefficient</th>
<th>Probit Pooled QMLE Coefficient</th>
<th>APE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Literacy</td>
<td>0.0792**</td>
<td>0.2041**</td>
<td>0.0580***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.003)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0126**</td>
<td>-0.0467**</td>
<td>-0.0133**</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.021)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Age^2</td>
<td>0.00008*</td>
<td>0.0032**</td>
<td>0.0009***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Dummy for married</td>
<td>0.0128</td>
<td>0.0429**</td>
<td>0.0197**</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.007)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Dummy for female</td>
<td>-0.0344</td>
<td>-0.0732</td>
<td>-0.0208</td>
</tr>
<tr>
<td></td>
<td>(0.0575)</td>
<td>(0.064)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Wealth quintiles</td>
<td>0.0474***</td>
<td>0.1790***</td>
<td>0.0509***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.013)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Income quintiles</td>
<td>0.0126***</td>
<td>0.0457***</td>
<td>0.0130***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.011)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Dummy for no social interaction</td>
<td>-0.0273***</td>
<td>-0.1656***</td>
<td>-0.0470***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.033)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>Trust</td>
<td>0.0050*</td>
<td>0.0187**</td>
<td>0.0053**</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Self-perceived social status</td>
<td>0.00001</td>
<td>0.0002</td>
<td>0.00003</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Self-perceived health status</td>
<td>0.0006</td>
<td>0.0020</td>
<td>0.00005</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>
Who owns stocks in England: a panel analysis

<table>
<thead>
<tr>
<th>Education quality</th>
<th>-0.0187**</th>
<th>-0.0708***</th>
<th>-0.0202***</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.009)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Sharpe-ratio</td>
<td>0.1519***</td>
<td>0.5839***</td>
<td>0.1669***</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.051)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>v2</td>
<td>0.0671</td>
<td>-0.1588</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.093)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>21440</td>
<td>21440</td>
<td></td>
</tr>
<tr>
<td>Number of persons</td>
<td>4264</td>
<td>4264</td>
<td></td>
</tr>
</tbody>
</table>

Source: authors' calculations.

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 \( \nu_2 \), 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 lower than 1%).

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.

Appendix 1

Estimation under endogeneity:

The traditional instrumental variable method which could provide results by eliminating $ci$ cannot be attempted here (as these common estimation methods eliminates $ci$ along with any time-constant explanatory variables) and therefore the CRE approach of modelling the distribution of unobserved heterogeneity, $D(ci | zi)$, is again attempted. Additionally, one has to model how $yt2$ is related
to vit1. 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 ai1 to be independent of zi, we can write: 
\[ c1 = \psi_1 + \beta_1 z_i + \alpha_i, \quad ai1 | zi \sim \text{Normal}(0, \sigma^2 a_1) \]  
(A)

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

\[ E(yit1 | yt2, zi, ai1, vit1) = \Phi(\alpha_1 yit2 + \beta_2 vit1 + \psi_1 z_i + \alpha_i + vit1) = \Phi(\alpha_1 yit2 + \beta_2 vit1 + \psi_1 z_i + \alpha_i + vit1) \]

(B)

Assuming a linear reduced form of the suspected endogenous variable yit2 we get:

\[ yit2 = \psi_2 + \beta_2 z_i + \varepsilon_2 + \alpha_i + vit1, \quad t = 1, \ldots, T \]

with rit1 = ai1 + vit1. The addition of time averages of strictly exogenous variables Z_i in eq. (B) follows a Mundlak (1978) device. As for the source of endogeneity of yit2, it stems from the relationship between vit2, the reduced-form error term and the new term rit1 in eq. (6). Thus yit2 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 (B), while also allowing for possible feedback from unobserved idiosyncratic changes in yit1 as captured by vit1. 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 yit2 (pooled across t) and obtain the residuals;
2) Use the probit QMLE of yit1 on yit2, zit1, Z_i.

References


Who owns stocks in England: a panel analysis


151

