Beyond Connectivity: Stock Market Participation in a Network*

Anastasiia Parakhoniak Olga Balakina Claes Bäckman

April, 2025

Abstract

What are the aggregate and distributional consequences of the relationship between an individual's social network and financial decisions? Motivated by several well-documented facts about the influence of social connections on financial decisions, we build and calibrate a model of stock market participation with a social network that emphasizes the interplay between connectivity and network structure. Since connections to informed agents influence peers through utility and learning, there is a pivotal role for homophily. An increase in the average number of connections raises the average participation rate, mostly due to richer agents. Higher homophily benefits richer agents by creating clusters where information spreads more efficiently. We also show that social utility is crucial for matching stock market participation among poorer agents. Finally, we provide empirical evidence consistent with the importance of connectivity and sorting.

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

*Balakina: Leibniz Institute for Financial Research SAFE. Email: balakina@safe-frankfurt.de. Bäckman: Mannheim University, Department of Economics and the Leibniz Institute for Financial Research SAFE. Email: claes.backman@gmail.com. Parakhoniak: Durham University Business School, Department of Economics. Email: anastasiia.parakhoniak@durham.ac.uk. UPDATE We thank Peter Andre, Michael Haliassos, Kim Peijnenburg, Stephan Siegel, Anna Pavlova, Spyros Galanis and Tobin Hanspal for helpful comments. We gratefully acknowledge research support from the Leibniz Institute for Financial Research SAFE.

1. INTRODUCTION

This paper explores the aggregate and distributional consequences of several intuitive and well-documented facts about stock market participation and social networks: peers influence stock market participation (see e.g. Hong *et al.* , 2004; Kaustia & Knüpfer, 2012), there is selection and sorting in social networks (Jackson, 2021), individuals can both learn from their peers and derive utility directly from following the same strategy (Bursztyn *et al.* , 2014), and participation costs exist and decrease in know-how (Vissing-Jørgensen, 2002). We include these features in a theoretical model to study the aggregate and distributional impact of peer effects and network structure.

Specifically, we ask three questions: First, how does the average number of connections affect equilibrium stock market participation? We refer to the average number of connections in the model as *connectivity*. Second, how does network structure affect stock market participation? With network structure, we specifically mean the tendency of individuals to sort based on similarities, often called *homophily*. 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 who are like them. Third , does social learning and social utility have different implications for stock market participation? For all questions, we consider the answer at the aggregate level and for different income groups. These questions are crucial for understanding how social networks affect stock market participation and help us formulate several intriguing avenues for future research.

With these goals in mind, we build and calibrate a model of stock market participation where all agents can share information in a network. Agents in the model have to pay a fixed cost to enter the stock market, a common approach to modeling the decision to invest in stocks. Fixed costs capture monetary, behavioral, and non-pecuniary costs that make stock ownership uncomfortable for some households (Campbell, 2006).[1] The agent must decide whether to enter the stock market and, conditional on entry, determine the optimal portfolio allocation.

Both the entry decision and the optimal portfolio allocation depend on the number of peers who invest in risky assets. Motivated by the previous literature (e.g. Bursztyn *et al.* , 2014), we assume that agents can both learn from their peers (social learning) and that they derive utility from making the same choice as their peer (social utility). Social learning in our model affects the precision of the expected return. We model social utility as a reduction in the fixed participation cost, making the benefit of having more peers independent of the return. In the baseline model, we include both these channels, but later we examine each separately.

All agents in the economy are connected with an ex-ante connectivity parameter that determines each agent's expected number of links. The likelihood of connecting with another agent depends on a homophily parameter, defined as 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. 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] While the model is parsimonious, the setting is rich enough that we require simulations to answer the questions we are interested in.[3]

The model endogenously generates an S-shape relationship between connectivity and stock market participation. At low levels of connectivity, information sharing is limited, and

[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).

[2] Although we have chosen to focus on income homophily, we later show results when homophily is unrelated to factors that determine stock market participation. 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] We ensure that the simulations are robust by including a large number of agents and running results for several different assumptions over various model parameters.

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. When the average number of links per agent becomes large, a giant component of connected and informed agents form in the network. At this point, the information propagation becomes more efficient and the potential of individual agents to block information flow significantly diminishes. Consequently, adding more connections has little impact on stock market participation. We also show that a marginal increased connectivity mainly benefits richer agents, who are closer to participating and thus need fewer informed connections.

Homophily in social networks alters the relationship between connectivity and stock market participation in nuanced ways. Suppose connectivity is low and homophily increases. Agents then become more likely to be connected to others with similar incomes and the model endogenously 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. In sparse networks, higher homophily has a small but positive impact on average stock market participation because of more efficient information transmission among rich agents.[4] The increase in participation is concentrated among wealthier agents, leading to an increase in 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. We also consider the effect of homophily attributed only to high-income agents, *gated community* effect, and find that it negatively affect stock market participation level in the economy overall.

Higher ex-ante inequality also affects the relationship between connectivity and stock market participation. Assuming that agents exhibit homophily in income, income inequality affects the probability that agents with information are connected to other agents. Higher

[4]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.

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, decreasing the likelihood of being connected to informed peers. Furthermore, while keeping average income the same, higher inequality shrinks the share of agents who can pay the fixed participation costs and decreases stock market participation.

In the final part of the paper, we study the implications of learning and social utility for stock market participation. Recall that social learning affects expected returns, whereas social utility affects the fixed participation costs. While both models can match the average participation rates, only the model with social utility allows for a more balanced distribution of stock market participation (SMP) across different income groups that is aligned with the data. The model with only social learning channel generates either too high participation among rich agents or too low participation among low-income agents. We show that the social utility channel is particularly important for encouraging participation among low-income agents, as it helps overcome the high participation costs. In contrast, learning more stock market returns has a relatively smaller impact for these agents, as they typically have limited resources to invest. For high-income agents, however, the social learning channel becomes increasingly more significant in the decision to participate in the stock market.

Empirically, 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). Economic connectedness explains 43 percent of the variation in stock market participation across US counties, whereas the SCI only explains 2 percent. These results are robust to including controls (e.g., income, age, education) and state-fixed effects and is economically significant.

This pattern is an aggregate consequence that our model can explain. Viewed through the lens of our model, economic connectedness measures connections between individuals who can benefit from information about the stock market.

We believe important questions immediately follow from the ideas in this paper. For instance, quantifying how important rising polarization or increases in homophily are for explaining flat stock market participation rates since 2000 is an exciting possibility for future research in the United States or other countries. Another intriguing question is whether homophily in social networks has changed in a manner relevant to stock market participation. Our results highlight that homophily matters for participation decisions only if it is related to factors determining stock market participation, such as income. The results also provide a theoretical framework for linking rising polarization to wealth inequality, given the importance of differences in returns for wealth inequality (Bach *et al.* , 2020; Fagereng *et al.* , 2020). Finally, if informed peers are an important source of information, as the literature would suggest, then it is important to ask who has access to such information. To what extent can we explain the lack of stock market participation among poorer households simply because they have no one to ask about how to invest? These questions provide a new way forward for a literature that has established the importance of informed peers for stock market participation but has not yet studied the aggregate and distributional impact of the individual-level results.

Consistent with the rest of the literature, our model highlights that access to peers is an important determinant of stock market participation. In a more unequal society or in a society with a higher degree of homophily, an important source of information will be unable to low-income households. While it seems difficult to reverse trends in homophily through policy interventions, policymakers should be aware that increased sorting by income or wealth can lead to negative outcomes in the lower parts of the income distribution. To the extent that financial education is effective, policymakers should be aware that financial education may have to replace peer effects, and should aim to target these interventions in highly-connected but uninformed parts of the distribution to maximize their impact.

Related literature. Our paper intersects with several strands of the large literature on household financial decision-making and peer effects. 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, 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 (Brown *et al.* , 2008; Kaustia & Knüpfer, 2012; Bursztyn *et al.* , 2014; Changwony *et al.* , 2014; Hvide & Östberg, 2015; Knüpfer *et al.* , 2017; Arrondel *et al.* , 2022; Patacchini & Rainone, 2017; Haliassos *et al.* , 2020; Ouimet & Tate, 2020; Balakina, 2022; Balakina *et al.* , 2024). Recent papers have also studied how social interactions affect portfolio composition and financial mistakes (Ammann & Schaub, 2021; Heimer & Simon, 2015; Hvide & Östberg, 2015; Heimer, 2016; Lim *et al.* , 2020; Han *et al.* , 2022), suggesting that peer effects do not always lead to better outcomes.[5] We argue that a potentially important yet overlooked aspect of peer effect in finance is to examine the distribution and clustering of informed agents.[6] Most of the empirical literature on peer effects in stock market participation focuses on the challenging question of documenting that peer effect exists but spends little time investigating who has access to informed peers.

[5]Our model is primarily about participation decisions, and so is not ideally suited to studying peer effects and financial mistakes.

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

Third, the limited stock market participation puzzle has been a major subject in finance dating back to Arrow (1965).[7] 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 non-pecuniary costs to participation (Andersen & Nielsen, 2010; Briggs *et al.* , 2021), and that peers can lower these costs (Duraj *et al.* , 2024). By allowing participation costs to vary with social connectivity, we generate heterogeneity in the costs unrelated to income or financial education. This assumption can help explain limited stock market participation rates in the cross-section.

Fourth, our paper relates to the literature on the linear threshold models (LTM), which are widely used to study the spread of behaviors, ideas, and adoption decisions across social networks (Morris, 2000; Watts, 2002; Granovetter, 1978; Vega-Redondo, 2007). Differently from many papers which study the linear threshold models in the context of optimal seeding (Kempe *et al.* , 2003; Chen *et al.* , 2010), we emphasize the importance of network structure, particularly the roles of homophily and inequality, in shaping information propagation in the network with limited participation (Lelarge, 2012). Our findings are consistent with Acemoglu *et al.* (2011), who demonstrate that networks with high clustering and short-range links may be less effective at spreading information compared to networks with lower clustering and more long-range connections.

Finally, a large literature in banking has documented the importance of network structure for financial stability and contagion.[8] Nonetheless, we view the ideas in this paper as important and novel in the context of *household* financial decision-making. The empirical

[7] 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).

[8] See e.g. Bernard *et al.* (2022), Allen & Gale (2000), Morris (2000), Elliott *et al.* (2014), and Amini *et al.* (2016).

literature in household finance has focused on the (hard) question of providing plausible evidence that peers affect financial decisions but has not yet tackled the aggregate or distributional implications. Our results suggest that *ceter paribus*, higher homophily leads to a larger gradient in stock market participation, which increases inequality. In addition, higher inequality leads to less stock market participation among low-income individuals because inequality affects network structure in the presence of homophily. Both these feedback loops are important to consider when discussing the causes and consequences of rising inequality.

The rest of the paper proceeds as follows. Section 2 provides the model, and Section 3 provides the results from the model simulations. Section 5 discusses the empirical evidence for the cross-section of US counties, Section 6 concludes.

2. THE MODEL

In this section, we propose a stylized model to examine how the interaction between connectivity, homophily, and other economic factors influences both average stock market participation and participation across different income groups. We cover two main model components in this section.

First, we introduce utility maximization problem in which an individual agent faces stock market participation costs, in accordance with the literature on stock market participation and information acquisition in financial markets, particularly the works of Vissing-Jørgensen (2002); Cocco *et al.* (2005); Peress (2004, 2005). Participation costs can be interpreted as expenses associated with opening an investment account, paying brokerage fees, and similar activities (Vissing-Jørgensen, 2002), or as psychological barriers to entering the stock market (Andersen & Nielsen, 2010; Briggs *et al.* , 2021; Duraj *et al.* , 2024). The agent must decide whether to enter the stock market and, conditional on entry, determine the optimal portfolio allocation. Both decisions depend on the number of peers who invest in risky assets and potentially share information about stock market returns.

Next, we extend the model analysis to the aggregate level. We describe how agents are embedded in a social network and how their social connections influence stock market participation.[9] In the baseline specification, we incorporate two primary mechanisms through which social influence can affect stock market participation. First, an agent's utility from owning a stock or stock index may depend on whether their peers also own the asset. We refer to this mechanism as *social utility* (Abel, 1990; Gali, 1994; Campbell & Cochrane, 1999; Taylor, 2011).[10] Second, an agent receives additional information about stock market performance from peers who are already investing (Peress, 2004, 2005), and agents who receive more signals benefit from greater accuracy in their posterior beliefs about stock market returns (Arrondel *et al.* , 2022). We refer to this mechanism as *social learning*. We introduce heterogeneity among agents through initial endowment and financial sophistication.

We conclude this section by discussing the diffusion of social influence in the model and by examining its equilibrium properties.

2.1 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, the agent is endowed with initial wealth, which we can interpret as discretionary income, and allocates it between a risk-free asset and a risky asset, such as a stock index. At the end of the period, the agent consumes the proceeds from her investment portfolio in the form of a non-durable consumption good.

We study the optimal portfolio choice of an agent with constant relative risk aversion (CRRA) utility function who decides how much to invest in a risky asset with log-normally

[9] We do not distinguish between entry and participation cost of the stock market due to the fact that our model has only one period.

[10] The social utility channel is closely related to imitation, where an agent observes and adopts their peers' investment behavior, inferring that it is beneficial.

distributed return factor (Merton, 1969; Samuelson, 1975). The utility function of agent i is:

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

where $W_{1,i}$ 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 $W_0 = \{W_{0,1}, ..., W_{0,j}, ...W_{0,n}\}$ distributed as $\mathscr{F}_w\left(\cdot\right)$, $W_{0,j} \sim \mathscr{F}_w\left(\cdot\right)$.[11]

The economy offers two investment opportunities. An agent can choose between investing her initial endowment in a risk-free asset with a net return of R_f or investing in a risky asset with a log-normally distributed return factor R_a, where $r_a = log(R_a)$ is normally distributed. If agent i decides to invest in the risky asset, she faces a participation cost, F_i, at the beginning of the period.[12]

Given that the terminal wealth $W_{1,i}$ is equal to proceeds from the investment portfolio, we can define $W_{1,i}$ as

$$W_{1,i} = \alpha_i(W_{0,i} - F_i)R_a + (1-\alpha_i)(W_{0,i} - F_i)R_f = (W_{0,i} - F_i)\underbrace{(\alpha_i(R_a - R_f) + R_f)}_{\equiv R_p}, \tag{2}$$

where α_i is the portfolio share allocated to the risky asset, and R_p is a portfolio return where the portfolio consists of risk-free and risky assets.

When deciding whether to invest in the risky asset, agent i considers both the magnitude of the fixed entry cost and her beliefs about the asset's performance. We assume that peer effects influence both factors.

[11]The inequality parameter is implicitly captured by the particular functional form of the function $\mathscr{F}_w\left(\cdot\right)$. We do not make any specific assumption on function $\mathscr{F}_w\left(\cdot\right)$ for the general setting of the model. However, we will assume a log-logistic wealth distribution in the simulation part.

[12]The costs include factual values and perceived expenses associated with stock market investment (Duraj *et al.* , 2024).

2.2 SOCIAL INFLUENCE

A key component of the model is the assumption that agents in the economy belong to a social network and face peer influence. We model two different mechanisms how an agent's investment choice is affected by her peers' stock market participation: *social utility* and *social learning*.

Social utility refers to the direct effect of peer j's stock market participation on agents i's utility. This effect may arise for various reasons widely discussed in finance literature. For example, agents may be concerned with their incomes or consumption levels relative to their peers, commonly referred to as "Keeping up with the Joneses" (Abel, 1990; Campbell & Cochrane, 1999; Georgarakos & Inderst, 2014). Alternatively, peer j's stock market investment may affect agent i's utility through "joint consumption" of the assets: peers can follow and discuss financial news together or track their joint returns (Taylor, 2011; Bursztyn *et al.* , 2014). Both of these explanation imply a higher utility from investing in stocks, irrespective of the return to investing.

We formally model the *social utility* channel through the participation costs. When deciding whether to invest in the stock market, agent i experiences lower fixed costs if more of her peers invest in stocks, thus making the agent more likely to start investing herself. The entry cost can be described as follows:

$$F_i = F_{i,0} - \theta p_i \tag{3}$$

The participation costs paid by agent i, F_i, is a function of initial participation costs $F_{i,0}$ and the number of agent i's peers who already invest in the risky asset, p_i. In (3), θ is an exogenous parameter that controls for how much the participation cost diminishes with every additional peer-investor. If an agent i has no peers already investing in the risky asset, the cost function is given by $F_i(\theta, F_{i,0}, p_i = 0) = F_{i,0}$. Consequently, an agent who has more informed peers faces lower participation costs, $\frac{\partial F_i}{\partial p_i} = -\theta < 0$. Modeling *social utility* channel through participation costs with more peer-investors decreasing costs, consequently increasing utility, offers a simplified representation of *the keeping up with the Joneses* effect (Abel, 1990) that helps keep the model tractable.

Social learning occurs when an agent acquires valuable information about stock market investments from her peers. Consider an agent i deciding whether to purchase a risky asset under uncertainty. When making this decision in isolation, agent i relies solely on her prior beliefs about market performance and invests only if the signal has sufficiently high precision. Within a social network, however, agent i can learn from her investing peers, allowing her to update beliefs and increase the signal precision. As a result, information shared by peers increases the likelihood of agent i's stock market participation compared to making the decision alone.[13]

We assume that in the economy each agent, when deciding to enter the stock market, faces a noisy signal about the expected return on risky asset, with $r_a = \mu + \epsilon$, $\epsilon \sim N(0, \sigma_0^2)$. Additionally, agents anticipate that upon entry, they will acquire signals from their informed peers, peers-investors, at no cost.[14] Among peers who share information about the stock market, each peer j provides an independent signal s_j such that $s_j = r_a + \varepsilon_j$, where $\varepsilon_j \sim N(0, \sigma_p^2)$.[15]

Suppose agent i has $p_i = k_i + l_i$ peers, of whom only k_i generate informative signals about stock market returns.[16] Let $\mathscr{I}_i$ denote the information structure of agent i representing her posterior belief about r_a given her aggregated signal, $r_a | \mathscr{I}_i \sim N(\mu_{post,i}, \sigma^2_{post,i})$. The posterior

[13]Canonical models of herding behavior and asset-price bubbles are rooted in social learning and extensively analyze its implications for financial markets (Bikchandani & Sharma, 2000; Chari & Kehoe, 2004).

[14]The modeling approach is similar to Peress (2003). However, we ignore the possibility of the agents to purchase better signals, as discussed in section 6.3 of Peress (2003). The possibility of acquiring an additional signal would contribute to even higher differences in incentives for stock market participation between rich and poor agents since the wealthy can afford to pay more for the information.

[15]We will define the exact rule of how and which peers share this information in Section 2.5.

[16]Alternatively, we could distinguish between k_i high- and l_i low-precision informative signals. However, to minimize the number of parameters, we interpret low-precision signals as completely uninformative.

mean and variance of the risky-asset return are as follows:

$$\mu_{post,i} = \frac{\frac{\mu}{\sigma_0^2} + \frac{\sum_{j=1}^{k} s_j}{\sigma_p^2}}{\frac{1}{\sigma_0^2} + \frac{k_i}{\sigma_p^2}}$$

$$\sigma_{post,i}^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{k_i}{\sigma_p^2}}. \tag{4}$$

In line with the conjecture in Arrondel *et al.* (2022), agents with a higher number of peers-investors receive more signals and benefit from greater precision in their posterior beliefs about stock market returns. Thus, due to *social learning*, agents with more informed peers are more likely to participate in the stock market. Thornton (2021) find evidence that peers influence financial expectations, consistent with this channel.

2.3 FINANCIAL EDUCATION

In addition to heterogeneity in initial endowment, we assume that agents differ in their financial sophistication: Two types of agents populate the economy: Financially Educated and Non-Financially Educated (Van Rooij *et al.* , 2011; Behrman *et al.* , 2012). For tractability, the type of agent i is denoted by t_i, where t_i equals one if agent i is Financially Educated and zero if she is Non-Financially Educated. The two types of agents differ in their participation cost $F_{i,0}$. Financially Educated agents have ex-ante knowledge about the investment in the risky asset, meaning their fixed entry cost is zero, $F_{0,i}(t_i = 1) = 0$. Consequently, if a Financially Educated agent expects a positive return on stock market investment, she will with certainty enter the stock market.[17] Whereas Non-Financially Educated agents do not have ex-ante knowledge about the stock market and face high participation cost, $F_{i,0}(t_i = 0) = F_0 >> 0$. Both types of agents are subject to social influences, both *social utility* and *social learning*. However, due to their low fixed costs, Financially Educated

[17] 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.

agents enter the stock market immediately, rendering social influence negligible.

2.4 THE AGENT'S OPTIMAL INVESTMENT DECISION

We first consider the decision problem of an individual agent i who chooses whether to enter the stock market and, upon entry, how much to invest in the risky asset within the previously described economy. At first, based on her prior beliefs, type, and income, agent i observes the number of peers who participate in the stock market and anticipates how many of them will provide her with informative signals upon entry.[18] She then decides whether to enter the stock market. Next, after entering the market, agent i, conditional on the realized signals, determines the optimal portfolio share for the risky asset, α_i.[19]

We solve agent's investment problem using backward induction. Conditional on entry, agent i solves the following portfolio optimization problem

$$max_{\alpha_i} E[U(W_{1,i})|\mathscr{I}_i] = max_{\alpha_i} E\left[\frac{W_{1,i}^{1-\gamma}}{1-\gamma}\middle|\mathscr{I}_i\right], \tag{5}$$

$$s.t \quad W_{1,i} = \alpha_i(W_{0,i} - F_i)R_a + (1-\alpha_i)(W_{0,i} - F_i)R_f = (W_{0,i} - F_i)\underbrace{(\alpha(R_a - R_f) - R_f)}_{\equiv R_p},$$

where $R_p = e^{r_p}$ denotes the portfolio return factor where the portfolio consists of risk-free and risky assets, and $\mathscr{I}_i$ denotes agent i's information set based on a particular signal realization. In line with standard approximation in financial literature (Campbell & Viceira, 2002), $r_p = \log(R_p)$ can be expressed as follows.

$$r_p = r_f + \alpha_i(r_a - r_f) + \frac{1}{2}\alpha_i(1-\alpha_i)\sigma^2_{post,i}.$$

[18]Specifically, we assume that agent i expects k_i of her investing peers to share information about stock market returns, helping her make more informed financial decisions and optimize her future portfolio.

[19]Since we impose no restrictions on her investment strategy, we expect this share to be almost surely nonzero.

Solving (5) for α_i, we get the optimal portfolio share allocated to the risky asset:[20]

$$\alpha_i^* = \frac{1}{\gamma} \frac{E(r_a - r_f|\mathscr{I}_i) + \frac{1}{2} Var(r_a|\mathscr{I}_i)}{Var(r_a|\mathscr{I}_i)}, \tag{6}$$

The optimal risky share, α_i^*, depends on the agent's risk-aversion, γ, and on her posterior beliefs about the asset performance, $E(r_a - r_f|\mathscr{I}_i)$ and $Var(r_a|\mathscr{I}_i)$.

Now, knowing the optimal portfolio share, we can rewrite the expected utility function and consider the entry decision. Conditional on market entry, obtaining signals, and choosing optimal portfolio share α_i^*, the expected utility of the agent is

$$E[U(W_{1,i}, \alpha_i^*)|\mathscr{I}_i] = \frac{(W_{0,i} - F_i)^{1-\gamma}}{1-\gamma} e^{\frac{(1-\gamma)\left(4\sigma^2_{post,i}\left(\mu_{post,i}+(2\gamma-1)r_f\right)+4(r_f-\mu_{post,i})^2+\left(\sigma^2_{post,i}\right)^2\right)}{8\gamma\sigma^2_{post,i}}}.$$

Notice that before entering the stock market, the agent does not know the specific realization of the signals. However, she expects that, conditional on obtaining signals, the posterior distribution of the risky asset return follows $r_a|\mathscr{I}_i \sim N(\mu_{post,i}, \sigma^2_{post,i})$. Since $\mu_{post,i}$ depends on the actual realization of the signals, it is unknown ex-ante. Therefore, provided that agent i receives at least one informative signal $k_i \geq 1$, she forms beliefs about the distribution of $\mu_{post,i}$, which follows[21]

$$\mu_{post,i} \sim N\left(\mu, \sigma^2_{\mu,i}\right), \text{where } \sigma^2_{\mu,i} = \sigma^2_{post,i} + \sigma^2_0(\sigma^2_{post,i})^2 \left(\frac{k_i}{\sigma^2_p}\right)^2,$$

If the agent does not expect to receive any signals ($k_i = 0$), her posterior distribution remains identical to her prior, meaning that $\mu_{post,i} = \mu$.

As a result, agent i enters the stock market if her expected utility is above $\frac{W_{0,i}^{1-\gamma}}{1-\gamma}$, and she doesn't enter otherwise. Let's denote by σ_i the strategy of agent i, where σ_i is 1 if the agent decides to enter the stock market, and 0 otherwise. The optimal decision to enter the stock

[20] Appendix D.1 describes the detailed solution of the problem in (5).

[21] See Appendix D.1 for the proof.

market for agent i, which depends on her type, initial wealth and number of peers-investors, generating informative and uninformative signals, can be described as follows:

$$\sigma_i(W_{0,i}, t_i, k_i, l_i) = \begin{cases} 1, & \text{if } f_i(W_{0,i}, t_i, k_i, l_i) \geq 0 \\ 0, & \text{if } f_i(W_{0,i}, t_i, k_i, l_i) < 0, \end{cases} \tag{7}$$

where

$$f_i(W_{0,i}, t_i, k_i, l_i) = E\left[\frac{W_{1,i}^{1-\gamma}}{1-\gamma}\middle| t_i, k_i, l_i\right] - \frac{W_{0,i}^{1-\gamma}}{1-\gamma}. \tag{8}$$

$E\left[\frac{W_{1,i}^{1-\gamma}}{1-\gamma}\middle| t_i, k_i, l_i\right]$ represents the expected utility conditional on entering the market before signals realization.[22]

Important to notice that with the number of peers-investors providing informative signals approaching infinity, $k_i \to \infty$, the variance of the aggregated signal converges to zero. In this case, agent i always prefers to enter the market, as she expects to obtain highly precise information about market returns. If the expected return falls below the risk-free rate, agent i instead opts for a short position in the risky asset.

2.5 INFLUENCE PROPAGATION IN THE SOCIAL NETWORK

All agents in the economy belong to a social network. The structure of the network is described by $\{N, G, W, T, SP\}$, where N is a set of agents-nodes of cardinality n, the number of agents in the economy. G is an $n \times n$ adjacency matrix describing connections between agents in the network: $G = \{g_{ij}, \forall i, j \in N$ such that $g_{ij} = g_{ji} = 1$ if i and j are linked, and $g_{ij} = g_{ji} = 0$ otherwise$\}$. A link between agents i and j signifies that they can communicate, and thus, they can exert *influence* on each other. We use the term "influence" to describe the process through which agent i affects agent j's stock market participation. It can go through the *social utility* or *social learning* or both both. Ex-ante influence can propagate in

[22] The details of derivation of $E\left[\frac{W_{1,i}^{1-\gamma}}{1-\gamma}\middle| t_i, k_i, l_i\right]$ are provided in Appendix D.1.

either direction, but the actual communication patterns and the direction in which influence propagates – whether from i to j or vice versa – emerge as equilibrium outcome of the model. Section 3.5 details how the matrix G is generated given specific model parameters, connectivity and homophily.

$W = W_{0,1}, ..., W_{0,n}$ is a vector of length n that describes the initial discretionary income allocated to each agent in the network, while $T = t_1, ..., t_n$ is a binary vector identifying agent's type based on financial education. For agent i, both $W_{0,i}$ and t_i are exogenously given and independent of the influence propagation process, yet they directly affect the decision to participate in the stock market. Finally, vector SP captures signal precision, where $sp_i = 1$ indicates that agent i generates an informative signal when investing and communicating with peers, whereas $sp_i = 0$ means agent i does not generate informative signals. We assume that all financially educated agents always produce informative signals. In contrast, among non-financially educated agents who participate in the stock market, only a proportion r_h generate informative signals. We treat the type of signals agents can generate, given a specific network realization $\{N, G, W, T, SP\}$, as exogenously determined. However, the influence propagation process determines if they actually generate the signals.

To reduce the number of parameters in the model, we introduce the concept of a *Participation Threshold Set*. For any agent i, investment in the risky asset now depends on the composition of her peers in terms of information quality. Specifically, we define a *Participation Threshold Set* as a set of pairs $(\hat{k}_i, \hat{l}_i)$, such that

$$\begin{aligned} \hat{S}_i = &\{(\hat{k}_i, \hat{l}_i) \in \mathbb{N} \times \mathbb{N} \mid \forall (k,l) \text{ with } k \geq \hat{k}_i, l \geq \hat{l}_i, f_i(k,l) \geq 0, \\ &\text{and } \forall (k,l) \text{ with } k \leq \hat{k}_i, l \leq \hat{l}_i, \text{ and } (k < \hat{k}_i \text{ or } l < \hat{l}_i), f_i(k,l) < 0\}, \end{aligned} \tag{9}$$

where $\hat{k}_i$ represents the number of peers generating high-precision signals about the stock market, and $\hat{l}_i$ represents the number of peers investing but not generating any signals, with function $f_i(k,l)$ defined in equation (8). Agent i invests in the risky asset if she has a peer-group composition that meets at least one of the combinations in $\hat{S}_i$. In other words, the *Participation Threshold Set* defines a set of (k_i, l_i) pairs, where each pair represents a pivotal mix of the agent i's peers-investors, generating informative signals and not, that guarantees the agent's stock market entry. The composition of pairs in $\hat{S}_i$ depends on the agent's discretionary income, $W_{0,i}$, and type, t_i.[23]

For Non-Financially Educated agents with high income, the fixed entry costs are smaller relative to income than for agents with low income. Consequently, high-income agents require lower social influence from their peers to start investing compared to the low-income agents. All Financially-Educated agents have a "zero" participation threshold pair, $\hat{S}_i(t_i = 1) = \{(0,0)\}$, since they already possess all the necessary information about the stock market. The introduction of participation threshold sets allows us to redefine the initial network structure as $\{N, G, S, SP\}$, where $\hat{S} = (\hat{S}_i)_{i \in N}$ is a vector of participation thresholds sets.

Agents who are connected can communicate with one another. Each agent determines whether to invest based on the expected number of informative signals received and the number of her peers-investors. An agent's strategy as a binary choice of whether to invest in the stock market or not is formally described by Equation (7) in the previous section. This decision depends on the agent's type, characterized by the participation threshold set $\hat{S}_i$, as well as the total number of peers investing $(k_i + l_i)$ and those providing informative signals (k_i). Let $P^+(i)$ denote the set of agent i's participating peers. The number of peers influencing agent i through the *social utility* channel is given by the size of this set, $p_i = |P^+(i)|$. Let $K(i) \subset P^+(i)$ represent the subset of agent i's peers who in addition to investing in stocks also share informative signals, thereby influencing agent's decision through the *social learning* channel. The size of this set is $k_i = |K(i)|$.

Now we can define a strategy of player i as a mapping $\sigma_i(\hat{S}_i, p_i, k_i) : \mathbb{N} \times \mathbb{N} \mapsto \{0,1\}$. We have $\sigma(\hat{S}_i, p_i, k_i) = 1$ if the agent decides to participate, and $\sigma(\hat{S}_i, p_i, k_i) = 0$, otherwise.

[23]We formally investigate key properties of this set in Appendix D.1

According to equation (7), the best response function of player i requires $\sigma(\hat{S}_i, p_i, k_i) = 1$ if $\exists(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$, s.t. $k_i \geq \hat{k}_i$ and $p_i - k_i \geq \hat{l}_i$. Otherwise, we require that $\sigma(\hat{S}_i, p_i, k_i) = 0$.

An agent who decides to participate in the stock market subsequently exerts influence on her peers. This interaction gives rise to a stock market participation influence propagation network, represented by the adjacency matrix X. While the adjacency matrix G captures potential connections between agents, matrix X represents the actual transmission of social influence. The influence propagation process, and thus matrix X, must satisfy the properties described in Definition 1.

Definition 1. ***The stock market participation influence propagation process*** *describes how agents in a network influence each other's decisions regarding stock market participation. This process is represented by the adjacency matrix X, where $x_{ij} \in X$ equals 1 if agent i exerts influence on agent j and $x_{ij} = 0$ otherwise. The influence propagation process satisfies the following properties:*

1. *Agents i and j can only exert influence on each other if they are peers according to network $\{N, G, S, SP\}$. Formally, $x_{ij} = 1$ only if $g_{ij} = 1$.*

2. *Influence flows in only one direction between any pair of connected agents. Specifically, for any i and j, either i influences j or j influences i, or there is no social influence between them, $x_{ij} + x_{ji} \leq 1$.*

3. *If agent i exerts influence on j ($x_{ij} = 1$), then agent j experiences a reduction in stock market participation costs either trough social utility channel or through both social utility and social learning.*

4. *If agent i participates in the stock market, she influences on all her connected peers who do not already exert influence on her. Formally, if $\sigma_i = 1$ and agents i and j are connected, $g_{ij} = 1$, then either j exerts influence on i or i exerts influence on j, $x_{ij} + x_{ji} = 1$.*

A set of agent i's participating peers who exert influence on i through the social utility channel is defined as $P_i^+ = \{j \in N | x_{ji} = 1, x_{ji} \in X\}$, and the total number of such peers is given by $p_i = \sum_{j=1}^n x_{ji}$. Similarly, the set of agent i's participating peers who provide additional influence through the social learning channel is $K_i = \{j \in N | x_{ji} \times sp_j = 1, x_{ji} \in X\}$, and $k_i = \sum_{j=1}^n x_{ji} \times sp_j$. These sets are endogenously determined for each agent as an equilibrium outcome.

Importantly, we assume that influence propagation is one-directional. If agent i participates in the stock market and is connected to agent j, one of two outcomes occurs: either agent i influences agent j's decision, reducing j's stock market participation costs, or agent j influences agent i's decision, reducing i's participation costs. This assumption rules out mutual reinforcement, where agent i decides to participate solely because she expects j to participate, while agent j does the same based on her expectation that i will participate.[24] In Appendix E, we examine how allowing two-way influence propagation alters the equilibrium and analyze the best possible outcome for a given network structure under full cooperation. However, in the context of stock market influence propagation, such cooperation is unlikely, so our primary focus is on non-cooperative outcomes.

However, we show that imposing a one-directional influence propagation condition is not sufficient to eliminate all potential cooperative outcomes. To illustrate this, consider a network consisting of 6 agents. For simplicity and clarity, we assume that agents do not share any informative signals, $\forall sp_i \in SP, sp_i = 0$, meaning they influence each other's decisions solely through the social utility channel. Consequently, the vector of relevant participation

[24] Our main analysis follows a one-way influence flow approach, similar to Bala & Goyal (2000) minimal network formation idea. In a dynamic setting, this approach would correspond to an agent making a participation decision based only on peers who have already invested, rather than on the expected number of future participants whose choices depend on the agent's own decision. Conceptually, one-way influence propagation does not imply that participating peers cannot later exchange information or exert social utility on one another. Rather, it means that, at the moment of deciding to participate, agents cannot rely on potential future connections to peers who have not yet joined. Our approach captures this idea within a static framework, without explicitly modeling a dynamic process. We also explore additional results for a two-way information flow model in Appendix E.

thresholds is given by

$$\forall i \in N, \{(0, \hat{l}_i) \mid (0, \hat{l}_i) \in \hat{S}_i\} = \{\{(0,0)\}, \{(0,1)\}, \{(0,2)\}, \{(0,2)\}, \{(0,1)\}, \{(0,1)\}\}.$$

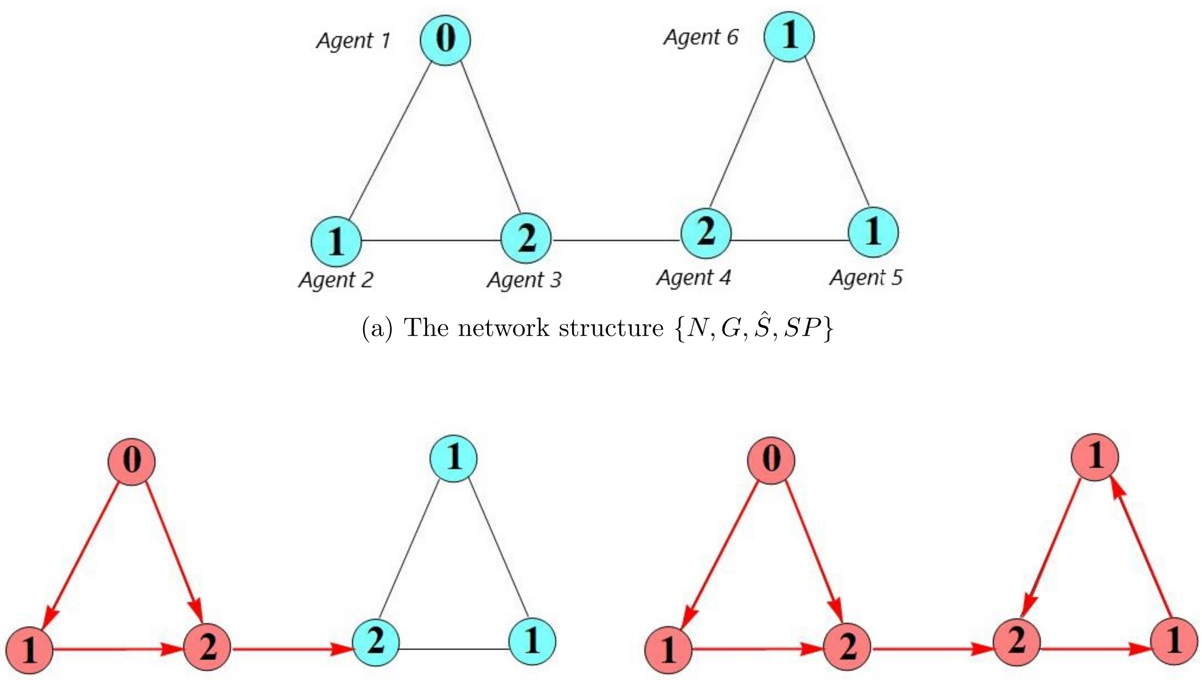
The following adjacency matrix G represents the connections between the agents:

$$G = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Figure 1a illustrates the network structure $\{N, G, \hat{S}, SP\}$. The colored circles correspond to agents who can potentially invest in the risky asset. The assigned numbers represent the participation thresholds $\hat{l}_i$ for each agent given that $\hat{k}_i = 0$. Black edges indicate connections between agents through which they can exert influence on each other's stock market participation decisions. Figures 1b and 1c depict two possible ways in which communication flows within the network. In these figures, red-colored nodes correspond to agents who participate in the stock market, while red edges illustrate the actual direction of influence. A red edge from agent i to agent j signifies that agent i exert influence on agent j. The corresponding adjacency matrices, which formally capture the realized flow of influence, can be constructed for each case. Case 1 (1b) adjacency matrix X, and Case 2 (1c) adjacency matrix X' are as follows:

$$X = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad X' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

In both cases, agents participate in the stock market if $p_i = l_i \geq \hat{l}_i$. Any agent who does not participate has no outgoing edges. Any agent who does participate has either incoming or outgoing edges to their peers, as defined by matrix G. However, agent 4 with a participation



(a) The network structure $\{N, G, \hat{S}, SP\}$

(b) SMP network: Case 1

(c) SMP network: Case 2

Figure 1. An example.

threshold equal to 2 does not participate in the stock market in Case 1 but participates in Case 2. To understand why the outcomes for agent 4 are different in two settings, let us take a closer look at Case 2. In Figure 1c, agent 2 has two incoming edges. Therefore, agent 2 participates only because she is influenced by agents 3 and 6. Agent 6, in turn, participates due to the influence of agent 5, while agent 5 is influenced by agent 4's decision to participate. This creates a loop: agent 4 participates because she anticipates that her decision will influence agent 5, which will then influence agent 6, ultimately reinforcing agent 4's own participation.[25]

Thus, Case 2 represents a scenario where a collaborative influence loop exists among agents, allowing participation to be reinforced through mutual expectations. We aim to exclude such situations from consideration. We assume that an agent will only participate and exert influence on others if she has already received sufficient influence herself—specifically, at least as much as her participation threshold requires. This ensures that participation is based on actual influence rather than anticipatory cooperation. Since we do not explicitly model the formation dynamics of the influence propagation matrix, we introduce the following notion of a non-cooperative equilibrium. This concept formalizes our assumption and guarantees the absence of influence loops.

Definition 2. *The strategy profile σ is a* ***non-cooperative Nash Equilibrium*** *profile if there exists an acyclic directed graph represented by the adjacency matrix X such that*

1. $x_{ij} = 1$ *only if* $g_{ij} = 1$.
2. $\sigma_i = 1$ *if and only if there exists* $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ *s.t.* $\sum_{j \in N} x_{ji} \geq \hat{k}_i + \hat{l}_i$ *and* $\sum_{j \in N} x_{ji} sp_j \geq \hat{k}_i$*, and* $\sigma_i = 0$ *otherwise.*
3. *If* $\sigma_i = 1$*, for all* $j \in N$*, if* $g_{ij} = 1$*, it follows that* $x_{ij} + x_{ji} = 1$.
4. *If* $\sigma_i = 0$*, for all* $j \in N : x_{ij} = 0$.

A key additional requirement in the definition of a non-cooperative equilibrium is the existence of an *acyclic* directed graph, represented by the adjacency matrix X.[26] In the

[25] Similarly, we can construct a situation with a loop going from agent 4 to agent 6, from agent 6 to agent 5, and from agent 5 back to agent 4.

[26] Note that the existence of an acyclic graph is required, although it may not necessarily be unique. In contrast, we demonstrate later that the strategy profile σ is uniquely determined for a given network structure.

previous example, constructing a matrix X that describes an influence propagation process without influence loops — while still allowing agents 4, 5, or 6 to participate in the stock market and exert influence — is impossible. However, Figure 1b illustrates a non-cooperative equilibrium in which agents 1, 2, and 3 participate. Moreover, we cannot construct any matrix X that corresponds to the properties described in Definition 1, such that agents 1, 2, and 3 are playing their best responses and any of them does not participate in equilibrium. This leads us to the result introduced in Proposition 1.

Proposition 1. *For any given network structure* $\{N, G, \hat{S}, SP\}$ *there exists a unique non-cooperative equilibrium.*

Proof. See Appendix D.2 □

The proof of the existence of a non-cooperative equilibrium is constructive. We propose an algorithm that finds a set of participating agents (Appendix D). Notably, the uniqueness of the equilibrium means that there is a unique set of participating agents, such that we can construct a stock market participation influence propagation matrix X that satisfies the properties of a non-cooperative equilibrium. However, it does not mean there is a unique way to identify matrix X. We provide an example in Appendix D.3.

2.6 PARAMETER VALUES IN SIMULATIONS

We now describe the parameter values used in the simulations. We study how simultaneous changes in model parameters affect stock market participation. The values of the fixed model parameters primarily come from U.S. financial and macroeconomic data from 2014 and are summarized in Table B3.

Fixed parameters – We perform some preliminary computations for model estimation. We assume that the income distribution is log-logistic (Atkinson, 1975), and agents in the model can invest their income minus a minimum wage in the stock market. We obtain historical data on annual stock market risk premium r_m and risk free rate, r_f. We set expected risky asset return to $\mu = r_m + r_f$. We also use VIX S&P 500 index to approximate both volatility of prior beliefs, σ_0^2, and volatility of informative signals, σ_p^2.[27] Moreover, we add information about the income distribution from the U.S. 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 and insurance sectors (52 NAICS) in 2015 from the Quarterly Census of Employment and Wages as a proxy for financial education.

Connectivity and homophily – The connectivity parameter, c, controls for the expected number of links (peers) for each agent in the population. We assume that every agent has the same expected number of peers, independent of their other characteristics. However, the actual number of peers varies across agents. We split all agents into five income groups based on income distribution quintiles.

Each agent forms a link with an agent 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 for the difference between unconditional probabilities to form the link with agents within and outside the agent's income group, $h = p_{In} - p_{Out}$. If homophily parameter h equals 0, each agent is equally likely to form a link with any other agent inside and outside her income group. If the homophily parameter h equals 1, each agent forms a connection only inside her income group. It is important to note that homophily and connectivity parameters are independent. In the baseline model for the parameter calibrations, we set the homophily parameter equal to 0.5. This value implies that agents are more likely to connect to agents within their income group as to agents outside their income group. Empirical evidence on segregation between low and high-income households suggests that homophily may be even higher, especially

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

among high-income households (Reardon & Bischoff, 2011; Massenkoff & Wilmers, 2023). In the calibration exercise, we will allow homophily to vary to explore how different levels of homophily change the effect of connectivity.

All economic agents in the model belong to a social network represented by an adjacency matrix G. To construct the network, we model the degree distribution using a log-normal distribution. While the literature suggests that social networks can typically be well-represented by either power-law or log-normal degree distributions (Mitzenmacher, 2004), some evidence indicates that log-normal distributions provide a better fit for a larger proportion of networks (Broido & Clauset, 2019). This is often found in social networks where maintaining meaningful connections requires at least some effort (Sheridan & Onodera, 2018). In our context, it is relevant because the majority of individuals tend to share financial information only with a limited number of close friends and relatives (Balakina *et al.* , 2024; Lusardi *et al.* , 2010). Furthermore, the log-normal distribution allows us to align our results more closely with the empirical evidence presented by Arrondel *et al.* (2022), with a power-law degree distribution generating unrealistically large hubs.[28]

To construct a connectivity matrix, we first assign a number of peers to each agent following a log-normal distribution with an expected mean of c.[29] Second, at each iteration, we start with an agent with the number of formed links below the preassigned number of peers. Next, we compute conditional probabilities to form additional links to other peers given the number and the outside/inside income group nature of already formed links. We update the conditional probabilities for each agent at each iteration such that the unconditional probabilities to form links inside and outside the agent's income group remain the same for

[28]Nonetheless, we also conducted similar analyses using both binomial degree distributions and power-law distributions implemented through a preferential attachment mechanism adjusted for homophily. The results remain qualitatively similar.

[29]We also set the shape parameter to 0.75. The literature provides different estimates based on various types of networks, but they typically range from 0 to 1 (Smith, 2021). The choice of shape parameter is motivated by the empirical findings of Arrondel *et al.* (2022), where the largest reported number of connections through which an agent shares financial information is 100. An agent with such a large number of connections is almost certainly participating in the stock market. Our choice of the shape parameter allows us to generate hubs close to 100 connections.

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.

Calibrated Parameters Our model includes three key parameters that we estimate through calibration: the cost function parameters F_0 and θ, as well as the proportion of non-financially educated agents who generate informative signals, denoted as $r_h \equiv \sum_i^n (sp_i)/n$.

To estimate these parameters, we use data on participation rates across different income groups. We calibrate the model to match observed participation patterns using a maximum likelihood approach.[30]

The estimated cost F_0 is \$20,400, and the estimated value of θ is \$3,900. Previous literature provides a wide range of estimates, from \$260 in Vissing-Jørgensen (2002) to \$134,000 in Andersen & Nielsen (2010). Notably, past estimates of F_0 do not account for information received from peers. For instance, the lower estimate of \$260 from Vissing-Jørgensen (2002) can be reconciled with our estimate of \$20,400 by assuming that individuals obtain information from approximately 5.2 peers. Given this, we consider our estimates to be reasonable.[31]

Network size and structure – Given the complexity of the calculations, we approximate the population size as 10,000 agents in all simulations. Each agent is assigned an

[30]We use data from the Survey of Consumer Finances, provided by the Board of Governors of the Federal Reserve System. The calibration relies on reported stock market participation rates for different income groups in 2013 and 2016, averaging the two periods to approximate participation in the 2014 U.S. stock market. Our estimation follows these steps: (1) we compute stock market participation in our model for the same income groups as in the survey, varying θ in steps of 100 and r_h in steps of 0.1; (2) for each combination of θ and r_h, we determine the participation cost F using a gradient descent method to ensure the model's average participation rate aligns with the empirical data within a ± 0.03 margin; and (3) we identify the best-fitting values for θ and r_h using a maximum likelihood approach, and take the corresponding value of F.

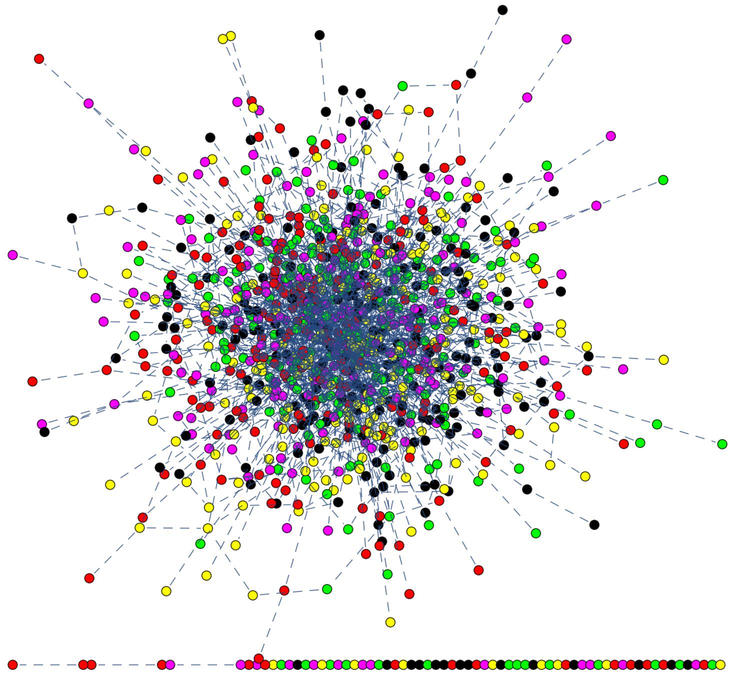
[31]It is also important to note that in estimating the cost, we do not impose any restrictions on agents' investment strategies, allowing them to borrow and potentially take short positions. Introducing additional constraints would significantly lower our cost estimate.

income parameter, randomly drawn from a log-logistic distribution, as described earlier. Additionally, each agent is randomly classified as either financially educated or not, with no correlation assumed between education level and wealth in the benchmark model. However, we discuss in the results section how introducing a correlation between these factors may impact the results. Non-financially educated agents are also randomly assigned the type of signal they will generate, using the estimated parameter r_h. The adjacency matrix is formed following the procedure described above.

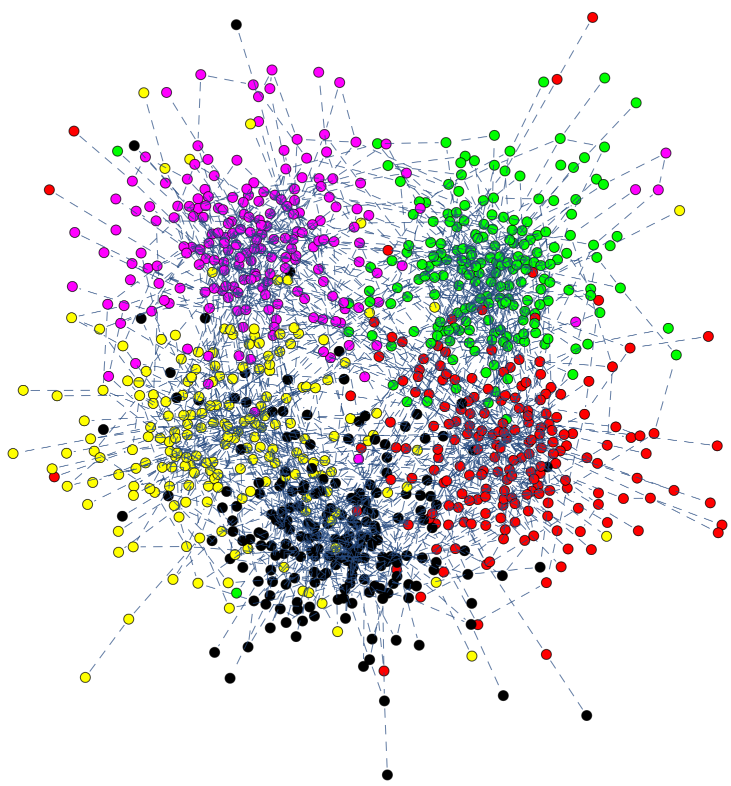
The large size of the network allows us to have robust simulation results for each set of the parameters, despite the random network structure. However, to provide a more reliable estimates, we apply the following procedure. We begin by sampling values for the income vector according to the specified income distribution parameters, and then we construct a random network based on the sampled values. After determining the equilibrium outcome for the network, we repeat the process of constructing a random network 10 times. Finally, we report the average outcomes of these 10 iterations.

3. MODEL RESULTS

We now present the main results for the model simulations. We run simulations of the full model, including both social utility and social learning channels, and use parameter values described in Table B3. Overall, we consider 1,815 combinations of parameters for the main results. 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 the homophily parameter, which governs how likely agents are to connect with agents from other income groups. Figure 2 illustrates the network structure with low homophily in panel a) and high homophily in panel b). Various colors signify agents with different income levels. In an economy with low homophily, panel a), there are no clear clusters of income groups. In contrast, the economy with high homophily, in panel b), shows distinct clusters among



(a) Low homophily



(b) High homophily

Figure 2. Network structure visualisation

Notes: The figures plot networks generated by our algorithm. For illustration purposes, we constructed networks with 1,000 agents and average number of connections of 4. Vertices of different colours represent agents of different income groups. We set the homophily index of 0.1 for low homophily and 0.9 for high homophily.

different income groups.

We focus on how the number of connections, homophily, and ex-ante income inequality affect the share of stock market investors, starting with the entire population. We also explore how each parameter of interest affects different income groups. The groups coincide with the five income groups used to construct the network.

3.1 THE EFFECT OF CONNECTIVITY

Figure 3 presents the relation between connectivity and stock market participation. The solid orange line plots the average participation, and black, gray, blue, green, and light purple dashed lines present participation among five income groups, starting from the lowest. In the figure, we vary the average number of connections on the x-axis and fix all other parameters.

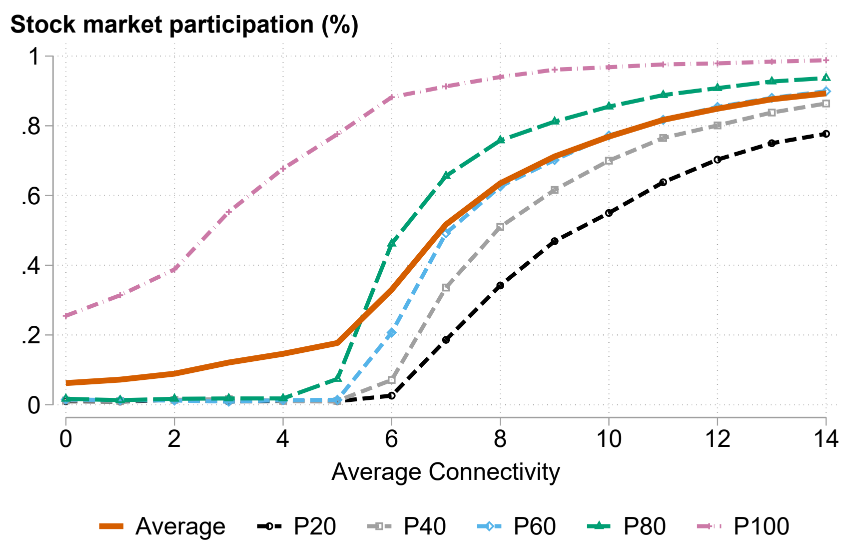


Figure 3. The effect of connectivity by model parameters

Notes: The figure plots stock market participation (y-axis) against the average number of connections (x-axis). The orange solid line plots the average stock market participation among all agents. The dashed black, gray, blue, green, and light purple lines plot average participation among income quintile groups, starting with the lowest. We set the homophily parameter to 0.5 and the ex-ante Gini coefficient to 0.4 for the simulations.

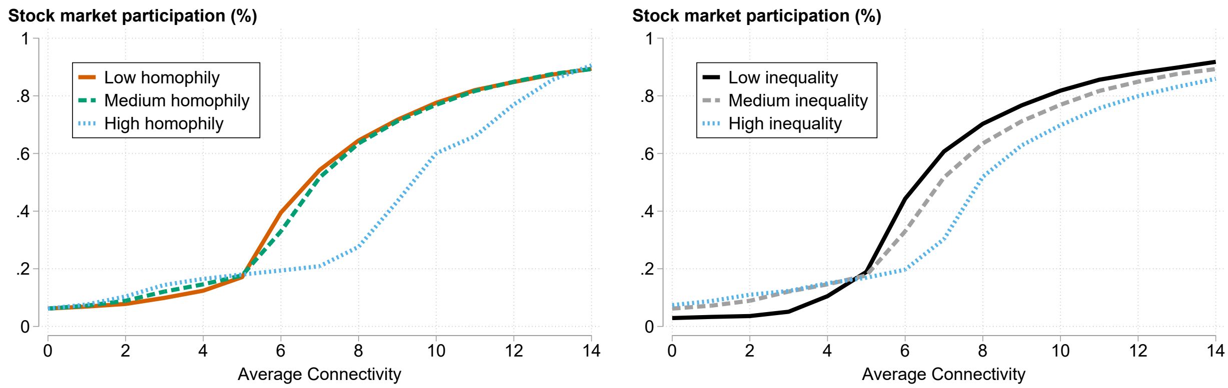
The relation is S-shaped, in particular for middle income agents (incomes between P40 and P80). The effect of adding one more peer is small for low levels of ex-ante connectivity.

However, as the number of connections increases, the marginal effect of adding one more peer grows. The figure also shows that not all income groups benefit equally from higher connectivity. Intuitively, lower-income agents are further from the participation threshold and thus need more connections to participate.

Since high-income agents are more likely to be close to the participation threshold already, connectivity positively impacts their participation. With more than five connections on average, the effect diminishes, again giving an S-shaped pattern. With more than ten connections, essentially all high-income agents participate. In contrast, middle income agents, denoted by the green, blue and gray dashed line, need a larger number of connections before connectivity starts to have an impact. These 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 far from the participation threshold and thus need many peers before participating.

This first result of our model implies that we expect to see rising stock market participation from higher connectivity. Due to increasing social media usage over the last 20 years, individuals have had ample opportunity to widen their social networks, leading to a likely increase in social connectivity. However, stock market participation has been flat for all income groups over the last twenty years according to data from the Survey of Consumer Finances. A natural question is why rising connectivity has yet to lead to rising stock market participation, as we would expect from the first model results, and, indeed, from a large empirical literature on peer effects in financial decisions. Below, we argue that homophily and inequality mediate the effect of connectivity, which helps explain the lack of response in stock market participation.

3.2 THE MEDIATING EFFECT OF HOMOPHILY AND INEQUALITY FOR CONNECTIVITY



(a) Homophily (b) Inequality

Figure 4. 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 in the economy. 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 different values for the Gini coefficient: 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.

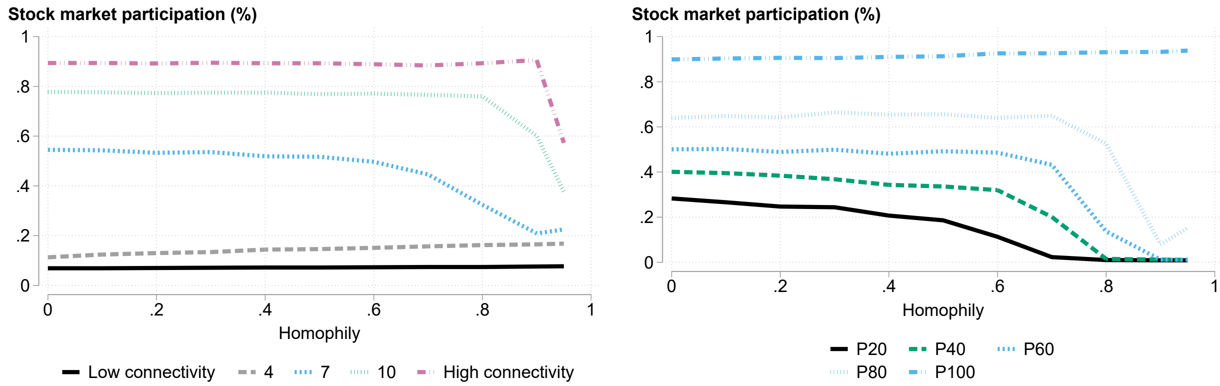
We now examine how the levels of homophily and inequality affect the relationship between connectivity and stock market participation. Recall that the homophily parameter measures how likely two agents from different income groups are to connect. Panel a) of Figure 4 plots the connectivity and stock 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 the average number of connections below eight: the blue line with a high level of homophily is consistently above the other lines. However, the effect flips if connectivity is high. If connectivity is above five in panel a) in Figure 4, 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 cover participation costs, 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. We further discuss results for homophily in Section 3.3.

A similar pattern appears for inequality. Panel b) of Figure 4 plots the connectivity and stock market participation for different levels of the Gini coefficient. We choose Gini coefficients of 0.25, 0.4, and 0.6 for low, medium, and high inequality, respectively. We keep the average income level constant in the simulations but adjust other distribution parameters.[32] As the Gini coefficient increases, more income 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. The S-shaped pattern starts to flatten due to information sharing being limited by a high share of agents with low income. 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. We further discuss results for inequality in Section 3.4.

3.3 THE EFFECT OF HOMOPHILY

Figure 5 shows that homophily has a small but 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 the stock market is more likely to spread to agents far from the participation threshold, who, consequently, do

[32] We assume that income is distributed according to a 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 5. 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, bottom quintile, is marked with a solid black line, the second, third, and forth quintiles are marked with various patterns of light blue color, and the top income quintile group, high income, 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.

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 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 cover the costs, 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 less of an impact on participation.

To illustrate who benefits from higher homophily, Panel b) plots stock market participation against homophily for five income groups with the level of connectivity set to seven peers, 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 slightly increase their stock market participation. For low and medium-income groups, homophily has little impact on participation rates, while homophily is at a moderate level. However, when homophily increases, the low-income agents are negatively affected first, followed by medium-level income agents. High levels of homophily limit information transmission from one income group to another, thus diminishing the access of low and medium-income agents to participating peers from high-income groups who can share the information.

In the above results, we assume that homophily is related to income and that income directly affects stock market participation. There is ample evidence that homophily is directly related to income or wealth (e.g. Fagereng *et al.* , 2022; McPherson *et al.* , 2001). Figure C1 shows that homophily does not have an effect if it is unrelated to factors that determine stock market participation. These results suggest that rising homophily will only affect stock market participation if sorting is based on the factors directly affecting the stock investment decision. For example, imagine that homophily has increased based on political preferences, but stock market participation and income are similar among different political parties. In this scenario, the results in Figure C1 predict that the increase in homophily will not impact aggregate stock market participation. This prediction could be tested empirically.

Finally, we explore whose homophily matters more for stock market participation. Until now, we have assumed that the homophily parameter is constant across agents. We now want to explore how higher level of homophily for one group, specifically high-income households, affects average stock market participation in the economy. We find that homophily among high-income agents has an impact on other agents in the model. Due to high homophily high-income agents form a tight cluster, resulting in other agents not getting access to them. We can think of this as the *gated-community* effect: homophily among high-income agents affects all other agents, even if the sorting is present only in the preferences of high-income agents. In the gated-community example, community members limit their interactions with other agents outside the community by segregating themselves, even if the other agents outside would like to socialize across groups.

Figure 6 plots stock market participation among middle-income agents against homophily level among high-income agents. A zero value on the x-axis corresponds to the scenario when high-income agents are equally likely to connect to any other agent, inside and outside high-income group. A value of one on the x-axis describes the case when high-income agents only connect to other high-income agents. Disregarding the difference between the colored bars for a moment, we see that high homophily among rich agents leads to lower stock market participation for everyone else. When rich agents segregate themselves, they inhibit information sharing among all other agents too.

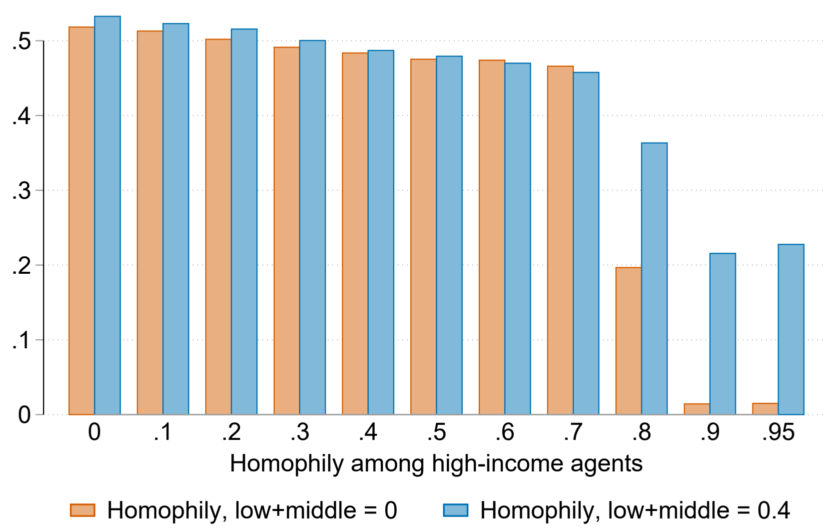


Figure 6. The effect of homophily among high income agents

Notes: The figure plots stock market participation depending on homophily among high- and low- and middle-income agents. High income group includes agents from top 20 percent of earners in the income distribution. Middle income group includes all agents with income above the 20_{th} and below the 80_{th} percentile of the distribution. Orange bars describe the stock market participation with zero homophily among low and middle income agents; light-blue bars describe the environment with homophily of 0.4 among low and middle income agents. We set ex-ante Gini coefficient to 0.4 for the simulations.

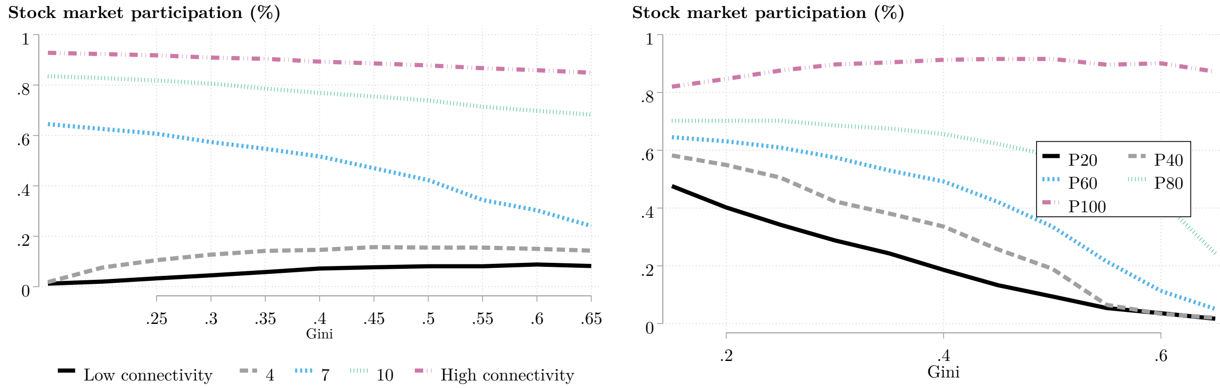
We further allow for two values of homophily for the rest of the agents, distinguished by the color of the bars. Note that "low+middle" homophily measures the likelihood that agents from the bottom four quintiles of income distribution connect to each other. A value of zero indicates that low income agents are equally likely to connect to low and middle-income agents. A value of 0.4 indicates that middle-income agents are more likely to connect to other middle-income agents than to low-income agents. This latter scenario, with low+middle homophily equal to 0.4, leads to higher participation rates when high-income agents are locked away in a gated community. If high-income agents are isolated in the network, middle-income agents are better off if they are also locked into a gated community.

The intuition is that higher homophily creates clusters among *low- and middle-income* agents, which helps information sharing. Moving from left to right in the figure, we further see that low homophily among the rich helps stock market participation among middle-income agents for both bars. In a gated-community scenario, when homophily among high-income agents is equal to one, middle-income agents will never meet with high-income agents who have information to share, and consequently stock market participation is low. In this scenario, it does not matter how much homophily there is in the rest of the society. The implication is that we need to consider different measures of homophily among agents who have information to share and that *average* homophily is potentially misleading. Empirically, we need to understand homophily among the group with information to share.

3.4 THE EFFECT OF INEQUALITY

We now examine how inequality affects stock market participation. In the simulations, we keep the average income fixed but vary other parameters of the income distribution. 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 afford to pay the fixed participation costs ex-ante before considering the effect of informative connections. Second, inequality also affects the probability that informed agents are connected to other agents through homophily.

Figure 7 provides the results for different levels of connectivity. For low levels of connectivity, black and gray dashed lines at the bottom of the figure, fewer agents have the income necessary to reach participation threshold at low levels of inequality, and there is little information sharing. As inequality increases, we take money from the poor and give it



(a) Population-level (b) By income groups

Figure 7. 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 bottom quintile income group is marked with black solid line, dashed grey, blue, and green lines correspond to average participation of the second, third, and forth quintiles, respectively. The light purple line marks the average stock market participation of the top income quintile of agents. Connectivity in panel b) is set to 7. The homophily coefficient in both panels is set to 0.5.

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 aggregate level of stock market participation.

We see a negative relationship between ex-ante inequality and participation for high connectivity. Higher inequality leads to clustering in the network and less efficient information diffusion. Consequently, stock market participation is lower. For instance, 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.

Inequality affects the distribution of participation thresholds across the network. The combination of threshold heterogeneity and network structure results in uneven diffusion of information. Richer agents, with lower participation thresholds, benefit more from their connections and from homophily. Karampourniotis *et al.* (2015) explore how the shape of threshold distributions influences informational cascades, comparing fixed and uniformly distributed thresholds. They find either model can generate relatively larger cascades, depend-

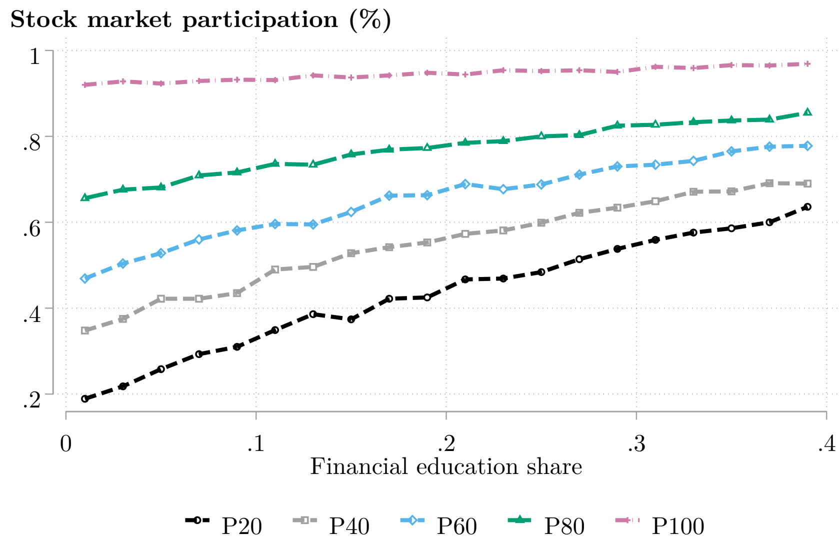


Figure 8. The effect of Financial Education

Notes: The figure plots participation rates by income groups against the share of financially educated agents. We set the Connectivity to 7, the homophily parameter to 0.5 and the ex-ante Gini coefficient to 0.4 for all simulations.

ing on the size of the initiator set. In our setup, the threshold distribution itself influences the number of initiators — those wealthy enough to invest without additional connections — who can spark the diffusion process. Nevertheless, our findings resonate with those of Karampourniotis *et al.* (2015), as we similarly observe that greater inequality in participation thresholds can either amplify or limit information cascades, but primarily based on network connectivity. Karampourniotis *et al.* (2019) further explore non-monotone relation of the shape of threshold distribution and the network degree assortativity. In our analysis, we primarily consider the impact of higher assortativity in participation thresholds driven by homophily, which arises from clustering among agents with similar income levels.

3.5 THE EFFECT OF HIGHER FINANCIAL EDUCATION

Financial education has a strong impact on stock market participation, in particular for low-income agents. We show this by exogenously varying the share of financially educated agents in the model. Figure 8 plots aggregate stock market participation against the share of financial educated agents. The effect of increasing the share is approximately linear for all income groups. The slope on financial education differs across income groups, however, and the stock market participation rates increases more for lower income groups for a given change in financial education. Intuitively, network effects imply that lower income groups benefit more from informed peers. With network effects, financial education targeting poorer households is more effective at boosting stock market participation.

Thus far, we have assumed that the probability of being financially educated is uniform across the wealth distribution. However, the literature suggests that financial sophistication is positively associated with wealth accumulation, with wealthier households possessing specialized skills that enable them to invest more effectively and earn persistently higher returns (Fagereng *et al.* , 2020). To investigate the implications of this relationship, we simulate the full model under varying degrees of correlation between wealth and financial education, ranging from zero to one, across different levels of homophily. A zero correlation implies that financially educated individuals are equally likely to be found among both poor and rich agents, whereas a correlation of one indicates that only the wealthy have access to financial education.

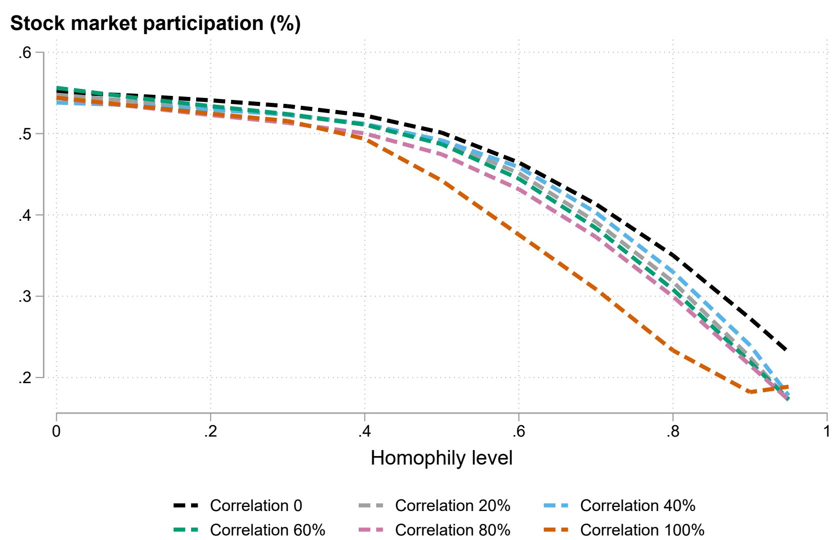


Figure 9. The effect of Correlation between Financial education and Wealth

Notes: The figure plots participation rates at zero, 20%, 40%, 60%, 80%, and 100% correlation between financial education and wealth against the level of homophily in the economy. We set the Connectivity to 7 peers, entry cost F to \$ 20,400, the probability of informed signal to 0.4, and the ex-ante Gini coefficient to 0.4 for all simulations.

The results, presented in Figure 9, show that when homophily is below 0.4, the correlation between financial education and wealth has a negligible effect on average stock market participation. However, as the level of homophily increases, the impact becomes more pronounced. For example, at a homophily level of 0.8, increasing the correlation from zero to one results in a decline in average stock market participation of more than 10 percentage points, from 35 percent to below 25 percent.

Figures 8 and 9 demonstrate that policies to improve financial education targeted towards low-income individuals would be especially effective in high-homophily economies.

4. THE IMPLICATIONS OF SOCIAL UTILITY AND SOCIAL LEARNING

In section 3 we present the results for the baseline model with both social utility and social learning channel in place. We now seek to understand the implication of different channels of social influence. Recall that the social utility channel affects the fixed costs of participation and that social learning affects the beliefs about returns. Empirical evidence suggests that both these factors are relevant for households decision making (Bursztyn *et al.* , 2014).

To compare the consequences of two channels in isolation, we simulate the model by "turning off" one of the channels at the time. To simulate the model with only social utility channel as active, we fix probability of Non-Financially Educated agents to spread informative signal to zero, $r_h = 0$. To isolate the effect of social learning, we simulate another version of the model with marginal benefit of having an informed peer set to zero, $\theta = 0$. The remaining parameter values are estimated in the model to match the average stock market participation.[33]

Figure 10 plots average and income-group specific stock market participation levels for simulation with only social utility (blue), only social learning (green), and both channels

[33] Running the simulations of the new versions of the model inevitably includes the re-estimation of the ex-ante value of participation cost, F. The resulting estimates remain in the range proposed by the literature.

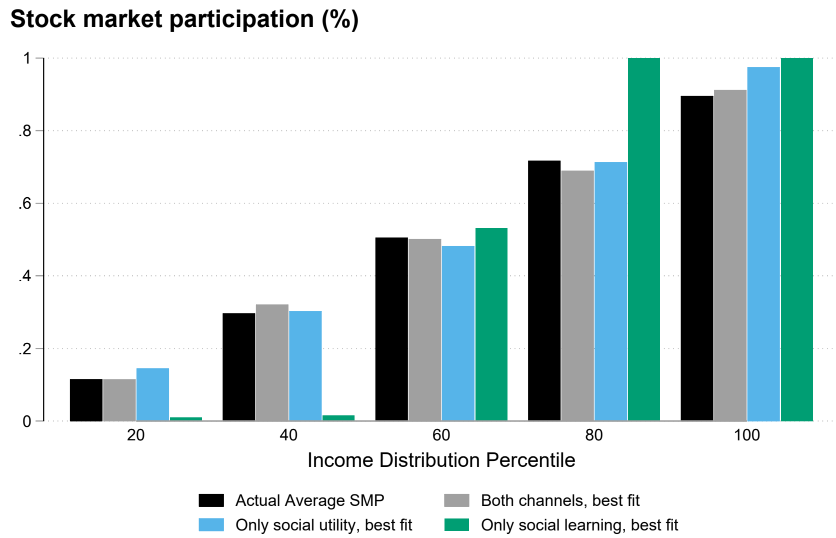


Figure 10. Distributional impact of Social Utility and Social Learning

Notes: The figure plots stock market participation for different income groups and different simulations for the best-fit models, based on ML estimation. The empirical stock market participation rates by income from the Survey of Consumer Finances are denoted by black bars. The gray bars provides the results from the baseline model, the blue bars provides the results from the model with only social utility, and the green bars provide the results for the model with only social learning. The results are split by income quintiles. We set the Connectivity to 7, the homophily parameter to 0.5 and the ex-ante Gini coefficient to 0.4 for all simulations.

(gray) against actual levels (black). We find that, while it is always possible to choose parameter values to match aggregate participation level, only the model with social utility provides a reasonably good fit to the actual stock market participation rates by income quintiles. Including social utility allows us to correctly match the participation among both low and high income agents. In contrast, the model with social learning predicts very low participation among the bottom 40 percent, which is compensated by very high participation rates in the 80^{th} percentile.[34] Intuitively, social utility affects the fixed cost of participation, which is especially crucial for poorer agents. Social learning affects the expected return, which is less important for poorer agents. Even if the expected return is high for poor agents, a fixed participation cost will deter their entry into the stock market. Overall, we find that social learning, given that it affects the expected return on investing, does not allow us to match the empirical patterns of participation. These results are consistent with recent

[34]Simulations with only social learning channel reinforce the stock market participation puzzle, with actual stock market participation among wealthy being too low and participation and among poor - too high compare to model predictions.

research on the stock market participation puzzle (Duraj *et al.* , 2024), which highlights that individuals helps overcome fixed participation costs with the help of family, friends, or financial advisors they trust.

For the effect of homophily, under social-learning model low-income agents do not participate, as a result homophily has a very limited effect. Figure C2 in the appendix show that social utility creates a strong effect of homophily, whereas social learning fails to generate much effect and can actually help to boost overall stock market participation through wealthier agents (P60). This occurs because wealthier agents benefit from having fewer connections with low-income (non-participating) agents and more connections with other wealthier agents who are more likely to participate and, therefore, generate informative signals.

As shown in Bursztyn *et al.* (2014), both channels are present in social influence. Thus to understand better the overall effect of social influence we need to investigate the combined effect of social utility and social learning channels on stock market participation, particularly when both channels are present and reinforce each other. To do this, in the model with, with both channels present we create two new datasets. In the first dataset, we run simulations while varying the parameter θ in increments of 100 within the interval $[2900, 4800]$, while adjusting the homophily parameter within the set $[0, 0.2, 0.4, 0.6, 0.8, 0.95]$. We simultaneously keep the learning parameter, r_h,constant at the calibrated value. In the second dataset, we fix θ at the calibrated level, and vary r_h in steps of 0.05 within the interval $[0, 1]$, and adjust homophily as before. The first and second datasets allow us to measure the sentitivity of the stock market participation level to increasing the impact of the social utility and social learning channels, respectively.

Using the generated data, we compute the elasticities of aggregate and income-group specific stock market participation levels to changes in θ and r_h at every level of homophily from simulations. Let $P_{j,i}$ denote stock market participation in group j, where j represents one of the five income quintiles, for a given value of the factor $factor_i$. We then compute the elasticity as:

$$\varepsilon_{j,i} = \frac{P_{j,i+1} - P_{j,i}}{factor_{i+1} - factor_i} \times \frac{factor_{i+1} + factor_i}{P_{j,i+1} + P_{j,i}}.$$

Next, we calculate the average elasticity across all observations for each group:

$$\bar{\varepsilon}_j = \frac{1}{m-1} \sum_{i=1}^{m-1} \varepsilon_{j,i},$$

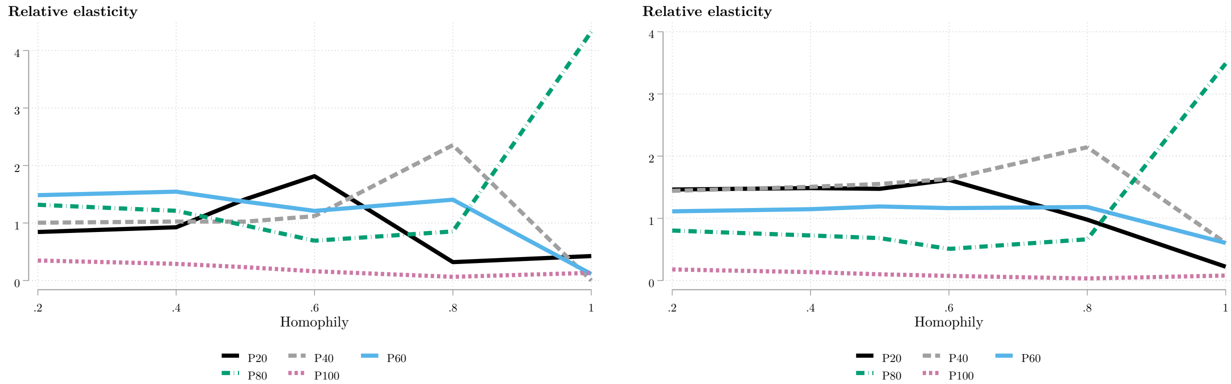
where m is the number of different levels for the corresponding factor.

In the final step, to facilitate comparisons across income groups, we normalize estimated income-group elasticities by dividing them on the average elasticity across groups:

$$\tilde{\varepsilon}_j = \frac{\bar{\varepsilon}_j}{\sum_{j=1}^{5} \bar{\varepsilon}_j}.$$

$\tilde{\varepsilon}_j > 1$ indicates that income group j is more responsive to changes in the corresponding factor than the average group. Conversely, $\tilde{\varepsilon}_j < 1$ suggests that the income group j is less responsive than the average. Figure 11 illustrates the results for relative elasticities across different levels of homophily for both factors, θ (representing social utility) and r_h (representing the social learning channel).

We observe that at lower levels of homophily, low-income groups benefit more when the social utility is more impactful. Increasing importance of social learning, however, benefits more medium- and high-medium income groups. The highest-income group experiences the smallest benefit overall in both scenarios, primarily due to the wealthiest agents already participating, regardless of their connections. For medium levels of homophily, low-income groups benefit the most from both channels. However, this is likely driven by spillover effects,



(a) Social Learning Channel

(b) Social Utility Channel

Figure 11. The relative elasticities for SMP across income groups

Notes: The figure plots the relative elasticities for stock market participation across different income groups with respect to θ and r_h at various levels of homophily. We set the ex-ante Gini coefficient to 0.4, the average connectivity to 7, θ to 3900 for the dataset where r_h varies, and r_h to 0.4 for the dataset where θ varies.

where the social utility channel amplifies the benefits for low-income groups. Essentially, the direct effect of higher informativeness primarily impacts higher-income groups, but as more individuals from high-income groups begin to participate, they also spread more social utility to other income groups. Finally, At very high levels of homophily, it is primarily the medium-high income group that benefits fully from social learning and utility.

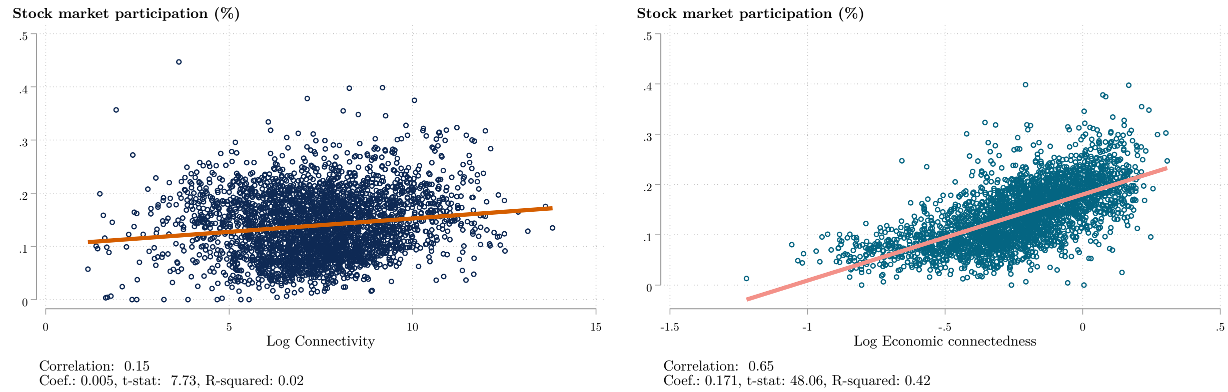
Including and analysing different channels for social influence shows that both channels are important with social learning alone not enough to explain the actual levels of stock market paticipation. In addition, our findings demonstrate that low-income groups benefit the most from social utility, especially at below high levels of homophily.

5. EMPIRICAL EVIDENCE

A key idea in the 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 now show that a general social connectivity measure, SCI, does not predict cross-sectional variation in stock market participation in U.S. counties (Bailey *et al.* , 2018). 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. 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 U.S. 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. Considering that 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, descriptive statistics are available in Table B1, and correlations between select variables are available in Table B2.

We examine the relationship between stock market participation and social connectivity in Figure 12. The figure reports scatter plots between connectivity measures and stock market participation. Panel a) plots the average SCI against stock market participation 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



(a) Log Connectivity index

(b) Log Economic Connectivity

Figure 12. 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.

four times higher than that between SCI and stock market participation. A one standard deviation increase in Economic Connectedness is associated with a 0.66 standard deviation increase in stock market participation.[35] 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. The estimated effect of Economic Connectedness is positive as well as statistically and economically significant even after we add control variables, state fixed effect, and primary industry of employment fixed effects in Column 1. The coefficient on the SCI index in Columns 2 is also positive and significant, but the economic significance is somewhat lower: a one standard deviation increase in the SCI index is associated with a 0.12 standard deviation increase in stock market participation. Columns 3 provide results where we include both Economic Connectedness and SCI. Both variables are now statistically significant and positive across specifications. We interpret these results in the following way. Holding Economic Connectedness fixed,

[35] Following Mitton (2022), we define economic significance as $E = \beta s_x/s_y$, where β is the coefficient of interest, s_x adn s_y denotes the standard deviation of the independent variable x and dependent variable y, respectively.

Table 1. Social Connectivity and Stock market participation

	Within-county			Non-local peers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log economic connectedness	0.0502***		0.0575***				
	(0.0080)		(0.0087)				
Log connectivity		0.00413***	0.00551***				
		(0.0011)	(0.0011)				
Log economic connectedness, non-local peers				0.284***			0.181*
				(0.061)			(0.10)
SMP of peers, non-local peers					1.050***		0.559
					(0.23)		(0.39)
Financial education of peers, non-local peers						0.929	0.404
						(0.62)	(0.72)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.14	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	0.06
Economic significance	0.19	0.12	0.22/0.17	0.19	0.17	0.00	–
Observations	2949	2949	2949	2949	2949	2949	2949
R-squared	0.813	0.808	0.818	0.810	0.810	0.805	0.811

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, the county's share of a bachelor-level education or above, county population, per capita number of schools in the county, dummy indicator for metropolitan areas, and the ratio of the mean income for the top 20% of earners divided by the mean income of the bottom 20% of earners in a county. in Fixed effects for the 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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. These results show that it is important to connect to individuals with information to share.

An empirical concern with the results is that there exists a variable correlated with both Economic Connectedness and stock market participation within the county. To partially address this concern, we instead examine the economic connectedness of *non-local friends*. We use the Facebook connectivity data to measure the connectivity of each county with all others counties in the United States, and calculate the weighted average of economic connectedness and financial education among non-local peers, similar to Cannon *et al.* (2024). We only select counties more than 250 miles away, to detach the measure from local economic conditions. The results are presented in Columns 4-6. First, economic connnectedness of non-local peers has a large impact on local stock market participation. The estimated economic significance (which takes the differences in variation between local and peer economic connectedness into account) is identical in Columns 1 and 4. We also examine two other variables that our model predicts should mediate the effect of connectivity. First, we add the stock market participation rate of non-local counties, which has a large and positive effect on partipation rates. In terms of economic significance, the impact is similar at 0.17 to Economic Connectedness. In column 3, we examine the effect of financial education in peer counties, showing that this has next to no relationship to stock market participation. While this may appear puzzling, we note that the correlations between Economic Connectedness and Financial education in peer counties has a correlation of 0.03 in the data. While some counties have a high share of financially educated households, they are not counties where rich and poor tend to interact. Finally, we show that only Economic Connectedness is significant if we include all variables in one regression, albeit only at 10% level.

Finally, we analyze the effects of economic connectedness and financial education across counties stratified by income level. Our model predicts that connectivity is important for all income groups, regardless of whether the primary benefit arises through the social utility or social learning channel. Which group benefits more depends largely on the overall level of connectivity. Simulation results indicate that at very low levels of connectivity, associated with low average stock market participation, high-income groups benefit the most from increased connectivity. In contrast, at moderate levels of connectivity, corresponding to moderate average stock market participation, all income groups begin to benefit from increases in connectivity. With respect to financial education, our model predicts that it would benefit the low-income group the most.

To examine whether the model's predictions for different income groups are supported by the data, we divide the sample of counties into three equally sized groups based on average county income. We then estimate the effect of economic connectedness with non-local peer counties on stock market participation within each income group separately. The results in Figure C3, panel (a), show that the magnitudes of the effect are similar across income groups. That is, the economic connectedness of peer counties has a comparable impact on stock market participation across the three income groups, with a slightly larger, though not statistically different, effect for high-income counties. This finding is consistent with Cannon *et al.* (2024), who document a similar within-county pattern. In contrast, Figure C3, panel (b), shows that financial education in peer counties, while generally positively associated with stock market participation, has a significantly stronger effect in lower-income counties compared to high-income counties, corroborating our model's predictions.

5.1 ROBUSTNESS

We now examine several plausible explanations for the correlation between Economic Connectedness and stock market participation. While we can plausibly rule out some channels, others are inherently more difficult to examine, and we stress again that it is difficult to establish causality. With this caveat in mind, it is reassuring that we find similar results to Cannon *et al.* (2024), who also study the link between Economic Connectedness and stock market participation. They use a quasi-experimental approach based on non-local friends' income changes to argue for a causal relationship. Cannon *et al.* (2024) also find that interactions with high-SES individuals facilitate stock market participation, consistent with our theoretical model.

Reverse causality – A plausible explanation for the correlation between Economic Connectedness and stock market participation is reverse causality, where stock market participation may lead to the formation of connections between individuals. For example, this effect could arise if greater stock market participation leads to interactions at shareholder meetings due to ownership in local stocks. However, we can rule out this concern by examining childhood Economic Connectedness. Since the childhood links are formed before the household starts investing, there should be no issue of reverse causality. The results are presented in Appendix Table B4, where we show that the coefficients on Childhood Economic Connectedness are similar in magnitude and economic significance to our main results in Table 1. Note that there can still be issues related to omitted variables, which could affect the participation rates of both parents and children in a given county. For example, parents connection could shape their children's connections through school choice, and could be correlated with an omitted variable.

Causal effects of place versus selection – A second empirical concern is selection into areas or residential sorting. Simply put, families living in areas with high Economic Connectedness may have higher stock market participation for some inherent reason, such as financial literacy or wealth, and not because of Economic Connectedness *per se*.

As noted by Chetty *et al.* (2022a), segregation by race or ethnicity is a prominent source of residential sorting in the United States. Areas with a larger share of the black population tend to have lower Economic Connectedness (Chetty *et al.* , 2022a). Data from the Survey of Consumer Finances also reveal that Black Americans are less likely to invest in stocks (Derenoncourt *et al.* , 2023). A simple way to assess whether residential sorting by race drives the correlation between stock market participation and Economic Connectedness is to condition the analysis on racial composition in counties. We provide such a test in Appendix Table B5, where we split the sample into counties with below or above median share of different ethnicities. The coefficients are somewhat lower in areas with above median Black, Hispanic, and Asian shares but remain statistically and economically significant.

The ideal experiment for ruling out selection would be to randomly allocate individuals across neighborhoods with low and high Economic Connectedness and examine their stock market participation later. This strategy is reminiscent of Haliassos *et al.* (2020), who use random allocation of immigrants in Sweden, and find strong effect of exposure to financially literate neighbors on financial behavior. In principle, a similar empirical strategy could be used to study selection versus the causal effect of place. We lack the data to do so, but note that the results in Haliassos *et al.* (2020) are consistent with our theoretical model – access to more informed peer raises stock market participation only for educated households, who are arguably closer to the participation threshold.

Connectedness versus other factors – Higher Economic Connectedness could also be associated with higher stock market participation because of other factors. For instance, areas with higher Economic Connectedness could have better school or schools that teach financial literacy. It is reassuring that the estimates on Economic Connnectedness are robust to controls for various variables that explain stock market participation, like income, the share of individuals working in the financial sector, the share of households with a bachelor degree, and age. Cannon *et al.* (2024) find that Economic Connectedness is a stronger predictor of stock market participation than cultural values such as cohesiveness or civic engagement. Added to this, we find that Economic Connectedness is more important than general sociability, as measured by the SCI. Furthermore, Cannon *et al.* (2024) find that non-local income shocks derived from friendship links affect stock market participation. These income shocks help address concerns over other factors as they are based on social links based on county-connectivity pairs (Bailey *et al.* , 2018), instead of local economic conditions.

6. CONCLUSION

We present a parsimonious 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 U.S. 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 also 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?

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. Policymakers should be aware of these negative side-effects of increased homophily, and may wish to consider targeted interventions in groups without access to informed peers.

Many of us have moved to different countries throughout our careers. 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 we have been able to learn 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.

REFERENCES

Abel, Andrew B. 1990. Asset prices under habit formation and catching up with the Joneses. *The American Economic Review*, **80**(2), 38–42.

Acemoglu, Daron, Ozdaglar, Asuman, & Yildiz, Ercan. 2011. Diffusion of innovations in social networks. *Pages 2329–2334 of: 2011 50th IEEE Conference on Decision and Control and European Control Conference.*

Allen, Franklin, & Gale, Douglas. 2000. Financial contagion. *Journal of Political Economy*, **108**(1), 1–33.

Amini, Hamed, Cont, Rama, & Minca, Andreea. 2016. Resilience to contagion in financial networks. *Mathematical finance*, **26**(2), 329–365.

Ammann, Manuel, & Schaub, Nic. 2021. Do individual investors trade on investment-related internet postings? *Management science*, **67**(9), 5679–5702.

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.

Bala, Venkatesh, & Goyal, Sanjeev. 2000. A Noncooperative Model of Network Formation. *Econometrica*, **68**(5), 1181–1229.

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. 2024. *Personal Financial Advice and Portfolio Quality*. SAFE Working Paper No. 353.

Behrman, Jere R, Mitchell, Olivia S, Soo, Cindy K, & Bravo, David. 2012. How financial literacy affects household wealth accumulation. *American Economic Review*, **102**(3), 300–304.

Bernard, Benjamin, Capponi, Agostino, & Stiglitz, Joseph E. 2022. Bail-ins and bailouts: Incentives, connectivity, and systemic stability. *Journal of Political Economy*, **130**(7), 1805–1859.

Bikchandani, Sushil, & Sharma, Sunil. 2000. Herd behavior in financial markets: a review. *International Monetary Fund, WP/00/48*.

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.

Broido, Anna D., & Clauset, Aaron. 2019. Scale-free networks are rare. *Nature Communications*, **10**, 1181–1229.

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, J. Y., & Viceira, L. M. 2002. *Appendix to 'Strategic Asset Allocation: Portfolio Choice for Long-Term Investors'*. Oxford University Press. Accessed: 2025-03-11.

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

Campbell, John Y, & Cochrane, John H. 1999. Vby force of habit: A consumption based explanation of aggregate stock market behaviorv. *Journal of Political Economy*, **107**(2), 205–251.

Cannon, Brad, Hirshleifer, David, & Thornton, Joshua. 2024. *Friends with benefits: Social capital and household financial behavior*. Tech. rept. National Bureau of Economic Research.

Changwony, Frederick K, Campbell, Kevin, & Tabner, Isaac T. 2014. Social engagement and stock market participation. *Review of Finance*, 317–366.

Chari, Varadarajan V, & Kehoe, Patrick J. 2004. Financial crises as herds: overturning the critiques. *Journal of Economic Theory*, **119**(1), 128–150.

Chen, Wei, Yuan, Yifei, & Zhang, Li. 2010. Scalable Influence Maximization in Social Networks under the Linear Threshold Model. *Pages 88–97 of: 2010 IEEE International Conference on Data Mining*.

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.

Cocco, Joao F, Gomes, Francisco J, & Maenhout, Pascal J. 2005. Consumption and portfolio choice over the life cycle. *The Review of Financial Studies*, **18**(2), 491–533.

Derenoncourt, Ellora, Kim, Chi Hyun, Kuhn, Moritz, & Schularick, Moritz. 2023. *Unemployment risk, portfolio choice, and the racial wealth gap*. ECONtribute.

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

Duraj, Kamila, Grunow, Daniela, Haliassos, Michael, Laudenbach, Christine, & Siegel, Stephan. 2024. Rethinking the Stock Market Participation Puzzle: A Qualitative Approach. *Available at SSRN.*

Elliott, Matthew, Golub, Benjamin, & Jackson, Matthew O. 2014. Financial networks and contagion. *American Economic Review*, **104**(10), 3115–3153.

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.

Gali, Jordi. 1994. Keeping up with the Joneses: Consumption externalities, portfolio choice, and asset prices. *Journal of Money, Credit and Banking*, **26**(1), 1–8.

Georgarakos, Dimitris, & Inderst, Roman. 2014. Financial advice and stock market participation. *Available at SSRN 1641302*.

Granovetter, Mark. 1978. Threshold Models of Collective Behavior. *American Journal of Sociology*, **83**(6), 1420–1443.

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.

Han, Bing, Hirshleifer, David, & Walden, Johan. 2022. Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, **57**(1), 390–412.

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

Heimer, Rawley Z. 2016. Peer pressure: Social interaction and the disposition effect. *The Review of Financial Studies*, **29**(11), 3177–3209.

Heimer, Rawley Z, & Simon, David. 2015. *Facebook finance: How social interaction propagates active investing.* Tech. rept.

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

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.

Karampourniotis, P. D., Sreenivasan, S., Szymanski, B. K., & Korniss, G. 2015. The Impact of Heterogeneous Thresholds on Social Contagion with Multiple Initiators. *PloS one*, **10**.

Karampourniotis, P. D., Szymanski, B. K., & Korniss, G. 2019. Influence Maximization for Fixed Heterogeneous Thresholds. *Scientific Reports*, **9**.

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

Kempe, David, Kleinberg, Jon, & Tardos, Éva. 2003. Maximizing the spread of influence through a social network. *Pages 137–146 of: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.*

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

Lelarge, Marc. 2012. Diffusion and cascading behavior in random networks. *Games and Economic Behavior*, **75**(2), 752–775.

Lim, Sonya, Lane, Jacqueline, & Uzzi, Brian. 2020. Biased information transmission in investor social networks: Evidence from professional traders. *Page 18198 of: Academy of Management Proceedings*, vol. 2020. Academy of Management Briarcliff Manor, NY 10510.

Lusardi, Annamaria, Mitchell, Olivia S, & Curto, Vilsa. 2010. Financial literacy among the young. *Journal of consumer affairs*, **44**(2), 358–380.

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.

Massenkoff, Maxim, & Wilmers, Nathan. 2023. Rubbing Shoulders: Class Segregation in Daily Activities. *Available at SSRN 4516850.*

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

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

Merton, Robert C. 1969. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 247–257.

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

Mitzenmacher, Michael. 2004. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics*, **2**.

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.

Morris, Stephen. 2000. Contagion. *The Review of Economic Studies*, **67**(1), 57–78.

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.

Peress, Joel. 2004. Wealth, information acquisition, and portfolio choice. *The Review of Financial Studies*, **17**(3), 879–914.

Peress, Joel. 2005. Information vs. entry costs: What explains us stock market evolution? *Journal of Financial and Quantitative Analysis*, **40**(3), 563–594.

Peress, Joël. 2003. Wealth, Information Acquisition, and Portfolio Choice. *The Review of Financial Studies*, **17**(3), 879–914.

Reardon, Sean F, & Bischoff, Kendra. 2011. Income inequality and income segregation. *American journal of sociology*, **116**(4), 1092–1153.

Samuelson, Paul A. 1975. Lifetime portfolio selection by dynamic stochastic programming. *Stochastic optimization models in finance*, 517–524.

Sheridan, Paul, & Onodera, Taku. 2018. A Preferential Attachment Paradox: How Preferential Attachment Combines with Growth to Produce Networks with Log-normal In-degree Distributions. *Scientific Reports*, **8**, 2811.

Smith, Keith Malcolm. 2021. Explaining the emergence of complex networks through log-normal fitness in a Euclidean node similarity space. *Scientific Reports*, **11**.

Taylor, Chris. 2011. Tight Budgets, Wild Markets Hurt Investment Clubs. *available at http*.

Thornton, Joshua. 2021. Peer Effects in Financial Expectations. *Available at SSRN 3803506*.

Van Rooij, Maarten, Lusardi, Annamaria, & Alessie, Rob. 2011. Financial literacy and stock market participation. *Journal of Financial Economics*, **101**(2), 449–472.

Vega-Redondo, Fernando. 2007. *Complex Social Networks*. Econometric Society Monographs. Cambridge University Press.

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.

Watts, Duncan J. 2002. A Simple Model of Global Cascades on Random Networks. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(9), 5766–5771.

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[36], where the average number of links is normalized to the largest number of connections for a Los Angeles County - Los Angeles County pair.[37]

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).[38] 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.

[36] 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.

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

[38] 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.

B. TABLES

Table B1. 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
Schools per capita	0.001	0.000	0.001	0.000	0.007
Income Inequality	13.043	12.409	3.325	6.065	47.101
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 B2. Correlation table

	SMP	Log EC	Log Con	Age	Inc.	Fin.Ed	Black	Female	Hisp.	Asian share	Schools p.c.	Ineq.	Pop
SMP	1												
Log EC	0.650***	1											
Log Con	0.150***	-0.178***	1										
Age	0.218***	0.151***	-0.456***	1									
Inc.	0.630***	0.586***	0.293***	-0.120***	1								
Fin.Ed	0.409***	0.223***	0.370***	-0.187***	0.318***	1							
Black	-0.326***	-0.586***	0.201***	-0.180***	-0.269***	-0.0245	1						
Female	0.111***	-0.0524**	0.289***	0.0125	0.0108	0.148***	0.105***	1					
Hisp.	-0.163***	-0.190***	0.153***	-0.339***	0.0540**	0.0545**	-0.102***	-0.130***	1				
Asian share	0.297***	0.139***	0.482***	-0.259***	0.423***	0.350***	0.0477**	0.0906***	0.181***	1			
Schools p.c.	0.0409*	0.260***	-0.611***	0.249***	-0.125***	-0.0869***	-0.236***	-0.138***	-0.0462*	-0.187***	1		
Ineq.	-0.177***	-0.402***	0.222***	-0.239***	-0.305***	0.166***	0.435***	0.104***	0.0975***	0.170***	-0.106***	1	
Pop	0.151***	0.00740	0.498***	-0.178***	0.260***	0.309***	0.0788***	0.107***	0.208***	0.564***	-0.164***	0.159***	1

Notes: The table displays pairwise correlations between variables. Variable abbrevation are as follows. SMP = Stock market participation. Log EC = Log Economic Connectedness. Log Con = Log Inside Connectivity. Age = median age in the county. Inc. = Log median household income in the county. Fin.Ed = the share of financially educated households. Black, Female, Hisp, and Asian denotes the share of the county population that belongs to each ethnicity. Schools pc = schools per county population. Pop. = county population. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B3. Model Parameters

Description	Parameter	Benchmark Value	Range
Wealth distribution	$F(W_i)$	Log-logistic(α, β)	
Relative risk aversion	γ	2	
Risk-free rate	r_f	0.025	
Average return on the risky asset	μ	0.08	
Expected volatility of the return (based on VIX)	σ_0^2	0.032	
Informative signal precision (based on VIX)	$1/\sigma_p^2$	1/0.032	
Size of economy	n	10,000	
Share financially educated		1.2%	
Gini coefficient		0.4	$[0.15, ..., 0.65]$
Average income		41,905	
Minimum wage		\$15,120	
Homophily parameter	h	0.5	$[0, ..., 0.95]$
Average connectivity	c	7	$[0, ..., 14]$

Notes: We choose the log-logistic wealth distribution because it is consistent with the data we use for the analysis. 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 minimum wage rate corresponds to the minimum wage per hour of \$ 7.5 multiplied by 8 working hours per day multiplied by 252 working days based on 2014 data.

Table B4. Childhood Economic Connectedness and stock market participation

	Economic connectedness		Connectivity index		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)
Log child economic connectedness	0.124***	0.0357***			0.122***	0.0358***
	(0.010)	(0.0064)			(0.011)	(0.0065)
Log connectivity			0.00594***	0.00259*	0.00181	0.00261**
			(0.0019)	(0.0013)	(0.0014)	(0.0012)
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	0.60	0.17	0.03	0.01	0.59 / 0.01	0.17 / 0.01
Observations	2671	2671	2671	2671	2671	2671
R-squared	0.358	0.824	0.023	0.820	0.360	0.825

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 the state and main industry of employment for the county are indicated. Economic significance is equal to $\beta s_x/s_y$, where β 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B5. Social Connectivity and Stock market participation by race and ethnicity

	White Share (1) Above	(2) Below	Black share (3) Above	(4) Below	Asian share (5) Above	(6) Below	Hispanic share (7) Above	(8) Below
Log economic connectedness	0.0741***	0.0330***	0.0303***	0.0728***	0.0357***	0.0507***	0.0399***	0.0623***
	(0.016)	(0.0091)	(0.0095)	(0.013)	(0.011)	(0.011)	(0.0091)	(0.012)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	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	0.28	0.13	0.12	0.28	0.14	0.19	0.15	0.24
Observations	1463	1482	1499	1441	702	2243	1504	1442
R-squared	0.762	0.859	0.886	0.753	0.901	0.785	0.856	0.805

Notes: The table provides results where we regress stock market participation at the county level against connectivity measures and controls depending on the racial composition of the county. Column *"Above"* describes the results for counties with *above-median* shares of White, Black, Asian, and Hispanic residents. Column *"Below"* describes the results for counties with *below-median* shares of White, Black, Asian, and Hispanic residents. 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 the 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C. FIGURES

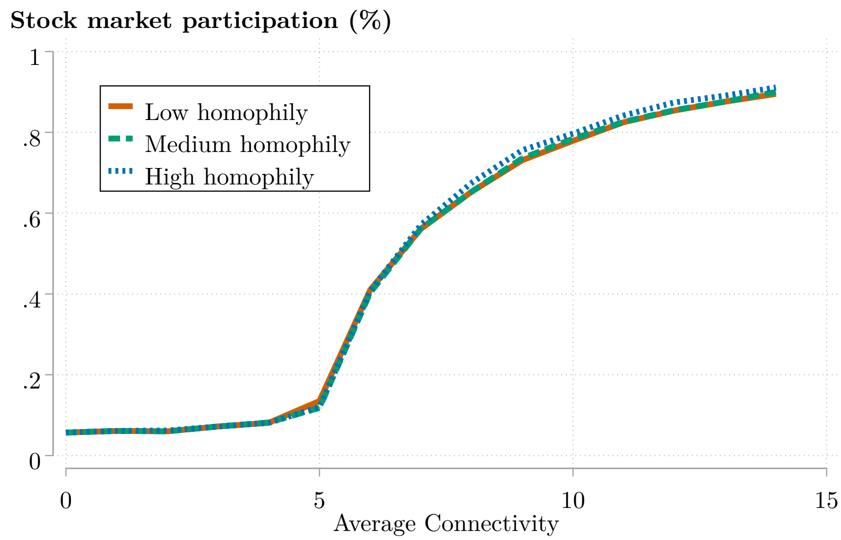


Figure C1. The effect of connectivity when homophily is unrelated to income

Notes: The figure plots stock market participation against connectivity. The ex-ante Gini coefficient is set to our baseline value of 0.4. Homophily is set to 0.1 for Low homophily (solid orange line) to 0.5 for Medium homophily (dashed green line) and to 0.9 for High homophily (dotted blue line).

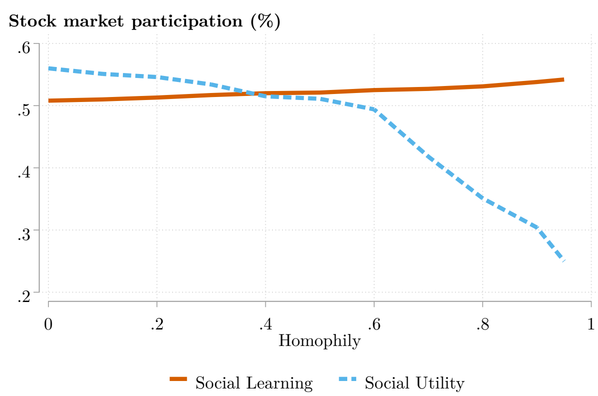
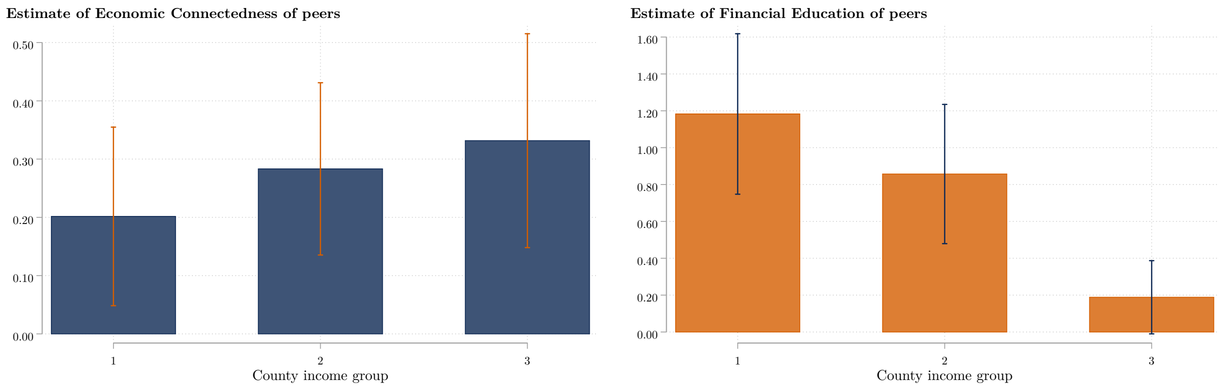


Figure C2. The effect of homophily on stock market participation with social utility and social learning

Notes: The figure plots stock market participation against homophily in the model with only Social Learning and the model with only Social Utility. Connectivity is set to 7 and the Gini coefficient is set to 0.4 in both models.



(a) Economic Connectedness

(b) Financial Education

Figure C3. Effect of Economic Connectedness of peers across income groups

Notes: The figure plots the coefficient on Economic Connectedness by county income groups. We use the median county income level to assign each county to one of five groups, and interact Economic Connectedness of non-local peers with the five income groups. The regressions control for the same variables as in Table 1, along with a dummies for each income group.

D. APPENDIX: SOLUTIONS AND PROOFS

D.1 INDIVIDUAL INVESTOR PROBLEM

Deriving Optimal Portfolio Composition Let's rewrite the expected utility maximization problem conditional on entry as follows:

$$\max_{\alpha_i} E\left[\frac{W_{1,i}^{1-\gamma}}{1-\gamma}\bigg|\mathscr{I}_i\right],$$

where $W_{1,i} = (W_{0,i} - F_i)e^{r_p}$.

Then,

$$E\left[\frac{((W_{0,i}-F_i)e^{r_p})^{1-\gamma}}{1-\gamma}\bigg|\mathscr{I}_i\right] = \frac{(W_{0,i}-F_i)^{1-\gamma}}{1-\gamma}E[e^{r_p(1-\gamma)}|\mathscr{I}_i] =$$
$$= \frac{(W_{0,i}-F_i)^{1-\gamma}}{1-\gamma}E[e^{(r_f+\alpha_i(r_a-r_f)+\frac{1}{2}\alpha_i(1-\alpha_i)\sigma^2)(1-\gamma)}|\mathscr{I}_i] =$$
$$= \frac{(W_{0,i}-F_i)^{1-\gamma}}{1-\gamma}e^{\left(r_f(1-\alpha_i)+\frac{1}{2}\alpha_i(1-\alpha_i)\sigma^2_{post,i}\right)(1-\gamma)}E[e^{\alpha_i r_a(1-\gamma)}|\mathscr{I}_i] =$$
$$= \frac{(W_{0,i}-F_i)^{1-\gamma}}{1-\gamma}e^{\left(r_f(1-\alpha_i)+\frac{1}{2}\alpha_i(1-\alpha_i)\sigma^2_{post,i}\right)(1-\gamma)}e^{(1-\gamma)\alpha_i\mu_{post,i}+\frac{1}{2}(1-\gamma)^2\alpha_i^2\sigma^2_{post_i}}$$

The expected utility maximization problem is now equivalent to

$$\max_{\alpha_i}\left(r_f(1-\alpha_i)+\frac{1}{2}\alpha_i(1-\alpha_i)\sigma^2_{post,i}\right)(1-\gamma)+(1-\gamma)\alpha_i\mu_{post,i}+\frac{1}{2}(1-\gamma)^2\alpha_i^2\sigma^2_{post_i}.$$

F.O.C.s:

$$-r_f+\frac{1}{2}\sigma^2_{post,i}-\alpha_i\sigma^2_{post_i}+\mu_{post,i}+(1-\gamma)\alpha_i\sigma^2_{post,i}=0\Rightarrow$$

$$\alpha_i^* = \frac{\mu_{post,i}-r_f+\frac{1}{2}\sigma^2_{post,i}}{\gamma\sigma^2_{post,i}}$$

Deriving the Distribution of the Posterior Mean We can compute σ_μ^2 using the law of total variance.

$$Var(\mu_{post,i}) = \mathbb{E}[Var(\mu_{post,i} \mid \mathscr{I}_i)] + Var\left(\mathbb{E}[\mu_{post,i} \mid r_a]\right),$$

where

$$Var(\mu_{post,i} \mid \mathscr{I}_i) = \sigma^2_{post,i} \Rightarrow \mathbb{E}[Var(\mu_{post,i} \mid \mathscr{I}_i)] = \sigma^2_{post,i}.$$

$$\mathbb{E}[\mu_{post,i} \mid r_a] = \mathbb{E}\left[\frac{\frac{\mu}{\sigma_0^2} + \frac{\sum_{i=1}^{k_i} s_i}{\sigma_p^2}}{\frac{1}{\sigma_0^2} + \frac{k_i}{\sigma_p^2} +}\middle| r_a\right] = \frac{\frac{\mu}{\sigma_0^2} + r_a\left(\frac{k_i}{\sigma_p^2}\right)}{\frac{1}{\sigma_0^2} + \frac{k_i}{\sigma_p^2}},$$

therefore, we can rewrite $\mathbb{E}[\mu_{post,i} \mid r_a]$ as follows

$$\mathbb{E}[\mu_{post,i} \mid r_a] = r_a \underbrace{\left(\frac{k_i}{\sigma_p^2}\right)\sigma^2_{post,i}}_{\equiv A} + \underbrace{\frac{\mu}{\sigma_0^2}\sigma^2_{post,i}}_{\equiv B}$$

Then,

$$Var\left(\mathbb{E}[\mu_{post,i} \mid r_a]\right) = Var(Ar_a + B) = A^2 Var(r_a) = A^2\sigma_0^2.$$

Summarizing the results,

$$\sigma^2_{\mu,i} = Var(\mu_{post,i}) = \sigma^2_{post,i} + \sigma_0^2(\sigma^2_{post,i})^2\left(\frac{k_i}{\sigma_p^2}\right)^2$$

Expected Utility at the Entry Stage We compute the expected utility of agent i, conditional on market entry but before receiving any signals, as follows

$$E\left[\frac{W_{1,i}^{1-\gamma}}{1-\gamma}\Big|k_i \geq 1, l_i, t_i, W_{0,i}\right] = E\left[\frac{(W_{0,i}-F_i)^{1-\gamma}}{1-\gamma} e^{\frac{(1-\gamma)\left(4\sigma^2_{post,i}\left(\mu_{post,i}+(2\gamma-1)r_f\right)+4(r_f-\mu_{post,i})^2+\left(\sigma^2_{post,i}\right)^2\right)}{8\gamma\sigma^2_{post,i}}}\right] =$$

$$= \frac{(W_{0,i}-F_i)^{1-\gamma}}{1-\gamma} \frac{e^{\frac{(1-\gamma)\left(\left(2\mu+\sigma^2_{post,i}\right)^2+4r_f^2-4r_f\left((1-2\gamma)\sigma^2_{post,i}+2\mu\right)+8(\gamma-1)r_f\sigma^2_{\mu,i}\right)}{8(\gamma-1)\sigma^2_{\mu,i}+8\gamma\sigma^2_{post,i}}}}{\sqrt{\sigma^2_{\mu,i}}\sqrt{\frac{\gamma-1}{\gamma\sigma^2_{post,i}}+\frac{1}{\sigma^2_{\mu,i}}}} \tag{10}$$

Notice that this expression is true only if $k_i \geq 1$. If the agent doesn't expect to receive any additional signal, then her posterior is simply equivalent to her prior $\mu_{post,i} = \mu$. The posterior variance $\sigma^2_{post,i}$ is simply the variance of the prior σ^2_0. In this case we can compute the expected utility conditional on not expecting to receive any signal as

$$E\left[\frac{W_{1,i}^{1-\gamma}}{1-\gamma}\Big|k_i = 0, l_i, t_i, W_{0,i}\right] = \frac{(W_{0,i}-F_i)^{1-\gamma}}{1-\gamma} e^{\frac{(1-\gamma)\left(4\sigma^2_0\left(\mu+(2\gamma-1)r_f\right)+4(r_f-\mu)^2+\left(\sigma^2_0\right)^2\right)}{8\gamma\sigma^2_0}}$$

Properties of the Participation Threshold Set Here we discuss properties of the *Participation Threshold Set*.

First, we show that for any agent i with $W_{0,i} > 0$ and type t_i, the set is non-empty.

If $t_i = 1$, meaning agent i is financially educated, then by definition, her Participation Threshold Set is $\hat{S}_i = \{(0,0)\}$.

Now, consider $t_i = 1$ and suppose $f_i(W_{0,i}, t_i = 1, (0,0)) < 0$. Consider the sequence $\{(k_i, 0)\}_{k_i \in \mathbb{N}}$. Since f_i is monotonically increasing in k_i and $\lim_{k_i \to \infty} \sigma^2_{post,i} = 0$, for any positive θ and signal precision σ^2_p, we can always find $\tilde{k}_i$ large enough such that $f_i(W_i, t_i = 1, (\tilde{k}_i, 0)) > 0$. Therefore, there exists $\hat{k}_i$ such that:

$$f_i(W_{0,i}, t_i = 1, (\hat{k}_i, 0)) \geq 0, \quad \text{and} \quad f_i(W_{0,i}, t_i = 1, (\hat{k}_i - 1, 0)) < 0.$$

Thus, $\hat{S}_i$ contains at least one element $(\hat{k}_i, 0)$, and, therefore, it is not empty.

Next, we show that for any two elements $(\hat{k}_i, \hat{l}_i), (\hat{k}'_i, \hat{l}'_i) \in \hat{S}_i$, it must hold that either $\hat{k}_i > \hat{k}'_i$ and $\hat{l}_i < \hat{l}'_i$, or $\hat{k}_i < \hat{k}'_i$ and $\hat{l}_i > \hat{l}'_i$.

Towards contradiction, suppose that there exist $(k_i, l_i), (k'_i, l'_i) \in \hat{S}_i$ such that $k_i > k'_i$ but $l_i \geq l'_i$. By the definition of $\hat{S}_i$, if $(k_i, l_i) \in \hat{S}_i$, then for any $(k, l) \neq (k_i, l_i)$ with $k \leq k_i$ and $l \leq l_i$, we must have $(k, l) \notin \hat{S}_i$, contradicting the assumption that $(k'_i, l'_i) \in \hat{S}_i$. A similar argument holds for the case $l_i > l'_i$ and $k_i \geq k'_i$.

D.2 NON-COOPERATIVE EQUILIBRIUM

First, we need to define Algorithm 1 which constructs a non-cooperative equilibrium.

Algorithm 1. *Non-cooperative equilibrium with one-directional influence propogation*

Step 1 *Initialize the set of participants* $P \leftarrow \emptyset$.

Step 2 *Initialize the adjacency matrix* $X_0 \leftarrow \mathbf{0}_{n\times n}$ *where each element* $x^0_{ij} = 0$.

Step 3 *Set the iteration counter* $t \leftarrow 0$.

Step 4 ***Repeat*** *the following steps until equilibrium is reached:*

- *For each node* $i \in \{1, ..., n\}$*:*
  - ***If*** *there exists* $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ *such that* $\hat{k}_i + \hat{l}_i \leq \sum_{j\in N} x^t_{ji}$ *and* $\hat{k}_i \leq \sum_{j\in N} sp_j \times x^t_{ji}$ *(i.e., node* i *has met its participation threshold,* $f(W_i, t_i, k_i(X_t), l_i(X_t)) \geq 0$*):*
    - *Increment the iteration counter:* $t \leftarrow t + 1$.
    - *Update the participant set:* $P \leftarrow \{i\} \cup P$.

- *Update the adjacency matrix X_t as follows:*

$$x_{hj}^t := \begin{cases} 1 & \textit{if } h = i,\ g_{hj} = 1, \textit{ and } j \notin P \\ x_{hj}^{t-1} & \textit{otherwise} \end{cases}$$

Here, for each node $h \in N$ and $j \in N$:

· *Set $x_{hj}^t = 1$ if node h is the newly added participant i, there exists a link $g_{hj} = 1$, and node j is not already in the participant set P.*

· *Otherwise, retain the previous value x_{hj}^{t-1} from the previous iteration.*

- ***If** the set of participants P has not changed after evaluating all nodes:*
  - ***Exit** the loop, as an equilibrium has been reached.*

Step 5 *The final adjacency matrix is denoted as $X := X_t$.*

We label all intermediate steps of the adjacency matrix X as X_t, where each element $x_{ij}^t \in X_t$. This notation is used for convenience in the proof of Proposition 1. The final set of participants, P, is defined such that if $i \in P$, then $\sigma_i = 1$, and if $j \in N \setminus P$, then $\sigma_j = 0$.

Proof of Proposition 1. ***Existence.*** We first prove the existence of a non-cooperative equilibrium. The proof is constructive, using Algorithm 1. To show that this algorithm identifies a non-cooperative equilibrium, we verify that it determines a set of participating agents P and an SMP information diffusion matrix X that satisfy all the necessary properties.

First, by definition, the algorithm adds a link from agent i to agent j only if they are connected according to matrix G, so $x_{ij} = 1$ only if $g_{ij} = 1$.

Second, the algorithm adds an agent i to the set of participants P at iteration t only if there exists $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ such that the number of incoming links for agent i, as defined by X_t, is at least $\hat{k}_i + \hat{l}_i$, and the number of links generating informative signals is at least $\hat{k}_i$.

Thus, for all $i \in P$, there exist $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ such that $\sum_{j \in N} sp_j x_{ji} \geq \hat{k}_i$ and $\sum_{j \in N} x_{ji} \geq \hat{k}_i + \hat{l}_i$, and by definition of the set P, $\forall i \in P : \sigma_i = 1$.

Third, for all $i \notin P$, at least one of the following must hold for all $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$: $\sum_{j \in N} x_{ji} < \hat{k}_i + \hat{l}_i$ or $\sum_{j \in N} sp_j x_{ji} < \hat{k}_i$. By construction, if at iteration t there exists a node $i \in N \setminus P$ such that $\sum_{j \in N} sp_j x^t_{ji} \geq \hat{k}_i$ and $\sum_{j \in N} x^t_{ji} \geq \hat{k}_i + \hat{l}_i$, then Algorithm 1 must add node i to the set P at some iteration $t' \geq t$.

Fourth, the algorithm adds an outgoing link for node i only if node i is included in the set P. Therefore, if $j \notin P$, it does not have any outgoing links. Once a node i is added to the set of participants, all links from i to its connected peers in G are formed by construction.

Finally, we demonstrate that matrix X is acyclic. Towards contradiction, suppose there exists a cycle involving nodes i and i' such that there is a direct link from i' to i. Notice that, if $g_{ij} = g_{ji} = 1$ and j is added to the set of participants after i, then $x_{ij} = 1$ and $x_{ji} = 0$. Thus, if there is a path from i to i' according to adjacency matrix X, then i' must be added to P after i. However, if there is a direct link from i' to i, it implies that i' is added to the set of participants before i. This contradiction shows that the resulting graph characterized by X must be acyclic.

Therefore, we have proven that Algorithm 1 finds a non-cooperative equilibrium for any network $\{N, G, SP, \hat{S}\}$.

Uniqueness. Next, we prove the uniqueness of the equilibrium. Suppose there exists a non-cooperative equilibrium characterized by the strategy profile σ and participation set P, and there exists an acyclic influence propagation matrix X consistent with this equilibiurm strategy profile.

We define the following sequence of sets: the set $S_0 = \{i \in N \mid \hat{S}_i = \{0, 0\}\}$ contains all agents with zero participation thresholds. These agents must be participants in any equilibrium, so $\forall i \in S_0, i \in P$. Now let's construct a set S_1 such that

$$S_1 = \left\{ i \in N \setminus S_0 \mid \exists(\hat{k}_i, \hat{l}_i) \in \hat{S}_i \text{ such that } \sum_{j \in S_0} g_{ij} \geq \hat{k}_i + \hat{l}_i \text{ and } \sum_{j \in S_0} sp_j g_{ij} \geq \hat{k}_i \right\}.$$

That is, S_1 consists of agents who can satisfy their participation constraints solely through connections to agents in S_0.

Notice that any agent in set S_1 should participate in the stock market in equilibrium. Towards contradiction, suppose $i \in S_1$ and $\sigma_i = 0$. By definition of S_1, there exists $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ such that $\sum\limits_{j \in S_0} g_{ij} \geq \hat{k}_i + \hat{l}_i$ and $\sum\limits_{j \in S_0} sp_j g_{ij} \geq \hat{k}_i$. Now, if $\sigma_i = 0$, there must exist some agent $j \in S_0$ such that $g_{ij} = 1$ and $x_{ji} = 0$. However, since $\sigma_j = 1$, the constraint $x_{ij} + x_{ji} = 1$ must hold. This implies $x_{ij} = 1$. But if $x_{ij} = 1$, then it follows that $\sigma_i = 1$. So, we get a contradiction.

Now suppose that $\forall i \in \bigcup\limits_{h=0}^{t} S_h$, we have $\sigma_i = 1$. We define

$$S_{t+1} = \left\{ i \in N \setminus \bigcup_{h=0}^{t} S_h \mid \exists(\hat{k}_i, \hat{l}_i) \in \hat{S}_i \text{ such that } \sum_{j \in \bigcup\limits_{h=0}^{t} S_h} g_{ij} \geq \hat{k}_i + \hat{l}_i \text{ and } \sum_{j \in \bigcup\limits_{h=0}^{t} S_h} sp_j g_{ij} \geq \hat{k}_i \right\}.$$

Then, if $i \in S_{t+1}$, it must be that $i \in P$. Towards contradiction, suppose $\exists i \in S_{t+1}$ and $i \notin P$. Therefore, for any $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ it must be that $\sum_{j \in \bigcup\limits_{h=0}^{t} S_h} x_{ji} < \hat{k}_i + \hat{l}_i$ or $\sum_{j \in \bigcup\limits_{h=0}^{t} S_h} sp_j x_{ji} < \hat{k}_i$. At the same time, from definition of S_{t+1}, there exists $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ such that $\sum_{j \in \bigcup\limits_{h=0}^{t} S_h} g_{ij} \geq \hat{k}_i + \hat{l}_i$ and $\sum_{j \in \bigcup\limits_{h=0}^{t} S_h} sp_j g_{ij} \geq \hat{k}_i$. Hence, there exist $j \in \bigcup\limits_{h=0}^{t} S_h$ such that $g_{ij} = g_{ji} = 1$, but $x_{ji} = 0$. If $i \notin P$, then $\sigma_i = 0$, therefore it must also be that $x_{ij} = 0$ according to the definition of non-cooperative equilibrium. However, the definition also requires that if j is a participant and $g_{ij} = g_{ji} = 1$ then $x_{ij} + x_{ji} = 1$. Therefore, we get a contradiction, and if $i \in S_{t+1}$ then it must be that $i \in P$.

Since the number of agents is finite, there exists a $t' < \infty$ such that $S_{t'} = \emptyset$, which means that $\forall t > t', S_t = \emptyset$. Let T be such that $S_T \neq \emptyset$ and $S_{T+1} = \emptyset$. By the earlier argument, we have by induction that $\bigcup_{h=0}^{T} S_h \subseteq P$.

Next, we show that for any $i \in P$ it must be that $i \in \bigcup_{h=0}^{T} S_h$. Again, we prove this by contradiction. If $i \notin \bigcup_{h=0}^{T} S_h$, it means for any $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ either $\sum_{j \in \bigcup_{h=0}^{T} S_h} g_{ij} < \hat{k}_i + \hat{l}_i$ or $\sum_{j \in \bigcup_{h=0}^{T} S_h} sp_j g_{ij} < \hat{k}_i$. Therefore, if $i \in P$, it must be that, according to the corresponding influence propagation matrix X there exists some node i' such that $i' \in N \setminus \bigcup_{h=0}^{T} S_h$ and $x_{i'i} = 1$, which means $\sigma_{i'} = 1$. i' belonging to $N \setminus \bigcup_{h=0}^{T} S_h$ implies that for any $(\hat{k}_{i'}, \hat{l}_{i'}) \in \hat{S}_{i'}$ either $\sum_{j \in \bigcup_{h=0}^{T} S_h} g_{i'j} < \hat{k}_{i'} + \hat{l}_{i'}$ or $\sum_{j \in \bigcup_{h=0}^{T} S_h} sp_j g_{i'j} < \hat{k}_{i'}$. This means we must find another node $i'' \in N \setminus \bigcup_{h=0}^{T} S_h$ such that $x_{i''i'} = 1$ and $\sigma_{i''} = 1$. Repeating this argument for successive nodes, we either encounter a node that violates the participation conditions for the SMP (i.e., it has an insufficient number of participating peers) or we find a cycle, which contradicts the requirement that the matrix X be acyclic. Thus, we reach a contradiction, and it follows that $P = \bigcup_{h=0}^{T} S_h$. Therefore, the non-cooperative equilibrium is unique.

□

D.3 IDENTIFYING MATRIX X: EXAMPLE

Let us consider an example. Suppose there are three agents, each connected to the others. For simplicity, assume that any participating agent also generates an informative signal. Consequently, in any equilibrium, it must hold that $p_i = k_i$, $l_i = 0$. This allows us to use p_i as the relevant participation threshold instead of considering combinations of two variables to illustrate the idea. Now, assume agent 1 has a participation threshold of 0, while agents 2 and 3 have participation thresholds equal to 1. In this case, the only equilibrium satisfying the properties of the influence propagation process is the one where all three agents participate, and it is also a non-cooperative equilibrium. However, the matrix X may have different configurations. For instance, it could contain the links $x_{1,2} = x_{1,3} = x_{2,3} = 1$ and $x_{2,1} = x_{3,1} = x_{3,2} = 0$, or alternatively, it could have $x_{1,2} = x_{1,3} = x_{3,2} = 1$ and $x_{2,1} = x_{3,1} = x_{2,3} = 0$ configuration.

E. APPENDIX: MAXIMAL SOCIAL COOPERATION

Similar to a non-cooperative equilibrium, we can construct an equilibrium with maximum participation, assuming two-way influence propagation. This means that social influence and information can be transmitted through the same link in both directions. Intuitively, this allows agents to strategically collaborate in both groups and pairs to learn and exchange different pieces of information and influence each other's decision to participate. As before, we assume that agents participating in the stock market share information with their peers, and any agent who is influenced by participating peers experiences a reduction in the SMP participation cost through the social utility channel.

To accommodate two-way influence propagation, we need to slightly redefine the influence propagation process compared to the previous analysis. For convenience, we will characterize it by the matrix $\hat{X}$, similar to the earlier approach. However, there will be a few important changes in how we define the matrix $\hat{X}$ compared to the one-way influence propagation game.

First, the link in matrix $\hat{X}$ between agents i and j indicates that information and influence flow in both directions – from i to j and from j to i – thereby reducing the SMP participation cost and improving signal precision for both agents, and that both agents participate in the stock market. Consequently, if agent i participates and is connected to agent i', who does not participate, we say that $\hat{x}_{ii'} = \hat{x}_{i'i} = 0$. It is important to note that the absence of a link no longer implies that there is no actual information or influence propagation from i to i'; rather, it simply means that this influence propagation does not lead to agent i' participating. Nonetheless, we maintain the assumption that if agent i participates, it affects its peer i' decision to participate by reducesing the SMP cost for agent i' and, potentially, improving stock return signal precision. This represents a key difference from the definition of the influence propagation matrix used in the case of one-way influence propagation. While this difference is technical, it allows us to account for two-way influence propagation when proving the key results. Therefore, matrices X and $\hat{X}$ should not be directly compared, as they represent different concepts. Matrix X captures the actual influence propagation between agents in a one-way influence propagation game, while $\hat{X}$ captures only the links that lead to two-way influence and information flow between participating agents.

Therefore, we can define the influence propagation process for the two-way influence propagation game as follows:

Definition 3. ***Two-Way SMP Influence Propagation Process:*** *This process describes how agents in a network influence each others' decision regarding stock market participation. It is represented by the adjacency matrix $\hat{X}$, where $\hat{x}_{ij} \in \hat{X}$ equals 1 if both agents i and j participate ($\sigma_i = \sigma_j = 1$) and exert influence on each others' decision to participate through social utility and social learning channels; otherwise, it equals 0. The SMP influence propagation process is characterized by the following properties:*

1. *Agents i and j can only exert influence on each other if they are connected in the network $\{N, G, SP, \hat{S}\}$, which means that $g_{ij} = 1$.*

2. *Influence propagates in both directions between any two connected agents who participate in the stock market. Specifically, for any agents i and j, $\hat{x}_{ij} = \hat{x}_{ji} = 1$ if and only if $g_{ij} = g_{ji} = 1$ and $\sigma_i = \sigma_j = 1$. Otherwise, $\hat{x}_{ij} = \hat{x}_{ji} = 0$.*

3. *The stock market participation cost for agent i decreases in proportion to the number of its participating peers, calculated as $\sum_{j \in N} g_{ij}\sigma_j$; and signal precision depends on number of participating peers who generate informative signals, calculated as $\sum_{j \in N} sp_j g_{ij}\sigma_j$. This cost reduction and increase in signal precision are not determined by the matrix $\hat{X}$ which accounts only for reciprocal links.*

Now, we can define a Nash Equilibrium strategy profile in the game with two-way influence propagation game:

Definition 4. *The strategy profile $\hat{\sigma}$ is a Nash Equilibrium profile in two-way influence* *propagation game if and only if for each $i \in N$, $\sigma_i = 1$ if and only if there exists $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ such that $\sum_{j \in \mathbf{N}} g_{ji}\sigma_j \geq \hat{k}_i + \hat{l}_i$ and $\sum_{j \in \mathbf{N}} sp_j g_{ji}\sigma_j \geq \hat{k}_i$; $\sigma_i = 0$ otherwise.*

Notice that the definition of a Nash equilibrium is not reliant on the definition of the matrix $\hat{X}$.

Given that we now allow for two-way influence propagation and all possible cooperation among groups of nodes, this can lead to multiple equilibria. Referring back to the previous example in Figure 1, we observe that in the two-way influence propagation game, both sets of participating agents represented in Figures 1b and 1c can be supported as equilibrium outcomes in the two-way influence propagation game. Therefore, we seek to identify the most beneficial social outcome that maximizes the number of participating agents through maximum cooperation, which corresponds to Case 2 in Figure 1c. We demonstrate that the following algorithm identifies this equilibrium profile and the corresponding matrix $\hat{X}$ for a given network structure $\{N, G, SP, \hat{S}\}$ in the two-way influence propagation game with social cooperation:

Algorithm 2. *Maximum Benefit Cooperative Equilibrium with Two-Way Influence Propagation*

Step 1 *Initialize the vector of potential participants $P \leftarrow N$.*

Step 2 *Initialize the iteration counter $t \leftarrow 0$ and the adjacency matrix $\hat{X}_0 \leftarrow G$.*

Step 3 ***Repeat*** *the following steps until equilibrium is reached:*

- *For each node $i \in \{1, \ldots, n\}$:*
  - ***If*** *for all $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ either $\hat{k}_i + \hat{l}_i > \sum\limits_{j \in N} \hat{x}^t_{ji}$ or $\hat{k}_i > \sum\limits_{j \in N} sp_j \hat{x}^t_{ji}$ (i.e., node i doesn't meet its participation threshold):*
    - *Increment the iteration counter: $t \leftarrow t + 1$.*
    - *Update the vector of potential participants by excluding this node: $P \leftarrow$*

$P \setminus \{i\}$.

∗ *Update the adjacency matrix $\hat{X}_t$ as follows:*

$$\hat{x}^t_{hj} := \begin{cases} 0, & \text{if } h = i \\ \hat{x}^{t-1}_{hj}, & \text{otherwise} \end{cases}$$

$$\hat{x}^t_{jh} := \hat{x}^t_{hj}$$

Here, for each node $h \in N$ and $j \in N$:

· *Set $\hat{x}^t_{hj} = 0$ and $\hat{x}^t_{jh} = 0$ if node h is the newly removed participant i, meaning all links from i must be removed.*

· *Otherwise, retain the previous value $\hat{x}^{t-1}_{hj}$ from the previous iteration.*

- ***If** the set of potential participants P has not changed after evaluating all nodes:*
  - ***Exit** the loop, as an equilibrium has been reached, and the final set of participants P has been found.*
- *The final adjacency matrix is denoted as $\hat{X} := \hat{X}_t$.*

Proposition 2. *For any given network structure $\{N, G, SP, \hat{S}\}$, Algorithm 2 finds a Nash Equilibrium profile characterized by the set of participants P and information flow matrix $\hat{X}$ such that if there exists another Nash Equilibrium profile with the set of participating agents $P' \neq P$, then $P' \subset P$.*

Proof. We first demonstrate that Algorithm 2 returns the equilibrium set of participating agents according to the equilibrium notion defined in Definition 4. Initially, the set of potential participating agents includes all nodes, and the matrix $\hat{X}_0$ is equivalent to the adjacency matrix G. Let $P_0 = N$ represent the initial set of potential participants, and let P_t denote the set of potentially participating agents returned by Algorithm 2 at step t. P

represents the final set of participating agents when the algorithm terminates.

At each step t, an agent i is removed from the set of potential participants P_t if and only if for all $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ either $\hat{k}_i + \hat{l}_i > \sum_{j \in N} \hat{x}^t_{ji}$ or $\hat{k}_i > \sum_{j \in N} sp_j \hat{x}^t_{ji}$. So, the participation threshold cannot be met even if all remainig peers according to the current matrix $\hat{X}_0$ exert influence on player i. Once removed, the matrix $\hat{X}_t$ is updated by deleting all links connecting agent i to its neighbors. Consequently, if the participation threshold condition is not met at step t, it can't be met at any step t' where $t' > t$. Thus, when the algorithm converges in finite time T, it holds for any $i \in N \setminus P$ that for all $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ either $\hat{k}_i + \hat{l}_i > \sum_{j \in N} \hat{x}^t_{ji}$ or $\hat{k}_i > \sum_{j \in N} sp_j \hat{x}^t_{ji}$. At the same time, for every $i \in P$, it must be that there exists $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ such that $\hat{k}_i + \hat{l}_i \leq \sum_{j \in N} \hat{x}^t_{ji}$ and $\hat{k}_i \leq \sum_{j \in N} sp_j \hat{x}^t_{ji}$, as this is the stopping criterion of the algorithm. Therefore, the algorithm indeed returns the equilibrium set of participating agents P, where $\hat{\sigma}_i = 1$ if $i \in P$ and $\hat{\sigma}_i = 0$ if $i \in N \setminus P$.

Next, we prove that if there exists another Nash equilibrium with a set of participating agents $P' \neq P$, then $P' \subset P$. We will show this by induction.

Initially, $P_0 = N$ and $\hat{X}_0 = G$. If agent i is excluded from the set of potential participants, so that $P_1 = P_0 \setminus \{i\}$, it follows that for any $(\hat{k}_i, \hat{l}_i) \in \hat{S}_i$ either $\hat{k}_i + \hat{l}_i > \sum_{j \in N} g_{ji}$ or $\hat{k}_i > \sum_{j \in N} sp_j g_{ji}$. Therefore, agent i cannot be a participant in any equilibrium, implying that $i \notin P'$. Moreover, if agent i does not participate in any equilibrium, it will not exert influence on her peers in any equilibrium. Hence, all links connecting agent i to her peers can be removed from matrix G without affecting the equilibrium outcomes of the game. This allows us to replace the adjacency matrix G with $\hat{X}_1$, where:

$$\hat{x}^1_{hj} := \begin{cases} 0, & \text{if } h = i \\ g_{hj}, & \text{otherwise} \end{cases}$$

and

$$\hat{x}^1_{jh} := \hat{x}^1_{hj}.$$

Now, suppose that at step t, the set $N \setminus P^t$ contains all nodes that never become participants in any equilibrium. The matrix $\hat{X}_t$ is the reduced version of matrix G, with all links connecting nodes in $N \setminus P^t$ removed. By the same reasoning, if agent i is removed at step $t+1$, she cannot be a participant in any equilibrium. Therefore, all links connecting agent i to its peers can be deleted, as no information will flow through these links in any equilibrium under the two-way information flow process. The adjacency matrix can then be updated to $\hat{X}_{t+1}$ as follows, without altering the equilibrium outcomes:

$$\hat{x}^{t+1}_{hj} := \begin{cases} 0, & \text{if } h = i \\ \hat{x}^t_{hj}, & \text{otherwise} \end{cases}$$

and

$$\hat{x}^{t+1}_{jh} := \hat{x}^{t+1}_{hj}.$$

Thus, we have shown that any node excluded from the set of participating agents by Algorithm 2 cannot be a participant in any other equilibrium. Therefore, if there exists another equilibrium characterized by the set of participating agents $P' \neq P$ then $P' \subset P$.

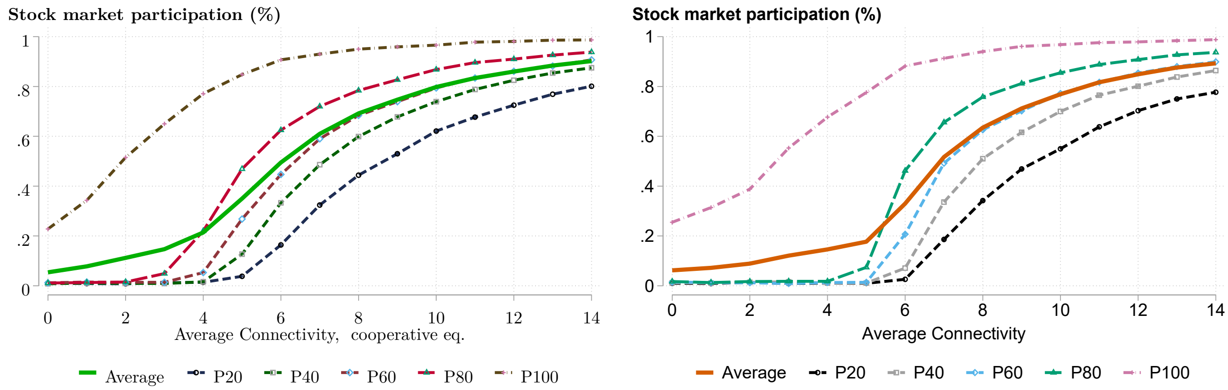
$\square$

It is also straightforward to show that for a fixed network structure $\{N, G, SP, \hat{S}\}$, the set of participants in a non-cooperative equilibrium of the one-way information flow game can be supported as an equilibrium outcome in the two-way information flow game. This equilibrium outcome corresponds to the scenario with the minimum possible number of participating agents among all equilibria of the two-way information flow game.

E.1 COMPARING COOPERATIVE AND NON-COOPERATIVE EQUILIBRIA

We now compare the outcomes for the cooperative and non-cooperative equilibria. We begin by comparing the effect of connectivity in Figure E1. The left panel a) provides the results from the cooperative equilibrium panel b) on the right provides the results for the non-cooperative equilibrium. Panel b) is the same figure as 3, to ease comparison. Overall, we see that the effect of connectivity is very similar across the two figures. The average SMP in both figures exhibits the same S-shape, where the marginal effect of adding one more connection is small for low levels of connectivity. The marginal effect of adding one more connection increases rapidly as we move to the right in the figures, but slows down again as the average number of connections become sufficiently high. The cooperative equilibrium exhibits a slightly higher level of SMP for a given level of connectivity, but exhibits a similar marginal effect of adding one more connection. The marginal effect of connectivity for different income groups is also similar. The steep rise in the effect of connectivity naturally starts at earlier levels of average connectivity in the cooperative equilibrium, due to the more efficient information sharing between cooperating agents. However, once information sharing has started to play a role, the S-shape pattern is similar for a given income group in the two different figures.

We compare the simulated SMP in the cooperative and non-cooperative equilibria for all permutations of the model in Figure E2. Overall, we have eleven different values for both Gini and Homophily, and 15 values for Connectivity, giving us $11 * 11 * 14 = 1,815$ different permutations for both the cooperative and non-cooperative equilibria. The correlation coefficient is 0.9344 between the level of stock market participation in the cooperative and non-cooperative equilibria. The below figure shows a close correspondence between the cooperative and non-cooperative equilibrium at both low and high average levels of stock market participation. The participation rates in the cooperative equilibrium is consistently higher than for the non-cooperative equilibria in the middle of the graph. This is precisely



(a) Cooperative equilibrium

(b) Non-cooperative equilibrium (baseline)

Figure E1. Stock market participation and connectivity in a cooperative equilibrium

Notes: The figure in both panels plots stock market participation (y-axis) against the average number of connections (x-axis). Panel a) describes a cooperative equilibrium with the green solid line plotting the average stock market participation among all agents; the dark navy, dark green, maroon, cranberry, and olive dashed lines plotting average participation among agents in below 20^{th}, between 20^{th}-40^{th}, between 40^{th}-60^{th}, between 60^{th}-80^{th}, and above 80^{th} percentile of income distribution, respectively. Panel b) describes a non-cooperative equilibrium with the orange solid line plotting the average stock market participation among all agents; the black, gray, blue, green, and light purple dashed lines plotting average participation among agents in below 20_{th}, between 20^{th}-40^{th}, between 40^{th}-60^{th}, between 60^{th}-80^{th}, and above 80^{th} percentile of income distribution, respectively. We set the homophily parameter to 0.5 and the ex-ante Gini coefficient to 0.4 for the simulations.

the region where connectivity and homophily have the largest impact in both models, but the cooperative equilibrium generates more efficient information sharing and so leads to higher stock market partipation rates. We can see this by examining the colors of the dots, where we see that the difference in participation rates between the cooperative and non-cooperative equilibria occurs when connectivity is low.

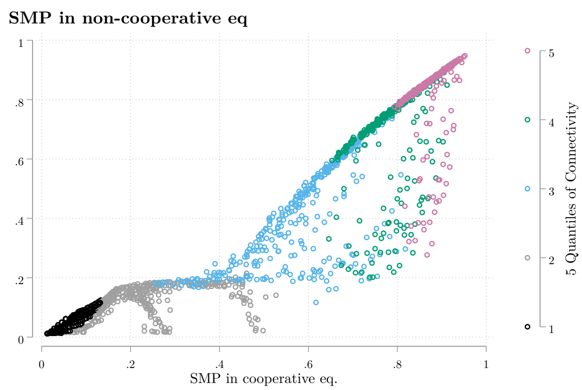


Figure E2. Stock market participation in cooperative and non-cooperative equilibria

Notes: The figure plots stock market participation in the baseline non-cooperative equlibrium (y-axis) against stock market participation in the cooperative equlibrium. The color of the dots indicates the level of connectivity, whose values are noted on the right-hand side of the figure.