The Value of Connections: Network Effects and Stock Market Participation*

Olga Balakina Claes Bäckman Anastasiia Parakhoniak

April, 2023

Abstract

The past twenty years have seen an explosion in our ability to share financial information on social networks, yet stock market participation has barely changed. We introduce an equilibrium model of stock market participation with a social network and study how the impact of connectivity on stock market participation differs in the presence of other factors. The effect of connectivity depends on how efficient information spreads, which is linked to how agents are connected, inequality, and the share of informed agents. High-income agents benefit more from connectivity, leading to increased inequality. Increased connectivity is not generally associated with higher participation rates.

Keywords: Social networks; Peer effects; Stock Market Participation; Connectivity

JEL classification: D14, D85, G11

*Balakina: Department of Economics and Business Economics, Aarhus University, and Danish Finance Institute. Email: olga.balakina@econ.au.dk. Bäckman: Department of Economics and Business Economics, Aarhus University. Email: claes.backman@econ.au.dk. Parakhonyak: Durham University Business School, Department of Economics and Finance. Email: anastasiia.parakhoniak@durham.ac.uk. We thank Tobin Hanspal, Petra Thiemann, Peter Norman Sørensen, Steffen Andersen, and Lena Jaroszek as well as seminar participants at the Copenhagen University, Bocconi University, Lund University, University of Trento, the 2nd SAFE Household Finance Workshop, 6th Annual CIRANO-Sam M. Walton College of Business Workshop on Networks in Trade and Finance, KWC-CFF Workshop, Belgian Network Research Meeting 2016, EDGE Jamboree and DGPE Workshop. Support from the Danish Finance Institute is gratefully acknowledged.

1. INTRODUCTION

A large literature in economics and finance links social networks to individual stock market participation. However, a general connectivity measure, the Social Connectivity Index (SCI) from Facebook, does not predict cross-sectional variation in stock market participation across US counties. Similarly, the past twenty years of the explosive growth of social media had seemingly no effect on stock market participation. While the share of American households using social media went from 5 percent in 2005 to 89 percent in 2019 (Ortiz-Ospina, 2019), data from the Survey of Consumer Finances (SCF) shows a flat or even declining average share of households who directly invest in the stock market (Figure 1). However, stock market participation is strongly correlated with *economic connectedness*, a measure that conveys information about the connectivity between individuals with a high- and low socioeconomic status (Chetty et al., 2022a). It seems like it is not general sociability that determines stock market participation, but instead connections to informed peers. To the extent that this mechanism is at work, better understanding connectivity and information sharing in social networks can help shed light on the surprising lack of movements in stock market participation.

This paper argues that network structure is important for understanding stock market participation. The mechanism that we posit is that access to informed peers depends on inequality and homophily (the tendency of individuals to associate with others of the same group), and that these two parameters are key to understanding how connectivity affects participation decisions. Our approach uses a theoretical model with a social network to explore three important questions: (1) how does connectivity affect stock market participation? (2) Which group benefits from increased connectivity? (3) How does inequality and homophily work to mediate the effect of connectivity? These issues have not been previously explored in a theoretical setting, and our answers thus shed light on the important issue of stock market participation and social networks.

We build and calibrate a model of stock market participation where all agents can share information about the stock market in a network. Agents in the model have to pay an agent-specific fixed cost to participate in the stock market. The fixed cost captures both the monetary cost and behavioral and psychological factors that make stock ownership uncomfortable for some households (Campbell, 2006). We let the fix cost depend on the number of *informed* agents that each agent is connected to. An informed agent is any agent that participates in the stock market. Moreover, we introduce two types of agents into the model who differ in their participation costs. *Financially educated agents* face low fixed costs and thus participate at a high rate. *Non-Financially educated agents* face ex-ante high stock market participation costs, which can be lowered through learning from peers who invest in stocks. Agents in the model are connected with an ex-ante connectivity probability that determines the expected number of links for each agent. We also introduce a homophily parameter, defined as the difference between the probability of connecting with an individual with a similar income.When the homophily parameter is equal to zero, each individual is equally likely to form a link with any other agent independently of their income. When the homophily parameter equals one, each individual forms a connection only inside her income group.

The model endogenously generates an S-shape relationship between connectivity and stock market participation. At low levels of connectivity information sharing is limited, meaning that the only participants are agents rich enough to afford the fixed participation cost sand Financially educated agents. As connectivity increases, more agents participate, become informed and spread information. Participation rapidly increases. At some point, however, the information diffusion process slows down, and increased connectivity has a smaller effect on overall participation. Increased connectivity mainly benefits richer agents, who are closer to the participation threshold and thus need fewer informed connections to participate. Increased connectivity therefore generates more ex-post inequality within our model.

Higher homophily has a positive impact on stock market participation in the model because of more efficient information transmission, but the effect depends on the level of connectivity. As agents become more likely to be connected to agents with similar incomes, the model generates clusters of high-income agents who are closer to the participation threshold. This leads to more efficient information sharing and higher participation rates among richer agents, with little impact on the participation by middle and low income agents. However, the effect of homophily also depends on the level of connectivity. In simulations with very low or high number of average connections, stock market participation is flat across homophily. This nicely illustrates the idea that homophily has an effect through connectivity: if all agents are connected, homophily does not have a big impact on participation.

Higher ex-ante inequality also affects stock market participation in the model through two channels. First, since we keep average income the same in simulations, inequality affects the share of agents who have the ability to pay the fixed participation costs. Second, inequality also affects the probability that agents with information are connected to other agents because of homophily. Similar to the results for homophily, the impact of inequality on stock market participation depends on the average level of connectivity. In worlds with low connectivity, information sharing is limited and participation is low. Higher inequality can generate higher participation rates, since income is concentrated among agents who are close to the participation threshold. In worlds with higher connectivity, however, increased inequality instead generates lower participation, since inequality affects the likelihood that agents are connected.

We make several contributions to the literature. First, a large literature investigates the limited stock market participation puzzle, dating back to Arrow (1965).[1] Standard models of stock market participation show that moderate participation costs can explain the non-participation of many US households but not the richest ones (Haliassos and Bertaut,

[1] See e.g. Mehra and Prescott (1985), Fama and French (2002), Mankiw and Zeldes (1991), Haliassos and Bertaut (1995), Heaton and Lucas (2000), Brav et al. (2002) and Vissing-Jørgensen (2002).

1995; Vissing-Jørgensen, 2002). Conversely, participation of the low-income individuals is inconsistent with the high participation costs estimated in Andersen and Nielsen (2010) and Briggs et al. (2015). By allowing the costs to vary with social connectivity, we generate heterogeneity in the level of stock market participation unrelated to income or financial education. In essence, social connectivity acts as an omitted variable. Second, our paper provides a theoretical framework for the empirical literature that studies how social interactions between agents affect financial decisions (Kaustia and Knüpfer, 2012; Bursztyn et al., 2014; Changwony et al., 2014; Patacchini and Rainone, 2017). Finally, participation in risky asset markets is one of the main drivers of difference between wealthy and less wealthy households (Bach et al., 2020; Fagereng et al., 2020), and access to informed peers varies across the wealth distribution is a natural empirical extension of our model. Moreover, our model suggests that rising income inequality directly impacts wealth inequality if increased inequality is associated with less socializing between groups.

2. MOTIVATING EVIDENCE

The key idea in our model is that if a particular group has no stock market participant who can share information, then the degree to which they socialize will not matter. We begin by showing that a general measure of connectivity does not predict cross-sectional variation in stock market participation in US counties, but that stock market participation is strongly correlated with *economic connectedness*, a measure that conveys more information about connectivity between individuals with a high- and low socioeconomic status (Chetty et al., 2022a).

We combine several county-level dataset to examine the correlation between connectivity and stock market participation. Below we describe the main sources and variables of interest. We first collect county-level data on connectivity the Social Connectivity Index (SCI) from Facebook Bailey et al. (2018). The SCI measures the social connectedness between and within US county-pairs. This measure is constructed as an index based on the number of friendship links on Facebook, where the average number of links is normalized to the county-pair with largest number of connections – the Los Angeles County - Los Angeles County pair. We augment this connectivity data with data on economic connectedness from Chetty et al. (2022a,b). Economic connectedness is defined as two times the share of high socio-economic status (SES) friends among low-SES individuals, averaged over all low-SES individuals in the county. Since high income agents are more likely to have information to share about the stock market simply because their participation is higher, economic connectedness captures the idea that low information agents need access to high information agents in order to benefit. Finally, we calculate the county-level participation share as the fraction of tax returns claiming ordinary dividends. Hung (2021) provides a detailed validation of the measure. Details on other data sources and definitions are available in Appendix A and descriptive statistics are available in Table 1.

We examine the relationship between stock market participation and social connections in Figure 3. The left side of the figure reports scatter-plots between connectivity variables of interest and stock market participation. The right side of the figure plots binned scatterplots that control for county-level age, age squared, median household income, the share of financially educated households, and the share of the county has a bachelor-level education or above, as well as fixed effects for state and the main industry of employment in the county. More detailed results are presented in Table 2.

Panel a) and b) of Figure 3 provide results for the SCI and stock market participation. The relationship between log inside connectivity from the SCI and stock market participation is positive and significant in both figures, but SCI only explains 2 percent of the variation in stock market participation and the statistical significant is low. The relationship is also not generally robust to changing control variables or to different transformations. For example, column 5 in Table 2 is equivalent to the results in the binned scatterplot in panel b) of Figure 3. Adding a control for the major employment industry in the county reduces the magnitude of the coefficient and the variable is no longer statistically significant.

In panel c) and d) of Figure 3, we see that economic connectedness is more strongly correlated with stock market participation. Moreover, economic connectedness explains 42 percent of the variation in stock market participation. This result is robust to controls and to variable transformations such as taking logs. We interpret these results as evidence that while connectivity in general seem to matter, it is more important to be connected to individuals who have information to share.

Finally, Column 7-9 of Table 2 provides results where we include both economic connectedness and connectivity. Both variables are now statistically significant and positive across specifications.

We interpret these results in the following way. Holding economic connectedness fixed, a higher connectivity has a positive impact on stock market participation. If we fix the information content in the county by holding the economic connectedness constant, having more connections will help spread information.

3. THE MODEL

In this section, we propose a stylized model to study how the interaction between connectivity and other economic factors affects the average SMP in the population and across different income groups. The model set-up has two main components. First, we formulate the utility maximization problem of an individual investor who has some stock market participation cost. This part closely follows the previous literature and is mostly inspired by the framework proposed in Vissing-Jørgensen (2002). However, unlike in the previous literature, we endogenize individual investors' stock participation costs assuming that they depend on the number of participating peers who share financial information. Therefore, the second component of the model set-up describes how individuals embedded in a social network share information about the financial market and how the information diffusion process works.

3.1 GENERAL SETTING

We introduce a one-period, closed-economy model that describes the financial behavior of an agent within a social network. At the beginning of the period, agents allocate their endowment in the form of discretionary income between a risk-free and a risky asset, such as a stock, and at the end of the period, they consume the proceeds from the investment portfolio in the form of a non-durable consumption good. The economy is populated by risk-averse agents with identical CRRA preferences. The utility function of agent i is:

$$U_i\left(W_{i,1}\right) = \frac{{W_{i,1}}^{1-\gamma}}{1-\gamma},\ \gamma > 0,$$

where $W_{i,1}$ defines the level of wealth of agent i at the end of the period, and γ is the level of relative risk aversion of the agent. Agents have initial endowment $W_0 = \{W_{1,0}, ..., W_{j,0}, ...W_{n,0}\}$ distributed as $\mathscr{F}_w(\cdot)$, $W_{j,0} \sim \mathscr{F}_w(\cdot)$. The economy offers two investment opportunities. An agent can choose between investing her initial endowment in a risk-free asset with the net return equal to zero, $r^f = 0$, or investing in a risky asset with a higher return. If agent i decides to invest in the risky asset, she faces the costs of participation, F_i, at the beginning of the period.[2] The net return on the risky asset r is a random variable with the binomial distribution, such that

$$r = \begin{cases} r_u \text{ , with probability } \pi \\ r_d \text{ , with probability } (1-\pi), \end{cases}$$

where $r_d < 0 < r_u$. The expected net return on the risky asset is positive, that is:

$$\pi r_u + (1-\pi)\, r_d > 0$$

[2]The costs F_i in the case of stock market investments measure the cost of time or money spent understanding basic investment principles as well as acquiring enough information about risks and returns, the cost of time spent setting up an account, brokerage commission, and the time spent implementing the trade (Vissing-Jørgensen, 2002).

Because the terminal discretionary income $W_{i,1}$ is equal to proceeds from the investment portfolio, we can define $W_{i,1}$ as

$$W_{i,1} = (W_{i,0} - F_i)(1 + \lambda r_j), \text{ where } j = \{u, d\},$$

where r_j is a realization of the net risky-asset return at the end of the period, and λ is the share of income invested in the risky asset. If the agent decides not to invest in the risky asset her discretionary income does not change from period 1 to period 2 , $W_{i,1} = W_{i,0}$.

We assume that only agents whose initial discretionary income is larger than participation cost decide to participate and invest in the risky asset. Therefore, if $F > W_0$, the agent does not invest in the risky asset, and thus $W_{i,1} = W_{i,0}$.

3.2 THE AGENT'S OPTIMAL INVESTMENT DECISION

We first consider the problem of an individual agent i who decides how much to invest in the risky asset. Every agent in the economy solves the following optimization problem:

$$\max_{\lambda} E(U(W_{i,1})) = \max_{\lambda} E\left(\frac{W_{i,1}^{1-\gamma}}{1-\gamma}\right), \ \gamma > 0, \tag{1}$$

$$s.t \quad W_{i,1} = (W_{i,0} - F_i)(1 + \lambda r_j), \quad \text{for} \quad j = \{u, d\}, \quad 0 \leq \lambda \leq 1. \tag{2}$$

The assumption $\lambda \leq 1$ implies that an agent allocates at most all of her discretionary income to the risky asset and hence does not borrow to invest. Constraint (2) should be satisfied with equality, thus we can incorporate it in the equation for expected utility.

$$\max_{0 \leq \lambda \leq 1} \frac{\pi \left[(W_{i,0} - F_i)(1 + \lambda r_u)\right]^{1-\gamma} + (1-\pi)\left[(W_{i,0} - F_i)(1 + \lambda r_d)\right]^{1-\gamma}}{1-\gamma}.$$

The first-order condition for this problem is:

$$\pi r_u \left[(W_{i,0} - F_i)(1 + \lambda r_u)\right]^{-\gamma} + (1-\pi) r_d \left[(W_{i,0} - F_i)(1 + \lambda r_d)\right]^{-\gamma} = 0. \tag{3}$$

Solving (3) for λ we get the optimal fraction of the portfolio allocated to the risky asset, λ^*:

$$\begin{aligned}\lambda^* &= \min\left\{\frac{(1-m)}{(mr_u - r_d)}, 1\right\}, \text{ where} \\ m &= \left(\frac{\pi r_u}{(\pi - 1)\, r_d}\right)^{-\frac{1}{\gamma}}.\end{aligned}$$

We assume that $r_d < 0 < r_u$ and $\pi r_u + (1-\pi)\, r_d > 0$. As a result, we have that $0 < m < 1$ and $\lambda \geq 0$.

As the next step, we introduce two types of agents in the economy: Financially Educated and Non-Financially Educated. We define the type of agent i as t_i, where t_i equals 1 if the agent is Financially Educated, and 0 if an agent is Non-Financially Educated. The two types of agents differ in their participation cost functions, $F(t_i)$. Financially Educated agents have ex-ante knowledge about the investment in the risky asset, meaning their participation costs are equal to zero.[3] Non-Financially Educated agents do not have ex-ante knowledge about the stock market and face high participation costs. Non-Financially Educated agents can attain knowledge by learning from their peers who invest in the risky asset – Non-Financially Educated agents face costs of participating that are decreasing in the number of peers in their social network that invest in the risky asset. We assume the participation costs paid by a Non-Financially Educated agent i, $F(t_i = 0)$, are equal to some function $C(\theta, k_i)$ where k_i is the number of peers of agent i already investing in the risky asset, θ is an exogenous parameter that controls for the general level of participation cost in the population, $C'_\theta(\theta, k_i) > 0$. An agent who has more informed peers faces lower participation cost, $C'_{k_i}(\theta, k_i) < 0$.

[3]The necessary assumption is that Financially Educated agents have lower participation costs than Non-Financially Educated agents. Zero costs always satisfy this condition and guarantee maximum participation of Financially Educated agents. Any positive costs will generate a lower participation level among Financially Educated agents and a lower equilibrium participation level in the economy.

$$F(t_i) = \begin{cases} C(\theta, k_i), & \text{if } t_i = 0: \text{ agent } i \text{ is Non-Financially Educated} \\ 0, & \text{if } t_i = 1: \text{ agent } i \text{ is Financially Educated} \end{cases} \tag{4}$$

Note that all agents who invest in the risky asset can spread the information about it, not just the Financially Educated agents.

3.3 INFORMATION DIFFUSION IN THE SOCIAL NETWORK

All agents are a part of a network. The structure of the network is described by $\{\mathbf{N}, G, W, T\}$, where $\mathbf{N}$ is a set of agents-nodes of power n, the number of agents in the economy, G is a $n \times n$ adjacency matrix describing connections between agents in the network, $W = \{W_{1,0}, ..., W_{n,0}\}$ is a vector of length n describing the level of the initial discretionary income allocated to each agent in the network, and $T = \{t_1, ..., t_n\}$ is a binary vector that identifies types of agents.

Before moving to the discussion of potential equilibria and how they can be reached through the information diffusion process, we find it technically convenient to construct a new variable which we will call a Participation Threshold. For any agent i, we can determine the minimum number of peers already investing in the risky asset, $\hat{k}_i$, such that if the agent i has a number of peers-investors larger or equal to $\hat{k}_i$, she will decide to invest in the risky asset herself. In other words, a Participation Threshold $\hat{k}_i$ of agent i shows how many participating peers should share information about the stock market with agent i, such that her stock market participation cost becomes sufficiently low to enter the stock market. The variable $\hat{k}_i$ depends on agent's individual characteristics such as her discretionary income, W_i, and type, t_i. Intuitively, non-participating agents with high income need to collect less information from their peers before they start to invest in the risky asset than agents with low income. All financially-educated agents have a zero participation threshold because they possess all the necessary information.

We can now reformulate our problem and consider a network structure where each agent-node i has a randomly assigned number $\hat{k}_i$ with some discrete probability distribution function $\mathscr{F}(\hat{k}_i)$ instead of W_i and t_i. In equilibrium, each agent is a Participant if and only if the number of her first-degree peers (agents in the network that an agent is directly linked to) who are *Participants* is higher than or equal to $\hat{k}_i$. Note that for all Financially Educated agents $\hat{k}_i = 0$ if $W_{i,0} > 0$.

We illustrate the idea with an example presented in Figure 4. In the figure, we assign a number $\hat{k}_i$ to each agent. Coloured circles correspond to agents who invest in the risky asset. In this set-up, three possible equilibria exist (*i*) only one agent participates, (*ii*) three agents participate, (*iii*) five agents participate. In Figure 4a, we initially have one agent participating and investing in the risky asset (right lower corner). The Participant is connected only to one agent (right upper corner), who in turn requires a minimum of two Participants in her network to start investing in the risky asset. Because the agent in the right upper corner has only one peer who invests in the risky asset, her participation costs stay high and she does not invest herself. As a result, the rest of the agents in the economy do not invest in the risky asset, because they have zero Participants among their peers. The resulting equilibrium level of the risky asset investments for Figure 4a is 1 out of 6, or about 17 percent.

Before we move to the technical details, let us briefly discuss the economic intuition behind the minimum equilibrium. First, we focus on the following information diffusion process. Consider the situation where no agent initially invests in the stock market. All financially educated agents with positive discretionary income will enter the stock market. This must be true for any possible equilibrium. These new participants spread information further to their peers. Among those who get information some have a sufficiently large income and, therefore, hit their participation threshold and enter the stock market. They spread information to their peers. We can continue this process further until no new agent enters the market. What we construct as a result is the equilibrium with the minimum number of participating agents. In the next section, we will provide formal proof that the set of participants obtained through the described information diffusion process is always a subset of participants in any type of static equilibrium.

3.4 THE EQUILIBRIUM WITH THE MINIMUM NUMBER OF PARTICIPANTS

This equilibrium is unique and can be reached through a dynamic information-diffusion process where information goes from participating agents to non-participating agents through active links as described above.[4] We have a matrix of linked agents G and a stack of participation thresholds $K = \{\hat{k}_1, ..., \hat{k}_n\}$ for each agent in the economy. Matrix $G = \{g(i,j), \forall i, j \in \mathbf{N}$ such that $g(i,j) = 1$ if i and j are linked, and $g(i,j) = 0$ otherwise$\}$ represents links between agents, where every active link allows for information sharing.

We apply the following algorithm to find an equilibrium.

Definition 1. *Algorithm 1:*

Step 1 *Create vector* $P = \underbrace{\{0, 0, ..., 0\}}_{N\ times}$ *of Participants.*

Step 2 *Compute the vector* $PN = \{pn_1, pn_2, ..., pn_N\}$*, where* pn_i *is the number of neighbors of agent* i *that are marked as Participants,* $pn_i = \sum_{j=1}^{n} P_j \times G(i,j)$

Step 3 *For* $\forall i \in \{1, N\}$*, if* $\hat{k}_i \leq PN_i$*, we mark this node as Participant, or* $P_i := 1$*. If vector* P *has changed after all iterations, we proceed to* ***Step 2****; otherwise, we have found an equilibrium and the algorithm stops.*

Proposition 1. *Algorithm 1 finds an equilibrium with the minimum number of agents investing in the risky asset.*

Proof. See Appendix C.

[4]By an active link, we mean a link through which agents transmit information relevant to risky asset investment such as investment in stocks. We call each agent who invests in the risky asset an active node.

In our previous example, the equilibrium with the minimum participating agents is illustrated in Figure 4a, where only one initially informed node invests in the risky asset, and information does not pass further.

3.5 PARAMETER VALUES IN SIMULATIONS

We now describe the parameter values used in the simulations. Here it is important to highlight that the main goal of the paper is not to calibrate the model in order to estimate the stock market participation cost similar to the previous literature (see e.g. Vissing-Jørgensen, 2002; Andersen and Nielsen, 2010; Khorunzhina, 2013). We mainly focus on comparative statics to study how changes of combinations of model parameters affect stock market participation. Our values for the fixed model parameters is primarily taken from financial and macro data for United States in 2014.

Fixed parameters – We perform some preliminary computations for model estimation. We assume that the income distribution is log-logistic (Atkinson, 1975).We obtain historical data on annual risk premium $r_{m,t}$ and volatility σ_t.[5] Using these data, we calculate r_u and r_d parameters in our model, assuming equal probabilities for the stock market index to go up or down, $\pi = 0.5$. We obtain stock market participation rates from the Internal Revenue Service's (IRS) Statement of Income (SOI) for individual income tax return (Form 1040) statistics.

Moreover, we add information about the income distribution from the US Census Bureau's 2010-2015 American Community Survey (ACS). The data contains information for the lower bound, upper bound, and mean household income for 2010-2015. We use employment in the financial sector as a proxy for financial education. The number of individuals employed in Finance is obtained from the Quarterly Census of Employment and Wages. We calculate the number of individuals employed in the Finance and Insurance Sector (52 NAICS) in 2015.

[5] Data is obtained from IESE, Social Science Research Network, 2015 (https://www.statista.com/statistics/664840/average-market-risk-premium-usa/).

Given the complexity of calculations, we will approximate the population size by 10,000 in all simulations.

Connectivity and homophily – The connectivity parameter, c, controls for the expected number of links (peers) for each agent in population. We assume that the expected number of peers for each individual is the same, and it is independent of other individual's characteristics. Each individual forms a link with a peer who belongs to her income group with unconditional probability p_{In}, and individual who doesn't belong to the same income group with unconditional probability p_{Out}. The homophily parameter, $h \in [0, 1]$, controls for the difference between unconditional probabilities to form the link with other agents within and outside individual's income group. If the homophily parameter h equals 0, then each individual is equally likely to form a link with any other agent independently of their income. If the homophily parameter h equals 1, then each individual forms connection only inside their own income group. It is important to notice that the homophily parameter and the connectivity parameter are independent.[6] Previous research suggests that individuals on average have maximum of 50 active connections (Arrondel et al., 2021; Mac Carrona et al., 2016). However, only the small number of these links is used to share financial information. Arrondel et al. (2021) find that on average individuals have 7 peers in their financial circles. This is the reference point that we choose for our simulation analysis.

Stock market participation cost – In the model, we endogenize stock market participation cost assuming that individual's stock market participation cost depend on number of informed peers. For simulation analysis, we assume the linear functional form of

[6] We do not describe the entire technical procedure of constructing a connectivity matrix here. However, the general procedure is as follows. We first assign to each agent a number of peers following a binomial distribution with an expected mean of c. Then, at each iteration we take an agent who still need to form some links (the number of formed links is below the assigned number of peers), and considering the number and the types of links (outside or inside the agent's income group) that has been already formed we compute conditional probabilities to form additional links to other peers. Conditional probabilities are recomputed for each agent at each iteration such that the unconditional probabilities to form links inside or outside agent's own income group remain the same for all individuals.

$C(k_i)$ function for Non-Financially Educated agents:[7]

$$C(k_i) = \theta - \Delta\theta k_i$$

The parameter θ controls for stock market participation cost of a non-financially educated agent who is not connected to any informed peers. The parameter $\Delta\theta$ shows the marginal decrease in stock market participation cost when a number of informed links is increasing. Given the data of SMP cost across entire population and participation across different income groups, we calibrate the model to estimate the magnitude of the stock market calibration cost θ and $\Delta\theta$. We get the estimation of cost market around 2000 and the estimation of $\Delta\theta$ around 100 to 200 USD.[8] This estimation give us an idea about the reasonable magnitude of the parameters. However, in the simulation analysis we will allow them to vary in certain ranges. See Table 3 for details.

4. MODEL RESULTS

We now present the main results for the model simulations. We run multiple simulations of our model with fixed market parameters described in Table 3. We focus on how the number of connections, homophily and inequality affect the risky asset investment level, starting with the full popuation. We later explore how each parameter of interest affects different income groups.

[7]As a part of the robustness check analysis, we have tried different functional forms of $C(k_i)$ function. In particular, we ran simulations assuming $C = \theta/k_i^{\alpha}$, where $\alpha > 0$ is an exogenous parameter. We got very similar qualitative results and, therefore, decide to include to the main analysis only the simplest linear form stock market participation cost function.

[8]The previous literature presents the variation in estimates ranging from $260USD$ in Vissing-Jørgensen (2002) to $134000USD$ in Andersen and Nielsen (2010). Our estimate is comparable to the cost of a full Forex trading course.

4.1 THE EFFECT OF CONNECTIVITY

The following section presents the baseline results of the model for selected model parameters. Figure 5 presents the results for how stock market participation varies for several parameters of the model. In the figure, we vary the parameter marked on the x-axis but let all other parameters stay the same. We begin by examining connectivity. Panel a) of Figure 5 shows that stock market participation in increasing in connectivity. Without any connections, the only agents who participate are Financially-educated agents and agents who are wealthy enough to pay the fixed participation costs. If more agents are connected, information spreads more efficiently throughout the economy, and more agents participate.

The effect is S-shaped in connectivity – with few connections, the effect of adding one more connection is small. As the number of connections increases, however, the effect of adding one more connection becomes larger. Intuitively, with higher number of connections, a new connection will create more links with a higher probability, which increases information transmission. As the number of connections increases, more and more agents have the necessary links starts to participate and the effect of additional connectivity is lower.

Panel b) of Figure 5 shows that not all income groups benefit equally from higher connectivity. For agents with high income, denoted by the blue dotted line, the baseline participation rate is somewhat higher, as it is more likely that income is higher than the fixed participation cost regardless of the number of peers. Moreover, since high income agents are more likely to be close to the participation threshold, connectivity has a positive impact on participation. For high income agents, more connections help spread information more efficiently. At some level the effect starts to diminish, again giving an S-shaped pattern. With more than 10 connections, all high income individuals participate.

In contrast, agents with medium income, denoted by the gray dashed line, need a larger number of connections before connectivity starts to have an impact. Middle-income agents are further from the participation threshold, and thus do not benefit as much from an increase in connectivity initially. Once connectivity reaches a sufficient level, however, stock market participation is strongly increasing. At the right side of the graph, the gap in participation between middle and high income agents is small. Finally, the effect of connectivity for low-income agents is small. These agents are very far from the participation threshold, and thus need a very large number of connections before they start to participate.

The level of homophily and inequality also affects the relationship between connectivity and stock market participation. Recall that the homophily parameter measures how likely two agents from different incomes are to connect. Panel a) of Figure 6 plots the connectivity and stock market participation for different levels of homophily. For low levels of connectivity, high homophily leads to more efficient information sharing among agents with relevant information to share. At low levels of connectivity, this leads to higher participation rates. It is only at higher levels of connectivity that high levels of homophily results in lower stock market participation.

A similar pattern appears for inequality. Panel a) of Figure 6 plots the connectivity and stock market participation for different levels of the GINI coefficient. In the simulations we keep the average income level constant but change the distribution. As the GINI coefficient increases, more wealth is concentrated among high-income households. At high levels of inequality, the effect of connectivity is muted, since many agents are far away from the participation threshold. There is no longer an S-shaped relationship between connectivity and participation, as information sharing is inhibited by a high number of agents with little wealth. The relationship is instead approximately linear, with a low level of participation even in a highly-connected society.

The S-shaped relationship between stock market particcipation and connectivity is most pronounced for low levels of inequality. In this economy agents have a more equal share of the same pie, which leads to many agents being far away from the participation threshold. As connectivity increases, however, there are more agents who can benefit from access to information. This leads to a rapid rise in participation as connectivity increases.

4.2 THE EFFECT OF HOMOPHILY

We proceed to examine other important variables related to network structure and information diffusion, beginning with homophily. Figure 7 shows that, somewhat surprisingly, homophily has a positive impact on stock market participation in simulations with low or medium connectivity. To see why, note that for a low level of homophily connectivity is independent of income. Information about stock market participation is then more likely to spread to agents who are far from the participation threshold and thus do not benefit from the information. As homophily increases, information spreads to more similar agents, allowing connectivity to have a higher marginal impact on stock market participation. With a homophily of 1, however, agents are only connected to agents within their own income group, leading to less efficient information sharing. Information will still spread throughout the network, but the effect will be localized to specific income groups. Since only high income groups have enough wealth to start participating, information sharing is limited to this group and participation rates for the population drops rapidly. If connectivity is high, however, all agents are likely connected to informed peers and homophily has no impact on participation.

To illustrate who benefits from higher homophily, Panel b) plots stock market participation against homophily for three separate income groups. We set the level of connectivity to 7, the middle level of connectivity in panel a). Naturally, the high income group participates at a higher level than the low and middle income group. As homophily increases, it is the high income group who increases their stock market participation. Higher homophily increases the likelihood that they connect with someone from their own income group who has information to share. For the other groups, homophily has little impact on their participation rates. The positive relationship between homophily and stock market participation in panel a) is thus driven by higher participation among high-income agents.

4.3 THE EFFECT OF INEQUALITY

We then examine examine how inequality affects stock market participation. In the simulations, we keep average income fixed but vary the distribution of income. Inequality has two effects on participation in the model. First, since we keep average income the same, inequality affects the share of agents who have the ability to pay the fixed participation costs. Second, inequality also affects the probability that agents with information are connected to other agents because of homophily. We now show that the effect of inequality depends on the level of connectivity in the economy.

Figure 8 provide the results for different levels of connectivity. For the black and gray lines at the bottom of the figure, agents have few connections. For these simulations, few agents have the income necessary to pay the fixed cost of participation at low levels of inequality and there is little information sharing. As inequality increases, we take money from the poor and give to the rich, allowing more agents to participate and spread information. This increases stock market participation. In societies with low connectivity, inequality has a positive effect on participation.

In simulations with higher connectivity, we see a negative relationship between inequality and participation. This model features homophily, and connectivity therefore depends on the difference in income. Higher inequality leads to clustering in the network and less efficient information diffusion. This lower stock market participation. For instance, we see that stock market participation is above 80 percent for simulations with high connectivity and a low level of inequality. The share is reduced to less than 40 percent for the most unequal simulation.

4.4 OTHER MODEL PARAMETERS

Finally, we report results for several other model parameters in Figure 9 in Appendix B. These mostly have expected effects. Higher average income implies that more agents meet the fixed cost of investing and stock market participation rises for them. Importantly, however, higher income also creates more informed agents, which in turn helps spread information and increases stock market participation. The effect of higher income therefore works not only through direct stock market participation, but also through more efficient information sharing through connectivity. The share of Financially Educated agents increases stock market participation, as they both directly participate themselves and help seed the network with information. Finally, higher participation costs leads to lower participation.

5. CONCLUSION

In this paper, we introduce a theoretical model describing the portfolio decisions of agents in an economy with a social network. The model describes a mechanism behind the effect of social connectivity on the level of stock market participation in the economy. In the model, social connectivity affects the stock market participation directly and indirectly through interactions with the income distribution and financial-education level.

Empirical tests of the model demonstrate that social connectivity is an important element that helps explain the heterogeneity of the equilibrium stock market participation rates. We show that introducing non-linear stock market participation costs that depend on the network structure in the economy improves the theoretical model predictions and sheds light on and improves our understanding of non-participation among wealthy households and participation of poorer individuals.

First, the theoretical model shows that a non-linear participation-cost function is necessary to explain the heterogeneity of the stock market participation rates. This result is important for understanding the wide range of estimates of participation costs in the literature. According to the model, estimates that do not take connectivity into account will be biased. In essence, it is an omitted-variable bias. For counties with high connectivity, the average estimated costs will be low, and for the economies with low connectivity, the estimated costs will be high. Taking connectivity into account, generates heterogeneity of costs and closely matches the unexplained heterogeneity in participation rates, which helps explain the wide range of estimates for stock market participation costs. It is not the size of the costs that matters, but the functional form.

Second, although in the paper we often refer to peers as a source of information that decreases fixed participation costs, we do not explicitly take a stand on exactly what this information contains. Participation costs that are decreasing with the number of investing peers can be related to the other non-monetary costs of stock investment. For example, a larger number of investing peers can improve the agent's trust in the financial system (Guiso et al., 2008), or increase the "social" utility of investing (Bursztyn et al., 2014).

Third, the results from the theoretical model provide us with important insights for the policy. In the full graph, where all the agents are connected to each other, education does not matter, the effective costs are minimum, and even an agent with a low income can afford to invest in a risky asset to improve her welfare. However, if the network is sparse and not all agents are connected, stock market participation does not reach 100 percent among agents who have positive discretionary income. In reality, we are not connected to everyone else; therefore, clusters will be a necessary part of a social network. In some clusters, there are enough educated agents to spread knowledge, but some clusters of agents will disconnect from the stock market because of a lack of knowledge. If the policy goal is to improve stock market participation, understanding which groups are disadvantaged and targeting them specifically either with financial education or with programs improving their connectivity with the rest of the agents in the economy is important.

We present a simple theoretical model with a number of simplifying assumptions. For example, we assume no correlation between financial education and income, or we assume no correlation between connectedness and wealth or connectedness and income. Although these assumptions may well be violated in the data, introducing them would simply create a higher degree of clustering in the social network and make it more sparse. As a result, connectivity would play a larger role, and our model would generate equilibrium stock market participation rates closer to the ones observed in the data. Our assumptions therefore provide a lower bound for the predictive power of the model. In the future, the model could include more realistic assumptions about correlations between factors and it can be extended to a dynamic case with endogenous network formation. However, the basic message will likely remain: Connectivity to informed peers matters for stock market participation.

REFERENCES

Steffen Andersen and Kasper Meisner Nielsen. Participation constraints in the stock market: Evidence from unexpected inheritance due to sudden death. *The Review of Financial Studies*, 24(5):1667–1697, 2010.

Luc Arrondel, Hector F. Calvo Pardo, Chryssi Giannitsarou, and Michael Haliassos. Informative social interactions. *Working paper*, 2021.

Kenneth Joseph Arrow. *Aspects of the theory of risk-bearing*. Yrjö Jahnssonin Säätiö, 1965.

AB Atkinson. The distribution of wealth in britain in the 1960s the estate duty method reexamined. In *The personal distribution of income and wealth*, pages 277–328. NBER, 1975.

Laurent Bach, Laurent E Calvet, and Paolo Sodini. Rich pickings? risk, return, and skill in household wealth. *American Economic Review*, 110(9):2703–47, 2020.

Claes Bäckman and Tobin Hanspal. Participation and losses in multi-level marketing: Evidence from an ftc settlement. Technical report, 2019.

Michael Bailey, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. Social connectedness: measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80, 2018.

Alon Brav, George M Constantinides, and Christopher C Geczy. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy*, 110(4):793–824, 2002.

Joseph S Briggs, David Cesarini, Erik Lindqvist, and Robert Östling. Windfall gains and stock market participation. *National Bureau of Economic Research*, 2015.

Jeffrey R Brown, Zoran Ivković, Paul A Smith, and Scott Weisbenner. Neighbors matter: Causal community effects and stock market participation. *The Journal of Finance*, 63(3): 1509–1531, 2008.

Leonardo Bursztyn, Florian Ederer, Bruno Ferman, and Noam Yuchtman. Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, 82(4):1273–1301, 2014.

John Y Campbell. Household finance. *The Journal of Finance*, 61(4):1553–1604, 2006.

Frederick K Changwony, Kevin Campbell, and Isaac T Tabner. Social engagement and stock market participation*. *Review of Finance*, page rft059, 2014.

Raj Chetty, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenberg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barbera, Monica Bhole, and Nils Wernerfelt. Social capital i: Measurement and associations with economic mobility. *Nature*, 608(7921):108–121, 2022a.

Raj Chetty, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenberg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barbera, Monica Bhole, and Nils Wernerfelt. Social capital ii: Determinants of economic connectedness. *Nature*, 608(7921):122–134, 2022b.

Y Chien, Paul Morris, et al. Household participation in stock market varies widely by state. *The Regional Economist*, 3:4–5, 2017.

Maeve Duggan, Nicole B Ellison, Cliff Lampe, Amanda Lenhart, and Mary Madden. Demographics of key social networking platforms. *Pew Research Center*, 9, 2015.

Andreas Fagereng, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. Heterogeneity and persistence in returns to wealth. *Econometrica*, 88(1):115–170, 2020.

Eugene F Fama and Kenneth R French. The equity premium. *The Journal of Finance*, 57 (2):637–659, 2002.

Luigi Guiso, Paola Sapienza, and Luigi Zingales. Trusting the stock market. *The Journal of Finance*, 63(6):2557–2600, 2008.

Michael Haliassos and Carol C Bertaut. Why do so few hold stocks? *the economic Journal*, pages 1110–1129, 1995.

Michael Haliassos, Thomas Jansson, and Yigitcan Karabulut. Financial literacy externalities. *The Review of Financial Studies*, 33(2):950–989, 2020.

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

Harrison Hong, Jeffrey D Kubik, and Jeremy C Stein. Social interaction and stock-market participation. *The Journal of Finance*, 59(1):137–163, 2004.

Chih-Ching Hung. Does social connectedness affect stock market participation? *Available at SSRN 3724407*, 2021.

Hans K Hvide and Per Östberg. Social interaction at work. *Journal of Financial Economics*, 117(3):628–652, 2015.

Markku Kaustia and Samuli Knüpfer. Peer performance and stock market entry. *Journal of Financial Economics*, 104(2):321–338, 2012.

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

Samuli Knüpfer, Elias Henrikki Rantapuska, and Matti Sarvimäki. Why does portfolio choice correlate across generations. *Bank of Finland Research Discussion Paper*, (25), 2017.

P Mac Carrona, K. Kaski, and R. Dunbar. Calling dunbar's numbers. *Social Networks*, 47: 151–155, 2016.

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

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

Rajnish Mehra and Edward C Prescott. The equity premium: A puzzle. *Journal of monetary Economics*, 15(2):145–161, 1985.

Esteban Ortiz-Ospina. The rise of social media. Published online at OurWorldInData.org. Retrieved from: https://ourworldindata.org/rise-of-social-media [Online Resource], 2019. Accessed: 2022-09-14.

Paige Ouimet and Geoffrey Tate. Learning from coworkers: Peer effects on individual investment decisions. *The Journal of Finance*, 75(1):133–172, 2020.

Eleonora Patacchini and Edoardo Rainone. Social ties and the demand for financial services. *Journal of Financial Services Research*, 52(1-2):35–88, 2017.

Annette Vissing-Jørgensen. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy*, 110(4):825–853, 2002.

Yves Zenou. Networks in economics. *Available at SSRN 2153453*, 2012.

6. FIGURES

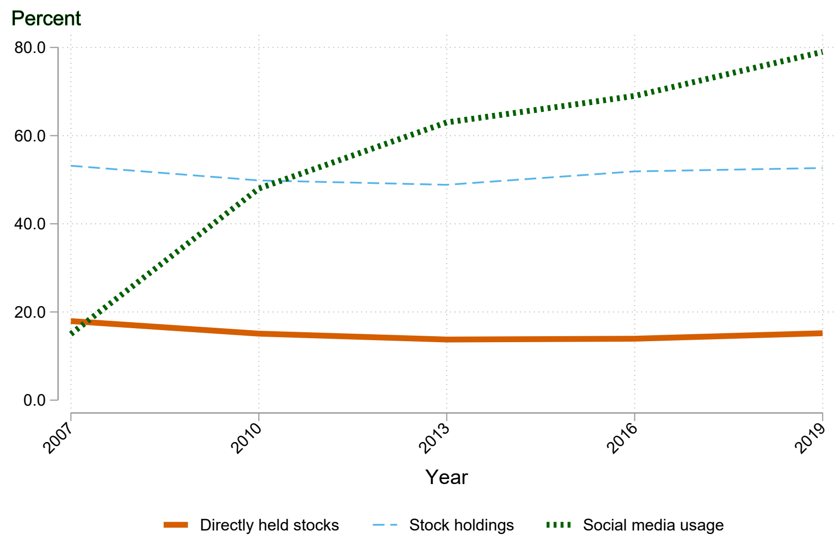


Figure 1: Stock market participation and social media usage over time

Note: The figure plots the share of household who directly own stock (orange solid line), the share of household who have some direct or indirect stock holding (blue dashed line), and the percentage of US adults who use at least one social media site (green dotted line) over time.
Source: Survey of Consumer Finances and Ortiz-Ospina (2019)

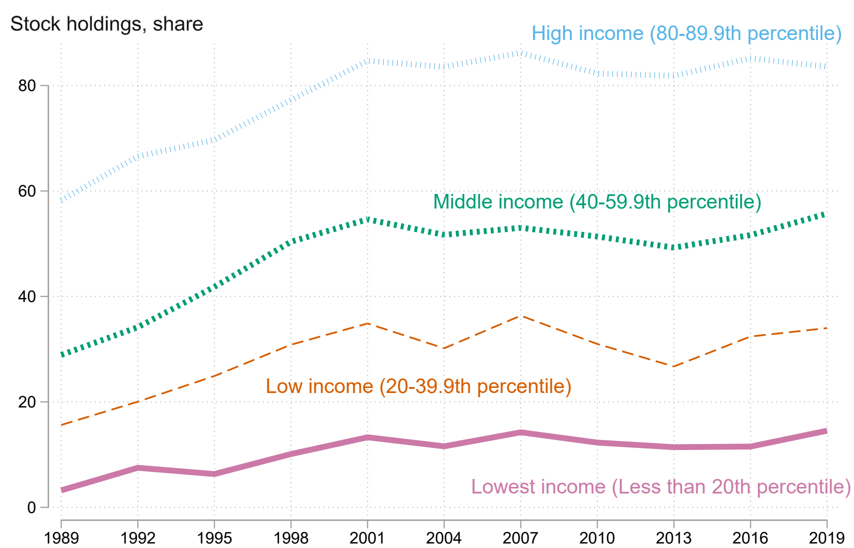
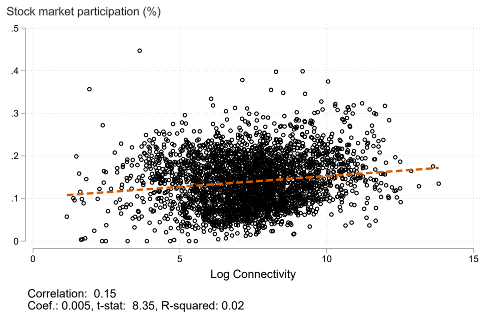
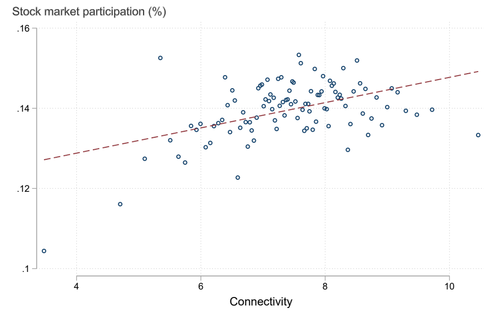


Figure 2: Stock market participation by different income groups over time

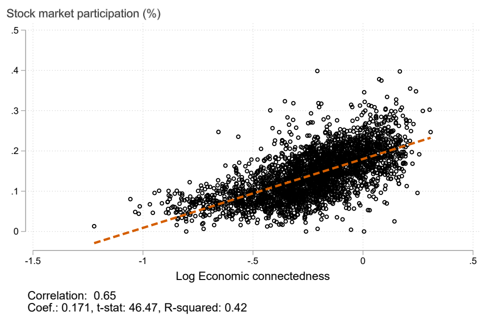
Note: The figure plots the share of household who directly own stock (orange solid line) by income. Source: Survey of Consumer Finances.



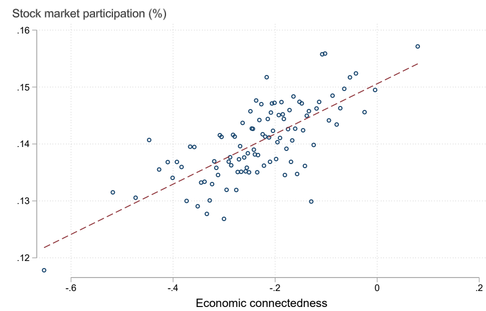
(a) Log Connectivity index



(b) Log Connectivity index with controls



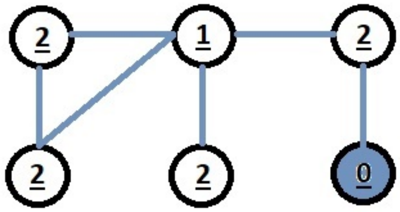
(c) Log Economic Connectivity



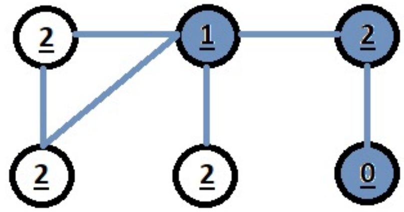
(d) Log Economic Connectivity with controls

Figure 3: Stock market participation and different measures of Connectivity

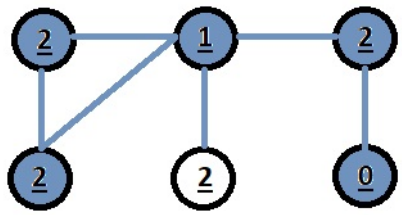
Note: All figures plot stock market participation on the county level on the y-axis. All figures on the right side include controls for county-level age, age squared, median household income, the share of financially educated, and the share of the county has a bachelor-level education or above, as well as fixed effects for state. Panel a) and b) plot Log Connectivity on the x-axis, where Connectivity is an index of the within-county connectivity measure from Facebook. We remove counties in the 99th percentile of connectivity. Panel c) and d) instead plots Log Economic Connectedness on the x-axis.



(a) $SMP = \frac{1}{6}$



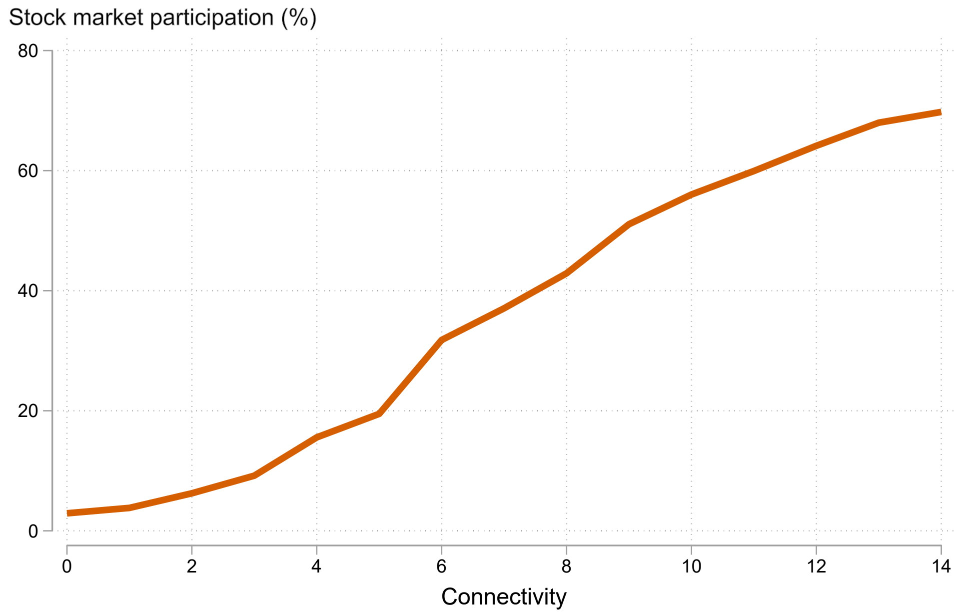
(b) $SMP = \frac{3}{6}$



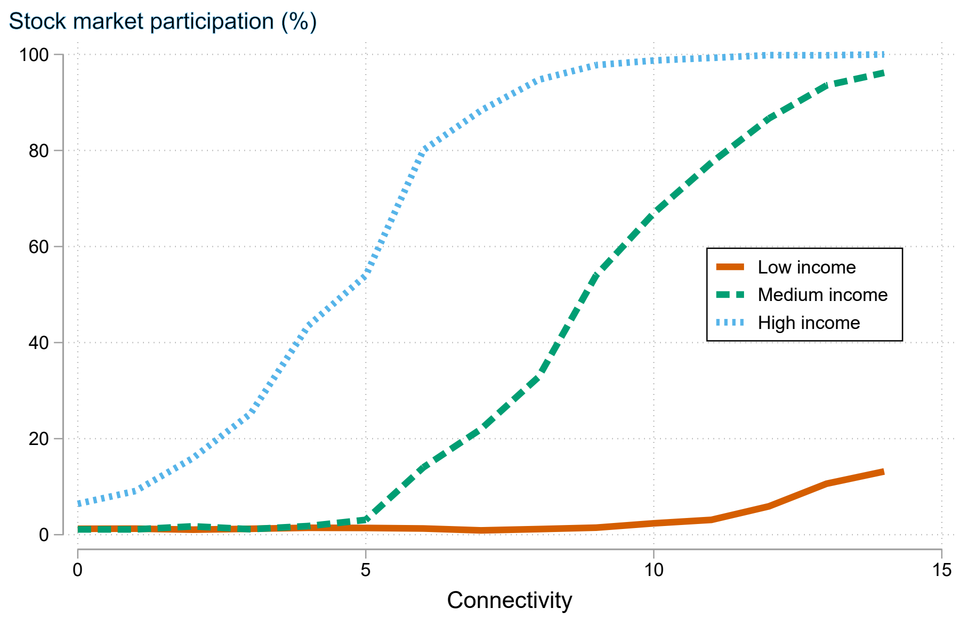
(c) $SMP = \frac{5}{6}$

Figure 4: Possible equilibria

Note:



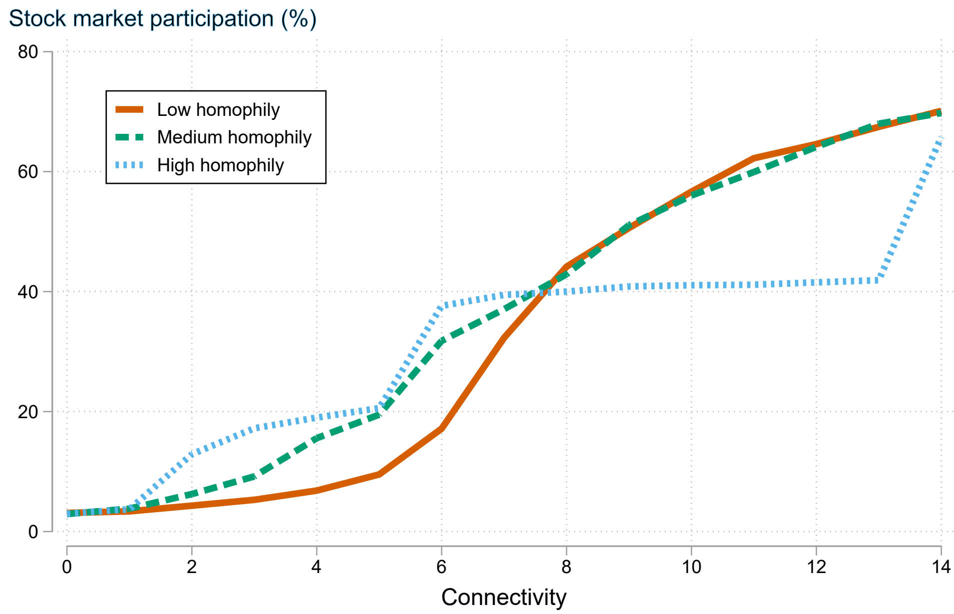
(a) Connectivity



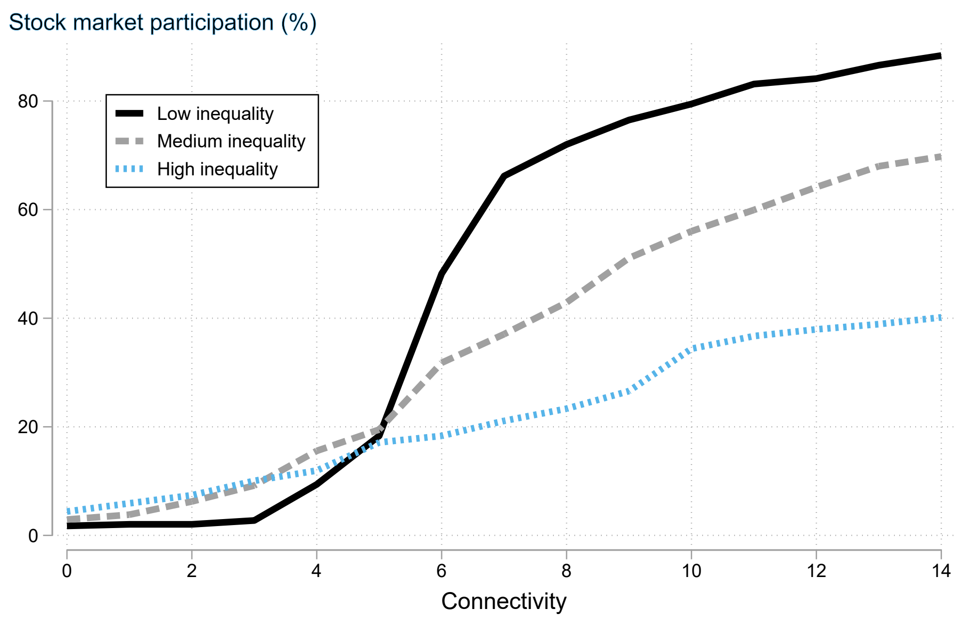
(b) By income groups

Figure 5: The effect of connectivity by model parameters

Note:



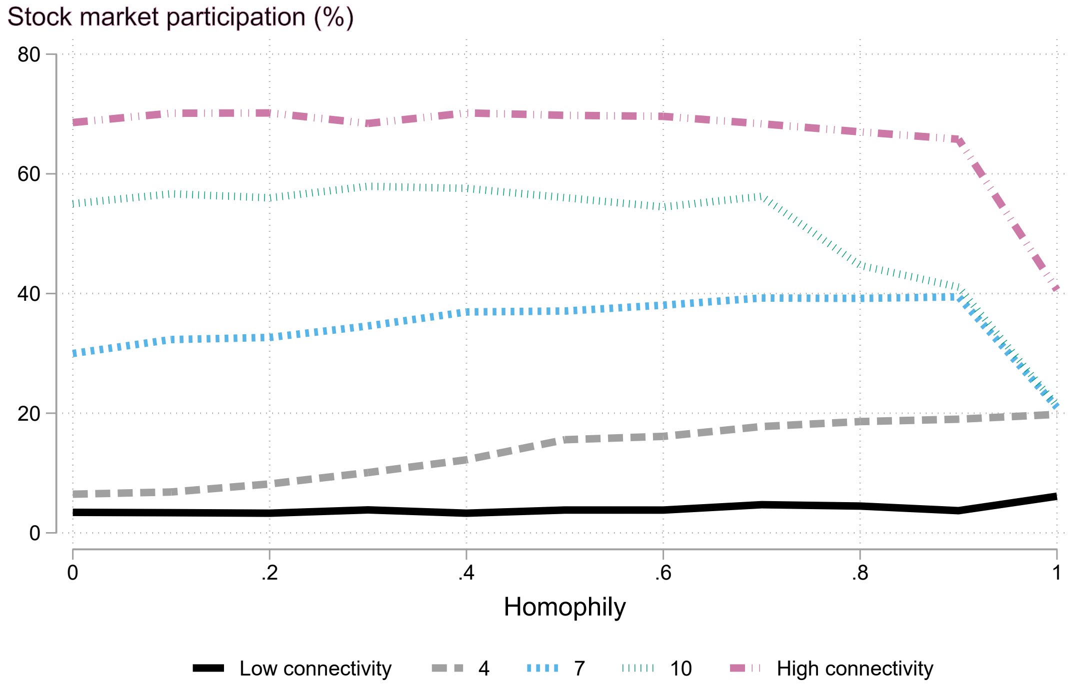
(a) Homophily



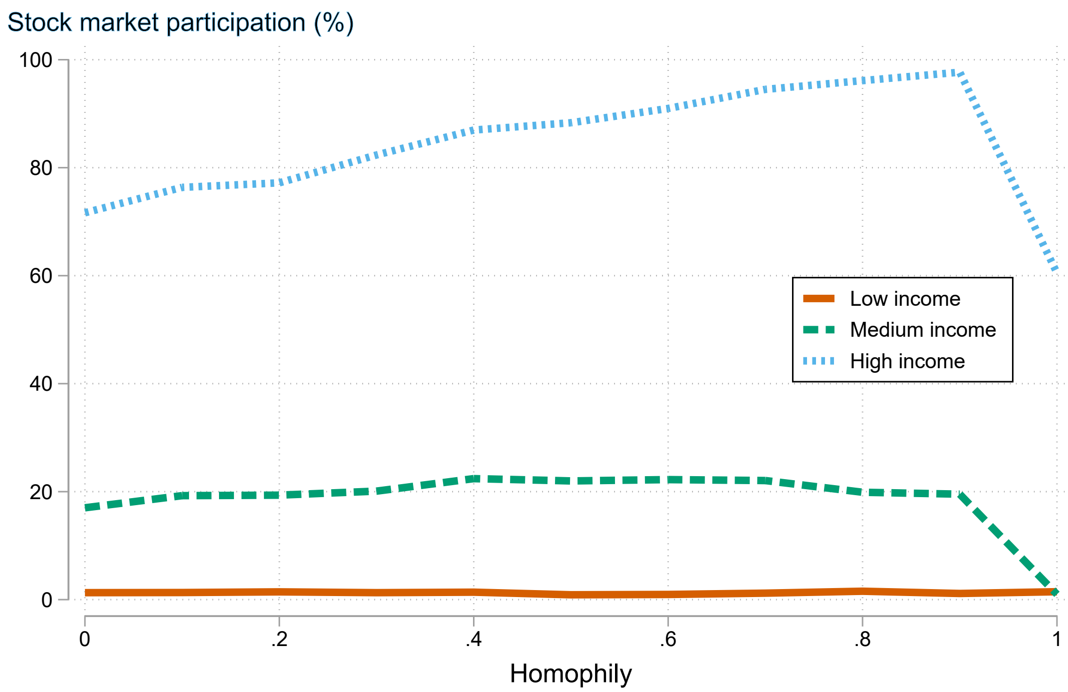
(b) Inequality

Figure 6: The effect of connectivity by levels of homophily and inequality

Notes:



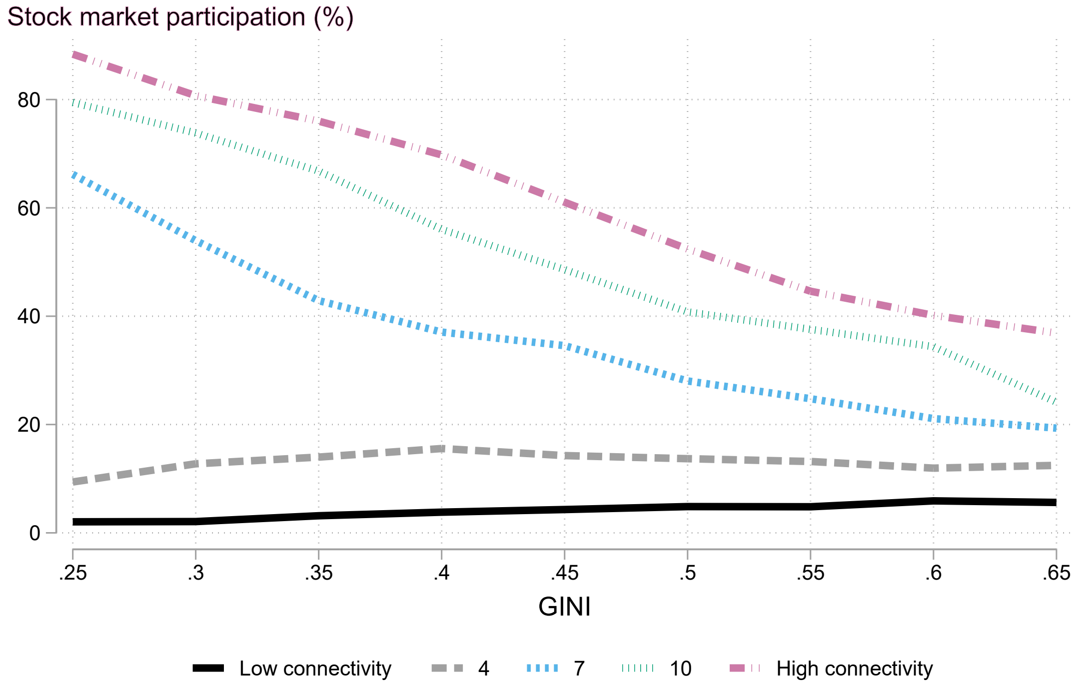
(a) Population-level



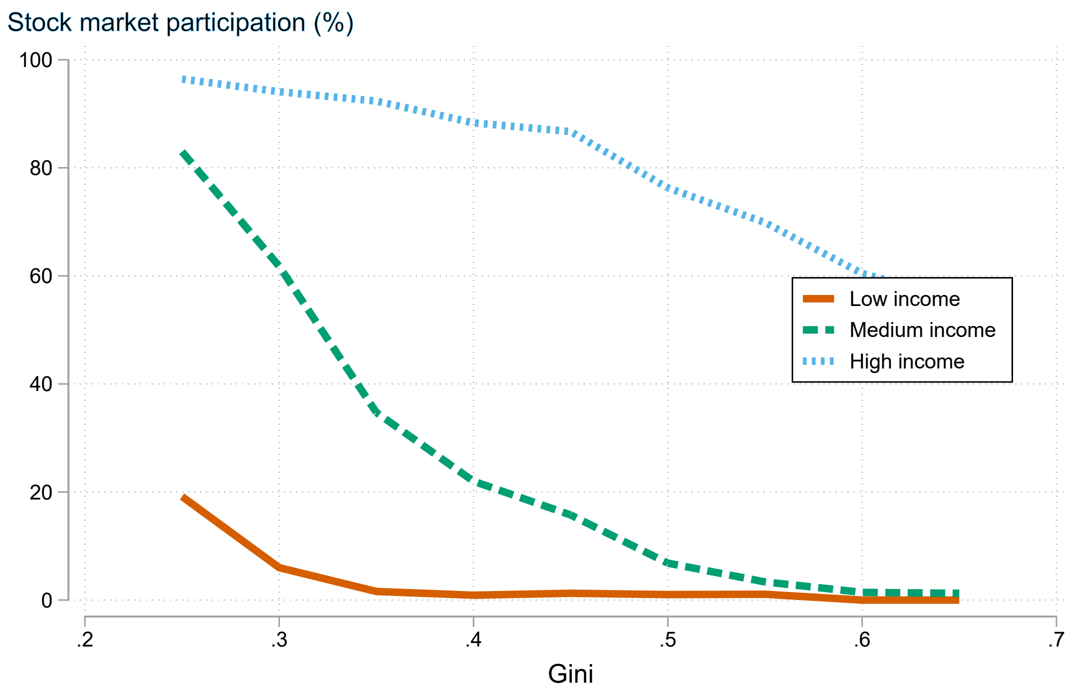
(b) By income groups

Figure 7: The effect of homophily on stock market participation

Notes:



(a) Population-level



(b) By income groups

Figure 8: The effect of inequality on stock market participation

Notes:

7. TABLES

Table 1: Descriptive statistics

	Mean	Median	Std. dev.	Min	Max
Stock Market Participation	0.140	0.137	0.061	0.000	0.447
Connectivity measures					
Economic connectedness	0.812	0.806	0.176	0.295	1.360
Ec. Connectedness - high SES among low SES	0.848	0.839	0.213	0.187	1.476
Ec. Connectedness - high SES among high SES	1.252	1.257	0.177	0.701	1.715
Friendship exposure	0.904	0.902	0.212	0.270	1.486
Friending bias	0.064	0.064	0.050	-0.108	0.335
Friendship clustering	0.116	0.115	0.020	0.072	0.222
Inside Connectivity index	8,539.620	1,693.562	33,669.911	3.162	1000000.000
Inside Connectivity index per capita	0.066	0.060	0.034	0.001	0.241
Demographics					
Median Age, county	41.300	41.300	5.268	23.200	66.600
Log Median Household Income	10.663	10.654	0.240	9.870	11.658
Financial Employment	0.013	0.011	0.011	0.000	0.224
Share of African Americans	0.084	0.008	0.147	0.000	0.861
Share of Women	0.501	0.505	0.022	0.304	0.575
Share of Hispanic Americans	0.070	0.020	0.132	0.000	0.983
Metropolitan Area	0.370	0.000	0.483	0.000	1.000
County Population, in 1000s	97.794	26.087	312.754	0.489	9,758.256

Notes: Economic connectedness is two times the share of high-SES friends among low-SES individuals, averaged over all low-SES individuals in the county. Friendship exposure is the mean exposure to high-SES individuals by county for low-SES individuals. Friendship clustering is the average fraction of an individual's friend pairs who are also friends with each other.

Table 2: Social connectivity and Stock market participation

	Economic connectedness			Connectivity index			All		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log economic connectedness	0.17***	0.046***	0.051***				0.18***	0.055***	0.056***
	(0.011)	(0.0092)	(0.0088)				(0.012)	(0.0095)	(0.0091)
Log connectivity				0.0051**	0.0021**	0.00076	0.0095***	0.0037***	0.0024**
				(0.0020)	(0.0010)	(0.00095)	(0.0011)	(0.00100)	(0.00091)
Control variables									
Median Age, county		0.0044*	0.0033		0.0043*	0.0035		0.0041*	0.0032
		(0.0025)	(0.0022)		(0.0025)	(0.0023)		(0.0024)	(0.0022)
Age Squared		-0.000017	-0.0000019		-0.0000072	0.0000024		-0.0000082	0.0000016
		(0.000033)	(0.000029)		(0.000032)	(0.000029)		(0.000031)	(0.000028)
Log Median Household Income		0.036***	0.037***		0.055***	0.057***		0.032***	0.033***
		(0.0064)	(0.0060)		(0.0060)	(0.0056)		(0.0067)	(0.0062)
Financial Employment		0.44***	0.47***		0.31**	0.38***		0.34***	0.40***
		(0.12)	(0.11)		(0.12)	(0.12)		(0.12)	(0.12)
Bachelor degree over 25, share		0.32***	0.32***		0.34***	0.34***		0.30***	0.30***
		(0.020)	(0.020)		(0.018)	(0.018)		(0.021)	(0.020)
State FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Industry FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	2950	2950	2949	2950	2950	2949	2950	2950	2949
R-squared	0.423	0.794	0.806	0.020	0.788	0.797	0.492	0.798	0.808

Notes: The table provides results where we regress stock market participation at the county-level against connectivity measures and controls. Control variables include county-level age, age squared, median household income, the share of financially educated, and the share of the county has a bachelor-level education or above. Fixed effects for state and the main industry of employment for the county are indicated. Standard errors are clustered by state are presented in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Model Parameters

Description	Parameter	Benchmark Value	Range
Wealth distribution	$F(W_i)$	Log-logistic(α, β)	
Relative risk aversion	γ	2	
Prob. high return	π	0.5	
High return	r_u	0.1629	
Low return	r_d	-0.0549	
Size of economy	n	10000	
Gini index		0.36	$[0.15, ..., 0.45]$
Average income		41905	$[21905, ..., 61905]$
Minimum wage		$\$30,763$	
Constant in fixed cost	θ	2000	$[1000, ..., 4000]$
Exponent in fixed cost	$\Delta\theta$	200	$[50, ..., 500]$
Homophily parameter		0.5	$[0, ..., 1]$

Notes: We choose the log-logistic wealth distribution because it is consistent with the data we use for the analysis. π is the probability of the net return equal to r_u. r_u and r_d are the realizations of the net return of the risky asset in two states, where $r_d < 0 < r_u$. γ is the relative risk aversion coefficient. The disposable income which an agent can invest in the stock market is equal her labor income minus minimal cost of living, approximated by minimum wage. The fixed participation cost for Financial Non-educated agents is given by $F(t_i) = \theta - k_i\Delta\theta$.

INTERNET APPENDIX

FOR ONLINE PUBLICATION

A. DATA SOURCES AND DESCRIPTION

We use detailed USA county-level data for income, financial employment, stock market participation, and social connectivity. First, we collect data for the average within-county connectivity levels. Specifically, we use the Social Connectedness Index (SCI) from Bailey et al. (2018), where authors construct a measure of social connectedness between US county-pairs. This measure is constructed as an index based on the number of friendship links on Facebook[9], where the average number of links is normalized to the largest number of connections for a Los Angeles County - Los Angeles County pair.[10]

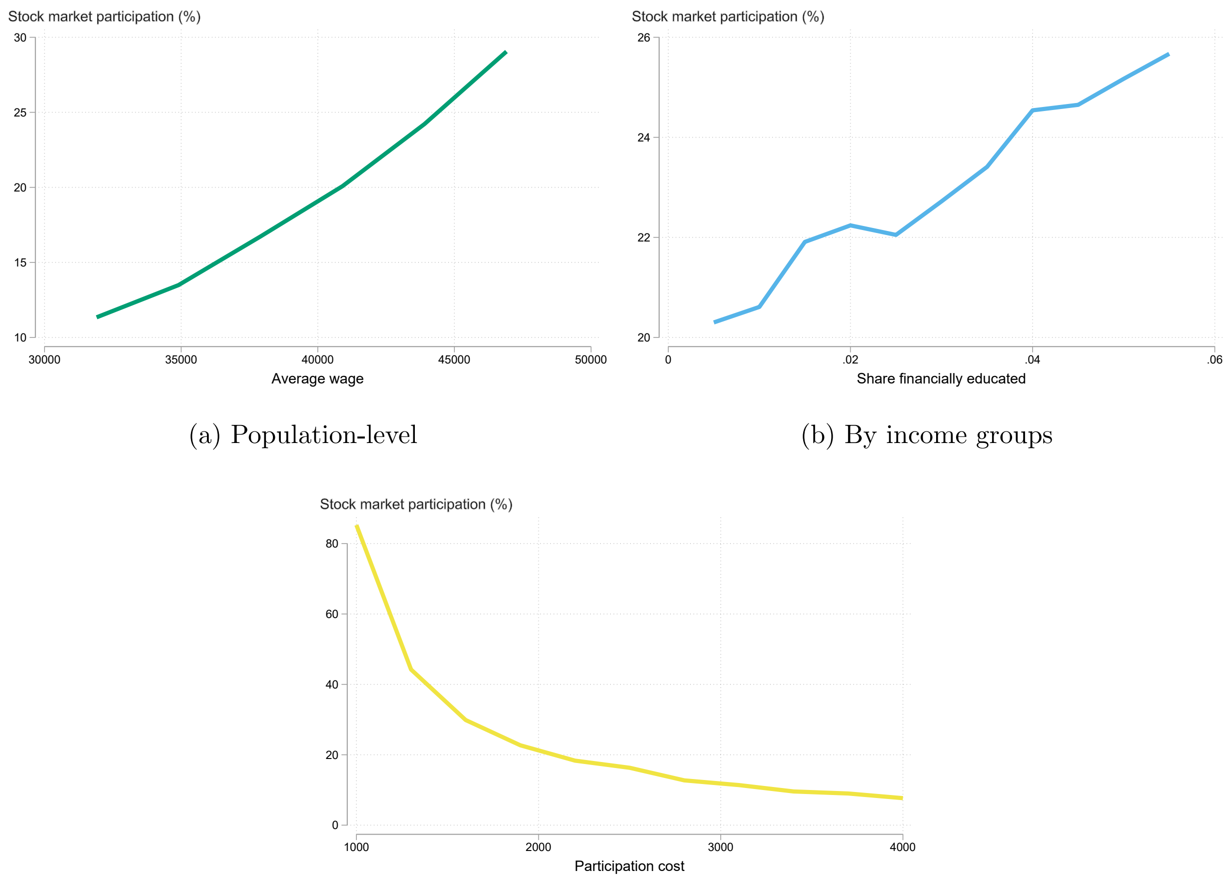
We obtain stock market participation rates from the Internal Revenue Service's (IRS) Statement of Income (SOI) for individual income tax return (Form 1040) statistics following the procedure described in Bäckman and Hanspal (2019), where the fraction of tax returns claiming ordinary dividends are used as an indication of stock market participation within the county. See also Chien et al. (2017), who uses the same data to calculate state-level participation, and Hung (2021), who calculates county-level participation and provides a detailed validation of the measure. We add information about the income distribution in each county from the US Census Bureau's 2010-2015 American Community Survey (ACS).[11] We use employment in Finance and Insurance Sector (52 NAICS) in 2015 from the Quarterly Census of Employment and Wages as a proxy for financial education.

[9] Duggan et al. (2015) report that as of September 2014, more than 58 percent of the US adult population and 71 percent of the US online population used Facebook. The same source reports that, among online US adults, Facebook usage rates are relatively constant across income groups, education groups, and racial groups.

[10] The SCI for Los Angeles County - Los Angeles county is equal to 1,000,000

[11] The data contains information for the lower bound, upper bound, and mean household income. We assume that the income distribution for different counties in the US is log-logistic (Atkinson, 1975). This assumption is consistent with the income distribution that we observe in the data. For more details, see Appendix **??**

B. APPENDIX: FIGURES



(a) Population-level

(b) By income groups

(c) By income groups

Figure 9: The effect of model parameters on stock market participation

C. APPENDIX: PROOFS

Proof of Proposition 1. The algorithm returns the set of participating agents that we denoted by P. Let's denote the set of all agents, who don't invest in the risky asset as $NP = \mathbf{N} \setminus P$. All agents are marked as non-participants at the initial stage. Let's denote the set of participants at the initial stage as $P_0 = \emptyset$. We denote by P_j the updated set of participants at iteration j, $j = 0, ..., T$. Hence, at the final iteration T, the algorithm returns the set $P \equiv P_T$. We also denote by $pn_{i,j}$ the number of participating neighbours of agent i at iteration j.

First, we show that the set P is an equilibrium set. We show the result in two steps:

1. We show that each agent belonging to set P doesn't want to reverse her decision and stop investing. By construction, we are adding each agent i, who is initially marked as a non-participant, to set P_j only if $\hat{k}_i > pn_{i,j}$. As it is always true that $P_j \in P_{j+1}$, then if condition $\hat{k}_i < pn_{i,j}$ is satisfied for agent i at step j, it can't be reversed at iteration $j' > j$, so $\hat{k}_i < pn_{i,j'}$.Therefore, each agent who belongs to set P has a strictly positive payoff and doesn't want to reverse her investing decision.

2. We show that each agent, who belongs to set NP, doesn't want to invest into a risky asset. At each stage, the algorithm adds to set P any agent i who is marked as non-participant if condition $\hat{k}_i > pn_{i,j}$ is satisfied. The algorithm stops when it is not possible to add any agents satisfying the property. Therefore, for all remaining agents who belong to set NP, it must be true that $\hat{k}_i \leq pn_i$ where $i \in NP$. Therefore, for each agent from set NP, it is not profitable to reverse an investing decision.

Second, we show that there is no equilibrium where smaller number of agents invest into risky asset. Suppose that there is such equilibrium defined by the set of participating agents P' where $|P'| < |P|$. We denote by $NP' = \mathbf{N} \setminus P'$.

Then set P' should include all agents with $\hat{k}_i = 0$. If these agents are included then their neighbors with $\hat{k}_i \leq 1$ must be included in P' and so on. Thus, we get all agents from the set P being included to set P', thus $P \in P'$. It is a contradiction. Hence set P has the minimum power among all possible equilibrium partitions. Then set P' should include all agents with $\hat{k}_i = 0$. If these agents are included then their neighbors with $\hat{k}_i \leq 1$ must be included in P' and so on. Thus, we get all agents from the set P must be included to set P', thus $P \in P'$. We got a contradiction. Hence, set P has the minimum power among all possible equilibrium partitions.

Proof of Proposition **??**. The algorithm returns the set of participating agents that we denoted by P. Let's denote the set of all agents, who don't invest in the risky asset as $NP = \mathbf{N} \setminus P$. All agents are marked as participants at the initial stage. Let's denote the set of participants at the initial stage as $P_0 = \mathbf{N}$. We denote by P_j the updated set of participants at iteration j, $j = 0, ..., T$. Hence, at the final iteration T, the algorithm returns the set $P \equiv P_T$. We also denote by $pn_{i,j}$ the number of participating neighbours of agent i at iteration j. First, we show that the set P is an equilibrium set. We show the result in two steps:

1. We show that each agent belonging to set P doesn't want to reverse her decision and stop investing. By construction, the algorithm deletes all agents, for whom condition $\hat{k}_i > pn_{i,j}$ is violated, from the set P_j at iteration j. The algorithm stops when at some iteration it cannot further delete any agent from the set of participants. Therefore, for all agents belonging to P_T, condition $\hat{k}_i > pn_i$ is not violated by construction. All these agents have positive payoffs, and therefore, they do not want to reverse their investing decisions.

2. We show that each agent, who belongs to set NP, doesn't want to invest into a risky asset. Again, we proof the result by construction. All agents belong to set P_0 at the initial stage. The algorithm deletes all agents, for whom condition $\hat{k}_i < pn_{i,j}$ is violated at some iteration. As it is always true that $P_{j+1} \in P_j$, then if condition $\hat{k}_i < pn_{i,j}$ is

violated for agent i at step j, it can't be reversed at iteration $j' > j$, so $\hat{k}_i > pn_{i,j'}$. Therefore, each agent from set NP can't reverse her investment decision as she would get a negative payoff.

Second, we show that there doesn't exists an equilibrium with larger number of participating agents. Let's denote our original game with the set of agents $\mathbf{N}$ as G_0. Suppose, that initially all agents are informed and share information. If under this condition everybody wants to participate, then we have an equilibrium which includes all agents. If there is an agent i who does not want to participate even if all her neighbours are informed, then this agent can not belong to a set of participants in any equilibrium. As she does not belong to the set of participants, she also can not share any information. So we can exclude this agent and all her links without affecting the outcome because this agent and all her links are not active in any equilibrium. Therefore, we now consider the game G_1 with the set of agents $\mathbf{N} \setminus \{i\}$. Now we apply the same exclusion procedure to the game G_1. At each iteration i, we delete an agent and get an updated game G_i, where the game G_i has the same equilibria[12] as game G_0. At some step T, we cannot remove any agent because either the set of agents is empty, or all remained agents want to participate given that all of them share the information about the stock market. So, we get a game G_T which has an equilibrium where all agents want to participate. This is an equilibrium with the maximum number of participants of game G_T, and therefore, of the original game G_0. Hence, we proved the result by construction.

[12]Notice that by an equilibrium we mean here an equilibrium set of participating agents.