Wealth and Stock Market Participation: Estimating the Causal Effect From Swedish Lotteries *

Joseph Briggs† David Cesarini‡ Erik Lindqvist § Robert Östling¶

Preliminary
January 16, 2015

Abstract

In this paper we estimate the causal effect of wealth on stock market participation. The positive cross-sectional relationship between participation and wealth is well-established, with previous work suggesting that moderate costs of stock market participation are capable of rationalizing the decision of most non-participants. In our study we use a large sample of Swedish lottery players whom were randomly assigned over 1 billion USD, linked to administrative tax records of asset holdings, to precisely identify both the effect of wealth and the costs necessary to explain non-participation. Although we estimate a positive effect of wealth on participation, our estimate is much smaller than that implied by the cross-section. Furthermore, our estimates of participation costs are 10-20 times higher than those proposed in previous studies. We interpret these results within a structural model of life-cycle stock market participation, and use participation responses following random wealth assignment to estimate entry and participation costs conditional on a variety of demographic and individual characteristics. We conclude that it is unlikely that fixed financial costs are credible explainations for equity market non-participation.

1 Introduction

Canonical life-cycle models of consumption and savings make the strong prediction that all individuals, irrespective of their degree of risk aversion, should invest some non-zero fraction of their wealth in stocks (Samuelson (1969); Merton (1971)). Because this prediction is not borne out empirically – a substantial fraction of household do not own stock directly or through mutual funds (Friend and Blume (1975); King and Leape (1984); Mankiw and Zeldes (1991)) – a large literature in household finance formulates and tests hypotheses about the causes of the "non-participation puzzle" (Haliassos and

*This research was financially supported by NBER Household Finance working group research grant number 22-2382-13-1-33-003.

†Corresponding author. Joseph Briggs, Department of Economics, New York University, Research Institute for Industrial Economics. Email: jsb493@nyu.edu.

‡David Cesarini, Department of Economics, New York University, Research Institute for Industrial Economics. Email: david.cesarini@nyu.edu

§Erik Lindqvist, Stockholm School of Economics, Research Institute for Industrial Economics. Email: erik.lindqvist@hhs.se

¶Robert Östling, Institute for International Economic Studies. Email: robert.ostling@iies.su.se

Bertaut (1995)). Insights into the causes of the non-participation puzzle may lead to the formulation of better models of household finance and may also help guide efforts to promote more effective financial decision making by households (Campbell (2006)).

To account for the high rates of non-participation, most models introduce a fixed cost of stock ownership, either a one-time entry cost or an ongoing, per-period, participation cost. Because the potential gains increase with greater wealth, whereas the costs do not, these models predict that a household will enter the stock market if their wealth exceeds an endogenously determined threshold. The models therefore predict that a wealth shock should increase the probability of participation, and thus provide an intuitive explaination for the robustly established positive cross-sectional correlation between wealth and participation (Mankiw and Zeldes (1991); Poterba and Samwick (2003); Calvet, Campbell, and Sodini (2007)). In an important paper, Vissing-Jorgensen (2002) showed that for all but the richest US households, a very modest per-period for most of the observed non-participation. Several papers have sought to get inside the black box of "participation costs" by examining how participation is related to an individual's psychological characteristics and cognitive resources van Rooij, Lusardi, and Alessie (2011); Grinblatt, Keloharju, and Linnainmaa (2011a), which could influence participation through their impact on attention, the costs of information acquisition, or beliefs about the benefits of stock market participation. Clearly, the extent to which it is desirable and feasible to influence households' participation decisions depends critically on the exact nature of the participation costs.

In this paper we estimate the causal effect of wealth on stock market participation by exploiting the randomized assignment of wealth in three Swedish samples of lottery players who have been matched to administrative records with high-quality information about financial portfolios. Theories of household finance make predictions about the impact of a windfall gain on subsequent stock market participation and estimates of the causal impact of wealth on participation are therefore useful for testing and refining theories of household finance. A fundamental challenge when estimating the effect of wealth is that it is not feasible to randomly assign substantial amounts of wealth to individuals. As a result, researchers interested in the relationship between wealth and participation are usually confined to studying observational data Calvet and Sodini (2014); Brunnermeier and Nagel (2008); Calvet, Campbell, and Sodini (2009); Alan (2006); Khorunzhina (2013); Fagereng, Gottlieb, and Guiso (2013), where the possibility of omitted variable bias and reverse causation looms large (though see Andersen and Nielsen (2011) for an example of a quasi-experimental study exploiting unexpected parental deaths as a plausibly exogenous source of wealth variation). The overall conclusion from this literature is that changes in wealth are associated with a greater likelihood of participation, but the magnitudes of the estimated effects vary. A second branch of the literature Vissing-Jorgensen (2002); Gomes and Michaelides (2005); Alan (2006); Khorunzhina (2013) uses structural models to identify the magnitudes and patterns of such costs, with most finding that moderate financial costs are capable of accounting for the majority of non-participation.

Our paper contributes to both of these literatures. In our reduced form analyses, we report the results of a comprehensive set of analyses examining how wealth impacts participation. Our sample satisfies a number of methodological desiderata that allow us to make stronger inferences about the effect of wealth on participation than in previous work. First, we observe the factors (such as number of tickets owned) conditional on which the lottery wealth is randomly assigned. As a result, we can be uniquely confident that our estimates have a causal interpretation. Second, because the size of the prize pool is almost one billion dollars, 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. Finally, because our lottery and financial data are drawn from administrative records, our sample is virtually free from attrition, and any sample selection biases should be negligibly small.

In our structural analyses, we use the exogenous wealth variation to estimate a structural model of portfolio choice over the lifecycle. The lottery data provide us with an additional source of variation that can be leveraged to obtain more credible estimates. To convey the intuition behind the principal result in our structural analyses, consider again the result that a modest per-period participation cost can explain the bulk of non-participation of US households. The theory used to generate this conclusion makes quantifiable predictions about the impact of a substantial wealth shock on subsequent participation and it is natural to test these predictions using our causal estimates as a benchmark. When we make this comparison, we find that the observed participation response is much smaller than the theoretical prediction: participation costs twenty times larger than those reported by Vissing-Jorgensen (2002) are required to account for the relatively modest response we observe. Our basic conclusion – that substantial entry costs are required to account for non-participation in a model identified using the exogenous wealth variation – holds also in a more richly specified lifecycle model. The finding shows up robustly and does not appear to be an artifact of a highly non-representative sample. When we instead estimate the model using the cross-sectional wealth, we obtain a distribution of fixed costs that resembles those in the previous literature Vissing-Jorgensen (2002); Gomes and Michaelides (2005); Alan (2006); Khorunzhina (2013).

The remainder of the paper is structured as follows. Section 2 describes how we constructed our sample of lottery players by matching administrative data on participants in three lotteries to Statistics Sweden's register data on wealth.[1] In describing our lottery samples, we address several important issues about external validity that are often raised about studies of lottery players. Importantly, Section 2.3 lays out our basic identification strategy, and Section 2.6 sets the stage by showing participation costs implied by the cross-section. In Section 3 we report the results from our reduced form analyses. Section 4 presents a structural model of life-cycle asset market participation and estimates participation costs implied by our variation in wealth. In Section 5 extends this model using reduced form heterogeneity analyses to allow for and estimate heterogeneous costs of entry and participation. Section 6 concludes with a discussion about what our findings imply about the magnitude, nature and heterogeneity of participation costs.

[1]The usefulness of this wealth panel has been previously demonstrated in the in a set of influential papers (Calvet, Campbell, and Sodini (2007, 2009); Calvet and Sodini (2014)). It is fortuitous that we are able to link these two data sets, and doing so permits a unique data set to study the effect of wealth on equity market participation.

2 Data and Identification Strategy

Our analyses are based on three samples of lottery players who have been matched, using personal identification numbers (PINs) to administrative records covering the entire Swedish population. Below, we begin brief description of the register variables that play a key role in our analyses. We draw primarily on high-quality information about year-end financial portfolios (assets and debt) which are available 1999–2007 (the "study period"). We next turn to a description of the lottery data. Our basic 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 wealth is randomly assigned. Because the construction of the cells varies by lottery, we discuss each of the three lotteries.

2.1 Administrative Data

2.1.1 Household Financial Data

Until 2007, household wealth was taxable under Swedish tax law. To implement this tax, Statistics Sweden collected information from other branches of government, as well as banks and other financial institutions. A register known as the *Swedish Wealth Registry* contains detailed information about the year-end financial portfolios of the entire Swedish population during the study period. The register contains individual-level variables measuring bank account balances, mutual funds, directly held stocks, bonds, money market funds, debt, residential and commercial real estate, and other financial and real assets. Using these variables, it is straightforward to define several outcome variables that we will use for the remainder of this paper, including our measures of stock market participation. Although originally, collected for tax purposes, the high quality of the records has made them extremely useful resource for researchers, as exemplified by several recent and influential studies(Calvet, Campbell, and Sodini (2007, 2009); Calvet and Sodini (2014)).

2.1.2 Other Covariates

Several of other analyses also make use of the a rich set of demographic covariates available in Statistics Sweden's *Integrated Database for Labour Market Research*, which contains annual information (1990-2010) about a number of demographic characteristics which include income, employment, educational attainment, region of residence, retirement status and household composition.

2.1.3 Household Definition

A prior it is difficult to determine whether the appropriate unit of analysis is a household or an individual. The answer depends on the extent to which it is reasonable to assume that the adults in a household make joint financial decisions. In our data, the wealth of a winning player's spouse or partner increases by about 20% of the total prize amount in the year of the win, often because the prize money won is deposited into a joint account. In the main analyses that follow, we therefore make the household the unit of analysis. A household always comprises one or two adults. Following Statistics Sweden, we say that two adults form a household if they are either married or cohabiting with an individual with whom they have at are either married or cohabiting with an individual with whom they have at least one child. All other adults are treated as one-person households. In one of our sensitivity checks, we show that our main conclusions are substantively identical if we instead restrict the sample to individuals who

2.2 Lottery Data

Our final estimation sample was constructed from three samples of lottery players. The first is a monthly Swedish subscription lottery called Kombilotteriet ("Kombi"). Our second sample, Triss, contains of scratch lottery players who qualified for a TV show where they could win substantial amounts of money. Our final sample is a panel of individuals with prize-linked savings (PLS) accounts. PLS accounts are savings accounts which, instead of just paying interest, also incorporate a lottery element by enrolling account holders in lotteries Kearney, Tufano, Guryan, and Hurst (2008).

Throughout, we restrict attention to lottery prizes won during the sample period (1999-2007). In Kombi and Triss, we have data on draws conducted throughout the entire sample period, whereas the PLS data are only available until 2003. We describe all three samples briefly below, and refer the reader to Cesarini, Lindqvist, Östling, and Wallace (2013) for a richer description.

Kombi

Kombi is a monthly subscription lottery whose proceeds are given to the Swedish Social Democratic Party, by far the most dominant political force in Sweden during the post-war era. Subscribers choose their desired number of subscription tickets and are billed monthly usually by direct debit. Using the data provided to us by Kombi, we constructed an unbalanced panel covering our entire sample period. For each draw, the panel contains has one entry per eligible participant, and lists the players' PIN, number of tickets purchased and the prize amount won (for all prizes exceeding 1M SEK, net of taxes). For a small number of individuals ($\sim 1\%$) the PIN is missing and we do not include these individuals when constructing our final estimation sample.[2]

In each draw, every purchased ticket is assigned a unique number by Kombi, and the winning tickets are then drawn randomly from the set of purchased tickets. Therefore, two individuals (or households) who purchased the same number of tickets in a given draw face the exact same probability of winning the large prize. Because our main analyses are conducted at the household level, our empirical strategy is to compare each household winning a large prize with "matched control households" who did not win a large prize but who purchased exactly the same number of tickets in the month of the draw. To construct the cells, we began by computing the number of tickets owned by the household of each winning player. We then matched each large-prize winner to (up to) 100 non-winning households who did not win a large prize in the month of the draw. When more than 100 controls are available,

[2] Because missingness is determined entirely by whether the participant supplies a valid PIN at enrollment, this restriction does not introduce any sample selection biases that would jeopardize the interpretation of our parameter estimates as causal. The restriction does change the composition of the sample for which we are estimating the treatment effect.

we choose the controls most similar in household size, sex, age and marital status.[3] The Kombi data contain XXX large prizes won during our sample period: we were able to match 297 to the administrative registers, leaving us with 297 prizes and a matched estimation sample of 29,524.

Triss Sample

Our second sample is called Triss, a scratch-ticket lottery run since 1986 by Svenska Spel, the Swedish government-owned gambling company. One (of many prizes) participants can win is the opportunity to participate in a TV show (TV-Triss) where they can win a substantial lump-sum prizes. Each month, around 25 TV-Triss prizes are awarded on television.

At the show, participant draws a prize from a stack of tickets. This stack is determined by a public prize plan that is subject to occasional revision. Because the tickets in the stack are shuffled and look identical, the prize won by the participant in the show is random conditional on the prize plan. Tv-Triss Prizes are paid out as a lump-sum and vary in size from 50,000 SEK to 5 million SEK (net of taxes).

Svenska Spel supplied us with information about all individuals who participated in the TV show between 1999 and 2010. With the help of Statistics Sweden, we were able to to use the information in the spreadsheet (name, age, region of residence, and often also the names of close relatives), to reliably identify the PINs of 98.7% of show participants. In the Online Appendix, we provide a detailed account of the processing of the data. We also report the results from several quality controls. For example, we ordered 500 randomly drawn TV-tapings (each show featuring one draw) of the show from the Swedish Television Archives and verified that the overlap between the information reported on TV and the spreadsheet is almost perfect. The spreadsheet also notes any instances where the participant shared ownership of the ticket. Our analyses below are based exclusively on participants who did not indicate that they shared ownership of the winning ticket, but our main results do not change appreciably with these individuals in the sample.

Our empirical strategy makes use of the fact that, conditional on the prize plan and winning exactly one prize, the nominal prize won is plausibly independent of pre-determined characteristics. To account for small changes in the real value of the prizes induced by inflation, we further restrict our comparison to individuals who won in the same year. Thus, our empirical strategy is to exploit the prize variation between individuals who won in the same "draw", where we define each unique combination of year and prize plan as a separate draw. If the members of a household win more than one prize in any given draw. In principle, households winning two prizes could be compare to other households who won two prizes in the same draw, but multiple wins are so rare that it is never possible to identify a successful match.

We begin with a sample of X,XXX prizes won during our sample period. We drop prizes won by individuals whose PIN could not be reliably identified and players whose tickets were jointly owned. The final sample comprises X,XXX won by 1852 households during our sample period.

[3]We match on age and gender in order to reduce the amount of noise due to random differences in demographic characteristics between winners and non-winners. The exact matching procedure is described in the Online Appendix.

PLS Sample

A PLS account is a savings account whose owner is enrolled in regular lotteries with monetary prizes (Kearney, Tufano, Guryan, and Hurst (2010)) paid in addition to (sometimes in lieu of) interest payments. Such accounts have existed in Sweden since the 1950s (Regeringen, 1972). The subsidies ceased in 1985, at which point the government authorized banks to offer prize-linked-savings products. Two systems were put into place. The savings banks (Sparbankerna) started offering their clients PLS-products through a system known as the Million Accounts ("Miljonkontot"), whereas the remaining banks joined forces and offered a PLS product known as Winner Accounts ("Vinnarkontot"). Approximately one in two Swedes held a PLS account.

During the period we study, PLS account holders could win two types of prizes: odds prizes and fixed prizes. The probability of winning either type of prize was proportional to the account balance (an account holder got one lottery ticket per 100 SEK in the account). Fixed prizes were prizes whose magnitude was not determined by the account balance of the winning account. The size of the odds prizes, on the other hand, depended on account balance. In each draw, an account's balance was proportional to the number of lottery tickets assigned to the account. An overwhelming fraction of the prizes awarded were small fixed prizes (typically 1000 SEK), but about XX% of the total sum of prizes awarded came from large prizes (100K SEK or more).

Our final lottery sample is obtained by combining data from two sources of information about the Winner Accounts: a set of printed lists with information about all prizes won 1979–2003 and microfiche images with information (account number, account owner's PIN, and number of tickets purchased) on all accounts in the draws between December 1986 and December 1994 (the "fiche period"). Both sources were retrieved from the Swedish National Archives.

Our final estimation sample is restricted to fixed prizes won during the sample period. It was constructed in three steps. In a first step, we digitized all the information on the prize lists pertaining to our sample period (1999-2003). For each draw, these list all winning accounts (account number) and prize(s) won (type of prize, prize amount). Because the prize lists do not contain information about the account balances (and hence number of tickets) of the accounts, the Kombi strategy of matching is not feasible. Moreover, whereas the fiches list the account number and the PIN of each account owner, the prize lists only list account numbers. There it is only possible to map each account number of prize-winning account during our sample period to a PIN if the account was open for at least part of the fiche period. In a second step, we therefore dropped prizes won by individuals whose PINs could not be identified.

Our identification strategy instead exploits the fact that in the population of households who won exactly n fixed prizes in a particular draw, the total sum of fixed prizes won is independent of account balances (and all other predetermined characteristics, including the date on which the account was opened). For each draw, we therefore assign two winning households to the same cell if they won an identical number of fixed prizes in that draw. This strategy is similar to that used by Imbens, Rubin, and Sacerdote (2001), Hankins, Hoestra, and Skiba (2011), and Hankins and Hoestra (2011) but unfortunately is only valid for fixed prizes. In a third step, we therefore drop odds-prizes from the sample period. In the Online Appendix, we show why our identifying assumption is valid for the subsample of fixed-prize-winning accounts that were in existence during the fiche period.

During the subperiod of our study period with PLS prizes (1999-2003) there were a total of XXXX fixed prizes. We were able to identify and match to the registers the PINs of 37066 winning households.

2.3 Identification Strategy

Our identification strategy thus uses the available data and knowledge about the institutional details of each lottery to define subsamples/cells within which wealth is assigned independently of potential outcomes. Table 1 summarizes the previous section's discussion of how these cells are constructed in each of the three lotteries. Normalizing the time of the lottery to $s = 0$, our main estimating equation is given by,

$$Y_{i,s} = \beta_{1,s} \times L_{i,0} + \mathbf{Z}_{i,-1} \times \gamma_s + \mathbf{X}_i \times M_s + \eta_{i,s} \tag{1}$$

where i indexes households, $L_{i,0}$ denotes the prize amount won (in 2010 SEK), X_i is a vector of cell fixed effects and Z_i is a vector controls observed the year before the lottery. The key identifying assumption needed for β_1 to have a causal interpretation is that the prize amount won is independent of $\eta_{i,s}$ conditional on the cell fixed effects. In practice, we often control for several additional characteristics ($\mathbf{Z}_{i,-1}$) which are always measured in the year before the lottery. These controls are included to absorb more variance of the residual and hence improve the precision of our estimates. We estimate separate equations for participation in the year of the lottery ($s = 0$) and each of the five years ($s = 1, .., 5$) after the lottery.

To get a better sense of the source of our identifying variation, Table 2 provides basic information about the distribution of prizes won by the players in our samples. For each lottery, and the pooled sample, the table shows the total number of small (<100,000 SEK), medium (<1,000,000 SEK) and large prizes (>1,000,000 SEK). To put these prize amounts in perspective, the median annual after-tax earnings of a Swede working full time was roughly 170,000 SEK, in 2010 prices, in 1998 (the first year of the sample period). A first important message from Table 2 is that even though the number of prizes won vary dramatically across the lotteries, all lotteries contribute substantial identifying variation to our study. The total value of the after-tax prize money disbursed to the winners in our samples is almost 3.5 billion SEK (about 500 million dollars) – two orders of magnitude larger than previous studies of lottery players (e.g., Ettner (1996); Lindahl (2005); Gardner and Oswald (2007); Apouey and Clark (2013)) .

A second important message is that effects we report in the paper therefore assign relatively little weight to the marginal effects of small lottery prizes, even though these account for a large fraction of the number of prizes won. The reason is that even though a large number of prizes are small, they account for only a modest fraction of identifying variation. For example, dropping all the prizes below 10,000 SEK from the sample reduces the total amount of treatment variation by 19%. In Kombi, all of the identifying variation comes from comparisons of players who win large prizes to players who did not win a prize. In Triss, most identifying variation comes from comparisons of winners of large prizes to winners of small or modest prizes. Finally, in PLS, virtually all of the identifying variation comes from winners of medium or large prizes to winners of small prizes. Consequently, our estimates are most informative about the impacts of wealth shocks equal to several years of income, for example the effect that major changes to capital income taxes or pension systems can have on lifetime wealth.

2.4 Internal Validity

To test our key identifying assumption, we ran quasi-randomization tests premised on the simple idea that if lottery wealth is random conditional on the cell fixed effects, it should not be possible to predict the lottery outcome using covariates determined before the lottery in a regression that controls for the cell fixed effects. We estimate the following regression equation,

$$L_{i,0} = X_{i,0} \times \Gamma + Y_{i,-1} \times \rho_{-1} + \epsilon_i$$

where X_i denotes the individual's assigned cell, $Y_{i,-1}$ is a set of time-invariant characteristics (sex and birth year), as well time-varying characteristics measured in the year before the lottery. These lagged characteristics include marital status, educational attainment, income and a host of financial characteristics. As shown in Table 3, none of the predetermined characteristics are significant predictors of the prize amount, individually or jointly, once the cell fixed effects are included as controls. This result holds across all three lotteries and in the pooled sample.

2.5 External Validity

2.5.1 Representativeness

One concern frequently voiced about studies of lottery players is that individuals who play the lottery may not be representative of the population. To investigate the representativeness of our samples, we compare the players in each of the three samples to a sample of adult Swedes drawn randomly from the Swedish population in December 2000. To avoid comparisons of outcomes that may be endogenous to the outcome of the lottery, we measure all covariates at year-end in 1999 (the first year for which we have financial data) and limit the lottery samples to players who won (or were matched to a winner) in a draw that took place in 2000 or later.

As Table 4 shows, there are several notable differences between the lottery players and the population. Most striking is that our sample of lottery winners are older, more likely to be retired, and less likely to have completed college than the general population. Because of these age differences, we also compared each of the three lottery populations to a representative samples reweighted to match each lottery sample's sex- and age distribution. After sex- and reweighting, the differences on observable characteristics are invariably small. Most importantly, differences in financial characteristics are invariably small very after adjustment for sex- and age differences.

Cross-Sectional Analysis

To probe further into our sample's representativeness, we investigated whether the cross-sectional relationships between stock market participation and household characteristics observed in our sample of lottery players resembles those in previous studies with representative samples. An influential paper by Calvet, Campbell, and Sodini (2007) used data on all Swedish households in the year 2002 to explore the demographic and financial correlates of non-participation. We used data from year-end in 1999 to estimate the same Probit specification in two samples: our representative sample and our sample players from draws conducted in 2000 or later. We use data from the 1999 cross section to avoid using wealth variation that was induced by the lottery (and may therefore change the coefficient estimates even if the a lottery population were representative).

In column 1 of Table 5 we report estimates of the marginal effects obtained from our restricted lottery sample. Column 2 shows the same estimates after the lottery-sample has been reweighted to match the sex- and age distribution of our representative sample. Column 3 reports the marginal effects we obtain in our representative sample observed in 1999. For comparability, Column 4 shows Sodini et al's estimates for the year 2002. Overall, the estimated relationships are quite similar across the four samples. For the key financial variables, we observe that the lottery sample estimates are similar to those obtained from a representative sample. In addition, the estimated marginal effects for our lottery sample are very close to those estimated in Calvet, Campbell, and Sodini (2007), with the only notable difference being related to immigrant status.

2.6 Generalizing Beyond Sweden

Finally, an important concern about external validity is that insights from Sweden may not generalize to other countries. There is surely some merit to the view, which is also discussed by Calvet, Campbell, and Sodini (2007) (p. 712). We nevertheless believe there are compelling reasons to expect the findings we report here to be relevant beyond the Swedish setting. For example, previous work has noted that the predictors of non-participation in Sweden are surprisingly similar the United States (Calvet, Campbell, and Sodini (2007), Table I). Cross-sectional analyses also find that the composition of Swedish household wealth is no outlier when compared to what has been observed for other industrial countries. For example, the fraction of non-participation households was 62% in Sweden in 1999, compared to 59% in the US the same year. These similarities suggest to us that it is plausible to expect that the causal processes that give rise to non-participation in the two countries are not altogether different.

To provide some more indirect evidence on generalize-ability, and set the stage for the rest of the paper, we now give a simple illustration of how researchers have sought to improve our understanding of non-participation by augmenting the standard household finance model with participation costs. We also show that when the models are calibrated to match the Swedish distribution of wealth and degree of non-participation, the required costs are of similar magnitude to those reported for the US by Vissing-Jorgensen (2002).

Vissing-Jorgensen (2002)'s influential calibration exercise is based on a model with time separable and homothetic preferences. Individuals allocate their lifetime wealth to a risk free and a risky asset. Individuals who own a non-zero quantity of the risky asset incur an ongoing per-period fixed cost. Entering the stock market is not costly. Vissing-Jorgensen (2002) calculates participation costs in two ways. First, define the certainty equivalent of return r_{ce} as

$$\mathbb{E}\left[U\left(W_i(1+r_s)\right)\right] = U\left(W_i(1+r_{ce})\right).$$

Here, r_{ce} is the certain return that renders an individual indifferent between participating in equity markets and receiving stochastic return r_s and not participating but receiving r_{ce} for sure. An individual's per period benefit from participating in equity markets can then be expressed as

$$Benefit_{i,t} = W_{i,t} \times \alpha_i \times (r_{ce,t} - r_f \qquad (2)$$

where $W_{i,t}$ denotes individual i's wealth, α_i denotes individual i's portfolio allocation, and r_f denotes the risk free rate. Vissing-Jorgensen (2002) assumes $r_{ce} = .05$, $r_f = .01$, and $\alpha_{i,1989} = .359$ ($\alpha_{i,1993} = .566$) is calibrated to the median of PSID financial portfolios. Benefits are then a linear function of wealth, and given that the 75^{th} percentile of wealth amongst non-participants is approximately $10,000$ in the 1989 and 1993 PSID, this implies an annual cost $F_i^P = \$260$ is sufficient to explain non-participation for 75% of the population.

It is straightforward to repeat this calculation in our 1999 cross-sectional data of lottery players. Estimates of the historical equity premium the Swedish stock market range from 6-9%, and so we follow Vissing-Jorgensen (2002) and specify $r_{ce} - r_f = .04$. We also calibrate $\alpha_i = .43$ to match the median cross-sectional equity share of financial wealth in our sample. In figure 1 we replicate figure 6 of Vissing-Jorgensen (2002) and plot the implied minimum stock market participation costs for our Swedish non-participants. Here we observe that for the 75^{th} percentile of our wealth distribution the necessary per period cost necessary explain non-participation is 2400 SEK (350 USD), not dramatically different from the \$260 estimated in the US population.

A second estimation of fixed costs of participation in Vissing-Jorgensen (2002) allows for population heterogeneity under the assumption that costs are uncorrelated with wealth. Assuming participation is optimal iff

$$W_{i,t} \times \alpha_i \times (r_{ce,t} - r_f) \geq F_i^P \qquad (3)$$

a non-parametric estimate of the distribution of participation costs (denoted G_{F^P}) is the percentage of households with a given wealth level that participate in the stock markets. For example, if a given wealth level $\bar{W}$ implies a participation benefit $Benefit_t(\bar{W})$, then

$$G_{F^P}(Benefit_t(\bar{W})) = \frac{\#\ participants\ with\ W = \bar{W}}{\#\ of\ individuals\ with\ W = \bar{W}}. \qquad (4)$$

To implement this, Vissing-Jorgensen (2002) divides the sample into ten deciles and plots the implied CDF against the median of each wealth decile (see figure 7). We replicate this exercise in figure 2. Here, we see that the median participation cost is XXX and 75% of non-participants are estimated to have participation costs lower than XXX SEK (XXX USD). Again, this estimation applied to our Swedish sample supports the finding of Vissing-Jorgensen (2002) and others that moderate participation costs are capable of accounting for the majority of non-participation.

We will revisit these cross-sectional cost estimates in section 4 to demonstrate their implications for participation responses for our sample of lottery winners. Except for this exercise, the remainder of this paper we will use within individual variation of wealth to estimate participation responses and costs of entry and participation. We will show that this variation in wealth results in higher estimated costs of participation than have been previously found in the literature, and as a result bring into question the structural interpretation of previous cost estimates based on cross-sectional variation.

3 Empirical Results

In this section we use equation 1 to estimate the causal impact of wealth on stock market participation over various time horizons. In our primary specification, we estimate a separate regression for participation at year-end in the year of the lottery and of the five ensuing years (at $s = 0, ..., 5$). Participation in year s is an indicator variable equal to 1 if in year s the household's year-end portfolio included any directly or indirectly held stocks (and 0 otherwise). In all specifications, we control for cell fixed effects and a handful of predetermined characteristics measured at $s = -1$.

Figure 3 depicts the estimated causal effect of wealth from a linear probability model estimated using our pooled lottery sample. We observe an immediate, and seemingly permanent, increase in the participation probability of 3.50-4.00 percentage points per 1M SEK won, implying a marginal effect of approximately 6% given that the proportion of players with stocks is 60%. If we instead adopt an event study framework, and impose the restriction that $\beta_{1,s} = \beta$ for all $s = 0, 1, 2, 3, 4, 5$ the estimate is 3.63 with a standard error of XXX. It is useful to benchmark this estimate against the cross-sectional relationship between wealth and participation, or the relationship in a carefully specified panel in which the effect of wealth on participation is identified from changes in a person's wealth over time. In our data, the cross-sectional relationship is 21.88% and the panel estimates is YYY.

The small aggregate effect may mask substantial heterogeneity, and we turn now to an exploration of this heterogeneity. Given the earlier inertia in participation, we began by investigating how responses vary when the sample is stratified by stock participation in $s = -1$. For those individuals that were not participants before wealth assignment, the estimated impact is around 11 percentage points (s.e.=.02) in the year of the lottery and the ensuing three years. In the population of participants, the estimated response is positive but very small. Thus, virtually the entire participation response is driven by nonparticipants, a finding consistent with the predictions of models in which, large, one-time, fixed costs of entry feature prominently.

Given the evidence that the treatment response is explained almost entirely by a positive effect of wealth in the population of nonrespondents, all tests for heterogeneous treatment effects in the analyses that follow are based on nonparticipants. Furthermore, because the effect appears to be one time and permanent, we only present results for the year following win. Effects for non-participants as well as other horizons are presented in appendices B and C. We test for heterogeneous responses in subsamples stratified along financial characteristics (debt, labor income risk, household portfolio composition) and miscellaneous characteristics (sex, age, educational attainment, recent stock market performance).

In each heterogeneity analysis, we are conceptually interested in comparing the estimated effect of wealth in subsamples stratified along one of the dimensions, for example winners with and without a college degree. Procedurally, we run a single regression in which all regressors are interacted with indicator variable(s) for the subpopulations. The pooled regression recovers exactly the same coefficient estimates as those obtained when Equation 1 is estimated separately in each of the subsamples. To test for heterogeneity, we conduct an F-test of the null hypothesis that the coefficients are identical.

Because only wealth is randomly assigned, evidence of treatment effect heterogeneity along some dimension X need not imply that varying X exogenously will change participation costs. For example, treatment effect heterogeneity by college attainment could in principle arise because college completion is correlated with some factor (pre-college ability) that reduces participation costs independently of college completion. However, the heterogeneity analyses nevertheless provide useful information about how participation costs are distributed across individuals with observable characteristics. Such information is valuable for formulating new hypotheses about the sources of heterogeneity in participation costs, and is a key input into to our treatment of cost heterogeneity in the structural model.

The upper panel of Table 6 shows the results of heterogeneity analyses performed in various subsamples stratified by financial characteristics. We begin with debt. The rate at which individuals can borrow typically exceeds the risk free rate and it has been proposed that this "borrowing wedge" may explain why indebted households elect to repay debt our participating in equity markets. Column 1 shows that the estimated effect of 1M SEK on the participation probability is 0.19 in households classified as debt-free <10K SEK in debt), compared to 0.08 in all other households. Our next heterogeneity analyses is inspired by research suggesting that individuals may rationally choose to not participate because of uninsurable labor income risks Viceira (2001) Heaton and Lucas (2000). We therefore estimate the effect in nonparticipants with and without their own business. As shown in Column 3, we find that no evidence of an increase in participation probabilities in households where the winner was self-employed, and a strong response in all other households. Third we test for heterogeneous responses in households with and without real estate. The role of housing wealth in portfolio choice was first highlighted in influential work by Grossman and Laroque (1990) and Flavin and Yamashita (2002). However, housing price risk is a significant source of background risk that might discourage equity market participation. In our data, the responses are strikingly similar in households with and without real estate the year before the lottery. Finally, Column 5 shows the results in subsamples stratified by recent stock market performance, distinguishing players who won during a bear market (value of stock market declined in the 12 month-period preceding the lottery) or a bull market. If individuals' subjective beliefs about the benefits of participation move with aggregate conditions, then the participation response may too. We find that the effect of wealth on participation is indeed weaker amongst lottery players who during a bear market.

In the lower panel of Table 6 we report heterogeneity analyses for the remaining characteristics. Column 1 shows that the impact of 1M SEK on the participation probability of college-educated winners is 23%, compared to .08% in individuals without a college degree. This difference is highly significant, and consistent with theories that cognitive constraints are an important source of participation Grinblatt, Keloharju, and Linnainmaa (2011b), Van Rooij, Lusardi, and Alessie (2012), assuming that attending college relaxes these cognitive constraints. Columns 2 and 3 show the results for samples stratified by age (18-50, 50-70, 70+) and sex (women and men). Theories in which there is a one-time fixed cost of entry make the prediction that, all else equal, older individuals who are holder should exhibit weaker responses, as they have fewer remaining years in which to harvest the gains of participation.

3.1 Are the Effects Nonlinear?

Under our identifying assumption, our estimator gives en unbiased estimate of a weighted treatment effect, but the linear estimator will assign most weight to the marginal effect of wealth at modest to large wealth shocks, as such prizes account for most of our identifying variation. One possible interpretation of the discrepancy between our causal estimates (0.110 per 1M in nonparticipants) and the cross-sectional estimates (.219 per 1M in nonparticipants) is that the effects of wealth may be non-linear. Additionally, models with participation costs Vissing-Jorgensen (2002) have the additional property that households should follow a threshold strategy: for each household there exists some wealth level above which participation is always optimal. Such a threshold rule is likely to show up in the form of non-linear effects.

To test for non-linear effects, we modify our basic estimating equation so that it can accommodate non-linear responses in a fairly transparent and easy-to-interpret way. Specifically, we replace the continuous prize variable by indicator variables for the prize amount won. By estimating the effect of wealth at different thresholds, it is possible to identify non-linear effects. The third panel in Table 6 presents results from the regression with thresholds at 100K SEK, 1M SEK and 2M SEK. We also report results from linear spline regressions with knots at these prize thresholds. The results strongly suggest that the marginal effect of wealth is rapidly diminishing. A prize of 100K SEK increases the participation probability of 0.071 relative to individuals winning small prizes. This estimate is comparable to the estimated .110 (.101) effect of receiving 1M SEK presented in figure 4.

3.2 Calibrating Participation Costs to Match Causal Estimates

The causal effects we estimate in nonrespondents are not easy to reconcile with the hypothesis that most of non-participation is due to households facing a modest ongoing participation cost (of the order a few hundred dollars per year). Under this theory, many households decline to participate for the simple reason that at their level of wealth, the gains from participation (which are proportional to wealth) do not offset the costs of participation (which are fixed). But as we now show, our estimated participation responses to the wealth shocks are far smaller than is predicted by a model with an annual cost of a few hundred dollars.

To illustrate, we use the effects we estimate in our sample of nonparticipants. We conservatively assume that all households have zero wealth prior to the lottery. From Equation 2, we have that the per-period benefit of participation is 20,000 SEK at a wealth level of 1M and 40,000 SEK at a wealth level of 2M. From Table 6, we have that the effect of a wealth shock of 1-2 M SEK increases the probability of participation by 18 percentage points. The analogous estimate for at least a 2M wealth shock is 27 percentage point.

A natural question to ask is: what participation costs would be required to make a household with 0 wealth indifferent between participating and not participating after a windfall gain of 1M or 2M SEK? Straightforward calculations show that in Vissing-Jorgensen (2002)'s framework, a per-period cost of 40,000 SEK (5,700 USD), is needed to explain our estimate that 73% of non-participants do not enter the stock market after a shock of 2M SEK or more. A cost of 20,000 SEK (2,850 USD) is required to explain why 82% of non-participants do not purchase stocks following a windfall of 1M SEK. These calibration exercises are conservative for two reasons. First, most nonparticipating households start with positive wealth, implying greater gains from participation than assumed above. Second, the effect-size estimate of 18.2 (27.0) probably overstates the participation response to a 1M SEK wealth shock, as it is based on a comparison of participation rates of households who won prizes in the range 1M-2M (2M+) to households winning small prizes. For both reasons, the calculation above is likely to understate the benefits of the wealth shocks, implying even larger costs would be required to account for the observed non-participation rates.

This finding is in sharp contrast to the original results in Vissing-Jorgensen (2002) which concludes that an ongoing participation cost of 260 USD is sufficient to explain non-participation amongst 75% of non-participants. It similarly contradicts the cost estimates from Section 2.6 that we get when we apply the method used by Vissing-Jorgensen in Swedish data.

3.3 Other Robustness Checks

We conducted a number of sensitivity checks to explore the robustness of our results to sample selection criteria, the definition of the dependent variable, and choice of estimator. We report the results from these robustness analyses in Table 7.

To facilitate comparisons, Column 1 shows our original estimate of the effect of 1M SEK on participation in the pooled lottery sample and the two subsamples stratified by participation status. We report coefficients for $s = 0$ and $s = 3$. Column 2 shows that the estimated marginal effects from the Probit analogue, and their standard errors, are very similar to the linear probability model estimates reported in our primary specification. Column 3 shows that the results from a specification in which only directly owned stocks are included in the definition of participation. Here, the estimated effects on this more narrowly defined dependent variable are overall much smaller. The most important lesson from this robustness check is that the results we report are explained primarily by households purchasing mutual funds. In addition, the effect of wealth on participation for non-participants increases for four years following receipt of wealth. It is possible that the delayed stock market entry may reflect learning processes. Columns 4-6 show results for each of the three lotteries considered separately. In all three lotteries, the estimated effect is larger in non-participating households than in participating households. In the Kombi sample, the estimated effect on participation is larger, but overall, the congruence of the findings from the three lotteries show that the conclusions are not driven by any one of the lotteries.

3.3.1 Summary of Reduced Form Findings

To summarize, our reduced form analyses shows that the effect of wealth on participation is immediate, permanent, small, heterogeneous and small.

The effect is *immediate*, as it is usually discernible in the year of the lottery and it appears *permanent*, as the estimated effects of wealth on participation in the years following the lottery are of similar magnitude.

The effects are *small* both when bench-marked both against cross-sectional estimates and the participation responses predicted by a standard household finance model augmented with modest participation costs. The participation costs that would be required to explain the participation responses we estimate are thus strikingly high compared to earlier work.

The effects are *heterogeneous*. The heterogeneous effects we observe are intuitive and usually easy to reconcile with standard theories of participation. Most importantly, we find that the response of 3.5% per million SEK in our pooled sample masks substantial heterogeneity. In the linear model, the effect of 1M SEK is negligibly small in households who already participated, whereas it increases the probability of participation in the approximately ˜40% nonparticipating households by 12 percentage points. We also find a more elastic response in individuals in college-educated winners, in winners with lower background risk, and in households with lower debt and who win during a bull market.

4 A Structural Model

The previous section demonstrated significant differences between our cross-sectional estimates of costs amongst non-participants and our causal estimates of the effect of wealth. To better understand these differences, in this section we present a structural model of equity market participation. The structural model has several advantages that help to clarify these differences. First, in our estimation of the effect of wealth assignment we distinguish between wealth that is assigned and pre-existing wealth. In the cross-section low pre-existing wealth amongst non-participants suggests low costs of participation, while in our causal estimates non-participation amongst recipients of large amounts of wealth suggest high participation costs. Given the collinearity between assigned wealth, existing wealth, and total wealth force, we are forced to omit total wealth in our estimation of equation 1. However, most economic theories, including our structural model, treat total wealth as the key state variable in participation decision. Second, although we include a rich set of covariates and interaction terms in our lottery estimation, we assume a parametric relationship between participation and these covariates. Given that we estimate significant non-linear effects, it is not clear that the assumed relationships accurately reflect consumer decisions. Finally, the estimates presented in previous sections do not map cleanly to participation costs without strong assumptions. In particular, absent a modeling framework, it is impossible to distinguish between entry and participation costs. To allow for and estimate both costs of entry and participation, we need a structural model.

In this section we present a model of life-cycle market participation and undertake a series of exercises to better understand our results. First, we calibrate the model using the cross-sectional implied costs of participation calculated in section 1 and simulate random assignment of wealth. Given that we observe all state variables and our model and the values of wealth assigned, we can simulate each individuals' post-wealth assignment participation decision and compare to observed decisions. We repeat estimations presented in section 3 on our simulated data set, and find not-surprisingly that section 1's participation costs suggest a larger effect of wealth on participation probability than we estimate.

In a second exercise, we estimate costs of participation and entry that best replicate the participation responses of lottery winners. Our estimation strategy uses the method of indirect inference, choosing costs that best replicate the estimates shown in section 3. We identify the population distribution of participation costs, as well as the distribution of entry costs amongst non-participants before random assignment of wealth. We are clearly not able to identify the population distribution of entry costs given that a significant proportion of our sample participate in the stock market prior to random assignment of wealth. However, we find that introducing entry costs significantly improves our ability to replicate estimates in section 3. Furthermore, this identification challenge is not unique to our study (e.g., Khorunzhina (2013)), and identification of the conditional distribution is a significant distribution to the literature on life-cycle portfolio choice.

Modern life-cycle portfolio choice literatures typically allow for more heterogeneity,costs, and motives than we permit in this exercise[4]. Given that it is not the intention of this study to innovate on or extend these models, we make several simplifying assumptions. This permits us to highlight the difference between participation costs estimated from the cross-section and individual variation in wealth. In the remainder of this section we present the model and results.

4.1 Model

We assume a very standard model of life-cycle savings, market participation, and portfolio choice. Each period an age t agent chooses how much to consume, save, and invest in equity markets according to their current resources. An agent has finite lifespan, living to age T, but faces mortality risk with exogenous survival probability from period t to $t+1$ denoted s_t. Upon death, an agent receives terminal payout of zero. If an agent decides to participate in equity markets they face separate financial costs of participation and entry. Participation costs, denoted κ, are payed each period an agent allocates non-zero wealth to equity holdings. Entry costs, denoted χ are only paid once during the first period an agent decides to hold non-zero equity. Even if an equity market participant exits equity markets, there is no required cost for subsequent re-entries beyond the per-period participation costs. Equity provides a return r_s with $\mathbb{E}(r_s) > r_f$, but the return is risky. In addition, each period an agent is endowed with age specific labor income y_t.

We express each agent's dynamic decision problem as a Bellman equation. Let I_t be an indicator for whether an agent has never previously participated in equity markets. For an agent that decides to not-participate in equity markets, their value function is standard. Given a continuation value V_{t+1},

[4]see Haliassos and Bertaut (1995), Viceira (2001) Gomes and Michaelides (2005),Cocco, Gomes, and Maenhout (2005), Alan (2006), and Khorunzhina (2013), amongst others

the agent simply decides how much to consume and how much to save at risk free rate r_f.

$$V_t^{NP}(W_t, I_t) = \max_{c_t, W_{t+1}} u(c_t) + \beta s_t \mathbb{E}_{y_{t+1}} \left[V_{t+1}(W_{t+1}, I_{t+1}) \right]$$
$$W_{t+1} = r_f \left(W_t - c_t \right) + y_{t+1}$$
$$I_{t+1} = I_t$$

For an agent that decides to participate in the equity market, the agent must pay costs of participation and decide how to allocate wealth between stocks and bonds. Given continuation value V_{t+1}, an equity market participants problem can be expressed as

$$V_t^P(W_t, I_t) = \max_{c_t, W_{t+1}, \alpha_t} u(c_t) + \beta s_t \mathbb{E}_{y_{t+1}, r_{s,t+1}} \left[V_{t+1}(W_{t+1}, I_{t+1}) \right]$$
$$W_{t+1} = r_f \left(W_t - c_t - \kappa - I_t \chi \right) + (r_{s,t+1} - r_f) \alpha_t \left(W_t - c_t - \kappa - I_t \chi \right) + y_{t+1}$$
$$I_{t+1} = 0$$

Finally, each period an agent's decision to participate or not is determined by the max of the above two value functions. An agent's full decision problem is specified as

$$V_t(W_t, I_t) = \max\{V_t^{NP}(W_t, I_t), V_t^P(W_t, I_t)\}. \tag{5}$$

To fully characterize the decision problem, we need to specify stochastic processes for s_t, $r_{s,t}$ and y_t. Although our data is collected at the household level, we treat s_t as the individual survival probability as a function of age t. This can be interpreted as either that the household dies upon death of the specified agent or that there is no preference for remaining household members upon death. These survival probabilities are calculated based on the age-specific survival rates of the Swedish population, and are presented in figure 5.

We assume that the log of equity returns follow a stationary MA(1) process, with mean $\bar{r}$ and standard deviation σ_s. We represent the equity return process as

$$ln r_{s,t} = \bar{r} + \sigma_s \epsilon_t.$$

We estimate $\bar{r}$ and σ_s from the post-war equity return distribution of the Swedish stock exchange. This is done using a linear regression, resulting in estimates of $\bar{r} = X$ and $\sigma_s = X$. We calibrate our equity returns using these values, and calibrate $r_f = .01$ at the historical average of Swedish treasury bills.

In addition, we assume an age-dependent income process, with log income being expressed as

$$ln y_t = f(t) + \sigma_{y,t} \eta_t.$$

We assume that $f(t)$ is a quadratic in age and $\sigma_{y,t}$ is chose to match the the age-specific dispersion in earnings. We omit any persistence in earnings and any heterogeneity in earnings. We estimate this equation using the 1999 cross-section of the Swedish population, and present the estimated earnings profile in figure 6. Finally, we assume that utility is CRRA, and calibrate risk aversion parameter $\nu = 3$. We could in principal estimate this, but choose to instead focus on estimating participation costs. We assume that entry and participation costs are linear in observables, with each individual's time invariant parameter being represented by the following expressions.

$$
\begin{aligned}
ln\kappa_i =& \zeta + \epsilon_{\kappa,i} && (6)\\
& \epsilon_{\kappa,i} \sim \mathbb{N}(0, \sigma_\kappa) \\
ln\chi_i =& \psi + \epsilon_{\chi,i} && (7)\\
& \epsilon_{\chi,i} \sim \mathbb{N}(0, \sigma_\chi)
\end{aligned}
$$

In our first exercise, the constants ζ and ψ and the corresponding standard deviations are chosen to match the CDF presented in figures 2. In our second exercise, we estimate the model using the method of indirect inference. For each individual we sample a large number of draws from the cost distribution implied by parameters $\Theta = (\zeta, \psi, \sigma_\kappa, \sigma_\psi)$. We then simulate participation decisions implied by the model. We then repeat estimation of equation 1 for several specifications, and choose cost parameters that match the causal effect coefficients for unconditional participation responses (figure 3), pre-win participants and non-participants (figure 4), the pre-existing wealth coefficient pre-existing wealth coefficient for both of these estimates, and the non-linear coefficients for participants and non-participants (presented in table 6, panel 3, columns 1-3 for non-participants). The estimation procedure can be formally expressed as:

$$\hat{\Theta} = \arg\min_{\Theta} (\hat{\beta} - \tilde{\beta}(\Theta))' W (\hat{\beta} - \tilde{\beta}(\Theta)) \tag{8}$$

where $\hat{\beta}$ represents the three lottery response and two pre-existing wealth coefficients we seek to match. For further information on the implementation of our structural estimation procedure we refer the reader to appendix E

4.2 Results

Given that the calculation presented in figures 7 only provide costs of participation, for our initial two exercises we set $\chi = 0$ and $\sigma_\chi = 0$. To calibrate the model to these cross-sectional estimates, we set $\kappa = 360$ and $\sigma_\kappa = 100$. The resulting implied CDF is presented in figure 7.

In the first column of table 8 we present the results from this calibration. from this calibration. Not surprisingly, we observe that our model dramatically overpredicts the effect receipt of wealth has on participation. Amongst non-participants, the model predicts that receipt of wealth should have a large impact on participation, with an estimated coefficient of .85. This is almost eight times the effect we see in the data. The model similarly ovepredicts the effects of wealth on participants as well as the unconditional effect, as these individuals would likely continue participating even absent winnings. It performs especially poorly amongst non-partiipant winners of larger prizes , predicting near universal participation despite this not being observed empirically. In short, this exercise confirms that modest costs of participation are incompatible with the causal effects wealth we estimate.

In the second column of table 8, we present results from estimating equation 8. We estimate the costs of participation to be slightly larger than those implied by the calibration would suggest, but not drastically different. The most striking feature, however, is our estimation of extremely large, widely dispersed, costs of entry. We estimate $\hat{\chi} = 2,200,000$ and $\hat{\sigma_\chi} = 420,000$, suggesting that entry into the stock market incurs a financial cost that is on average approximately 320,000 USD and has a cross-sectional standard deviation of 60,000 USD. This cost is perhaps absurdly high, but when examining the simulated reduced form coefficients we see that they match their targets decently well. Given that the model fit is undeniably improved, we conclude that very high costs of entry are necessary to match observed responses to random assignment of large amounts of wealth when a model is restricted to fixed costs of entry and participation.

5 Structural Model with Heterogeneous Costs

As mentioned in section 3, heterogeneity analyses are slightly difficult to interpret given that the underlying distribution of participants and non-participants changes is correlated with the conditioning variables. Because the heterogeneous characteristics for which we examine participation responses are not exogenously assigned, when we observe higher participation responses in one group compared to another it is unclear if the differential responses reflect different costs or differences in pre-assignment states. To better understand what these results imply for costs of participation, in this section we estimate our structural model allowing for heterogeneous costs of entry and participation. Thus, in this section we build upon the specification of costs presented in 6 and re-estimate equation 8 in an effort to identify how the implied costs differ in our population.

To allow for heterogeneous costs, we denote a set of individual characteristics x_i and alter the costs as follows:

$$
\begin{aligned}
ln\kappa_i =& \zeta x_i + \epsilon_{\kappa,i} \\
& \epsilon_{\kappa,i} \sim \mathbb{N}(0, \sigma_\kappa) \\
ln\chi_i =& \psi x_i + \epsilon_{\chi,i} \\
& \epsilon_{\chi,i} \sim \mathbb{N}(0, \sigma_\chi).
\end{aligned}
$$

Thus, ζx_i now denotes the expected (log) participation costs of an individual with characteristic set x_i , while ψx_i denotes the expected (log) entry costs of an individual with characteristic set x_i.

We are primarily interested in replicating the results from Table 6, and so we augment our previous matched coefficients with the results from this section. Thus, in equation 8 we augment β to include all coefficients from the baseline estimation, as well as coefficients of assigned wealth for participants and non-participants conditional on home ownership, debt, entrepreneurial risk, prior stock market returns, education, gender, and age. To generate these differential patterns, we thus include these covariates in x_i to allow for differential costs.

The resulting parameter estimates are presented in table 9. Here we observe [Need to complete heterogeneous cost estimation and update this section].

6 Conclusion

In this paper we have used random assignment of significant amounts of wealth to estimate the effect of wealth on stock market participation. Generally, we estimate a small, significant unconditional effects, a significant effect for individuals that did not own equity prior to assignment of wealth, and practically no effect for individuals that owned equity prior to assignment of wealth. Surprisingly, amongst individuals that did not own equity prior to wealth assignment, receiving 1M SEK has a fairly small effect on participation probability relative to prize size. In no specification, including amongst individuals that received more than 2M SEK (approximately 250k USD), do we observe the effect of wealth on participation probability rise above .3, and most specifications suggest an effect of less than .2. These effects suggest per annum costs of equity market participation above 20000 SEK (2800 USD) are necessary to explain non-participation for 75% of non-participants. This number not only is far different from what previous cross-sectional studies have found, but simply too large for recurring participation costs to be a valid explanation for observed non-participation. In addition, our estimates of no effect of wealth on participation for pre-assignment participants further suggests that participation costs are not suitable for modeling equity market participation.

In estimating our structural model, we find participation costs alone are not able to match participation responses, but that entry costs are necessary. We estimate these costs using the method of indirect inference, and find again that these costs of entry are much higher than have previously been found in other studies. This suggests that other mechanisms are likely driving non-participation. Finally, when conducting our heterogeneity analysis, we find that low education, high debt, and background risks are associated with higher costs of entry. Although we estimate the impact of these characteristics as costs and do not model the actual channels through which they likely affect participation, our estimates suggest that future structural models of life-cycle participation should focus on correct modeling of the channels through which they operate.

More broadly, our paper shows the dangers of using cross-sectional variation in characteristics to make inference regarding individual behavior. Identification from the cross-section is complicated by reverse-causality and unobserved heterogeneity, and we estimate extremely different implications from cross-sectional variation than within individual variation of wealth. Assumptions required to identify individual effects in cross-sectional studies are often quite strong (as seen with our assumption of no correlation between cost of participation and wealth in section 1, and the identification strategy should be carefully considered rather than taken at face value. Future work with this project will use our generated individual variation to examine not only the external margin effect of wealth on financial risk taking, but also the internal margin of adjustment.

In short, we utilize an extremely unique data set to estimate a behavior that has implications for several branches of economics, and our resulting estimates call into question a common belief that moderate costs of participation are capable of accounting for the lack of participation observed empirically.

References

Alan, S. (2006): "Entry costs and stock market participation over the life cycle," *Review of Economic Dynamics*, 9(4), 588–611.

Andersen, S., and K. M. Nielsen (2011): "Participation constraints in the stock market: Evidence from unexpected inheritance due to sudden death," *Review of Financial Studies*, 24(5), 1667–1697.

Apouey, B., and A. E. Clark (2013): "Winning Big but Feeling No Better? The Effect of Lottery Prizes on Physical and Mental Health," Working paper, Institute for the Study of Labor.

Brunnermeier, M. K., and S. Nagel (2008): "Do Wealth Fluctuations Generate Time-Varying Risk Aversion? Micro-Evidence on Individuals' Asset Allocation," *The American Economic Review*, pp. 713–736.

Calvet, L., J. Y. Campbell, and P. Sodini (2009): "Fight or flight," *Portfolio rebalancing by individual investors, Forthcoming in The Quarterly Journal of Economics.*

Calvet, L. E., J. Y. Campbell, and P. Sodini (2007): "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes," *Journal of Political Economy*, 115(5), 707–747.

Calvet, L. E., and P. Sodini (2014): "Twin Picks: Disentangling the Determinants of Risk-Taking in Household Portfolios," *The Journal of Finance*, 69(2), 867–906.

Campbell, J. Y. (2006): "Household Finance," *Journal of Finance*, 61(4), 1553–1604.

Cesarini, D., E. Lindqvist, R. Östling, and B. Wallace (2013): "Estimating the causal impact of wealth on health: Evidence from the Swedish lottery players," *Unpublished manuscript.*

Cocco, J. F., F. J. Gomes, and P. J. Maenhout (2005): "Consumption and portfolio choice over the life cycle," *Review of Financial Studies*, 18(2), 491–533.

Ettner, S. L. (1996): "New Evidence on the Relationship Between Income and Health," *Journal of Health Economics*, 15, 67–85.

Fagereng, A., C. Gottlieb, and L. Guiso (2013): "Asset market participation and portfolio choice over the life-cycle," *WP*, NA, NA.

Flavin, M., and T. Yamashita (2002): "Owner-occupied housing and the composition of the household portfolio," *American Economic Review*, pp. 345–362.

Friend, I., and M. E. Blume (1975): "The demand for risky assets," *The American Economic Review*, pp. 900–922.

Gardner, J., and A. J. Oswald (2007): "Money and Mental Wellbeing: A Longitudinal Study of Medium-Sized Lottery Wins," *Journal of Economic Health*, 26, 49–60.

Gomes, F., and A. Michaelides (2005): "Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence," *The Journal of Finance*, 60(2), 869–904.

Grinblatt, M., M. Keloharju, and J. Linnainmaa (2011a): "IQ and Stock Market Participation," *Journal of Finance*, 66(6), 2121–2164.

——— (2011b): "IQ and stock market participation," *The Journal of Finance*, 66(6), 2121–2164.

Grossman, S. J., and G. Laroque (1990): "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, 58(1), 25–51.

Haliassos, M., and C. C. Bertaut (1995): "Why do so few hold stocks?," *Economic Journal*, 105(432), 1110–1129.

Hankins, S., and M. Hoestra (2011): "Lucky in Life, Unlucky in Love? The Effect of Random Income Shocks on Marriage and Divorce," *Journal of Human Resources*, 46(2), 403–426.

Hankins, S., M. Hoestra, and P. M. Skiba (2011): "The Ticket to Easy Street? The Financial Consequences of Winning the Lottery," *Review of Economics and Statistics*, 93(3), 961–969.

Heaton, J., and D. Lucas (2000): "Portfolio choice and asset prices: The importance of entrepreneurial risk," *The journal of finance*, 55(3), 1163–1198.

Imbens, G. W., D. B. Rubin, and B. I. Sacerdote (2001): "Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players," *American Economic Review*, 91(4), 778–794.

Kearney, M. S., P. Tufano, J. Guryan, and E. Hurst (2008): "Making Savers Winners: An Overview of Prize-Linked Savings Products," Working Paper 16433, NBER.

Kearney, M. S., P. Tufano, J. Guryan, and E. Hurst (2010): "Making savers winners: An overview of prize-linked savings products," Discussion paper, National Bureau of Economic Research.

Khorunzhina, N. (2013): "Structural estimation of stock market participation costs," *Journal of Economic Dynamics and Control*, 37(12), 2928–2942.

King, M. A., and J. I. Leape (1984): "Wealth and portfolio composition: theory and evidence," *Journal of Public Economics*.

Lindahl, M. (2005): "Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income," *Journal of Human Resources*, 40(1), 144–168.

Mankiw, N. G., and S. P. Zeldes (1991): "The consumption of stockholders and nonstockholders," *Journal of Financial Economics*, 29(1), 97–112.

Merton, R. C. (1971): "Optimum consumption and portfolio rules in a continuous-time model," *Journal of economic theory*, 3(4), 373–413.

Poterba, J. M., and A. A. Samwick (2003): "Taxation and household portfolio composition: US evidence from the 1980s and 1990s," *Journal of Public Economics*, 87(1), 5–38.

Samuelson, P. A. (1969): "Lifetime portfolio selection by dynamic stochastic programming," *The review of economics and statistics*, pp. 239–246.

van Rooij, M., A. Lusardi, and R. J. Alessie (2011): "Financial Literacy, Retirement Planning, and Household Wealth," NBER Working Papers 17339, National Bureau of Economic Research, Inc.

Van Rooij, M. C., A. Lusardi, and R. J. Alessie (2012): "Financial Literacy, Retirement Planning and Household Wealth*," *The Economic Journal*, 122(560), 449–478.

Viceira, L. M. (2001): "Optimal Portfolio Choice for Long-Horizon Investors with Nontradable Labor Income," *The Journal of Finance*, 56(2), 433–470.

Vissing-Jorgensen, A. (2002): "Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures," Discussion paper, National Bureau of Economic Research.

Tables

	Period	Treatment	Cells
PLS	1999-2003	Total prize, fixed	No. wins
Kombi	1999-2008	Prize	Prize Draw × Balance
Tv-Triss	1999-2008	Prize	Year

Table 1: Overview of Identification Strategy

Variables	**PLS**	**Kombi**	**TV-Triss**	**Klover**
$x \leq 100000$	37066	29524	408	0
$100000 < x \leq 500000$	0	0	1255	0
$500000 < x \leq 1000000$	0	0	96	0
$x \geq 1000000$	2	297	93	416

Table 2: Prize Distribution

Variables	**Pooled**	**PLS**	**Kombi**	**TV-Triss**	**Klover**
Female	.003	2.0e-4	-1.6e-4	.073	.180
	(.002)	(1.1e-4)	(.004)	(.037)	(.142)
Age	5.0e-4	1.6e-5	7.1e-4	.004	-.007
	(3.0e-4)	(1.4e-5)	(.001)	(.006)	(.036)
Age^2	-4.4e-6	-1.7e-7	-7.5e-6	-4.5e-5	1.1e-4
	(2.6e-6)	(1.5e-7)	(9.0e-6)	(5.2e-5)	(3.6e-4)
Capital Income	3.7e-7	-1.48e-9	1.80e-6	-1.7e-5	1.3e-5
	(3.0e-7)	(2.9e-9)	(1.8e-6)	(1.5e-5)	(1.9e-5)
Marital status	-3.6e-6	2.3e-7	5.0e-6	-3.3e-4	-8.3e-4
	(4.7e-6)	(5.5e-7)	(7.8e-6)	(1.3e-4)	(7.6e-4)
Wealth	5.6e-4	-2.0e-5	-5.2e-4	.011	.022
	(3.0e-4)	(1.8e-4)	(6.5e-4)	(.013)	(.060)
Pre-Win	-.001	1.3e-4	-.003	.006	.105
Market Participation	(.002)	(9.e-4)	(.002)	(.0348)	(.156)
Labor Income	-1.8e-7	-3.8e-9	-1.0e-6	-8.7e-6	-3.9e-7
	(.010)	(5.6e-9)	(5.3e-7)	(8.4e-6)	(7.1e-5)
R^2	.104	.001	.003	.012	.080
$P(F > f)$	.247	.674	0.125	.080	.303

Table 3: Testing for random assignment

Variables	**Population**	**Full Sample**	**PLS**	**Kombi**	**TV-Triss**	**Klover**
Size	100000	69157	37068	29821	1852	416
Disposable Income	2555.797	2555.745	2605.627	2474.063	2539.907	2596.64
ln(Financial Assets)	11.25626	10.95085	11.55651	10.09041	8.048509	7.906221
ln(Real Estate)	10.01015	9.939964	9.755607	10.3609	8.632976	8.869826
ln(Total Debt)	8.14286	8.101101	7.487512	9.063475	7.819228	8.091861
Age	46.83393	60.1189	61.18995	59.58445	49.77484	49.04567
Household Size	2.33265	2.102939	2.074441	2.12065	2.413158	2.44186
High School	.42905	.4445869	.4142524	.4876367	.5157895	.5348837
Post High School	.23471	.2725777	.3028328	.2259861	.2333333	.255814
Retired	.19872	.3872021	.4049976	.3782565	.177193	.124031
Entrepreneur	.03219	.0282327	.0262353	.0300961	.0429825	.0465116
Unemployed	.10959	.0712471	.0628118	.0814054	.1114035	.1124031
Student	.137	.0426999	.0512635	.0246603	.0868421	.0775194
Immigrant	.08951	.0714407	.0750443	.0613855	.1166667	.1124031

Table 4: Comparison of sample to population

Variables	Lottery Sample			Matched Population Sample			CCS 2007		
	Coeff	Z	ME	Coeff	Z	ME	Coeff	Z	ME
Constant	-2.01	-27.15		-3.12	-57.36		-4.852	-98.80	
Disposable Income	8.7e-5	8.90	6.02%	3.3e-5	8.91	4.41%	10.6	15.30	8.80%
Private Pension Premia/Income							3.053	14.00	6.10%
Financial Assets	.264	71.16	21.88%	.158	78.84	55.07%	.491	115.00	22.9%
Total Liabilities	.019	2.80	3.49%	-.033	-6.02	21.55%	.017	13.50	4.6%
Total Real Estate	.010	1.38	1.93%	.041	7.85	-14.88%	.021	22.80	3.3%
Age	-.012	-11.15	-5.95%	-.007	-12.77	-5.15%	-.014	-26.30	-9.8%
Retired	.023	.72	0.74%	-.172	-7.03	-6.36%	.137	6.04	-3.8%
Entrepreneur	-.019	-.35	-0.61%	-.022	-.72	-0.84%	-.030	-.94	-1.0%
Unemployed	.084	2.32	2.63%	.002	.12	0.09%	-.065	-3.61	-2.3%
Student	.065	1.29	2.05%	.109	5.12	4.23%	.028	1.08	1.0%
Immigrant	-.092	-2.61	-0.95%	-.162	-7.59	0.40%	-.384	-26.30	-14.4%
Household Size	-.029	-2.74	-3.05%	.011	1.96	-6.00%	-.009	-1.62	-.4%
High School education	.121	5.22	3.76%	.132	9.14	5.14%	.195	13.50	7.1%
Higher Degree	.214	7.82	6.41%	.193	11.66	7.54%	.130	9.12	4.3%

Table 5: Probit Estimation of Participation. Marginal effects are calculated as the effect of increasing each variable by one standard deviation or by changining an indicator variable to one when all other variables are set to their median level.

	Financial Characteristics					
Characteristic	**Debt Free**	**Income Risk**	**Entrepreneur**	**Real Estate Holdings**	**Positive Prior Year Return**	
Effect	.19		.019	.12	.18	
	Other Characteristics					
Characteristic	**Education**	**Male**	**Age 18-50**	**Age 51-70**	**Age 70+**	
Effect	.23	.18				
	Prize Size					
Characteristic	**100k SEK Threshold**	**1M SEK Threshold**	**2M SEK Threshold**	**100k SEK Spline**	**1M SEK Spline**	**2M SEK Spline**
Effect	.07	.18	.27			

Table 6: Heterogeneous Effects. All presented coefficients are significant at a 1% level.

Table 7: Robustness Checks

Parameter	**Reduced Form Estimate**	**Calibration Value**	**Estimation Value**
κ		460	2013
σ_κ		100	403
ψ			2.2e6
σ_ψ			4.2e5
Unconditional Effect	.04	.3	.08
Pre-existing Wealth	1.3e-6	.8e-5	2.4e-6
Effect on P	.00	.01	.08
Effect on NP	.11	.85	.15
Effect of 100k-1M (P)	.01	.07	.06
Effect of 1M-2M (P)	.03	.11	.08
Effect of 2M+ (P)	.06	.20	.12
Effect of 100k-1M (NP)	.07	.84	.11
Effect of 1M-2M (NP)	.18	.89	.22
Effect of 2M+ (NP)	.27	.90	.28

Table 8: Structural estimation results . ** Note, these numbes are very preliminary. Qualitative results unlikel to change, but quantitative results are very likely to move**

Parameter	Reduced Form Estimate	Value	Estimation Value
κ_{Debt}			
κ_{Ent}			
κ_{Home}			
κ_{Ret}			
κ_{Ent}			
$\sigma_{Education}$			
κ_{Gender}			
κ_{Age}			
σ_{κ}			
ψ_{Debt}			
ψ_{Ent}			
ψ_{Home}			
ψ_{Ret}			
ψ_{Ent}			
$\psi_{Education}$			
ψ_{Gender}			
ψ_{Age}			
σ_{ψ}			
Unconditional Effect	.04		
Pre-existing Wealth	1.3e-6		
Effect on P	.00		
Effect on NP	.11		
Effect of 100k-1M (P)	.01		
Effect of 1M-2M (P)	.03		
Effect of 2M+ (P)	.06		
Effect of 100k-1M (NP)	.07		
Effect of 1M-2M (NP)	.18		
Effect of 2M+ (NP)	.27		
Effect of Debt (NP)	.19		
Effect of Entrepreneur (P)	.02		
Effect of Home Ownership (P)	.12		
Effect of Prior Market Return >0 (P)	.18		
Effect of Education (P)	.23		
Effect of Male (P)	.18		
Effect of Age (P)			
Effect of Debt (NP)			
Effect of Entrepreneur (NP)			
Effect of Home Ownership (NP)			
Effect of Prior Market Return >0 (NP)			
Effect of Education (NP)			
Effect of Male (NP)			
Effect of Age (NP)			

Table 9: Structural estimation results permitting heterogeneous costs

Figures

Figure 1: Replication of AVJ estimation A

Figure 2: Replication of AVJ estimation B

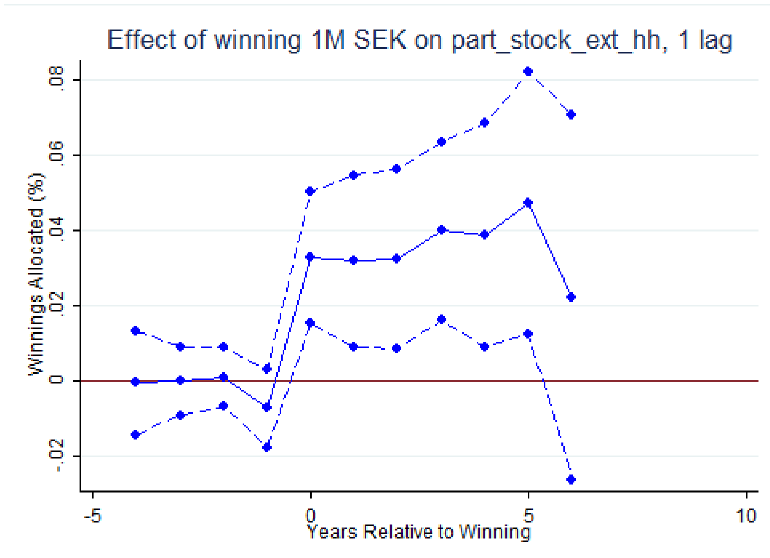

Figure 3: Unconditional effect on participation probability.

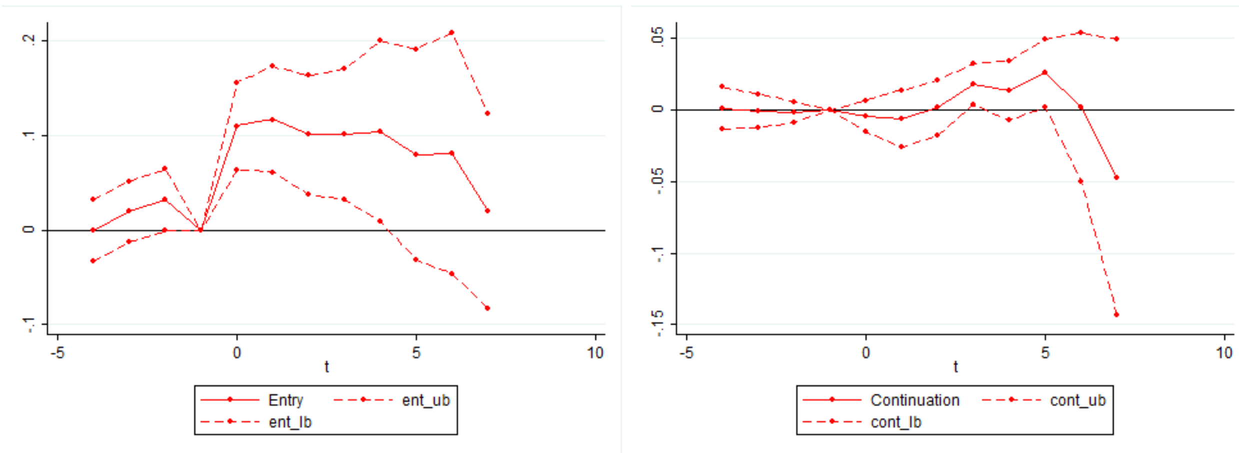

Figure 4: Effect of wealth on participation, when participation is defined as mutual fund or direct ownership. The left hand side displays the estimated effect for non-participants, while the right displas the effect for paricipants

Figure 5: Survival Probabilities

Figure 6: Income Profiles

Figure 7: calibrated CDFs.

A Data

See Cesarini, Lindqvist, Östling, and Wallace (2013) for more detail. To be added here before posting online.

B Direct Participation only Results

To be added.

C Other Results

To be added.

D Model Solution

Available on request. In addition, code will be posted online shortly.

E Estimation Details

Available on request. In addition, code will be posted online shortly.