Exposure to bankers: networks and stock market participation

Arian Borgers, Rachel Pownall, Louis Raes *†

Abstract: In this paper we investigate the importance of acquaintance networks for financial decisions by households. We construct a variable capturing the expected proximity or social closeness to a sub-population of financially savvy people using an overdispersed Poisson model. This measure captures the exposure to people with financial knowledge in an investor's acquaintance network. We find that investors with a higher exposure to financial savvy people are more likely to invest in stocks. This holds after controlling for a wide range of known stock market participation determinants. Moreover when restricting to households with elevated levels of trust or wealth the impact of exposure to financial savvy people increases. Our main findings continue to hold in several analyses to uncover an exogenous effect from proximity on stock market participation. Furthermore we show that the importance of acquaintance networks extends to other financial decisions as well.

* Arian Borgers: Maastricht University, ECCE. Rachel Pownall: Maastricht University, ECCE, Tias Business School. Louis Raes: CentER and EBC, Tilburg University. Email: l.b.d.raes@tilburguniversity.edu *(corresponding author).*

† We thank participants in the Tilburg Economics workshop, in the Maastricht Finance seminar, in the Mannheim SFB seminar, at the Research in Behavioral Finance 2016 Conference, at the EEA 2016 conference, at the Rimini 2016 Conference in Economics and Finance, at the CEA 2016 conference, at the FMA 2016 conference and Dimitris Georgarakos for comments and suggestions. We are grateful to Shirin Goichi for discussing the implementation of the overdispersed model and graciously sharing code. We thank Gerrit Bloothooft from the Meertens Institute for providing data on first names in the Netherlands.

How many people do you know named Jennifer, Nicole, David or Robert? How many people do you know who work for a financial institution? How many people do you know working as a private banker? We expect that readers of this paper would know, on average, more people working for a financial institution or as a private banker then a randomly chosen U.S. citizen -even after controlling for how many people both know. This is less likely to be the case for name questions.

In this paper we use survey answers to questions of the type *"How many X's do you know?"* to capture aspects of individuals' social networks. In particular we construct a proximity variable capturing how close one is to financially savvy people. We then argue that people with more financially savvy people in their network are more likely to participate in the stock market. Our results indicate that the impact of being exposed in your personal network to people with financial knowledge on stock market participation is substantial. Moreover the effect holds after controlling for variables which have been found to be predictive of stock market participation in previous studies.

The decision to participate in the stock market and to hold stocks, has received considerable attention because it is relevant for policy makers and it has theoretical implications. For various policy debates related to the financial health of households and saving for old age, an important consideration is the extent to which stock market non-participation is voluntary. If households do not participate due to frictions then policies which force exposure to the stock market e.g. through the pension system may have some merits, see Abel (2001a). The theoretical interest is related to the ongoing academic debate on the equity premium. Various authors have argued that the participation rate has implications for the stock market premium, e.g. Mankiw and Zeldes (1991), Brav, Constantinides, and Geczy (2002).

In this paper we use methods originally developed to learn about hard-to-count populations such as people suffering from drug abuse or homeless people. These methods were developed because collecting complete network data is financially and logistically impossible. These methods rely on survey questions of the type "How many X's do you know?", where X can be a name, a profession or a trait. The data generated by this type of questions is called aggregate relation data or ARD.

The focus of this paper is social interaction. The idea is that you are more likely to hold stocks if you know many people with a certain degree of financial sophistication. This seems intuitive. However an empirical analysis is not straightforward for various reasons and part of this paper is devoted to demonstrating how one can measure the extent to which people are exposed to other people with financial knowledge.

Humans are social species and social interactions with friends and acquaintances are key in understanding how humans form an opinion, which goods they buy or what they invest in. How information spreads depends ultimately on the full network, see Jackson (2014). Since Granovetter (1973) we know that not only close friends matter, but that also acquaintances (weak ties) play a substantial role in shaping opinions and ultimately behavior.

Combined with increased access to data sources and computing power, we see an increase in research on the importance of networks in various fields. In economics and finance however, empirical studies still have the tendency to proxy networks by peer groups or geography. Jackson (2014) argues that while such approaches can be valuable, often insight can be gained by delving into network data.

In this paper we construct a *proximity measure* which captures how close one is to financial savvy people. To do so, we estimate a general model of social ties. We then use the difference between the prediction of this model and the actual answer as a measure of how close one is to a particular subpopulation (in our case financially savvy people). The model takes into account how large an individual's social network is, the size of the subpopulation of interest and the propensity of individuals to get to know people of certain subpopulations. This proximity measure then forms the key element in our subsequent analysis of stock market participation. In our analysis we control for a range of well established variables as well as variables which recent research has found to be related to stock market participation. We acknowledge the possibility of endogeneity in our first set of estimation results. From the onset we therefore collected additional information that can serve as instruments to our proximity variable. In addition we also perform placebo analyses to test for unobserved factors within groups of individuals with similar socio-economic profiles and we propose a battery of additional control variables. The estimation results using instrumental variables, placebo variables, and additional controls confirm earlier results.

The contribution of this paper is twofold. First we show how one can construct a network variable using questions which are easy to include in a survey. This approach is very versatile and not too costly. Existing research on network effects in household finance typically relies on particular data sets which may be difficult to transfer to other contexts. The approach we discuss here, can be applied in many settings. The development of statistical methods seems to be a precursor to applications. Most applications are related to so-called hard-to-count groups such as people at risk for drug abuse or homeless people. The only other application in the Finance literature is Hong and Xu (2015). The latter paper develops a measure capturing whether a mutual fund manager is a member of a social network. Managers with geographically concentrated stock pics and a high propensity to be connected to a city tend to have university alumni connections in that city and outperform other managers. The paper by Hong and Xu (2015) is similar in flavour to this paper and also builds on similar insights into the theory and statistics of networks. Our paper differs from Hong and Xu (2015) in the following ways. First we focus on household finance rather than on the mutual fund industry. We are concerned with the impact of social networks on financial decisions by households whereas Hong and Xu (2015) analyze investment decisions and performance in the mutual fund industry. Second, there are also important methodological differences due to the differential nature of the data. Despite these differences this paper shares a key intuition with Hong and Xu (2015). Hong and Xu (2015) consider the relative propensity to have contacts and calculate this as the ratio of the actual number of acquaintances of a manager in a given city divided by the predicted number of acquaintances in a given city, where the prediction comes from an overdispersed Poisson model. In this paper we use a proximity measure which is the difference between the actual number of financially savvy people known by an individual and the predicted number. In both cases the intuition is that if you know more people with a certain trait than what can be expected on the basis of a general model of social ties, then this reflects the existence of a (latent) social network. One attractive property of the proximity measure we use is that there is a link with a network theory which is often lacking in empirical research on peer effects and social networks in the household finance literature, see the discussion Georgarakos and Pasini (2011) and the literature review below.[1]

Second, we add to the literature on stock market participation or the *stock-holding puzzle* (Haliassos and Bertaut (1995), Campbell (2006)). We argue that social interaction is important for understanding stock market participation. Social interaction is especially relevant when the interaction involves financial savvy people. So what matters is not so much with how many people an investor interacts rather with how many financially knowledgeable people an investors interacts. The intuition is straightforward. Consider two investors, A and B. Investor A belongs to the lower quartile of sociability (a measure often used in the literature on household finance, see Hong, Kubik, and Stein (2004)) but the vast majority of his social contacts work in the financial industry. Investor B belongs to the upper quartile of sociability but none of his contacts has any financial knowledge. Our argument is that investor A, all else equal, has a larger probability of owning stocks.[2]

[1]In economics, there are papers which connect empirical research of peer effects to network theory. Prime examples are Calvó-Armengol, Patacchini, and Zenou (2009) or Alatas, Banerjee, Chandrasekhar, Hanna, and Olken (2012). There is a trade-off however between the demands on the data and the connection to theory. The approach we follow is fairly flexible in the sense that one can easily add the type of questions we use to existing surveys but it also has a connection to theoretical work on the formation of social ties. We are able to capture weak ties which is important for our purposes (Hong, Kubik, and Stein (2004)) but often difficult to achieve when using measures of network centrality.

[2]A related idea follows from the model in Hong, Kubik, and Stein (2004). While not the focus of the paper, the authors show that the marginal effect of sociability is higher in areas with a higher density of stock market participants.

Literature Our paper fits into the literature on stock market (non)participation. As pointed out by Campbell (2006), it is still an open issue why so many households do not hold stocks. Haliassos and Bertaut (1995) suggested that income risk, inertia and departures from expected-utility maximization could partially explain this. The past decade the literature has emphasized behavioral, social and psychological explanations. Several papers point out that investor characteristics play a major role, even after controlling for a wide range of demographic characteristics. One strand of this literature considers variables related to (mental) ability. Grinblatt, Keloharju, and Linnainmaa (2011) explicitly point to investor IQ, others focus more broadly on cognitive ability e.g. Benjamin, Brown, and Shapiro (2013), Christelis, Jappelli, and Padula (2010). Barnea, Cronqvist, and Siegel (2010) argue that risk preferences may have a genetic component and find that a genetic component explains one third of the the cross-sectional variation of investor behavior of a sample of identical, fraternal twins. Related is the literature on financial literacy. This literature shows that financial illiteracy can be a major hurdle for investing in the stock market (Van Rooij, Lusardi, and Alessie (2011)) and is also related to a range of other financial decisions (Van Rooij, Lusardi, and Alessie (2012), Lusardi and Mitchell (2014)). Another strand of this literature considers a range of psychological explanations. Guiso, Sapienza, and Zingales (2008) suggest that trust matters as the perception of risk is a function of objective characteristics of stocks *and* subjective investor characteristics. Kaustia and Torstila (2011) find that political orientation is also related to stock market participation. Puri and Robinson (2007) propose a measure of optimism which correlates with beliefs about future economic conditions. They find that optimists tend to invest more in individual stocks. Malmendier and Nagel (2011) point to past experiences. Individuals who experienced low stock market returns are less likely to invest in stocks. Andersen, Hanspal, and Nielsen (2014) show that bad experiences result in lower subsequent risk raking.

A last strand in the literature acknowledges the importance of peer effects, social networks and social interactions. Hong, Kubik, and Stein (2004) and Georgarakos and Pasini (2011) consider sociability. Sociability is believed to reduce fixed participation costs through cheaper information sharing or observational learning. This is related to the idea that social interactions allow for word-of-mouth diffusion of information (Jackson (2010)). Hong, Kubik, and Stein (2004) provide empirical evidence that there is a higher participation rate among social investors compared to non-social investors. The authors use data from the Health and Retirement Survey and proxy for sociability by asking respondents about whether they go to church or whether they know their neighbours. While an important contribution in the literature on stock market participation, the research design in Hong, Kubik, and Stein (2004) reveals an issue when analyzing the influence of (latent) social networks. We typically do not have data on these networks because collecting these data is not feasible. Hong, Kubik, and Stein (2004) circumvent this issue by using proxies of sociability such as knowing neighbors or attending church. Such proxies capture relevant dimensions of people's social lives but are fairly crude measures. Another approach to capturing network effects is to make use of special groups and asset classes observed by peers, e.g. Duflo and Saez (2002). A related approach is to relate the investments of individuals to the investments of neighbors or the local community directly, see Ivković and Weisbenner (2007) or Brown, Ivković, Smith, and Weisbenner (2008). Ivković and Weisbenner (2007) find that neighbors have larger influence in more social states in line with the findings of Hong, Kubik, and Stein (2004). Our measure complements this literature by providing an approach to measure with a number of advantages. It is flexible and can be added to existing surveys, therefore it does does not depend on a specific setting. Moreover it also captures weak ties and is connected to a fairly general model of social ties.[3] The network variable we construct is an individual level variable and hence sheds information on micro characteristics of networks, Jackson (2014).

[3] The approach used in this paper allows the researcher whether to focus on weak or on strong ties. DiPrete, Gelman, McCormick, Teitler, and Zheng (2011) use a similar approach to analyze the differences between weak and strong ties.

This paper proceeds as follows. In Section 1 we discuss the construction of our proximity measure in detail. In Section 2 we introduce our data. In Section 3 we present the results of our analysis. In Section 4 we conclude.

1 Modeling social structure and the formation of ties

In this section we introduce the model we estimate to obtain the proximity variable i.e. closeness of individuals to certain subpopulations. This model was developed by Zheng, Salganik, and Gelman (2006) and generalizes a celebrated result by Erdős and Rényi (1959).

The intuition of the model is as follows. We model the answers of people on questions of the type *"How many X's do you know?"*, which are counts, with a negative binomial model. The model has three types of parameters. First we have individual specific parameters which capture network size. Then we have question specific parameters which capture (1) the size of subpopulation X and (2) the overdispersion one can expect in the answers on subpopulation X. Combined this model assumes that the answers to *"How many X's do you know?"* can be modeled as depending on how many people someone knows, how many X's there in fact are in the population and the propensity to know people in group X. To identify the parameters we ask people how many people they know with a certain first name. From administrative data we know how many people there in fact are with that surname and hence we can establish the individual parameters capturing the network size. To make this explicit, suppose that you know two persons named Kim in a population where 0.1 % has that name. A simple estimate would be that you know $2/0.001 = 2000$ people. Average the estimates across different subpopulations then yields a more precise estimate. This procedure assumes that everyone has an equal propensity to know someone from a given subpopulation. The model outlined below relaxes this.

We also ask questions on subpopulations on which we have less information. Assume you want to know how many homeless people there are. If we have determined that someone knows 2000 people and that person knows 1 homeless person then a crude estimate would be that about 0.05% of the population is homeless. Average over a random sample of individuals drawn from a population and we get a more precise estimate of the number of homeless people. Here we make a few tacit assumptions but a similar idea is employed in our procedure.

Our model is a model of how social ties are formed. We estimate this model and then use this model to create proximity variables. These proximity variables are the residuals of our model or the difference between the answers on the questions and the model predictions. If this residual is positive, it means that you know more people in a given subcategory than could be deduced based on your network size, the subpopulation size and your propensity to know people of that subpopulation. We interpret this proximity variable as how close you are to the subpopulation of interest. In other words, how important that subpopulation is in your social network.

This is the intuition behind our approach. Before introducing the data in Section 2, we go into greater detail about the model for the interested reader.

If this intuitive explanation is sufficient, you may skip the remainder of this subsection and go directly to Section 2. The remainder of this Section spells out the details of our model and estimation procedure.

1.1 Notation

We have a population of size N with $k = 1, \ldots, K$ subpopulations S_k each of size N_k. We write the probability that individual i knows individual j in the population as: p_{ij}. Summing this probability across all people in the population gives the *expected network size* of person i:

$$a_i = \sum_{j=1}^{N} p_{ij}. \tag{1}$$

The expected number of people known by person i in subpopulation S_k is written as:

$$\lambda_{ik} = \sum_{j \in S_k} p_{ij}. \tag{2}$$

We write the proportion of connections involving a member of subpopulation k as:

$$b_k = \frac{\sum_{i \in S_k} a_i}{\sum_{i=1}^{N} a_i}. \tag{3}$$

The numerator in this fraction is the proportion of *total links* involving subpopulation k whereas the denominator is sum of all individual network sizes. Individual i's relative propensity to known a person in group k is then:

$$g_{ik} = \frac{\lambda_{ik}}{a_i b_k} = \frac{\sum_{j \in S_k} p_{ij} / \sum_{j=1}^{N} p_{ij}}{b_k}. \tag{4}$$

The last equality is obtained by substution of the expression for a_i. This reveals that g_{ik} is the ratio of the proportion of links involving group k in individual i's network over the proportion of links that involve group k in the entire population. Therefor, if $g_{ik} > 1$ then individual i has a higher propensity to form ties with individuals from group k than the average person in the population, hence the name *relative propensity*.

1.2 Model

In the classical model by Erdős and Rényi (1959) ties between individuals i and j are believed to be formed completely at random. This means that the probability of a tie between person i and j is equal for all pairs (i, j). This leads to equal gregariousness parameters for all persons, $a_i = a$ for all i and relative propensities $g_{ik} = 1$ for all i, k.

The generalized model by Zheng, Salganik, and Gelman (2006) relaxes this setup in two important ways. First, the model allows variation in the gregariousness parameters a_i. This means that individuals in our sample can have different propensities to know other people. Differences in the network sizes of people can be very large. For example politicians or religious leaders may have acquaintance networks in the thousands whereas the average network size, using a similar definition as we do in this paper, is around five hundred fifty in the United States (DiPrete, Gelman, McCormick, Teitler, and Zheng (2011)).

Second, we allow for variation in the relative propensity to know people in different groups, g_{ik}. In statistical notation we write this model as:

$$y_{ik} \sim \text{Poisson}(\lambda_{ik}), \quad \lambda_{ik} = a_i b_k g_{ik}. \tag{5}$$

Zheng, Salganik, and Gelman (2006) label this model the *overdispersed* model, because the variation in g_{ik} results in overdispersion of the count data.

1.3 Model Estimation

We rewrite the overdispersed model given in equation 5 as:

$$y_{ik} \sim \text{Poisson}(\exp(\alpha_i + \beta_k + \gamma_{ik})), \tag{6}$$

with $\alpha_i = \log(a_i)$, $\beta_k = \log(b_k)$, $\gamma_{ik} = \log(g_{ik})$. We assume that for each subpopulation S_k, the multiplicative factor $g_{ik} = \exp(\gamma_{ik})$ follows a gamma distribution with a mean of 1 and a shape parameter of $1/(\omega_k - 1)$. This choice allows the γ parameters to be integrated out and yields the following overdispersed model:[4]

$$y_{ik} \sim \text{Neg.Bin.}(\text{mean} = \exp(\alpha_i + \beta_k), \text{overdispersion} = \omega_k). \tag{7}$$

Higher values of ω_k correspond to more overdispersion. The intuition for overdispersion is that the frequency of people who know exactly 1 person in subpopulation k as compared to the frequency of people who know none in subpopulation k decreases. So the more overdispersion, the less likely it is for a person to have *an isolated acquaintance* in that subpopulation.

The parameters α_i, β_k and ω_k are estimated with a hierarchical, Bayesian approach. The α parameters are assumed to follow a normal distribution with mean μ_α and standard deviation σ_α, corresponding to a lognormal distribution for the gregariousness parameters. Zheng, Salganik, and Gelman (2006) point out that this is in line with existing research on acquaintanceship networks. The β parameters are also assumed to follow a normal distribution with μ_β and standard deviation σ_β. Following Zheng, Salganik, and Gelman (2006) we assign to the overdispersion parameters ω_k independent uniform$(0, 1)$ priors on the inverse scale. This constrains the overdispersion parameters to the range $(1, \infty)$. We assign weakly informative priors to the hyperparameters $\mu_\alpha, \sigma_\alpha$. All (hyper)priors are provided in the appendix. The joint posterior density can be written as:

$$\begin{aligned} p(\alpha, \beta, \omega, \mu_\alpha, \mu_\beta, \sigma_\alpha, \sigma_\beta | y) \propto & \prod_{i=1}^{n} \prod_{k=1}^{K} \binom{y_{ik} + \xi_{ik} - 1}{\xi_{ik} - 1} \left(\frac{1}{\omega_k}\right)^{\xi_{ik}} \left(\frac{\omega_k - 1}{\omega_k}\right)^{y_{ik}} \\ & \times \prod_{i=1}^{n} N(\alpha_i | \mu_\alpha, \sigma_\alpha^2) \prod_{k=1}^{K} N(\beta_k | \mu_\beta, \sigma_\beta^2), \end{aligned} \tag{8}$$

with $\xi_{ik} = \frac{\exp(\alpha_i + \beta_k)}{\omega_k - 1}$, (Zheng, Salganik, and Gelman (2006)).

Normalization The model discussed above is not identified. Any constant C can be added to all of the α_i's and substracted from all of the β_k's without affecting the likelihood. We identify the model by providing tight priors on some parameters β_k such that these represent the proportions of links in the population that go to members of subpopulation k. In this paper we will renormalize to names of which we know the proportions in the general population.

Fitting The model is estimated using the No-U-Turn Sampler due to Hoffman and Gelman (2014) and implemented in the probabilistic programming language Stan (Stan Development Team (2014)).

1.4 Assumptions

The approach outlined above builds on a range of assumptions. Some of which may be less obvious to the reader. We discuss three assumptions as to be transparent about what the above approach entails.

[4] This disussion follows Zheng, Salganik, and Gelman (2006). A more detailed treatment of this Poisson-gamma mixture is given by Cameron and Trivedi (2005) chapter 20, p.674-677.

- We assume that respondents do not systematically know more or less individuals than would be expected under random mixing. A violation of this assumption is referred to as barrier effects. For example, young people might know more people with a certain name than older people solely because that specific name is more popular in their cohort. This could lead to an overestimation of the network size of young people and an underestimation of the network size of older people. McCormick, Salganik, and Zheng (2010) thoroughly analyzed this problem and provided a few guidelines to alleviate this problem. These guidelines are: (1) balance gender (do not only ask male or female names), (2) balance popularity of names (choose names which were popular sixty years ago as well as names which were popular thirty years ago), (3) choose names which compromise about 0.1%-0.2% of the population (this is mainly to alleviate recall error, see below), (4) avoid names which are typically used as nicknames. The authors show in a simulation study that when these guidelines are followed that barrier effects are negligible. Moreover 12 names are sufficient (McCormick, Salganik, and Zheng (2010)). In our design we took all these considerations into account. An overview of the names can be found in the appendix.

- We assume that respondents are fully aware that a given acquaintance belongs to a certain subpopulation. For example, a respondent may have an acquaintance in his network who works as a private banker but he is unaware. A violation of this assumption is called transmission error. This error is very hard to quantify but substantially this is *irrelevant* for our purposes for two reasons. First, of the 16 ARD questions we are going to use in this paper, 12 refer to first names. We explicitly define *to know someone* and one of the criteria is knowing that persons name. By definition, this error can not apply to these 12 questions. The remaining four questions refer to professions and traits and here we could have a transmission error. However, our interest in this study is in constructing measures of social exposure (see further). If you are not aware that someone possesses a certain trait then it is less likely that you are exposed to that trait to the extent that it influences you. Typically the transmission error matters mostly when estimating network size based on questions where a transmission error is likely, McCormick, Salganik, and Zheng (2010). In this paper we follow McCormick, Salganik, and Zheng (2010) by calibrating our parameters using first name questions as to avoid this issue.

- We assume perfect recall. Each respondent is fully able to recall all acquaintances in subpopulation S_k. This is unlikely to be true. In particular there is evidence that respondents tend to underrecall people they know in large subpopulations and overrecall people in small subpopulations, see Killworth, McCarty, Bernard, Johnsen, Domini, and Shelley (2003), Zheng, Salganik, and Gelman (2006), McCormick, Salganik, and Zheng (2010). For the name questions, the solutions lies in choosing names which are not too popular nor too impopular. There is a consensus that names should be chosen that comprise about 0.1%-0.2% of the population, see Zheng, Salganik, and Gelman (2006), McCormick, Salganik, and Zheng (2010). We follow that consensus and we use these names to identify the parameters of our model. These are also the questions we use to identify the parameters in our model.

Having addressed the assumptions we are ready to discuss the data.

2 Data

Our data source is the DNB Household Survey (or DHS) to which we added a module with questions to construct proximity variables for which we provided some intuition at the beginning of Section 1 and which is defined by equation 9 (see further). This survey is conducted by CentERdata at Tilburg

Table 1: Matching of different CentER data surveys.

	#Household heads	Age	Education cat.	Male
Networks data	1377	56.44	4.04	0.71
Drop answers over 100	1350			
Drop no variability network data	1329			
Drop missing network data	1329			
Drop missing network data	1329			
DHS data	2054	52.70	4.10	0.72
Matched sample with complete data	1043	57.50	4.04	0.72

This table shows the number of households as well as mean age, mean education derived from a categorical variables where higher values correspond with a higher level of education (categories are defined by the Central Bureau of Statistics), and mean fraction of male household heads in our data before and after the matching proces.

University and provides high quality data on household saving behavior.[5]

We combine our customized survey with other data from the household panel. In line with the literature (e.g. Van Rooij, Lusardi, and Alessie (2012)) we use responses from household financial decision makers. They are considered to be most relevant for household financial decision making (Smith, McArdle, and Willis (2010)). The implications of the data cleaning and matching procedure for our sample are presented in Table 1. We report the number of household financial decision makers and three general demographic characteristics that were fully available in all datasets used, namely age, education categories (higher values correspond with higher levels of education), and gender. Our survey was sent out in November 2014 and was completed by 1377 household heads. We then cleaned the questions from the customized survey as described in Subsection 2.1 (see below) and are left with 1329 individuals. Our control variables are taken from two sources, the DHS data from 2014 with financial and socio-economic information and an extra survey question on generalized trust sent out early 2014 (weeks 2 and 13). These data entail 2054 households. Finally we construct a sample of 1043 household heads with complete data on stock market participation, network information and background variables. When looking at the characteristics of the household heads, we observe that the average respondent in our final dataset is about five years older than the average household head in the complete 2014 DHS data. Mean gender and education are very similar. These numbers suggest that our matching procedure results in a dataset that largely retains the representativeness of the DHS.

In our analyses we test if network characteristics influence the decision to hold stocks. To this end we create an indicator for stock market participation which is one if the household owns individual stocks, mutual or growth funds, or holds options. Of the individuals in our sample 20% participates in the stock market, see Table 2. [6]

A number of facts on the participation in the stock market by households are well documented, see

[5]Additional information on the DNB Household Survey and the data collection process is provided in the appendix.

[6]For an international comparison of stock market participation rates, see Giannetti and Koskinen (2010). The Netherlands seem to fall in the middle of the countries surveyed. Note that the Dutch have compulsory defined benefit pension schemes that invest money on their behalf. In this study we investigate the active decision for stock market participation by households.

Table 2: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Proximity	1043	.018	.992	-4.755	5.446
Stock market	1043	.2	.4	0	1
Male	1043	.724	.447	0	1
Age	1043	57.5	15.573	25	91
TotalAssets	1043	248534.7	222129.8	0	2107407
Primary school	1043	.031	.173	0	1
Lower secondary	1043	.214	.41	0	1
Higher secondary	1043	.108	.311	0	1
Lower vocational	1043	.152	.36	0	1
Higher vocational	1043	.318	.466	0	1
University	1043	.176	.381	0	1
Optimism	1043	3.51	1.042	1	5
Household size	1043	2.166	1.124	1	8
Partner	1043	.672	.47	0	1
Retired	1043	.366	.482	0	1
Self employed	1043	.039	.194	0	1
Rural	1043	2.953	1.31	1	5
Income $<$ 1151	1043	.067	.25	0	1
Income 1151 - 1800	1043	.165	.371	0	1
Income 1801 - 2600	1043	.285	.452	0	1
Income $>$ 2600	1043	.483	.5	0	1
Religion none	1043	.449	.498	0	1
Religion Catholic	1043	.273	.446	0	1
Religion Protestant	1043	.186	.389	0	1
Religion Other	1043	.092	.289	0	1
Generalized trust	1043	6.273	2.119	0	10
Risk tolerance self	1043	13.65	4.948	5	29

This table shows summary statistics for the variables used in the regression analysis. The variable *Proximity* is the Proximity measure given by equation 10 which we subsequently normalized.

for example Bertaut and Starr (2000), Vissing-Jorgensen (2002), Hong, Kubik, and Stein (2004), Guiso, Sapienza, and Zingales (2008). The DNB household survey is fairly rich and allows us to control for the important determinants identified in the literature (e.g. wealth) as well as commonly used control variables. In Table 2 we provide summary statistics for these variables. Most of these variables are self-explanatory but a few require further elaboration.

First, we want to control for trust, following the work of Guiso, Sapienza, and Zingales (2008) who show that trust has a positive impact on stock market participation. Our generalized trust question is similar to theirs: *"Generally speaking, do you feel that most people can be trusted or that you have to be very careful in dealing with people?"*.

Second, we include a risk tolerance variable following the literature, see Guiso, Sapienza, and Zingales (2008). Risk tolerance is constructed from respondents' assessments of 5 statements (on a 1 to 7 scale) concerning taking financial risks which are not directly related to stock market investments. We invert the answers of two questions so that all variables measure risk tolerance and define "Risk tolerance self" as the sum of the individual constructs.

Third, we also use a variable capturing the optimism of respondents. Puri and Robinson (2007) show that optimism is related to various work and life choices as well as investment behaviour. The DNB household survey asks respondents to evaluate themselves by providing a number of statements where respondents have to indicate to what extent they agree (on a 1 to 5 scale). To capture optimism we use the statement: *"I seldom feel blue."*.

2.1 Proximity variables

The focus of this paper are the proximity measures we develop by estimating overdispersed models as described in section 1. To do this we asked a range of questions to the survey participants. First the respondents were informed that we would ask them about people they know. We explained that: *"Knowing means that you know that persons name and you would give a sign of recognition when you ran into this person. Please limit yourself to people who currently live in the country and who you expect to be aged sixteen or older."* This definition was repeated throughout this part of the survey.
The survey asks respondents how many people they know with certain first names as well as how many people they know with certain characteristics. These questions are:

How many people do you know who

- *actively invest.*
- *have a lot of knowledge of financial matters.*
- *work for a financial institution such as a bank, a pension fund or an insurance company.*
- *have a private banker.*

We use these sixteen questions (the twelve name questions and the four above) to construct our proximity variables.[7]

In line with previous surveys using this type of data there were some minor irregularities. We cleaned the data as follows. Respondents who provided implausibly large answers (more than 100) on one of the questions were dropped. For the remaining respondents we truncated the answers at 30, following Zheng, Salganik, and Gelman (2006). As a robustness check we changed the truncation point to 50, which had no noticeable effect on the results. We also removed respondents who answered each question with the same number.[8]
Then we estimated the model given by equation 7. We identified the model by restricting the β_k parameters corresponding to the names to match the true population proportions.[9] The β_k parameters corresponding to the other ARD questions are given uninformative priors around the mean of the questions for which we know the proportions. An overview of the proportions and exact prior specifications is also provided in the appendix.
The overdispersed model as presented in equation 5 is a stylized model of how social ties are formed. The probability of knowing someone in a subpopulation depends on one's network size, the size of that subpopulation and the propensity to know someone in a given subpopulation. The model we estimate, presented in equation 7, is therefor parameter rich. To get a feel for the estimation results we present in Figure 1 the distribution of estimated network sizes. This histogram resembles the results found for the distribution of individuals network sizes in the United States. DiPrete, Gelman, McCormick, Teitler, and Zheng (2011) report that the median person in the United States is acquainted with 550 people -based on an estimation of the same model, with a similar definition of *to know* and General Social Survey data. We find that an invidual knows on average 534 people in the Netherlands. The interquartile

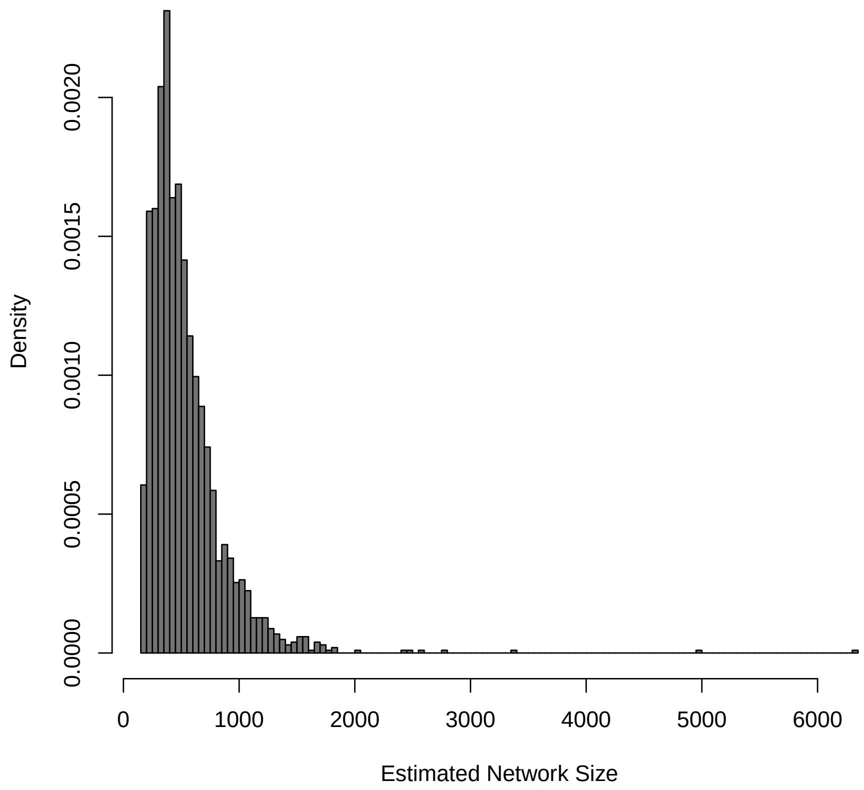
[7] The survey contains other ARD questions. In total there were 30 ARD questions asked of which we use 16 in this paper. The other questions pertain broadly to issues such as drug use, environmental concerns and are not related to the topic of this paper.

[8] To estimate the parameters from which we construct our proximity measure we use the entire sample including people who are not financial decision makers in the household.

[9] We obtained the true proportions from the Meertens Institute. This research institute on Dutch language and culture maintains, with support from the Royal Dutch Scientific Academy (KNAW), a database of names in the Netherlands. This database is constructed from administrative records and provides a complete overview of birth names of Dutch citizens. Researchers can make specific queries such as the number of people with a certain name, born in a certain year etc.

range we find 330-644, is however a bit smaller than what DiPrete, Gelman, McCormick, Teitler, and Zheng (2011) report for the United States i.e. 400-800. The far right tail with some people knowing several thousand people also align with results reported in DiPrete, Gelman, McCormick, Teitler, and Zheng (2011) or Zheng, Salganik, and Gelman (2006). People with such large acquaintance networks can be predominantly found in professions which require to interact with many people on daily basis such as certain religious or political positions.

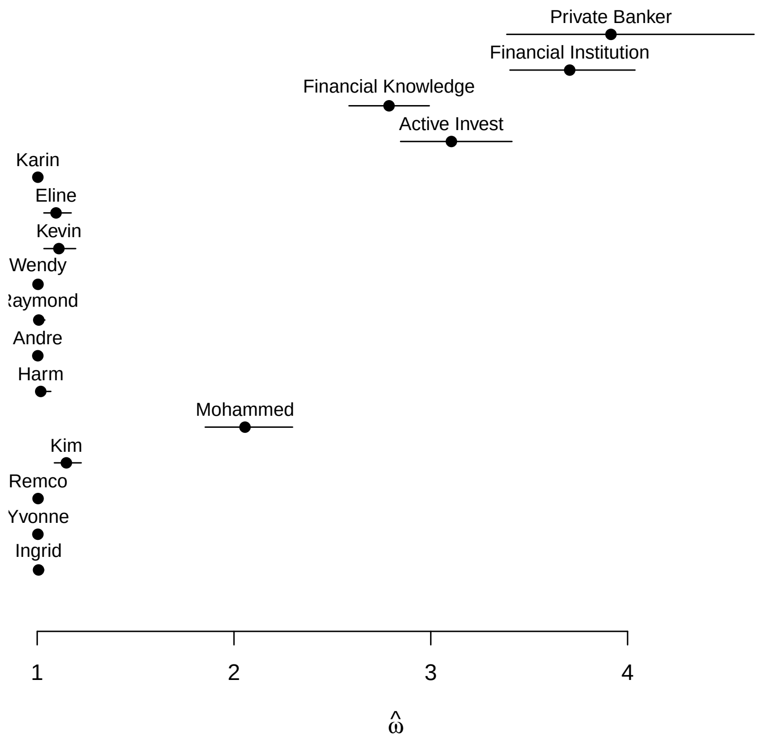
Figure 1: Network sizes



The histogram shows the distribution of the estimation individual network sizes.

In our survey we have added four questions probing at the proximity to subpopulations with an above average knowledge of financial affairs. We expect substantial variation in the relative propensity to form ties with these groups. Figure 2 shows the point estimates $\hat{\omega}_k$ and 95% confidence intervals around the point estimates for these categories. Figure 2 reveals that the estimated overdispersions are close to 1 for all names except Mohammed, indicating that there is little variation in the relative propensity to form ties with people having these names. This is in line with what we would expect. In the Netherlands, the name Mohammed is ethnically loaded. The name is relatively common among immigrants of Arabic origins and so a higher overdispersion parameter makes sense. For our four other groups the estimated overdispersion parameters are even larger suggesting that having an isolated acquaintance of those groups is even less likely and hinting at the existence of social networks.

Figure 2: Dispersion parameters



This plot shows the estimates of the dispersion parameters for all the variables we use in the estimation of the model given by equation 7. The dots are the point estimates and the lines represent the 95% confidence intervals.

The goal of estimating model 7 is not so much the parameters per se but rather a proximity variable that we can construct with these parameters. Following a suggestion in Zheng, Salganik, and Gelman (2006) and McCormick, Moussa, Ruf, DiPrete, Gelman, Teitler, and Zheng (2013) we construct a proximity variable r_{ik}:

$$r_{ik} = \sqrt{y_{ik}} - \sqrt{a_i b_k}, \tag{9}$$

where y_{ik} is the answer by respondent i on subpopulation k, α_i is the network size of individual i and b_k is the size of subpopulation k. This proximity variable captures how much more (or less) people you know then what could be expected based on our model of social contacts.[10]
Throughout this paper we use a proximity index which is the average of the proximity variables constructed from the four categories:

$$\text{Proximity Index}_i = \sum_{k=1}^{4} r_{ik} \tag{10}$$

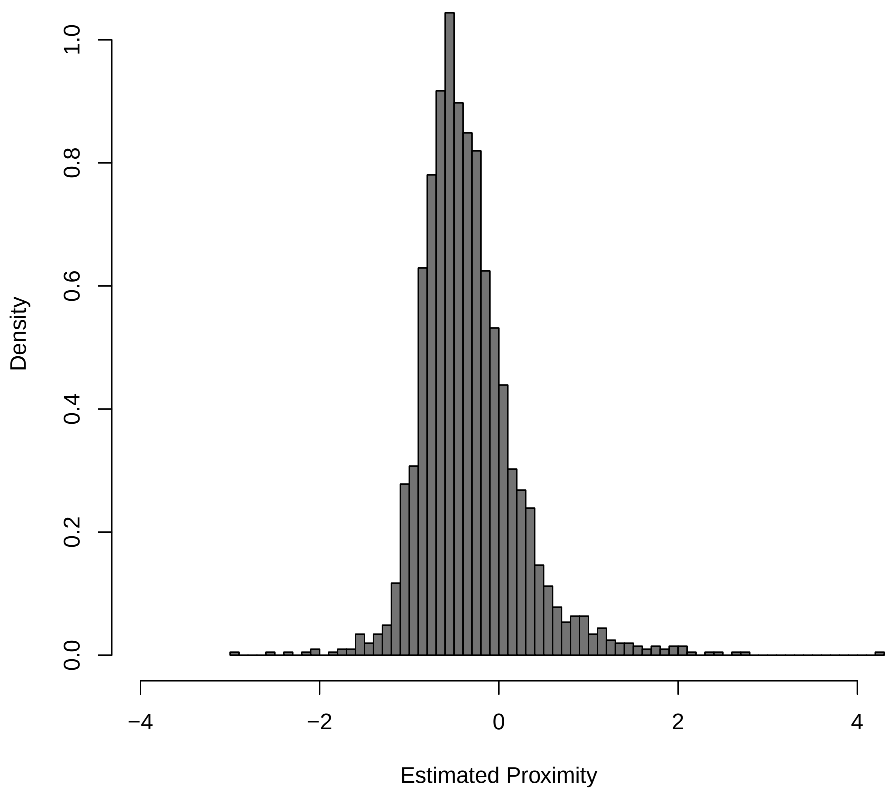
[10]Computing the residuals on a square root scale is commonly done with count data in order to stabilize the variance. See also Zheng, Salganik, and Gelman (2006) who provide a further reference.

where r_{ik} is given by equation 9. The proximity index gives a sense of how financial savvy people's acquaintance networks are. We could as well have asked questions on how many portfolio managers someone knows, how many members of an investment club, how many people with a finance degree, etc. We have chosen these questions as we expected some variation in the extent to which these categories are overdispersed.

These categories also capture different aspects. The question on financial knowledge probes to the perception by individuals of their networks. The question about people who are active investors refers to activities undertaken by acquaintances. The questions about people working at financial institutions and people working as private bankers captures acquaintances with active in the financial industry. The former is a broad group of professions and encompasses many people whereas the latter is much more narrow and focused measure of a social network of finance professionals.

In Figure 3 we show the distribution of the proximity index. The distribution shows a slight right skew with the bulk of the mass below zero. In our empirical analysis we normalize this proximity variable (see also Table 2) and we use trichomotized versions of this variable which facilitate interpretation.

Figure 3: Proximity



This plot shows the distribution of the proximity index calculated according to equation 10.

3 Results

3.1 Baseline results

The aim of this paper is to shed light on the connection between holding stocks and one's social network. Underlying this aim is the puzzle that relatively few people hold stocks, an important *puzzle* in the literature on household finance, see Van Rooij, Lusardi, and Alessie (2011). In Figure 4 we present stock ownership for various demographic groups. The Figure highlights various stylized facts on stock ownership. Men tend to hold stocks more often than women. Stock ownership goes up as one ages and stock ownership goes up as one has more wealth. Education also matters, the higher educated the more likely one holds stocks. The degree of urbanization where someone lives does not seem to play an important role.[11] Religious denomination plays a minor role. In particular Catholics hold stocks more often than Protestants, other religious denominations or non-religious people. At the bottom of Figure 4 we see that few people in the lower tertile of our proximity measure (see Section 1) hold stocks. As we move to the upper tertile, stock ownership shoots up. This is suggestive of the link between characteristics of one's social network and stock ownership. The next step in our analysis is to verify whether this remains after controlling for a range of important covariates.

In Table 3 we show estimates from multivariate analyses using 4 specifications. The first two are OLS and Logit models that relate an indicator for stock market participation to our normalized proximity variable as well as a battery of control variables. The second two are OLS and Logit models where we have replaced the proximity variable by two dummy variables capturing individuals in the middle and upper tercile of the proximity variable. Across the four specifications age seems to matter as well as wealth. Larger households seem less likely to hold stocks whereas more trusting and more risk tolerant people are more likely to hold stocks. The effects of household size, trust and risk tolerance seem small. Education has no effect conditionally. Note that all these specifications contain dummies for different income levels, religious affiliation and different degrees of urbanization.

Our normalized proximity measure is positively related and significant at the 1% level considering both OLS and logit models. In columns (5) through (8) we see that being in a higher tercile of the proximity measure is positively related to the probability of owning stocks. The results from the logit analysis show that going from the first to the second tercile increases stock market participation by 6% while moving from the lowest tercile to the upper tercile increases stock market participation with 12%.

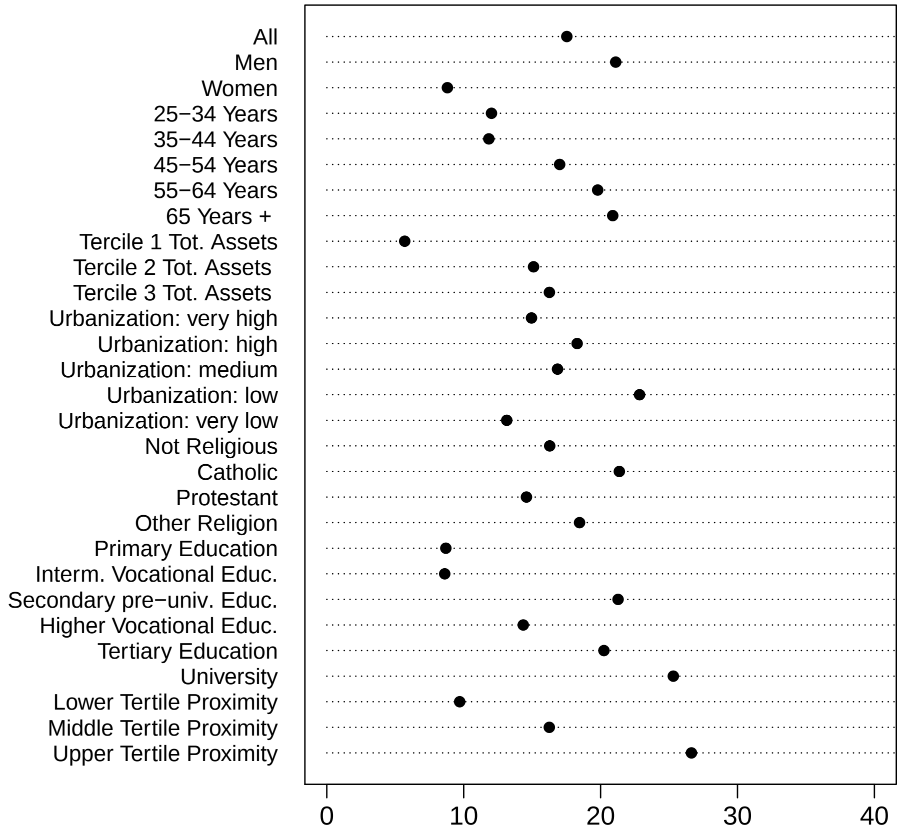
Taken together our results strongly show that having a more financially savvy network is positively associated with stock market participation either directly or through mutual and growth funds. Moreover, the economic effect is substantial.

The results so far concern the entire sample. One aspect of the stock-holding puzzle is that even among the wealthier part of the population stock market participation is fairly low. This can be seen in Figure 4. In Table 4 we show the regression results when restricting the sample to the bottom, middle and top wealth tercile respectively. For the bottom tercile the effects of knowing financial savvy people in your network is small and statistically insignificant. People in this group have total assets (excluding stock market holdings) of less then 158,000 euro.[12] Not many people in this group have stock market investments (see Table 4), presumably due to financial constraints. Households in the middle tercile have total assets between 158.000 euro and 285.000 euro. For this group we find an effect in line with the results presented in Table 3. For people with total assets above 285.000 euro the importance of having a financial savvy network seems even higher.

[11] The Netherlands is fairly small country with a high population density. Even though rural areas exist, people living in these areas are fairly close to more urbanized places and vice versa. So this finding may not hold in larger countries such as the U.S. where the labels *very highly urbanized* and *very lowly urbanized* may well proxy for other things.

[12] In the construction of total assets we follow Von Gaudecker (2015).

Figure 4: Stock market participation across subgroups



This Figure shows the Stock market participation across different levels of gender, age, wealth, urbanization, religion, education and proximity.

Recent work in the literature on stock market participation stresses the role of trust, see Guiso, Sapienza, and Zingales (2008). Less trusting individuals are less inclined to buy stocks. This raises the question whether the impact of one's social network holds for trusting people. In Table 5 we present regression results when restricting the sample to individuals belonging to Low, Mid, and High terciles of generalized trust. The results provide interesting insights into the effects of networks on financial decision making. According to our findings, the effects of being exposed to financially savvy people are more pronounced for individuals with middle to high levels of trust. Put differently, distrusting individuals are less likely to let people they know influence their financial choices. These findings imply that even for individuals who are expected to be more inclined to participate in the stock market, people with elevated trust levels, one's social network predominantly plays a role in deciding to participate in the stock market.

Table 3: Baseline results

VARIABLES	(1) OLS	(2) se	(3) Logit	(4) se	(5) OLS	(6) se	(7) Logit	(8) se
Proximity	0.04***	(0.01)	0.03***	(0.01)				
Proximity Mid 1/3rd					0.05*	(0.03)	0.06*	(0.03)
Proximity High 1/3rd					0.11***	(0.03)	0.12***	(0.03)
Male	0.09***	(0.03)	0.08***	(0.02)	0.09***	(0.03)	0.08***	(0.02)
Age 35 - 44	0.08*	(0.04)	0.10	(0.08)	0.07*	(0.04)	0.09	(0.07)
Age 45 - 54	0.12**	(0.05)	0.14	(0.09)	0.12**	(0.05)	0.14	(0.08)
Age 55 - 64	0.16***	(0.04)	0.20**	(0.08)	0.16***	(0.04)	0.19**	(0.08)
Age $>$ 65	0.18***	(0.06)	0.21**	(0.08)	0.17***	(0.06)	0.20**	(0.08)
Assets 158001 - 285000	0.07**	(0.03)	0.09**	(0.04)	0.07**	(0.03)	0.09**	(0.04)
Assets $>$ 285000	0.16***	(0.03)	0.17***	(0.04)	0.16***	(0.03)	0.16***	(0.04)
Lower secondary	-0.02	(0.06)	-0.04	(0.06)	-0.01	(0.06)	-0.02	(0.07)
Higher secondary	0.08	(0.07)	0.08	(0.10)	0.08	(0.07)	0.10	(0.10)
Lower vocational	0.06	(0.06)	0.05	(0.09)	0.06	(0.06)	0.06	(0.09)
Higher vocational	0.08	(0.06)	0.08	(0.08)	0.09	(0.06)	0.09	(0.08)
University	0.09	(0.07)	0.08	(0.09)	0.10	(0.07)	0.10	(0.10)
Optimism	0.01	(0.01)	0.01	(0.01)	0.01	(0.01)	0.01	(0.01)
Household size	-0.03**	(0.01)	-0.03*	(0.01)	-0.03**	(0.01)	-0.03**	(0.01)
Partner	-0.03	(0.04)	-0.03	(0.04)	-0.03	(0.04)	-0.02	(0.04)
Retired	-0.03	(0.05)	-0.04	(0.04)	-0.03	(0.05)	-0.04	(0.04)
Self employed	0.01	(0.07)	0.01	(0.06)	-0.00	(0.07)	-0.00	(0.06)
Generalized trust	0.01*	(0.01)	0.01*	(0.01)	0.01*	(0.01)	0.01*	(0.01)
Risk tolerance self	0.01***	(0.00)	0.01***	(0.00)	0.01***	(0.00)	0.01***	(0.00)
Constant	-0.31***	(0.10)			-0.36***	(0.10)		
Observations	1,043		1,043		1,043		1,043	
Income dummies	Yes		Yes		Yes		Yes	
Urbanization dummies	Yes		Yes		Yes		Yes	
Religion dummies	Yes		Yes		Yes		Yes	
(Pseudo) R-squared	0.173		0.182		0.174		0.186	

This table shows results from regressions of stock market participation on proximity measures and control variables. Columns (1) and (3) show coefficients using the standardized proximity measure and columns (5) and (7) use indicator variables for terciles of the proximity measure where Low proximity is the baseline. In Columns (1) and (5) we report simple OLS results and in (3) and (7) we apply a logit model and report the marginal effects at the means. Controls include dummies for CentERdata groups of age, education, urbanization, religion, and income of which we suppress the coefficients of the latter three groups. We also include household size and dummies for gender, wealth groups, living with a partner, being retired, being self employed, optimism, self assessed risk tolerance and generalized trust. Heteroskedasticity robust standard errors are reported in parentheses. We indicate significance at 1%, 5%, and 10% by ***, **, and * respectively.

Table 4: Results reported by wealth tercile

VARIABLES	(1) Low	(2) Low	(3) Mid	(4) Mid	(5) High	(6) High
Proximity	0.02		0.06**		0.05**	
	(0.02)		(0.02)		(0.02)	
Proximity Mid 1/3rd		0.03		0.08*		0.02
		(0.03)		(0.04)		(0.06)
Proximity High 1/3rd		0.06		0.14***		0.13**
		(0.05)		(0.05)		(0.06)
Observations	329	329	347	347	367	367
Controls as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes

This table shows results from regressions of stock market participation on proximity measures and control variables by wealth tercile as indicated in the column headings. Columns (1), (3), and (5) show coefficients using the standardized proximity measure and columns (2), (4), and (6) use indicator variables for terciles of the proximity measure where Low proximity is the baseline. Controls are suppressed because they are similar to those in Table 3 excluding wealth controls. Heteroskedasticity robust standard errors are reported in parentheses. We indicate significance at 1%, 5%, and 10% by ***, **, and * respectively.

Table 5: Results reported by terciles of Generalized Trust

VARIABLES	(1) Low	(2) Low	(3) Mid	(4) Mid	(5) High	(6) High
Proximity	0.03		0.07**		0.04	
	(0.02)		(0.03)		(0.02)	
Proximity Mid 1/3rd		-0.01		0.10		0.11**
		(0.04)		(0.06)		(0.05)
Proximity High 1/3rd		0.04		0.19***		0.15**
		(0.05)		(0.06)		(0.06)
Observations	442	442	282	282	319	319
Controls as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes

This table shows results from regressions of stock market participation on proximity measures and control variables by terciles of the generalized trust measure as indicated in the column headings. Columns (1), (3), and (5) show coefficients using the standardized proximity measure and columns (2), (4), and (6) use indicator variables for terciles of the proximity measure where Low proximity is the baseline. Controls are suppressed because they are similar to those in Table 3. Heteroskedasticity robust standard errors are reported in parentheses. We indicate significance at 1%, 5%, and 10% by ***, **, and * respectively.

3.2 Uncovering exogenous effects

3.2.1 Instrumental variables

So far we have relied on basic approaches OLS and logit estimates. Our results may suffer from endogeneity bias. One could argue that participation in the stock market leads to new acquaintances with financial knowledge and hence a higher value for our proximity variable. On the other hand we could have measurement bias. Nor the degree nor the direction of the bias is clear, see the discussion in Yoong (2011). In any case we would like to corroborate our results by means of instrumental variables. We have two variables which we consider as instrument. High financial literacy of the parents is thought to be exogenous to stock market participation as argued by Van Rooij, Lusardi, and Alessie (2011). Inflation literacy probes to understanding what inflation entails. Inflation literacy should not determine stock market participation but is likely to be correlated with our proximity variable. In the appendix we provide the financial literacy questions that we have access to and use in the analysis. We estimate a linear probability model which yields a heteroskedastic error term. For this reason we use Generalized Method of Moments estimation when we perform the instrumental variables estimation. The first stage results show statistically significant instruments and F-statistics which are high and in the recommended range to avoid weak instruments. The second stage estimates confirm the positive, significant effect of proximity on stock market participation. The Hansen J-test indicates that the null cannot be rejected and so there is no evidence that the overidentifying restrictions are invalid. The results of this analysis square with our earlier findings: having a network of acquaintances with many financial savvy people makes it more likely that one holds stocks.

Table 6: Results from instrumental variables analyses

VARIABLES	(1) 1st stage	(2) se	(3) 2nd stage	(4) se
Proximity			0.26**	(0.10)
Male	-0.10	(0.09)	0.04	(0.04)
Age 35 - 44	-0.30	(0.25)	-0.03	(0.18)
Age 45 - 54	-0.22	(0.24)	-0.03	(0.18)
Age 55 - 64	-0.19	(0.24)	0.02	(0.17)
Age $>$ 65	-0.21	(0.26)	0.03	(0.18)
Assets 158001 - 285000	0.09	(0.09)	0.05	(0.04)
Assets $>$ 285000	0.25**	(0.10)	0.07	(0.06)
Lower secondary	0.23	(0.17)	-0.20**	(0.08)
Higher secondary	0.29	(0.19)	-0.15	(0.10)
Lower vocational	0.31*	(0.17)	-0.14	(0.09)
Higher vocational	0.40**	(0.17)	-0.19*	(0.10)
University	0.74***	(0.20)	-0.27**	(0.13)
Optimism	0.02	(0.03)	-0.00	(0.02)
Household size	0.00	(0.04)	-0.03	(0.02)
Partner	-0.15	(0.11)	0.00	(0.06)
Retired	0.07	(0.12)	-0.05	(0.06)
Self employed	0.04	(0.20)	-0.04	(0.10)
Generalized trust	-0.00	(0.02)	0.01*	(0.01)
Risk tolerance self	0.03***	(0.01)	0.01**	(0.00)
Fin. literacy risk stocks	0.20	(0.14)	-0.00	(0.06)
Fin. literacy interest	0.03	(0.20)	0.04	(0.07)
Fin. literacy bonds	0.27***	(0.08)	0.04	(0.05)
Fin. literacy parents high	0.20***	(0.07)		
Fin. literacy inflation	0.36***	(0.12)		
Observations	646		646	
Income dummies	Yes		Yes	
Urbanization dummies	Yes		Yes	
Religion dummies	Yes		Yes	
(Pseudo) R-squared	0.231		0.0102	
F-stat instruments			9.253	
P of Hansen J			0.267	

This table shows results from Instrumental Variables regressions of stock market participation on proximity measures and control variables.

3.2.2 Placebo tests

In this section we investigate whether there are unobserved factors which influence our proximity measure and stock market participation for people with a similar socio-economic profile - which would result in endogeneity bias. To do this we partition our sample in a number of cells of people with similar social circles.[13] Within these cells we randomly assign the proximity measure of another individual in that cell. If it is the case that there are unobserved factors driving the results (preferences of individuals in the same social circle are correlated) we should expect to see a significant effect of the *reallocated proximity measure* on the probability of owning stocks. To capture the variability in the estimated coefficient due to the randomization, we repeated this process for each placebo test a 100 times. In Figure 5 we show the results of four different placebo tests. We have split the sample respectively according to (1) age, region and education, (2) age, region and wealth, (3) province and education, (4) age and wealth.[14] Each graph shows the estimated coefficient on the permuted proximity measure as well as the associated 95% confidence interval. We see that the estimated coefficients now tend to be small and that in almost all cases the confidence interval covers zero. We could have constructed different placebo tests based on different crossings of household characteristics. In the appendix to this paper we mention three additional placebo tests leading to the similar results.

3.2.3 Additional control variables

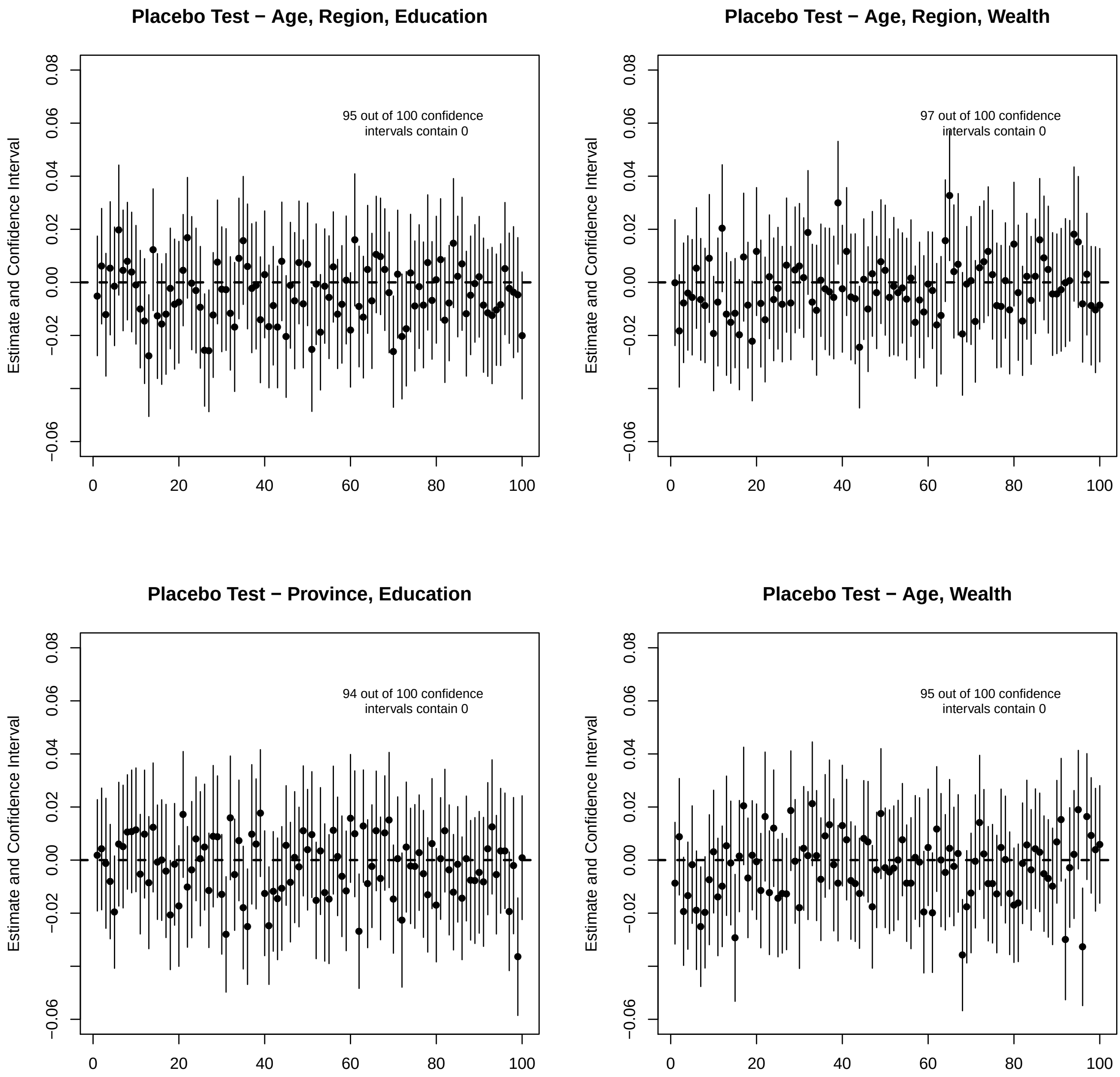
Another potential concern is that there are factors missing from our regression specification that influence both the decision to participate in the stock market and the explanatory variables. For this reason we include a battery of alternative control variables that are available in the DHS data. Some papers argue that financial literacy affects the decision to hold stocks as well as other financial decisions, see Van Rooij, Lusardi, and Alessie (2011) or Van Rooij, Lusardi, and Alessie (2012). In our data we have a a self-assessed indicator of financial literacy we add as an additional covariate to our baseline regression. We also consider future orientation because more future oriented individuals likely have a more long term financial planning and consequently a higher propensity to own stocks. Future orientation is constructed from 12 questions that measure the extent to which people consider distant versus immediate consequences of possible behaviour as in Strathman, Gleicher, Boninger, and Edwards (1994). Impatience could negatively influence stock market participation too. We follow Van Rooij, Lusardi, and Alessie (2012) and use variables that capture smoking and drinking behavior to proxy for myopic behavior.[15] We also control for bequest motive since bequest motives have been argued to influence the decision to hold stocks through retirement, see Abel (2001b). We generate an indicator for individuals who have indicated to have the intention of leaving a bequest (conditionally or unconditionally) following a DHS statement choice framework about bequest motives. Finally, we add a variable that measures if individuals have control over their expenditures since we argue that individuals need control over their finances to accumulate money and subsequently invest it in stocks. We create a measure for financial control using the answers to the question: *"Do you find it easy or difficult to control your expenditures?"*. Because the answer categories run from "Very easy" to "Very difficult" on a 7 point scale, we invert these answers so that higher values correspond with higher levels of control. The results from adding the variables described above are reported in columns (1) to (4) of Table 7. From the added set of controls only one variable turned out to be statistically significant. We find that people with better financial control are more likely to invest in the stock market.

[13] These placebo tests follow the placebo test undertaken by Georgarakos, Haliassos, and Pasini (2014).

[14] Region is provided by CentERdata and separates between "The three biggest cities in the west", "Rest west", "North", "East", and "South".

[15] Fuchs (1980) relates several types of health choices to patience.

Figure 5: Placebo tests



In these graphs we show the results of running the placebo test as described in the text. Each graph shows the results from a different crossing of cells. In each case we have re-estimated the baseline model a hundred times and saved the coefficient of interest as well as the associated standard error. In the different placebo tests the confidence intervals cover zero in almost all cases.

The financial literacy module sent out by CentERdata in 2011 provides us with a second set of additional control variables. The analyses are done separately since we lose some observations in matching our dataset to this extra module. From this model we use two questions where respondents have to indicate to what extent they agree with a statement (on 1 to 5 scales). The first statement captures a more focused trust variable: *"I trust financial advisors and agree with their advice."*. The second statement captures (self-assessed) mathematic skill: *"I am good at math."*. In columns (5) through (8) of Table 7 we show the results from our baseline specification when we add this second set of control variables in addition to all original controls and the first set of extra controls. We find that in addition to financial control, having high self-assessed math skills is statistically significantly related to stock market participation. More important is that across the board our original results remain consistent. If anything, our results become even stronger when we control for additional explanations of stock market participation. When adding the first (and second) set of controls individuals in the third proximity tercile are 14% (20%) more likely to hold stocks compared to individuals from the first tercile keeping all else equal compared to an 11% effect in column (5) of Table 3.

3.3 Different outcome variable

The focus of this paper is understanding the decision to invest in the stock market, a longstanding puzzle in the financial literature. However, it is likely that the importance of one's social network affects more aspects of financial decision making. In the literature on household finance, the stock market participation has traditionally received the most attention but recent work also looks at aspects such as the diversification of investment portfolios, see Von Gaudecker (2015) or the debt taken by households, see Georgarakos, Haliassos, and Pasini (2014). Guiso, Sapienza, and Zingales (2008) mainly focus on stock market participation but also present results on investments in all risky assets. Van Rooij, Lusardi, and Alessie (2012) argue that more financially literate individuals are more likely to plan for their retirement.

We focus on financial decisions that involve a high degree of complexity because this is arguably where one's financial network could have an impact. Because the exact data on investment portfolios is mostly missing in our sample and taking out debt might be a financially optimal choice, we analyze an indicator for the presence of any risky financial investment as in Von Gaudecker (2015) and an indicator for having made retirement savings calculations taken from the customized literacy survey of 2011.[16] The first four columns in Table 8 show that the impact generalizes to all risky assets as well.[17] The next four columns present results when the dependent variable captures whether respondents have already made calculations for the amount of money needed for retirement. In this case the effect is significant for the normalized proximity variable (column 5) but and for the indicator capturing the upper tercile (column 7). Taken together these results suggest that exposure in ones network to financial savvy people may influence a wide range of financial decisions.

[16] The exact wording of this question is *Did you ever calculate how much money you need for your old age?*

[17] We only present OLS results. The results of logit models are very similar to the ones reported here.

Table 7: Results with additional control variables

VARIABLES	(1) Extra 1	(2) se	(3) Extra 1	(4) se	(5) Extra 2	(6) se	(7) Extra 2	(8) se
Proximity	0.06***	(0.02)			0.09***	(0.02)		
Proximity Mid 1/3rd			0.05	(0.03)			0.09**	(0.04)
Proximity High 1/3rd			0.14***	(0.04)			0.20***	(0.05)
Male	0.09**	(0.04)	0.09**	(0.04)	0.05	(0.05)	0.05	(0.05)
Age 35 - 44	0.03	(0.07)	0.03	(0.07)	-1.01***	(0.11)	-1.07***	(0.11)
Age 45 - 54	0.14*	(0.08)	0.14*	(0.08)	-0.93***	(0.11)	-0.97***	(0.11)
Age 55 - 64	0.15**	(0.07)	0.14*	(0.07)	-0.95***	(0.11)	-1.00***	(0.11)
Age $>$ 65	0.21**	(0.09)	0.19**	(0.09)	-0.92***	(0.13)	-0.97***	(0.13)
Assets 158001 - 285000	0.06	(0.04)	0.05	(0.04)	0.07	(0.05)	0.07	(0.05)
Assets $>$ 285000	0.17***	(0.04)	0.17***	(0.04)	0.16***	(0.05)	0.16***	(0.05)
Lower secondary	0.02	(0.07)	0.04	(0.07)	-0.09	(0.08)	-0.08	(0.08)
Higher secondary	0.14*	(0.08)	0.15*	(0.08)	0.02	(0.10)	0.04	(0.09)
Lower vocational	0.11	(0.07)	0.12	(0.07)	-0.04	(0.09)	-0.03	(0.09)
Higher vocational	0.14*	(0.07)	0.15**	(0.07)	-0.03	(0.09)	-0.01	(0.09)
University	0.06	(0.08)	0.08	(0.08)	-0.13	(0.10)	-0.10	(0.10)
Optimism	0.02*	(0.01)	0.02*	(0.01)	0.01	(0.02)	0.02	(0.02)
Household size	-0.03	(0.02)	-0.03	(0.02)	-0.05*	(0.03)	-0.05*	(0.03)
Partner	0.01	(0.05)	0.01	(0.05)	0.04	(0.06)	0.03	(0.06)
Retired	-0.07	(0.05)	-0.07	(0.05)	-0.08	(0.06)	-0.09	(0.07)
Self employed	-0.05	(0.08)	-0.06	(0.08)	-0.05	(0.10)	-0.06	(0.10)
Generalized trust	0.01	(0.01)	0.01	(0.01)	0.00	(0.01)	0.00	(0.01)
Risk tolerance self	0.01***	(0.00)	0.02***	(0.00)	0.02***	(0.00)	0.02***	(0.00)
Fin. literacy self	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.03)	-0.01	(0.03)
Future orientation	0.02	(0.02)	0.03	(0.02)	0.01	(0.02)	0.02	(0.02)
Bequest intention	-0.00	(0.04)	-0.01	(0.04)	-0.01	(0.05)	-0.01	(0.05)
Financial control	0.02***	(0.01)	0.02***	(0.01)	0.03*	(0.01)	0.03*	(0.01)
Smoking: sometimes	0.02	(0.05)	0.03	(0.06)	-0.04	(0.08)	-0.03	(0.08)
Smoking: daily $<$ 20	-0.01	(0.05)	-0.01	(0.05)	-0.03	(0.06)	-0.02	(0.06)
Smoking: daily $\geq$ 20	-0.01	(0.08)	-0.01	(0.08)	0.04	(0.10)	0.04	(0.10)
Drinking: daily $>$4 drinks	0.06	(0.06)	0.06	(0.06)	0.08	(0.07)	0.07	(0.07)
Fin. advice trust					0.02	(0.02)	0.03	(0.02)
Good at math					0.04*	(0.02)	0.04**	(0.02)
Constant	-0.58***	(0.18)	-0.66***	(0.18)	0.50**	(0.25)	0.42	(0.26)
Observations	709		709		450		450	
Income dummies	Yes		Yes		Yes		Yes	
Urbanization dummies	Yes		Yes		Yes		Yes	
Religion dummies	Yes		Yes		Yes		Yes	

This table shows the OLS results as in Table 3 with additional controls. The first set of extra controls is derived from the DHS 2014 wave while the second set of extra controls comes from a customized survey sent out in 2011. Heteroskedasticity robust standard errors are reported in parentheses. We indicate significance at 1%, 5%, and 10% by ***, **, and * respectively.

Table 8: Results for different outcome variables

VARIABLES	(1) Risky assets	(2) se	(3) Risky assets	(4) se	(5) Retirement plan	(6) se	(7) Retirement plan	(8) se
Proximity	0.05***	(0.01)			0.06***	(0.02)		
Proximity Mid 1/3rd			0.06**	(0.03)			0.04	(0.04)
Proximity High 1/3rd			0.13***	(0.03)			0.09*	(0.05)
Male	0.08***	(0.03)	0.09***	(0.03)	0.02	(0.05)	0.02	(0.05)
Age 35 - 44	0.08*	(0.04)	0.07	(0.04)	-0.14	(0.20)	-0.16	(0.20)
Age 45 - 54	0.12**	(0.05)	0.12**	(0.05)	-0.11	(0.20)	-0.12	(0.20)
Age 55 - 64	0.18***	(0.04)	0.17***	(0.04)	0.00	(0.20)	-0.01	(0.20)
Age > 65	0.19***	(0.06)	0.18***	(0.06)	-0.17	(0.20)	-0.18	(0.21)
Assets 158001 - 285000	0.08***	(0.03)	0.08***	(0.03)	0.07	(0.04)	0.07	(0.04)
Assets > 285000	0.18***	(0.03)	0.19***	(0.03)	0.08	(0.05)	0.09*	(0.05)
Lower secondary	-0.05	(0.06)	-0.04	(0.06)	0.06	(0.08)	0.07	(0.08)
Higher secondary	0.06	(0.07)	0.07	(0.07)	0.19**	(0.09)	0.21**	(0.09)
Lower vocational	0.03	(0.06)	0.04	(0.07)	0.07	(0.08)	0.08	(0.09)
Higher vocational	0.06	(0.06)	0.06	(0.06)	0.07	(0.08)	0.09	(0.08)
University	0.05	(0.07)	0.06	(0.07)	0.21**	(0.09)	0.24**	(0.09)
Optimism	0.01	(0.01)	0.01	(0.01)	0.00	(0.02)	0.00	(0.02)
Household size	-0.03**	(0.01)	-0.03**	(0.01)	-0.01	(0.02)	-0.01	(0.02)
Partner	-0.04	(0.04)	-0.04	(0.04)	-0.01	(0.06)	-0.02	(0.06)
Retired	-0.03	(0.05)	-0.03	(0.05)	0.08	(0.07)	0.08	(0.07)
Self employed	-0.01	(0.07)	-0.02	(0.07)	0.15	(0.09)	0.15	(0.10)
Generalized trust	0.01*	(0.01)	0.01	(0.01)	0.02*	(0.01)	0.02*	(0.01)
Risk tolerance self	0.01***	(0.00)	0.01***	(0.00)	-0.01***	(0.00)	-0.01***	(0.00)
Constant	-0.25**	(0.10)	-0.32***	(0.10)	0.16	(0.23)	0.09	(0.23)
Observations	1,043		1,043		646		646	
Income dummies	Yes		Yes		Yes		Yes	
Urbanization dummies	Yes		Yes		Yes		Yes	
Religion dummies	Yes		Yes		Yes		Yes	

This table shows results from regressions of alternative outcome variables on proximity measures and control variables. We consider an indicator for retirement savings calculations and an indicator for risky financial assets. Controls are similar to those used in Table 3. Heteroskedasticity robust standard errors are reported in parentheses. We indicate significance at 1%, 5%, and 10% by ***, **, and * respectively.

4 Conclusion

In this paper we have demonstrated how one can use questions of the type *How many X's do you know?* to measure social exposure. We find this measure appealing because it is tied to a general model of social tie formation. We see this proximity measure as a useful complement to existing survey variables used to capture various aspects of social networks. We have demonstrated that exposure to financially savvy people matters for stock market participation. People with a substantial exposure are more likely to own stocks after controlling for a wide range of control variables. This remains true when restricting the sample to groups of people we deem to be more likely to own stocks in the first place, highly literate and highly trusting people. To confirm the robustness of these results we have applied an instrumental variables approach as well as placebo tests. In both cases our results remained. Finally we have explored different outcome variables such as *owning risky assets* and whether *respondents had already made calculations of their retirement needs*. In both cases we found that our proximity measure had a positive effect although the results were less strong for the latter variable.

An interesting feature of the approach we used in this paper is that it lends itself to various other applications easily.

A Appendix

A.1 Data Collection CentERdata

In this work we use household panel data collected by CentERdata which is administered and run by CentER at Tilburg University. The CentERdata household panel is representative of the Dutch society and collects data on individuals aged 16 and above. Administrators update the panel every year to maintain this representativeness. Generally interviews are done via the internet however when a household does not have a device that can connect to the internet a television set-top box is provided (with a television if necessary) that can be used to fill out surveys. Once a year that collect background data and during the year customized surveys are sent out amongst the panel members. Generally, they send a reminder in case a household member does not complete a survey. Nevertheless, households cannot be forced since participation is voluntary. The background information is extensive and covers seven topics: 1. General Information on the Household; 2. Household and Work; 3. Accommodation and Mortgages; 4. Health and Income; 5. Assets and Liabilities; 6. Economic and Psychological Concepts.
In our baseline analyses we use the 2014 wave of their main panel data that is freely available online, this data is also referred to as the DNB Household Survey data (or DHS data). CentERdata offers the opportunity to field customized surveys in their panel. We add customized networks survey data as well as data on trust, both collected in 2014, and customized financial literacy data collected in 2011.

A.2 Overview of financial literacy questions

One of our instruments is financial literacy inflation. This is an indicator for answering a question on inflation correctly. For completeness, we also include the other available literacy questions in the analyses where we use bond literacy. The first two questions are from the basic literacy index of Lusardi and Mitchell (2008) while questions 3 and 4 are taken from the advanced literacy survey of Van Rooij, Lusardi, and Alessie (2011). Note that one important difference is that in our case individuals do not have the option to answer "Do not know". The wording of the questions is listed below:

- Interest: Suppose you had €100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow? (i) More than 102; (ii) Exactly 102; (iii) Less than 102.
- Inflation: Imagine that the interest rate on your savings account was 1 percent per year and inflation was 2 percent per year. After 1 year, how much would you be able to buy with the money in this account? (i) More than today; (ii) Exactly the same; (iii) Less than today.
- Bonds: If the interest rate falls, what should happen to bond prices? (i) Rise; (ii) Fall; (iii) Stay the same; (iv) None of the above.
- Risk stocks: Buying a company stock usually provides a safer return than a stock mutual fund. True or false? (i) True; (ii) False.

A.3 Overview of population proportions and exact priors

To estimate the overdispersed model in the paper we specified the following priors:
$\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$, $\beta_k \sim N(\mu_\beta, \sigma_\beta^2)$, $\mu_\alpha \sim N(0, 25)$, $\sigma_\alpha \sim N(0, 5)$, $\frac{1}{\omega_k} \sim \text{Uniform}(0, 1)$. For μ_β we choose hyperpriors $\mu_\beta \sim N(\text{popval}_k, .2)$ for the name questions. Here popval_k refers to the population values of the names (see below) exponentiated (as $\alpha_i = \exp(a_i)$) -to identify the parameters (see paper). To the four other questions we assign the average of popval_k to the mean of the Normal hyperprior and a variance of 25. So we assign very diffuse priors to the finance related questions and very tight priors to the names questions as to normalize the model. A further discussion on providing priors is given in Zheng, Salganik, and Gelman (2006). Note that there are additional complexities in their setup. We were able to avoid these by choosing names in line with the recommendations of McCormick, Salganik, and Zheng (2010).

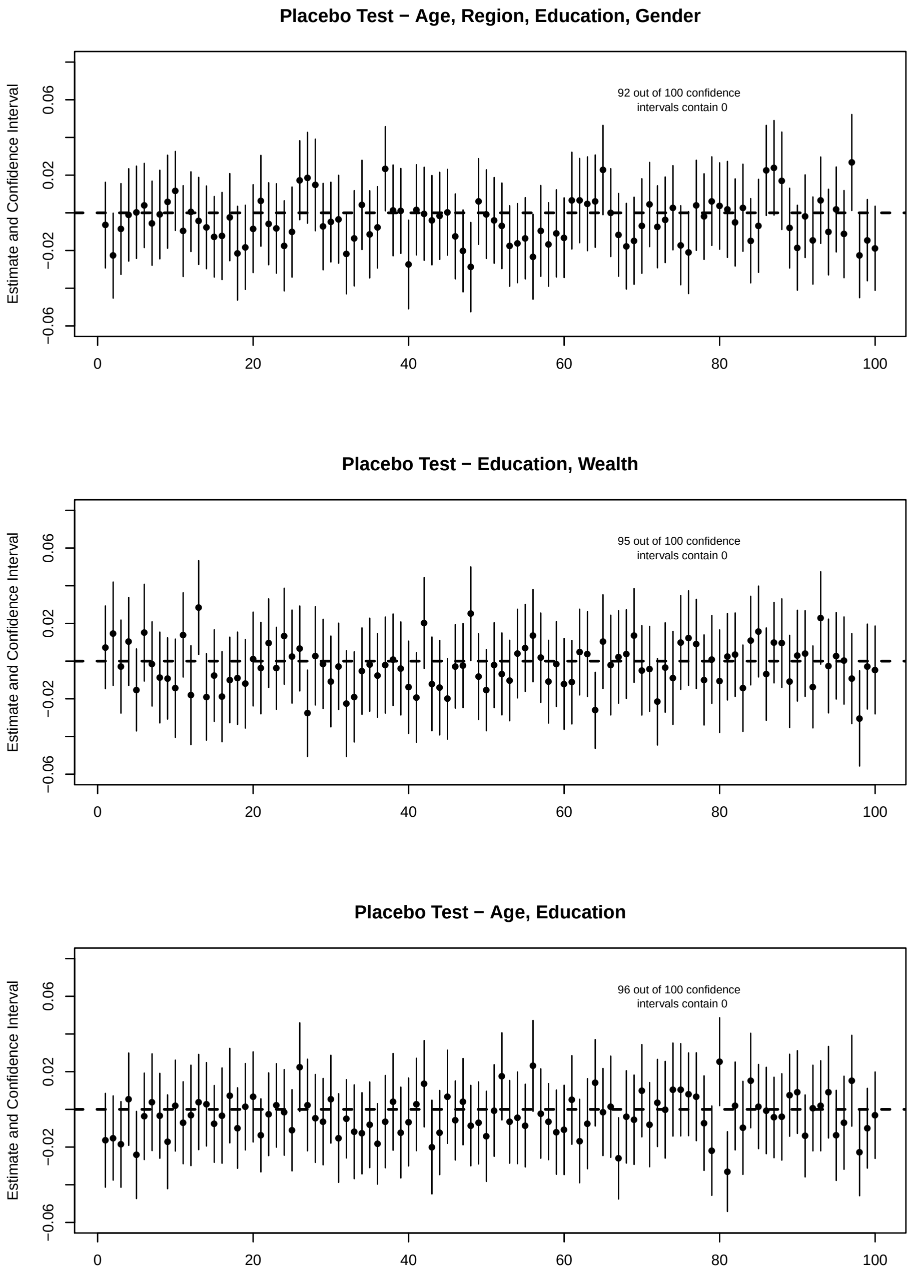
Table 9: Names and population values

Name	Population	Population Fraction
Ingrid	33882	0.25%
Yvonne	44955	0.33%
Remco	15109	0.11%
Kim	21210	0.16%
Mohammed	21956	0.16%
Harm	18378	0.14%
Andre	13201	0.1%
Raymond	13397	0.1%
Wendy	19231	0.14%
Kevin	17977	0.13%
Eline	7541	0.06%
Karin	23942	0.18%

A.4 Additional Placebo tests

Here we perform three additional placebo tests. As mentioned in the text we can construct more placebo tests. We feel that the seven we have presented (4 in the body of the paper and 3 here) provide sufficient evidence. One constraint on the construction of the cells is that cells need to contain sufficient observations. We randomly assign within each cell the proximity value to *another* household, hence the cells cannot be too small. For this reason we require minimally 10 observation within each cell.

Figure 6: Placebo tests



In these graphs we show the results of running the placebo test as described in the text. Each graph shows the results from a different crossing of cells. In each case we have re-estimated the baseline model a hundred times and saved the coefficient of interest as well as the associated standard error. We see in the different placebo tests that the confidence intervals in most cases cover zero.

References

ABEL, A. B. (2001a): "The Effects of Investing Social Security Funds in the Stock Market When Fixed Costs Prevent Some Households from Holding Stocks," *American Economic Review*, pp. 128–148.

——— (2001b): "Will bequests attenuate the predicted meltdown in stock prices when baby boomers retire?," *Review of Economics and Statistics*, 83(4), 589–595.

ALATAS, V., A. BANERJEE, A. G. CHANDRASEKHAR, R. HANNA, AND B. A. OLKEN (2012): "Network structure and the aggregation of information: Theory and evidence from Indonesia," Discussion paper, National Bureau of Economic Research.

ANDERSEN, S., T. HANSPAL, AND K. M. NIELSEN (2014): "Once Bitten, Twice Shy: Do Personal Experiences or Wealth Changes Affect Risk Taking?," *working paper*.

BARNEA, A., H. CRONQVIST, AND S. SIEGEL (2010): "Nature or nurture: What determines investor behavior?," *Journal of Financial Economics*, 98(3), 583–604.

BENJAMIN, D. J., S. A. BROWN, AND J. M. SHAPIRO (2013): "Who is 'behavioral'? Cognitive ability and anomalous preferences," *Journal of the European Economic Association*, 11(6), 1231–1255.

BERTAUT, C. C., AND M. STARR (2000): "Household portfolios in the United States," .

BRAV, A., G. M. CONSTANTINIDES, AND C. C. GECZY (2002): "Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence," *Journal of Political Economy*, 110(4), 793–824.

BROWN, J. R., Z. IVKOVIĆ, P. A. SMITH, AND S. WEISBENNER (2008): "Neighbors matter: Causal community effects and stock market participation," *The Journal of Finance*, 63(3), 1509–1531.

CALVÓ-ARMENGOL, A., E. PATACCHINI, AND Y. ZENOU (2009): "Peer effects and social networks in education," *The Review of Economic Studies*, 76(4), 1239–1267.

CAMERON, A. C., AND P. K. TRIVEDI (2005): *Microeconometrics: methods and applications*. Cambridge university press.

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

CHRISTELIS, D., T. JAPPELLI, AND M. PADULA (2010): "Cognitive abilities and portfolio choice," *European Economic Review*, 54(1), 18–38.

DIPRETE, T. A., A. GELMAN, T. MCCORMICK, J. TEITLER, AND T. ZHENG (2011): "Segregation in Social Networks Based on Acquaintanceship and Trust1," *American Journal of Sociology*, 116(4), 1234–83.

DUFLO, E., AND E. SAEZ (2002): "Participation and investment decisions in a retirement plan: The influence of colleagues' choices," *Journal of public Economics*, 85(1), 121–148.

ERDŐS, P., AND A. RÉNYI (1959): "On random graphs," *Publicationes Mathematicae Debrecen*, 6, 290–297.

FUCHS, V. R. (1980): "Time preference and health: an exploratory study," *NBER working paper*.

GEORGARAKOS, D., M. HALIASSOS, AND G. PASINI (2014): "Household debt and social interactions," *Review of Financial Studies*, 27(5), 1404–1433.

GEORGARAKOS, D., AND G. PASINI (2011): "Trust, sociability, and stock market participation," *Review of Finance*, 15(4), 693–725.

GIANNETTI, M., AND Y. KOSKINEN (2010): "Investor protection, equity returns, and financial globalization," *Journal of Financial and Quantitative Analysis*, 45(1), 135–168.

GRANOVETTER, M. S. (1973): "The strength of weak ties," *American journal of sociology*, pp. 1360–1380.

GRINBLATT, M., M. KELOHARJU, AND J. LINNAINMAA (2011): "IQ and stock market participation," *The Journal of Finance*, 66(6), 2121–2164.

GUISO, L., P. SAPIENZA, AND L. ZINGALES (2008): "Trusting the stock market," *the Journal of Finance*, 63(6), 2557–2600.

HALIASSOS, M., AND C. C. BERTAUT (1995): "Why do so few hold stocks?," *the economic Journal*, pp. 1110–1129.

HOFFMAN, M. D., AND A. GELMAN (2014): "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo," *Journal of Machine Learning Research*, 15(Apr), 1593–1623.

HONG, H., J. D. KUBIK, AND J. C. STEIN (2004): "Social interaction and stock-market participation," *The journal of finance*, 59(1), 137–163.

HONG, H. G., AND J. XU (2015): "Inferring latent social networks from stock holdings," *Available at SSRN 2375461*.

IVKOVIĆ, Z., AND S. WEISBENNER (2007): "Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices," *Review of Financial Studies*, 20(4), 1327–1357.

JACKSON, M. O. (2010): "An overview of social networks and economic applications," *The handbook of social economics*, 1, 511–85.

——— (2014): "Networks in the understanding of economic behaviors," *The Journal of Economic Perspectives*, pp. 3–22.

KAUSTIA, M., AND S. TORSTILA (2011): "Stock market aversion? Political preferences and stock market participation," *Journal of Financial Economics*, 100(1), 98–112.

KILLWORTH, P. D., C. MCCARTY, H. R. BERNARD, E. C. JOHNSEN, J. DOMINI, AND G. A. SHELLEY (2003): "Two interpretations of reports of knowledge of subpopulation sizes," *Social networks*, 25(2), 141–160.

LUSARDI, A., AND O. S. MITCHELL (2008): "Planning and Financial Literacy: How Do Women Fare?," *The American Economic Review*, pp. 413–417.

——— (2014): "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, 52(1), 5–44.

MALMENDIER, U., AND S. NAGEL (2011): "Depression babies: Do macroeconomic experiences affect risk taking?," *The Quarterly Journal of Economics*, 126, 373–416.

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

MCCORMICK, T. H., A. MOUSSA, J. RUF, T. A. DIPRETE, A. GELMAN, J. TEITLER, AND T. ZHENG (2013): "A practical guide to measuring social structure using indirectly observed network data," *Journal of Statistical Theory and Practice*, 7(1), 120–132.

MCCORMICK, T. H., M. J. SALGANIK, AND T. ZHENG (2010): "How many people do you know?: Efficiently estimating personal network size," *Journal of the American Statistical Association*, 105(489), 59–70.

PURI, M., AND D. T. ROBINSON (2007): "Optimism and economic choice," *Journal of Financial Economics*, 86(1), 71–99.

SMITH, J. P., J. J. MCARDLE, AND R. WILLIS (2010): "Financial decision making and cognition in a family Context*," *The Economic Journal*, 120(548), F363–F380.

STAN DEVELOPMENT TEAM (2014): "Stan: A C++ Library for Probability and Sampling, Version 2.5.0," .

STRATHMAN, A., F. GLEICHER, D. S. BONINGER, AND C. S. EDWARDS (1994): "The consideration of future consequences: Weighing immediate and distant outcomes of behavior.," *Journal of personality and social psychology*, 66(4), 742.

VAN ROOIJ, M., A. LUSARDI, AND R. ALESSIE (2011): "Financial literacy and stock market participation," *Journal of Financial Economics*, 101(2), 449–472.

——— (2012): "Financial Literacy, Retirement Planning and Household Wealth," *The Economic Journal*, 122(560), 449–478.

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.

VON GAUDECKER, H.-M. (2015): "How does household portfolio diversification vary with financial literacy and financial advice?," *The Journal of Finance*, 70(2), 489–507.

YOONG, J. (2011): "Financial Illiteracy and Stock Market Participation: Evidence from the RAND American Life Panel," *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*, p. 76.

ZHENG, T., M. J. SALGANIK, AND A. GELMAN (2006): "How many people do you know in prison? Using overdispersion in count data to estimate social structure in networks," *Journal of the American Statistical Association*, 101(474), 409–423.