Saskia van Rheenen

The association between social capital and stock market participation

Abstract

In this paper, I investigate the relationship between social capital and stock market participation. Using privacy-protected data on 21 billion Facebook friendships, I find that, on average, counties with high social connectedness – people with low socioeconomic status are highly connected with people with high socioeconomic status – have higher stock market participation. Also, social connectedness matters even after controlling for income measures, inequality measures and other determinants of stock market participation. Other measures of social capital (social cohesion and civic engagement) are not strongly associated with stock market participation.

1. Introduction

Stock market participation rates vary substantially across the United States. More specifically, participation is significantly lower in low-income households.[i,ii,iii] This is considered irrational behavior because these households fail to take advantage of the high equity premium of stock markets. For example, the average annualized equity premium in the US stock market averaged 5.9% between 1928 and 2019.[1] In other words, all households should invest at least a portion of their wealth in the stock market to reduce the income gap.[iv] Since behavior is misaligned with theory, researchers have called this phenomenon the non-participation puzzle.

Previous research proposed three explanations for the paradox: participation costs (fiscal costs or limited cognitive resources), information barriers and behavioral biases (loss aversion, narrow framing and ambiguity aversion).[v] The role of the second pillar, information barriers, has

[1] The equity premium equals the difference between the yearly holding period return of the value-weighted CRSP US Total market Index and the annualized holding period of the 90-day treasury bill. The maximum spread in this period was 7.8%.

been studied in previous work using social capital, which is defined as the positive product of human interaction that includes information, innovative ideas and future opportunities.[vi] Social capital has been recognized as a crucial influence in stock market participation.[vii] One study showed that self-reported socially active individuals are more likely to invest in the stock market.[viii] Similarly, stock performance of peers was a strong predictor of individuals' entry decision in Finland.[ix] Other studies have used electoral participation (in Italy), blood donations (in Italy) and credit scores (in the US) as an indicator of a community's social capital.[x, xi] All concluded that areas that have more social capital have higher stock market participation, even after controlling for a rich set of socioeconomic, preferential, neighborhood and demographic characteristics. A recent study used a Facebook Social Connectedness Index (SCI) and IRS tax filing data to show that social network and word-of-mouth communications play an important role in stock market participation.[xii]

While these studies reported that social capital plays a central role in stock market participation decisions, the measures used don't directly observe social interactions between different types of people, a distinction that has been shown to be empirically important.[xiii] More specifically, the association between social capital and stock ownership is more pronounced among the lower educated, and social capital levels where one grew up have a lasting influence on future stock investment.[viii] For this reason, in this research I use data on the social networks of 72.2 million users of Facebook between 25 and 44 years from the Social Capital Atlas to correlate three new measures of social capital with data on stock market participation. These measures are cross-type economic connectedness (the extent to which different types of people are friends with each other), social cohesion (the degree to which friendship networks are clustered into cliques) and civic engagement (participation in civic organizations). This dataset was developed by Chetty et

al and used to analyze the relationship between social capital and upward mobility.[xiv] This paper builds off of that work.

The paper proceeds as follows. Section 2 describes the methodology, including social capital measures, stock market participation data, control variables and demographic data. Section 3 provides the results of the correlation and regression analyses. I begin by examining correlations between social capital and stock market participation. I then consider the extent to which those correlations are driven by omitted variables. Section 4 discusses the findings, lists future research opportunities and concludes.

2. Methodology

2.1 Social Capital Measures

In this paper, I measure social capital using six variables that are divided in three categories by the Social Capital Atlas. These categories were developed using a privacy-protected dataset on 21 billion friendships between Facebook users aged 25 to 44 in the United States.[2] These online friendships serve as a proxy for real-life relationships. The first category of social capital is Social Connectedness, defined as two times the share of high-socioeconomic status friends among low-socioeconomic individuals, averaged over all low-socioeconomic status individuals in a county. The second social capital category is Cohesiveness. This category is measured by two variables: Clustering (the average fraction of an individual's friend pairs who are also friends with each other) and Support Ratio (the proportion of within-county friendships where the pair of friends share a mutual friend within the same county). The third social capital category is Civic Engagement,

[2] Two papers on this research, Social Capital I: Measurement and Associations with Economic Mobility and Social Capital II: Determinants of Economic Connectedness were published in Nature on August 1, 2022. These papers include detailed descriptions and equations of how the social capital measures are computed.

measured by two variables: Volunteering (the percentage of Facebook users who are members of a group which is predicted to be about volunteering or activism based on group title and other group characteristics), Civic Organizations (the number of Facebook pages that are considered "Public Good" pages based on page title, category and other page characteristics, per 1,000 users in the county) and the Penn State Index (a composite score of civic engagement comprised of the number of membership organizations per 1,000 population, voting rate in presidential elections, the response rate to the Census Bureau's decennial census, and the number of non-profit organizations per 10,000 population).[xv] The Penn State Index is the only social capital measure that wasn't put together using the Facebook dataset.

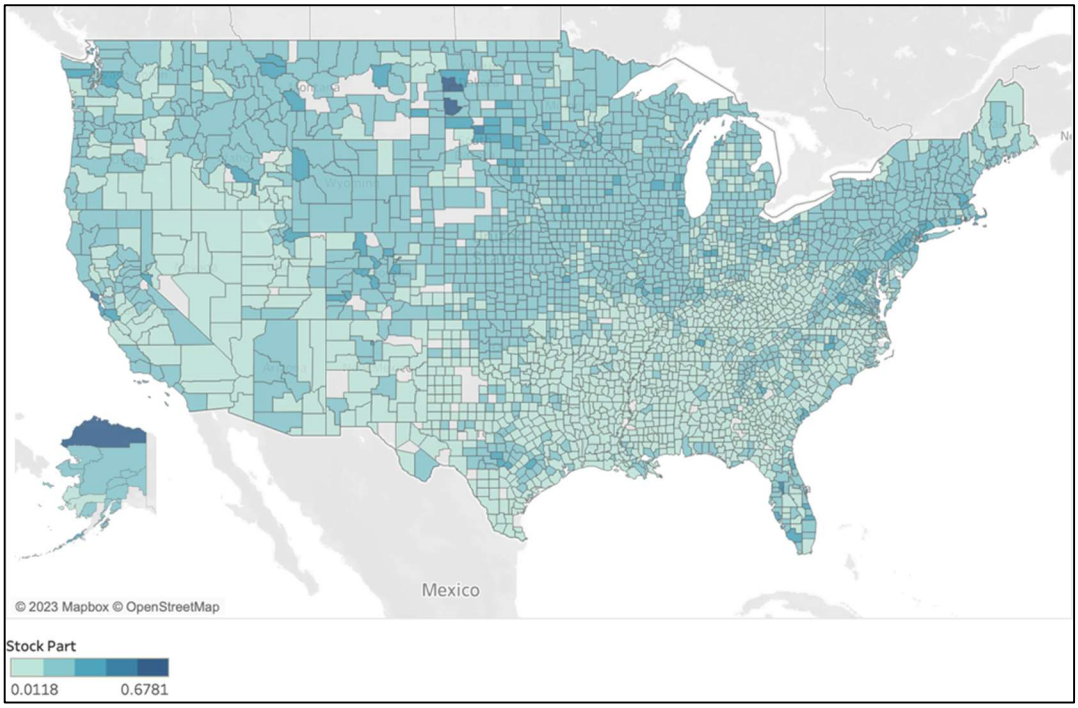
2.2 Stock Market Participation

In order to measure stock market participation, I use the Internal Revenue Services (IRS) Statistics of Income, a publicly available dataset that provides aggregated personal income tax information.[xvi] In this paper, I define the stock market participation rate in a county as the ratio of number of tax filings with dividend tax payment in the county in 2019 over the total number of tax filings in that county in 2019. The reasoning behind this methodology is that people who pay dividend tax hold some type of equity that yields dividends. Figure 1 displays the US county-level stock market participation rate in 2019. The average participation rate is 15.78%, ranging from 0.01% to 67.81%, and the standard deviation is 6.66%.

It's important to note that the IRS taxation data doesn't capture 100% of people's stock market participation. First, people may hold stocks that don't pay dividends. Second, people may indirectly participate in the stock market through mutual or pension funds. In 2022, more than 80% of the companies listed in the S&P500 paid dividends. In addition, 20% of the mutual funds

consistently distribute stock dividends down to the investors. [xvii] In other words, the IRS data reflects the lower bound of US stock market participation. [3]

Figure 1: The geography of stock market participation in the United States (2019)



2.3 Alternative drivers of stock market participation

As mentioned earlier, previous research listed participation costs, information barriers and behavioral biases as potential determinants of stock market participation.[xviii] While the social capital variables reflect information barriers, I need to control for participation costs and behavioral biases, which includes taxes, trust and risk attitude.[xix, xx] One study showed that after the revelation of corporate fraud in a state, household stock market participation in that state decreased.[xxi] Another study showed that Dutch and Italian individuals with a high perceived probability of being cheated tend to have lower stock market participation.[xxii] To control for these indicators, I use three variables. First, I look at the number of complaints filed at the Federal

[3] The average population-weighted participation rate is 17.96%.

Communications Commission (FCC). The higher the number of complaints, the higher the level of social distrust. I aggregated the number of complaints per 100 population at county level. On average, about 2.98 complaints were filed to the FCC per 100 residents, with significant cross-county variation; the number of complaints filed ranged from 0 to 121 per county.

Second, I control for the residents' risk appetite by including the National Risk Index for each county. This score is a baseline relative risk measurement designed by the Federal Emergency Management Agency (FEMA) to illustrate which communities in the United States are most at risk.[xxiii] The index is based on data for expected annual loss due to natural hazards, social vulnerability and community resilience. While previous research used individual self-reported levels of risk-aversion, this type of data is unavailable at county level. A county's score lies between 0 and 100 and describes its relative position among all other communities. The average risk score in the US is 10.8, with extreme cross-county variation (e.g. a score of 0 for Loudoun (VA) and Chattahoochee (GA), and 100 for Los Angeles (CA). There are two challenges with interpreting the National Risk Index only as risk aversion. First, not everyone is aware the risks of the location where they live. Second, some of the most affected areas are relatively poor, and it's hard for people to move out to a safer location if the risk is higher than their risk appetite. Examples of such counties are the Bronx (NY), Cameron (TX) and Hidalgo (TX).

Third, I look at the 2022 capital gains tax at state level.[xxiv] MacKie-Mason et al (1998) proved that taxes significantly affect corporate and personal financing decisions.[xxv] In other words, the higher the capital gains tax – the rate at which dividends and other stock market gains are taxed – the lower the expected tendency to start or increase stock market participation. The rates range from 0% (e.g. Alaska, Wyoming and Texas) to 13.30% in California.

2.4 Demographic Data

Obviously, one would expect stock market participation to be higher in high income communities. Therefore, I used demographic data in multiple categories to correct for this effect: population, income (e.g. mean and median household income, poverty rate, income equality), education (fraction of population with a college degree), race (e.g. share of black, Hispanic and white individuals in a given county, race inequality), employment (e.g. job growth rate, employment rate). This data was obtained from the 2000 Census and the Opportunity Atlas.[xxvi,xxvii] Appendix 1 describes all variables in detail.

3. Results

3.1 The association between social capital and stock market participation

Table 1 reports univariate correlations (weighted by population) across counties between stock market participation and each measure of social capital that was constructed in the Social Capital Atlas.[4] These measures are split in three categories: Economic Connectedness [EC], Network Cohesiveness (Clustering and Support Ratio), and Civic Engagement (Penn State Index, Civic Organizations and Volunteering Rate). Appendix 2 shows the scatter plots of stock market participation against the three main social capital measures: EC, Clustering and Civic Organizations, respectively.

[4] This is limited to the measures that were publicly accessible. The paper "Social Capital and Upward Mobility" had access to the full dataset, including social capital measures such as "Facebook Childhood EC", "Instagram Childhood EC", "EC Restricted to Top 10 Friends".

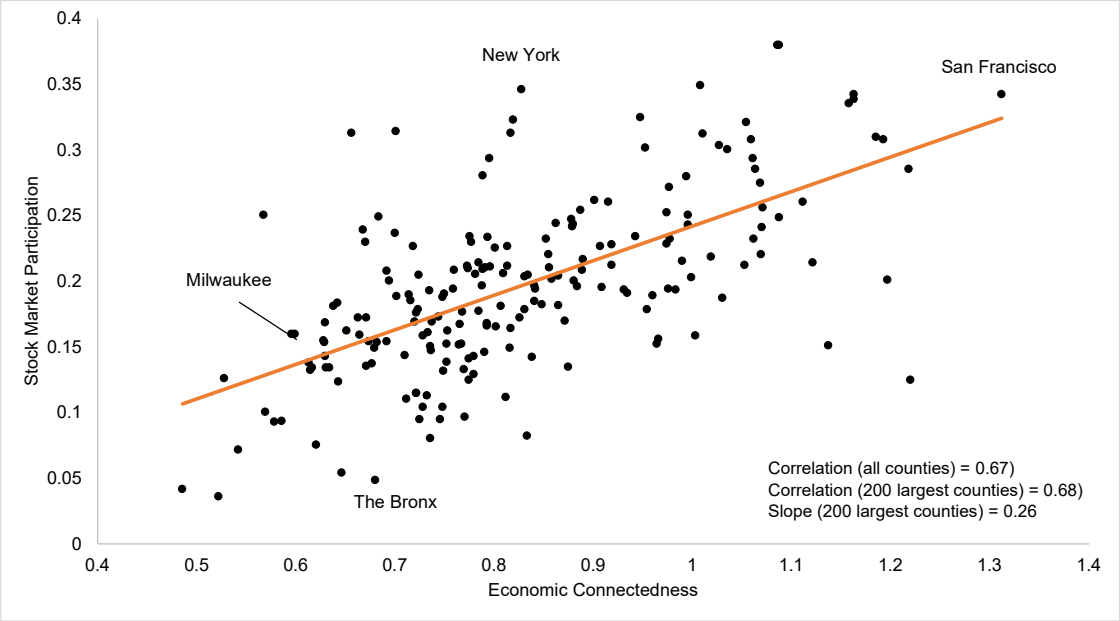
Table 1: County-level correlations between stock market participation and measures of social capital

	Correlation Coefficient
Economic Connectedness	0.67***
Social Cohesion	
Clustering	-0.05***
Support Ratio	-0.25***
Civic Engagement	
Penn State Index	0.42***
Civic Organizations	0.34***
Volunteering Rate	0.32***

*** = Statistically significant at the 1% level

EC is strongly positively correlated with stock market participation (0.67). Figure 2 shows the relationship between EC and stock market participation for the 200 most populous counties through a scatter plot. Similar to the story told by the positive correlation of 0.67 in the entire sample, Figure 2 shows that people in counties where low-SES individuals have more high-SES friends tend to have higher rates of stock market participation. As an example, low-SES individuals have a much larger share of high-SES friends in San Francisco (66%, corresponding an EC of 1.31) compared with Milwaukee (33%, corresponding an EC of 0.65). Correspondingly, the average stock market participation rates in San Francisco and Milwaukee are 34% and 16%, respectively. On average, an increase in EC of 0.5 units (equivalent to raising the share of high-SES friends among low-SES people from 25% to 50%, and approximately equal to the difference in EC between the 10^{th} and 90^{th} percentile counties) is associated with an increase in stock market participation of 13 percentiles.

Figure 2: Association between stock market participation and EC across counties



Measures of network cohesion are less strongly associated with stock market participation. The correlation coefficients for clustering and support ratio are -0.05 and -0.25, respectively. In other words, counties with high support ratios – proportion of within-county friendships where the pair of friends share a third mutual friend within the same county – tend to have lower rates of stock market participation. This considers areas that exhibit highly cohesive networks – and thus might be thought of as tightly knit communities – but nevertheless have low levels of EC and corresponding low levels of stock market participation. A potential explanation for this pattern is that although those communities have strong social connections among their predominantly low-income residents (high cohesion), they are not well connected to individuals from higher-SES backgrounds who can provide the resources or information required to start participating in the stock market.

Finally, stock market participation and measures of civic engagement are moderately associated, with correlations being 0.32 (Volunteering Rate), 0.34 (Civic Organizations) and 0.42

(Penn State Index). Levels of civic engagement in societies have been previously measured using self-reported levels of trust.[xxviii] According to multiple studies, trust is an important indicator in stock market participation. In other words, a potential explanation for the pattern is that counties with high civic engagement show high levels of trust, and therefore experience higher stock market participation rates.

Figure 3: Multivariate regression of stock market participation on measures of social capital

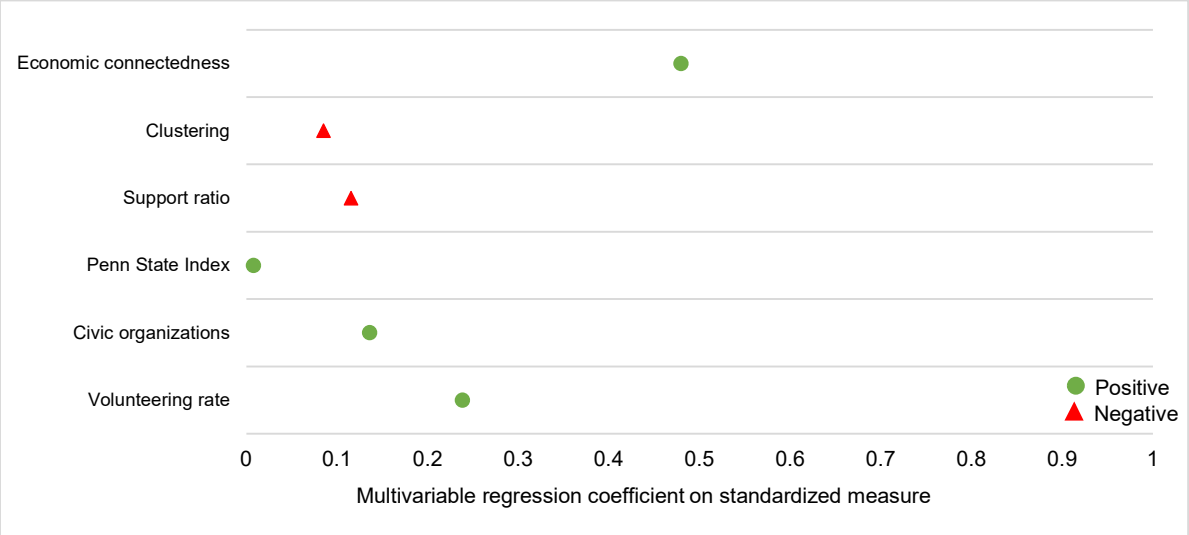


Figure 3 shows the results of a regression of stock market participation on standardized versions of all the social capital measures that are featured in Table 1. EC remains the strongest predictor of stock market participation. In line with the negative associations between measures of social cohesiveness (clustering and support ratio) in Table 1, the regression coefficients for these variables are negative.

3.2 Why EC is related to stock market participation

The remainder of the results section will explore the reason behind the strong correlation between EC and stock market participation, which is either causal or non-causal. In case of a causal effect, stock market participation might be facilitated by connections to high-SES people who can shape

aspirations or provide access to information on the stock market. This theory is consistent with the information barrier that has been reported as one of the main reasons why people stay away from the stock market. A non-causal relationship has three potential explanations: reverse causality, selection effects or omitted variables. In the next sub-sections, I will evaluate all three options.

3.2.1 Reverse Causality

The first alternative explanation for the correlation between EC and stock market participation is reverse causality. In this case, higher stock market participation could lead to higher EC. In this scenario, which goes against intuition, stock market participation brings people from different socioeconomic background closer to each other. In other words, the engagement in stock market participation can be seen as the common denominator or bonder in these communities. The correlation analysis was performed with EC measured among adults. Since the friendships and SES in the dataset are measured in adulthood, EC may itself be influenced by stock market participation. For example, in places with high stock market participation, many children from low-SES families have higher incomes as adults and may retain friendships with individuals who remain at a low SES. This would lead to high stock market participation areas having a high rate of friendships among people with different SES in adulthood, even in the absence of any effect of EC on stock market participation. To test this theory, I examined the association between stock market participation and childhood EC, on the basis of childhood friendships and parental SES. Because childhood friendships are made before people start selling and buying stocks, they can't be directly influenced by rates of stock market participation. Childhood EC is based on high school friends and parental SES of individuals. While slightly lower than the 'adulthood' correlation between stock market participation and EC, the correlation between stock market participation and

childhood EC across counties remains strong at 0.57. Since stock market participation remains strongly correlated with childhood EC, any causal effect of stock market participation on connectedness can only account for, at most, a small share of the correlation between the two variables.

3.2.2 Selection Effects

Selection bias is a second alternative to a causal relationship between social capital and EC. In other words, the type of families that live in high-EC areas may inherently have higher rates of stock market participation, independent of where they live. Put differently, these selection effects lead people in high-EC counties to have higher stock market participation rates, even in the absence of a causal effect of EC on these rates. The most important demographic variables that define residence are education and race. Education can reduce the information barrier to engage in stock market trade. Race also plays a role. For example, areas with larger Black populations, on average, have lower rates EC; the average EC scores for counties in which black and white individuals are dominant are 0.54 and 0.83, respectively. Put differently, the degree to which low-income and high-income people are friends with each other is higher in areas in which people are predominantly white. At the same time, previous research has shown that Black people have lower stock market participation than White people and that this is likely to widen the post-pandemic wealth gap.[xxix] The combination of these effects may have led to the strong correlation between EC and stock market participation in Table 1.

Lacking data on individual education level, I grouped counties by the fraction of the population that has completed a college degree. A county was labeled ‘Lower Educated’ if the fraction of a county’s population that completed at least a college degree was lower than 10%. A

county was labeled 'Medium Educated' if the fraction of a county's population with a college degree was between 10 and 30%. Finally, a county was labeled 'Higher Educated' if the fraction of the county's population with a college degree exceeded 30%. Next, these buckets of counties were correlated with stock market participation. Table 2 reports the results of this analysis. Column 1 shows that the correlation between stock market participation and EC is 0.47 for lower educated individuals. The average fraction of the population with a college degree in these 226 counties was 8.6%. The correlation is slightly higher (0.55) for the sample of 2,471 counties in which 10 to 30% of individuals have completed a college degree. In this group, which includes the vast majority of counties, the average fraction was 17%. Column 3 shows slightly lower but very similar correlation (0.51) for the 'Higher Educated' counties. The average share of people with a college degree in those 323 counties is 38%. All correlations are statistically distinguishable at the 1% level.

Table 2: Correlation between Stock Market Participation and Economic Connectedness

Stock Market Participation for:	Lower Educated (1)	Medium Educated (2)	Higher Educated (3)
Economic Connectedness	0.47***	0.55***	0.51***
Requirement	<10%	10-30%	>30%
Observations	226	2471	323
Fraction with a College Degree	8.6%	17%	38%

A similar exercise was completed for race. In absence of information on race on individual level, I grouped counties by their most dominant ethnicity using Census data. In other words, a county was labeled 'White' if the majority of a county's population in 2018 self-identified as White. A similar approach was taken to identify 'Black', and 'Hispanic' counties. Next, these counties were correlated with stock market participation, weighted by total population. Table 3 reports the results

of this analysis. Column 1 shows that the correlation between stock market participation and EC is 0.59 for white individuals. The average share of white individuals in these counties was 78%. The correlation is similar (0.56) for the sample of counties in which at least 70% of individuals report as White, which is still large with over 2000 observations. Column 2 shows lower correlation (0.50) for the sample of counties in which the majority reports as Black. That said, the average share of Black people in those 104 counties is only 57%. There are only 14 counties in which the share of Black individuals exceeds 70%. In this sample, the correlation between stock market participation and EC is 0.41. In counties in which Hispanics are the dominant race there is a correlation of 0.48 in the overall sample of 125 counties and an almost equal correlation (0.49) in the sample of 28 in which at least 70% of the population self-identifies as Hispanic. All correlations are statistically distinguishable at the 1% level.

Table 3: Correlation between Stock Market Participation and Economic Connectedness

Stock Market Participation for:	White Counties (1)	Black Counties (2)	Hispanic Counties (3)
Economic Connectedness	0.59***	0.50***	0.48***
Observations	2,788	104	125
Average Race Share in Sample	78%	57%	60%
Economic Connectedness	0.56***	0.41**	0.49***
Requirement	> 70% White	> 70% Black	> 70% Hispanic
Observations	2,003	14	28
Average Race Share in Sample	86%	76%	83%

The results in Table 2 and 3 show that though not as strong as in the overall sample, EC remains correlated with stock market participation across all education and racial makeup samples. This

implies that segregation by race or education is unlikely to be the primary driver of the overall correlation between EC and stock market participation. This is in line with a study conducted by the Federal Reserve Bank of St. Louis in 2022, which found that the participation gap between white and black households prevails when controlling for household income.[xxx]

3.3.3 Omitted Variables

The third alternative to a causal effect of EC on stock market participation is that the correlation is the result of omitted variables. To find this out, I compared the relative explanatory power of EC and all the other predictors of stock market participation that have been used in previous studies. Figure 4 shows univariate county-level correlations between stock market participation and measures of income and various other neighborhood characteristics that have been proven to drive stock market participation. These can be bucketed in three categories: measures of income, inequality and other factors impacting the tendency to engage in stock market participation.

Figure 4: County-level correlations between stock market participation and county characteristics

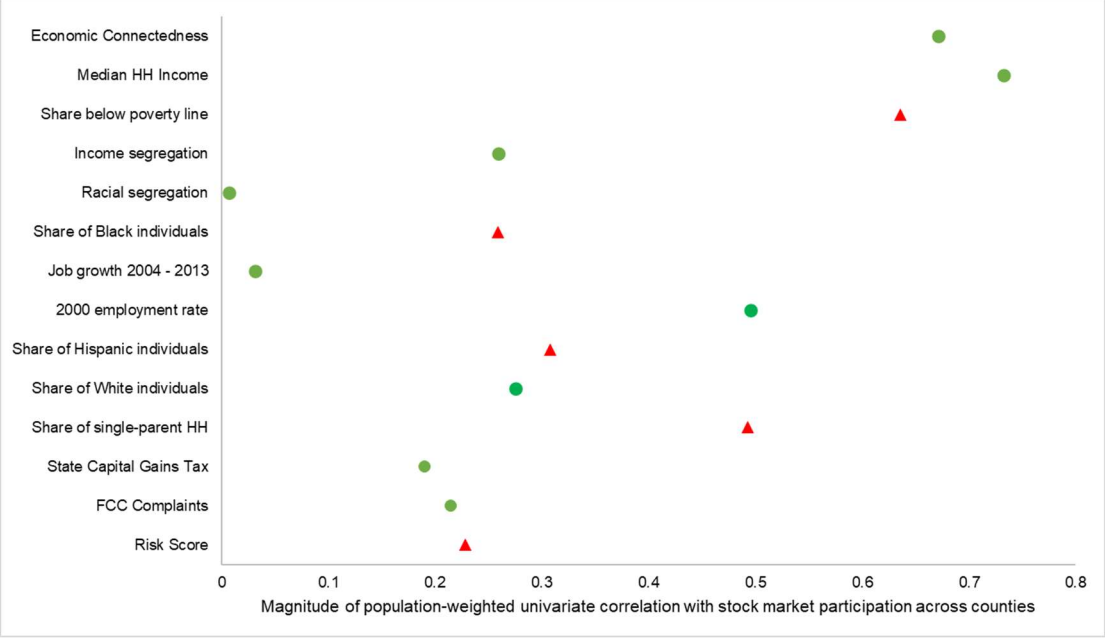
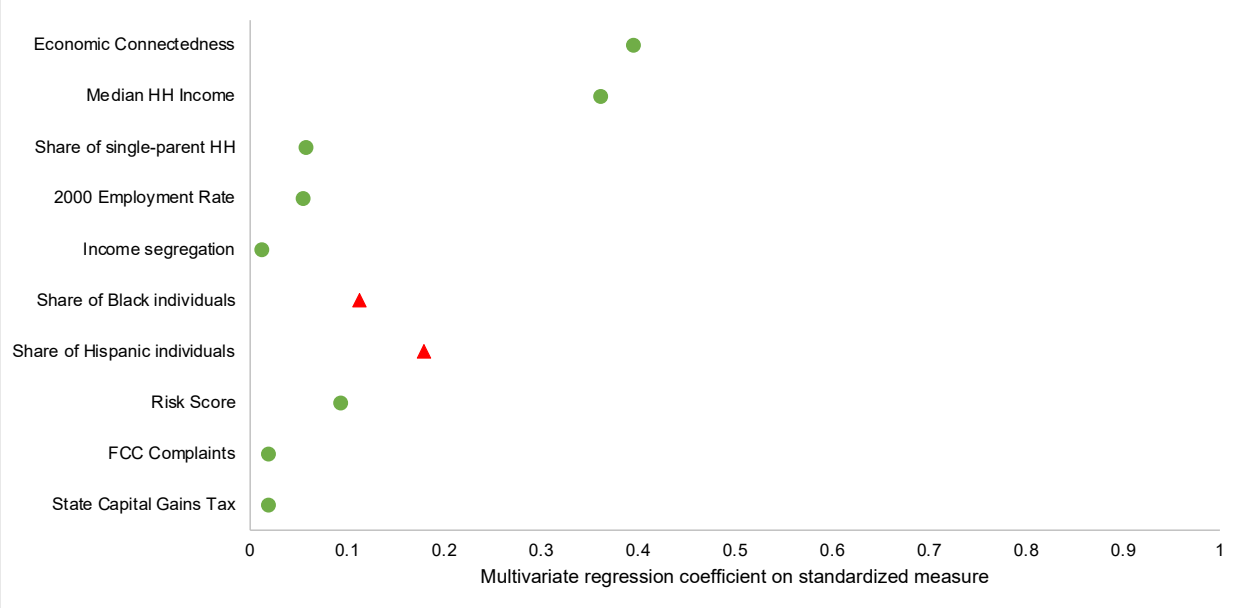


Figure 5 displays the results of a regression of stock market participation on standardized versions of a subset of 10 variables from Figure 4. The variables that were used have the largest univariate correlation with stock market participation (except for share of poverty, which is highly correlated with the share of single-parent households, and Share of White individuals, which is highly correlated with share of Black and Hispanic individuals). EC remains the strongest predictor of stock market participation. In line with the negative associations between stock market participation and the share of Black and Hispanic individuals in Figure 4, the regression coefficients for these variables are negative. Except for income segregation and the number of FCC Complaints, all variables are statistically significant (State Capital Gains Tax at 5%, all other variables at 1%).

Figure 5: Multivariate regression of stock market participation on a subset of variables



Next, I analyzed the variable buckets in more detail, starting with measures of income across counties. The share of individuals below the poverty line, median household incomes and the employment rate have strong absolute correlations of 0.64, 0.73 and 0.50 with stock market

participation across counties, respectively. Since these correlations are close to the correlation between EC and stock market participation, which is 0.67, this is a first indication that income measures play a role in stock market participation beyond their effect on EC. When I regressed stock market participation on both EC and these measures of local income, EC remains a strong predictor of stock market participation. In other words, after controlling for income, EC still matters. Unsurprisingly, median household income nor employment rate lose its explanatory power either. Columns 2 and 6 in Table 4 show that both independent variables remain statistically significant at the 1% level. The poverty rate loses some of its explanatory power, as it drops to a 5% significance level when EC is included in the regression (column 4). These results suggest that social ties may predict stock market participation over and above average incomes and employment rates. Put differently, living in a lower income county with lower employment rates inhibits stock market participation as it reduces interaction with people with a higher socioeconomic status.

Figure 6: Associations between stock market participation, EC and median household income by county

Figure 6 demonstrates this dual effect (effect of income measures both through and beyond EC) in more detail through a scatter plot of EC against median household income by county. The dots are colored according to the level of stock market participation. Dark blue dots represent counties with high levels of stock market participation, while red dots represent counties with low levels of stock market participation. Horizontal slices of the graph – counties with different levels of median household income but similar levels of EC tend to have somewhat similar levels of economic mobility. For example, the majority of counties with an EC between 0.4 and 0.8 have a low-to-medium stock market participation (orange dots). That said, there are areas (e.g. EC between 0.8 and 1.0) that show a large variety of stock market participation. Similarly, vertical slices of the graph – counties with similar income but different levels of EC – show stock market participation across the entire spectrum for a household income between $40,000 and $60,000, but fewer variation for counties with income above $80,000. These results imply that living in an area with high EC isn't the only requirement for higher stock market participation; having a high income, or being surrounded with high-income people also plays a role.

Previous research reported that stock market participation is lower in areas with higher income or racial inequality. Income segregation and racial segregation have lower absolute correlations – 0.2592 and 0.007 – with stock market participation than EC (0.67). When I regress stock market participation on both EC and income and race segregation measures, EC remains a strong predictor of stock market participation and adds significantly to the explanatory power through an increased R^2. This is shown in column 2 and 4 of Table 4. That said, both income and race segregation remain statistically significant at 1% when EC is added to the regression. Like with the median household, poverty rate and employment measures, this shows that segregation measures have a strong influence on stock market participation beyond its influence on EC.

Finally, previous work has looked at other factors impacting stock market participation: risk attitude, capital gains tax and trust in financial and governmental institutions. In this paper, these are measured through the National Risk Index, the 2022 capital gains tax rate, and the number of FCC complaints, respectively. As can be seen in Figure 4, EC is more strongly correlated with stock market participation than any of these three variables. As expected, though, all these variables remain a statistically significant predictor at the 1% level when EC is regressed on stock market participation in combination with these variables. These results imply that living in an area with high EC isn't the only requirement for higher stock market participation; people's risk attitude, tax rates and the level of trust play an important role too.

Table 4: Associations between stock market participation, EC and other county characteristics

EC versus median income and poverty rates						
Dependent variable	**Stock Market Participation**					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Median Income	0.620***	0.322***				
	(0.000)	(0.000)				
Poverty Rate			-0.424***	0.013**		
			0.129	(0.013)		
Employment Rate					0.525***	0.203***
					(0.000)	(0.000)
Economic Connectedness		0.497***		0.698***		0.577***
		(0.006)		(0.006)		(0.000)
Observations	3,017	3,017	3,017	3,017	3,017	3,017
R^2	0.385	0.543	0.180	0.4763	0.2754	0.5046

EC versus segregation and inequality				
Dependent variable	**Stock Market Participation**			
	(1)	**(2)**	**(3)**	**(4)**
Income segregation	0.236***	0.159***		
	(0.000)	(0.000)		
Racial segregation			-0.181***	0.108***
			(0.000)	(0.000)
Economic Connectedness		0.672***		0.731***

		(0.005)		(0.000)		
Observations	3,016	3,016	2,998	2,998		
R^2	0.056	0.501	0.0326	0.4839		

EC versus stock market determinants						
Dependent variable	**Stock Market Participation**					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Risk Score	-0.107***	0.091***				
	(0.000)	(0.000)				
Complaints			0.064***	0.035***		
			(0.001)	(0.010)		
State Tax Capital Gains					0.172***	0.061***
					(0.000)	(0.000)
Economic Connectedness		0.715***		0.683***		0.680***
		(0.000)		(0.000)		(0.000)
Observations	3,017	3,017	2,843	2,843	3,017	3,017
R^2	0.111	0.484	0.004	0.4703	0.030	0.4798

4. Discussion

In this paper, I investigated the relationship between social connectedness and stock market participation. In the past, this has proven to be more challenging due to the lack of cross-type data on social capital. By using the newly developed Social Capital Atlas database, I was able to explore the relationship between stock market participation and three types of social capital. While economic connectedness is strongly associated with stock market participation, social cohesion and civic engagement are not.

While there is a strong relationship between social connectedness and stock market participation, it can't be concluded that the relationship is causal. The analysis ruled out reverse causality and selection bias. The key result, however, is that social capital matters for stock market participation over and above factors like income, racial inequality, risk appetite, trust and fiscal costs. This is consistent with previous research, which reported three different areas of stock

market participation drivers: participation costs, information barriers (=social capital) and behavioral biases.

The results in this paper create multiple questions for future research. First, some of the analyses in this paper (Tables 2-4 and Figure 6), should be repeated on zip-code level. While this was beyond the scope of this paper, the additional datapoints on zip-code level will provide a second-opinion for the results in this paper. Second, it would be valuable to leverage the wealth of this dataset by repeating the methods used in this paper to determine which forms of social capital – economic connectedness, social cohesion or civic engagement – matter for other socioeconomic outcomes (e.g. health choices or political outcomes). Third, it would be valuable to repeat this study using data in the same year. Due to limited availability of data on county level, some demographic data stems from 2000, while median income stems from 2016 and the social capital measures are constructed with even more recent data. Cross-county changes over time may have impacted the results presented in this paper.

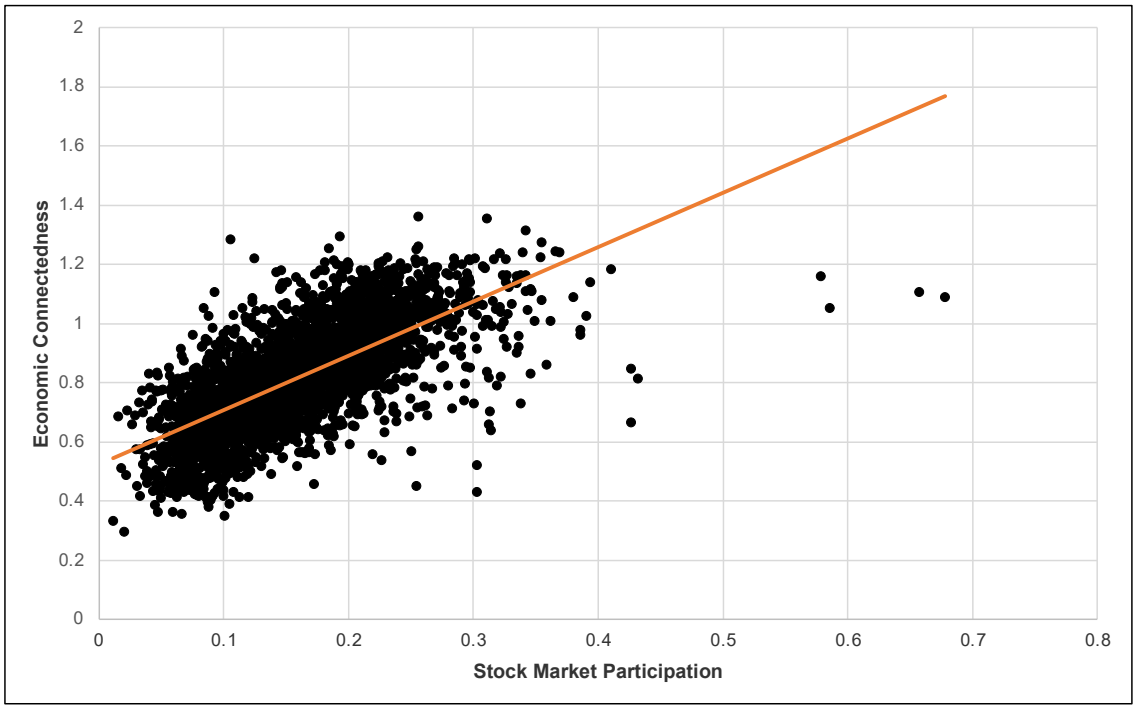
Appendix 1: Variable Codebook

Variable	Description
state	Two-digit state 2010 FIPS code
county	Three-digit county 2010 FIPS code
county name	Name of the county and state.
ann_avg_job_growth_2004_2013	Average annualized job growth rate over the time period 2004 to 2013. Constructed using Local Area Unemployment Statistics (LAUS) released by the Bureau of Labor Statistics
emp2000	The rate of employment computed as total employed population (the sum of employed females and employed males) divided by the total population 16 years and over. Obtained from 2000 Decennial Census
frac_coll_plus2010	Number of people aged 25 or older who have a bachelor's degree, master's degree, professional school degree, or doctorate degree, divided by the total number of people aged 25 or older in a tract.
hhinc_mean2000	Mean household income. Obtained from 2000 Decennial Census.
med_hhinc2016	Median household income. Obtained from the 2012-2016 American Community Survey.
Poor_share2010	Share of individuals in the county below the federal poverty line, measured in the 2006-2010 American Community Survey.
singleparent_share2010	The number of households with females heads (and no husband present) or male heads (and no wife present) with own children under 18 years old

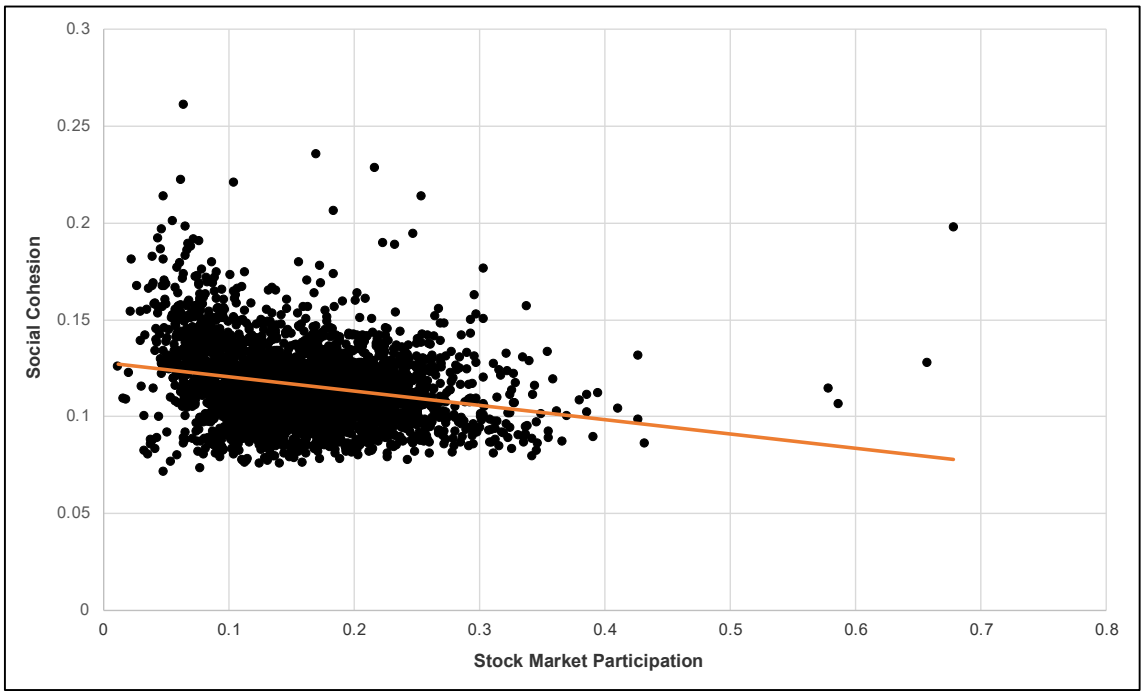
	present divided by the total number of households with own children present. Obtained from the 2006 -2010 estimate.
pop2018	Population in 2018. This variable is not constructed using Facebook data; it is obtained from publicly available data posted at the Census website (American Community Survey).
ec_county	Baseline definition of economic connectedness: two times the share of high-SES friends among low-SES individuals, averaged over all low-SES individuals in the county. See equations (1), (2), and (3) of Chetty et al. (2022a) for a formal definition. We calculate SES as in Supplementary Information B.1 of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document. This variable is mapped in Figure 2A of Chetty et al. (2022a).
child_ec_county	Childhood economic connectedness: two times the share of highparental-SES friends among low-parental-SES individuals averaged over all low-parental-SES individuals in the county, calculated using only individuals' high school friends.
clustering_county	The average fraction of an individual's friend pairs who are also friends with each other. See equations (4) and (5) of Chetty et al. (2022a).
support_ratio_county	The proportion of within-county friendships where the pair of friends share a third mutual friend within the same county. See equation (6) of Chetty et al. (2022a).
volunteering_rate_county	The percentage of Facebook users who are members of a group which is predicted to be about 'volunteering' or 'activism' based on group title and other group characteristics.
civic_organizations_county	The number of Facebook Pages predicted to be "Public Good" pages based on page title, category, and other page characteristics, per 1,000 users in the county.
penn_state_index	Composite score of civic engagement comprised of the number of membership organizations per 1,000 population, voting rate in presidential elections, the response rate to the Census Bureau's decennial census, and the number of non-profit organizations per 10,000 population.
hispanic	Number of individuals who report as Hispanic. Obtained from Decennial Census.
black	Number of individuals who report as Black. Obtained from Decennial Census.
white	Number of individuals who report as White Obtained from Decennial Census.
other	Number of individuals who don't report as Hispanic, black, or white.
theil_race_index	Measure of racial inequality; entropic "distance" the population is away from the everyone having the same race. Calculated using methodology of Iceland, J. (2004).
theil_income_index	Measure of economic inequality; entropic "distance" the population is away from the "ideal" egalitarian state of everyone having the same income. Calculated using methodology of Iceland, J. (2004).
state_capital_tax	The levy on the profit that an investor makes when an investment is sold.
risk_score	*Risk Index = Expected Annual Loss × Social Vulnerability ÷ Community Resilience*, where *Expected Annual Loss = Exposure × Annualized Frequency × Historical Loss Ratio.* Obtained from FEMA National Risk Index.
fcc_complaints	The number of complaints filed at the Federal Communications Commission (FCC). Obtained from https://www.fcc.gov/consumer-help-center-data.

Appendix 2: Correlation

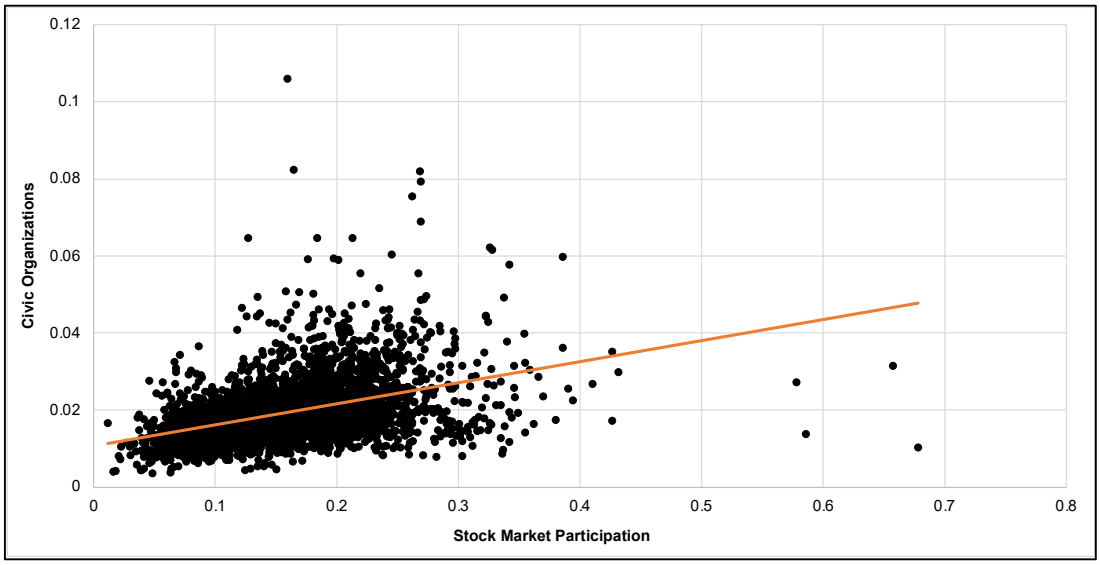
a) Association between Economic Connectedness and Stock Market Participation



b) Association between Social Cohesion and Stock Market Participation



c) Association between Civic Organizations and Stock Market Participation



[i] Vissing-Jørgensen, Annette, and Orazio P. Attanasio. 2003. "Stock-Market Participation, Intertemporal Substitution, and Risk-Aversion ." American Economic Review, 93 (2): 383-391.
[ii] Campbell, J.Y. (2006), Household Finance. The Journal of Finance, 61: 1553-1604. https://doi.org/10.1111/j.1540-6261.2006.00883.x
[iii] Guiso, L., & Sodini, P. (2013). Household finance: An emerging field. In Handbook of the Economics of Finance (Vol. 2, pp. 1397-1532). Elsevier.
[iv] Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. The review of Economics and Statistics, 247-257.
[v] Bricker, J., Dettling, L. J., Henriques, A., Hsu, J. W., Jacobs, L., Moore, K. B., ... & Windle, R. A. (2017). Changes in US family finances from 2013 to 2016: Evidence from the Survey of Consumer Finances. Fed. Res. Bull., 103, 1.
[vi] Kenton, W. (2022, November 27). *What Is Social Capital? Definition, Types, and Examples.* Investopedia. https://www.investopedia.com/terms/s/socialcapital.asp
[vii] Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1992). Making democracy work: Civic traditions in modern Italy. Princeton university press.
[viii] Hong, H., Kubik, J., Stein, J., 2004. Social interaction and stock-market participation. The Journal of Finance 59, 137-163.
[ix] Kaustia, M., & Knüpfer, S. (2012). Peer performance and stock market entry. Journal of Financial Economics, 104(2), 321-338.
[x] Bricker, J., & Li, G. (2017). Credit scores, social capital, and stock market participation. Social Capital, and Stock Market Participation (2017-02). FEDS Working Paper, (2017-008).
[xi] Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. "The Role of Social Capital in Financial Development." American Economic Review, 94 (3): 526-556.

[xii] Hung, C. C. (2021). Does Social Connectedness Affect Stock Market Participation?. Available at SSRN 3724407.
[xiii] Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., ... & Wernerfelt, N. (2022). Social capital II: determinants of economic connectedness. Nature, 608(7921), 122-134.
[xiv] Chetty, R., Jackson, M.O., Kuchler, T. et al. Social capital I: measurement and associations with economic mobility. Nature 608, 108–121 (2022). https://doi.org/10.1038/s41586-022-04996-4
[xv] Rupasingha, A., Goetz, S. J., & Freshwater, D. (2006, with updates). The production of social capital in US counties. Journal of Socio-Economics, 35, 83-101. doi:10.1016/j.socec.2005.11.001.
[xvi] IRS (2019). SOI Tax Stats County Data 2019 [Data set]. https://www.irs.gov/statistics/soi-tax-stats-county-data-2019
[xvii] Titan. (2022, February 10). Does the S&P 500 Pay Dividends? https://www.titan.com/articles/does-the-sp-500-pay-dividends
[xviii] Bricker, J., Dettling, L. J., Henriques, A., Hsu, J. W., Jacobs, L., Moore, K. B., ... & Windle, R. A. (2017). Changes in US family finances from 2013 to 2016: Evidence from the Survey of Consumer Finances. Fed. Res. Bull., 103, 1.
[xix] Mishra, R. (2018). Financial literacy, risk tolerance and stock market participation. Asian Economic and Financial Review, 8(12), 1457-1471.
[xx] Gumbo, M., & Sandada, M. (2018). The Determinants of Stock Market Participation: Evidence from Individual Investors in Zimbabwe. Acta Universitatis Danubius: Oeconomica, 14(4).
[xxi] Giannetti, M., & Wang, T. Y. (2016). Corporate scandals and household stock market participation. The Journal of Finance, 71(6), 2591-2636.
[xxii] Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. the Journal of Finance, 63(6), 2557-2600.
[xxiii] FEMA. (2023). Data Resources [Data set]. https://hazards.fema.gov/nri/data-resources
[xxiv] Realized. (2023) *Capital Gain Tax Rates by State.* https://www.realized1031.com/capital-gains-tax-rate
[xxv] Walsh, W. E., Wellman, M. P., Wurman, P. R., & MacKie-Mason, J. K. (1998, May). Some economics of market-based distributed scheduling. In Proceedings. 18th International Conference on Distributed Computing Systems (Cat. No. 98CB36183) (pp. 612-621). IEEE.
[xxvi] U.S. Census Bureau (2000). US decennial Census 2000. https://data.census.gov/
[xxvii] Opportunity Insights. (2023). Neighborhood Characteristics by County. [Data set]. https://opportunityinsights.org/data/
[xxviii] Glaeser, E.L., Laibson, D.I., Scheinkman, J.A. & Scoutter, C.L. Measuring trust. Q.J. Econ. **115**, 811-846 (2000).

[xxix] Li, Y. (2022, February 2). Black Americans' lack of participation in the stock market likely to widen post-pandemic wealth gap. CNBC. https://www.cnbc.com/2022/02/02/black-americans-lack-of-participation-in-the-stock-market-likely-to-widen-post-pandemic-wealth-gap.html

[xxx] Julie Bennett and YiLi Chien, "The Large Gap in Stock Market Participation Between Black and White Households," Economic Synopses, No. 7, 2022. https://doi.org/10.20955/es.2022.7