Windfall Gains and Stock Market Participation*

Joseph Briggs†
Federal Reserve Board of Governors

David Cesarini
New York University, IFN & NBER

Erik Lindqvist
SOFI, Stockholm University & IFN

Robert Östling
Stockholm School of Economics

We exploit the randomized assignment of lottery prizes in a large, administrative, Swedish data set to estimate the causal effect of wealth on stock market participation. A $150,000 windfall gain increases stock market participation probability by 12 percentage points among pre-lottery nonparticipants but has no discernible effect on pre-lottery stock owners. For entry costs below $31,000, a structural life-cycle model significantly overpredicts entry rates. Additional analyses implicate pessimistic beliefs regarding equity returns as a major source of this overprediction, and suggest that both recency- and early-life biases affect the belief formation process.

*This paper is part of a project hosted by the Research Institute of Industrial Economics (IFN). We are grateful to IFN Director Magnus Henrekson for his strong support of the project and to Marta Benkestock for superb administrative assistance. For helpful comments, we thank Steffen Andersen, Dan Benjamin, Claudio Campanale, David Laibson, Annette Vissing-Jørgensen, Enrichetta Ravina, Paolo Sodini, and Roine Vestman, as well as seminar participants at NYU Stern, IFN, Institute for International Economic Studies, London Business School, Copenhagen Business School, the Federal Reserve Bank of New York, the Federal Reserve Board of Governors, University of Southern California, the 2015 NBER SI (Household Finance), and the 2015 SED. We also gratefully acknowledge financial support from the Alfred P. Sloan Foundation through the NBER Household Finance small grant program (22-2382-13-1-33-003), the NSF (1326635), the Swedish Council for Working Life and Social Research (2011-1437), the Swedish Research Council (B0213903), Riksbankens Jubileumsfond (P15-0615:1), and Handelsbanken's Research Foundations (P2011:0032:1). The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Board of Governors.
JEL No: D1,D14,D91, G02,G11, G12

†Corresponding author: Joseph Briggs. Board of Governors of the Federal Reserve System, Constitution Ave & 20th St NW, Washington, DC 20551. 202-245-4214. joseph.s.briggs@frb.gov.

1. Introduction

Canonical life-cycle models of consumption and saving (see, e.g., Samuelson (1969); Merton (1971)) predict that all individuals should invest a positive fraction of their wealth in equities. However, a sizable fraction of households in most countries do not own equity. A large literature in household finance formulates and tests hypotheses about the causes of this "nonparticipation puzzle."[1] As Campbell (2006) notes, insights into the causes of equity market nonparticipation could guide efforts to promote efficient financial decision-making.

Limited stock market participation is often analyzed using models in which agents weigh the benefits of owning equities against its costs.[2] Early work by Vissing-Jørgensen (2003) posited a simple model with two types of costs: per-period participation costs and a one-time entry cost. Since the gains from participation increase with wealth, whereas costs remain fixed, this framework can explain why participation increases with wealth. The framework has been subsequently adopted by a large structural literature which models household saving and portfolio decisions over the life-cycle. A common finding in this literature is that under standard calibrations, a modest per-period participation cost is enough to match participation rates at most wealth levels.[3]

These models make precise, quantitative predictions about the effect of wealth on stock market participation. Stringently testing these predictions is challenging, however, since most studies of wealth effects (see, e.g., Brunnermeier and Nagel (2008); Calvet et al. (2009); Calvet and Sodini (2014)) rely on observational data where, even applying the best methods, it is difficult to eliminate concerns about omitted variables and simultaneity. A notable exception is Andersen and Nielsen (2011), which uses Danish inheritances from sudden deaths to study the effect of a financial windfall on stock market participation.

In this paper, we estimate the effect of lottery wealth on stock market participation by exploiting the randomized assignment of lottery prizes in three samples of Swedish lottery players

[1]See Haliassos and Bertaut (1995), Guiso et al. (2002), Vissing-Jørgensen (2003), Campbell (2006), and Guiso and Sodini (2013), among others, for discussions of limited stock market participation.

[2]Examples of such models include Mulligan and Sala-i Martin (2000), Vissing-Jørgensen (2003), Paiella (2007), and Attanasio and Paiella (2011).

[3]Examples of structural models featuring cost-based disincentives to stock market participation include Gomes and Michaelides (2005), Cocco (2005), Alan (2006), Khorunzhina (2013), Cooper and Zhu (2016), and Fagereng et al. (2017). Campbell (2006) notes that matching nonparticipation rates of wealthy households is a challenge to models with cost disincentives. Extending models to include housing (see, e.g., Cocco (2005); Flavin and Yamashita (2011); Vestman (2018)), outstanding debt (see, e.g., Davis et al. (2006); Becker and Shabani (2010)), and private business equity (see, e.g., Heaton and Lucas (2000)) improve model fit along this dimension.

who have been matched to high-quality, administrative financial records.[4] Our research design has several attractive features. First, we observe the factors conditional on which lottery prizes are randomly assigned (e.g. number of tickets owned), as is necessary for a credible causal estimation strategy. Second, because the size of the prize pool is over 500 million USD, our study has excellent power to detect even modest effects of wealth on participation over various time horizons. Third, the prizes won by the players in our sample vary in magnitude, allowing us to explore and characterize nonlinear effects of wealth. Fourth, because our lottery and financial data are drawn from administrative records, our sample is virtually free from attrition.

Our study proceeds in three stages. We first report reduced-form estimates of the effects of wealth on stock ownership. According to our quasi-experimental estimates, a 1M SEK (approximately 150K USD) windfall from lottery wealth increases the probability of stock ownership in post-lottery years by 4 percentage points. This effect is driven almost entirely by an immediate and seemingly permanent 12 percentage point effect among households that did not participate in equity markets prior to winning the lottery.

We next use a structural model to interpret the quasi-experimental estimates and provide insights into the economic forces underlying equity participation decisions (Kahn and Whited (2017)). When the model parameters are estimated from observational data, the model predicts rates of entry much larger than the reduced-form estimates. Accounting for participation responses to lottery wins requires extremely large entry costs: when model parameters are estimated to match our quasi-experimental estimates, the average entry cost for pre-lottery equity market nonparticipants is over 31K USD, or approximately 10 times larger than the average cost estimated from non-experimental data. Our structural analysis thus demonstrates the challenge our reduced-form estimates pose to standard models of stock market participation.

A third set of analyses explore potential explanations for the significant discrepancy between reduced form estimates and model predictions. We consider three broad classes of explanations: economic explanations (e.g., investment in other assets), alternative preferences (e.g., status-quo bias, loss aversion, and present-bias), and non-standard beliefs. While these explanations are not

[4] A key methodological difference between our reduced-form analyses and Andersen and Nielsen (2011) is that a bequest is conceptually different from a windfall gain to lifetime wealth. Although unexpected inheritances clearly increase liquid wealth, their net impact on lifetime wealth is difficult to quantify (or even sign correctly) absent further assumptions on the parent's saving, investment and consumption decisions under the counterfactual scenario in which the parent dies at an older age. In contrast, our study's estimates can be interpreted unambiguously as reflecting the causal impact of a positive wealth shock induced by lottery winnings.

mutually exclusive and there is some support for each, the evidence points to non-standard beliefs as a major source of the model's overprediction. For example, the difference between empirical and model predictions is much smaller, albeit still positive, in subsamples of individuals with higher education and cognitive test scores. Additionally, survey measures suggest that lottery winners' future equity return beliefs are overly pessimistic relative to historical returns. We conservatively estimate that half of the discrepancy between reduced-form estimates and model predictions vanishes when the model is calibrated to match the subjective belief-distribution.

In light of the evidence that non-standard beliefs are a potentially important source of the model's overprediction, we conduct further analyses to shed light on the underlying belief formation process. Estimated effects of lottery wealth on participation are larger both among households that won in years following positive equity returns and among households that experienced higher returns during formative years. These patterns are consistent with theories in which recent experiences (see, e.g., Greenwood and Shleifer (2014); Gennaioli et al. (2016)) and early personal experiences (see, e.g., Kaustia and Knüpfer (2008); Malmendier and Nagel (2011); Kuhnen (2015)) bias beliefs. Further evidence suggests that these two belief biases add to each other, and that even the highly educated are affected by both types of biases.

The paper is structured as follows. Section 2 describes the lottery and wealth data, our identification strategy, and addresses several issues regarding external validity that are often raised about studies of lottery players. Section 3 reports reduced-form estimates of the effect of lottery wealth on equity market participation, while Section 4 uses a structural life-cycle model to interpret the causal estimates. Section 5 presents a set of empirical and structural analyses to evaluate the credibility of alternative explanations of our results. Finally, Section 6 discusses our findings and concludes.

2. Data and Identification Strategy

Our analyses are conducted in a sample of lottery players who have been matched to administrative demographic and financial records using personal identification numbers (PINs).

2.1 Register Data

Our outcome variables are all derived from the Swedish Wealth Register, which contains high-quality information about the financial portfolios of all Swedes. The register was discontinued when Sweden abolished its wealth tax, but has annual year-end financial information for 1999-

2007. This information includes total assets and debt, and relevant subcategories such as bank account balances, mutual funds, directly held stocks, bonds, money market funds, debt, and residential and commercial real estate. Beginning with a landmark paper by Calvet, Campbell and Sodini (2007), the data have been used in several influential studies and are generally of very high quality. Section 3.2 discusses and addresses several data limitations that are important to consider in our specific context.

We supplement the portfolio data from the Wealth Register with basic demographic information available from Statistics Sweden. The unit of analysis in our main specification is a household, defined as the observed winner and, if present, his or her spouse. All lottery winners in our sample are aged 18 and above.

2.2 Lottery Data

Our identification strategy is to use the available data and knowledge about the institutional details of each of the lotteries to define cells within which the lottery prizes are randomly assigned. We control for cell fixed effects in all our analyses, thus ensuring all identifying variation comes from players in the same cell. Because the exact construction of the cells varies across lotteries, we describe each lottery separately. For a more detailed description of the data, including how the original lottery data were preprocessed and quality-controlled, see Section 2 and the Online Appendix of Cesarini et al. (2016). Unless otherwise noted, prizes are paid as a one-time lump-sum and all amounts are after tax. In this paper, all prize amounts (and other financial variables) are adjusted for inflation and expressed in year-2010 SEK and USD, assuming the Dec. 31, 2010 exchange rate of 6.72 SEK/USD.

2.2.1 Kombi

Kombi is a monthly subscription lottery whose proceeds are given to the Swedish Social Democratic Party, Sweden's main political party during the post-war era. Kombi provided us with a longitudinal data set with information about all draws conducted between 1998 and 2011. For each draw, the panel contains an entry per lottery participant, with information about the number of tickets held, any large prizes won, and the player's PIN.

In a given Kombi draw, each prize is awarded by randomly selecting a unique ticket. Two individuals who purchased the same number of tickets are equally likely to win a large prize. To construct the cells, each winning player is matched to (up to) 100 non-winning players with the same number of tickets in the month of the draw. To improve precision, we choose controls similar to the winner on sex and age whenever more than 100 matches are available. This matching procedure leaves 346 large prize-winners matched to a total of 31,180 controls.

2.2.2 *Triss*

Triss is a scratch-ticket lottery run since 1986 by Svenska Spel, the Swedish government-owned gambling company. Since 1994, Triss players can win an opportunity to participate in a TV show in which they draw a prize by selecting a ticket from a shuffled stack. In our main analyses, the Triss sample consists of 3,404 TV show participants who won lump-sum prizes between 7.8K USD (52K SEK) and 909K USD (6.1M SEK). However, one analysis in Section 3 compares estimates for lump-sum prize winners to a "Triss-monthly" sample of 476 participants who received prizes paid in monthly installments for 10 to 25 years (see Appendix Table B.1 for descriptive statistics). We convert the installments to net present value to make them comparable to lump-sum prizes.

Svenska Spel supplied the basic demographic information (name, age, and address) of all TV show participants between 1994 and 2011, allowing us to identify 99% of participants. Our analyses are based exclusively on the 93% of winners that did not indicate they shared ownership of the winning ticket. Our empirical strategy makes use of the fact that, conditional on winning the right to participate in the TV show, the nominal prize amount is random. Thus, two players are assigned to the same cell if they won the same type of prize, in the same year, and under the same prize plan.

2.2.3 *PLS*

Prize-linked savings (PLS) accounts are savings accounts whose owners participate in regular lotteries with monetary prizes paid on top of (or sometimes in lieu of) interest payments. In Sweden, PLS accounts were subsidized by the government until 1985, at which point the government ceased subsidies but authorized banks to continue offering PLS accounts. Two systems were put into place, one operated by savings banks and one by commercial banks and the state bank. The two systems were approximately equally popular and participation was widespread across broad strata of Swedish society, with every other Swede owning an account in the late 1980s.

The PLS sample was obtained by combining prize lists and monthly data on account balances from the PLS accounts maintained by commercial banks and the state bank. These data allow us to identify the account owner, account balance, and amount won in each draw. Overall, we were able to reliably identify the owner's PIN for 99% of prize-winning accounts. PLS account holders could win odds prizes or fixed prizes. The probability of winning either type of prize was proportional to the number of tickets associated with an account: account holders were assigned one lottery ticket

Table 1: Overview of Identification Strategy. Period indicates the years that lottery prizes were paid. Prize Type indicates whether prizes were fixed prizes of a set level or odds prizes paid as a multiple of account balance. Cells indicates the factors that were used to construct the groupings which are included as fixed effects in Equation 1 to control for random assignment.

Lottery	Period	Prize Type	Cells
PLS	1989-2003	Fixed Prize	Draw × # Fixed Prizes
PLS	1989-1994	Odds Prize	Draw × Balance
Kombi	1994-2007	Fixed Prize	Draw × # Tickets
Triss Lump-sum	1994-2007	Fixed Prize	Year × Prize Plan
Triss Monthly	1997-2007	Fixed Prize	Year × Prize Plan

per 100 SEK in account balance. Fixed prizes were prizes whose magnitude did not depend on the balance of the winning account. Odds prizes, on the other hand, were awarded as a multiple of the balance of the prize-winning account.

For fixed prize winners, our identification strategy, which is the same as in Imbens et al. (2001) and Hankins et al. (2011), exploits the fact that in the population of players who won exactly the same number of fixed prizes in a particular draw, the total amount is independent of the account balance. We therefore assign two individuals to the same cell if they won an identical number of fixed prizes in that draw. To construct odds prize cells, we match individuals who won exactly one odds-prize between 1989 and 1994 in a draw to individuals with a near-identical account balance who also won exactly one prize (odds or fixed) in the same draw. This matching procedure ensures that within a cell, the prize amount is independent of potential outcomes. In total, the sample includes 332,647 PLS prizes, of which 478 are larger than 150K USD (1M SEK).

2.3 Identification Strategy

Table 1 summarizes the previous section's discussion of how we construct the cell fixed effects in each of the three lotteries. Normalizing the time of the lottery to $s = 0$, the main estimating equation is given by,

$$Y_{i,s} = \beta_s L_{i,0} + \boldsymbol{X}_{i,0}\mathrm{M}_{\mathbf{s}} + \boldsymbol{Z}_{i,-1}\boldsymbol{\gamma}_s + \eta_{i,s}, \tag{1}$$

where i indexes households, $L_{i,0}$ denotes the prize size (in million SEK), $\boldsymbol{X}_{i,0}$ is a vector of cell fixed effects, and $\boldsymbol{Z}_{i,-1}$ is a vector of controls measured in the year before the lottery. The controls are included only to improve the precision of our estimates. Standard errors are clustered at the level of the player. The key identifying assumption needed for β_s to have a causal interpretation is that the prize amount won is independent of $\eta_{i,s}$ conditional on the cell fixed effects.

We estimate Equation 1 in our pooled sample and in the subsample of players who participated in draws between 2000 and 2007. In what follows, we refer to these samples as the *all-year* and the *post-1999* samples. The post-1999 sample plays an important role in subsample analyses conditioned on pre-lottery participation status, which is first observed in 1999. In the all-year sample regressions, the set of pre-lottery controls include age, sex, marital status, higher education, household size, household income, and Nordic born. In the post-1999 sample regressions, additional controls include net wealth, gross debt, and an indicator for real estate ownership.

2.3.1 Prize Variation

To get a better sense of the source of our identifying variation, Table 2 summarizes the distribution of prizes. The total value of the after-tax prize money disbursed to the winners in our samples is over 500M USD (3.4B SEK). Although most prizes are small, our reduced-form estimates are mostly informative about the effect of winning large sums of money. Most of the identifying variation in all three lotteries comes from within-cell comparisons of non-winners, or winners of small or moderate amounts, to large-prize winners. One way to see this is to consider the change in the total treatment variation (defined as the the within-cell demeaned total sum of squares of prizes) when prizes of different sizes are dropped from the data. Dropping the 308,948 prizes below 1.5K USD (10K SEK) in the all-year sample reduces the treatment variation by 1.4% while dropping the 1,012 prizes above 150K USD (1M SEK) reduces the treatment variation by 91.1%.[5] Triss, Kombi, and PLS all contribute substantial identifying variation to the all-year sample (57%, 14%, and 29%, respectively), while Triss and Kombi account for most identifying variation in the post-1999 sample (64% and 35%, respectively).

2.3.2 Testing for Random Assignment

To test our key identifying assumption, we again normalize the time of lottery to $s = 0$ and run the following regression:

$$L_{i,0} = \boldsymbol{X}_{i,0}\boldsymbol{\Gamma}_0 + \boldsymbol{Z}_{i,-1}\boldsymbol{\rho}_{-1} + \epsilon_i. \tag{2}$$

[5] We retain non-winners in Kombi in the sample when dropping small prizes. Because all players in the Kombi lottery won a large prize or nothing, dropping the non-winners eliminates any identifying variation.

Table 2: Prize Distribution. This table shows the number of lottery prizes in the indicated prize-size categories for the pooled all-year and post-1999 samples, and their respective lottery-specific subsamples. Prize amounts are in year-2010 USD and net of taxes. In the all-year regressions (Columns 1 and 2) controls include age, sex, marital status, higher education, household size, household income, and an indicator for being Nordic born. In the post-1999 regressions (Columns 3-7) we additionally control for time net wealth, gross debt, and an indicator for real estate ownership, all measured at time $s = -1$.

Prize Amount	**A. All-year**				**B. Post-1999**			
(K USD)	**Pooled**	**PLS**	**Kombi**	**Triss**	**Pooled**	**PLS**	**Kombi**	**Triss**
$L = 0$	31,180	0	31,180	0	26,126	0	26,126	0
$L \leq 1.5K$	308,948	308,948	0	0	41,578	41,578	0	0
$1.5 < L \leq 15$	22,082	21,097	0	985	734	368	0	366
$15 < L \leq 75$	4,009	1,935	0	2,074	1,237	0	0	1,237
$75 < L \leq 150$	346	189	0	157	89	0	0	89
$150 < L \leq 300$	822	443	330	49	297	2	273	22
$300 < L$	190	35	16	139	78	0	16	62
N	367,577	332,647	31,526	3,404	70,139	41,948	26,415	1,776

Under the null hypothesis of conditional random assignment, the characteristics determined before the lottery ($\boldsymbol{Z}_{i,-1}$) should not predict the lottery outcome ($L_{i,0}$) conditional on the cell fixed effects ($\boldsymbol{X}_{i,0}$). We run these randomization tests in the pooled all-year and post-1999 samples, and for each lottery separately in the post-1999 sample. As expected, Table 3 shows that the lagged characteristics have no statistically significant predictive power in the specifications that include cell fixed effects. If they are omitted however (Columns 2 and 4), the null hypothesis of random assignment is rejected.

2.4 *Representativeness of the Lottery Sample*

The main concern about the external validity of our sample is that individuals who play the lottery might not be representative of the population at large. To investigate representativeness, we compare the lottery samples, weighted by prize size, to randomly drawn population samples of adult Swedes matched on sex and age.

Columns 1 and 2 of Table 4 show that the demographic characteristics of our lottery players closely resemble those of the representative sample. Columns 3 and 4 compare the financial

Table 3: Testing for Random Assignment. Results are obtained by estimating Equation 2 in our all-year sample, in the post-1999 sample, and in the post-1999 lottery-specific subsamples. F-statistics and corresponding p-values result from testing the joint significance of the indicated controls.

	All-year		**Post-1999**				
	Pooled		**Pooled**		**PLS**	**Kombi**	**Triss**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Fixed Effects	Cells	None	Cells	None	Cells	Cells	Cells
Demographic Controls							
F-stat	0.80	9.92	1.13	8.41	0.69	1.41	1.34
p	0.61	0.00	0.33	0.00	0.72	0.22	0.21
Financial Controls							
F-stat			1.29	17.38	0.77	0.87	1.22
p			0.28	0.00	0.51	0.46	0.30
Demographic + Financial Controls							
F-stat			1.52	14.95	0.81	1.65	1.43
p			0.11	0.00	0.64	0.11	0.15
N	367,577	367,577	70,139	70,139	41,948	26,415	1,776

characteristics of members of the post-1999 sample to a matched population sample. The pooled lottery sample has slightly less wealth than the matched population sample, slightly more debt, and is slightly more likely to own real estate. Notably, the equity market participation rate (the main outcome in our study) in the pooled sample is 66%, close to the 63% participation rate in the matched population sample. Columns 5-7 provide the corresponding descriptive statistics for the post-1999 sample broken down by lottery. PLS participants, who are selected on bank account ownership, have significantly more wealth than the representative sample.

Another way to gauge representativeness is to compare the cross-sectional relationships between stock market participation and household characteristics in the lottery samples to the relationships estimated in a representative sample. We conduct such a comparison by estimating a cross-sectional probit equation similar to that presented in Calvet et al. (2007)'s study of the Swedish population. To avoid including wealth variation that was induced by the lottery, we restrict the estimation sample to the post-1999 sample and use observations the year prior to the lottery. We then repeat this regression for the matched representative sample. Appendix Table B.2

Table 4: Representativeness of All-Year and Post-1999 Samples. This table compares our prize-weighted all-year and post-1999 samples to representative samples matched on sex and age. The summary statistics shown are all means and measured at $s = -1$. All variables except female, age, and Nordic born are measured at the household level. Households are classified as equity market participants if the own equity either directly or indirectly via mutual funds. Continuous financial variables are winsorized at the .5 and 99.5 percentiles.

	All-Year		Post-1999				
	Pooled (1)	Pop (2)	Pooled (3)	Pop (4)	PLS (5)	Kombi (6)	Triss (7)
Demographic							
Female	.50	.50	.52	.52	.58	.44	.56
Age (years)	56.6	56.6	56.2	56.2	62.9	61.7	51.9
Nordic Born	.96	.93	.96	.92	.95	.98	.94
Household Members	.38	.41	.43	.42	.24	.22	.59
Household Income (K USD)	48	45	54	54	49	51	57
Married	.56	.57	.52	.54	.52	.48	.54
College	.23	.24	.24	.31	.27	.22	.26
Financial							
Net Wealth (K USD)			131	158	205	123	128
Gross Debt (K USD)			53	49	27	37	67
Home Owner			.75	.69	.73	.78	.73
Equity Market Participant			.66	.63	.74	.69	.63
N	367,577	367,577	70,139	70,139	41,948	26,415	1,776

shows that the results from these regressions are quite similar.

While the absence of large differences in pre-lottery financial and demographic characteristics between the lottery sample and the representative sample is reassuring, the possibility that selection into lotteries is based upon unobserved factors that limit the external validity of our results cannot be completely ruled out.

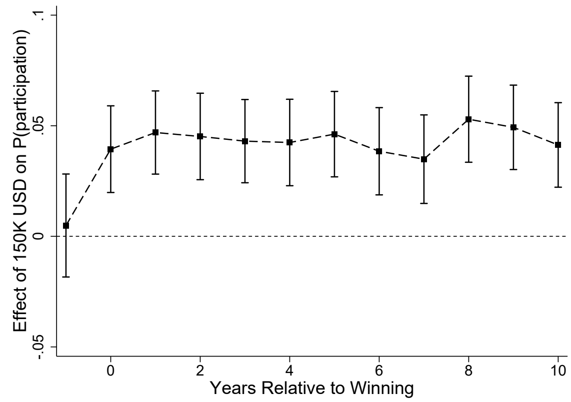


Figure 1: Effect of 150K USD (1M SEK) of Lottery Wealth on Participation Probability. Coefficients and 95% confidence intervals are obtained by estimating Equation 1 in the all-year sample. See Appendix Table B.3 for the underlying estimates.

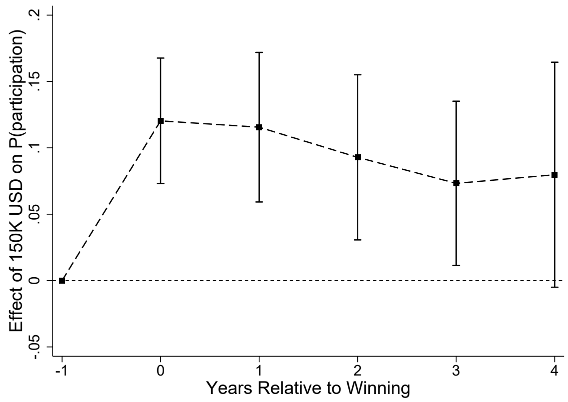
3. Quasi-experimental Estimates

3.1 Baseline Results

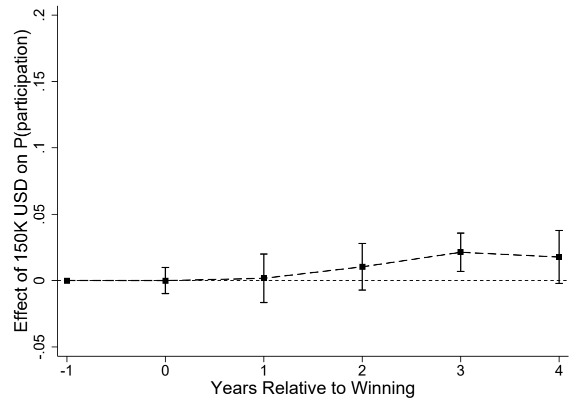
The primary outcome variable is year-end participation, defined (as is standard in the literature) as an indicator variable equal to 1 for households that own stocks either directly or indirectly via mutual funds. Figure 1 presents the estimated coefficients for $s = -1, ..., 10$ from the all-year lottery sample. Each 150K USD (1M SEK) causes a near-immediate and permanent increase in the participation probability of around 3.9 percentage points. As expected, lottery wealth does not predict participation prior to the lottery.

We next investigate treatment effect heterogeneity with respect to equity market participation prior to the lottery. Figure 2 shows the estimated treatment effects on participation at $s = -1, ..., 4$ in the post-1999 sample stratified by pre-lottery participation status. Among pre-lottery nonparticipants, each 150K USD (1M SEK) increases participation probability by 12.0 percentage points at $s = 0$. The estimated treatment effect among nonparticipants is similar in the four years following the lottery, though less precisely estimated as we extend the time horizon.[6] In contrast, the estimated effect for pre-lottery participants (for whom lottery wealth might increase participation by discouraging equity market exit) is small and mostly not statistically distinguishable from zero.

[6]There are two reasons why confidence intervals widen. First, participation is only observed during a nine-year period and we condition on prior participation status, so the sample size decreases with time horizon. Second, the predictive power of lagged financial and demographic characteristics falls with time, increasing the standard errors.



(a) Nonparticipants



(b) Participants

Figure 2: Effect of 150K USD (1M SEK) of Lottery Wealth on Participation Probability by $s = -1$ Participation Status. Coefficients and 95% confidence intervals are obtained by estimating Equation 1 in the post-1999 sample of nonparticipants (a) and participants (b). See Appendix Table B.4 for the underlying estimates.

Hence, the aggregate effect of 3.9 percentage points we observe in the pooled sample appears to be driven nearly entirely by a positive effect on nonparticipants.

3.1.1 Effects by Prize Size

Because large prizes account for most of the identifying variation, our linear estimator assigns most weight to the marginal effect of lottery wealth at modest to large levels of wealth. To test for nonlinear effects, we replace the lottery-wealth variable in Equation 1 by indicator variables for five categories defined according by prize size and run regressions with the smallest prize category omitted.

Figure 3 presents the estimated coefficients for each of these categories, with coefficients marked at the mean prize size in each category. Relative to small prize winners (<1.5K USD, 10K SEK), a prize in the range 1.5 to 15K USD (10K-100K SEK) increases the participation probability of pre-lottery nonparticipants by 1.4 percentage points. The corresponding estimates for winners of prizes in the 15 to 150K (100K-1M), 150 to 300K (1M-2M), and 300K+ (2M+) are 8.2, 17.7 and 39.9 percentage points. Thus, the marginal effect (defined as the slope between points in Figure 3) is everywhere positive, but largest for winners of small prizes. Among pre-lottery participants, none of the prize-category coefficients are statistically distinguishable from zero.

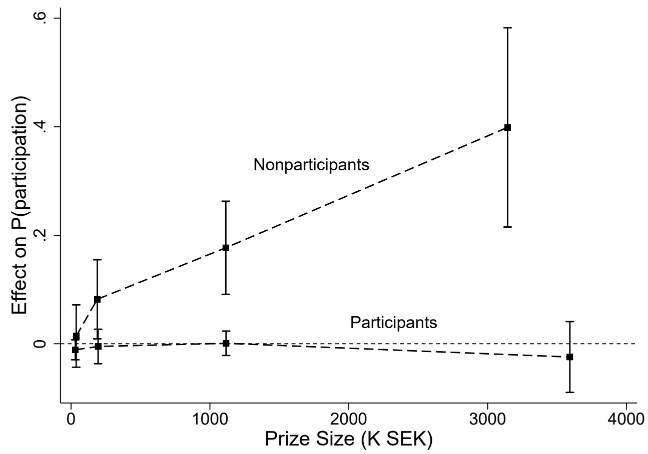


Figure 3: Effect of Lottery Wealth on Participation Probability by Prize Size. Coefficients are obtained by estimating Equation 1 in the post-1999 sample with the lottery wealth variable replaced by indicators for five mutually exclusive prize categories: 0 to 1.5K USD (0 to 10K SEK), 1.5 to 15K (10K to 100K), 15 to 150K (100K to 1M), 150 to 300K (1M to 2M), and 300K+ (2M+). Coefficient estimates and the 95% confidence bands are plotted at the mean prize in each category. See Appendix Table B.5 for the underlying estimates.

3.1.2 Effects by Prize Payment: Lump Sum vs. Monthly Installments

The finding that a majority of pre-lottery nonparticipants who won the largest prizes (300K+ USD (2M+ SEK)) do not buy stocks suggests a large disincentive to equity market entry. Such disincentives are often modeled as either a one-time entry cost or per-period participation costs. To help distinguish between these explanations, we exploit the "Triss-monthly" subsample that received monthly installments instead of a lump-sum prize. If up-front costs determine stock market participation and winners cannot perfectly borrow against future installments, a liquid lump-sum prize would result in a larger effect on participation than illiquid monthly installments.

Figure 4 shows the effect for nonparticipants by type of prize in the Triss lottery (the results for participants are shown in Appendix Table B.6). For winners of monthly installments, the effect per 150K USD (1M SEK) in net present value is close to zero at all horizons. In contrast, each 150K USD (1M SEK) paid as a lump sum increases participation probability by 10.5 percentage points at $s = 0$ and the estimated effect is positive (though not always statistically significant) in all subsequent years. These differences by payment plan suggest that up-front costs are more likely

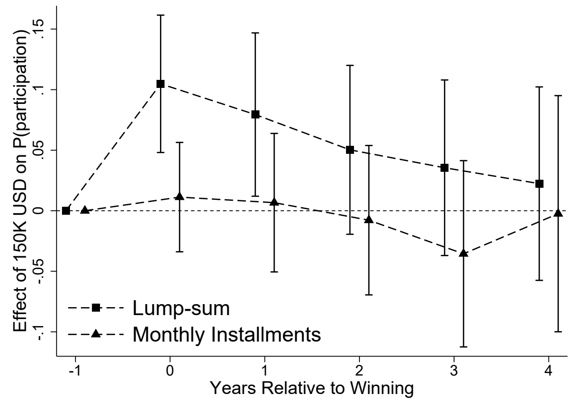


Figure 4: Effect of 150K USD (1M SEK) of Lottery Wealth on Participation by Payment Form. Coefficients and 95% confidence intervals are obtained by estimating Equation 1 in the post-1999 sample of Triss winners stratified by type of payment plan (lump-sum or monthly installments) for $s = -1$ equity market nonparticipants. See Appendix Table B.6 for the underlying estimates and corresponding estimates for of $s = -1$ participants.

to disincentive participation than continued costs of participation.[7]

In classical models with complete markets (e.g., Samuelson (1969)), participation and entry costs are equivalent and the household problem can be simplifed to a static setting. However, stark differences in the effects on entry by payment plan suggest that a simplified model (e.g., Vissing-Jørgensen (2003)) is insufficient to identify the structure of participation disincentives. Correct inference instead requires application of an appropriate economic framework, which we turn our attention to in Section 4.

3.2 Robustness

We conduct a number of sensitivity checks to explore the robustness of our results.

A first set of robustness analyses examine the sensitivity of our results to alternative definitions of participation. In our main analyses, a player is classified as participants if they or their spouse own stocks or mutual funds. Estimates do not change appreciably if we only classify households with directly held stocks as participants (Appendix Table B.3, Panel B) or exclude spousal assets

[7]One complicating factor when comparing Triss-Lumpsum and Triss-Monthly is that the support of the prize distribution in the two lotteries differ (50,000 to 6M SEK in Triss-Lumpsum, a net present value of 1.1 to 10.5M SEK in Triss-Monthly). We therefore exclude Triss-Monthly prizes above 6M SEK in our analyses. Unshown analyses confirm the difference between Triss-Monthly and Triss-Lumpsum is robust to an alternative estimation strategy that uses the panel dimension of the data and compares winners before and after the lottery.

from the participation definition (Appendix Table B.7, Columns 1-3). The main results are also robust to alternative treatments of two types of securities, capital insurance and structured products, which are composed of other assets that might have equity exposure. The Wealth Register only records the total value of these assets, but not the composition of underlying assets. Appendix Table B.7, Columns 3-6 and 7-9 show that estimated effects of wealth on participation are slightly larger after broadening the definition of participation to include individuals with structured products and capital insurance, respectively. In Section 5.2.2 we discuss further what inferences can be made from entry into the structured product market.

Our next analyses address concerns that some individuals with private pension plans may be misclassified as nonparticipants, since private pension assets are not measured in the Wealth Register. Fortunately, private pension plans were rare during our study period, and our data set does contain annual measures of private pension income. We therefore reran our main analysis in a subsample of players who had reached retirement age and had zero private pension income at the time of win. As shown in Columns 10-12 of Appendix Table B.7, misclassification due to unobserved private pension wealth is unlikely to meaningfully affect our results.[8] Private business equity, which does not constitute stock market participation but does reflect equity ownership, is also unmeasured in the Wealth Register (see Nekoei and Seim (2018) for details). Columns 13-15 of Appendix Table B.7 show that classifying individuals whose main source of income comes from their own incorporated business as equity owners has virtually no effect on our estimates, while Section 5.1 shows that self employment income actually falls following a wealth shock. Although we don't observe investment in private businesses in which the individual is not employed, observed indicators of private business investment do not suggest results are sensitive along this dimension.

Our next analyses address potential concerns about selection and external validity. Appendix B.7, Columns 16-21 shows that the results are similar across lotteries.[9] Since selection into lot-

[8]In addition to private pension plans, part of our sample may hold equity via the public or occupational pension systems. A reform in 1999 allowed workers born in 1938 or later to decide how pension funds corresponding to 2.5% of their salary were to be managed. By the late 1990's, most private sector workers were also able to choose the management of a small share of their occupational pensions, a possibility that in 2003 was extended to workers in centralized and local government. Neither of these types of pension funds are observable in our data. However, 55% of the winners in our data were born prior to 1938 and were thus unaffected by the reform to the public pension system. Appendix Table B.8 also shows the results do not vary appreciably with age.

[9]We exclude PLS from this comparison because, as noted in Section 2.3, PLS contributes little identifying variation to the post-1999 sample we focus on here.

teries is different, this similarity is reassuring. Yet, concerns regarding selection extend beyond selection into lotteries. For example, it is well established in the literature that standard frameworks do not capture the behavior of individuals who are older and wealthier, and yet elect not to own stocks (see, e.g., Vissing-Jorgensen (2002); Campbell (2006)). To address concerns that selection of wealthier and older nonparticipants drive our results, we reran the analyses in subsamples stratified by pre-lottery wealth and age quartiles. The results are summarized in Appendix Table B.8. Despite marked differences in pre-lottery participation rates, the estimated effects are generally similar across all age and the bottom three wealth quartiles. The estimated effect of lottery wealth is significantly smaller in the top-wealth quartile. However, since the households in this subsample only account for 7.1% of our nonparticipant sample, they contribute little to the overall estimate.

Finally, Columns 22-24 of Appendix B.7 show that probit marginal effects are similar to the OLS estimates reported in the main analyses. Results are also robust to dropping small (<7.5K USD, <50K SEK) prizes (Columns 25-27), but estimates increase slightly when we drop large (>225K USD, >1.5M SEK) prizes (Columns 28-30). The latter effect reflects the decreasing marginal effect of lottery wealth shown in Figure 3. Overall, our results appear robust to alternative participation definitions, sample restrictions, and estimation strategies.

4. Structural Analysis

Previous structural work have shown that modest costs of entry and/or participation – which proxy for the totality of time costs, financial costs, and behavioral disincentives – are sufficient to disincentivize low-wealth households from purchasing equity and match observed participation patterns.[10] In this section, we estimate a structural model to analyze whether this conclusion holds up also in our quasi-experimental data.

4.1 Model Specification

Each period, an agent of age t chooses how much to consume C_t, save A_t, and what fraction α_t to invest in equities given their normalized cash on hand X_t, prior equity market participation status I_t, permanent income P_t, and lottery prizes L_t.

[10]Examples in this literature include Gomes and Michaelides (2005), Cocco et al. (2005), Cocco (2005), Alan (2006), Benzoni et al. (2007), Khorunzhina (2013), Cooper and Zhu (2016), and Fagereng et al. (2017).

4.1.1 Demographics

Each agent in our model is a single household with a fixed marital status $m \in \{0, 1\}$. Households fall into one of three education groups: high school education ($e = 0$); some post-secondary education ($e = 1$), and college degree or higher ($e = 2$). Life lengths are stochastic and finite – households survive from age t to $t + 1$ with probability π_t, and die with certainty at age $T = 100$ if they survive to that age.

4.1.2 Preferences

Agents have Epstein and Zin (1991) preferences

$$V_t = \left\{ (1 - \beta\pi_t) C_t^{1-1/\psi} + \beta \mathbb{E} \left[\pi_t V_{t+1}^{1-\rho} + (1 - \pi_t) b(X_{t+1})^{1-1/\psi} \right]^{\frac{1-1/\psi}{1-\rho}} \right\}^{\frac{1}{1-1/\psi}}, \tag{3}$$

where C_t is consumption, β is the time discount factor, ρ is risk aversion, ψ is the intertemporal elasticity of substitution, and b is a bequest multiplier.

4.1.3 Income, Assets, and Housing

Each year alive, agents receive labor income Y_t. Before retirement, income is risky and follows the standard specification

$$\begin{aligned} Y_t &= exp(f(t, m, e)) P_t U_t \\ P_t &= P_{t-1} N_t, \end{aligned} \tag{4}$$

where $f(t, m, e)$ is a deterministic function of age, education, and marital status, P_t is a permanent income component with innovation N_t, and U_t is a transitory income shock. We assume that lnN_t and lnU_t are normally distributed with education-dependent variances, respectively denoted $\sigma_{N,e}$ and $\sigma_{U,e}$, and means such that their exponent has mean one. Furthermore, lnN_t is allowed to covary with equity returns as detailed below.

At retirement age $t_R = 65$ income becomes non-stochastic and is defined by a replacement rate λ of the age-65 permanent component of income, where λ varies with education and marital status. Thus, $Y_t = \lambda P_{t_R}$ for all $t \geq t_R$.

Agents have two assets in which they can invest: a risk-free asset that pays out certain return R_f and a risky equity that pays stochastic return R_t^s. Equity returns are assumed to be lognormal, with mean excess return μ_s. Log equity returns are denoted

$$r_t^s - r_f = \mu_s + \epsilon_{s,t}, \tag{5}$$

where $\epsilon_{s,t}$ is distributed normally with standard deviation σ_s, and $corr(lnN_t, \epsilon_{s,t}) = \rho_{n,r}$. The share of savings a household allocates to equities is denoted by α_t. We assume that households cannot hold short positions in either asset, so $\alpha_t \in [0, 1]$.

We do not formally model housing investment or utility, but follow Gomes and Michaelides (2005) in modeling housing expenditures as an age-dependent mandatory payment expressed as a share of income. Thus, housing expenditures of amount $H_t = h(t)Y_t$ are subtracted from cash on hand at the start of each period.

4.1.4 Entry and Participation Costs

Households investing in equities pay two types of financial costs. The first time a household invests in equities (i.e., $\alpha_t > 0$), they must pay an entry cost χ. In addition, a per-period participation cost κ is paid in each period an agent allocates non-zero wealth to equity holdings. Participation statuses are denoted as I_t and $Part_t$, where $I_t = 1$ indicates whether a household has ever owned equities and $Part_t$ denotes the current period's participation decision. The total cost of investment is written

$$((1 - I_t) \times \chi + \kappa)) \times Part_t. \tag{6}$$

In our baseline model we assume that costs are constant across the population, but in Section 4.7 we extend the model to allow for entry cost heterogeneity.

4.1.5 Lottery Prizes and Wealth Accumulation

To align the model with our empirical setting, households can receive unanticipated lottery winnings L_t. Households do not form expectations over the prize distribution, meaning that prizes are exogenous and unexpected. Whenever lottery winnings L_t are positive, they enter additively into the budget constraint.

The intertemporal budget constraint is the difference between the sum of income, lottery prizes, and returns on the previous year's non-consumed cash on hand and the sum of housing expenditures and investment costs:

$$X_{t+1} = \left[R_f + \alpha_t(R^s_{t+1} - R_f)\right](X_t - C_t) + \\ Y_{t+1}(1 - h_t) - ((1 - I_t) \times \chi + \kappa)\, Part_t + L_{t+1}. \tag{7}$$

4.1.6 Decision Problem and Model Solution

The household decision problem is formally specified in Appendix A.1. To solve the model we exploit the model's homotheticity and normalize all other states and controls by P_t (normalized variables are subsequently indicated as lower cased). The model is then solved by backward induction for each education and marital status. More details on the model solution are presented in Appendix A.3

4.2 First Stage Calibration

Table 5 presents parameters calibrated externally from the model. Panel A shows parameters that characterize asset returns. The risk-free rate is $r_f = .02$, and the excess return and standard deviation on equities are $\mu_s = .04$ and $\sigma_s = .21$, respectively. The assumed equity premium is thus approximately 4.4%, below the historical 6.5% equity premium in Sweden (see Waldenström (2014)). Calibrating a lower than historically observed equity premium is common in the literature to reflect unmodeled asset management fees, which are estimated to reduce returns to Swedish households by 2% (Calvet et al. (2007)). Because of this calibration choice, participation costs κ should be thought of as excluding investment fees.

The procedure used to calibrate income processes is described in Appendix A.5. Income processes, including age profiles ($f(t, m, e)$) and parameters governing income risk ($\sigma_{U,e}$, $\sigma_{N,e}$, and $\sigma_{N,R,e}$) differ by group and marital status. Appendix Figure A.2 shows that average income profiles are hump-shaped and differ in level across education groups. Panel B presents the remaining parameters that characterize the income processes. Income innovation parameters are similar across education groups, and in all groups the estimated correlation between equity returns and permanent income innovations is negligible. Overall, our estimates of income risk are comparable to values estimated in the United States (e.g., Carroll (1997); Gourinchas and Parker (2002); Cocco et al. (2005)). Retirement replacement rates (λ_{t_R}) are approximated as proposed in Laun and Wallenius (2015), with details included in Appendix A.6.

Other calibrated parameters include survival probabilities, which are calibrated to observed mortality rates (see Appendix A.4), and housing expenditures, which are calibrated to be 30% of

Table 5: First-Stage Calibration. This table presents parameter values determined separately from our structural estimation procedure. Panel A presents the risk-free rate and the mean and standard deviation of the excess equity return distribution defined in Equation 5. Panel B shows the standard deviation of transitory and permanent income innovations, the correlation of permanent income innovations with equity returns, and the replacement rates of retirement income for each education group. Appendix A.5 details the estimation of income parameters.

A. Asset Returns		**B. Income Processes by Education Level**			
			No College ($e = 0$)	**Some College** ($e = 1$)	**College** ($e = 2$)
Equity Mean: μ_s	.04	**Transitory Risk:** σ_U	.188	.188	.205
Equity Risk: μ_s	.21	**Permanent Risk:** σ_N	.110	.106	.110
Risk Free Return:r_f	.02	**Equity correlation:** $\rho_{n,r}$	.002	-.001	-.008
		Rep. Rate, Single: λ_{t_R}	.685	.641	.617
		Rep. Rate, Married: λ_{t_R}	.644	.608	.589

income while working and 20% of income in retirement.

4.3 Estimation Methodology

We estimate the remaining preference parameters and costs, namely the time discount factor β, risk aversion ρ, intertemporal elasticity of substitution ψ, entry cost χ, and participation cost κ. Hereafter, this vector of parameters is referred to as $\theta = [\beta, \rho, \psi, \chi, \kappa]$. To estimate θ we follow the empirical policy function (EPF) approach proposed in Bazdresch et al. (2017).

An EPF is an estimate of the relationship between state variables and policy functions in a structural model. EPFs provide useful benchmarks to evaluate model fit and to identify structural parameters by minimizing the distance between approximations of the model-defined policy functions and their corresponding estimates from the data. Formally, the consumption and participation policy functions from our structural model are written as functions of normalized state variables (t, x, I, l)

$$\begin{aligned} c_i &= c(t_i, x_i, I_i, l_i) \\ Part_i &= Part(t_i, x_i, I_i, l_i). \end{aligned} \tag{8}$$

These policy functions are approximated via a semi-parametric regression using a sequence of approximating functions $(h_j(t,x,P,I,l))_{j=1}^J$ such that

$$c_{i,s} \approx \sum_{j=1}^{J} b_j^C h_j(t_{i,s}, x_{i,s}, I_{i,s}, l_{i,s}) + \eta_{i,s}^C \tag{9}$$
$$Part_{i,s} \approx \sum_{j=1}^{J} b_j^{Part} h_j(t_{i,s}, x_{i,s}, I_{i,t,s}, l_{i,s}) + \eta_{i,s}^{Part}$$

where $s = 0$ denotes the year of the lottery event. We include linear and quadratic terms for continuous variables (t, x), indicator variables for discrete states (I), and a constant. l_s is omitted from our pre-lottery EPFs as $l_{i,s} = 0$ globally, but included as a linear term in years $s \geq 0$.[11] Details on the exact specification of EPFs for all estimation exercises are included in Appendix A.7.

Registry data from Statistics Sweden is used to construct the variables in Equation 9. All right-hand side variables are observed directly, as is participation. Although not observed directly, consumption is constructed from the budget constraint as

$$c_{i,s} = \left[R_f + \alpha_s(R_{t+1}^s - R_f)\right] x_{i,s} + \frac{y_{i,s+1} + l_{i,s} - x_{i,s+1}}{\left[R_f + \alpha_s(R_{t+1}^s - R_f)\right]} \tag{10}$$

and permanent income, which normalizes all continuous variables, is constructed as described in Appendix A.5.

Using the EPFs defined above, Bazdresch et al. (2017) adapt the indirect inference procedure proposed in Smith (1993) and Gourieroux et al. (1993) to estimate θ. Define $\nu_{i,s}$ as a vector of empirical observations and let $\nu_{i,s}^k(\theta)$ be the corresponding vector of observations from model-simulation $k = 1, ..., K$ given θ. Our identifying moments are coefficients b_j from Equations 9, and moment conditions are specified as the vector of differences between model-implied and

[11]Because income processes differ by education and marital status these are also state variables. We do not include these in our baseline EPF specification to maintain model parsimony and symmetry to estimating Equation 1. Subsequent preference parameter estimates can be thought of as the average preferences across education and marital groups. Similarly, we only consider linear effects of lottery wins l_s in our baseline estimation, but allow for nonlinear effects of lottery wins later in this section. Qualitative results are similar if we allow for richer and higher order approximating series in our EPFs, but model fit is worse.

empiricial coefficients:

$$g(\nu_{i,s},\theta) = \mathbb{E}\left[b_j(\nu_i) - \frac{1}{K}\sum_{k=1}^{K} b_j(\nu_{i,s}^k(\theta))\right]_{\forall j}. \tag{11}$$

Parameter estimates $\hat{\theta}$ are determined by

$$\hat{\theta} = \arg\min g(\nu_{i,s},\theta)'\hat{W}g(\nu_{i,s},\theta), \tag{12}$$

where $\hat{W}$ is the optimal weighting matrix estimated using the procedure described by Erickson and Whited (2002). Specifically, $\hat{W}$ is the inverse of the clustered covariance matrix $\hat{\Omega}$ of $m(\nu_{i,s})$'s stacked influence functions (denoted $\phi_{m(\nu_{i,s})}$):

$$\hat{\Omega} = \frac{1}{NS}\sum_{i=1}^{N}\left(\sum_s \phi_{m(\nu_{i,s})}\right)\left(\sum_s \phi_{m(\nu_{i,s})}\right)'. \tag{13}$$

Because the moment vector m consists of coefficients from an OLS regression and Equation 13 does not depend on θ, the influence functions (and thus the optimal weighting matrix) need only be calculated once as the standard OLS influence functions from the empirical estimates of b_j.

Our initial estimation exercise only uses observations prior to the lottery event. Each household in our post-1999 is sampled in periods $s = -4, ..., -1$ (or earliest observed period if first observation $s_i > -4$) and all state variables are recorded (including observed lottery prizes, where $l_i = 0$ since $s < 0$). Using these observations, we estimate Equation 9 to generate empirical moments $b(\nu_i)$. To generate the model implied moments, we use these same observations and the optimal policy functions to simulate the one-period ahead data set, and then estimate Equation 9 using this simulated data set to recover the 12 coefficients targeted in our baseline estimation. We repeat this simulation $K = 5$ times, construct moment conditions as defined by Equation 11, and calculate the objective function defined in Equation 12. We iterate on this procedure until the objective function converges to its minimum value.

Subsequent estimation exercises simulate household responses to lottery wins. When simulating lottery wins, the procedure is the same except we sample households only in period $s = 0$ and simulate responses assuming sampled prize $l_{i,0}$ enters the budget constraint as detailed in Equation 7. Lottery prizes are shuffled within prize group $X_{i,0}$ across simulations so that the simulated distribution of lottery prizes corresponds exactly to the observed distribution. In addition, we add lottery-cell fixed effects to Equation 9 as detailed in Appendix A.7. Finally, to evaluate model fit we use the standard Wald test for overidentification as well as the Wald test for external validity proposed by Bazdresch et al. (2017) (see Appendix A.2 for details).

4.4 Structural Estimation with Pre-lottery Data

Our estimation results based on pre-lottery decisions are presented Table 6, Column 1. Panel A shows estimates and standard errors for the preference parameters, entry cost, and participation costs. To facilitate comparison to other studies, the following text discusses these parameter estimates in the context of two recent studies, one with a similar sample and one with a similar model. Fagereng et al. (2017) (hereafter FGG) estimate a model with CRRA preferences using a representative sample from Norway (where institutions are similar to Sweden), while Cooper and Zhu (2016) (hereafter CZ) estimate a model with Epstein-Zin preferences and income heterogeneity by education status using an American sample.

Turning to the estimated preference parameters, a time-discount factor ($\beta = .869$) that is lower than most macro models, is necessary to limit wealth accumulation. FGG (estimates between $.75 - .83$) and CZ ($.76 - .90$) also estimate low time-discount factors for the same reason. Our estimates also suggest – again similar to FGG and CZ – that a bequest motive ($b = 5.191$) is needed to slow asset decumulation during retirement. Finally, risk aversion ($\rho = 3.162$) and the elasticity of intertemporal substitution ($\psi = .645$) estimates are comparable to the baseline estimates in CZ ($\rho = 4.409$, $\psi = .601$). Because $1/\rho = .316$ is significantly lower than $\psi = .645$, the estimates reject a time separable CRRA model in which $1/\rho = \psi$.

Estimated entry and participation costs are modest relative to total wealth. Per-period costs of stock market participation are economically insignificant at only 10 USD per year. The low costs reflect the persistence in equity market participation: if per-period participation costs were higher, a higher fraction of equity market participants would leave equity markets than what we see in the data. The entry cost, which is identified by the entry decisions of nonparticipant households, is estimated to be 3,217 USD. For comparison, FGG estimate per-period participation costs of 65 to 344 USD, while CZ estimate an entry cost of 684 USD and a transaction cost of 1,368 USD. Our slightly higher entry cost estimate relative to FGG and CZ reflects a difference in the estimation procedure. In our case, the entry cost reflects the average cost for nonparticipants (presumably participant households in our sample had lower costs of entry) instead of the cost required to generate life-cycle participation rates.

Table 6: Structural Estimation Results and Predictions. Column 1 presents results when the model is estimated using only pre-lottery observations and matching pre-lottery EPF coefficients, Column 2 using only post-lottery observations and matching post-lottery EPF coefficients, Column 3 using observations both pre- and post-lottery data and matching pre- and post-lottery EPF coefficients, and Column 4 using post-lottery observations to estimate the entry cost distribution (Figure 5) that matches linear and nonlinear EPF coefficients on lottery wealth assuming other parameters are fixed at their values in Column 1. Panel A presents the estimated parameters, Panel B presents the model's predictions of the effect of lottery wins on participation probability, Panel C presents tests of external fit for the indicated sets of lottery coefficients in Panel B. In all cases the post-1999 sample is used, and the corresponding coefficients from regressions on consumption are presented in Appendix Table A.1.

	Pre-Lottery (1)	Post-Lottery (2)	Pre- & Post (3)	Nonlinear (4)
A. Parameter Estimates ($\hat{\theta}$)				
Time Discounting - β	.869	.902	.896	.869
	(.019)	(.012)	(.006)	–
Bequest - $\mathbf{b}$	5.191	1.32	3.106	5.191
	(1.668)	(.688)	(1.700)	–
Risk Aversion - ρ	3.162	2.360	2.342	3.162
	(.097)	(.091)	(.211)	–
IES - ψ	.645	.595	.669	.645
	(.077)	(.070)	(.063)	–
Entry Cost (K USD)- χ	3.217	31.262	12.503	–
	(1.668)	(.688)	(.859)	
Participation Cost (K USD) - κ	.001	.004	.036	.001
	(.003)	(.003)	(.006)	–
Overidentifying χ^2 (d.f.).:	35.1 (6)	93.3 (10)	1525.6 (22)	–
N	$192,524$	$70,139$	$262,663$	$70,139$

B. Lottery Estimates vs. Model Predictions

	Benchmark	Model Predicted Effect (1)	(2)	(3)	(4)
i. Linear Effect (150K USD)					
All	.028	.101	.030	.067	.029
Nonparticipants	.104	.313	.113	.209	.104
Participants	.002	.000	.000	.000	.000
ii. Nonlinear, Nonparticipants					
$1.5K < L \leq 15K$	-.012	.013	.013	-.012	.006
$15K < L \leq 150K$	.078	.172	.017	-.016	.080
$150K < L \leq 300K$	.156	.644	.026	.462	.158
$300K < L$	.359	.953	.591	.976	.357

C. External Validity Test, χ^2 (d.f.)
(untargeted coefficients, Panel B)

	(1)	(2)	(3)	(4)
Linear and nonlinear (B.i,ii)	1084.5 (7)	–	–	–
Nonlinear (B.ii)	441.5 (4)	127.8 (4)	390.3 (4)	–

The model's EPF moments reasonably approximate their empirical counterparts (see Appendix A.7). Given the overidentification test has excellent power to detect even small differences between the model and data generating processes (Bazdresch et al. (2017)), it is unsurprising that the standard overidentification test statistic $\chi^2 = 35.1$ is rejected at all significance levels. Despite this rejection, Appendix A.7 shows the model replicates empirical coefficients with reasonable accuracy. As a further credibility check, Appendix A.8 compares the model's predicted wealth and participation profiles to the empirical age-profiles of wealth and stock market participation. These profiles, commonly targeted in other studies, are not targeted in our estimation procedure. Nevertheless, they are matched reasonably well. Our estimates slightly overpredict wealth accumulation early in life and decumulation later in life, but otherwise decently approximate life-cycle saving and participation patterns.

Table 6, Panel B compares the model's predictions of the effect of lottery wins on participation to their empirical counterparts (displayed at the left-hand side of panel B).[12] Panel B.i shows that the model predicts each 150K USD (1M SEK) increases stock market participation probability by 10.1 percentage points in the full sample, 3.6 times larger than the empirical estimate of 2.8 percentage points. This overprediction is driven by a predicted 31.3 percentage point effect on participation probability among nonparticipants, as entry costs are not large enough to discourage enough large prize winners from entering the stock market (see Panel B.ii). The model does match the near-zero effect of lottery wealth on participation for participants who, given the negligible participation costs, are predicted to continue participating regardless of lottery prize size. Overall, the baseline estimation exercise predicts responses to lottery wins that are qualitatively consistent with the main results in Section 3, but quantitatively much larger. Panel C formally documents this poor fit and shows that the test for external validity proposed by Bazdresch et al. (2017) is strongly rejected (see Panel C, row 1).

4.5 Structural Estimation with Lottery Data

To understand what model parameters – in particular entry costs – are needed to account for our lottery results, we re-estimate our model using only participation decisions after the lottery event. The model targets 16 benchmarks: all coefficients except for cell fixed effects from the post-lottery participation and consumption EPFs, and the lottery coefficients from participation and

[12] Appendix A.7 details the exact regressions we estimate to obtain the model-predicted lottery coefficients. EPF coefficients on lottery wins slightly differ from lottery coefficients presented in Section 3 due differences in specification between Equations 1 and 19.

consumption regressions by pre-lottery participation status. The exact regressions, coefficients, and resulting model fit are presented in Appendix A.7 and Table A.1. The optimal weighting matrix is again calculated as the inverse of the influence function from these regressions.

Column 2 of Table 6 presents the results from this estimation. Preference parameter estimates are mostly similar to those from pre-lottery data. Entry costs are, however, estimated to be 31,262 USD, an order of magnitude larger than our baseline estimate. This cost is quite significant economically and corresponds to approximately 30% of average wealth or 70% of annual income in our sample. It is difficult to reconcile such a high cost of entry with any reasonable cost that households might pay to enter equity markets. However, these large costs are intuitive: matching low rates of equity market entry after receiving large lottery prizes requires a large disincentive, which in our model is best reflected by the entry cost χ.

The standard overidentification test statistic $\chi^2 = 98.2$ is rejected at all significance levels. A main reason for this rejection is that the model is unable to generate MPCs from lottery wealth as high as those observed in the data and still match the consumption policies of households that did not win large lottery prizes. Despite the statistical rejection, Appendix A.7 shows that the model generally matches the empirical coefficients. Furthermore, Appendix A.8 shows that the model reasonably approximates life-cycle wealth profiles, although large entry costs reduce stock market entry over the life-cycle to virtually zero. Finally, we test and reject the external validity of the model's predicted nonlinear effects of lottery win on participation in Panel C. A one-time cost of 31K USD disincentivizes virtually all winners except those of more than 300K USD (2M+ SEK) from entering equity markets, while empirical estimates suggest larger effects on winners of smaller prizes and smaller effects on winners of larger prizes. Thus, the model has difficulty replicating effects on consumption and prize size heterogeneity with a single large cost, but can replicate most other patterns.

4.6 Structural Estimation with Pre-Lottery and Lottery Data

A correct model of stock market participation should be able to account for participation decisions both before and after lottery wins. Therefore, in Column 3 of Table 6 we estimate our model targeting the combined pre- and post-lottery benchmarks matched separately in Columns 1 and 2. In practice, we stack the two moment vectors from our previous two estimations, and re-estimate Equation 12 with the optimal weighting matrix defined by the inverse of the covariance matrix of these stacked influence functions.

The resulting parameter estimates are mostly similar to those obtained from targeting pre-lottery and lottery coefficients separately (Columns 1 and 2, respectively). The main parameter of interest, the one-time entry cost, is estimated to be 12,503 USD. This estimate is closer to the baseline estimate of 3,217 USD than the lottery estimate of 31,262 USD because the standard error of the lottery wealth coefficient is relatively large and the optimal weighting matrix assigns more weight to better identified moments. However, including the lottery estimates and their larger implied barriers to entry does increase the entry cost estimates by over 9K USD relative to the baseline.

The overidentification test statistic shown in Panel A indicates that the model's fit is strongly rejected, with predictions in Panel B generally falling between the pre-lottery and lottery predictions (Columns 1 and 2). Furthermore, our test of external validity for untargeted, nonlinear effects in Panel C is rejected at all significance levels. These rejections and relatively poor fit highlight the challenge faced by our model in simultaneously matching non- and quasi-experimental consumption and participation policies.

4.7 Structural Estimation with Entry Cost Heterogeneity

The model's predicted effects by prize size are soundly rejected in the first three structural estimation exercises (Panel B.iii). These rejections highlight an unanswered economic question: What size and structure of participation disincentives enable our model to match the full distribution of household participation responses to lottery wins?

To answer this question, in Column 4 we conduct a final structural exercise that extends the model to allow for heterogeneity in entry costs as determined by the cost distribution

$$\chi_i \sim G_\chi(x). \tag{14}$$

We approximate this distribution by seven equi-distant points between 2K and 70K USD and estimate the the probability mass for each point of this distribution, holding other parameters fixed at their baseline pre-lottery estimates. Our moment vector includes the estimated effects of lottery winnings on participation, namely the seven coefficients from Panel B.i-ii of Table 6.[13] The resulting entry cost distribution thus reflects the entry disincentives needed to match our lottery estimates, including the effects by prize size.

[13] The simulation procedure is almost the same as that in Section 4.3, except in each simulation we sample an entry cost for each household from the proposed cost distribution. Our estimation procedure is also unchanged – we estimate Equation 12 with optimal weighting matrix determined by the stacked influence functions. Test statistics are undefined because this system is just identified.

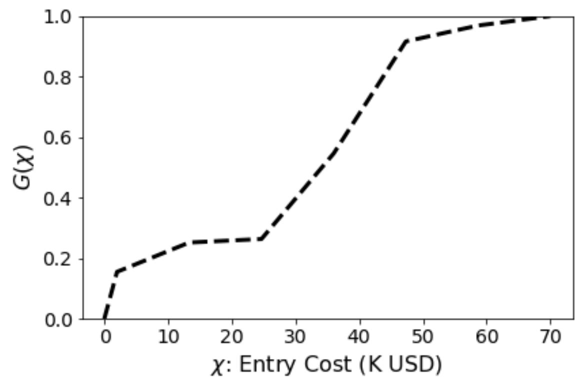


Figure 5: Structural Estimates of Entry Cost Distribution. This figure presents the estimated CDF of entry costs defined in Equation 14 in the estimation exercise that allows for entry cost heterogeneity (Table 6, Column 4). The model is estimated using using post-lottery observations from the post-1999 sample, with parameters besides entry costs held fixed at pre-lottery estimates in Table 6, Panel A, Column 1.

Panel B shows that given the flexibility of the assumed entry cost distribution the model predictions nearly exactly match their empirical counterparts. The resulting estimated cost CDF G_χ is presented in Figure 5. The mean (43K USD) and median (36K USD) implied costs of entry are estimated to be quite large, as is needed to match low entry rates. However there is significant heterogeneity in entry costs, with approximately 23% of our sample have entry costs $\leq$ 10K USD while approximately 40% have entry costs $\geq$ 40K USD. Furthermore, the shape of the estimated entry cost distribution mirrors the empirical estimates of effects by prize size. Accounting for positive effects of lottery wins on entry for small and intermediate prize winners requires some households to have small entry costs, while matching the small rates of entry of winners in our largest prize categories requires a majority of households to have large entry costs.

5. What Explains Nonparticipation?

The upshot of Section 4 is that under standard assumptions about entry costs, traditional modeling approaches predict increases in stock market participation much larger than our quasi-experimental estimates in Section 3. A simple way to align the model-based predictions with the quasi-experimental estimates is to assume entry costs at least an order of magnitude larger than those that have been reported in the literature. Clearly, costs of such magnitude are hard to interpret, since they are far larger than any plausible financial costs. In this section, we conduct a number of analyses to explore the potential roles of several factors that might contribute to the discrepancy.

Our analyses consider three broad classes of explanations: economic explanations (e.g., investment in other assets), alternative preferences (e.g., status-quo bias, loss aversion, and present-bias), and non-standard beliefs and belief formation processes. Since the model-based predictions are only wildly at odds with our reduced-form estimates for pre-lottery nonparticipants, all analyses in this section are restricted to non-participants unless otherwise noted.

To preview the findings, there is strong evidence that non-standard expectations and belief-formation processes contribute to the discrepancy between empirical and model-implied estimates. In a survey fielded to a subset of our lottery sample, many people reported subjective beliefs that are more pessimistic than historical averages. Model-based predictions that account for these subjective beliefs reduce the discrepancy by 50%. Players are also more likely to enter equity markets if they win during a period of high returns or experienced high returns during their formative years, suggesting that both recency bias and personal experiences affect equity return beliefs.

5.1 Economic Explanations

The life-cycle model of Section 4 does not allow for some investment options that, if sufficiently attractive, could crowd out demand for equities. For example, it has been suggested that investments in housing (see, e.g., Cocco (2005); Flavin and Yamashita (2011); Vestman (2018)), private business equity (see, e.g., Heaton and Lucas (2000)), or a desire to reduce high-interest debt (see, e.g., Davis et al. (2006); Becker and Shabani (2010)) could limit stock market participation.

As a first test of such crowd-out effects, we ran heterogeneity analyses in subsamples stratified by pre-lottery home ownership, presence of debt, and presence of self-employment income. The results are shown in Table 7, Panel A. Column 1 shows that the estimated effect of each 150K USD (1M SEK) on the participation probability of players who did not own their home at $s = 0$ was 14.7 percentage points, compared to 10.5 percentage points for home owners (Column 2). The estimates are not statistically distinguishable. Columns 3 and 4 show that the estimated effect in households without debt (Column 3) is about twice as large as for households with debt (Column 4). Finally, Columns 5 and 6 show that the estimated effect is smaller among the 3.5% of households with self-employment income, although estimates are imprecise due to the small sample size. Overall, these patterns are consistent with a role for unmodeled investment opportunities, especially reduction of debt.

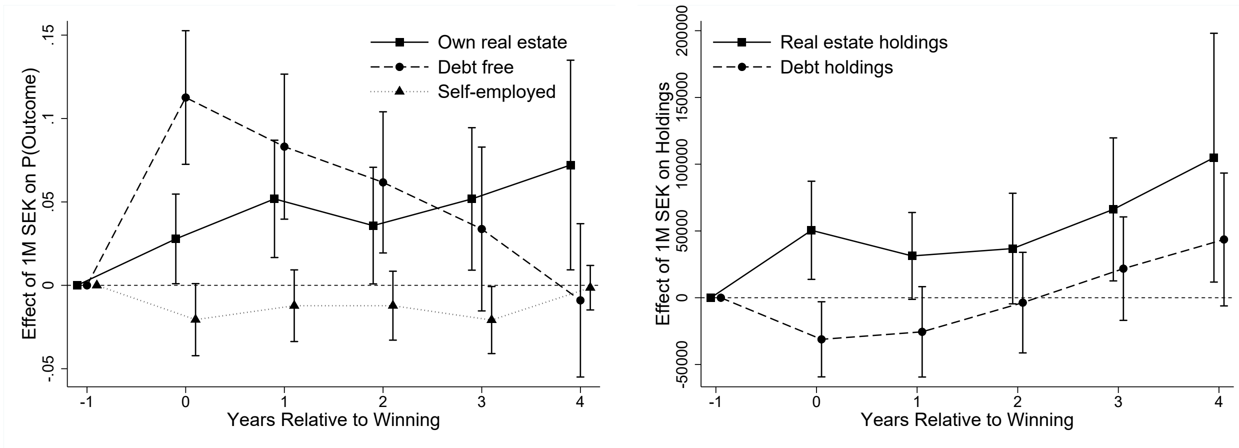
Our next analyses directly examine how lottery wealth impacts the probability of owning real estate, becoming debt free or having self-employment income in the post-lottery years (Figure 6, Panel A). We estimate that each 150K USD (1M SEK) increases the probability of being debt free

Table 7: Heterogeneous Effect of Wealth (1M SEK) on Participation Probability among $s = -1$ Equity Market Nonparticipants. Coefficients are obtained by estimating Equation 1 at time $s = 0$ in the post-1999 sample of equity market nonparticipants at time $s = -1$, stratified by the characteristics indicated in the column heads. Panel A stratifies households by financial characteristics: Columns 1 and 2 show effects for nonparticipants that do and do not own homes, Columns 3 and (4) for nonparticipants that do and do not have debt, and Columns 5 and 6 for nonparticipants that did and did not have self-employment income the year prior to the lottery. Panel B stratifies households by information proxies: Columns 7 and 8 show effects for nonparticipants that do and do not have college degrees while Columns 9 and 10 for nonparticipants that have above and below median cognitive skill among the subsample with conscription records available. Hetero p obtained from an F-test of the null hypothesis that the two lottery-wealth coefficients are identical. $\% \, Part_{-1}$ indicates the share of the post-1999 sample with the characteristic indicated by the column head that owned equity the year prior to the lottery. See Appendix Table B.9 for results for time $s = -1$ equity market participants.

	A. Financial Characteristics						**B. Information Proxies**			
	Home Owner		**Have Debt**		**Self-Employed**		**College Degree**		**Cognitive Skill**	
	No (1)	**Yes (2)**	**No (3)**	**Yes (4)**	**No (5)**	**Yes (6)**	**No (7)**	**Yes (8)**	**Low (9)**	**High (10)**
Effect	.147	.105	.212	.092	.131	.046	.107	.223	.039	.304
SE	.052	.027	.037	.025	.026	.040	.025	.053	.055	.147
p	.005	.000	.000	.000	.000	.246	.000	.000	.476	.038
Hetero p	.474		.007		.079		.050		.090	
N	8,022	11,256	9,545	9,733	18,628	650	16,510	2,768	804	957
$\% \, Part_{-1}$	.554	.784	.679	.759	.719	.832	.686	.842	.677	.790

in the year of win by 10 percentage points, but the effect appears to dissipate with time and is no longer statistically significant at $s = 4$. The estimated effect on the probability of real-estate ownership is 2.8 percentage points at $s = 0$ and rises to 7.2 percentage points in $s = 4$. We find no evidence that lottery winners are more likely to have self-employment income; if anything, the point estimates are in the opposite direction. Panel B shows the effects of lottery wealth on real estate and debt-levels.[14] On average, winners invest a small share of their lottery wealth in real estate or debt reduction: real estate wealth increases by about 4.5% of the amount won in year $s = 0$, whereas total debt falls by 3.1% of the amount won. Thus, the total share of lottery wealth allocated to real estate investments and debt reductions is less than 8%.

[14]The Swedish Wealth Register does not measure the value of private businesses, so intensive margin effects are not included in this figure.



(a) Extensive Margin **(b) Intensive Margin**

Figure 6: Effect of Wealth (1M SEK) on Real Estate, Debt Investment, and Self-employment. Coefficients and 95% confidence intervals are obtained by estimating Equation 1 in the post-1999 sample of $s = -1$ equity market nonparticipants. Panel (a) shows the effect of each 150K USD (1M SEK) on the probability of owning real estate, becoming debt free, and having self-employment income. Panel (b) shows the effect of each 150K USD (1M SEK) on real estate wealth and total debt. See Appendix Tables B.10 and B.11 for the underlying estimates and results for $s = -1$ equity market participants.

A final analysis, shown in Columns 1 and 2 of Table 8, compares the discrepancy between baseline estimates and model-based predictions in a subsample of households less likely to face investment opportunities that reduce incentives to enter equity markets. Specifically, we restrict the subsample to people who, at $s = -1$, did not have self-employment income, had low debt (<$15K) and were aged below 61 (the median age in our sample). In this subsample, the model predicts that for non-participants, each 150K USD (1M SEK) increases participation probability by 26.4 percentage points, compared to an estimated effect of 15.2 percentage points. This discrepancy is smaller but of a similar in magnitude to the discrepancy observed in the full sample.

Considered in their entirety, the results in this section therefore suggest that unmodeled investments are a small to modest factor in generating the discrepancy between our baseline estimates and model-based predictions.

5.2 Alternative Preferences

The analyses in this section are intended to shed some light on the possible role of status quo biases, loss averse preferences or present-biased preferences in accounting for the discrepancy between our empirical and model estimates. Each of these three factors has been proposed as a

Table 8: Structural Model Predictions, Alternative Specifications and Calibrations. Columns 1 and 2, respectively, present estimates and model predictions (assuming pre-lottery parameters from Table 6, Panel A, Column 1) after restricting the post-1999 sample to households with no self-employment income, debt less than 15K USD, net wealth less than 1M USD, and age less than 60. Columns 3 and 4 present estimates and model predictions from our post-1999 sample after assuming a present bias parameter $\beta = .6$ and re-estimating the model using pre-lottery data. Columns 5 and 6 present estimates and model predictions (assuming pre-lottery parameters from Table 6, Panel A, Column 1) after restricting the post-1999 sample to households with some secondary education and above median cognitive ability for those winners with available conscription records. Columns 7 and 8 present estimates and model predictions (assuming pre-lottery parameters from Table 6, Panel A, Column 1) from our post-1999 sample in which the subjective equity premium is sampled from the surveyed distribution presented in Figure 8.

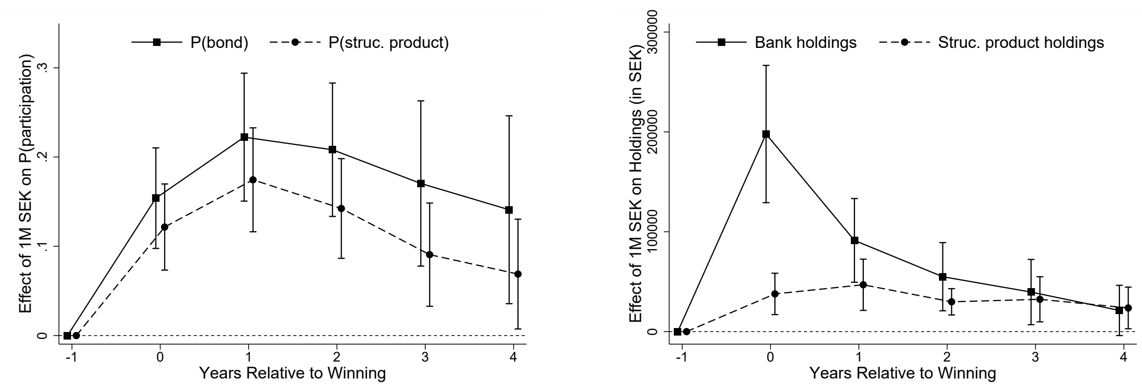
	Restricted Finances Subsample		Present-Bias Preferences		High Information Subsample		Subjective Beliefs	
	Benchmark	Model	Benchmark	Model	Benchmark	Model	Benchmark	Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
i. Linear Effect (150K USD)								
All	.040	.090	.028	.115	.013	.067	.028	.066
Nonparticipants	.145	.264	.104	.340	.163	.248	.104	.197
Participant	-.012	.000	.002	.000	-.026	.000	.002	.000
ii. Nonlinear, Nonparticipants								
$10K < L \leq 100K$	.013	.003	.-.012	-.012	–	–	-.012	-.014
$100K < L \leq 1M$	.107	.028	.078	.097	–	–	.078	.003
$1M < L \leq 2M$	.167	.465	.156	.680	–	–	.156	.452
$2M < L$	.564	.739	.359	.963	–	–	.359	.709
N	16,329		70,139		3,355		70,139	

source of non-participation.

5.2.1 Status quo bias

If households exhibit a general reluctance to actively invest their lottery wealth, such reluctance could contribute to the lower-than-predicted rates of equity participation that we observe. Status quo biases (see, e.g., Samuelson and Zeckhauser (1988)) could manifest themselves in several ways following a windfall gain.

A first is that we might expect to see large and sustained increases in account balances after the win, since prizes are automatically deposited into winners' bank accounts. Figure 7, Panel (b) shows that we do not observe such a pattern. Bank account balances increase by 20% of the prize amount at $s = 0$, but fall quickly in subsequent years (for reference, total wealth increases on average by 60% the amount won). These patterns suggest that players quickly transfer most of the



(a) Extensive Margin **(b) Intensive Margin**

Figure 7: Effect of Wealth (1M SEK) on Bank/Bond and Stuctured Product Investment. Coefficients and 95% confidence intervals are obtained by estimating Equation 1 in the post-1999 sample of $s = -1$ equity market nonparticipants. Panel (a) shows the effect of 150K USD (1M SEK) on the probability of owning bonds owning structured products. Panel (b) shows the effect of 1M SEK on total bank account balances and total structured product holdings 1M SEK on total structured product holdings (left axis). See Appendix Tables B.10 and B.11 for the underlying estimates and results for $s = -1$ equity market participants.

lottery wealth from their bank accounts.

A second manifestation of status quo bias could be that households shy away from financial products that they are unfamiliar with. If so, households would likely exhibit reluctance toward investment in any asset class that they have not previously invested in. Since fewer households own bonds than stocks, we would, if anything, expect small effects of lottery wealth on bond ownership under this hypothesis. Figure 7, Panel (a) shows that we observe the opposite: each 150K USD (1M SEK) received increases the probability of bond ownership by around 20 percentage points. Thus, winning the lottery induces many nonparticipants to invest their liquid wealth in financial assets, it is just that many players prefer financial assets other than equities.

Overall, the evidence in Figure 7 provides little evidence that status quo biases deter winners from entering equity markets.

5.2.2 Loss aversion

We next consider loss aversion, a preference specification in which individuals are more sensitive to losses than gains around a reference point (Tversky and Kahneman (1986)). Loss aversion is a commonly proposed explanation for limited equity demand (e.g., Berkelaar et al. (2004); Ang et al. (2005); Barberis et al. (2006)) with empirical support (Dimmock and Kouwenberg (2010)). To test for loss-aversion, we examine the effects of lottery wealth on retail structured products that offer capital protection against downside risk. As shown in Calvet et al. (2017), these products were widely purchased during our period of study, well-suited for loss-averse households, and popular among households that traditionally did not already participate in equity markets.

Figure 7 presents the effect of lottery wealth on structured product investment. Panel (a) shows that each 150K USD (1M SEK) increases structured product ownership by 10-17 percentage points in the years following the lottery win. However, Panel (b) shows that the level of investment in structured products is modest and never exceeds 5% of the total amount won. Furthermore, in unshown analyses we found that roughly one half of nonparticipants who entered the structured product market also entered equity markets. Thus, most nonparticipating households do not purchase assets with downside protection despite their being readily available, suggesting that loss aversion has limited scope in explaining our results.

5.2.3 Present-bias

A final alternative behavioral explanation we consider is present-biased time preferences. These preferences lower the value of future consumption, potentially making households less willing to pay entry costs and invest in equities despite their higher expected returns. To test whether present-biased preferences can account for our results, we extend our model to allow for naive quasi-hyperbolic discounting in the form of $\beta - \delta$ time preferences (Laibson (1997)). We then re-estimate the structural model (using pre-lottery data for both participants and nonparticipants) assuming a present-bias parameter of $\beta = .6$, and examine its predictions regarding lottery wins for the post-1999 sample.[15] Table 8, Column 2 shows this model still overpredicts the effect of lottery wins on participation by an amount comparable to our pre-lottery estimates.

5.3 Information- and Belief-based Explanations

A third possibility is that our structural model does not accurately characterize households' beliefs about stock market returns. Our model assumes all households believe that the logarithm of yearly stock returns is identically and independently distributed, and that the mean and variance parameters of this process are commonly inferred from historic data. However, as reviewed in Della Vigna (2009) and Benjamin (2019), people's actual belief formation processes are subject to a number of biases and thus likely to differ from the process implicitly assumed in our model. This section

[15]The assumed value of present-bias is consistent with experimental evidence in Angeletos et al. (2001) and is used by Love and Phelan (2015) in exploring the role of quasi-hyperbolic discounting in a life-cycle model with Epstein-Zin preferences.

reports a number of analyses that explore whether and how non-standard beliefs contribute to the discrepancy between the baseline estimates and the model-based predictions.

Our first analysis is motivated by prior work which has found that educational attainment and cognitive ability are positively related to financial literacy (Van Rooij et al. (2012)). A key finding in this literature is that individuals with higher and education and cognitive test scores are more likely to report beliefs that are close to what one might expect based on historical time series (e.g., Kézdi and Willis (2011); Kuhnen and Miu (2017)). If non-standard beliefs that deter stock market entry are less common among people with more years of schooling or above-median cognitive test scores, one might expect larger wealth effects in these groups. To test this hypothesis, we compare treatment effects in subsamples stratified by educational attainment. For many men in our sample, we also have cognitive test scores obtained from conscription records. A second analysis therefore compares men with above- and below-median cognitive test scores.

The results, shown in Table 7, Panel B are in the hypothesized direction. For education, we find that the each 150K USD (1M SEK) increases participation by 22.3 percentage point in households with college degrees, compared to 10.7 percentage points in remaining households. For men with above- and below-median cognitive test scores, the analogous estimates are 30.4 and 4.7 percentage points. The substantial differences in the estimated treatment effects are consistent with a major role for belief and information channels.[16] In a further analysis (Table 8, Columns 6 and 7) we restrict the sample to men with college and above-median cognitive scores. After this restriction the discrepancy between the model-predicted and reduced-form effects is 24.8-16.3=8.5 percentage points. Thus, the discrepancy is 60% smaller than our full sample, albeit still notable, in a subsample in which non-standard beliefs that deter stock entry are likely less prevalent.

To further explore the role of non-standard beliefs, we also analyzed data from a survey fielded in the Fall of 2016 to a subsample of lottery players (see Appendix C for survey details). The survey, which attained a response rate of 59%, elicited beliefs about the return on the Stockholm Stock Exchange index during the next 12 months. The distribution of responses is presented in Figure 8. Consistent with much prior work, we find that the subjective beliefs are highly heterogeneous and pessimistic on average.[17] The average expected return reported in the survey (5.9%) is

[16]We cannot rule out all alternative explanations for our results, particularly those that might be correlated or interact with the belief-formation process. For example, trust is correlated with education (Guiso et al. (2004)) and has been previously proposed as an explanation for nonparticipation (Guiso et al. (2008)).

[17]Hurd (2009) and Dominitz and Manski (2011) find substantial heterogeneity in equity return beliefs, with many households holding equity return beliefs substantially more pessimistic than historical data would suggest. In fact,

below the historical average (8.5%), over two thirds of respondents report beliefs below the historical average, and almost one in five expect negative returns. The pessimism raises the possibility that our structural analysis, which assumes expected returns calibrated to align with historical time series, substantially overstates the gains that many households perceive would accrue to them were they to enter.

To explore this possibility, we compare our baseline estimates to the predictions of a model calibrated assuming households have equity premium beliefs drawn from the distribution in Figure 8. We then generate model predictions assuming parameter values equal to the estimates obtained from pre-lottery data (Table 8, Column 1). Among nonparticipants, the original model predicted a 31.3 percentage-point increase in participation for each 150K USD (1M SEK) received, compared to our baseline estimate of 10.3. The revised model predicts an increase of 19.7 percentage points and thus reduces the discrepancy by approximately 50% from 31.3-10.3 = 20.3 to 19.7-10.3 = 9.4 percentage points. We consider the 50% figure a lower bound, because our exercise assumes all agents' beliefs are drawn from the same subjective-beliefs distribution. More likely, the distribution conditional on nonparticipation is further shifted toward the left. Accounting for this would further reduce the discrepancy between the baseline estimates and model predictions.[18]

It is natural to ask what non-standard belief-formation processes gave rise to the non-standard beliefs that can explain a substantial share of the discrepancy between our baseline estimates and the model predictions. Recent work has emphasized the possibility that some people rely on extrapolative belief-formation processes that overweight recent data (e.g. Vissing-Jørgensen (2003), Greenwood and Shleifer (2014), Gennaioli et al. (2016), Bordalo et al. (2019)) or data from formative periods (e.g., Kaustia and Knüpfer (2008), Malmendier and Nagel (2011), Kuhnen (2015)). In our final two analyses, we test for such recency- and early-life biases.

In our tests of recency bias, we compared players who won in years following positive equity returns on the Swedish Stock Exchange to those who won following negative returns. Larger treatment effects among households who win following positive returns are consistent with ex-

Hurd (2009) concludes that equity returns are sufficiently pessimistic for enough households to account for observed stock market nonparticipation. Additionally, Appendix C shows that the cross-sectional predictors of subjective equity return beliefs in our survey align with other studies. Consistent with findings in Das et al. (2019), households with higher socioeconomic status (as proxied by income and education) are generally more optimistic in their reported probability that the Stockholm Stock Exchange index would increase in value during the next 12 months, as well as being less likely to report overly negative expected returns.

[18]Unfortunately, our data are in a format that do not allow us to match the survey responses to information about participation. Therefore, we have no easy way of determining the conditional distributions.

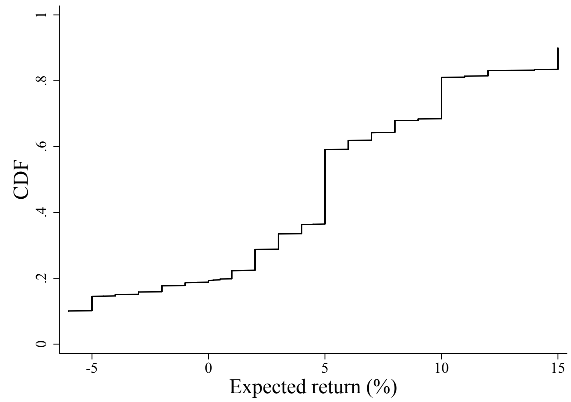


Figure 8: Subjective Distribution of Equity Returns. The above figure presents the CDF of survey respondents' expected market returns during the 12 months following the survey (i.e., Fall 2016-Fall 2017). For expositional purposes, we truncate the distribution at the 10^{th} and 90^{th} percentiles. The sample is composed of 1,749 lottery winners that responded to the survey. See Appendix C for a details on survey methodology.

trapolative belief-formation process in which returns are perceived to be persistent. In our tests of early-life bias, we compared players who experienced above average returns during formative years to players who did not. Following Fagereng et al. (2017), we define formative period is defined as the age range 18 through 25. Larger treatment effects in winners who experienced above average returns is consistent with overweighting of data from formative years.

Table 9 reports summarizes the results from several analyses of recency and early-life bias. Overall, we find support for both hypothesized belief biases, with larger effects both among players who experienced higher returns during their formative years (14.0 vs 5.3 percentage points in Columns 1-2) and players who won the lottery following a year with positive returns (17.6 vs 8.6 percentage points in Columns 3-4). Columns 5 through 8 show the results from analyses of subsamples stratified both by recent returns (low or high) and returns during formative years (low or high). Effects are largest for households that both won in years following positive equity returns and experienced high equity returns when young (18.7). Only winning after positive returns (11.1) or having experienced positive returns while young (13.8) is associated with smaller increases in participation probability, while being exposed to neither implies an effect close to zero (-0.7). These results suggest limited substitutability between these two of belief formation processes, and that some upward shift in beliefs is important in encouraging entry. While these estimates are generally too imprecise to reject equal effects with great statistical confidence, they point toward

Table 9: Heterogeneous Effect of Wealth (1M SEK) on Participation Probability among $s = -1$ Equity Market Nonparticipants, Belief and Information Channels. Coefficients are obtained by estimating Equation 1 at time $s = 0$ in the post-1999 sample of equity market nonparticipants at time $s = -1$, stratified by the characteristics indicated in the column headings. Recent equity return samples are defined by whether Stockholm Stock Exchange returns were negative or positive the year prior to the lottery. Early equity return samples are defined by whether a household experienced above or below average equity returns between ages 18-25. Education groups are defined by whether or not a household member has a college degree. Hetero p obtained from an F-test of the null hypothesis that the two lottery-wealth coefficients are identical. $\% \ Part_{-1}$ indicates the share of the post-1999 sample with the characteristic indicated by the column head that owned equity the year prior to the lottery. See Appendix Table B.13 for results for time $s = -1$ equity market participants.

	Recent Returns		**Early Returns**		**Recent Returns/Early Returns**			
	Low	**High**	**Low**	**High**	**Low/Low**	**Low/High**	**High/Low**	**High/High**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Effect	.053	.140	.086	.176	-.007	.138	.111	.187
SE	.039	.028	.030	.036	.021	.078	.037	.040
p	.167	.000	.004	.000	.730	.078	.002	.000
Hetero p	.069		.056					
N	10,402	8,876	10,591	8,687	5,678	4,724	4,913	3,963
$\% \ Part_{-1}$	.742	.703	.721	.730	.738	.747	.700	.707

	Recent Returns/College Degree				**Early Returns/College Degree**			
	Low/No	**High/No**	**Low/Yes**	**High/Yes**	**Low/No**	**High/No**	**Low/Yes**	**High/Yes**
	(9)	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
Effect	.050	.125	.111	.244	.082	.145	.126	.369
SE	.040	.030	.107	.062	.032	.037	.070	.085
p	.213	.000	.296	.000	.011	.000	.060	.000
Hetero p	.137		.283		.195		.025	
Δ Effect	.075		.132		.064		.243	
N	9,014	7,496	1,388	1,380	9,095	7,415	1,496	1,272
$\% \ Part_{-1}$	.699	.669	.865	.810	.674	.700	.851	.831

both recent and early returns as important determinants of beliefs and subsequent financial choices.

How pervasive are non-standard belief formation processes? One might hypothesize that only people with high information costs are afflicted by belief biases. To investigate, Columns 9-16 redo the analyses in Columns 1-4 separately for households with and without a college degree. We find that college-educated nonparticipants are more likely to enter equity markets if they win following positive equity returns (24.4 vs. 11.1 percentage points), as well as if they experienced high equity returns during formative years (36.9 vs. 12.6 percentage points). The differences in treatment effects by recent and early equity return experiences are in fact larger among college-educated (24.4-11.1=13.2 and 36.9-12.6=24.3 percentage points, respectively) than non-college-educated households (12.5-5.0=7.5 and 14.5-8.2=6.4 percentage points), although the statistical power does not allow us to reject the null hypothesis of no difference in effects across groups.[19] These results suggest that the highly educated are not immune to belief-formation biases, and non-standard beliefs are a potential explanation for the discrepancy between model and empirical estimates among the highly educated, cognitively-able households shown in Columns 6 and 7 of Table 8.

6. Conclusion

Widespread nonparticipation in the stock market is a much studied but imperfectly understood phenomenon. This study combines new quasi-experimental estimates of the effects of windfall gains on stock market participation and uses a rich structural life-cycle model to interpret and benchmark the estimates. The model predicts much larger rates of entry than those we observe when the structural parameters are estimated from observational data, and matching our quasi-experimental estimates thus requires entry costs that are implausibly large. Our quasi-experimental estimates therefore pose a challenge to standard modeling approaches.

Motivated by the large discrepancy between our quasi-experimental estimates and model predictions, we conduct a set of analyses to explore the credibility of unmodeled explanations for our results. Several converging lines of evidence suggest that non-standard beliefs and belief-formation processes are a major source of overprediction. We estimate that the discrepancy between predictions and quasi-experimental estimates shrinks by at least 50% when the model is calibrated to match the subjective distribution of beliefs rather than historical equity premia. Additional analyses also provide evidence that both recent equity return and prior-life experiences affect beliefs, consistent with two broad classes of belief-formation processes studied in the literature.

Our results suggest that better aligning equity return beliefs with historical data would likely raise stock market participation (especially if targeted at groups in which belief-pessimism is more prevalent), but provides limited insight into the feasibility of actually designing financial literacy programs or other interventions that do so. For example, although treatment effect heterogeneity by recent equity returns suggests beliefs are at least somewhat malleable, heterogeneity by early equity

[19] Comparing differences across groups is futher complicated because the distribution of participation incentives and beliefs is not independent of education status.

returns suggests a certain degree of belief rigidity. Additionally, although education (or something correlated with education) reduces the prevalence of non-standard beliefs, some college-educated households appear susceptible to belief biases. Our results thus send mixed-signals on the potential to improve financial outcomes by improving beliefs.

Finally, our paper echoes recent discussions by Kahn and Whited (2017), Lewbel (forthcoming), and Nakamura and Steinsson (2017) by demonstrating the value of both causal and structural estimates. Without our causal estimates, the structurally estimated entry costs are significantly understated. And without the structural model, it is impossible to quantify the size and structure of disincentives needed to account for our sample's behavior. Our research design thus demonstrates the methodological benefits of combining causal estimates and identification via economic theory in economic research.

References

Alan, S., 2006. Entry costs and stock market participation over the life cycle. Review of Economic Dynamics 9, 588–611.

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

Ang, A., Bekaert, G., Liu, J., 2005. Why stocks may disappoint. Journal of Financial Economics 76, 471–508.

Angeletos, G.-M., Laibson, D., Repetto, A., Tobacman, J., Weinberg, S., 2001. The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. The Journal of Economic Perspectives 15, 47–68.

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

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

Barillas, F., Fernández-Villaverde, J., 2007. A generalization of the endogenous grid method. Journal of Economic Dynamics and Control 31, 2698–2712.

Bazdresch, S., Kahn, R. J., Whited, T. M., 2017. Estimating and testing dynamic corporate finance models. The Review of Financial Studies 31, 322–361.

Becker, T. A., Shabani, R., 2010. Outstanding debt and the household portfolio. Review of Financial Studies 23, 2900–2934.

Benjamin, D. J., 2019. Errors in probabilistic reasoning and judgment biases. Handbook of Behavioral Economics-Foundations and Applications 2 p. 69.

Benzoni, L., Collin-Dufresne, P., Goldstein, R. S., 2007. Portfolio choice over the life-cycle when the stock and labor markets are cointegrated. The Journal of Finance 62, 2123–2167.

Berkelaar, A. B., Kouwenberg, R., Post, T., 2004. Optimal portfolio choice under loss aversion. The Review of Economics and Statistics 86, 973–987.

Bordalo, P., Gennaioli, N., La Porta, R., Shleifer, A., 2019. Diagnostic expectations and stock returns. Journal of Finance .

Brunnermeier, M. K., Nagel, S., 2008. Do wealth fluctuations generate time-varying risk aversion? micro-evidence on individuals' asset allocation. American Economic Review 98, 713–736.

Calvet, L., Campbell, J. Y., Sodini, P., 2009. Fight or flight: Portfolio rebalancing by individual investors. The Quarterly Journal of Economics 124, 301–348.

Calvet, L., Célérier, C., Sodini, P., Vallée, B., 2017. Can financial innovation solve household reluctance to take risk? Working Paper .

Calvet, L. E., Campbell, J. Y., Sodini, P., 2007. Down or out: Assessing the welfare costs of household investment mistakes. Journal of Political Economy 115, 707–747.

Calvet, L. E., Sodini, P., 2014. Twin picks: Disentangling the determinants of risk-taking in household portfolios. Journal of Finance 69, 867–906.

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

Carroll, C. D., 1997. Buffer-stock saving and the life cycle/permanent income hypothesis. The Quarterly journal of economics 112, 1–55.

Carroll, C. D., 2006. The method of endogenous gridpoints for solving dynamic stochastic optimization problems. Economics letters 91, 312–320.

Carroll, C. D., Samwick, A. A., 1997. The nature of precautionary wealth. Journal of Monetary Economics 40, 41–71.

Cesarini, D., Lindqvist, E., Östling, R., Wallace, B., 2016. Wealth, health, and child development: Evidence from administrative data on swedish lottery players. The Quarterly Journal of Economics 131, 687–738.

Cocco, J. F., 2005. Portfolio choice in the presence of housing. Review of Financial Studies 18, 535–567.

Cocco, J. F., Gomes, F. J., Maenhout, P. J., 2005. Consumption and portfolio choice over the life cycle. Review of Financial Studies 18, 491–533.

Cooper, R., Zhu, G., 2016. Household finance over the life-cycle: What does education contribute? Review of Economic Dynamics 20, 63–89.

Das, S., Kuhnen, C. M., Nagel, S., 2019. Socioeconomic status and macroeconomic expectations. The Review of Financial Studies .

Davis, S. J., Kubler, F., Willen, P., 2006. Borrowing costs and the demand for equity over the life cycle. Review of Economics and Statistics 88, 348–362.

Della Vigna, S., 2009. Psychology and economics: Evidence from the field. Journal of Economic literature 47, 315–72.

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

Dominitz, J., Manski, C. F., 2011. Measuring and interpreting expectations of equity returns. Journal of Applied Econometrics 26, 352–370.

Epstein, L. G., Zin, S., 1991. Substitution, risk aversion and the temporal behavior of consumption and asset returns. Journal of Political Economy .

Erickson, T., Whited, T. M., 2002. Two-step gmm estimation of the errors-in-variables model using high-order moments. Econometric Theory 18, 776–799.

Fagereng, A., Gottlieb, C., Guiso, L., 2017. Asset market participation and portfolio choice over the life-cycle. The Journal of Finance 72, 705–750.

Flavin, M., Yamashita, T., 2011. Owner-occupied housing: life-cycle implications for the household portfolio. American Economic Review 101, 609–614.

Gennaioli, N., Ma, Y., Shleifer, A., 2016. Expectations and investment. NBER Macroeconomics Annual 30, 379–431.

Gomes, F., Michaelides, A., 2005. Optimal life-cycle asset allocation: Understanding the empirical evidence. Journal of Finance 60, 869–904.

Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of applied econometrics 8, S85–S118.

Gourinchas, P.-O., Parker, J. A., 2002. Consumption over the life cycle. Econometrica 70, 47–89.

Greenwood, R., Shleifer, A., 2014. Expectations of returns and expected returns. Review of Financial Studies 27, 714–746.

Guiso, L., Haliassos, M., Jappelli, T., 2002. Household portfolios. MIT press.

Guiso, L., Sapienza, P., Zingales, L., 2004. The role of social capital in financial development. American Economic Review 94, 526–556.

Guiso, L., Sapienza, P., Zingales, L., 2008. Trusting the stock market. Journal of Finance 63, 2557–2600.

Guiso, L., Sodini, P., 2013. Household finance: An emerging field. Handbook of the Economics of Finance Part B, 1397–1532.

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

Hankins, S., Hoestra, M., Skiba, P. M., 2011. The ticket to easy street? the financial consequences of winning the lottery. Review of Economics and Statistics 93, 961–969.

Heaton, J., Lucas, D., 2000. Portfolio choice and asset prices: The importance of entrepreneurial risk. Journal of Finance 55, 1163–1198.

Hurd, M. D., 2009. Subjective probabilities in household surveys. Annu. Rev. Econ. 1, 543–562.

Imbens, G. W., Rubin, D. B., Sacerdote, B. I., 2001. Estimating the effect of unearned income on labor supply, earnings, savings and consumption: Evidence from a survey of lottery players. The American Economic Review 91, 778–794.

Kahn, R., Whited, T. M., 2017. Identification is not causality, and vice versa. Review of Corporate Finance Studies 7, 1–21.

Kaustia, M., Knüpfer, S., 2008. Do investors overweight personal experience? evidence from IPO subscriptions. The Journal of Finance 63, 2679–2702.

Kézdi, G., Willis, R. J., 2011. Household stock market beliefs and learning. Tech. rep., National Bureau of Economic Research.

Khorunzhina, N., 2013. Structural estimation of stock market participation costs. Journal of Economic Dynamics and Control 37, 2928–2942.

Kuhnen, C. M., 2015. Asymmetric learning from financial information. The Journal of Finance 70, 2029–2062.

Kuhnen, C. M., Miu, A. C., 2017. Socioeconomic status and learning from financial information. Journal of Financial Economics 124, 349–372.

Laibson, D., 1997. Golden eggs and hyperbolic discounting. The Quarterly Journal of Economics pp. 443–477.

Laun, T., Wallenius, J., 2015. A life cycle model of health and retirement: The case of swedish pension reform. Journal of Public Economics 127, 127–136.

Lewbel, A., forthcoming. The identification zoo: Meanings of identification in econometrics. Journal of Economic Literature 7.

Love, D., Phelan, G., 2015. Hyperbolic discounting and life-cycle portfolio choice. Journal of Pension Economics & Finance 14, 492–524.

Malmendier, U., Nagel, S., 2011. Depression babies: Do macroeconomic experiences affect risk taking? The Quarterly Journal of Economics 126, 373–416.

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

Mulligan, C. B., Sala-i Martin, X., 2000. Extensive margins and the demand for money at low interest rates. Journal of Political Economy 108, 961–991.

Nakamura, E., Steinsson, J., 2017. Identification in macroeconomics. Tech. rep., National Bureau of Economic Research.

Nekoei, A., Seim, D., 2018. How do inheritances shape wealth inequality? theory and evidence from sweden. Working Paper .

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

Samuelson, P. A., 1969. Lifetime portfolio selection by dynamic stochastic programming. Review of Economics and Statistics 51, 239–246.

Samuelson, W., Zeckhauser, R., 1988. Status quo bias in decision making. Journal of risk and uncertainty 1, 7–59.

Smith, A. A., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. Journal of Applied Econometrics 8.

Tversky, A., Kahneman, D., 1986. Rational choice and the framing of decisions. Journal of business pp. S251–S278.

Van Rooij, M. C., Lusardi, A., Alessie, R. J., 2012. Financial literacy, retirement planning and household wealth. The Economic Journal 122, 449–478.

Vestman, R., 2018. Limited stock market prticipation among renters and home owners. The Review of Financial Studies 32, 1494–1535.

Vissing-Jorgensen, A., 2002. Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures. Tech. rep., National Bureau of Economic Research.

Vissing-Jørgensen, A., 2003. Perspectives on behavioral finance: Does irrationality disappear with wealth? evidence from expectations and actions. NBER Macroeconomics Annual 18, 139–208.

Waldenström, D., 2014. Swedish stock and bond returns, 1856–2012. IFN Working Paper No. 1027 .

A. Structural Model Details

A.1 Household Decision Problem

The full household decision problem described in Section 4 is written as:

$$V_t(X_t, P_t, I_t, L_t, e,m) = \max_{C_t, Part_t, \alpha_t} \Bigg\{ (1-\beta\pi_t) C_t^{1-1/\psi} +$$

$$\beta \mathbb{E} \left[\pi_t V_{t+1}(X_{t+1}, P_{t+1}, I_{t+1}, e, m)^{1-\rho} + (1-\pi_t) b(X_{t+1})^{1-1/\psi} \right]^{\frac{1-1/\psi}{1-\rho}} \Bigg\}^{\frac{1}{1-1/\psi}}$$

$$X_{t+1} = \left[R_f + \alpha_t (R^s_{t+1} - R_f) \right] (X_t - C_t) + Y_{t+1} - \left[(1-I_t) \times \chi + \kappa \right] \times Part_t + L_t$$

$$0 \leq \alpha \leq 1$$

$$Y_t = \begin{cases} exp(f(t, Z_t)) P_t U_t \;\; if \;\; t \leq t_R \\ \lambda_{e,m} Y_{t_R} \;\; if \;\; t > t_R \end{cases}$$

$$P_t = P_{t-1} N_t$$

$$I_{t+1} = (1 - I_t) \times Part_t$$

$$\begin{pmatrix} r^s_t - r_f \\ log(N_t) \\ log(U_t) \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} \mu_s \\ -\sigma^2_N/2 \\ -\sigma^2_U/2 \end{pmatrix}, \begin{pmatrix} \sigma^2_s & \rho_{n,r} \times \sigma_n \sigma_s & 0 \\ \rho_{n,r} \times \sigma_n \sigma_s & \sigma^2_n & 0 \\ 0 & 0 & \sigma^2_u \end{pmatrix} \right]$$

A.2 Estimation and Test Statistics

We consider two test statistics to check for overidentifying restrictions and to evaluate model fit. First the standard overidentifying test used to test the model's fit of the empirical moments, correcting for simulation error, is given by

$$\frac{NK}{1+K} g(\nu_{i,s}, \theta)' \hat{\Omega}^{-1} g(\nu_{i,s}, \theta) \rightarrow \chi^2_{|g(\nu_{i,s},\theta)| - |\theta|}. \tag{15}$$

Second, we consider the Wald test for external validity presented in Bazdresch et al. (2017) that considers the model's fit of non-targeted moments $m^{\star}$. The null hypothesis of non-targeted fit,

$$g^{\star}(\nu_{i,s},\theta) = \mathbb{E}\left[m^{\star}(\nu_i) - \frac{1}{K}\sum_{k=1}^{K} m^{\star}(\nu_{i,s}^{k}(\theta))\right] = 0, \tag{16}$$

can be tested by a Wald statistic defined as

$$g^{\star}(\nu_{i,s},\hat{\theta})' avar(g^{\star}(\nu_{i,s},\hat{\theta}))^{-1} g^{\star}(\nu_{i,s},\hat{\theta}) \to \chi^2_{|g^{\star}(\nu_{i,s},\hat{\theta})|} \tag{17}$$

$$avar(g^{\star}(\nu_{i,s},\hat{\theta})) = \mathbb{E}\left[\phi_g^{\star}\phi_g^{\star}\right] \tag{18}$$

where $\phi_g^{\star}$ denotes the influence function for $g^{\star}$.

A.3 Model Solution

To solve the model, we follow Carroll (1997) and normalize the value function, state variables, and controls by the permanent component of income P_t to eliminate P_t as a state variable. We use lower case letters to denote the normalized variables (e.g., $v_t = V_t/P_t$, $x_t = X_t/P_t$). After these transformations, the model is solved by backwards induction. We assume that the last period's utility is as $v_T = b(x_T)^{1-\psi}$. We then use this to solve for the optimal saving policy $x_{T-1} - c_{T-1}$ using the endogenous grid method and portfolio allocation α_{t-1} using a grid search (100 grid points) (see, e.g., Carroll (2006) and Barillas and Fernández-Villaverde (2007)). For points that do not fall on next period's stored state-grid, we use cubic interpolation to evaluate the value function. To calculate the expected value of next period's value function, we follow the procedure described in Gomes and Michaelides (2005) to create a state transition matrix that makes integration less computationally costly. After having obtained the optimal saving and portfolio allocation policies, we are able to calculate the v_{T-1} value function. We then repeat this process and iterate backwards until reaching age t_0. We repeat this for all combinations of marital status and education level, and store the resulting policy functions.

A.4 Survival Probability Estimation

The survival probability (π_t) is calculated using the observed survival probabilities from years 1999-2000. We select 100,000 individuals in year 1998 from the Swedish population, and define a binary indicator equal to one if the individual is observed alive in 1999. We then regress a quartic in age on this indicator. We do not permit time or cohort effects in our estimation, and do not

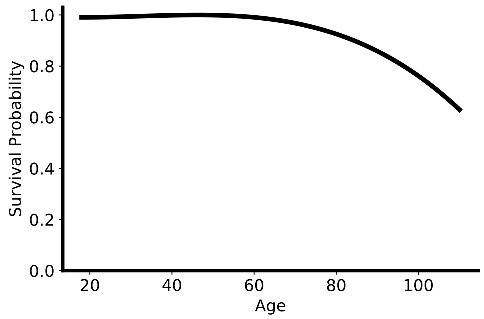


Figure A.1: Survival Probabilities. This figure presents the one year survival probability for each age. Survival probabilities are calculated as the average observed 1998-1999 survival probabilities for a random sample of the Swedish population.

allow survival probabilities to vary with wealth, income, or sex. There is no attrition or selection concerns in this sample as it is drawn randomly from the entire population. The resulting estimates are presented in Figure A.1.

A.5 Income Estimation

Our estimation of income profiles follows the procedure described in Cocco et al. (2005). Our definition of income is total income after taxes and transfers. As noted in Cocco et al. (2005), because there are (potentially endogenous) insurance mechanisms – including government transfers, family transfers, and spousal labor supply decisions – that provide a lower bound on income (perhaps especially in countries with strong social safety nets such as Sweden) this definition captures this insurance without explicit modeling of all income smoothing mechanisms. Our estimation sample is the sample of lottery winners in the thirty years (or as many as possible) prior to the lottery event.

Income processes are estimated separately for each of the education groups we consider. The estimation sample is the sample of lottery winners prior to the lottery. We regress the log of income on dummies of age and marital status. We then regress a third-order polynomial in age on the age dummies and marital status for housholds between ages 18-65 to recover an average income profile $f(t, m, e)$. The resulting average income profile estimates $exp(f(t, m, e))$ are shown in Figure A.2, with dotted lines representing married households and dashed lines representing single households. $P_{i,s}$ is then constructed as the ratio of observed to average income for each household in our sample.

We estimate income variance parameters again following Cocco et al. (2005), who closely

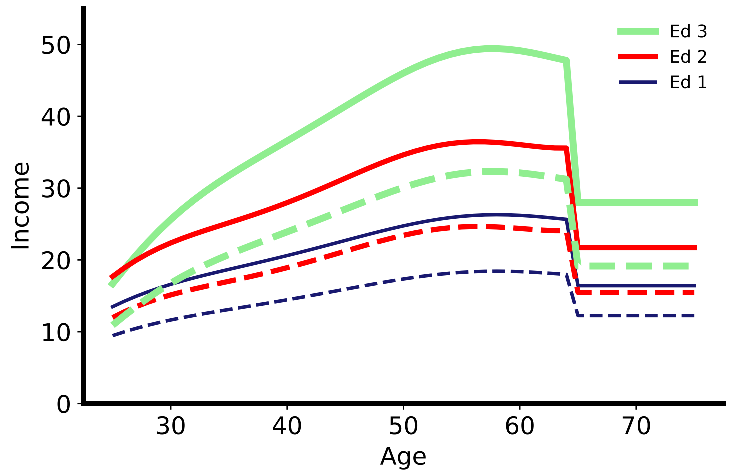


Figure A.2: Average Income Profiles. This figure presents the deterministic income component $f(t, m, e)$. Solid lines reflect married households while dashed lines reflect single households. Income profiles are estimated following the methodology in Cocco et al. (2005), which is summarized in Appendix A.5. Income in retirement is defined as the age 65 income times a replacement rate that depends on education and marital status. See Appendix A.6 for details on replacement rate calculations.

follow the procedure proposed by Carroll and Samwick (1997). In particular, defining

$$\epsilon_{i,t}^{Y} \equiv log(Y_{i,t}) - \hat{f}(t, m_i, e_i)$$
$$r_{i,d} \equiv \epsilon_{i,t+d}^{Y} - \epsilon_{i,t}^{Y},$$

then because

$$\text{Var}(r_{i,d}) = d\sigma_N^2 + 2 * \sigma_U$$

we can recover $\sigma_{N,e}$ and $\sigma_{U,e}$ via OLS regresison on $\text{Var}(r_{i,d})$ on d for each separate education group.

To estimate the correlation between income and equity returns, note that $\epsilon_{i,t}^{Y}$ can be written as

$$r_{i,1} = log(i, N_t) + log(U_{i,t}) - log(U_{i,t+1})$$

and taking the average yields

$$r_{i,1} = log(N_{i,t}) + log(U_{i,t}) - log(U_{i,t+1}).$$

Decomposing $N_{i,t}$ into aggregate and idiosynchratic components, letting s index year, and averaging (for each education group) yields:

$$\bar{r}_{i,1,s,e} = log(N_{s,e}^{Agg}).$$

The correlation between equity returns and $log(N_{i,t})$ for each education group is then recovered by the coefficient from an OLS regression of $\bar{r}_{i,1,s}$ on excess returns, where excess returns are defined as the difference between Stockholm Stock Exchange and short-term Swedish Treasury returns (Waldenström (2014)).

A.6 Retirement Income Replacement Rates

Retirement income replacement rates are approximated using the formulas described in Section 3 of Laun and Wallenius (2015), which conducts a detailed analysis of the Swedish pension system. Our formulas are slightly simplified due to the assumption that labor supply is exogenous. The pension has two parts. First, all households receive 96% of a basic amount (BA) of 43,600 SEK (6,500 USD). Second, an earning supplement is given by

$$.6 \times AP \times BA$$

where AP denotes pension points calculated from the fifteen years with highest observed income calculated recursively by the following formula:

$$AP_{t+1} = AP_t + \frac{1}{15} \max \left(0, \frac{\min(Y_t, 7.5BA) - BA}{BA} - AP_t \right).$$

Thus, retirement income is approximated as the ratio of the following formula

$$.6 \times AP \times BA + .96BA$$

to age 65 income.

To conserve state variables, we do not carry pension points as a state variable as in Laun and

Wallenius (2015). Instead, we simulate 20,000 income processes for each education and marital status, and calculate the average replacement rate for each group.

A.7 Model Benchmarks and Fit

Below we present the full specification of the regressions that form our EPF benchmarks. In addition, we indicate the corresponding panel for each regression in Table A.1, and, when appropriate, the location of selected coefficients presented in Table 6. Empirical estimates are presented in A.1, Column 1. Note in all lottery regressions we include cell-fixed effects that ensure all identifying variation comes from players in the same cell. The regressions we consider are:

1. Pre-lottery regressions (Table A.1, Panel A.i-ii):

$$\begin{aligned} c_{i,s} &= b + b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_I I_{i,s} + \eta_{i,s}^C \\ Part_{i,s} &= b + b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_I I_{i,s} + \eta_{i,s}^{Part} \end{aligned} \tag{19}$$

2. Post-lottery regressions (Table A.1, Panel B.i-ii; Table 6, Panel B.i):

$$\begin{aligned} c_{i,s} &= b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_I I_{i,s} + b_l l_{i,s} + MX_{i,0} + \eta_{i,s}^C \\ Part_{i,s} &= b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_I I_{i,s} + b_l l_{i,s} + MX_{i,0} + \eta_{i,s}^{Part} \end{aligned} \tag{20}$$

3. Post-lottery regressions by participation status (Table A.1; Panel B.iii-iv, Table 6, Panel B.ii)). These regressions are estimated separately in subsamples restricted to participants $I_{i,s} = 1$ and nonparticipants ($I_{i,s} = 0$):

$$\begin{aligned} c_{i,s} &= b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_l l_{i,s} + MX_{i,0} + \eta_{i,s}^C \\ Part_{i,s} &= b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_l l_{i,s} + MX_{i,0} + \eta_{i,s}^{Part} \end{aligned} \tag{21}$$

4. Post-lottery regressions, nonlinear (Table A.1; Panel B.v, Table 6, Panel B.iii):

$$\begin{aligned} Part_{i,s} = b_t t_{i,s} &+ b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + \mathbb{1}_{l_{i,s} \in [1.5,15)} + \\ &\mathbb{1}_{l_{i,s} \in [15,150)} + \mathbb{1}_{l_{i,s} \in [150,300)} + \mathbb{1}_{l_{i,s} \in [300,\infty)} + MX_{i,0} + \eta_{i,s}^{Part} \end{aligned} \tag{22}$$

Table A.1 presents the fits of are various estimation exercises. Our pre-lottery estimation (Table

Table A.1: Structural Estimation Model Fit. This table presents the model fit for our various structural estimation exercises. Column 1 presents the empirical policy function estimates for pre-lottery observations in Panel A and post-lottery observations in Panel B. Column 2 presents matched EPF coefficients when the model is estimated using only pre-lottery observations, Column 3 the matched EPF coefficients when the model is estimated using only post-lottery observations, Column 4 the matched EPF coefficients when the model is estimated using only pre- and post-lottery observations, Column 5 the matched EPF coefficients from estimating the entry cost distribution (Figure 5) when other parameters are fixed at their values in Column 1, and Column 6 the matched EPF coefficients when the model is augmented to allow for naive present-biased preferences and estimated using only pre-lottery observations. All estimations use the post-1999 sample of lottery winners.

	Estimate (1)	Pre-Lottery (2)	Post-Lottery (3)	Pre- & Post- (4)	Nonlinear (5)	Present-bias (6)
A. Pre-Lottery Benchmarks						
i. Consumption						
Age	.619	.077		.185		.164
Age^2	-.006	-.001		-.002		-.002
$Wealth/PI$	.164	.203		.156		.173
$(Wealth/PI)^2$	.000	.000		.000		.000
$Part_{-1}$	-.881	1.388		1.435		1.797
$Constant$	4.508	7.860		6.993		7.644
ii. Participation						
Age	.000	.000		.000		.003
Age^2	.000	.000		.000		.000
$Wealth/PI$	.000	.000		.000		.000
$(Wealth/PI)^2$	.000	.000		.000		.000
$Part_{-1}$	.883	.927		.938		.935
$Constant$	.114	.030		.019		-.013
B. Lottery Benchmarks						
i. Consumption						
Age	.614		.250	.223		
Age^2	-.005		-.003	-.003		
$Wealth/PI$	.039		.139	.155		
$(Wealth/PI)^2$	.000		.000	.000		
$Part_{-1}$	-1.618		1.487	1.253		
$Lottery$	.185		.123	.138		
ii. Participation						
Age	.001		.000	.001		
Age^2	.000		.000	.000		
$Wealth/PI$	.000		.000	.000		
$(Wealth/PI)^2$	.000		.000	.000		
$Part_{-1}$	.796		.993	.933		
$Lottery$	.028		.030	.067	.029	
iii. Effect on Consumption by Prior Participation Status						
$Lottery$, Nonparticipants	.239		.121	.136		
$Lottery$, Participants	.166		.124	.138		
iv. Effect on Participation by Prior Participation Status						
$Lottery/1M\ SEK$, Nonparticipants	.104		.137	.292	.104	
$Lottery/1M\ SEK$, Participants	.002		.000	.000	.000	
v. Effect on Participation by Prize Size (USD), Nonparticipants						
$1.5K < L_i \leq 15K$	-.012				.006	
$15K < L_i \leq 150K$	.078				.080	
$150K < L_i \leq 300K$	.156				.158	
$300K < L_i$	.359				.357	
$N =$		192,524	70,139	262,663	70,139	192,524

6, Column 1), which targets only pre-lottery regressions (Table A.1, Panel A.i and Panel A.ii) is presented in Column 2. Our post-lottery estimation (Table 6, Column 2), which targets only post-lottery regressions (Table A.1, Panel B.i-iv) is presented in Column 3. Our pre-/post-lottery combined estimation (Table 6, Column 3), which targets only both pre- and post-lottery regressions (Table A.1, Panels A.i-ii, B.i-iv) is presented in Column 4. Our entry-cost heterogeneity estimation (Table 6, Column 4), which targets selected post-lottery regression coefficients of the effect of lottery prizes on participation (Table A.1, Panel B.ii, iv-v) is presented in Column 5. Finally, our estimation with present-biased preferences (Table 8, Column 2), which targets only pre-lottery regressions (Table A.1, Panel A.i and Panel A.ii)) is presented in Column 6.

A.8 Life-cycle Profiles Comparison

In this section we compare the life-cycle profiles implied by our model estimates to their empirical counterparts. To estimate empirical life-cycle profiles of stock market participation and wealth, we use a simplified version of the estimation procedure described in Fagereng et al. (2017). Our estimation sample in this exercise consists of the matched population sample presented in Table 4, Column 2.

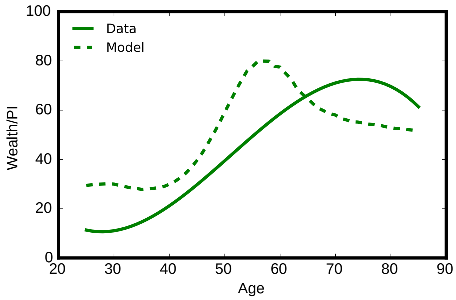
To estimate life-cycle profiles of the wealth/income ratio, we run an OLS regression of the registry defined wealth/income ratio on age indicators, year indicators, and a proxy of cohort effects defined by the average returns on the Stockholm Stock Exchange experienced between ages 18-25. We then regress the predicted wealth-to income ratios for each age on a cubic polynomial of age. The resulting wealth-to-income profiles are presented as the dotted line in Panel (a) of the below figures.

To estimate life-cycle profiles of stock market participation, we run a probit regression of household stock market participation on age indicators, year indicators, and a proxy of cohort effects defined by the average returns on the Stockholm Stock Exchange experienced between ages 18-25. We then regress the predicted stock market participation probabilities for each age on a cubic polynomial of age. The resulting participation probabilities profiles are presented as the dotted line in Panel (b) of the below figures. Overall, our estimated wealth and participation profiles are similar to those obtained by Fagereng et al. (2017) for a representative Norwegian sample.

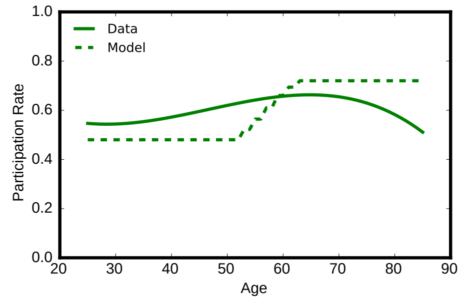
To generate model implied profiles, we draw a random sample of 10000 Swedish households aged 18-25 between 1999-2004. Because marital and education histories are incomplete by this age, we assign marital and education status as the highest values observed by 2009. We then record all model state variables, and simulate saving and participation decisions through age 85.

The average of these simulations for each age are presented as the dotted lines in the below figures.

Figure A.3 presents results from our model using parameter estimates from our estimation with pre-lottery data (Table 6, Column 1). Figure A.4 presents results from our model using parameter estimates from our estimation with post-lottery data (Table 6, Column 2). Figure A.5 presents results from our model using parameter estimates from our estimation with pre- and post-lottery data (Table 6, Column 3). Figure A.6 presents results from our model with quasi-hyperbolic discounting using parameter estimates from our estimation with pre- and post-lottery data (Table 8, Column 2).

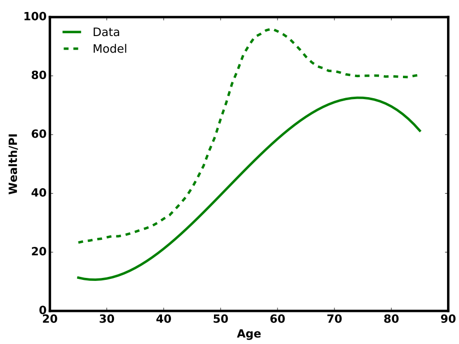


(a) Wealth/PI ratio

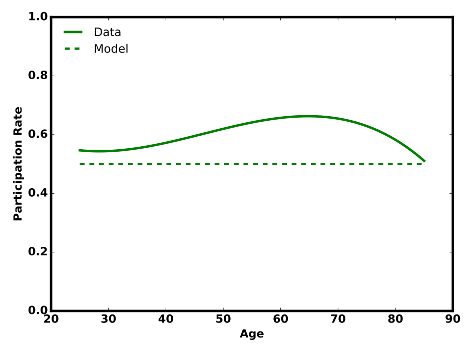


(b) Participation

Figure A.3: Life-cycle profiles - pre-lottery data. This figure compares the model-predicted and empirical wealth/permanent income ratio and participation rate over the life-cycle. The model is simulated using estimates obtained from pre-lottery data (Table 6, Column 1).

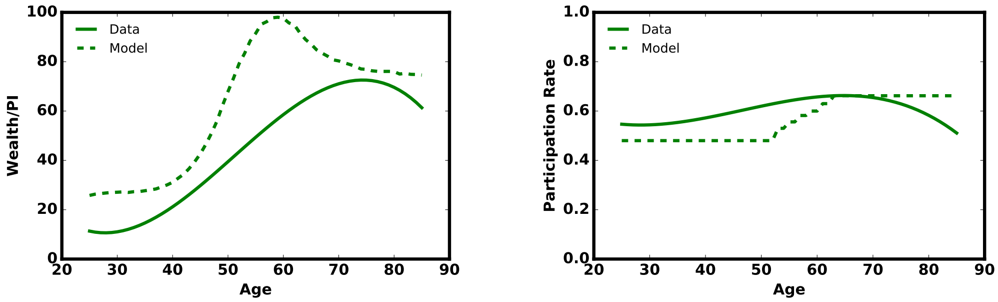


(a) Wealth/PI ratio



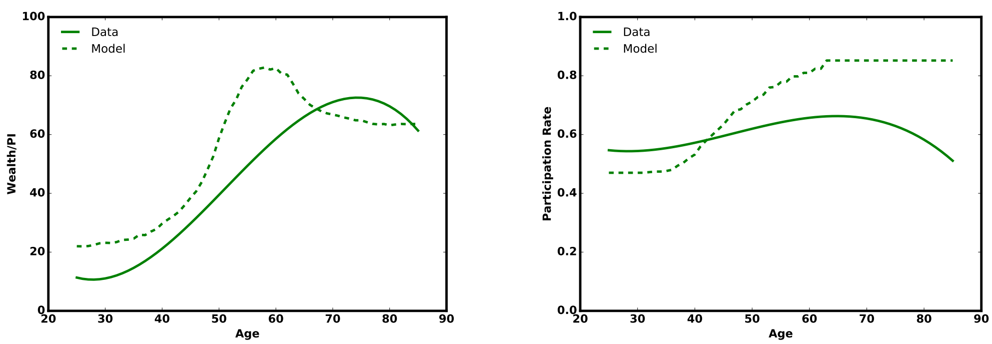
(b) Participation

Figure A.4: Life-cycle profiles - lottery data. This figure compares the model-predicted and empirical wealth/permanent income ratio and participation rate over the life-cycle. The model is simulated using estimates obtained from lottery data (Table 6, Column 2).



(a) Wealth/PI ratio **(b) Participation**

Figure A.5: Life-cycle profiles - pre- and post-lottery. This figure compares the model-predicted and empirical wealth/permanent income ratio and participation rate over the life-cycle. The model is simulated using estimates obtained from pre- and post-lottery data (Table 6, Column 3).



(a) Wealth/PI ratio **(b) Participation**

Figure A.6: Life-cycle Profiles - Pre-lottery with quasi-hyperbolic discounting. This figure compares the model-predicted and empirical wealth/permanent income ratio and participation rate over the life-cycle for our model that allows for $\beta - \delta$ preferences. Parameters are obtained by setting $\beta = .6$ and re-estimating the model with pre-lottery data (Table 8, Column 4).

B. Online Appendix - Supplemental Tables

Table B.1: Summary Statistics and Prize Distributions for Triss Lump-sum and Monthly Prize Winners. The summary statistics shown are all means and measured at $s = -1$. All variables except female, age, and Nordic born are measured at the household level. Households are classified as equity market participants if the own equity either directly or indirectly via mutual funds. Financial variables are winsorized at the .5 and 99.5 percentiles.

Summary Statistics

	A. All-Year		**B. Post-1999**	
Prize Amount	**Triss Lump-sum**	**Triss Monthly**	**Triss Lump-sum**	**Triss Monthly**
Female	0.54	0.49	0.56	0.48
Age (years)	52.52	49.78	51.86	50.50
Nordic Born	0.95	0.94	0.94	0.94
Household Members (#)	0.52	0.54	0.59	0.56
Household Income (K SEK)	335	374	382	392
Married	0.52	0.54	0.54	0.55
Higher Education	0.24	0.26	0.26	0.28
Financial				
Net Wealth (K SEK)			857	736
Gross Debt (K SEK)			448	386
Home Owner			0.73	0.66
Equity Owner			0.63	0.59
N	3,399	476	1,776	386

Prize Distribution

	A. All-Year		**B. Post-1999**	
Prize Amount	**Triss Lump-sum**	**Triss Monthly**	**Triss Lump-sum**	**Triss Monthly**
$L_i = 0$	0	0	0	0
$L_i \leq 10K$	0	0	0	0
$10K < L_i \leq 100K$	985	0	366	0
$100K < L_i \leq 500K$	2074	0	1237	0
$500K < L_i \leq 1M$	157	0	89	0
$1M < L_i \leq 2M$	49	130	22	110
$2M < L_i$	134	346	62	276
Total	3,399	476	1,776	386

Table B.2: Demographic and Financial Predictors of Participation in Post-1999 Sample and Sex- and Age-Weighted Swedish Representative Sample. The regression model is estimated using year-end net wealth in 1999 and is comparable to that used by Calvet et al. (2007). Marginal effects are calculated as the predicted effect of a one-standard deviation change on the probability of participation, holding fixed the value of all other variables at their median value.

	Post-1999 Lottery			**Matched Population**		
	Estimate	**SE**	**Marginal Effects.**	**Estimate**	**SE**	**Marginal Effects**
	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.064	0.006	5.43%	0.057	0.008	4.87%
Financial Assets	0.190	0.005	68.68%	0.177	0.007	76.58%
Total Real Estate	0.028	0.006	14.09%	0.036	0.011	17.64%
Total Liabilities	-0.014	0.007	-6.91%	-0.023	0.011	-11.14%
Retired	0.010	0.029	0.17%	0.056	0.045	0.76%
Self-Employed	0.046	0.043	0.35%	0.039	0.059	0.30%
Unemployed	0.020	0.028	0.24%	0.068	0.042	0.88%
Student	0.135	0.047	1.09%	0.068	0.055	0.87%
Age	-0.013	0.001	-7.74%	-0.013	0.001	-8.65%
Household Size	-0.044	0.012	-1.34%	-0.050	0.015	-1.88%
High School	0.145	0.024	2.77%	0.176	0.039	3.13%
Higher Degree	0.224	0.026	4.28%	0.294	0.042	5.34%
Missing Education	0.111	0.107	0.44%	0.421	0.127	1.52%
Immigrant	-0.166	0.052	-1.11%	-0.262	0.059	-2.71%
Constant	-3.884	0.123	.	-3.594	0.172	.
N	70,166			70,166		

Table B.3: Effect of Wealth (1M SEK) on Participation Probability. This table presents coefficients, standard errors, sample size, and mean predicted participation probability when lottery wealth is zero ($\hat{y}|_{L_i=0}$) obtained from estimating Equation 1 in the all-year sample. Columns 5 through 8 show analogous estimates with participation defined more narrowly to only include directly owned stocks.

Horizon	**A. Stock or Mutual Fund**				**B. Stock Only**			
(s)	β_s	**SE**	N	$\hat{y}_{\vert L_i=0}$	β_s	**SE**	N	$\hat{y}_{\vert L_i=0}$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
-1	0.005	0.012	70,166	0.725	0.002	0.012	70,166	0.425
0	0.039	0.010	75,773	0.727	0.021	0.011	75,773	0.433
1	0.047	0.010	91,940	0.729	0.036	0.011	91,940	0.434
2	0.045	0.010	113,879	0.749	0.041	0.011	113,879	0.462
3	0.043	0.010	141,878	0.761	0.043	0.011	141,878	0.478
4	0.042	0.010	149,324	0.770	0.041	0.012	149,324	0.494
5	0.046	0.010	153,464	0.773	0.033	0.012	153,464	0.502
6	0.038	0.010	168,061	0.778	0.039	0.012	168,061	0.512
7	0.035	0.010	182,380	0.788	0.044	0.012	182,380	0.523
8	0.053	0.010	197,045	0.794	0.046	0.012	197,045	0.533
9	0.049	0.010	216,787	0.794	0.043	0.012	216,787	0.537
10	0.041	0.010	214,208	0.797	0.023	0.012	214,208	0.545

Table B.4: Effect of Wealth (1M SEK) on Participation Probability by $s = -1$ Participation Status. This table presents coefficients, standard errors, sample size, and mean predicted participation probability when lottery wealth is zero ($\hat{y}_{|L_i=0}$) obtained from estimating Equation 1 in the post-1999 sample stratified by participation status.

Horizon	**Participants**				**Nonparticipants**			
(s)	β_s	**SE**	N	$\hat{y}_{\|L_i=0}$	β_s	**SE**	N	$\hat{y}_{\|L_i=0}$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
0	0.000	0.005	50,861	0.979	0.120	0.024	19,278	0.070
1	0.002	0.009	47,380	0.961	0.116	0.029	17,276	0.093
2	0.010	0.009	43,487	0.947	0.093	0.032	15,316	0.114
3	0.021	0.007	40,324	0.931	0.073	0.032	13,757	0.129
4	0.018	0.010	36,842	0.917	0.080	0.043	12,267	0.141

Table B.5: Effect of Wealth on Equity Market Participation Probability by Prize Size. Coefficients are obtained by estimating Equation 1 in the post-1999 sample with the lottery wealth variable replaced by indicators for five mutually exclusive prize categories: 0 to 10K (omitted category), 10K to 100K, 100K to 1M, 1M to 2M, and 2M+ SEK. Marginal effects are calculated by dividing the effect-size estimate by the mean prize in each category.

A. Participants

	$10K < L_i \leq 100K$	$100K < L_i \leq 1M$	$1M < L_i \leq 2M$	$2M < L_i$
Estimate	-0.011	-0.005	0.001	-0.024
SE	0.009	0.016	0.011	0.033
ME	-0.364	-0.026	0.001	-0.007
N	478	801	203	50

B. Nonparticipants

	$10K < L_i \leq 100K$	$100K < L_i \leq 1M$	$1M < L_i \leq 2M$	$2M < L_i$
Estimate	0.014	0.082	0.177	0.399
SE	0.029	0.037	0.044	0.094
ME	0.382	0.434	0.159	0.127
N	256	525	94	28

Table B.6: Effect of Wealth (1M SEK) on Participation Probability by Form of Payment. This table presents coefficients, standard errors, sample size, and mean predicted participation probability when lottery wealth is zero ($\hat{y}_{|L_i=0}$) obtained from estimating Equation 1 in the post-1999 sample of Triss-monthly winners stratified by participation status. Columns 1 through 8 show estimates for nonparticipants at time $s = -1$ that received annual and lump-sum payments, respectively. Columns 9 through 16 show analogous estimates for participants at time $s = -1$.

	A. Non-participants								B. Participants							
Horizon	**Annual Prizes**				**Lump-Sum Prizes**				**Annual Prizes**				**Lump-Sum Prizes**			
(s)	β_s	**SE**	N	$\hat{y}_{\vert L_i=0}$	β_s	**SE**	N	$\hat{y}_{\vert L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\vert L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\vert L_i=0}$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
0	0.011	0.023	164	0.115	0.105	0.029	710	0.083	0.003	0.012	222	0.920	-0.003	0.007	1066	0.960
1	0.007	0.029	135	0.159	0.079	0.034	591	0.133	0.012	0.015	200	0.854	-0.002	0.013	906	0.917
2	-0.008	0.031	112	0.183	0.050	0.036	482	0.177	0.034	0.014	163	0.790	0.015	0.011	733	0.891
3	-0.036	0.039	92	0.354	0.036	0.037	395	0.167	0.041	0.020	134	0.705	0.030	0.007	600	0.880
4	-0.002	0.050	73	0.267	0.022	0.041	314	0.185	0.030	0.026	107	0.727	0.028	0.008	473	0.871

Table B.7: Effect of Wealth (1M SEK) on Participation Probability, Robustness Checks. This table presents the effect of each 150K USD (1M SEK) wealth on participation at $s = 0$ by lottery in the pooled all-year sample and the post-1999 sample stratified by participants (P) and nonparticipants (NP). Individual Analysis: spousal ownership of equities excluded from participation definition. Capital Insurance: capital insurance ownership is included in participation definition. Structured Product: retail structured products are included in the equity market participation definition. Private Pension: restrict the sample to retired winners and exclude those that ever received nonzero income from private pensions. Unlisted Business Equity: exclude winners whose main source of income comes from their own incorporated business. Kombi: restrict to Kombi participants. Triss: restrict to Triss participants. Probit: marginal effects from probit instead of OLS. Small Prizes Excluded: prizes of size <50K SEK are dropped from analysis. Large Prizes Excluded: prizes greater than 1.5M SEK are dropped from analysis. presented.

	Individual Analysis			**Capital Insurance**			**Structured Product**		
	Pooled	**P**	**NP**	**Pooled**	**P**	**NP**	**Pooled**	**P**	**NP**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Effect	0.043	0.005	0.099	0.052	0.000	0.151	.055	-0.000	0.184
SE	0.011	0.004	0.021	0.010	0.005	0.029	0.010	0.005	0.029
N	75,773	46,703	23,436	75,773	52,339	17,800	75,773	51,425	18,714

	Private Pension Excluded			**Unlisted Business Equity**		
	Pooled	**P**	**NP**	**Pooled**	**P**	**NP**
	(10)	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**
Effect	0.059	-0.007	0.121	0.038	-0.000	0.120
SE	0.020	0.011	0.052	0.010	0.005	0.024
N	26,620	15,441	9,213	75,773	51,062	19,077

	Kombi			**Triss**		
	Pooled	**P**	**NP**	**Pooled**	**P**	**NP**
	(16)	**(17)**	**(18)**	**(19)**	**(20)**	**(21)**
Effect	0.042	0.005	0.151	0.036	-0.003	0.105
SE	0.018	0.007	0.036	0.012	0.007	0.029
N	28,571	19,154	7,261	1,968	1,066	710

	Probit			**Small Prizes Excluded**			**Large Prizes Excluded**		
	Pooled	**P**	**NP**	**Pooled**	**P**	**NP**	**Pooled**	**P**	**NP**
	(22)	**(23)**	**(24)**	**(25)**	**(26)**	**(27)**	**(28)**	**(29)**	**(30)**
Effect	0.133	0.009	0.532	0.036	-.003	.105	0.050	-.001	.170
SE	0.038	0.072	0.097	.013	.007	.031	0.021	.010	.038
Marginal Effect	0.040	0.000	0.067						
N	75,769	46,918	17,149	2,284	1,267	800	75,680	50,805	19,248

Table B.8: Effect of Wealth on Equity Market Participation Probability by Pre-lottery Wealth and Age Quartiles. Coefficients are obtained by estimating Equation 1 in the post-1999 sample stratified by quartiles of pre-lottery net wealth and age.

A. Wealth

	Participants				Nonparticipants			
Quartile:	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Estimate	-.012	-.004	-.001	.006	.154	.097	.114	.040
SE	.022	.010	.004	.003	.044	.035	.049	.014
N	8,222	12,151	14,316	16,172	9,299	5,387	3,223	1,369
$\hat{y}_{\vert L_i=0}$	.955	.974	.984	.991	.058	.073	.088	.089
$Part_{-1}$	.456	.675	.773	.907	.456	.675	.773	.907

B. Age

	Participants				Nonparticipants			
Quartile:	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Estimate	-.005	-.000	.006	-.018	.142	.080	.161	.110
SE	.012	.008	.007	.023	.035	.037	.052	.067
N	11,444	14,247	13,795	11,375	9,299	5,387	3,223	1,369
$\hat{y}_{\vert L_i=0}$	.978	.984	.977	.977	.000	.035	.002	.101
$Part_{-1}$	.595	.659	.779	.622	.595	.659	.779	.622

Table B.9: Heterogeneous Effect of Wealth (1M SEK) on Participation Probability among $s = -1$ Equity Market Participants Coefficients are obtained by estimating Equation 1 at time $s = 0$ in the post-1999 sample of equity market participants at time $s = -1$, stratified by the characteristics indicated in the column heads. Panel A stratifies households by financial characteristics: Columns 1 and 2 show effects for participants that do and do not own homes, Columns 3 and 4 for participants that do and do not have debt, and Columns 5 and 6 for participants that did and did not have self-employment income the year prior to the lottery. Panel B stratifies households by information proxies: Columns 7 and 8 show effects for participants that do and do not have college degrees while Columns 9 and 10 for participants that have above and below median cognitive skill among the sample of for whom conscription records are available. Hetero p obtained from an F-test of the null hypothesis that the two lottery-wealth coefficients are identical. $\% \ Part_{-1}$ indicates the share of the post-1999 sample with the characteristic indicated by the column head that owned equity the year prior to the lottery. See Table 7 for results for time $s = -1$ equity market nonparticipants.

	A. Financial Characteristics						**B. Information Proxies**			
	Home Owner		**Have Debt**		**Self-Employed**		**College Degree**		**Cognitive Skill**	
	No	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**Low**	**High**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Effect	.001	-.001	.000	-.001	.001	-.012	.002	-.005	-.030	.032
SE	.014	.005	.006	.006	.005	.017	.005	-.013	.047	.059
p	.954	.873	.939	.869	.888	.466	.649	.662	.517	.585
Hetero p	.912		.863		.460		.566		.980	
N	9,880	40,881	20,206	30,655	47,631	3,230	36,065	14,796	1,689	3,598
$\% \ Part_{-1}$	.554	.784	.679	.759	.719	.832	.686	.842	.677	.790

Table B.10: Effect of Wealth on Real Estate and Debt by $s = -1$ Equity Market Participation Status. This table presents coefficients, standard errors, sample size, and mean predicted change in wealth when lottery wealth is zero ($\hat{y}_{|L_i=0}$) obtained from estimating Equation 1 in the post-1999 sample, stratified by equity market participation status at $s = -1$. The coefficients are interpreted as the effect of 1 SEK – or equivalently the share of the amount won invested in – each asset category. Columns 1-8 show estimated effects of wealth on the value of real estate holdings. Columns 9-16 show estimated effects of wealth on the value of total debt.

	A. Real Estate								**B. Debt**							
Horizon	**Participants**				**Nonparticipants**				**Participants**				**Nonparticipants**			
(s)	β_s	**SE**	N	$\hat{y}_{\|L_i=0}$	β_s	**SE**	N	$\hat{y}_{\|L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\|L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\|L_i=0}$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
0	-0.032	0.046	50,861	1.185	0.045	0.018	19,278	0.498	-0.041	0.011	50,861	0.245	-0.031	0.014	19,278	0.160
1	-0.021	0.061	47,380	1.277	0.028	0.019	17,276	0.517	-0.056	0.010	47,380	0.247	-0.026	0.017	17,276	0.159
2	0.013	0.061	43,487	1.379	0.028	0.024	15,316	0.552	-0.028	0.017	43,487	0.250	-0.004	0.019	15,316	0.159
3	0.062	0.058	40,324	1.452	0.058	0.030	13,757	0.599	-0.018	0.021	40,324	0.260	0.022	0.020	13,757	0.161
4	0.031	0.065	36,842	1.560	0.103	0.047	12,267	0.654	-0.024	0.017	36,842	0.267	0.044	0.025	12,267	0.165

Table B.11: Effect of Wealth (1M SEK) on Probability of Owning Real Estate, Being Debt-Free, and Having Self-employment Income by $s = -1$ Equity Market Participation Status. This table presents coefficients, standard errors, sample size, and mean predicted participation probability when lottery wealth is zero ($\hat{y}_{|L_i=0}$) obtained from estimating Equation 1 in the post-1999 sample, stratified by equity market participation status at $s = -1$. Columns 1-8 show estimates for the effect of lottery wealth on real estate market participation, columns 9-16 show the estimates for the effect of lottery wealth on being debt-free, while columns 17-24 show the estimates for the effect of lottery wealth on having self-employment income.

	A. Real Estate								**B. Debt-Free**							
Horizon	**Participants**				**Nonparticipants**				**Participants**				**Nonparticipants**			
(s)	β_s	**SE**	N	$\hat{y}_{\mid L_i=0}$	β_s	**SE**	N	$\hat{y}_{\mid L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\mid L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\mid L_i=0}$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
0	-0.003	0.008	50,861	0.802	0.028	0.014	19,278	0.578	0.048	0.014	50,861	0.401	0.104	0.020	19,278	0.500
1	0.003	0.010	47,380	0.801	0.052	0.018	17,276	0.573	0.047	0.015	47,380	0.408	0.065	0.022	17,276	0.512
2	-0.005	0.014	43,487	0.799	0.036	0.018	15,316	0.574	0.056	0.016	43,487	0.414	0.043	0.023	15,316	0.520
3	0.012	0.010	40,324	0.800	0.052	0.022	13,757	0.578	0.030	0.015	40,324	0.413	0.008	0.022	13,757	0.525
4	0.007	0.010	36,842	0.798	0.072	0.032	12,267	0.580	0.032	0.018	36,842	0.416	-0.041	0.024	12,267	0.534

	C. Self-Employed							
Horizon	**Yes**				**No**			
(s)	β_s	**SE**	N	$\hat{y}_{\mid L_i=0}$	β_s	**SE**	N	$\hat{y}_{\mid L_i=0}$
	(17)	**(18)**	**(19)**	**(20)**	**(21)**	**(22)**	**(23)**	**(24)**
0	0.002	0.005	50876	0.069	-0.021	0.011	19290	0.035
1	0.001	0.005	49907	0.075	-0.012	0.011	18588	0.037
2	-0.007	0.009	48960	0.082	-0.012	0.011	17831	0.041
3	-0.007	0.009	48025	0.091	-0.021	0.010	17130	0.047
4	-0.011	0.010	44654	0.089	-0.001	0.007	15264	0.046

Table B.12: Effect of Wealth on Bank Account Balances, Structured Products Holdings and Ownership, and Bond Ownership by $s = -1$ Equity Market Participation Status. This table presents coefficients, standard errors, sample size, and mean predicted change in wealth when lottery wealth is zero ($\hat{y}_{|L_i=0}$) obtained from estimating Equation 1 in the post-1999 sample, stratified by equity market participation status at $s = -1$. Columns 1-8 show estimated effects of wealth on bank account balances, columns 9-16 show estimated effects of wealth on structured product holdings. These coefficients are interpreted as the effect of 1 SEK – or equivalently the share of the amount won invested in – each asset category. Columns 17-24 show estimated effects of wealth on participation in structured product markets, while columns 25-32 show estimated effects of wealth on participation in bond markets.

	A. Bank Holdings								**B. Structured Product Holdings**							
Horizon	**Participants**				**Nonparticipants**				**Participants**				**Nonparticipants**			
(s)	β_s	**SE**	N	$\hat{y}_{\|L_i=0}$	β_s	**SE**	N	$\hat{y}_{\|L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\|L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\|L_i=0}$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
0	0.241	0.040	50,861	.215	0.197	0.035	19,278	.147	0.039	0.008	50,862	0.012	0.037	0.011	19,278	0.002
1	0.121	0.029	47,380	0.222	0.091	0.021	17,276	0.150	0.049	0.012	47,382	0.013	0.047	0.013	17,276	0.003
2	0.781	0.024	43,487	0.227	0.055	0.017	15,316	.151	0.050	0.010	43,489	0.016	0.030	0.007	15,316	0.003
3	0.088	0.278	40,324	0.248	0.039	0.017	13,757	0.158	0.045	0.010	40,328	0.020	0.032	0.012	13,757	0.004
4	0.116	0.045	36,842	0.288	0.021	0.013	12,267	0.177	0.025	0.007	36,848	0.027	0.024	0.011	12,267	0.006

	C. Structured Product Ownership								**D. Bond Ownership**							
Horizon	**Participants**				**Nonparticipants**				**Participants**				**Nonparticipants**			
(s)	β_s	**SE**	N	$\hat{y}_{\|L_i=0}$	β_s	**SE**	N	$\hat{y}_{\|L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\|L_i=0}$	β_s	**SE**	**N**	$\hat{y}_{\|L_i=0}$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
0	0.103	0.016	50,862	0.163	0.130	0.024	19,278	0.040	0.123	0.017	50,861	0.509	0.173	0.031	19,278	0.225
1	0.105	0.019	47,382	0.172	0.180	0.029	17,276	0.043	0.149	0.017	47,380	0.532	0.239	0.038	17,276	0.243
2	0.122	0.019	43,489	0.188	0.147	0.029	15,316	0.048	0.135	0.017	43,487	0.556	0.227	0.040	15,316	0.269
3	0.124	0.018	40,328	0.196	0.097	0.032	13,757	0.051	0.125	0.021	40,324	0.581	0.188	0.050	13,757	0.288
4	0.099	0.026	36,848	0.206	0.067	0.031	12,267	0.051	0.111	0.029	36,842	0.605	0.148	0.056	12,267	0.300

Table B.13: Heterogeneous Effect of Wealth (1M SEK) on Participation Probability among $s = -1$ Equity Market Participants, Belief and Information Channels. Coefficients are obtained by estimating Equation 1 at time $s = 0$ in the post-1999 sample of equity market participants at time $s = -1$, stratified by the characteristics indicated in the column headings. Recent equity return samples are defined by whether Stockholm Stock Exchange returns were negative or positive the year prior to the lottery. Early equity return samples are defined by whether a household experienced above or below average equity returns between ages 18-25. College-educated groups are defined by whether or not a household member has a college degree. Hetero p obtained from an F-test of the null hypothesis that the two lottery-wealth coefficients are identical. $\% \ Part_{-1}$ indicates the share of the post-1999 sample with the characteristic indicated by the column head that owned equity the year prior to the lottery. Table 9 provides results for time $s = -1$ equity market nonparticipants.

	Recent Equity Returns		**Early Equity Returns**		**Recent Equity Returns/Early Equity Returns**			
	Low (1)	**High (2)**	**Low (3)**	**High (4)**	**Low/Low (5)**	**Low/High (6)**	**High/Low (7)**	**High/High (8)**
Effect	-.006	.003	.-.001	.003	-.007	.138	.111	.187
SE	.008	.006	.008	.005	.021	.078	.037	.040
p	.458	.646	.921	.633	.730	.078	.002	.000
Hetero p	.386		.723					
N	29,849	21,102	27,330	23,531	5,678	4,724	4,913	3,963
$\% \ Part_{-1}$	.742	.703	.721	.730	.738	.747	.700	.707

	College/Recent Equity Returns				**College/Early Equity Returns**			
	No/Low (25)	**No/High (26)**	**Yes/Low (27)**	**Yes/High (28)**	**No/Low (29)**	**No/High (30)**	**Yes/Low (31)**	**Yes/High (32)**
Effect	.050	.125	.111	.244	.082	.145	.126	.369
SE	.040	.030	.107	.062	.032	.037	.070	.085
p	.213	.000	.296	.000	.011	.000	.060	.000
Hetero p	.137		.283		.195		.025	
Δ Effect	.075		.132		.064		.243	
N	9,014	7,496	1,388	1,380	9,095	7,415	1,496	1,272
$\% \ Part_{-1}$	.699	.669	.865	.810	.674	.700	.851	.831

C. Online Appendix - Survey Details

C.1 Introduction

This appendix reports descriptive statistics about lottery players' beliefs about stock market returns using a survey that was sent to a subset of the lottery players during the fall of 2016. The survey focused on questions related to well-being, health, and political preferences and beliefs, but also included two questions about stock market returns. The survey was sent to 241 Kombi large-prize winners and 964 (241$\times$4) matched controls, 3,065 Triss-Lumpsum winners and 570 Triss-Monthly winners. To be consistent with the baseline sample analyzed in this paper, we exclude Triss-Monthly winners from all analyzes presented below. Among the Triss-Lumpsum and Kombi lottery players that received the survey, 59 percent responded to at least one of the two questions about stock market returns. Further details about the survey and the exact criteria used when selecting the survey population is provided in Östling, Lindqvist and Cesarini (2016). Data from the same survey has been used in Lindqvist, Östling and Cesarini (2018) which shows there is no evidence that the propensity to answer the survey was related to the amount won.

C.2 Survey Questions and Definitions

The first survey question about stock market returns asks respondents to assess the probability that the market index for the Stockholm Stock Exchange will appreciate during the coming 12-month period. Respondents are given 12 pre-specified response alternatives from 0 to 100 percent. The second question asks respondents to provide an estimate for how much the market index will depreciate or appreciate during the coming 12 months. Respondents could respond with any number between -99 and $+99$ percent.

Based on the respondents' answers to these two questions, we define three measures of "extreme" beliefs about the evolution of the stock market.

1. *Extreme subjective probability*. We first define reasonable beliefs about the probability of the stock market will appreciate (the first survey question). As a benchmark, we consider the MSCI Sweden index from the previous 20-year period (1996-2016). During these 20 years, the nominal stock market index appreciated 14 years and depreciated 6 years, implying that the stock market rate appreciate 70 percent of all years. We define subjective probabilities outside of 20 percent from this benchmark (i.e., below 50 or above 90 percent) as "extreme".

2. *Extreme expected returns*. We again use the evolution of the MSCI Sweden index 1996-2016

to provide a benchmark for expected returns (the second survey question). The nominal arithmetic average return during this period is 9.6 percent (8.5 percent after adjusting for inflation). We define expected returns below 0 or above +20 percent as “extreme”.

3. *Incoherent beliefs*. Our third measure focus on the consistency between the answers to the first and second survey question. In principle, any subjective probability of stock market appreciation between 0 and 100 percent can rationalize any expected return. However, when the subjective probability is close to 0 (100) percent, the distribution of expected return must be very skewed to the right (left) in order for the expected return to be positive (negative). We define beliefs as being “incoherently positive” if the subjective probability is weakly below 25 percent while the expected return is positive and as “incoherently negative” if the subjective probability is weakly above 75 percent while the expected return is negative.

C.3 Descriptive Statistics

We begin by presenting the distributions of beliefs. Because winning the lottery might affect expectations, we restrict attention to non-winners in the Kombi lottery and small-prize winners in the Triss lottery (prizes below 150K SEK), reducing the sample to 1,749 individuals. Figure C.1 shows the distribution of subjective probability of stock market appreciation. The distribution is skewed to the right with a median of 25 percent and a mean of 33 percent. According to our definition, 70.4 percent of respondents hold extremely negative beliefs, while only 2.6 percent hold extremely positive beliefs.

Figure C.2(a) shows the distribution of expected returns. To provide a more detailed view, C.2(b) shows the distribution of beliefs when the support of the distribution is limited to returns between -15 and $+25$ percent. Compared to the question about subjective probabilities, the distribution of expected returns is more in line with what well-informed respondents would answer. The average expected return is 5.9 percent, quite close to the close to the historical inflation-adjusted average of 8.5 percent. Only 25.7 percent of respondents hold extreme beliefs about expected returns. As for the subjective probability, unrealistically negative returns are more common (18.8 percent of respondents) than unrealistically positive returns (6.9 percent).

We now turn to incoherent beliefs, i.e., the relationship between expected return and the subjective probabilities. Figure C.3 plots the expected return against the subjective probability with the size of each circle being proportional to the number of respondents. As expected, respondents with a high subjective probability on average report a higher expected return ($\rho = 0.29$, $p < 0.0001$),

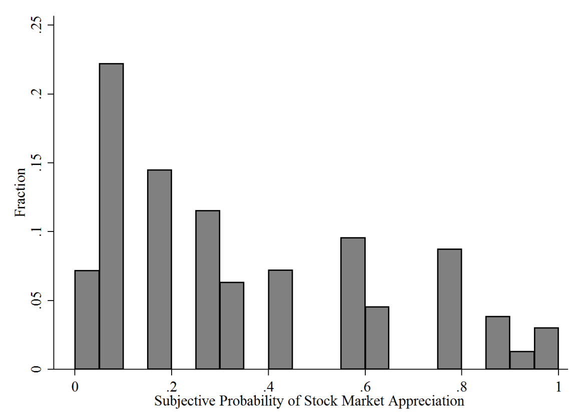
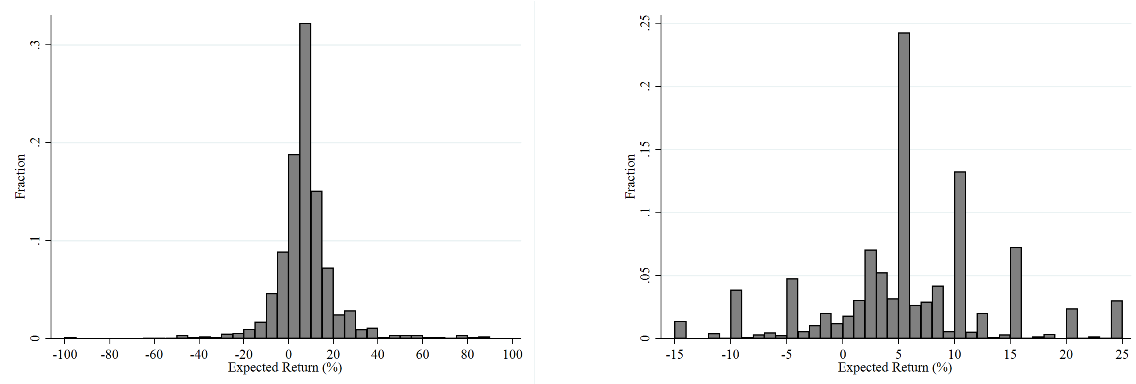


Figure C.1: Histogram of Subjective Probability of Stock Market Appreciation Histogram for the probability the subjects attach to the stock market appreciating during the next 12 months. N = 1,738.

though substantial variation in the expected return among respondents who report the same subjective probability. Regressing expected return on the subjective probability, we find an increase in subjective probability by 100 percentage point is associated with a 13.2 percentage point increase in the expected return.

Figure C.4 shows the distribution of expected return for respondents whose subjective probability of stock market appreciation is below 25 or above 75 percent, respectively. While the distribution for respondents who attach less than 25 percent chance of the stock market appreciating is clearly shifted to the left, it is noteworthy that the modal expected return is positive also for respondents with low subjective probability. Many respondents who report an unreasonably low subjective probability thus seem to have realistic expectations about stock market returns. A potential explanation for this finding and for the high proportion of respondents who report unreasonably low subjective probabilities is that respondents are used to thinking about the stock market in terms of expected returns and fail to properly grasp what is meant by the probability of the stock market appreciation.

Panel A of Table C.1 shows the different combinations of subjective probability and expected returns. The sample is restricted to the 1,587 individuals who won less than 150K SEK and for whom we observe both the expected return and the subjective probability. The most common combination of beliefs, held by about half of respondents (47.1 percent), is to believe in a realistic expected return but attach an overly negative probability to the stock market appreciating. Only about a quarter of respondents (24.8 percent) hold realistic beliefs about both the expected return and the



(a) Full Distribution **(b) Limited Support**

Figure C.2: Histogram of Expected Returns. Histogram of respondents' expected stock market return for the coming 12 months. The full distribution is presented in Panel (a), while Panel (b) plots this same distribution omitting extreme values of subjective expected returns. N = 1,589.

probability of stock market appreciating. Also common are overly negative views about both the expected return and subjective probability (18.1 percent). Panel B shows the share of respondents who hold incoherent beliefs, as defined in Section II above. While the fraction of respondents who hold "incoherently negative" beliefs is negligible, more than 40 percent of respondents hold beliefs we classify as "incoherently positive", meaning they state a positive expected return but attach a probability below 25 percent to the possibility of the stock market appreciating. There are two stylized facts to take from the descriptive analyses above. First, irrespective of whether we ask respondents about the expected return or subjective probability that the stock market will appreciate, beliefs are overly pessimistic. Second, a substantial fraction of respondents report answers which are mutually incoherent, suggesting that their basic financial literacy is low. The high fraction of respondents who state a low subjective probability indicate that people might be more used to thinking about the stock market in terms of expected returns.

We now consider which factors predict beliefs about stock market returns. To this end, we regress the different measures of beliefs discussed above on a set of basic demographic characteristics. To test whether beliefs are affected by the stock market return in formative years, we also include the average stock market return between age 18 and 25 as an independent variable. The results from these regressions are reported in Table C.2. We emphasize that our results in this part are purely descriptive and do not have a causal interpretation.

Columns 1-3 of Table C.2 shows the results for the subjective probability. The subjective

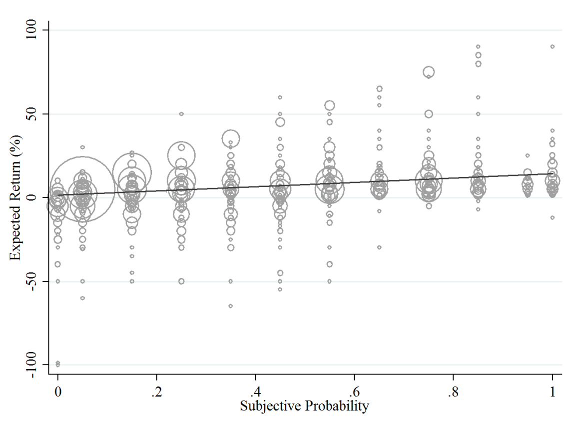


Figure C.3: Expected Return vs. Subjective Probability Each circle indicates a certain combination of subjective probability and expected return. The area of each circle corresponds to the number of respondents within each circle. N = 1,587.

probability is 9.6 percentage points lower for women, and women are more likely to be overly negative and less likely to be overly positive. The subjective probability is higher for respondents who are born in Sweden, have high labor earnings, many children, and a college degree. The same characteristics also predict a lower risk of overly negative beliefs, but are not associated with a higher risk of being overly positive. Columns 4-6 of Table C.2 shows the corresponding results for the expected return. Notably, none of the characteristics that predict the subjective probability predict the expected return. Perhaps surprisingly, women appear more likely to report an overly positive belief about the expected return. Labor income is negatively correlated both with overly positive and overly negative expected returns, indicating that people with high income have a better sense of what returns seem plausible. The stock market return in formative age appear to be unrelated to both the subjective probability and the expected return.

While we saw above that respondents, in general, are too pessimistic about stock market returns, no characteristics except labor income predict overly negative beliefs for both the subjective probability and the expected returns. One explanation is that pessimistic beliefs are simply weakly correlated with basic socio-economic characteristics. Yet, if so, it remains to explain why the subjective probability is quite strongly related to several socio-economic characteristics, so one possibility is that many respondents have trouble understanding the question about subjective probability, and that the correlations between the subjective probability and socio-economic characteristics are due to low financial literacy rather than pessimism per se. Consequently, the

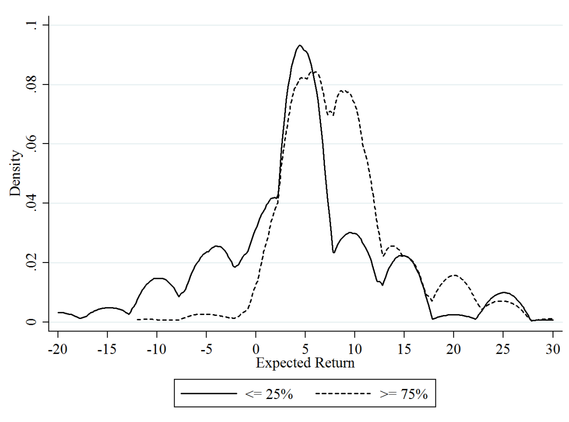


Figure C.4: Expected Return by Subjective Probability The figure shows the Kernel density plots (bandwidth = 1.25) for the distribution of expected returns by the subjective probability that the stock market will appreciate next year. The solid line indicates respondents whose probability is below 25 percent whereas the dashed line indicates respondents with a probability above 75 percent. N = 855 (subjective probability below 25 percent) and N = 271 (subjective probability above 75 percent).

large fraction of respondents who report overly negative beliefs about the subjective probability in Table C.1 may thus at least partly reflect a lack of financial literacy or basic numeracy skills. This interpretation is broadly consistent with previous literature which has revealed lower financial literacy among women (Lusardi and Mitchell 2008; Almenberg and Dreber 2015) and for people with low education (van Rooij, Lusardi and Alessie 2011). The results in Column 7 of Table C.2 – in which we regress an indicator variable for incoherent beliefs on the same set of socioeconomic characteristics – lend some support to this view. Female respondents are 12.5 percentage points more likely and respondents with a college degree 15.6 percentage points more likely to report incoherent beliefs, and both effects are strongly statistically significant.

C.4 Conclusion

Our analysis of the survey responses show that for both our survey questions aimed at eliciting the respondents' beliefs, a large fraction of respondents reports pessimistic beliefs about the evolution of the stock market. Moreover, both the large number of respondents who attach a probability close to zero to the stock market appreciating, and the large fraction whose responses are appear mutually incoherent, suggest financial literacy is low in our sample. The propensity to state extreme or incoherent beliefs also correlate with socio-economic factors which previous literature has shown predict low financial literacy.

Table C.1: Frequency of Belief Categories. Panel A presents the joint distribution of the three defined belief categories. The rows indicate respondents that reported overly positive, realistic, and overly negative subjective probabilities that the Stockholm Stock Exchange will appreciate during the coming 12-month period. The rows indicate respondents that reported overly positive, realistic, and overly negative expected equity returns over the coming 12 month period. Panel B presents the distribution of respondents that have coherent beliefs, incoherently negative beliefs, and incoherently positive beliefs. Incoherently positive beliefs are defined by reporting a positive expected return but a probability below 25 percent of positive stock returns in the next 12 months, incoherently negative beliefs are defined by reporting a negative expected return but assigning a probability of positive returns above 75%, and coherent beliefs are the remainder. Sample of non-winners and winners of prizes below 150K SEK for which both subjective probability and expected return is observed. N = 1,587.

Panel A. Combination of Beliefs

	Expected returns			
	Overly positive (>+20%)	**Realistic (0% to +20%)**	**Overly negative (<0%)**	**Sum**
Subjective probability				
Overly positive (>90%)	0.3	2.4	0.1	2.8
Realistic (50% to 90%)	2.7	24.8	0.6	28.2
Overly negative (<50%)	3.9	47.1	18.1	69.1
Sum	6.9	74.3	18.8	100

Panel B. Coherent vs Incoherent Beliefs

Percent of	**Coherent**	**Incoherently positive**	**Incoherently negative**	**Sum**
Respondents	59.2	40.5	0.3	100

Table C.2: Predictors of Stock Market Return Beliefs. All regressions estimated with OLS in the survey sample of non-winners and winners of 150K SEK or less. Column 1 considers the stated probability that the Stockholm Stock Exchange will appreciate during the coming 12-month period, Column 2 considers overly positive stated probabilities (stated probability $> 90\%$), and Column 3 considers overly negative stated probabilities (stated probability $< 50\%$). Column 1 considers the stated expected return on the Stockholm Stock Exchange over the coming 12 months, Column 2 considers overly positive expected returns (stated probability $> 20\%$), and Column 3 considers overly negative expected returns (stated probability $< 0\%$). Column 7 regresses on an indicator of incoherent beliefs, defined as either reporting a positive expected return but stated probability below 25 percent or a negative expected return but a stated probability above 75%. All time-varying independent variables refer to year 2014. Labor and Capital income have been winsorized at the 1st and 99th percentile. Robust standard errors are reported in parenthesis and the corresponding *p*-values in brackets.

	Subjective probability			**Expected return**			
	Stated probability (0-1)	**Overly positive (0/1)**	**Overly negative (0/1)**	**Stated return (%)**	**Overly positive (0/1)**	**Overly negative (0/1)**	**Incoherent beliefs (0/1)**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Female	-0.096	-0.021	0.147	0.115	0.027	0.023	0.125
	(0.014)	(0.008)	(0.022)	(0.724)	(0.014)	(0.020)	(0.025)
	[0.000]	[0.009]	[0.000]	[0.874]	[0.049]	[0.251]	[0.000]
Age/10	-0.091	-0.010	0.122	-1.710	0.029	0.095	0.028
	(0.046)	(0.028)	(0.072)	(2.256)	(0.042)	(0.066)	(0.081)
	[0.048]	[0.715]	[0.089]	[0.449]	[0.490]	[0.153]	[0.733]
Age/10 squared	0.007	0.000	-0.010	0.231	-0.003	-0.013	0.002
	(0.005)	(0.003)	(0.008)	(0.245)	(0.005)	(0.007)	(0.009)
	[0.160]	[0.973]	[0.186]	[0.346]	[0.476]	[0.082]	[0.817]
Married	0.004	-0.001	0.011	0.035	-0.001	-0.011	0.025
	(0.014)	(0.008)	(0.022)	(0.667)	(0.013)	(0.020)	(0.025)
	[0.747]	[0.869]	[0.603]	[0.959]	[0.928]	[0.600]	[0.314]
Born in Sweden	0.063	0.011	-0.100	0.962	-0.001	-0.002	-0.058
	(0.027)	(0.013)	(0.039)	(1.464)	(0.026)	(0.041)	(0.050)
	[0.019]	[0.384]	[0.010]	[0.511]	[0.982]	[0.956]	[0.251]
Capital income (in 100K SEK)	0.019	0.004	-0.032	0.375	-0.009	-0.015	-0.011
	(0.014)	(0.006)	(0.022)	(0.614)	(0.008)	(0.016)	(0.023)
	[0.184]	[0.524]	[0.144]	[0.541]	[0.285]	[0.335]	[0.645]
Labor income (in 100K SEK)	0.021	-0.001	-0.029	0.058	-0.011	-0.022	-0.011
	(0.005)	(0.003)	(0.008)	(0.245)	(0.004)	(0.007)	(0.008)
	[0.000]	[0.799]	[0.000]	[0.813]	[0.009]	[0.001]	[0.200]
Number of children	0.020	-0.002	-0.035	0.802	0.007	-0.020	-0.020
	(0.008)	(0.005)	(0.014)	(0.419)	(0.008)	(0.012)	(0.015)
	[0.018]	[0.613]	[0.010]	[0.056]	[0.416]	[0.103]	[0.180]
College degree	0.085	0.006	-0.136	-0.846	-0.015	0.006	-0.156
	(0.016)	(0.009)	(0.026)	(0.788)	(0.014)	(0.024)	(0.027)
	[0.000]	[0.502]	[0.000]	[0.283]	[0.271]	[0.800]	[0.000]
Stock market return in formative age	-0.035	-0.009	-0.001	-5.363	-0.054	0.197	-0.269
	(0.124)	(0.072)	(0.194)	(5.730)	(0.113)	(0.179)	(0.215)
	[0.779]	[0.901]	[0.997]	[0.349]	[0.635]	[0.273]	[0.210]
R^2	0.092	0.009	0.080	0.005	0.012	0.016	0.066
N	1,736	1,736	1,736	1,596	1,596	1,596	1,585