Beyond Connectivity: Stock Market Participation in a Network[*]

Olga Balakina   Claes Bäckman   Anastasiia Parakhoniak

July, 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 to show that the effect of connectivity on stock market participation depends on how efficient information spreads, which is linked to how agents are connected, homophily and inequality. High-income agents benefit more from connectivity, leading to increased inequality. We discuss the implications for access to financial information, wealth inequality, and stock market participation.

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

[*]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.

Conflict-of-interest disclosure statement

Olga Balakina

I have nothing to disclose

Claes Bäckman

I have nothing to disclose

Anastasiia Parakhoniak

I have nothing to disclose

1. INTRODUCTION

The past twenty years of explosive growth of social media have 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), Figure 1 shows that the share of direct or indirect stock holdings has been flat all income groups between 2001 and 2019. Given the large literature that argues that social networks help promote stock market participation (Brown *et al.*, 2008; Kaustia & Knüpfer, 2012; Hvide & Östberg, 2015; Knüpfer *et al.*, 2017; Arrondel *et al.*, 2022; Haliassos *et al.*, 2020; Ouimet & Tate, 2020), it is important to understand why an unprecedented increase in our ability to share information and connect with others have not led to an increase in direct stock market participation.

In this paper, we argue that the effect of connectivity is heavily dependent on homophily and inequality since these factors shape the information content in social networks. 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. Consistent with this assertion, we show that a general connectivity measure, the Social Connectedness Index (SCI) from Facebook (Bailey *et al.*, 2018), does not predict variation in stock market participation across US counties. Instead, stock market participation strongly correlates with *economic connectedness*, a measure that conveys information about the connectivity between individuals with a high- and low socioeconomic status (Chetty *et al.*, 2022a).

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 do inequality and homophily mediate the effect of connectivity? Specifically, 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. Fixed costs capture monetary, behavioral, and psychological costs that make stock ownership un-

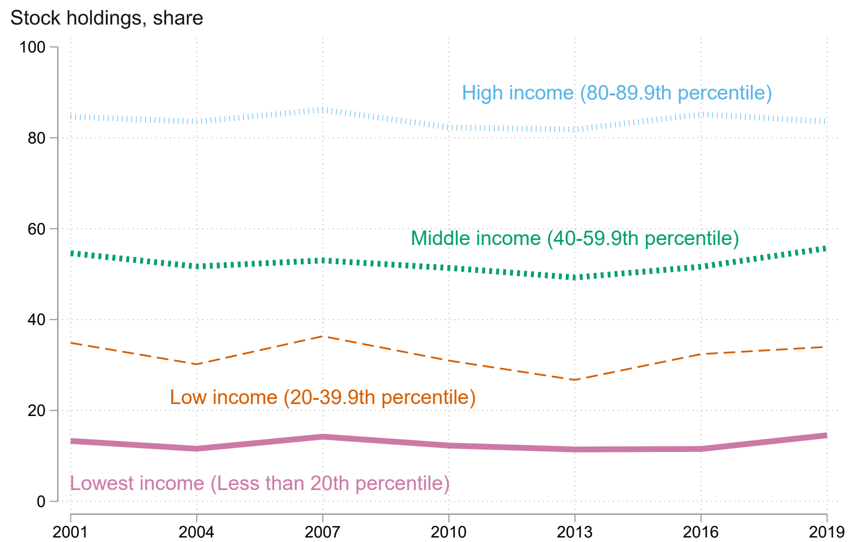


Figure 1. Stock market participation by income group

Notes: The figure plots the share of households with direct or indirect stock holding for different income groups over time. Source: Survey of Consumer Finances

comfortable for some households (Campbell, 2006).[1] We let the fixed cost depend on the number of *informed* agents in each agent's network. An informed agent is any agent that already participates in the stock market. Moreover, we introduce two types of agents into the model, which differ in their financial education and 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 informed peers. All agents in the economy are connected with an ex-ante connectivity parameter that determines each agent's expected number of links.

We also introduce a homophily parameter into the model, defined through the difference between the probability of connecting with an individual with a similar income and the probability of connecting with one from a different income group. Homophily in human interactions has long been studied in sociology and economics (Verbrugge, 1977; Jackson, 2014), and refers to the tendency of people to associate with others that are like them. In our

[1] Participation costs can be defined as money and time spent to invest in the stock market (Haliassos & Bertaut, 1995; Briggs *et al.*, 2021), or as an economist's representation of behavioral and psychological factors that make stock ownership uncomfortable for some households (Campbell, 2006). Both channels are likely to coexist; however, the previous literature mostly evaluates the first type.

model, high homophily means that agents with low income are more likely to be connected to low-income agents than to high-income agents, and vice versa.[2]

The model endogenously generates an S-shape relationship between connectivity and stock market participation. At low levels of connectivity, information sharing is limited and participation rates are low. Keeping all other model parameters constant, as connectivity increases, more agents participate, become informed and spread information, and participation rapidly increases. However, the information diffusion process slows down at higher levels of connectivity, which reduces the additional gain from increased connectivity. We show that increased connectivity mainly benefits richer agents, who are closer to the participation threshold and thus need fewer informed connections to participate. Increased connectivity generates more ex-post inequality within our model.

Homophily significantly affects the relationship between connectivity and stock market participation in nuanced ways. The positive correlation between homophily and stock market participation is the most pronounced when connectivity is low. In sparse networks, higher homophily almost always positively impacts stock market participation because of more efficient information transmission among rich agents.[3] High homophily compensates for low connectivity and allows rich agents to enter the stock market, improving average participation. The increase in average participation rates is driven by wealthier agents, which increases ex-post inequality. However, once all rich agents participate, the same low connectivity and high homophily prevent poor agents from starting to invest in stocks. As agents become more likely to be connected to agents with similar incomes, the model generates clusters of high-income agents who can cover the fixed costs and clusters of low-income agents without an opportunity to learn from informed peers. Thus, increasing homophily leads to more efficient information sharing and higher participation rates *among rich agents*, with little impact on the participation by middle and low-income agents. Homophily in our model is a double-edged sword, in the words of Jackson (2021).

[2]Although we have chosen to focus on income, there is evidence of homophily in many other characteristics, for example, age, gender, years of schooling, religion (Verbrugge, 1977), and there is also evidence of homophily in personality characteristics (Morelli *et al.*, 2017) and risk aversion (Jackson *et al.*, 2023).

[3]However, once homophily is very high or equal to one, the effect turns negative with higher homophily leading to lower stock market participation. In extreme cases, some disjoint components prevent information transmission between different income groups.

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, increasing inequality shrinks the share of agents who can pay the fixed participation costs, thus decreasing the stock market participation. Second, inequality also affects the probability that agents with information are connected to other agents because of homophily. Information sharing and participation are limited with low connectivity. Higher inequality can generate higher participation rates since income is concentrated among agents who can almost cover the fixed cost. With higher connectivity, however, increased inequality generates lower participation since inequality leads to higher clusterization and, therefore, affects the likelihood of being connected to informed peers.

In conclusion, our paper argues that while connectivity is important, we also need to examine the general network structure to understand why the rapid rise of online social networks has not led to an increase in stock market participation. Homophily is a powerful force in human interactions that can inhibit the effect of technological advances. It is also plausible that algorithms in social networks work to increase homophily because platforms can increase engagement by steering individuals toward others that share their backgrounds.

Our paper has three main contributions. First, a large literature investigates the economic and social drivers of inequality (see Jackson, 2021, and citations within). Homophily in social networks is linked to inequality through unequal access to jobs through social connections, unequal awareness of opportunities and unequal information on how to take advantage of opportunities, and differences in norms. Recent studies have documented that wealthier households earn higher returns (Bach *et al.*, 2020; Fagereng *et al.*, 2020) because of heterogeneity in individual skill, risk exposure, or access to information. Our framework suggests that differences in financial information can arise because of homophily in social networks.

Second, a large literature has documented that social interactions between agents affect financial decisions (Kaustia & Knüpfer, 2012; Bursztyn *et al.*, 2014; Changwony *et al.*, 2014; Patacchini & Rainone, 2017; Balakina, 2022; Balakina *et al.*, 2023). We argue that a potentially important yet overlooked aspect of peer effect in finance is to examine the distribution and clustering of informed agents.[4] Most of the empirical literature on peer effects in stock

[4] An exception is Fagereng *et al.* (2022), who examine sorting due to assortative mating.

market participation focuses on the challenging question of documenting that peer effect exists but spends little time investigating who has access to informed peers.

Third, the limited stock market participation puzzle has been a major subject in finance dating back to Arrow (1965).[5] 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 & Bertaut, 1995; Vissing-Jørgensen, 2002). Recent papers have also argued that entry- and exit rates are important for understanding the limited participation puzzle (Bonaparte *et al.*, 2018; Brandsaas, 2021). Empirical work on participation costs suggests that many households face high psychic costs to participation Andersen & Nielsen (2010); Briggs *et al.* (2021). By allowing participation costs to vary with social connectivity, we generate heterogeneity in stock market participation costs unrelated to income or financial education. This can help explain the flat participation rates over time, even in the face of falling *monetary* costs of investing in the stock market.[6]

The rest of the paper proceeds as follows. Section 2 discusses the motivating evidence, Section 3 provides the model and Section 4 provides the results from the simulations. Section 5 concludes.

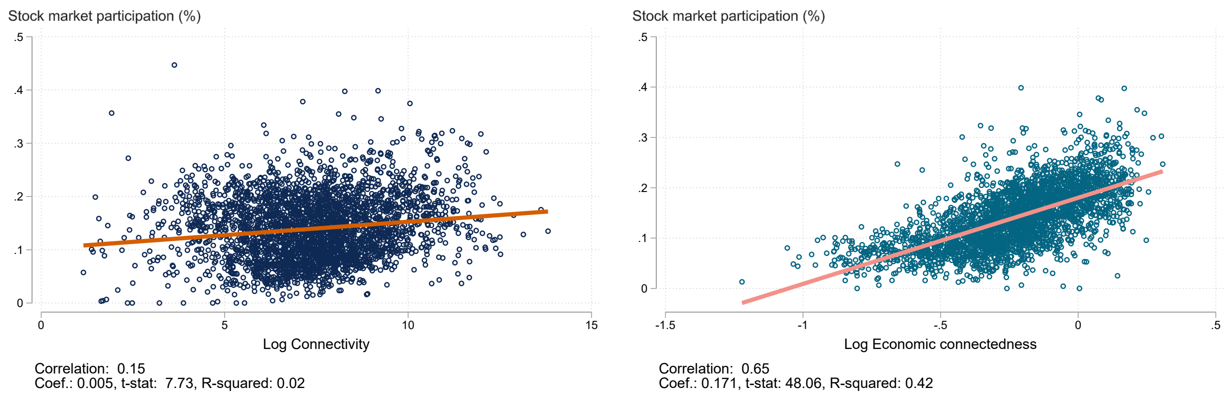
2. MOTIVATING EVIDENCE

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

We combine several county-level datasets to examine the correlation between connectivity and stock market participation. Below we describe the main sources and variables of interest.

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

[6] Many monetary costs for investing have likely fallen over time, with the increased availability of online financial education, low-cost trading platforms, and index funds.



(a) Log Connectivity index

(b) Log Economic Connectivity

Figure 2. Stock market participation and different measures of Connectivity

Notes: Both figures plot stock market participation on the county level on the y-axis. Panel a) plots Log Connectivity on the x-axis, where Connectivity is proxied by the within-county Social Connectedness Index from Facebook. Panel b) plots Log *Economic Connectedness* on the x-axis. We remove counties in the 99th percentile of connectivity. We report results regressions of the form $SMP = \alpha + \beta X + \epsilon_c$, where X is either Log SCI or Log Economics connectedness, and where we use robust standard errors.

We first collect county-level data on connectivity, the Social Connectedness Index (SCI) from Facebook Bailey *et al.* (2018). The SCI measures the social connectedness between and within US county pairs. This index measures the relative probability of a Facebook friendship link between Facebook users in two different or within one county. 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 socioeconomic status (SES) friends among low-SES individuals, averaged over all low-SES individuals in the county. Since high-income agents are more likely to invest in stocks and have information to share about the stock market, *economic connectedness* captures the idea that low-information agents need access to high-information agents 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 2.

We examine the relationship between stock market participation and social connectivity in Figure 2. The figure reports scatter plots between connectivity variables of interest and stock market participation. Panels a) plot the average SCI against stock market participation

Table 1. Social Connectivity and Stock market participation

	Economic connectedness		Connectivity index		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)
Log economic connectedness	0.171***	0.0511***			0.184***	0.0562***
	(0.011)	(0.0088)			(0.012)	(0.0091)
Log connectivity			0.00509**	0.000762	0.00950***	0.00236**
			(0.0020)	(0.00095)	(0.0011)	(0.00091)
Controls	No	Yes	No	Yes	No	Yes
State FE	No	Yes	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes	No	Yes
Mean Dep. Var.	0.14	0.14	0.14	0.14	0.14	0.14
Std. Dev. Dep. Var	0.06	0.06	0.06	0.06	0.06	0.06
Economic significance	0.66	0.20	0.02	0.00	0.70 / 0.04	0.22 / 0.01
Observations	2949	2949	2949	2949	2949	2949
R-squared	0.423	0.806	0.020	0.797	0.493	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 county's share of a bachelor-level education or above. Fixed effects for state and main industry of employment for the county are indicated. Economic significance is equal to $\beta s_x/s_y$, where *beta* is the coefficient of interest, s denotes the standard deviation of the independent variable x and dependent variable y. We cluster standard errors by state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

on the county level. The relationship between the logarithm of within-county connectivity from the SCI and stock market participation is positive and significant. However, SCI only explains 2 percent of the variation in stock market participation, and the relationship is not generally robust to adding control variables or to different transformations. Instead, panel b) shows that the correlation between economic connectedness and stock market participation is about four times higher than that between SCI and stock market participation. Moreover, economic connectedness explains 42 percent of the variation in stock market participation in a univariate regression.

We present regression results using the same data in Table 1. Column 1 presents a univariate regression without any control variables. Economic connectedness explains 42 percent of the variation in stock market participation across US counties. The effect is also economically significant: a one standard deviation increase in economic connectedness is associated with a 0.65 standard deviation increase in stock market participation. Following Mitton (2022), we define economic significance as $E = \beta s_x/s_y$, where *beta* is the coefficient of interest, s denotes the standard deviation of the independent variable x and dependent variable y. Economic connectedness is positive and significant even after we add control variables, state fixed effect and major industry of employment fixed effects in Column 2. The economic significance is lower but remains meaningful. The result for the SCI index in Column 3 shows a positive and significant effect. Still, the economic significance is relatively low: a one standard deviation increase in the SCI index is associated with a 0.02 standard deviation increase in stock market participation. The result is also not robust to including controls in Column 4. Finally, Columns 5-6 provide results where we include both economic connectedness and SCI. Both variables are now statistically significant and positive across specifications. However, in Column 6, economic connectedness is again highly economically significant ($E = 0.22$), whereas the SCI index has low economic significance ($E = 0.01$).

We interpret these results in the following way. Holding economic connectedness fixed, higher SCI positively impacts 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. We interpret these results as evidence that while connectivity generally seems to matter, for stock market participation, it is more important to connect to individuals with information to share.

3. THE MODEL

In this section, we propose a stylized model to study how the interaction between connectivity and other economic factors affects average stock market participation and participation across different income groups. The model setup has two main components. First, we formulate the utility maximization problem of an agent with a fixed stock market participation cost. This part follows the previous literature and is mostly inspired by the framework proposed in Vissing-Jørgensen (2002). However, unlike the previous literature, we endogenize the agent's stock participation cost assuming that it depends on the number of her peers who already invest in stocks and can share financial knowledge. Therefore, the second component of the model setup describes how agents 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 or stock index, and at the end of the period, they consume the proceeds from the investment portfolio in the form of a non-durable consumption good.

Risk-averse agents with identical CRRA preferences populate the economy. 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 $\mathscr{F}_w(\cdot)$, $W_{j,0} \sim \mathscr{F}_w(\cdot)$.[7]

The economy offers two investment opportunities. An agent can choose between investing her initial endowment in a risk-free asset with a 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 participation cost, F_i, at the beginning of the period.[8] The net return on the risky asset r is a random variable with a 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:

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

[7] The inequality parameter is implicitly captured by the particular functional form of the function $\mathscr{F}_w(\cdot)$. We don't need to make any specific assumption on function $\mathscr{F}_w(\cdot)$ for the general setting. However, in the simulation part, we will assume log-logistic wealth distribution.

[8] The cost F_i in the case of stock market investments includes the cost of time and 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 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 into 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 cost equals zero.[9] Non-Financially Educated agents do not have ex-ante knowledge about the stock market and face high participation costs. However, Non-Financially Educated agents can attain knowledge by learning from their peers who invest in risky assets. Thus, Non-Financially Educated agents face costs of participating that decrease with the number of peers in their social network that invest in the risky asset. We assume that the participation cost paid by a Non-Financially Educated agent i, $F(t_i = 0)$, is equal to a 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, $C'_\theta(\theta, k_i) > 0$. Consequently, an agent who has more informed peers faces lower participation cost, $C'_{k_i}(\theta, k_i) < 0$.

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

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

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 social 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. We discuss how matrix G is generated for given levels of connectivity and homophily in Section 3.5. $W = \{W_{1,0}, ..., W_{n,0}\}$ 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 discussing the information diffusion process, we find it technically convenient to construct a new variable called 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 for her participation cost to become sufficiently low to enter the stock market. The threshold $\hat{k}_i$ depends on the agent's 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 than agents with low income before investing in the risky asset. All Financially-Educated agents have a zero participation threshold because they already 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 she 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 $W_{i,0} > 0$.

Before we move to the technical details, let us briefly discuss the economic intuition. 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

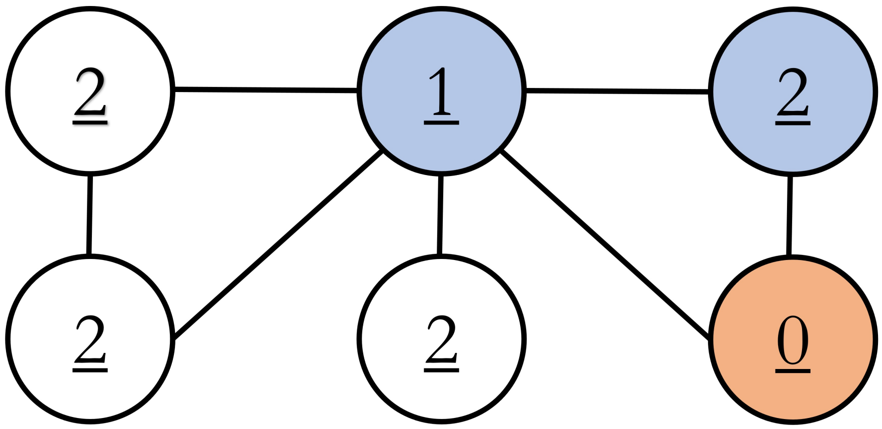


Figure 3. Possible equilibria

Notes: The figure plots possible equilibrium participation rates across different social networks. The number within each circle corresponds to the number of informed peers the individual needs to participate in the stock market. Financially Educated agents require zero informed peers to participate and are marked with an orange circle. Blue circles denote agents who participate.

be true for any possible equilibrium. These new participants spread information further to their peers. Some of those who get information have sufficiently large incomes 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.

We illustrate the idea with an example in Figure 3. In the figure, we assign a number $\hat{k}_i$ to each agent. Colored circles correspond to agents who invest in the risky asset. We initially have one agent participating and investing in the risky asset (right lower corner), the Financially Educated agent marked with an orange circle. This agent is connected to two other agents, indicated by lines between nodes. The agent in the top-middle row requires only one informed peer to participate and starts participating. The threshold for the agent in the top-right corner is two informed peers, meaning that this agent also participates. The agents on the left side have one informed peer but require two to participate. As a result, the rest of the agents in the economy lack enough connections and do not invest in the risky asset. The resulting equilibrium level of the risky asset investments is 3 out of 6, or about 50 percent. Notice the middle-bottom agent will never start investing within this network structure because she does not have sufficient connections: she requires at least two Participants among her connections to start investing but is only connected to one agent.

3.4 THE EQUILIBRIUM

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.[10] The economy has a matrix of linked agents G and a stack of participation thresholds $K = \{\hat{k}_1, ..., \hat{k}_n\}$ for each agent. 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.

It follows directly from the proof that the set of participating agents that the algorithm finds is a subset of a set of participating agents in any other equilibrium where agents can cooperate to extract information about the stock market.

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 estimate the stock market participation

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

cost (see, e.g. Vissing-Jørgensen, 2002; Andersen & Nielsen, 2010; Khorunzhina, 2013). We mainly focus on comparative statics to study how simultaneous changes in model parameters affect stock market participation. Our values for the fixed model parameters come from financial and macro data for the United States for 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.[11] 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 (Hung, 2021).

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 comes from the Quarterly Census of Employment and Wages. We calculate the number of individuals employed in the financial and insurance sectors (52 NAICS) in 2015. 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 the population. We assume that each agent's expected number of peers is the same and independent of the agent's other characteristics. We split all agents into five income groups based on income quantiles. Each agent forms a link with a peer who belongs to her income group with unconditional probability p_{In}, and an agent who does not belong to the same income group with unconditional probability p_{Out}. The homophily parameter, $h \in [0, 1]$, controls the difference between unconditional probabilities to form the link with other agents within and outside the agent's income group. If the homophily parameter h equals 0, then each agent is equally likely to form a link with any other agent independently of their income. If the homophily parameter h equals 1, then

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

each agent forms a connection only inside their income group. It is important to notice that the homophily and connectivity parameters are independent.

All agents in the economy belong to a social network described with a connectivity matrix. The general procedure for constructing a connectivity matrix is as follows. First, we assign a number of peers to each agent following a binomial distribution with an expected mean of c. Second, at each iteration, we consider an agent with the number of formed links below the assigned number of peers, and considering the number and outside/inside income group nature of the formed links, we compute conditional probabilities to form additional links to other peers. We update the conditional probabilities for each agent at each iteration such that the unconditional probabilities to form links inside or outside the agent's income group remain the same for all agents. Previous research suggests that individuals, on average, have a maximum of 50 active connections (Arrondel *et al.*, 2022; Mac Carrona *et al.*, 2016). However, only a small number of these links are used to share financial information. Arrondel *et al.* (2022) find, on average, individuals have seven peers in their financial circles. We therefore use seven peers as the baseline point in the simulation analysis.

Stock market participation cost − In the model, we assume that stock market participation cost depends on the number of informed peers. For simulation analysis, we assume the linear functional form of $C(k_i)$ function for Non-Financially Educated agents:[12]

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

The parameter θ represents the stock market participation cost of a Non-Financially Educated agent who is not connected to informed peers. The parameter $\Delta\theta$ measures how much stock market participation cost will decrease when the number of informed links increases. Given the data on participation cost across the entire population and participation across different income groups, we calibrate the model to estimate the magnitude of cost function parameters θ and $\Delta\theta$. We get an estimation of the cost θ of around \$2,000 and an estimated $\Delta\theta$ between \$100 to \$200. The previous literature present estimates ranging from \$260 in Vissing-Jørgensen (2002) to \$134,000 in Andersen & Nielsen (2010). Note that estimates of

[12]As a robustness check, we run simulations assuming $C = \theta/k_i^{\alpha}$, where $\alpha > 0$ is an exogenous parameter. The simulation results hold.

θ in the previous literature are net of information from peers. The estimates of \$260 from Vissing-Jørgensen (2002) can be justified by taking our estimate of \$2,000 and assuming that individuals received information from 8.5 peers, for example. We therefore view our estimates as reasonable. However, we will allow parameters to vary in certain ranges in the simulations. See Table 3 for details.

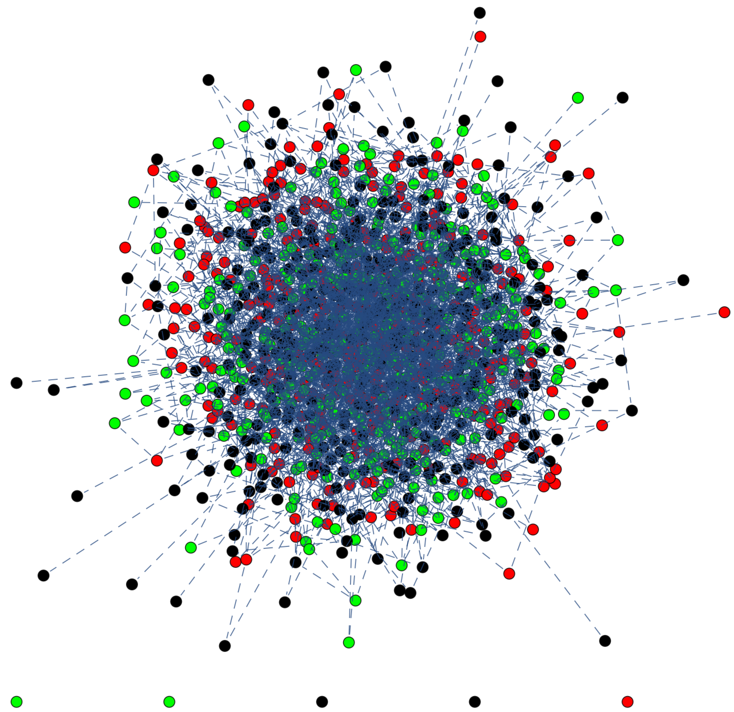
4. MODEL RESULTS

We now present the main results for the model simulations. We run simulations of our model with market parameters described in Table 3. Overall, we consider 1485 combinations of parameters. Computational capacity allows us to run simulations in networks with 10,000 agents. Given the average number of connections, we focus on a sparse network graph where agents exist in clusters. In the model, the composition of the clusters depends on homophily parameter since this parameter governs how likely agents are to be connected to agents from other income groups. Figure 4 illustrates the network structure with low homophily in panel a) and high homophily in panel b). Agents with different income levels are represented with different colors. In the economy with low homophily, there are no clear clusters of income groups. In contrast, the economy with high homophily shows clusters among different income groups.

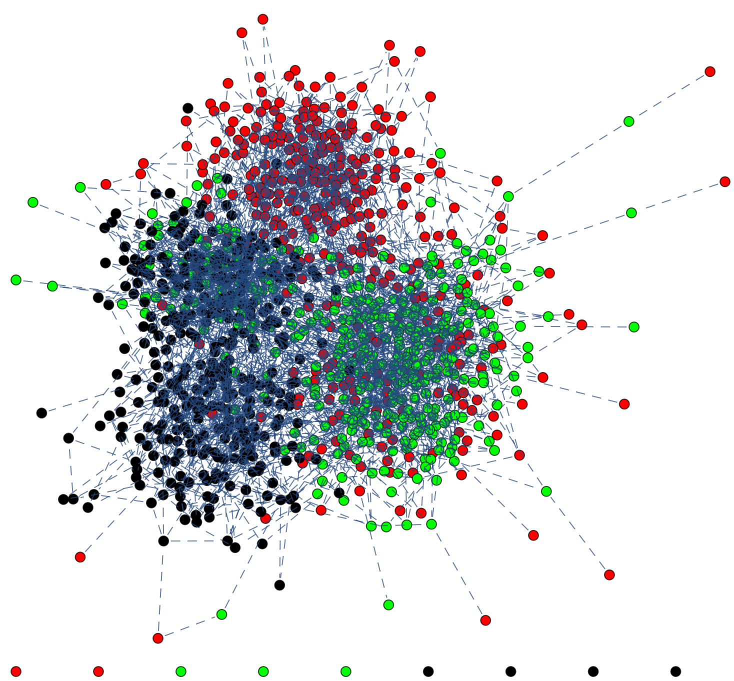
We focus on how the number of connections, homophily, and ex-ante income inequality affect the share of risky asset investors, starting with the full population. We later explore how each parameter of interest affects agents who belong to different income groups. For this purpose, we split agents into three equal-sized groups based on their income: low, middle, and high. The groups do not coincide with 5 income income groups used to construct the network. Each group's average and maximum income is not the same across different simulations where the inequality parameter varies.

4.1 THE EFFECT OF CONNECTIVITY

Figure 5 presents the relation between connectivity and stock market participation generated by the model simulations. The solid orange line plots the average participation, and black, gray, and blue dashed lines present participation among low, medium, and high income groups respectively. In the figure, we vary the parameter marked on the x-axis but fix all



(a) Low homophily



(b) High homophily

Figure 4. Network structure visualisation

Notes: The figures plot networks generated by our algorithm. For illustration purposes, we constructed networks with 1,000 agents. Red vertices represent agents with low income, green vertices represent agents with medium income, and black vertices correspond to agents with high income. We set the homophily index of 0.1 for low homophily and 0.9 for high homophily.

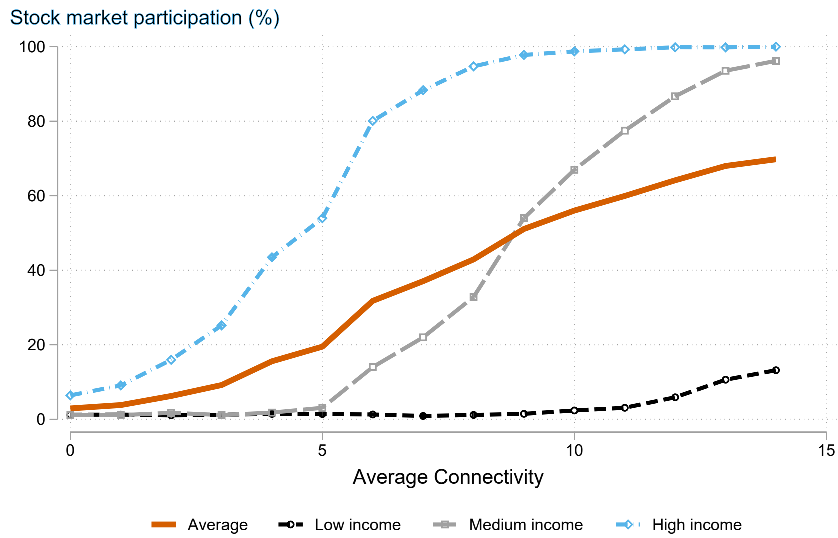


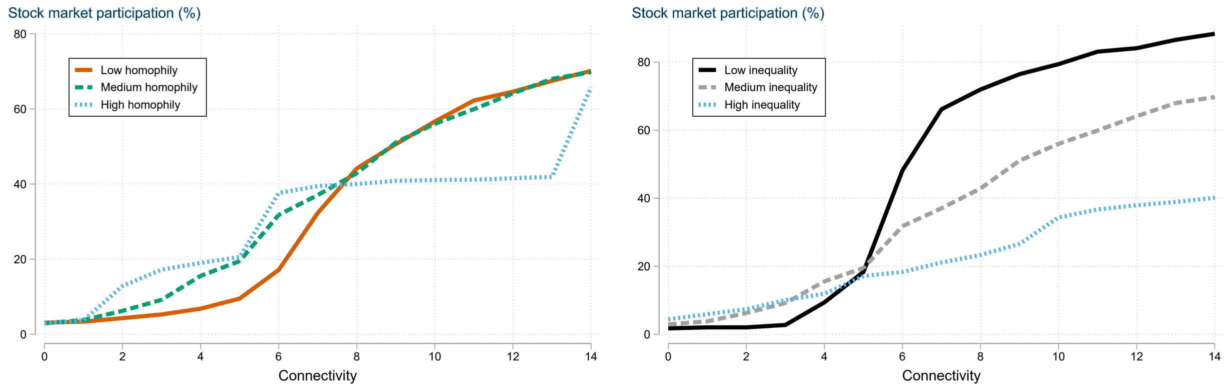
Figure 5. The effect of connectivity by model parameters

Notes: The figure plots stock market participation against connectivity, defined as the average number of connections of each agent. The orange solid line plots the average stock market participation among all agents. The black, gray, and blue lines plot average participation among high income agents, medium income agents and low income agents, respectively. We set the homophily parameter to 0.5 and the ex-ante GINI coefficient to 0.4 for the simulations.

other parameters. Panel a) of Figure 5 shows that stock market participation is increasing in connectivity. Without connections, the only agents who participate are Financially-Educated agents and agents who are wealthy enough to pay the fixed participation cost. If more agents are connected, information spreads more efficiently throughout the economy, and the share of risky asset investors grows.

The relation is S-shaped. The effect of adding one more peer is small for a few connections. However, as the number of connections increases, the marginal effect of adding one more peer grows. Intuitively, starting with a higher number of connections, a new peer will create more links with a higher probability, improving information transmission. However, as the number of connections increases, many agents have already started investing in the risky asset, and the effect of additional connectivity is decreasing.

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 dashed line, the baseline participation rate is higher, as it is more likely that they have enough income to cover the fixed



(a) Homophily

(b) Inequality

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

Notes: The figure plots stock market participation against connectivity, defined as the average number of connections of each agent. Panel a) plots the effect of connectivity on stock market participation for different levels of homophily. We use values for the homophily parameter of 0.1 for the solid orange line, 0.5 for the dashed green line and 0.9 for the blue dotted line. The GINI coefficient in panel a) is set to 0.4. Panel b) plots the effect of connectivity on stock market participation for different levels of ex-ante inequality. We use values for the GINI coefficient is 0.25 for the solid black line, 0.4 for the dashed gray line and 0.6 for the blue dotted line. The homophily parameter in panel b) is set to 0.5.

participation cost regardless of the number of peers. Moreover, since high-income agents are more likely to be close to the participation threshold, connectivity positively impacts their participation. For high-income agents, more connections help spread information more efficiently. Above level five, the effect diminishes, again giving an S-shaped pattern. With more than ten connections, all high-income agents 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. Medium-income agents are further from the participation threshold and, thus, do not initially benefit as much from increased connectivity. However, once connectivity reaches a sufficient level, stock market participation strongly increases. On the right side of the graph, the gap in participation between medium and high-income agents is small. Finally, the effect of connectivity for low-income agents is small. Low-income agents are very far from the participation threshold and thus need many peers before participating.

The levels of homophily and inequality also affect the relationship between connectivity and stock market participation. Recall that the homophily parameter measures how likely two agents from different income groups are to connect. Panel a) of Figure 6 plots the connectivity and stock market participation for different levels of homophily. Homophily affects the S-shaped relation between connectivity and stock market participation. For low levels of connectivity, high homophily leads to more efficient information transmission from informed to uninformed agents, generating higher participation rates. We see this by examining differences in participation for the three lines for levels of connectivity below 8: the blue line with a high level of homophily is consistent above the other lines.

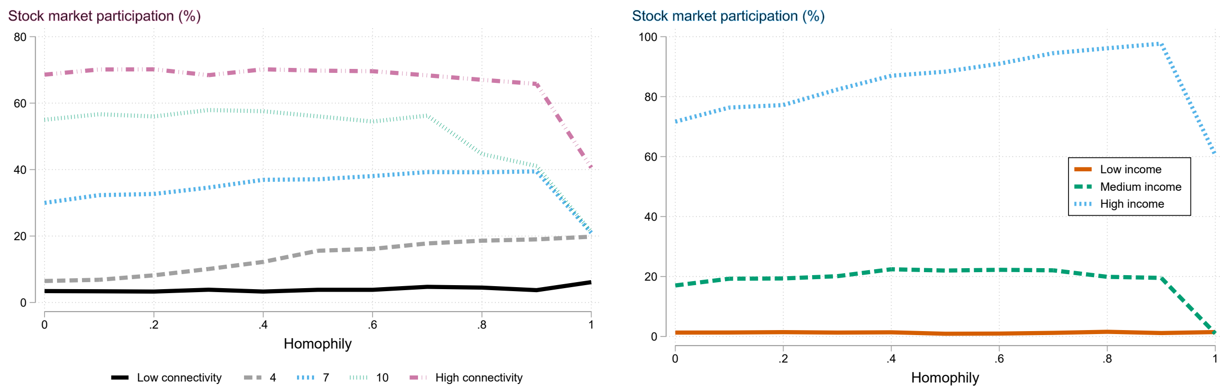
The effect flips if connectivity is high, however. If connectivity is above 8 in the figure, higher homophily results in lower stock market participation. High homophily makes it more likely for rich agents with few connections to form a link with another rich agent. Given that rich agents are more likely to be informed, high homophily promotes stock market participation among them. However, at a high level of connectivity, almost all rich agents already participate in the stock market. Thus, when connectivity is high, there is no room for homophily to affect 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.[13] In the simulations, we keep the average income level constant but change the distribution.[14] As the GINI coefficient increases, more wealth is concentrated among high-income households. At high levels of inequality, the effect of connectivity is muted because 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 limited by a high share 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 participation and connectivity is most pronounced for low levels of inequality. In this economy, agents have close to equal shares of the same pie, leading to many agents being far from the participation threshold. As connectivity increases, however, more agents can benefit from access to information, and participation increases rapidly.

[13] We choose GIN values of 0.25, 0.4, and 0.6 for low, medium, and high inequality respectively.

[14] We assume that income is distributed according to Log-Logistic distribution $\mathscr{F}_w\left(x; \log[\alpha], 1/\beta\right)$, where α is a scale and β is a shape parameter. $Mean\ Income = \frac{\alpha\pi/\beta}{\sin[\pi/\beta]}$, $\beta = 1/Gini$.



(a) Population-level

(b) By income groups

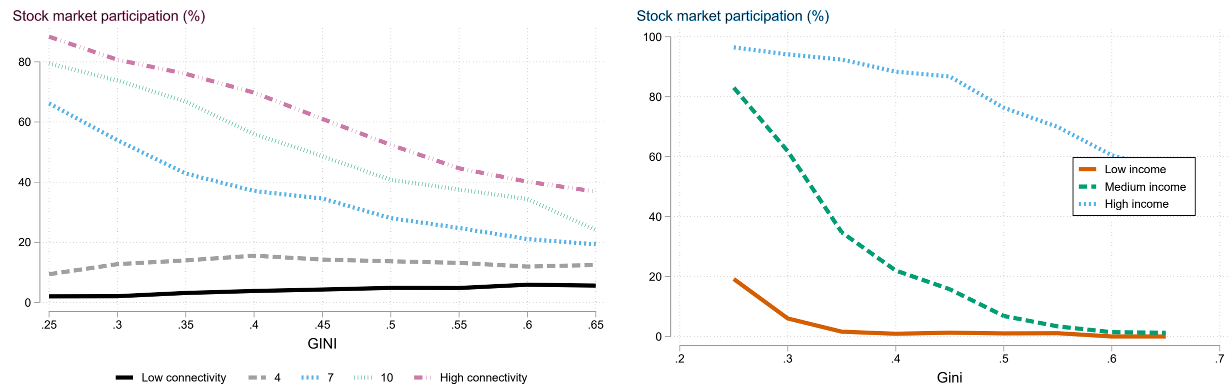
Figure 7. The effect of homophily on stock market participation

Notes: The figure plots stock market participation against homophily. Panel a) plots the effect of homophily on stock market participation for different levels of connectivity. Average connectivity ranges from 1 (Low connectivity, black solid line) to 14 (High Connectivity, pink dashed line). Panel b) plots the effect of homophily on stock market participation for different income groups. The low income group is marked with an orange solid line, the medium income group is marked with a dashed green line, and the high income group is marked with a dotted blue line. Connectivity in panel b) is set to 7. The GINI coefficient in both panels is set to 0.4.

4.2 THE EFFECT OF HOMOPHILY

Figure 7 shows that homophily positively impacts 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 the stock market is more likely to spread to agents far from the participation threshold, who, consequently, 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 homophily of 1, however, agents are only connected to agents within their income group, leading to less efficient information sharing. Information will still spread throughout the network, but the effect is limited to specific income groups. Since only high-income groups have enough income to start participating, information sharing is limited to this group, and the participation rate 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 partici-



(a) Population-level

(b) By income groups

Figure 8. The effect of inequality on stock market participation

Notes: The figure plots stock market participation against ex-ante inequality. Panel a) plots the effect of inequality on stock market participation for different levels of connectivity. Average connectivity ranges from 1 (Low connectivity, black solid line) to 14 (High Connectivity, pink dashed line). Panel b) plots the effect of ex-ante inequality on stock market participation for different income groups. The low income group is marked with an orange solid line, the medium income group is marked with a dashed green line, and the high income group is marked with a dotted blue line. Connectivity in panel b) is set to 7. The homophily coefficient in both panels is set to 0.5.

pation against homophily for three income groups. We set the level of connectivity to 7, the reference point from Arrondel *et al.* (2022). Naturally, agents for the high-income group participate at a higher level than the low and medium-income groups. As homophily increases, it is the high-income agents who increase their stock market participation. Higher homophily increases the likelihood that they connect with an informed agent from their income group. For low and medium-income groups, homophily has little impact on participation rates. The positive relationship between homophily and stock market participation in panel a) is thus driven by higher risky asset investment among high-income agents.

4.3 THE EFFECT OF INEQUALITY

We now 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 the average income constant, inequality adjusts the share of agents who can pay the fixed participation costs. Second, inequality also affects the probability that informed agents 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 provides the results for different levels of connectivity. For low levels of connectivity, black and gray lines at the bottom of the figure, fewer agents have the income necessary to cover participation costs at low levels of inequality, and there is little information sharing. As inequality increases, we take money from the poor and give it to the rich, allowing more agents to participate and spread information. As a result, stock market participation increases. In societies with low connectivity, inequality has a positive effect on participation.

We see a negative relationship between inequality and participation in simulations with high connectivity. This model features homophily and connectivity and therefore depends on the difference in income. Higher inequality leads to clustering in the network and less efficient information diffusion. Consequently, stock market participation is lower. For instance, we see that stock market participation is above 80 percent for simulations with high connectivity and low inequality. The share drops 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 more agents meet the fixed cost of investing, and stock market participation increases. Importantly, however, the higher income also creates more informed agents, which in turn helps spread information and increases stock market participation. Therefore, the effect of higher income works not only through direct stock market participation but also through more efficient information sharing through social connections. The share of Financially Educated agents increases stock market participation since Financially-Educated agents participate and help seed the network with information. Finally, higher participation cost leads to lower participation.

5. CONCLUSION

Many of us have moved in the course of our careers as academics. This is certainly true for all authors of this paper. Our career moves have led us to expand our social networks across several countries, with technology providing easy means to communicate and maintain contact with our friends. At the same time, we have often discussed how our many years of studying economics and finance do not always help deal with many practical aspects involved in making good financial decisions. Fortunately, being academics, we have moved to contexts where our new colleagues could help us with anything from how the pension system worked to how to invest in stocks, information that is invaluable when trying to make good financial decisions. To put these experiences in the context of this paper, our social networks have expanded over time, and since we move to other academic jobs, we have been able to benefit from learning from others with similar experiences. However, these networks are highly particular to our work, and we can only imagine that others did not have similar expertise within their social networks. The main idea in this paper is that increasing tendencies to associate only with others similar to us will leave some people without access to good sources of information within their networks, with detrimental effects on their financial situation, their wealth accumulation, and in the end, for society.

In this paper, we present a simple theoretical model of stock market participation to argue that the effect of increased connectivity depends heavily on the network structure. We provide evidence that economic connectedness strongly correlates with stock market participation in the cross-section of US counties, but social connectivity has little predictive power. Informed by this evidence, we show that connectivity leads to increased stock market participation but that the effect depends on homophily and ex-ante inequality. We show that higher-income agents are more likely to benefit from higher connectivity. The model suggests a new, previously unexplored avenue for future research: what is the distribution of financially informed peers in society, and how has this changed over the last twenty years? Can increased homophily explain why participation has not increased in twenty years?

REFERENCES

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

Arrondel, Luc, Calvo-Pardo, Hector, Giannitsarou, Chryssi, & Haliassos, Michael. 2022. Informative social interactions. *Journal of Economic Behavior & Organization*, **203**, 246–263.

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

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

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

Bäckman, Claes, & Hanspal, Tobin. 2022. Participation and losses in multi-level marketing: Evidence from a Federal Trade Commission settlement. *Financial Planning Review*, **5**(1).

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

Balakina, Olga. 2022. Peer Effects in Stock Trading: The Effect of Co-Workers, Family and Neighbors.

Balakina, Olga, Bäckman, Claes, Hackethal, Andreas, Hanspal, Tobin, & Lammer, Dominique M. 2023. Personal Recommendations and Portfolio Quality.

Bonaparte, Yosef, Korniotis, George M, & Kumar, Alok. 2018. Portfolio choice and asset pricing with investor entry and exit. *SSRN Electronic Journal*.

Brandsaas, Eirik Eylands. 2021. *Household stock market participation and exit: The role of homeownership*. Tech. rept. Working paper.

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

Briggs, Joseph, Cesarini, David, Lindqvist, Erik, & Östling, Robert. 2021. Windfall gains and stock market participation. *Journal of Financial Economics*, **139**(1), 57–83.

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

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

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

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

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

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

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

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

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

Fagereng, Andreas, Guiso, Luigi, & Pistaferri, Luigi. 2022. *Assortative mating and wealth inequality*. Tech. rept. National Bureau of Economic Research.

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

Haliassos, Michael, & Bertaut, Carol C. 1995. Why do so few hold stocks? *The Economic Journal*, 1110–1129.

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

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

Hung, Chih-Ching. 2021. Does Social Connectedness Affect Stock Market Participation? *Available at SSRN 3724407*.

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

Jackson, Matthew O. 2014. Networks in the understanding of economic behaviors. *Journal of Economic Perspectives*, **28**(4), 3–22.

Jackson, Matthew O. 2021. Inequality's Economic and Social Roots: The Role of Social Networks and Homophily. *Available at SSRN 3795626*.

Jackson, Matthew O, Nei, Stephen M, Snowberg, Erik, & Yariv, Leeat. 2023. *The Dynamics of Networks and Homophily*. Tech. rept. National Bureau of Economic Research.

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

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

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

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

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

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

Mitton, Todd. 2022. Economic Significance in Corporate Finance. *The Review of Corporate Finance Studies*, 02. cfac008.

Morelli, Sylvia A, Ong, Desmond C, Makati, Rucha, Jackson, Matthew O, & Zaki, Jamil. 2017. Empathy and well-being correlate with centrality in different social networks. *Proceedings of the National Academy of Sciences*, **114**(37), 9843–9847.

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

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

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

Verbrugge, Lois M. 1977. The structure of adult friendship choices. *Social forces*, **56**(2), 576–597.

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

6. TABLES

Table 2. 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
Log economic connectedness	-0.233	-0.216	0.227	-1.222	0.307
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
Log connectivity	7.415	7.435	1.805	1.151	13.816
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 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[15], where the average number of links is normalized to the largest number of connections for a Los Angeles County - Los Angeles County pair.[16]

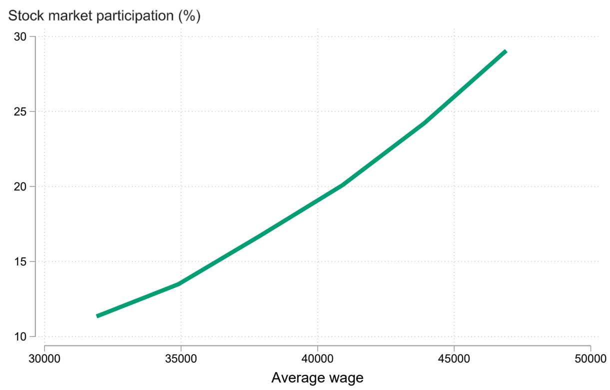
We obtain stock market participation rates from the Internal Revenue Service's (IRS) Statement of Income (SOI) for individual income tax return (Form 1040) statistics following the procedure described in Bäckman & Hanspal (2022), where the fraction of tax returns claiming ordinary dividends are used 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).[17] 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.

[15] 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.

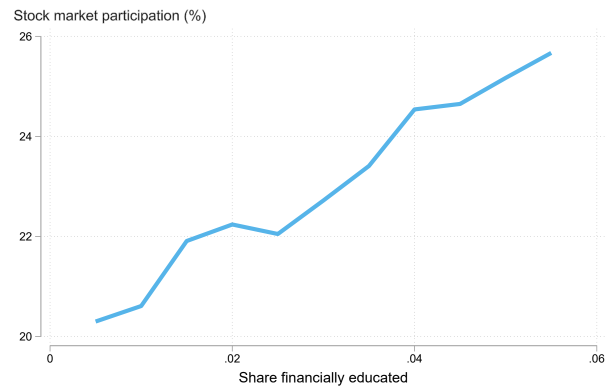
[16] The SCI for Los Angeles County - Los Angeles county is equal to 1,000,000

[17] 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 **??**

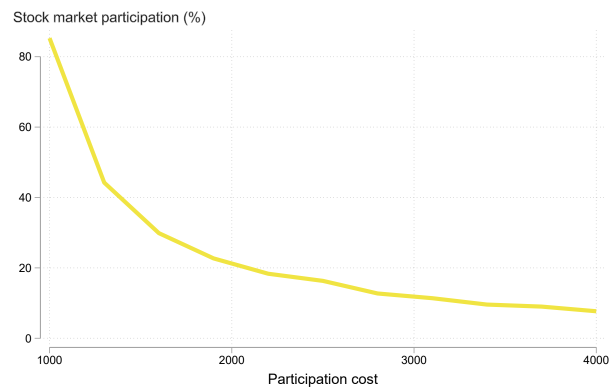
B. APPENDIX: FIGURES



(a) Population-level



(b) By income groups



(c) By income groups

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

Notes: The figures plots stock market participation against various levels of a) average income b) share financially educated and c) participation costs. Panel a) plots participation against the average income in Panel a) plots the effect of inequality on stock market participation for different levels of connectivity. Average connectivity ranges from 1 (Low connectivity, black solid line) to 14 (High Connectivity, pink dashed line). Panel b) plots the effect of ex-ante inequality on stock market participation for different income groups. The low income group is marked with an orange solid line, the medium income group is marked with a dashed green line, and the high income group is marked with a dotted blue line. Connectivity in panel b) is set to 7. The homophily coefficient in both panels is set to 0.5.

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[18] 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.

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