Journal of Financial Economics 139 (2021) 57–83

Contents lists available at ScienceDirect

Journal of Financial Economics

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

Windfall gains and stock market participation☆

Joseph Briggs[a,*], David Cesarini[b], Erik Lindqvist[c], Robert Östling[d]

[a] Federal Reserve Board of Governors, Constitution Ave NW and 20th St NW, Washington, DC 20551, United States
[b] New York University, IFN, and NBER, 19 W. 4th Street, New York, NY 10012, United States
[c] SOFI, Stockholm University and IFN, Universitetsvägen 10F, SE-106 91 Stockholm, Sweden
[d] Stockholm School of Economics, P.O. Box 6501, SE-113 83 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 3 November 2016
Revised 16 September 2019
Accepted 6 November 2019
Available online 24 July 2020

JEL classification:
D14
D15
G11
G51

Keywords:
Household saving and personal finance
Intertemporal consumer choice
Portfolio choice and investment decisions

ABSTRACT

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 the stock market participation probability by 12 percentage points among prelottery nonparticipants but has no discernible effect on prelottery stock owners. A structural life cycle model significantly overpredicts entry rates even for very high entry costs (up to $31,000). Additional analyses implicate pessimistic beliefs regarding equity returns as a major source of this overprediction and suggest that both recent and early-life return realizations affect beliefs.

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

☆ 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. 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 for helpful comments and Joel Becker and Nancy Wang for research assistance. We 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.

* Corresponding author.
E-mail address: jsbriggs87@gmail.com (J. Briggs).

[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.

https://doi.org/10.1016/j.jfineco.2020.07.014
0304-405X/Published by Elsevier B.V.

costs.[2] Early work by Vissing-Jørgensen (2003) posits 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 that 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), who use 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 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 postlottery 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 following lottery wins much larger than the reduced-form estimates. Consequently, 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 prelottery equity market nonparticipants is over 31K USD, approximately ten times larger than the average cost estimated from nonexperimental 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 nonstandard beliefs. While these explanations are not mutually exclusive and there is some support for each, the evidence points to nonstandard 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.

We next exploit temporal variation in equity returns to check for insights toward the underlying belief formation process. Most belief formation theories fall into one of two groups: experience-based learning (see, e.g., Malmendier and Nagel, 2011; Giuliano and Spilimbergo, 2013; Malmendier et al., 2020) and theories of overinference (see, e.g., Fuster et al., 2010; Fuster et al., 2012; Greenwood and Shleifer, 2014; Gennaioli et al., 2016). Both classes of theories are consistent with overweighting of recent returns, whereas only the experience-based models are consistent with overweighting of returns during early formative years. We find that effects on stock market entry are larger both following years with positive equity returns (recency bias) and among individuals who experienced positive returns during formative years (early life bias). Furthermore, these patterns are present even among the highly educated. Although rigorously distinguishing between belief-updating models is beyond the scope of our study,

[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.

[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.

our results suggest that these belief formation theories are relevant for explaining stock market nonparticipation.

The paper is structured as follows. Section 2 describes the lottery and wealth data and 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 et al. (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 lumpsum and all amounts are after tax. In this paper, all prize amounts (and other financial variables) are adjusted for inflation and are expressed in year-2010 SEK and USD, assuming the December 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 postwar 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 nonwinning 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 lumpsum prizes between 7.8K USD (52K SEK) and 909K USD (6.1M SEK). However, one analysis in Section 3 compares estimates for lumpsum prize winners to a "Triss monthly" sample of 476 participants who received prizes paid in monthly installments for 10 to 25 years (see Online Appendix Table B.1 for descriptive statistics). We convert the installments to net present value to make them comparable to lumpsum 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. Prize-linked savings

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

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 that are included as fixed effects in Eq. (1) to achieve conditional random assignment of lottery prizes.

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

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 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

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

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 Eq. (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 prelottery participation status, which is first observed in 1999. In the all-year sample regressions, the set of prelottery controls include age, sex, marital status, higher education, number of children in the household, 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 nonwinners, 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 1012 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

[5] We retain nonwinners in Kombi in the sample when dropping small prizes. Because all players in the Kombi lottery won a large prize or nothing, dropping the nonwinners eliminates almost all 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. See Table 4 for sample details. Prize amounts are in year-2010 USD and are net of taxes.

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
$15 < L \leq 75$	4009	1935	0	2074	1237	0	0	1237
$75 < L \leq 150$	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	3404	70,139	41,948	26,415	1776

Table 3
Testing for random assignment.
Results are obtained by estimating Eq. (2) in our all-year sample (Columns 1–2), in the post-1999 sample (Columns 3–4), and in the post-1999 lottery-specific subsamples (Columns 5–7). See Table 4 for sample details. *F*-statistics and corresponding *p*-values result from testing the joint significance of the indicated controls. Demographic controls include age, sex, marital status, higher education, household size, household income, and an indicator for being Nordic born. Financial controls include net wealth, gross debt, and an indicator for real estate ownership, all measured at time $s = -1$.

	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	1776

regression:

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

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. However, if they are omitted (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 represent 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 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, has 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.

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 are 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 they own equity either directly or indirectly via mutual funds. Continuous financial variables are winsorized at the 0.5 and 99.5 percentiles.

	All-Year		Post-1999				
	Pooled (1)	Pop (2)	Pooled (3)	Pop (4)	PLS (5)	Kombi (6)	Triss (7)
Demographic							
Female	0.50	0.50	0.52	0.52	0.58	0.44	0.56
Age (years)	56.6	56.6	56.2	56.2	62.9	61.7	51.9
Nordic born	0.96	0.93	0.96	0.92	0.95	0.98	0.94
Children in household (#)	0.38	0.41	0.43	0.42	0.24	0.22	0.59
Household income (K USD)	48	45	54	54	49	51	57
Married	0.56	0.57	0.52	0.54	0.52	0.48	0.54
College	0.23	0.24	0.24	0.31	0.27	0.22	0.26
Financial							
Net wealth (K USD)			131	158	205	123	128
Gross debt (K USD)			53	49	27	37	67
Homeowner			0.75	0.69	0.73	0.78	0.73
Equity market participant			0.66	0.63	0.74	0.69	0.63
N	367,577	367,577	70,139	70,139	41,948	26,415	1776

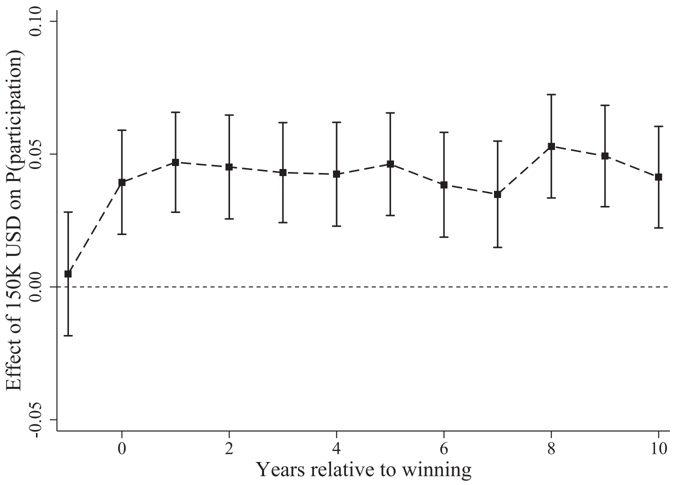


Fig. 1. Effect of 150K USD (1M SEK) of lottery wealth on participation probability. Coefficients and 95% confidence intervals are obtained by estimating Eq. (1) in the all-year sample. See Table 4 for sample details. See Online Appendix Table B.3 for the underlying estimates.

Another way to gauge representativeness is to compare the cross-sectional relation between stock market participation and household characteristics in the lottery samples to the relation 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.'s (2007) 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. Online Appendix Table B.2 shows that the results from these regressions are quite similar.

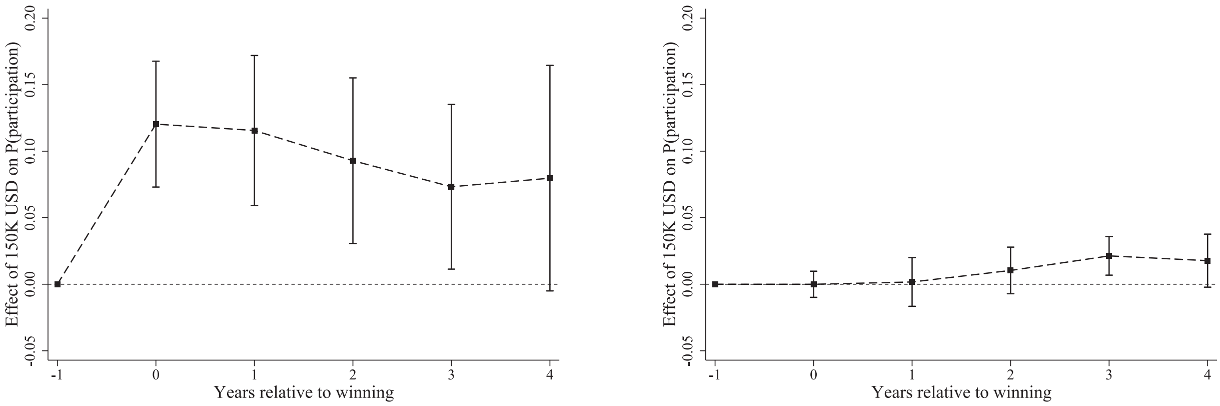
While the absence of large differences in prelottery financial and demographic characteristics between the lottery sample and the representative sample is reassuring, the possibility that selection into lotteries is based upon unobserved factors that limit the external validity of our results cannot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

3. Quasi-experimental estimates

In this section we present our main reduced-form 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 one for households that own stocks either directly or indirectly via mutual funds. Fig. 1 presents the estimated coefficients for $s = -1, \ldots, 10$ from the all-year lottery sample. Each 150K USD (1M SEK) causes a near-immediate and permanent increase in the partic-



(A) Nonparticipants (B) Participants

Fig. 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 Eq. (1) in the post-1999 sample of nonparticipants (A) and participants (B). See Table 4 for sample details. See Online Appendix Table B.4 for the underlying estimates.

ipation 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. Fig. 2 shows the estimated treatment effects on participation at $s = -1, \ldots, 4$ in the post-1999 sample stratified by prelottery participation status. Among prelottery nonparticipants, each 150K USD (1M SEK) increases participation probability by 12 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 prelottery participants (for whom lottery wealth might increase participation by discouraging equity market exit) is small and mostly not statistically distinguishable from zero. 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 Eq. (1) by indicator variables for five categories defined according by prize size and run regressions with the smallest prize category omitted.

Fig. 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 prelottery 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, respectively. Thus, the marginal effect (defined as the slope between points in Fig. 3) is everywhere positive but is largest for winners of small prizes. Among prelottery participants, none of the prize category coefficients are statistically distinguishable from zero.

3.1.2. *Effects by prize payment: lump sum versus monthly installments*

The finding that a majority of prelottery 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 lumpsum prize. If up-front costs determine stock market participation and winners cannot perfectly borrow against future installments, a liquid lumpsum prize would result in a larger effect on participation than illiquid monthly installments.

Fig. 4 shows the effect for nonparticipants by type of prize in the Triss lottery (the results for participants are shown in Online 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 the 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 to disincentive participation than continued costs of participation.[7]

[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.

[7] One complicating factor when comparing Triss lump sum and Triss monthly is that the support of the prize distribution in the two lotteries

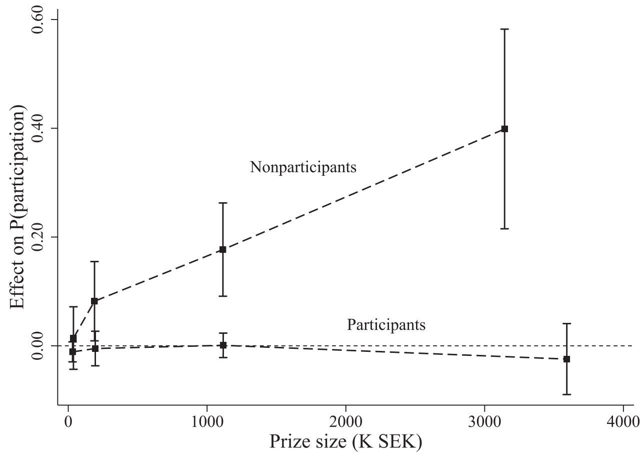


Fig. 3. Effect of lottery wealth on participation probability by prize size. Coefficients are obtained by estimating Eq. 1 in the post-1999 sample with the lottery wealth variable replaced by indicators for five mutually exclusive prize categories: 0 to 1.5K USD (0 to 10K SEK), 1.5 to 15K (10K to 100K), 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 Table 4 for sample details. See Online Appendix Table B.5 for the underlying estimates.

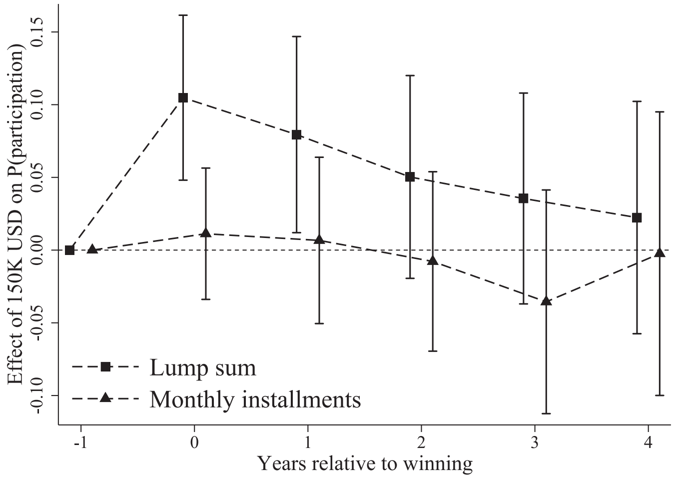


Fig. 4. Effect of 150K USD (1M SEK) of lottery wealth on participation by payment form. Coefficients and 95% confidence intervals are obtained by estimating Eq. (1) in the post-1999 sample of Triss winners stratified by type of payment plan (lumpsum or monthly installments) for $s = -1$ equity market nonparticipants. See Online Appendix Table B.1 for sample details. See Online Appendix Table B.6 for the underlying estimates and corresponding estimates for of $s = -1$ participants.

In classical models with complete markets (e.g., Samuelson, 1969), participation and entry costs are equivalent and the household problem can be simplified 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 applying an appropriate economic framework, which we turn our attention to in Section 4.

differ (52K to 6.1M SEK in Triss lump sum, 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 that the difference between Triss monthly and Triss lump sum is robust to an alternative estimation strategy that uses the panel dimension of the data and compares winners before and after the lottery.

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 (Online Appendix Table B.3, Panel B) or exclude spousal assets from the participation definition (Online 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. Online 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 Online 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 Online 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 do not 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. Online Appendix Table B.7, Columns 16–21 shows that the results are similar across lotteries.[9] Since selection into lotteries 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-Jørgensen, 2002; Campbell, 2006). To address concerns that selection of wealthier and older nonparticipants drive our results, we reran the analyses in subsamples stratified by prelottery wealth and age quartiles. The results are summarized in Online Appendix Table B.8. Despite marked differences in prelottery 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 Online Appendix Table B.7 show that probit marginal effects are similar to the linear 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 Fig. 3. Overall, our results appear robust to alternative participation definitions, sample restrictions, and estimation strategies.

4. Structural analysis

Previous structural work has 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 to match observed participation patterns.[10] In this section, we estimate a structural model to analyze whether this conclusion also holds up 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.

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 postsecondary 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.

[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 1990s, 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. Online 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.

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

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

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

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 standard deviations, 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 nonstochastic and is defined by a replacement rate λ of the age 65 permanent component of income, where λ varies with education level 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^s_t. Equity returns are assumed to be lognormal, with mean excess return μ_s. Log equity returns are denoted

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

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 nonzero wealth to equity holdings. Participation statuses are denoted as I_t and $Part_t$, where I_t equal to one 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 nonconsumed 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 $r_f = 0.02$, and the excess return and standard deviation on equities are $\mu_s = 0.04$ and $\sigma_s = 0.21$, 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 Fig. 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 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,

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 Eq. (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	0.04	Transitory risk: σ_U	0.188	0.188	0.205
Equity risk: μ_s	0.21	Permanent risk: σ_N	0.110	0.106	0.110
Risk-free return:r_f	0.02	Equity correlation: $\rho_{n,r}$	0.002	−0.001	−0.008
		Rep. rate, single: λ_{t_R}	0.685	0.641	0.617
		Rep. rate, married: λ_{t_R}	0.644	0.608	0.589

we follow the empirical policy function (EPF) approach proposed in Bazdresch et al. (2017).

An EPF is an estimate of the relation 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):

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

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

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

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 prelottery EPFs, as $l_{i,s} = 0$ globally, but is 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 Eq. (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 $k = 1, \ldots, K$ given θ. Our identifying moments are coefficients b_j from Eq. (9), and moment conditions are specified as the vector of differences between model-implied and empirical 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 Eq. (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 $s = -4, \ldots, -1$ (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 Eq. (9) to generate empirical moments $b(\nu_i)$.

[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 Eq. (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 the model fit is worse.

Table 6
Structural estimation results and predictions.
Column 1 presents results when the model is estimated using only prelottery observations and matching prelottery EPF coefficients, Column 2 using only postlottery observations and matching postlottery EPF coefficients, Column 3 using observations both pre- and postlottery data and matching pre- and postlottery EPF coefficients, and Column 4 using postlottery observations to estimate the entry cost distribution (Fig. 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, and 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 (see Table 4 for sample details), and the corresponding coefficients from regressions on consumption are presented in Appendix Table A.1.

		Prelottery (1)	Postlottery (2)	Pre- and post (3)	Nonlinear (4)
A. Parameter estimates ($\hat{\theta}$)					
Time discounting – β		0.869	0.902	0.896	0.869
		(0.019)	(0.012)	(0.006)	–
Bequest – b		5.191	1.327	3.106	5.191
		(1.668)	(0.688)	(1.700)	–
Risk aversion – ρ		3.162	2.360	2.342	3.162
		(0.097)	(0.091)	(0.211)	–
IES – ψ		0.645	0.595	0.669	0.645
		(0.077)	(0.070)	(0.063)	–
Entry cost (K USD) – χ		3.217	31.262	12.503	–
		(1.668)	(0.688)	(0.859)	
Participation cost (K USD) – κ		0.001	0.004	0.036	0.001
		(0.003)	(0.003)	(0.006)	–
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 versus model predictions					
	Benchmark	Model predicted effect			
		(1)	(2)	(3)	(4)
i. Linear effect (150K USD)					
All	0.028	0.101	0.030	0.067	0.029
Nonparticipants	0.104	0.313	0.113	0.209	0.104
Participants	0.002	0.000	0.000	0.000	0.000
ii. Nonlinear, nonparticipants					
$1.5K < L \leq 15K$	−0.012	0.013	0.013	−0.012	0.006
$15K < L \leq 150K$	0.078	0.172	0.017	−0.016	0.080
$150K < L \leq 300K$	0.156	0.644	0.026	0.462	0.158
$300K < L$	0.359	0.953	0.591	0.976	0.357
C. External validity test, χ^2 (d.f.) (untargeted coefficients, Panel B)					
Linear and nonlinear (B.i,ii)		1084.5 (7)	–	–	–
Nonlinear (B.ii)		441.5 (4)	127.8 (4)	390.3 (4)	–

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 Eq. (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 Eq. (11), and calculate the objective function defined in Eq. (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 Eq. (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 Eq. (9) as detailed in Appendix A.7. Finally, to evaluate the 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 prelottery data

Our estimation results based on prelottery 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 constant relative risk aversion 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 ($\beta = 0.869$) that is lower than most macro models, is necessary to limit wealth accumulation. FGG (estimates between 0.75 and 0.83) and CZ (0.76–0.90) 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 ($\psi = 0.645$) estimates are comparable to the baseline estimates in CZ ($\rho = 4.409,\ \psi = 0.601$). Because $1/\rho = 0.316$ is significantly lower than $\psi = 0.645$, 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 3217 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.

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 show decently approximated 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 are quantitatively much larger. Panel C formally shows 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 reestimate our model using only participation decisions after the lottery event. The model targets 16 benchmarks: all coefficients except for cell fixed effects from the postlottery participation and consumption EPFs, and the lottery coefficients from participation and consumption regressions by prelottery participation status. The exact regressions, coefficients, and resulting model fit are presented in Appendix A.7 and Appendix 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 prelottery 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 cannot generate marginal propensities to consume 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

[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 to differences in specification between Eqs. (1) and (A.6).

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 prelottery 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 postlottery benchmarks matched separately in Columns 1 and 2. In practice, we stack the two moment vectors from our previous two estimations and reestimate Eq. (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 prelottery 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 3217 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 prelottery 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

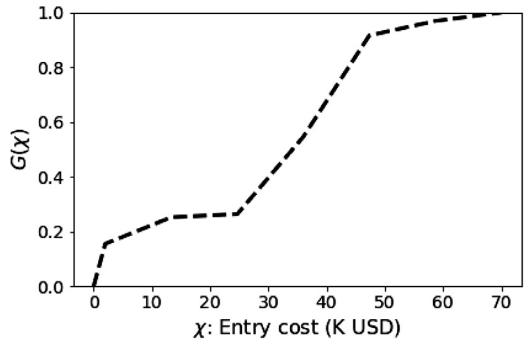


Fig. 5. Structural estimates of entry cost distribution. This figure presents the estimated CDF of entry costs defined in Eq. (14) in the estimation exercise that allows for entry cost heterogeneity (Table 6, Column 4). The model is estimated using using postlottery observations from the post-1999 sample (see Table 4 for sample details), with parameters besides entry costs held fixed at prelottery estimates in Table 6, Panel A, Column 1.

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 equidistant 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 prelottery 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.

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 distribution G_χ is presented in Fig. 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; 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.

[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 Eq. (12) with optimal weighting matrix determined by the stacked influence functions. Test statistics are undefined because this system is just identified.

Table 7
Heterogeneous effect of wealth (1M SEK) on participation probability among $s = -1$ equity market nonparticipants.

Coefficients are obtained by estimating Eq. (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 in households with and without college degrees, while Columns 9 and 10 show effects for nonparticipants that have above- and below-median cognitive skill among the subsample with conscription records available. Hetero p is 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 4 for sample details. See Online Appendix Table B.9 for results for time $s = -1$ equity market participants.

	A. Financial characteristics						B. Information proxies			
	Homeowner		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	0.147	0.105	0.212	0.092	0.131	0.046	0.107	0.223	0.039	0.304
SE	0.052	0.027	0.037	0.025	0.026	0.040	0.025	0.053	0.055	0.147
p	0.005	0.000	0.000	0.000	0.000	0.246	0.000	0.000	0.476	0.038
Hetero p	0.474		0.007		0.079		0.050		0.090	
N	8022	11,256	9545	9733	18,628	650	16,510	2768	804	957
% $Part_{-1}$	0.554	0.784	0.679	0.759	0.719	0.832	0.686	0.842	0.677	0.790

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 nonstandard beliefs and belief formation processes. Since the model-based predictions are only wildly at odds with our reduced-form estimates for prelottery nonparticipants, all analyses in this section are restricted to nonparticipants unless otherwise noted.

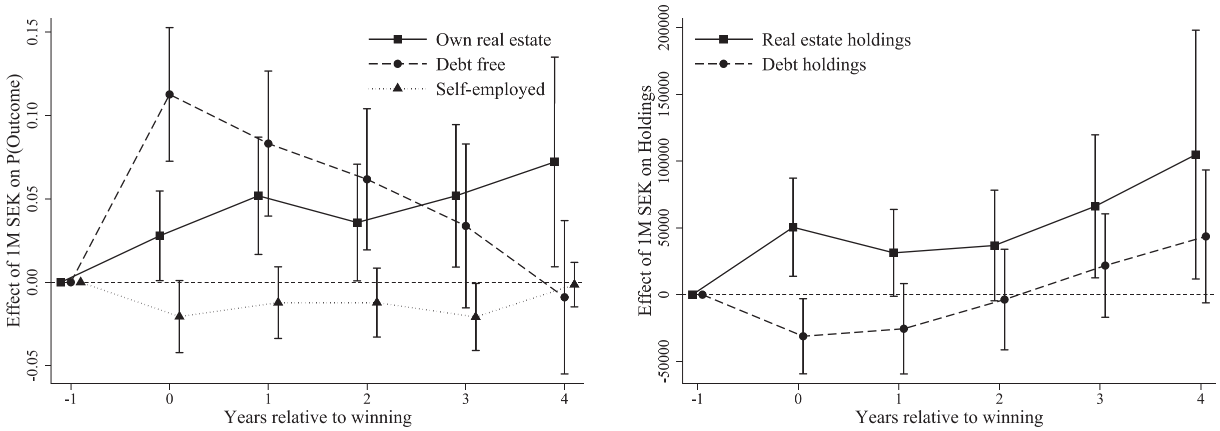
To preview the findings, there is strong evidence that nonstandard expectations and belief formation processes contribute to the discrepancy between empirical and model-implied estimates. In a survey fielded to a subset of our lottery sample, many people reported subjective beliefs that are more pessimistic 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) and 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 prelottery homeownership, 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 homeowners (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 postlottery years (Fig. 6, Panel A). We estimate that each 150K USD (1M SEK) increases the probability of being debt free in the year of win by 10 percentage points, but the effect appears to dissipate with time and is no longer



(A) Extensive margin (B) Intensive margin

Fig. 6. Effect of wealth (1M SEK) on real estate, debt investment, and self-employment. Coefficients and 95% confidence intervals are obtained by estimating Eq. (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 Table 4 for sample details. See Online Appendix Tables B.10 and B.11 for the underlying estimates and results for $s = -1$ equity market participants.

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%.

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 nonparticipants, 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.

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

5.2. Alternative preferences

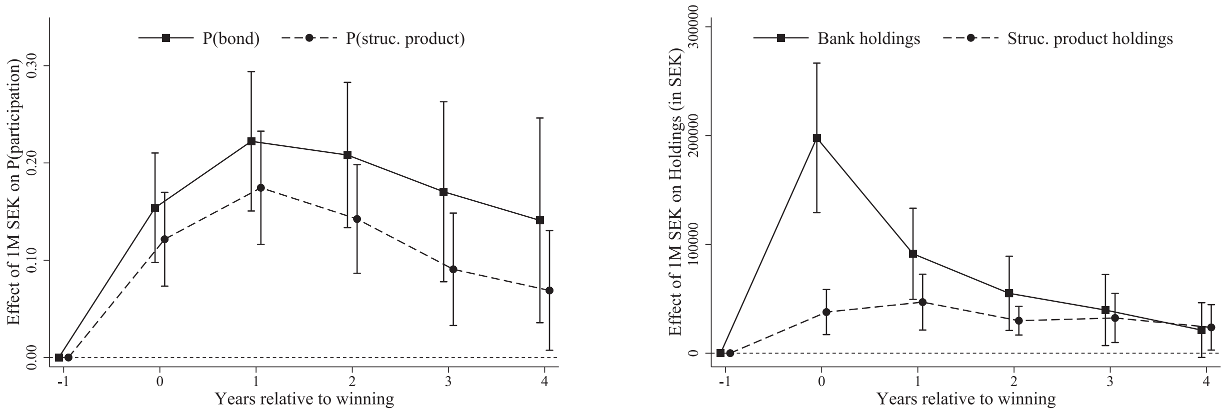
The analyses in this section are intended to shed some light on the possible role of status quo biases, loss-averse preferences, or present biased preferences in accounting for the discrepancy between our empirical and model estimates. Each of these three factors has been proposed as a source of nonparticipation.

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. Fig. 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 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. Fig. 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



(A) Extensive margin

(B) Intensive margin

Fig. 7. Effect of wealth (1M SEK) on bank/bond and structured product investment. Coefficients and 95% confidence intervals are obtained by estimating Eq. (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 and 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 Table 4 for sample details. See Online Appendix Tables B.10 and B.11 for the underlying estimates and results for $s = -1$ equity market participants.

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 Fig. 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 (see, 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.

Fig. 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 find 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 reestimate the structural model (using prelottery data for both participants and nonparticipants) assuming a present bias parameter of $\beta = 0.6$ 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 prelottery 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 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 are likely to differ from the process implicitly assumed in our model. This section reports a number of analyses that explore whether and how nonstandard

[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 quasihyperbolic discounting in a life cycle model with Epstein–Zin preferences.

Table 8
Structural model predictions, alternative specifications, and calibrations.
Columns 1 and 2, respectively, present estimates and model predictions (assuming prelottery 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 $\beta = 0.6$ and reestimating the model using prelottery data. Columns 5 and 6 present estimates and model predictions (assuming prelottery parameters from Table 6, Panel A, Column 1) after restricting the post-1999 sample to households with secondary education and above-median cognitive ability for those winners with available conscription records. Columns 7 and 8 present estimates and model predictions (assuming prelottery parameters from Table 6, Panel A, Column 1) from our post-1999 sample (see Table 4 for sample details) in which the subjective equity premium is sampled from the surveyed distribution presented in Fig. 8.

	Restricted finances subsample		Present bias preferences		High information subsample		Subjective beliefs	
	Benchmark (1)	Model (2)	Benchmark (3)	Model (4)	Benchmark (5)	Model (6)	Benchmark (7)	Model (8)
i. Linear effect (150K USD)								
All	0.040	0.090	0.028	0.115	0.013	0.067	0.028	0.066
Nonparticipants	0.145	0.264	0.104	0.340	0.163	0.248	0.104	0.197
Participant	−0.012	0.000	0.002	0.000	−0.026	0.000	0.002	0.000
ii. Nonlinear, nonparticipants								
$10K < L \leq 100K$	0.013	0.003	−0.012	−0.012	–	–	−0.012	−0.014
$100K < L \leq 1M$	0.107	0.028	0.078	0.097	–	–	0.078	0.003
$1M < L \leq 2M$	0.167	0.465	0.156	0.680	–	–	0.156	0.452
$2M < L$	0.564	0.739	0.359	0.963	–	–	0.359	0.709
N	16,329		70,139		3,355		70,139	

beliefs contribute to the discrepancy between the baseline estimates and the model-based predictions.

Our first analysis is motivated by prior work that has found individuals with higher education and cognitive test scores are more likely to report beliefs better aligned with historical data (e.g., Kézdi and Willis, 2011; Kuhnen and Miu, 2017). If nonstandard 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] We also examined if the discrepancy between the model's predictions and the data persists after restricting the sample to male winners with above-median cognitive scores living in households with secondary education (Table 8, Columns 5 and 6). In this restricted sample, the discrepancy is 24.8–16.3 = 8.5 percentage points, compared to 31.3–10.4 = 20.9 percentage points in the full sample. The discrepancy is thus smaller but still is sizable among highly educated and cognitively able households.

To further explore the role of nonstandard beliefs, we also analyzed data from a survey fielded in the fall of 2016 to a subsample of lottery players (see Online 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 Fig. 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 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 assum-

[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, Hurd (2009) concludes that equity returns are sufficiently pessimistic for enough households to account for observed stock market nonparticipation. Additionally, Online 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. (2020), 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.

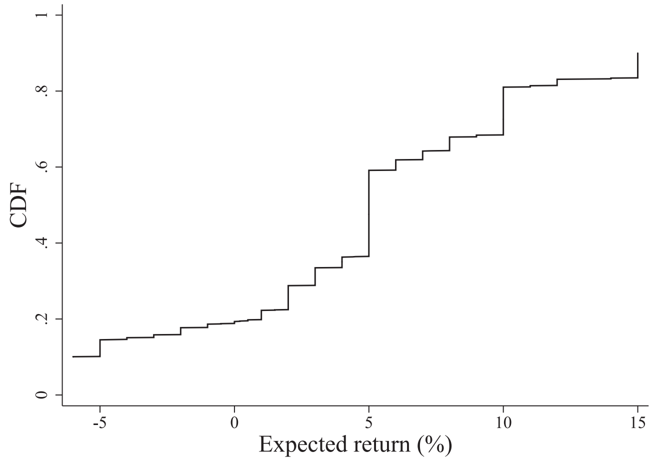


Fig. 8. Subjective distribution of equity returns. The above figure presents the CDF of survey respondents' expected market returns during the 12 months following the survey (i.e., fall 2016-fall 2017). For expositional purposes, we truncate the distribution at the 10th and 90th percentiles. The sample is composed of 1749 lottery winners that responded to the survey. See Online Appendix C for a details on survey methodology and sample details.

ing households have equity premium beliefs drawn from the distribution in Fig. 8. We then generate model predictions assuming parameter values equal to the estimates obtained from prelottery 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]

Our next set of analyses exploit time variation in equity returns to further test the hypothesis that nonstandard beliefs about future returns are an important explanation for the overprediction. We also seek to distinguish between two broad families of belief formation models that feature prominently in the literature: (1) models of experience-based learning in which individuals overweight personal experiences (see, e.g., Kaustia and Knüpfer, 2008; Malmendier and Nagel, 2011; Malmendier et al., 2020) and (2) models of natural expectations (see, e.g., Fuster et al., 2010; Fuster et al., 2012) or overextrapolation in which individuals overweight recent observations (see, e.g., Vissing-Jørgensen, 2003; Greenwood and Shleifer, 2014; Gennaioli et al., 2016; Bordalo et al., 2019). As explained below, both classes of theories are consistent with overweighting of recent returns, but only the experience-based models are consistent with overweighting of returns realized during an individual's formative years.

We first compare players who won in years following positive equity returns on the Swedish Stock Exchange to those who won following negative returns. Both belief formation theories generate recency bias in equity return beliefs (due to overweighting recent personal observations or recent realized returns) and therefore predict larger effects among households that win following positive returns. Table 9, Columns 1–2 confirm that effects are indeed larger among players who won the lottery following a year with positive returns (14 versus 5.3 percentage points), suggesting that individuals do overweight recent information when forming beliefs.

Motivated by research showing that equity returns experienced during formative years affect future equity return beliefs (see, e.g., Malmendier and Nagel, 2011; Giuliano and Spilimbergo, 2013; Fagereng et al., 2017; Malmendier et al., 2020), we next consider whether treatment effects vary by stock returns realized between ages 18 and 25. Table 9, Columns 3–4 show that effects are also larger among individuals with above-average returns during these ages (17.6 versus 8.6 percentage points), thus providing additional evidence that personal experiences affect beliefs. Columns 5 through 8 combine our first two tests and stratify jointly 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).

How pervasive are nonstandard 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 sep-

[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.

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 Eq. (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 at least one household member has a college degree. Hetero p is 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 4 for sample details. See Online Appendix Table B.13 for results for time $s = -1$ equity market participants.

	Recent returns		Early returns		Recent returns/Early returns			
	Low (1)	High (2)	Low (3)	High (4)	Low/Low (5)	Low/High (6)	High/Low (7)	High/High (8)
Effect	0.053	0.140	0.086	0.176	−0.007	0.138	0.111	0.187
SE	0.039	0.028	0.030	0.036	0.021	0.078	0.037	0.040
p	0.167	0.000	0.004	0.000	0.730	0.078	0.002	0.000
Hetero p	0.069		0.056					
N	10,402	8876	10,591	8687	5678	4724	4913	3963
% $Part_{-1}$	0.742	0.703	0.721	0.730	0.738	0.747	0.700	0.707

	Recent returns/College degree				Early returns/College degree			
	Low/No (9)	High/No (10)	Low/Yes (11)	High/Yes (12)	Low/No (13)	High/No (14)	Low/Yes (15)	High/Yes (16)
Effect	0.050	0.125	0.111	0.244	0.082	0.145	0.126	0.369
SE	0.040	0.030	0.107	0.062	0.032	0.037	0.070	0.085
p	0.213	0.000	0.296	0.000	0.011	0.000	0.060	0.000
Δ Effect	0.075		0.132		0.064		0.243	
N	9014	7496	1388	1380	9095	7415	1496	1272
% $Part_{-1}$	0.699	0.669	0.865	0.810	0.674	0.700	0.851	0.831

arately 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 versus 11.1 percentage points) as well as if they experienced high equity returns during formative years (36.9 versus 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 nonstandard 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 reports new and credible 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 these estimates. When the parameters of the model are estimated from observational data, the model predicts much larger rates of entry following a lottery windfall than we actually observe, and matching our quasi-experimental estimates thus requires implausibly large entry costs. Therefore, our estimates pose a challenge to standard modeling approaches.

We conduct a number of follow-up analyses to shed light on possible explanations for the discrepancy between the baseline model's predictions and the quasi-experimental estimates. While our analyses suggest that many factors are likely to contribute, several converging lines of evidence point to a major role for nonstandard beliefs and belief formation processes. We conservatively estimate that the discrepancy between the model and our estimates shrinks by 50% when the model is calibrated to match the subjective distribution of beliefs rather than historical equity premia. Additional analyses provide indirect evidence consistent with nonstandard belief formation processes. For example, higher rates of entry during periods of recent positive returns are consistent with belief formation processes that overweight recent returns.

Our analyses suggest that many households are currently deterred from entry because they hold overly pessimistic beliefs. It therefore seems plausible that many of these pessimistic households would choose to purchase stocks if their beliefs about future returns were better aligned with historical averages. Some of our results, including that returns early in life affect entry rates and a discrepancy between model and data persists even among highly educated households, suggest changing people's beliefs might be difficult. However, our study provides limited insight into the feasibility of actually orchestrating a realignment of beliefs (e.g., through education programs), although, if effective, such interventions would likely improve financial outcomes.

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

Finally, our paper highlights the insights obtainable from both causal and structural estimates (see, e.g., Kahn and Whited, 2017, Nakamura and Steinsson, 2017, and Lewbel, 2019 for recent surveys on this topic). The quasi-experimental evidence informs evaluations of the underlying theory by identifying areas where it works well and areas where it requires refinements. For example, our evidence suggests that structurally estimated cost parameters are dramatically underestimated by conventional approaches relying on observational data. Conversely, the structural model is needed to interpret and benchmark the causal findings and quantitatively explore various channels. Our research design thus demonstrates the methodological benefits of combining causal estimates and identification via economic theory.

Appendix A. Structural model details

A.1. Household decision problem

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

$$
\begin{aligned}
V_t(X_t, P_t, I_t, L_t, e, m) &= \max_{C_t, Part_t, \alpha_t} \Big\{ (1-\beta\pi_t)C_t^{1-1/\psi} \\
&\quad + \beta\mathbb{E}\big[\pi_t V_{t+1}(X_{t+1}, P_{t+1}, I_{t+1}, e, m)^{1-\rho} + (1-\pi_t) b(X_{t+1})^{1-1/\psi}\big]^{\frac{1-1/\psi}{1-\rho}} \Big\}^{\frac{1}{1-1/\psi}} \\
X_{t+1} &= \big[R_f + \alpha_t(R^s_{t+1} - R_f)\big](X_t - C_t) + Y_{t+1} - [(1-I_t)\times\chi + \kappa)]\times Part_t + L_t \\
0 &\le \alpha \le 1 \\
Y_t &= \begin{cases} exp(f(t, Z_t))P_t U_t & if \quad t \le t_R \\ \lambda_{e,m} Y_{t_R} & if \quad t > t_R \end{cases} \\
P_t &= P_{t-1}N_t \\
I_{t+1} &= (1-I_t)\times Part_t \\
\begin{pmatrix} r^s_t - r_f \\ log(N_t) \\ log(U_t) \end{pmatrix} &\sim \mathcal{N}\left[\begin{pmatrix} \mu_s \\ -\sigma_N^2/2 \\ -\sigma_U^2/2 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & \rho_{n,r}\times\sigma_n\sigma_s & 0 \\ \rho_{n,r}\times\sigma_n\sigma_s & \sigma_n^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}\right].
\end{aligned}
$$

A.2. Estimation and test statistics

We consider two test statistics to check for overidentifying restrictions and to evaluate the 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

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

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

$$g^\star(\nu_{i,s}, \theta) = \mathbb{E}\left[m^\star(\nu_i) - \frac{1}{K}\sum_{k=1}^{K} m^\star(\nu^k_{i,s}(\theta)) \right] = 0, \tag{A.2}$$

can be tested by a Wald statistic defined as

$$g^\star(\nu_{i,s}, \hat{\theta})' avar(g^\star(\nu_{i,s}, \hat{\theta}))^{-1} g^\star(\nu_{i,s}, \hat{\theta}) \to \chi^2_{|g^\star(\nu_{i,s},\hat{\theta})|} \tag{A.3}$$

$$avar(g^\star(\nu_{i,s}, \hat{\theta})) = \mathbb{E}\big[\phi^\star_g \phi^{\star\prime}_g\big]. \tag{A.4}$$

where $\phi^\star_g$ 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 backward 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; 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 can calculate the v_{T-1} value function. We then repeat this process and iterate backward 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 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 Fig. A.1.

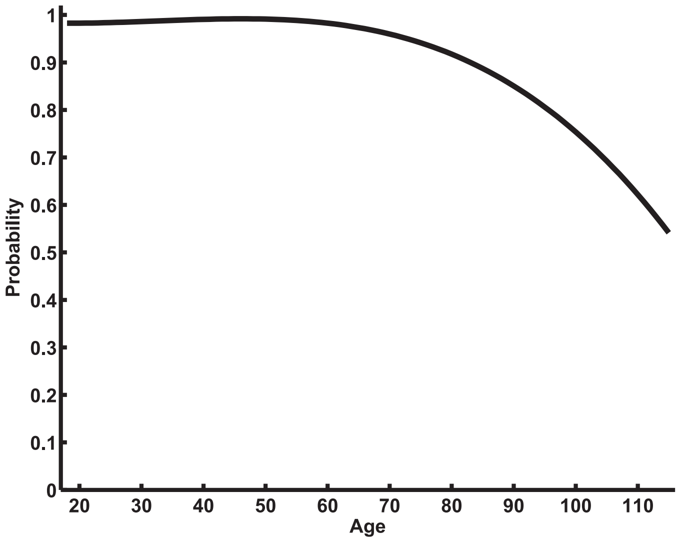


Fig. A.1. Survival probabilities. This figure presents the one-year survival probability for each age. Survival probabilities are calculated as the average observed 1998–1999 survival probabilities for a random 100,000 person sample from the Swedish population.

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

$$\epsilon^Y_{i,t} \equiv log(Y_{i,t}) - \hat{f}(t, m_i, e_i)$$

$$r_{i,d} \equiv \epsilon^Y_{i,t+d} - \epsilon^Y_{i,t},$$

then because

$$\mathrm{Var}(r_{i,d}) = d\sigma^2_N + 2 * \sigma_U,$$

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

To estimate the correlation between income and equity returns, note that $\epsilon^Y_{i,t}$ 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

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

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 $\bar{r}_{i,1,s,e}$ 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 (6500 USD). Second, an earning supplement is given by

$$0.6 \times AP \times BA,$$

where AP denotes pension points calculated from the 15 years with highest observed income calculated recursively by the following formula:

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

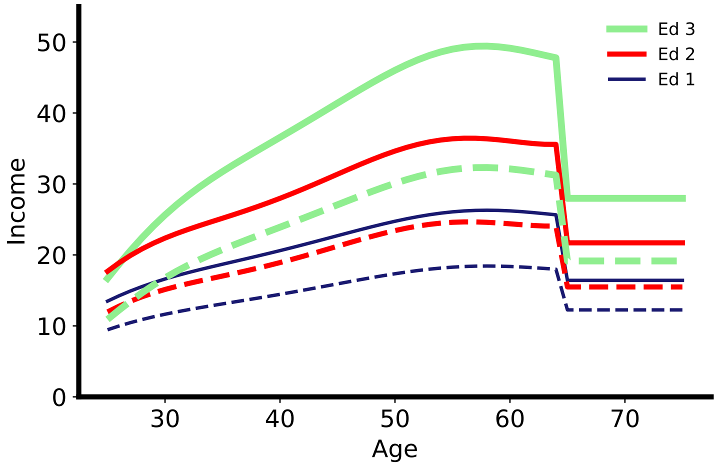


Fig. A.2. Average income profiles. This figure presents the deterministic income component $f(t, m, e)$. Solid lines reflect married households, while dashed lines reflect single households. Income profiles are estimated following the methodology in Cocco et al. (2005), which is summarized in Appendix A.5. The sample includes lottery winners in the 30 years (or as many as possible) prior to the lottery event. 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.

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

$0.6 \times AP \times BA + 0.96BA$

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 the following:

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

$$\begin{aligned} c_{i,s} &= b + b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_I I_{i,s} + \eta_{i,s}^c \\ Part_{i,s} &= b + b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_I I_{i,s} + \eta_{i,s}^{Part}. \end{aligned} \tag{A.5}$$

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

$$\begin{aligned} c_{i,s} &= b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_I I_{i,s} + b_l l_{i,s} \\ &\quad + MX_{i,0} + \eta_{i,s}^C \\ Part_{i,s} &= b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_I I_{i,s} + b_l l_{i,s} \\ &\quad + MX_{i,0} + \eta_{i,s}^{Part}. \end{aligned} \tag{A.6}$$

3. Postlottery 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 ($I_{i,s} = 0$):

$$\begin{aligned} c_{i,s} &= b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_l l_{i,s} + MX_{i,0} + \eta_{i,s}^c \\ Part_{i,s} &= b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + b_l l_{i,s} + MX_{i,0} + \eta_{i,s}^{Part}. \end{aligned} \tag{A.7}$$

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

$$\begin{aligned} Part_{i,s} =& b_t t_{i,s} + b_{t^2} t_{i,s}^2 + b_x x_{i,s} + b_{x^2} x_{i,s}^2 + \mathbb{1}_{l_{i,s} \in [1.5,15)} \\ &+ \mathbb{1}_{l_{i,s} \in [15,150)} + \mathbb{1}_{l_{i,s} \in [150,300)} + \mathbb{1}_{l_{i,s} \in [300,\infty)} \\ &+ MX_{i,0} + \eta_{i,s}^{Part}. \end{aligned} \tag{A.8}$$

Table A.1 presents the fits of our various estimation exercises. Our prelottery estimation (Table 6, Column 1), which targets only prelottery regressions (Table A.1, Panel A.i and Panel A.ii), is presented in Column 2. Our postlottery estimation (Table 6, Column 2), which targets only postlottery regressions (Table A.1, Panel B.i-iv), is presented in Column 3. Our pre/postlottery combined estimation (Table 6, Column 3), which targets only both prelottery and postlottery 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 postlottery 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 prelottery 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

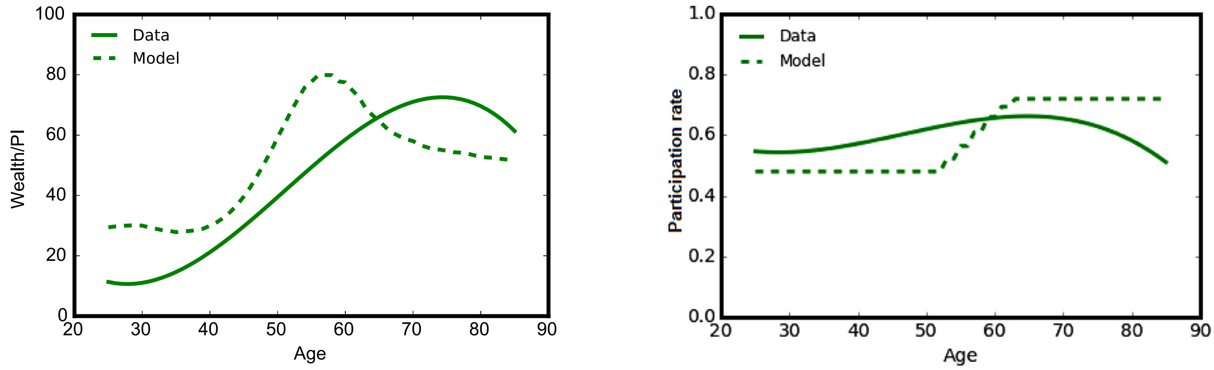
Table A.1
Structural estimation model fit.
This table presents the model fit for our various structural estimation exercises. Column 1 presents the empirical policy function estimates for prelottery observations in Panel A and postlottery observations in Panel B. Column 2 presents matched EPF coefficients when the model is estimated using only prelottery observations, Column 3 the matched EPF coefficients when the model is estimated using only postlottery observations, Column 4 the matched EPF coefficients when the model is estimated using only pre- and postlottery observations, Column 5 the matched EPF coefficients from estimating the entry cost distribution (Fig. 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 prelottery observations. All estimations use the post-1999 sample of lottery winners. See Table 4 for sample details.

	Estimate (1)	Prelottery (2)	Postlottery (3)	Pre- and post- (4)	Nonlinear (5)	Present bias (6)
A. Prelottery benchmarks						
i. Consumption						
Age	0.619	0.077		0.185		0.164
Age^2	−0.006	−0.001		−0.002		−0.002
Wealth/PI	0.164	0.203		0.156		0.173
$(Wealth/PI)^2$	0.000	0.000		0.000		0.000
$Part_{-1}$	−0.881	1.388		1.435		1.797
Constant	4.508	7.860		6.993		7.644
ii. Participation						
Age	0.000	0.000		0.000		0.003
Age^2	0.000	0.000		0.000		0.000
Wealth/PI	0.000	0.000		0.000		0.000
$(Wealth/PI)^2$	0.000	0.000		0.000		0.000
$Part_{-1}$	0.883	0.927		0.938		0.935
Constant	0.114	0.030		0.019		−0.013
B. Lottery benchmarks						
i. Consumption						
Age	0.614		0.250	0.223		
Age^2	−0.005		−0.003	−0.003		
Wealth/PI	0.039		0.139	0.155		
$(Wealth/PI)^2$	0.000		0.000	0.000		
$Part_{-1}$	−1.618		1.487	1.253		
Lottery	0.185		0.123	0.138		
ii. Participation						
Age	0.001		0.000	0.001		
Age^2	0.000		0.000	0.000		
Wealth/PI	0.000		0.000	0.000		
$(Wealth/PI)^2$	0.000		0.000	0.000		
$Part_{-1}$	0.796		0.993	0.933		
Lottery	0.028		0.030	0.067	0.029	
iii. Effect on consumption by prior participation status						
Lottery, Nonparticipants	0.239		0.121	0.136		
Lottery, Participants	0.166		0.124	0.138		
iv. Effect on participation by prior participation status						
Lottery/1M SEK, Nonparticipants	0.104		0.137	0.292	0.104	
Lottery/1M SEK, Participants	0.002		0.000	0.000	0.000	
v. Effect on participation by prize size (USD), nonparticipants						
$1.5K < L_i \leq 15K$	−0.012				0.006	
$15K < L_i \leq 150K$	0.078				0.080	
$150K < L_i \leq 300K$	0.156				0.158	
$300K < L_i$	0.359				0.357	
$N =$		192,524	70,139	262,663	70,139	192,524

Fagereng et al. (2017). Our estimation sample in this exercise consists of the matched population sample presented in Table 4, Column 2.

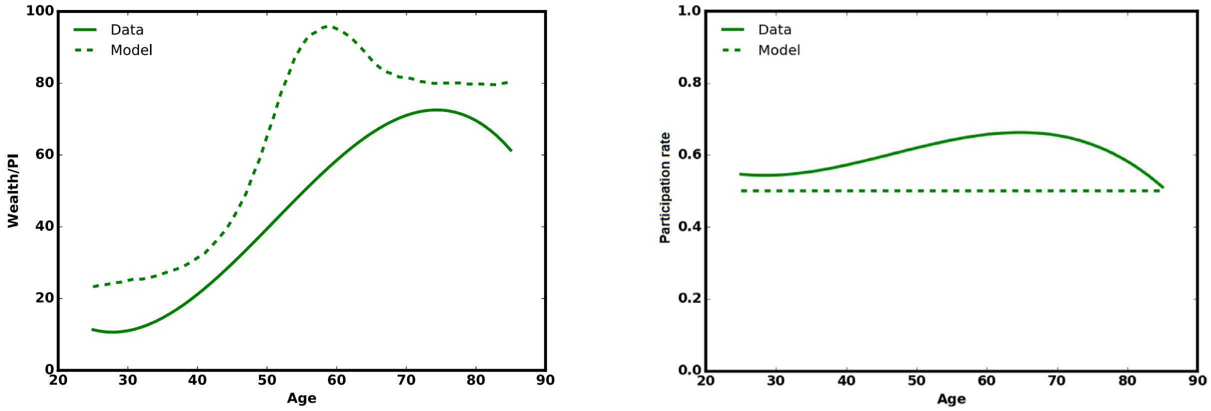
To estimate life cycle profiles of the wealth/income ratio, we run an OLS regression of the registry defined wealth/income ratio on age indicators, year indicators, and a proxy of cohort effects defined by the average returns on the Stockholm Stock Exchange experienced between ages 18–25. We then regress 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 solid line in Panel A of Figures A.3–A.6.

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 solid line in Panel B of Figs. A.3–A.6. Overall, our estimated wealth and participation profiles are similar to those obtained by Fagereng et al. (2017) for a representative Norwegian sample.



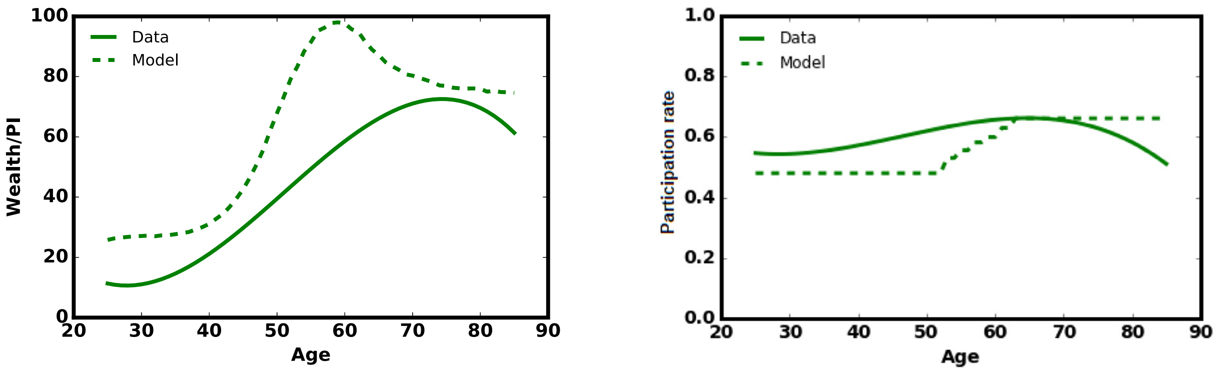
(A) Wealth/PI ratio (B) Participation

Fig. A.3. Life cycle profiles: prelottery 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 prelottery data (Table 6, Column 1). See Table 4 for sample details.



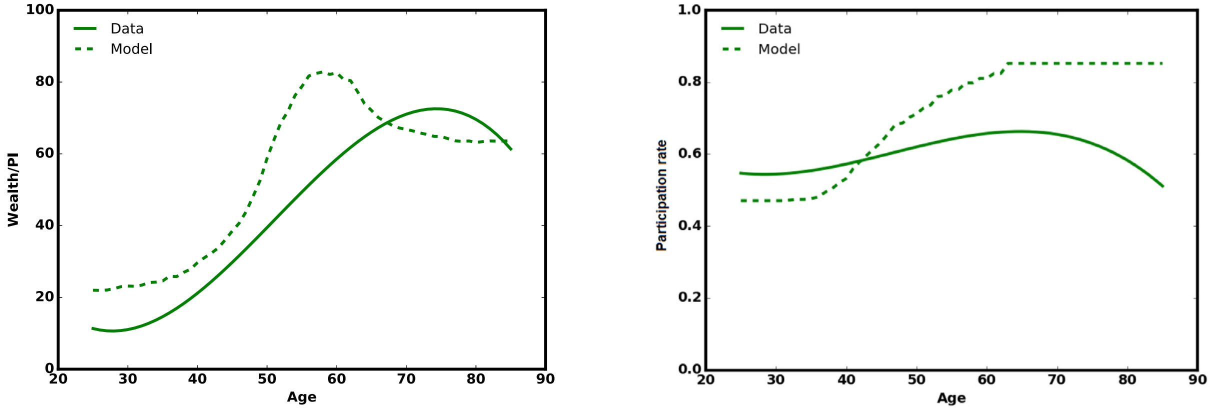
(A) Wealth/PI ratio (B) Participation

Fig. 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). See Table 4 for sample details.



(A) Wealth/PI ratio (B) Participation

Fig. A.5. Life cycle profiles: pre- and postlottery. 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 postlottery data (Table 6, Column 3). See Table 4 for sample details.



(A) Wealth/PI ratio

(B) Participation

Fig. A.6. Life cycle profiles: prelottery 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 $\beta=0.6$ and reestimating the model with prelottery data (Table 8, Column 4). See Table 4 for sample details.

To generate model-implied profiles, we draw a random sample of 10,000 Swedish households aged 18–25 between 1999 and 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 Figs. A.3–A.6.

Fig. A.3 presents results from our model using parameter estimates from our estimation with prelottery data (Table 6, Column 1). Fig. A.4 presents results from our model using parameter estimates from our estimation with postlottery data (Table 6, Column 2). Fig. A.5 presents results from our model using parameter estimates from our estimation with pre- and postlottery data (Table 6, Column 3). Fig. A.6 presents results from our model with quasihyperbolic discounting using parameter estimates from our estimation with pre- and postlottery data (Table 8, Column 2).

References

Alan, S., 2006. Entry costs and stock market participation over the life cycle. Rev. Econ. Dyn. 9 (4), 588–611.

Andersen, S., Nielsen, K.M., 2011. Participation constraints in the stock market: evidence from unexpected inheritance due to sudden death. Rev. Financ. Stud. 24 (5), 1667–1697.

Ang, A., Bekaert, G., Liu, J., 2005. Why stocks may disappoint. J. Financ. Econ. 76 (3), 471–508.

Angeletos, G.-M., Laibson, D., Repetto, A., Tobacman, J., Weinberg, S., 2001. The hyperbolic consumption model: calibration, simulation, and empirical evaluation. J. Econ. Perspect. 15 (3), 47–68.

Attanasio, O.P., Paiella, M., 2011. Intertemporal consumption choices, transaction costs and limited participation in financial markets: reconciling data and theory. J. Appl. Econom. 26 (2), 322–343.

Barberis, N., Huang, M., Thaler, R.H., 2006. Individual preferences, monetary gambles, and stock market participation: a case for narrow framing. Am. Econ. Rev. 96 (4), 1069–1090.

Barillas, F., Fernández-Villaverde, J., 2007. A generalization of the endogenous grid method. J. Econ. Dyn. Control 31 (8), 2698–2712.

Bazdresch, S., Kahn, R.J., Whited, T.M., 2017. Estimating and testing dynamic corporate finance models. Rev. Financ. Stud. 31 (1), 322–361.

Becker, T.A., Shabani, R., 2010. Outstanding debt and the household portfolio. Rev. Financ. Stud. 23 (7), 2900–2934.

Benjamin, D., 2019. Errors in probabilistic reasoning and judgment biase. In: Bernheim, D.B., DellaVigna, S., Laibson, D. (Eds.), Handbook of Behavioral Economics-Foundations and Applications 2. Elsevier, Amsterdam, pp. 69–186.

Benzoni, L., Collin-Dufresne, P., Goldstein, R.S., 2007. Portfolio choice over the life-cycle when the stock and labor markets are cointegrated. J. Financ. 62 (5), 2123–2167.

Berkelaar, A.B., Kouwenberg, R., Post, T., 2004. Optimal portfolio choice under loss aversion. Rev. Econ. Stat. 86 (4), 973–987.

Bordalo, P., Gennaioli, N., La Porta, R., Shleifer, A., 2019. Diagnostic expectations and stock returns. J. Financ. 74, 2839–2874.

Brunnermeier, M.K., Nagel, S., 2008. Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals' asset allocation. Am. Econ. Rev. 98 (3), 713–736.

Calvet, L., Campbell, J.Y., Sodini, P., 2009. Fight or flight: portfolio rebalancing by individual investors. Q. J. Econ. 124 (1), 301–348.

Calvet, L., Célérier, C., Sodini, P., Vallée, B., 2017. Can Financial Innovation Solve Household Reluctance to Take Risk?. EDHEC Business School, Stockholm School of Economics, University of Toronto, Harvard Business School. Unpublished working paper

Calvet, L.E., Campbell, J.Y., Sodini, P., 2007. Down or out: assessing the welfare costs of household investment mistakes. J. Polit. Econ. 115 (5), 707–747.

Calvet, L.E., Sodini, P., 2014. Twin picks: disentangling the determinants of risk-taking in household portfolios. J. Financ. 69 (2), 867–906.

Campbell, J.Y., 2006. Household finance. J. Financ. 61 (4), 1553–1604.

Carroll, C.D., 1997. Buffer-stock saving and the life cycle/permanent income hypothesis. Q. J. Econ. 112 (1), 1–55.

Carroll, C.D., 2006. The method of endogenous gridpoints for solving dynamic stochastic optimization problems. Econ. Lett. 91 (3), 312–320.

Carroll, C.D., Samwick, A.A., 1997. The nature of precautionary wealth. J. Monet. Econ. 40 (1), 41–71.

Cesarini, D., Lindqvist, E., Östling, R., Wallace, B., 2016. Wealth, health, and child development: evidence from administrative data on Swedish lottery players. Q. J. Econ. 131 (2), 687–738.

Cocco, J.F., 2005. Portfolio choice in the presence of housing. Rev. Financ. Stud. 18 (2), 535–567.

Cocco, J.F., Gomes, F.J., Maenhout, P.J., 2005. Consumption and portfolio choice over the life cycle. Rev. Financ. Stud. 18 (2), 491–533.

Cooper, R., Zhu, G., 2016. Household finance over the life-cycle: what does education contribute? Rev. Econ. Dyn. 20, 63–89.

Das, S., Kuhnen, C.M., Nagel, S., 2020. Socioeconomic status and macroeconomic expectations. Rev. Financ. Stud. 33 (1), 395–432.

Davis, S.J., Kubler, F., Willen, P., 2006. Borrowing costs and the demand for equity over the life cycle. Rev. Econ. Stat. 88 (2), 348–362.

Della Vigna, S., 2009. Psychology and economics: evidence from the field. J. Econ. Lit. 47 (2), 315–372.

Dimmock, S.G., Kouwenberg, R., 2010. Loss-aversion and household portfolio choice. J. Empir. Financ. 17 (3), 441–459.

Dominitz, J., Manski, C.F., 2011. Measuring and interpreting expectations of equity returns. J. Appl. Econom. 26 (3), 352–370.

Epstein, L.G., Zin, S., 1991. Substitution, risk aversion and the temporal behavior of consumption and asset returns. J. Polit. Econ. 99, 263–286.
Erickson, T., Whited, T.M., 2002. Two-step GMM estimation of the errors-in-variables model using high-order moments. Econom. Theory 18 (3), 776–799.
Fagereng, A., Gottlieb, C., Guiso, L., 2017. Asset market participation and portfolio choice over the life-cycle. J. Finance 72 (2), 705–750.
Flavin, M., Yamashita, T., 2011. Owner-occupied housing: life-cycle implications for the household portfolio. Am. Econ. Rev. 101 (3), 609–614.
Fuster, A., Hebert, B., Laibson, D., 2012. Natural expectations, macroeconomic dynamics, and asset pricing. NBER Macroeconom. Annu. 26 (1), 1–48.
Fuster, A., Laibson, D., Mendel, B., 2010. Natural expectations and macroeconomic fluctuations. J. Econ. Perspect. 24 (4), 67–84.
Gennaioli, N., Ma, Y., Shleifer, A., 2016. Expectations and investment. NBER Macroeconom. Annu. 30 (1), 379–431.
Giuliano, P., Spilimbergo, A., 2013. Growing up in a recession. Rev. Econ. Stud. 81 (2), 787–817.
Gomes, F., Michaelides, A., 2005. Optimal life-cycle asset allocation: Understanding the empirical evidence. J. Financ. 60 (2), 869–904.
Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. J. Appl. Econ. 8 (S1), S85–S118.
Gourinchas, P.-O., Parker, J.A., 2002. Consumption over the life cycle. Econometrica 70 (1), 47–89.
Greenwood, R., Shleifer, A., 2014. Expectations of returns and expected returns. Rev. Financ. Stud. 27 (3), 714–746.
Guiso, L., Haliassos, M., Jappelli, T., 2002. Household Portfolios. MIT Press, Cambridge.
Guiso, L., Sapienza, P., Zingales, L., 2004. The role of social capital in financial development. Am. Econ. Rev. 94 (3), 526–556.
Guiso, L., Sapienza, P., Zingales, L., 2008. Trusting the stock market. J. Financ. 63 (6), 2557–2600.
Guiso, L., Sodini, P., 2013. Household finance: an emerging field. Handb. Econ. Finance Part B, 1397–1532.
Haliassos, M., Bertaut, C.C., 1995. Why do so few hold stocks? Econ. J. 105 (432), 1110–1129.
Hankins, S., Hoestra, M., Skiba, P.M., 2011. The ticket to easy street? The financial consequences of winning the lottery. Rev. Econ. Stat. 93 (3), 961–969.
Heaton, J., Lucas, D., 2000. Portfolio choice and asset prices: the importance of entrepreneurial risk. J. Financ. 55 (3), 1163–1198.
Hurd, M.D., 2009. Subjective probabilities in household surveys. Annu. Rev. Econ. 1 (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. Am. Econ. Rev. 91 (4), 778–794.
Kahn, R., Whited, T.M., 2017. Identification is not causality, and vice versa. Rev. Corp. Financ. Stud. 7 (1), 1–21.
Kaustia, M., Knüpfer, S., 2008. Do investors overweight personal experience? Evidence from IPO subscriptions. J. Financ. 63 (6), 2679–2702.
Kézdi, G., Willis, R. J., 2011. Household Stock Market Beliefs and Learning. Unpublished working paper. National Bureau of Economic Research.
Khorunzhina, N., 2013. Structural estimation of stock market participation costs. J. Econ. Dyn. Control 37 (12), 2928–2942.
Kuhnen, C.M., Miu, A.C., 2017. Socioeconomic status and learning from financial information. J. Financ. Econ. 124 (2), 349–372.
Laibson, D., 1997. Golden eggs and hyperbolic discounting. Q. J. Econ. 112, 443–477.
Laun, T., Wallenius, J., 2015. A life cycle model of health and retirement: the case of Swedish pension reform. J. Publ. Econ. 127, 127–136.
Lewbel, A., 2019. The identification zoo: meanings of identification in econometrics. J. Econ. Lit. 57 (4), 835–903.
Love, D., Phelan, G., 2015. Hyperbolic discounting and life-cycle portfolio choice. J. Pension Econ. Financ. 14 (4), 492–524.
Malmendier, U., Nagel, S., 2011. Depression babies: do macroeconomic experiences affect risk taking? Q. J. Econ. 126, 373–416.
Malmendier, U., Pouzo, D., Vanasco, V., 2020. Investor experiences and financial market dynamics. J. Financ. Econ. 136 (3), 597–622.
Merton, R.C., 1971. Optimum consumption and portfolio rules in a continuous-time model. J. Econ. Theory 3 (4), 373–413.
Mulligan, C.B., Sala-i Martin, X., 2000. Extensive margins and the demand for money at low interest rates. J. Polit. Econ. 108 (5), 961–991.
Nakamura, E., Steinsson, J., 2017. Identification in Macroeconomics. Unpublished working paper. National Bureau of Economic Research.
Nekoei, A., Seim, D., 2018. How do Inheritances Shape Wealth Inequality? Theory and Evidence from Sweden. Unpublished working paper. Centre for Economic Policy Research.
Paiella, M., 2007. The forgone gains of incomplete portfolios. Rev. Financ. Stud. 20 (5), 1623–1646.
Samuelson, P.A., 1969. Lifetime portfolio selection by dynamic stochastic programming. Rev. Econ. Stat. 51 (3), 239–246.
Samuelson, W., Zeckhauser, R., 1988. Status quo bias in decision making. J. Risk Uncertain. 1 (1), 7–59.
Smith, A.A., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. J. Appl. Econom. 8 (S1), 563–584.
Tversky, A., Kahneman, D., 1986. Rational choice and the framing of decisions. J. Bus. 59, S251–S278.
Vestman, R., 2018. Limited stock market participation among renters and home owners. Rev. Financ. Stud. 32 (4), 1494–1535.
Vissing-Jørgensen, A., 2002. Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures. Unpublished working paper. National Bureau of Economic Research.
Vissing-Jørgensen, A., 2003. Perspectives on behavioral finance: does irrationality disappear with wealth? Evidence from expectations and actions. NBER Macroecon. Annu. 18, 139–208.
Waldenström, D., 2014. Swedish Stock and Bond Returns, 1856–2012. Unpublished working paper. Research Institute of Industrial Economics.