Does social connectedness affect stock market participation? *

Chih-Ching Hung

December 26, 2020

Latest Version

Abstract

Using IRS tax filing data, I show that social network and word-of-mouth communications play an important role in stock market participation decisions. Using a novel dataset from Facebook, I construct a measure of social network friends' participation for US counties and find that a one-percentage point increase in friends' participation increases the focal county participation by 14 to 25 basis points in the following year. For identification strategy, I employ the revelation of financial misconducts as an exogenous negative shock to local participation rate and show that the instrumented change in friend participation significantly and positively predicts the change in focal county participation rates. The increase in participation rates among the low-income households induced by friends' participation decreases the Gini coefficients in metropolitan counties in the following two years. The evidence suggests that social influences and peer effects contribute to the cross-sectional differences in the stock market participation rates across US counties and may lead to lower income inequality.

Keywords: Household Finance, Social Connectedness, Peer Effect, Non-Participation Puzzle

*Chih-Ching Hung: Department of Economics and Finance, Zicklin School of Business, Baruch College, CUNY. One Baruch way, Box B10-225. New York, NY 10010. Phone: (646) 312-3524, Email: Chih-Ching.Hung@baruch.cuny.edu. Thanks to Linda Allen, Stacey H. Chen, Yu-Kuan Chen, Chun-Che Chi, Sophia Gilbukh, Rawley Heimer, Jian Hua, Waldo Ojeda, Lin Peng, Johannes Stroebel, Dexin Zhou, and seminar participants at Baruch College, CUHK-Shenzhen, and National Chengchi University for valuable comments. I also thank Michael Bailey and Facebook for sharing the Facebook social connectedness index data and Mark Egan, Gregor Matvos, and Amit Seru for sharing the financial advisory data. All the errors are mine.

1 Introduction

The "non-participation puzzle" remains one of the biggest challenges in asset pricing and household finance. Between 1928 and 2019, the average annualized equity premium in the US stock market ran as high as 7.8 percent.[1] Financial theories such as the Merton (1969) model suggest that all households should invest at least a portion of their wealth in the stock market so as to take advantage of the high equity premium. In spite of such theories, empirical studies across various countries have shown that a substantial share of households do not, whether directly or indirectly, hold any stock at all; moreover, this phenomenon holds across all income levels (Haliassos and Bertaut (1995), Campbell (2006), Guiso and Sodini (2013)).

The non-participation puzzle is relevant not merely because it challenges the fundamentals of academic finance. Under-participating in the stock market can lead to lower aggregate output, decrease social welfare, and intensify social inequality. Bhamra and Uppal (2019), for instance, demonstrate in a theoretical model that sub-optimal investment decisions from household sector reduce aggregate investment in the equilibrium and decrease aggregate outputs. In addition, non-participation is especially significant in the low-income households (Vissing-Jorgensen (2003), Campbell (2006), Guiso and Sodini (2013)). Failure to take advantage of the high equity premium for the low income households exacerbates income inequality.

Researchers have linked the non-participation puzzle to factors such as participation cost (fiscal cost or limited cognitive resource), non-standard preference (loss aversion, narrow framing, or ambiguity aversion), as well as beliefs and trusts.[2] However, Chien and Morris (2017) document that, even after conditioning on income level, stock market participation rates vary significantly across different states in the US, a finding that the arguments above have a hard time accounting for. Following the approach in Chien and Morris (2017), I explore the IRS Statistics of Income and estimate county-level stock market participation rates as the ratio of tax filings with dividend tax over the total number of filings from 2010 to 2018. While stock

[1]The equity premium here is the arithmetic mean of the difference between yearly holding period return of the value-weighted CRSP US Total Market Index and the annualized holding period return of 90-day treasury bill. The annualized difference between the entire holding period returns is 5.9 percent.

[2]See Vissing-Jorgensen (2003), Barberis et al. (2006), Garlappi et al. (2007), Guiso et al. (2008), Cole et al. (2014), and Gurun et al. (2018) among others.

market participation includes both direct holding of equities and indirect participating through mutual funds, the estimation from IRS Statistics of Income captures the vast majority of direct participation and roughly 20% of the indirect participation. At the national level, the participation rate yielded from IRS Statistics of Income is highly consistent in various dimensions with the Survey of Consumer Finance (SCF), the standard database to reference when studying household finance in the US. More importantly, the precise and unique cross-sectional heterogeneity in IRS Statistics of Income provides promising perspectives in understanding profoundly how households make important financial decisions.

In this paper, I investigate the relationship between social connectedness and stock market participation. Word-of-mouth is one of the most important information sources for all types of human decision making. We rely on our friends, parents, and opinion leaders to acquire information, form beliefs, and make all kinds of economic decisions. A possible reason why people do not participate in the stock market is simply because they are unaware of the benefit of stock market participation. Using survey data from the Health and Retirement Study, Hong et al. (2004) show that self-reported socially active individuals are more likely to invest in the stock market. Kaustia and Knüpfer (2012) also find that, in Finland, stock performance of local peers is a strong predictor of individuals' entry decisions. However, there are two potential problems in the attempt of establishing a causal relationship between peer effect and stock participation. First, there may exist confounding factors that jointly determine one's social activeness and financial decision. According to Angrist (2014), it is "perilous" to make causal inference without proper research designs. Second, the process of information transmission from one person to another is not observable in the studies above. The failure to document explicit information diffusion raises concern when taking this peer effect route to explain participation decisions.

To overcome the problems above, I obtain the Facebook Social Connectedness Index (SCI) to investigate how local stock market participation rates may transmit to other regions through social connection. Facebook SCI is a county-to-county measure of social connectedness estimated from the number of Facebook users and their friends by the location where they logged in. Though it is a one-time snapshot at 2016, Bailey et al.

(2018b) have shown that SCI is very persistent and reflects trends of immigration and cultural linkage on top of geographic closeness. Recent literature have relied on the Facebook data to identify social connected structures to study the spread of COVID-19, to explain the commuting flows in urban areas, and to investigate how residents form beliefs about local mortgage price levels (Bailey et al. (2019), Bailey et al. (2020), Kuchler et al. (2020b)). The spatial nature of Facebook SCI makes it ideal to study how stock participation in one area may directly spillover to other connected regions. For each county i in year t, I construct a variable, the friends' participation ($FP_{i,t}$), as the SCI-weighted average participation rate from outer-state "friend" counties. The higher the FP, the more deeply the county is connected to high stock market participating counties. After controlling for the lagged local participation, the effects of demographic variables such as population, education attainment, unemployment rate, household income, local stock return, county fixed effects, and state*year fixed effects, I identify a positive and significant association between friends' participation and the local county participation rate in the following year. One percentage point increase in FP leads to 14 to 25 basis points increase locally. In terms of economic magnitude, the effect would translate into 53 to 93 non-participants starting to invest in the stock market in an average county of 37,000 population.

In a placebo test where I randomly simulate 1,000 sets of artificial Social Connectedness Indices, the coefficient estimated from the true SCI, 0.142, is larger than all the coefficients estimated from the simulated SCIs. The placebo test eliminates the concern that the results may be driven from potential global shocks. The positive and robust association between friends' participation and local participation indicates that word-of-mouth may be an important ingredient in stock market participation decisions.

Exogenous shocks are essential to take care of the endogeneity problem and establish causal inference. If a local shock transmits to other counties by affecting its local participation rate first, then to other counties connected through the social network, the word-of-mouth effect on stock participation may more likely to be causal. Literature have shown that trust is one critical factor in an individual's financial decision (Guiso et al. (2008), Gurun et al. (2018)). Gurun et al. (2018), for instance, find that stock market participation decreased in counties affected by the Madoff Ponzi scheme after the scandal was revealed. I follow Egan et al. (2019) to collect financial misconduct incidents and aggregate the incidents to county-year level.[3] I first verify the relationship between fraud revelation and local participation rate. One financial misconduct event reported out of a population of 1,000 households leads to 29.5 basis points decrease in the focal participation rate in the same year and a further 31.1 basis points in the subsequent year. The difference is insignificant during the year before the event (pre-trend) and statistically insignificant 2 years after. This result suggests that revelation of financial misconduct is a negative and exogenous shock to local stock market participation rates. The inclusion of the probability of financial misconduct into the regression model increase the adjusted R^2 from 3.2% to 6.0%. In addition, the F-stats for the probability of financial misconduct is 22.16. Therefore, the probability of financial misconduct serves as a nice instrument variable in the study of friends' participation and its effect on local participation. The instrumented friends' participation positively and significantly predicts the change in local participation rate in the following year. This result indicates that the relationship between the friends' participation and local participation rate is likely to be causal.

To further verify the connection between stock market participation induced by friends' participation and the "non-participation puzzle" as well as its implications to the society, I explore the cross-sections and welfare implications of the "FP" effect. If unawareness of the benefit of stock market participation is one reason behind the "non-participation puzzle", we should expect to see different responses in the change in participation from households of different income levels due to liquidity constraints. Ideally, households from all income levels would be willing to participate when they hear about the benefit of stock market through social interaction. However, households that are more liquidity constrained could not respond as flexible as the less constrained ones. I sort the population into high-, median-, and low-income groups and find that the local participation rate of the high-income group is always the most responsive to friends' participation. I also find that Metropolitan counties are more responsive and more influential comparing to non-metropolitan counties.

[3] According to Egan et al. (2019), financial misconduct reports include: employment separation after allegations, customer dispute (settled), customer dispute (award/judgment), regulatory (final), civil (final), and criminal (final disposition).

Lastly, I examine the welfare implication of friends' participation. Households may enjoy the high equity premium upon participating in the stock market. However, recent literature also suggest that behavioral biases may be amplified through social interaction (Heimer (2016), Hirshleifer (2020)). I calculate county-level Gini coefficient as a proxy for income inequality and find that the increase in participation rates from the low-income households driven by friends' participation leads to lower income inequality in metropolitan areas in the following two years, while the increase in participation among the high-income households does not worsen income equality.

This paper contributes to previous literature in both peer effects in financial markets and the non-participation puzzle. Recent papers have shown that social connectedness helps explain how individuals form their beliefs on the local housing market, how institutions conduct investment decisions, and how P2P lendings spread in the US (Bailey et al. (2018a), Kuchler et al. (2020a), Allen et al. (2020)). On the other hand, previous attempts to explain the non-participation puzzle suffer from endogeneity problem and external validity problem due to the usage of specific surveys. This paper applies a causally reliable approach to connect the literature and illustrate how participation decisions of connected counties can be an important factor in local stock market participation using a data with holistic coverage of the entire US counties.

This paper also adds to the literature of belief and expectation formations. Malmendier and Nagel (2011) show that individuals extrapolate their past experience to form expectations and conduct financial decisions. This paper complements to this line of research by showing that socially connected others are also important sources for financial decisions.

Last but not least, this paper contributes to the recent literature about social finance. Does peer effect on financial decisions lead to more optimal investment as well as better welfare, or does the contagion in behavioral biases outweigh the positive spillover from peer effect? This paper shows how stock market participation among low-income households induced by friends' participation may be associated with better future income equality. To my understanding, this paper is the first to provide evidence on the overall pros and cons of peer effects on financial decisions.

The paper proceeds as follows. Section 2 describes the IRS Statistics of Income data and the Facebook Social Connectedness Index. I compare both of the stock market participation rates estimated from IRS Statistics of Income and Survey of Consumer Finance (SCF). Section 3 is the benchmark results where I show the association between friends' participation and local participation rate as well as establish causal inference using financial misconduct as an instrument variable. Section 4 explores the cross-section of spillover effect from friends' participation and its welfare implication. Section 5 concludes.

2 Data

2.1 IRS Statistics of Income and Survey of Consumer Finance

In the literature, researchers have relied on high-quality surveys to study stock market participation in the US (Campbell (2006), Hong et al. (2004)). The Survey of Consumer Finance (SCF) conducted by the Board of Governors of the Federal Reserve System is the most commonly used reference[4]. Every three years, SCF selects around 6,000 households that are representative to the entire US households and ask about their financial decisions. However, stock market participation rates across states or counties are unlikely to be homogenous. To identify how financial decisions between socially connected regions may be correlated or even affected by each other, it is crucial to observe participation rates at finer scale.

The Internal Revenue Service (IRS) Statistics of Income is a publicly available dataset that provides aggregated personal income tax information at state, county, and zipcode levels[5]. Chien and Morris (2017) use IRS Statistics of Income to estimate state-level participation rate. They estimate it as the ratio of number of tax filings with dividend tax payment over the total number of tax filings within a state. If one is paying dividend tax, this individual inevitably has to hold equity that yields dividend. Chien and Morris (2017) document strong cross-sectional differences in participation rates in 2014 across all income level. However, they do not attempt to resolve the puzzle but merely raise doubt against traditional preference-based or

[4]The SCF data is public available on the Federal Reserve Board website: https://www.federalreserve.gov/econres/scfindex.htm.
[5]The IRS Statistics of Income data is publicly available: https://www.irs.gov/statistics/soi-tax-stats-statistics-of-income.

cost-based explanation for non-participation puzzle. I follow their approach to use IRS Statistics of Income to estimate local participation and examine how local participation rates may spillover to connected region.

In this paper, I estimate the stock market participation rates at county level since county-level estimation provides much richer variations comparing to state level and suffer less from estimation errors comparing to zipcode level. The sample period is between 2010 and 2018.[6] Figure 1 is the US county-level stock market participation rate in 2018. The average participation rate is 16.25%, ranging from 0 to 67.96%, and the standard deviation is 6.84%. The cross-sectional differences in stock market participation rates at county-level are very substantial.

While local demographic variables may partially explain the cross-sectional differences in stock market participation rates across regions, much of the heterogeneity remains after conditioning on demographic variables. Figure 2 shows the stock market participation rates in Connecticut (the highest participating state), Mississippi (the lowest participating state), and the entire nation across different income brackets. The differences between participation rates between Connecticut and Mississippi remain substantial across all income groups. The stock market participation rate among households with income between $100,000 and $200,000 in Mississippi (18.34%) is lower than the participation rate among households with income between $75,000 and $100,000 in Connecticut (21.85%). Similar discrepancies can also be observed in lower income brackets. These empirical facts suggest that neither the cost-based or the preference-based hypotheses is sufficient on their own to fully explain the non-participation puzzle.[7]

[6]The number of dividend filing is available only after 2010. Before 2009, IRS Statistics of Income only reports the total dollar value of dividend tax for each county.

[7]Figure 2 is an updated version of figure 1 in Chien and Morris (2017) in which they show the substantial differences in participation rates after conditioning on income using the 2014 IRS Statistics of Income data.

2.2 What does the IRS participation rate capture?

One may have concerns using taxation data to estimate stock market participation rate for two reasons. First, households file dividend tax only when they receive dividends. It is possible that the stocks or ETFs held by investors do not payout. Second, investors may participate in the stock market in various different channels. People can directly hold individual stocks or ETFs which directly pay out dividends. They can also participate indirectly in the stock market by investing in the mutual fund market, or even more commonly, in the pension fund market. Not all mutual fund yield distributions to fund investors. Furthermore, individual's wealth in pension fund would not be taxed until they withdraw it. Therefore, the participation rate estimated from IRS Statistics of Income should be regarded as the lower bound of stock market participate rate. The question is, does the gap between this lower bound and the true participation rate reflect some unobserved confounding factors? Or alternative, does the gap stay relatively consistent across geographical regions? If the later holds, the heterogeneity in the participation rate estimated from taxation data contains meaningful information to study.

The first concern should be less of a problem for several reasons. First, since 2010, the year when the analysis starts, more than 80% of the stocks in S&P 500 pay dividend every year. Moreover, according to SCF2016, the median number of different stocks individuals hold is 5, a significant improvement from 2 in the 20^{th} century and 3 in SCF2001 (Campbell (2006)). Considering that retail investors are more likely to hold stocks with better visibility due to limited attention (Kahneman (1973)) along with the increasing number of different stocks individuals hold, the IRS Statistics of Income data should capture vast majority of individual's equity holdings.

In response to the second concern, about 20% of the mutual funds consistently distribute stock dividends to the investors.[8] Technically speaking, all mutual funds that invest in the equity market should pass down dividends, and the dividends distributed to the mutual fund investors, whether directly reinvested or not, is taxable. However, practically only funds with a dividend objective (value stocks, high payout

[8] According to the Investment Company Institute, about 20% of the domestic equity fund focus on value/ high payout stocks. https://www.ici.org/

ratio stocks) distribute to investors. Other funds only collect dividends sporadically and would use it to offset certain operation-related expenses. Combined with the statistics in the last paragraph, the stock market participation rate estimated from IRS Statistics of Income captures the vast majority of direct equity investment activities and about 20% of mutual fund investments.

To further validate the usage of IRS Statistics of Income, I compare the stock market participation rates from IRS Statistics of Income and SCF at national level, the only scale participation rate is available in SCF. Figure 3 compares the participation rate from IRS Statistics of Income and the participation rate as well as the ratio of direct stock holding in SCF. The SCF direct stock hold ratio in figure 3 swings from 21.3% in 2001 to 13.9% in 2016. The SCF participation rate also exhibits a weakly decreasing trend, from 53% in 2001 to 48.8% in 2013 and slightly bouncing back to 51.9% in 2016. The larger drop in direct stock holding and comparably the smaller drop in participation rate suggest that a substantial fraction of investors switch from direct stock holdings to passive investment instruments. On the other hand, the participation rate from IRS Statistics of Income swings from 24.69% in 2001 to 18.61% in 2018. The participation rate yielded from IRS Statistics of Income is consistently larger than the direct stock hold ratio in SCF, suggesting that the investment activity captured by IRS Statistics of Income is beyond retail investor's direct stock holding. The participation rate from IRS Statistics of Income also exhibits a salient decreasing trend similar to direct stock holding in SCF. However, the gap between the two also increase from less than 3% in 2001 to more than 5% in 2016, indicating that IRS Statistics of Income still captures a fraction of indirect market participation and part of the increasing trend of passive investments. In the time-serial dimension, the change in participation rates from IRS appears to be reasonable according to our understanding of its composition.

To investigate more into the properties of IRS Statistics of Income, we look at income cross-sections of national participation rates. Figure 4 compares the participation rates across different income levels in 2016 for SCF and IRS Statistics of Income. Both of the participation rates increase along with income in a concave manner, a finding consistent with the cost-based explanation for non-participation puzzle (Vissing-Jorgensen (2003), Cole et al. (2014)). The correlation coefficient of the SCF participation and IRS participation across the six income groups in 2016 is 96.2%. The correlation coefficients in previous waves of SCF are also extremely high (97.1% in 2010 and 96.4% in 2013). The pooled correlation coefficient across income groups within the past 3 waves of SCF is 96.5%. At income cross-section, IRS participation rate co-moves almost perfectly with the SCF participation rate at national level.

At national level, the participation rate estimated from IRS Statistics of Income contains legitimate properties consistent with our expectation from the literature and our understanding of its composition. It is undeniable that IRS Statistics of Income participation only represents the lower bound of stock market participation. Nevertheless, the IRS Statistics of Income data possesses unique and decisive cross-sectional variations at regional level that can help us study the transmission of stock market participating decisions across regions. It would be hard to imagine that the social connectedness effect we identify from IRS Statistics of Income data would reverse when we turn into overall participation rate. If anything, the effect I will show later in empirical section should be partial to the true effect social connectedness can have on overall participation rate in different regions.

2.3 Facebook Social Connectedness Index

I measure social connectedness between US counties with the Facebook Social Connectedness Index (SCI) provided by Bailey et al. (2018b). The Facebook Social Connectedness Index (SCI) is a county-to-county connectedness measurement based on the login information of Facebook users and their connected friends as of April 2016. Facebook is the most popular social media platform in the world. By 2019, around 69% of US adults report that they ever use Facebook, far exceeding Instagram (38%), LinkedIn (27%), Snapshot (24%), and Twitter (22%).[9] For any given county pair, county i and county j, Facebook construct the Social Connectedness Index (SCI) as the relative number of friendship links between county i and j, then scaled by the product of the numbers of Facebook users in the two counties.

[9]Information source: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

There are two important features in the Facebook SCI that make it ideal to study stock market participation decisions. First, though SCI is a one-time snapshot in 2016, this connection is indeed very persistent across time. Instead of regarding it as how individuals interact with each other through Facebook, SCI reflects more of a concrete underlying connection across regions. For one example, the college towns across the nation are all strongly connected to each other. In addition, SCI also exhibits patterns of past immigration. Counties in Wisconsin, Minnesota, and the Dakotas where Norway immigrants locate have strong connectedness with each other (see Bailey et al. (2018b) for more examples). Thanks to the great coverage of Facebook userbase in the US and its general social media purpose, Facebook SCI reflects more of the persistent relationship between US counties rather than merely Facebook utilization in 2016.

Second, SCI does not necessary reflect geographic distance. Take the Monroe County and Miami-Dade County at Florida as example. All of the top 10 connected counties of Miami-Dade County are in Florida. This may be because Miami is the largest city locally and attracts local residents from all regions in Florida. However, 5 out the top 10 connected counties of the Monroe County, a county just sitting next to Miami-Dade, are not in Florida. This connection may be due to the vibrate hospitality industry in the Monroe County. The spatial nature in SCI helps us rule out the alternative hypothesis that the spillover effect in our analysis may be simply picking up relationship due to geographic distance.

2.4 Other Demographic Variables and Summary Statistics

Lastly, we collect county-level demographic variables and local stock performance from the American Community Survey (ACS) and Compustat. ACS provides 5-year tracking of demographic variables for all US counties.[10] The county-year variables I collect from ACS are the number of households, percentage of individuals with degree higher than Bachelor degree as a proxy for education attainment, unemployment rate for local economy condition, and median household income for local income level.

[10]ACS also provide the 1-year tracking demographic variables every year but only for counties with more than 65,000 population. Though the 5-year data is less timely, the data comes with better precision and coverage.

Stock performance of local companies may be more salient to residents and thus might affect their willingness to participation in the stock market. In addition, literature have documented the ”home bias” phenomenon that investors tend to place higher weights on stocks of local enterprises in their portfolio (Coval and Moskowitz (1999), Baik et al. (2010), Bernile et al. (2015)). Therefore, local stock market returns can weakly proxy for the performance of the portfolio of existing local investors. Controlling for local stock market returns can take care of the effects how local stock performance or local peer's performance affect stock market participation (Kaustia and Knüpfer (2012)). I assign all US-based stocks to US counties according to their headquarter addresses. I then calculate the annual total log return for each stock in each year (nature log of the sum of fiscal-year-end price and fiscal-year dividend over the fiscal-year-end price in the last year) and value-weight the returns using firms' market capitalization.

Table 1 presents the summary statistics for the estimated county-level direct stock market participation rates and other demographic variables. The sample period is from 2010 to 2018. Together, we obtain 28,195 county-year observations. The average participation rate is 16.5% and the median is 16.2%. A county has an average population of 37,119 households and median population of 9,857. The average and median ratio of individuals with higher than Bachelor degree are 19.3% and 17.3%. On average, the median household in the county earns $47,079.4 every year. The mean and median local stock market returns are 2.2% and 0%.

To provide welfare implications, I also estimate the Gini coefficients at county level to proxy for income inequality. IRS Statistics of Income provides Aggregated Gross Income (AGI) information at different incomes brackets for each county, <$1, $1-$10K, $10K-$25K, $25K-$50K, $50K-$75K, $75K-$100K, $100K-$200K, and >$200K. I calculate the cumulative percentage of population and the cumulative percentage of total AGI for each bracket and estimate the Gini coefficients. The estimated Gini coefficients have an average of 0.403, ranging from 0.195 to 0.726. Though the income brackets from IRS Statistics of Income is coarse, the estimations appear to be quite reasonable. The state-level Gini coefficients estimated from IRS Statistics of Income in 2018 are Florida (0.561), New York (0.555), and California (0.520) ranking in the top and West Virginia (0.396), Alaska (0.406), and Iowa (0.411) in the bottom. After establishing the relationship between stock market participation rates across socially connected counties and the casual relationship, I will explore how stock market participation may be related to the improvement of income inequality.

3 Empirical Method

To investigate how local participation rate may be affected by the participation rates of socially connected counties, I construct a variable, "friends' participation ($FP_{i,t}$)", that equals to the weighted average participation rate of counties in the whole nation excluding counties that reside in the same state with county i, using the SCI as the weights. This approach is similar to testing the relationship between outcomes of individuals and the so-called "leave-out mean" in the peer effect literature (See Townsend (1994) for example). However, there is a substantial difference between the "leave-out mean" approach and the SCI-weighting approach. For the leave-out mean approach, local individuals face almost the same "leave-out means" with only trivial or negligible differences that barely identify the testing models. On the other hand, as mentioned in Section 2, geographically close counties may have very different SCI structure and thus may be connected to counties with dramatically different participation rates. Moreover, by excluding not just county i but also all counties in the same state with i, we alleviate the concern that we may be simply picking up geographical spillovers such as Kaustia and Knüpfer (2012) showed. Any effect that we identify later must come from geographically distant but socially connected counties.

$$FP_{i,t} = \frac{\sum SCI_{j\in\Omega\setminus I} * Participation_{j,t}}{\sum SCI_{j\in\Omega\setminus I}} \tag{1}$$

, where Ω is the full universe of US counties, and I is the set of counties in the same state with county i. Figure 5 is the heatmap of friends' participation in US counties in 2018.

3.1 Benchmark Regression

The benchmark model I analyze is a panel regression of the local participation rate at year t on the friends' participation at year $t-1$ along with demographic variables and fixed effects. If investors do rely on their connected others for financial decisions, we should expect local participation rates to be positively associated with their friends' participation.

$$Participation_{i,t} = \alpha + \beta_1 FP_{i,t-1} + \beta_2 Participation_{i,t-1} + \tilde{X}'\tilde{\gamma} + Ret_{i,t-1} + \rho_i + \phi_{I,t} + \epsilon_{i,t} \tag{2}$$

$$\Delta Participation_{i,t} = \alpha + \beta_1 \Delta FP_{i,t-1} + \beta_2 \Delta Participation_{i,t-1} + \tilde{\Delta X}'\tilde{\gamma} + Ret_{i,t-1} + \rho_i + \phi_{I,t} + \epsilon_{i,t} \tag{3}$$

Equation 2 and equation 3 are the models I test: $Participation_{i,t}$ is the participation rate of county i in year t; $FP_{i,t-1}$ is the SCI-weighted average participation rate from counties other than i in year $t-1$; $\tilde{X}$ contains the demographic variables including population, education attainment, unemployment rate, and median household income for county i in year t; $Ret_{i,t-1}$ is the most recent available local stock market return; ρ_i is the county fixed effect for county i; and $\phi_{I,t}$ is the state*year fixed effect in year t for state I in which county i locates. In Equation 2, all the variables are level variables. In equation 3, all the variables are changes (first differences).

Participation cost is one of the most popular explanations to the non-participation puzzle. Financial market is not frictionless. It is costly, both fiscally and cognitively, for investors to participation in the stock market. Assume for a fixed fiscal participation cost, investors with higher income would be less constrained comparing to investors with lower income and thus exhibit higher participating behaviors. In addition, better-educated investors tend to have better financial literacy, can digest relevant financial information with less cognitive resources, and thus participate in the stock market more vibrantly. The macroeconomic condition is also an important determinant for participation. When the overall economic condition is bad, households tend to be more constrained, leading to a higher participation cost and lower stock market participation.[11] The demographic control variables can help us capture the participation variations due to the effect of participation cost in various demographic dimensions.

As mentioned in the data section, local stock performance have been shown to be deterministic to stock market participating decision (Kaustia and Knüpfer (2012)). In addition, the home bias tendency leads investors to place higher weights on stocks of local companies in their portfolio (Coval and Moskowitz (1999), Baik et al. (2010), Bernile et al. (2015)). Therefore, local stock market return may also weakly proxy for the performance of existing investor's performance within the same county. Controlling for local stock market return may help us mitigate the concern that the effect of friends' participation may be associated with local stock performance.

County fixed effect, state*year fixed effect, and the lagged own participation rate are also important factors for stock market participation. County fixed effect captures the persistent component in county's financial circumstances. For equation 3, county fixed effect captures the long-term trend of changes in local participation rates. State*year fixed effect absorbs potential common shock to counties within the same state and in the same year. In the study how colleagues in the same workplace may affect each other's employee stock purchase plans (ESPPs), Ouimet and Tate (2020) argue that the time*residential area fixed effect they controlled for absorbs common shocks employees that live in the same area encounter together and is a valid identification strategy for causal inference by itself. Lastly, investors exhibit the so-called extrapolative behavior in which they rely their current financial decision on the recent experience. Malmendier and Nagel (2011) find that investors extrapolate their past experience to form belief about financial markets and conduct financial decisions. Therefore, controlling lagged participation on top of county fixed effect can capture the effect how investors reply on their recent decisions.

The inclusion of the demographic variables, local stock return, lagged participation, and the sets of fixed effects allow us to account for multiple alternative channels that are related to current participation decision. To further alleviate the concern that the level of participation rates may be persistent and exhibit spurious

[11] See Vissing-Jorgensen (2003), Christelis et al. (2010), Grinblatt et al. (2011), Van Rooij et al. (2011), and Cole et al. (2014) among others.

association between local participation and friends' participation, the main specification throughout the paper will be equation 3 where all the variables are in changes rather than levels. I study how $\Delta FP_{i,t-1}$, the change of friends' participation from year $t-2$ to year $t-1$ affects $\Delta\text{Participation}_{i,t}$, the change of local participation rate from year $t-1$ to year t. Nevertheless, results in equation 2 is still worth discussing since the effects of demographic variables on the level of participation rates are deeply connected to the cost-based literature. Table 2 panel A and B presents the results for the level and change in participation rates correspondingly. In all the analysis afterwards, I focus on only the change in participation, but all the results are robust using the level of participation.

In Table 2 panel A column (1) where I include only the demographic variables, the signs and the significance of the coefficients are consistent with our prior. One percentage point increase in the ratio of population with degrees higher than bachelor leads to 0.31 percentage point increase in local stock market participation rate; one percentage point increase in the unemployment rate leads to 0.39 percentage point decrease in local stock market participation rate; \$1,000 increase in the median household income leads to 0.14 percentage point increase in local stock market participation rate. The negative and significant coefficient for population simply reflects the mathematical identity that population is the denominator of participation rate. The coefficient for local stock market return is positive but insignificant. We shall see the effect of local stock market return when we move on to panel B. The adjusted R^2 is 48.0%, suggesting that the demographic variables along can explain close to half of the variations in participation rates.

In column (2), we add the friends' participation into the regression. The coefficient is positive and statistically significant. Before elaborating about the magnitude of the coefficient, the focus of column (2) is the improvement in the adjusted R^2. By including friends' participation, the adjusted R^2 increase from 48.0% in column (1) to 61.2%. Considering the fact that both friends' participation and local participation rate are persistent (AR(1) coefficients of 0.62 and 0.51), it would be hard to quantify how much variation in local participation rate is explained purely by the lagged friends' participation. However, it is obvious that friends' participation is able to explain variations that other demographic variables cannot account for.

From column (3) to column (6), county fixed effects and lagged participation are added into the regression. The inclusion of county fixed effects leads to an adjusted R^2 of more than 97%. However, it's hard to argue what the county fixed effects really stand for intuitively. Column (6) is the full testing model. Surprisingly, even though the lagged participation and county fixed effects explain almost all the variations in stock market participation rates, the coefficient for friends' participation stays positive and statistically significant. The economic magnitude is also large. One percent higher in FP is associated with 54.84 basis points higher local participation rate in the following year after controlling for the effect of income, education, economic condition, population, local stock performance, lagged participation, county fixed effects, and state*year fixe effects. The thought experiment here is that: for 2 counties located at the same state and in the same year, on top of their long-term average participation rates, the extrapolation from their participation decision in the last year, and all the effects from population, economic condition, income, education, and local stock return, the county that is socially connected to more high-participation counties in other states tend to have higher participation rate as well.

Instead of the level of participation rates, table 2 panel B reports regressions of the first difference of participation rates, $\Delta \text{Participation}_{i,t}$, on $\Delta FP_{i,t-1}$, the SCI-weighted average of change in participation rates of connected counties from outer states. I study 2 specifications: controlling for state*year fixed effects only and controlling both county fixed effects and state*year fixed effects. It is not clear whether we should include county fixed effect or not. First, when the dependent variable is the first difference of participation rate, controlling for county fixed effect is assuming the existence of a long-term trend in the change of participation and may incur look-ahead bias. Second, though the raw R^2 improved from 9.5% to 18.6% from column (1) to column (3), the adjusted R^2 decreases from 8.0% to only 3.2% since we include fixed effects for more than 3,000 counties. Still, the coefficients for $\Delta FP_{i,t-1}$ are consistently significant across specifications, ranging from 0.142 to 0.247. In the following tests, I will present results of both specifications but emphasize more on the specification with county fixed effect since it is more conservative.[12]

[12]An interesting side result is that the local stock market return now positively and significantly predicts the change in stock market participation rate in the following year. One standard deviation (25%) increase in local stock market return leads to

The economic interpretation is different from panel A. In the thought experiment, on top of all the effects from other determinants, an average of one-unit increment in participation rate in connected counties from year $t-2$ to year $t-1$ contributes to roughly 0.2-unit increment in the local participation rate from year $t-1$ to year t. To access the economic magnitude, consider the average number of tax filing in a county. Out of 37,000 households, the 1 percentage increase in friends' participation encourages 53 to 93 non-participants to start investing in the stock market. This strong and positive association suggests that the local stock market participation rate is positively correlated with the financial decision of the socially connected counties. Though the social connection network between counties in the US may be endogenous, documenting cross-sectional differences in local participation rates and showing that local participation rates correlate with the participation rates in connected regions are already two prominent findings.

3.2 Placebo Test

A possible concern brought up by Angrist (2014) is that the leave-out mean approach may by affected by confounding global factors that shock all the observations simultaneously. Friends' participation should suffer less from this critic since its variation is mainly driven by the SCI structure. In addition, the construction of FP excludes the participation rate in the same state, mitigating the possibility that FP may be affected by regional shocks that affect local participation at the same time. Last but not least, the state*year fixed effect should absorb common shocks at a finer dimension. That being said, I decide to take Angrist's concern seriously.

To examine whether SCI is the main driving factor of the relationship between $\Delta FP_{i,t-1}$ and change in local participation rate or that alternatively, the positive association is just a reflection of global shocks, I simulate 1,000 fake SCIs and estimate the ΔFP coefficients accordingly. The pooled SCI has a very unique distribution with mean of 2,318, standard deviation of 111,863, skewness of 178, and kurtosis of 50,675. To accommodate the special distribution, I simulate the fake SCIs using the same distribution without

0.5 basis point increase in local stock market return. However, the effect is too small to have economic meanings.

replacements. Essentially, I randomly shuffle the SCI scores for county pairs. For each set of the simulated SCI, I construct the ΔFP accordingly and test it in the benchmark model with the same control variables and fixed effects in Table 2 panel B column (4). If it were the global shocks rather than the unique structure of SCI that drives the result, we should expect to see a lot of simulated ΔFP to pick up the effect. Figure 6 is the distribution of the 1,000 coefficients for ΔFP according to the simulated SCIs. The coefficients range from -0.057 to 0.049, with mean of 0.0003 and standard deviation of 0.0133. None of the 1,000 coefficients is larger than, or even close to 0.142, the coefficient estimated from the true SCI. At the very minimum, it is safe to conclude that local stock market participation decision correlates with social connected counties from other states.

3.3 The effects from geographic and socially connected friends

To avoid capturing geographical spillover of stock market participation decisions, I exclude the counties within the same state when constructing the friends' participation. Therefore, the effect of friends' participation is purely driven by geographically distant but socially connected counties. However, it is also reasonable to expect financial decisions to be affected by geographical nearness. Thus, I construct another friends' participation variable purely from the counties within the same state, $FP_{i,t}^{SameState}$, and test its predictability on county i's participation decision in the following year.

$$FP_{i,t}^{SameState} = \frac{\sum SCI_{j \in I} * Participation_{j,t}}{\sum SCI_{j \in I}} \tag{4}$$

Table 3 presents the results where I include $\Delta FP^{SameState}$ into the regression. The same-state friends' participation also positively and significantly predicts local county's participation decision. This is somewhat surprising since I control for the state*year fixed effects. This suggest that the within-state connection structure also plays an important role. For random three counties a, b, and c in the same state, assume county a has an exogenous and significant increase in participation rate, on top of the common shocks all three counties face, county b that is more connected to a will enjoy a higher increase in stock market participation rate in the following year comparing to county c which is socially more distant to a.

In table 3 column (3), I include both $\Delta FP^{SameState}$ and ΔFP. The coefficient for $\Delta FP^{SameState}$ is 0.326, more than two times of the sensitivity between ΔFP and the change in local participation rate. Considering that the standard deviation for $\Delta FP^{SameState}$ (0.90%) is about three times of the standard deviation for ΔFP (0.29%), the economic magnitude of friends' participation from geographically close counties is about 6 times of the effect of friends' participation from outer-state counties. However, both $\Delta FP^{SameState}$ and ΔFP positively predicts the change in local stock market participation rates in the subsequent year after controlling for all the demographic variables and fixed effects.

3.4 Identification Strategy: Negative Shocks from Financial Misconducts

This paper thus far has illustrated a concrete association of participation rates between socially connected counties. However, to argue for a causal relationship, an identification strategy is still essential. Since SCI is time-invariant, what I can do is to look for incidents that shock local stock market participation rates and utilize the shock as an instrument.

To invest in the stock market, individuals have to trust the financial system at first place. Guiso et al. (2008) find that Dutch households who in general have less trust on others are also less likely to participate in stock markets. Gurun et al. (2018) find that after the Madoff scandal was revealed in 2008, the counties where more Madoff victims reside in experience an higher outflow from the registered investment advisors (RIA) and an increase in local deposit, suggesting that the households living in areas which experience "trust busts" are more likely to withdraw money from the stock market. Thus, I follow Egan et al. (2019) to look into reports of financial advisor misconducts and test if the revelation of misconducts affect local participation rates first then spillover to socially connected counties.

The financial misconduct incidents are collected from the Financial Industry Regulatory Authority (FINRA), a self-regulated organization authorized by the congress to protect US investors. Egan et al. (2019) classify the following six types of FINRA reports as financial misconducts: employment separation after allegations, customer dispute (settled), customer dispute (award/judgment), regulatory (final), civil (final), and criminal (final disposition). I aggregate the misconduct incidents up to county-year level according to the address where the financial advisory firms locate. In sum, I obtain 3,998 county-year observations that experience any financial misconduct incident reported between 2010 and 2017. The average population (158,155 households), income level (median household income of $52,173), education attainment (with 29.6% of population holding higher than Bachelor degrees), and participation rates (average of 19.9%) are all significantly higher than the full sample. This is consistent with the finding in Gurun et al. (2018) that Madoff targets victim with higher income and education and the finding in Egan et al. (2019) that misconduct advisors tend to cluster in certain advisory firms that constantly hire advisors with misconduct records. Although where potential misconduct advisors locate in the US may not be random, the exogenous shock that I utilize is the randomness of misconduct incidents being reported. Figure 7 is the heatmap of number of misconduct revelations at county level in 2017.

I construct a variable, the *Probability of Misconduct*$_{i,t}$, as the number of financial misconduct revealed at county i in year t divided by the household population (in thousands). Table 4 panel A is the panel regression of change in participation rates on the *Probability of Misconduct* along with demographic controls and fixed effects. The *Probability of Misconduct* in the same year and the lagged *Probability of Misconduct* significantly and negatively predicts the change in participation rates. 1 more financial misconduct incident revealed among a population of 1,000 household leads to 29.5 basis points decrease in the local participation rate during the same year and a further 31.1 basis points decrease in the following year. Considering how dividends are taxed in real practice, this result appears to be very reasonable. Residents of a county in which financial frauds are revealed may choose to withdraw their investments from the stock market and receive no dividend at the same year. It is also possible that the residents may have already received stock dividends by the time the fraud is discovered, postponing the effect to the following year. The *Probability of Misconduct* has no effect in the prior year, suggesting that we have a parallel pre-trend before the event taking place. The effect is statistically insignificant 2 years afterwards. This effect is characterized in figure 8. The revelation of financial misconduct does exogenously decrease local stock market participation rates.

To act as a valid instrument variable, *Probability of Misconduct* has to satisfy the exclusion restriction as well. In other words, *Probability of Misconduct* could not affect other counties' participation except through the channel of local county's participation decision. Considering that the majority of financial misconduct reports are small incidents such as customer disputes that would not be reported by national media rather than severe events like final dispositions of advisors due to criminal actions, the exclusion restriction should not be a problem.

I use the *Probability of Misconduct* in the same year as an instrument variable and estimate the predicted change in local participation using the specification in Table 4 panel A column (2). Lastly, I weight the predicted change in local participation by SCI and estimate the ΔFP^{IV}.

$$\Delta FP_{i,t}^{IV} = \frac{\sum SCI_{j \in \Omega \setminus I} * \Delta \widehat{Participation}_{j,t}}{\sum SCI_{j \in \Omega \setminus I}} \tag{5}$$

In Table 4 panel A, the inclusion of probability of misconduct in the same year has an F-stats of 22.16, and the adjusted R^2 improves from 3.2% to 6.0% from column (1) to column (2). The high F-stats and huge R^2 improvement both suggest that the probability of misconduct is not a weak instrument. One may still have concern using the probability of misconduct in the same year. Financial misconduct may in certain way be reversely driven by a lower participation rate. However, the randomness I am exploring here is not the "happening" of financial misconduct but the "revelation" of financial misconduct. Logically, it would be hard to imagine in any way that lower participation might drive more revelation events.

Table 4 panel B reports the results of ΔFP (OLS) and ΔFP^{IV}. In column (2), the coefficient for ΔFP^{IV} is a positive and significant 0.455, much higher than the OLS coefficients of 0.142. IV estimation represents a local effect due to the episode of events we are exploiting, while OLS estimation is the average effect of all variations in the independent variable of our interest and the dependent variable. There may be 2 reasons why we find a stronger local effect when exploiting the exogenous variations due to financial misconduct revelations. First, as mentioned in the beginning of this subsection, counties that experienced financial misconducts tend to have higher income, education level, and also higher population. These counties are more likely to be more "influential". Second, behavioral finance and experimental literature have taught us that agents respond asymmetrically stronger to negative information comparing to positive information (Kuhnen (2015) for example). Considering that ΔFP^{IV} is constructed using negative shocks, it's also reasonable to see that the effect is stronger.

To sum up, I have shown in this subsection that: (1) revelation of financial misconduct leads to lower local stock market participation rates in the same year and in the following year; (2) using this negative shock as an instrument, the instrumented variable ΔFP^{IV} still positively predicts the stock market participation rate in the local county. These results together suggest that the relationship between local financial decision and the financial decision of their socially connected others is causal.

4 Cross-section and welfare implication of friends' participation

4.1 Income Cross-section of friends' participation

In this section, I investigate the income cross-section of the documented effect. In the previous section, I have shown that friends' participation affect focal county's participation. The implication for the "non-participation puzzle" behind this finding is that people may not be participating simply because of the unawareness of the benefit. If this the case, we should expect that households of all income-level to be willing to participate upon hearing about the benefit, but the ones with less liquidity constraints could response more flexible accordingly.

County-level IRS Statistics of Income data reports tax filings for the following income brackets: less than \$1, \$1-\$10,000, \$10,000-\$25,000, \$25,000-\$50,000, \$50,000-\$75,000, \$75,000-\$100,000, \$100,000-\$200,000, and above \$200,000. To avoid estimation errors and unbalanced population, I aggregate the tax filings into 3 income groups: Group 1 for income less than \$50,000, Group 2 for \$50,000 to \$100,000, and Group 3 for income above \$100,000. I then estimate the participation rates at county-year-income level, $Participation_{i,t,G}$.

I test how $\Delta FP_{i,t-1}$, the same explanatory variable in the previous analyses, may have different effect on participation of different groups.

$$\Delta Participation_{i,t,G} = \alpha + \beta_1 \Delta FP_{i,t-1} + \beta_2 \Delta Participation_{i,t-1,G} + \tilde{X}'\tilde{\gamma} + Ret_{i,t-1} + \rho_i + \phi_{I,t} + \epsilon_{i,t} \quad (6)$$

In table 5, we can see that ΔFP consistently predicts participation rates across all income groups. Column (1) and column (5) are the same regression in table 2 panel B column (2) and column (4). What's more interesting is that, when we compare column (2) to (4) and column (6) to (8), we can see that the ΔFP coefficient is always the largest when the dependent variable is the participation rate of the high-income group. This finding is consistent with the participation cost explanation of the non-participation puzzle, especially the liquidity-constrained argument. Given the same set of information, the high-income group in the local county always respond the most. This may be due to the inability to respond for the low-income households. Even though they may hear from their peers about the benefit of investing, the low-income households are too constrained to react and participate in the stock market. This result is consistent with the hypothesis that non-participation can be attributed to unawareness to the benefit of equity premium and/or inability to participate. Friends' participation leads to stock market participating decision and is more prominent among households who are less constrained to respond.

4.2 Metropolitan versus non-metropolitan areas

In the peer effect literature, it's nature to ask who are the ones more likely to response and who are the ones that are more influential. The income-cross-section test in subsection above addresses the first part. To answer the second part, we can sort counties into subsamples and test how each group have different strength to affect the rest of the counties in this nation. I follow the 2013 Rural-Urban Continuum Codes defined by the Office of Management and Budget (OMB) to classify counties into metropolitan counties and non-metropolitan counties.[13] I then construct the friends' participation from metropolitan counties and non-metropolitan counties and test their predictabilities on local participation rates.

$$\Delta FP_{i,t}^{Metro} = \frac{\sum SCI_{j\in\Omega\setminus I} * \Delta Participation_{j,t}^{Metro}}{\sum SCI_{j\in\Omega\setminus I}} \tag{7}$$

$$\Delta FP_{i,t}^{NonMetro} = \frac{\sum SCI_{j\in\Omega\setminus I} * \Delta Participation_{j,t}^{NonMetro}}{\sum SCI_{j\in\Omega\setminus I}} \tag{8}$$

Table 6 panel A reports the results using ΔFP^{Metro} and $\Delta FP^{NonMetro}$ as the independent variable predicting the participation rates of the entire US counties. When we control for only state*year fixed effects, both ΔFP^{Metro} and $\Delta FP^{NonMetro}$ positively and significantly predicts the change in local participation. However, when we control for county fixed effect that proxy for the long-term trend in the change of participation rates, only the coefficients for ΔFP^{Metro} remain statistically significant, implying that the friends' participation from metropolitan areas may be more influential.

Table 6 panel B reports the result using ΔFP, the friends' participation from all outer-state counties in the US to predict local participation rates in metropolitan or non-metropolitan counties only. ΔFP significantly predicts participation rates from both metropolitan or non-metropolitan counties. The coefficients for ΔFP when predicting metropolitan counties are much larger than the coefficients when predicting non-metropolitan counties, suggesting that metropolitan counties are more responsive to friends' participation decisions. In addition, the R^2s for metropolitan counties and non-metropolitan counties are very different.

[13] The classification can be found on United States Department of Agriculture - Economic Research Service website: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/

The adjusted R^2s in column (1) and column (2) are 33.2% and 37.3%, while the adjusted R^2s in column (3) and column (4) are 5.9% and -0.3%. The high R^2s in column (1) and column (2) suggest that the changes in participation rates in metropolitan areas are more attributable comparing to non-metropolitan areas. More studies on how households in non-metropolitan areas conduct their financial decision is needed.

Table 6 panel C reports the cross predictability of ΔFP^{Metro} and $\Delta FP^{NonMetro}$ on metropolitan or non-metropolitan counties with or without county fixed effects. All combinations except column (2) and column (8) possess positive and significant predictabilities. In column (2) where ΔFP^{Metro} predicts metropolitan-county participation rate and column (8) where $\Delta FP^{NonMetro}$ predicts non-metropolitan-county participation rate with county fixed effects, the coefficients are positive but insignificant. These results seem to suggest that county fixed effects, the long-run trend in the change of participation rates, for metropolitan counties (non-metropolitan counties) correlates with ΔFP^{Metro} ($\Delta FP^{NonMetro}$) but not the other way around.

To sum up, I find that: (1) friends' participation from metropolitan counties are more influential than that of non-metropolitan counties; (2) metropolitan counties are more responsive to friends' participation comparing to non-metropolitan counties; (3) changes in participation rates in metropolitan or non-metropolitan counties contain very different properties that need to be studied separately; and (4) friends' participation from metropolitan (non-metropolitan) counties correlates with the long-run trend in the change of participation rates in metropolitan (non-metropolitan) counties, suggesting a distinct property in participation decisions between metropolitan and non-metropolitan counties.

4.3 Income inequality and stock market participation

The core welfare implication behind the non-participation puzzle is that the lack of participation among the low-income households lead to larger income inequality. The literature regard stock market participation to be beneficial since households could not take advantage of the high equity premium had they not participated in the stock market (see Duflo and Saez (2002), Hong et al. (2004), and Ouimet and Tate (2020) for example). However, retail investors also suffer from behavioral biases such as overconfidence in their investing ability, skewed preference that leads retail investors to hold lottery-like stocks with negative average returns, and representativeness heuristic when discussing about their portfolio choices with their friends. The behavioral biases may be further augmented through social connection (Heimer (2016), Hirshleifer (2020), and Bali et al. (2020)). Kogan et al. (2020) also show in a rational setting that the benefits of technological improvement may not necessarily flow to stock holders and lead to less income inequality. In this subsection, I calculate Gini coefficients at county-year level using the IRS Statistics of Income data and test directly how participation of different income groups may be related to income inequality.

In order to identify the driving force behind the participation rates of different income groups, I estimate the predicted value of change in participation rates of different income groups from the change in friends' participation using the specifications in Table 5 column (6) to (8). In table 7, I run the panel regression of change in Gini coefficients on the predicted value as well as the residuals of changes in participation of different income groups along with demographic control variables, county fixed effect, state*year fixed effects, and the lagged change in Gini coefficient. Considering the findings in table 6 that metropolitan counties and non-metropolitan counties exhibit very different properties in stock market participation, I also study their income inequality separately. The average Gini coefficients for metropolitan counties and non-metropolitan counties are 0.406 and 0.401, not significantly different from each other though.

In column (1) to (3). the dependent variables are the change in Gini coefficients from year $t-1$ to year t in the entire US counties, the non-metropolitan counties, and the metropolitan counties. The only significant coefficient is the predicted change in participation rate among the low-income group in the metropolitan counties. The effect of stock market participation on income inequality may not take place very timely. In column (4) to (6), the dependent variables are the change in Gini coefficients one year forward. The coefficients for $\Delta\widehat{par_{i,t-1,low}}$ are all negative and significant for the change in Gini coefficients in metropolitan counties. On the other hand, the coefficients for $\Delta\widehat{par_{i,t-1,high}}$ are all positive and significant in column (4) and (6). In column (7) to (9), the dependent variables are the change in Gini coefficients from $t-1$ to year $t+1$. We can see that the change in participation rates in the low-income group driven by friends' participation do lead to lower income inequality in the longer horizon for metropolitan counties.

For household to benefit from the stock market, intensive margin is as important as extensive margin. I have shown in section 3 that friends' participation increases focal county's participation rate. However, I cannot tell whether households are investing wisely in response to their friends. According to section 4.2, stock market participation has dramatically different patterns between metropolitan counties and non-metropolitan counties. The fact that the increase in low-income households' participation contribute to better income equality in metropolitan counties may suggest that low-income households in metropolitan areas are investing more optimally. How investment decisions differ across different types of investors according to their opinion peers is one potential direction the following literature could investigate into.

The results from subsection 4.1 to subsection 4.3 provide very direct policy implication. While the high-income group is the most responsive, all income groups react to friends' participation. Albeit the fact that low-income households are less responsive to friends' participation, their participating decision due to friends' participation may lead to lower income inequality in metropolitan counties. One percentage point increase in the predicted low-income participation rate lead to 0.011 decrease in the Gini coefficient, about 2.7% to an average Gini coefficient of 0.406 in metropolitan areas. The results indicate that promoting a higher participation rate among low-income households may lead to better income equality.[14]

[14] A potential concern in this analysis is that the data sample period is between 2010 and 2018, an economic expansion cycle. Whether or not stock market participation among the low-income households during crisis period leads to lower income inequality remains a question.

5 Conclusion

The non-participation puzzle is one of the most intriguing question in household finance. In spite of the high equity premium, a substantial fraction of households across all income level do not invest. This under-participation may result in lower aggregate output and exacerbate income inequality. Using the IRS Statistics of Income data, I document geographic heterogeneity in stock market participation rates that traditional aims to explain the non-participation puzzle do not predict.

Using the Facebook Social Connectedness Index (SCI) as the weights, I show that friends' participation (FP), the weighted average participation rates of the socially connected counties is a strong predictor of local participation rates after controlling for past participation, demographic effects, county fixed effect, and the state*year fixed effect. The friends' participation also provides substantial improvement in R^2 on top of the explanatory power of demographic variables per se. To access causal inference, I illustrate how the revelation of financial misconducts, a negative local shock, can act as a valid instrument variable. The revelation of financial misconducts decreases local stock market participation, and the instrumented friends' participation still predicts local participation in the subsequent year.

I then explore the cross-section of FP effect and its welfare implication. The high-income households are more responsive to FP, a finding is consistent with the participation cost argument. The metropolitan areas are more influential and more responsive comparing to the non-metropolitan areas. Lastly, FP-induced participation in low-income groups leads to lower income inequality in metropolitan counties. These findings not only help us understand more about the fundamentals of individual financial decisions, but also provides policy implications how stock market participation may promote better social welfare.

Reference

Allen, L., Peng, L., and Shan, Y. (2020). Social networks and credit allocation on fintech lending platforms. working paper.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

Baik, B., Kang, J.-K., and Kim, J.-M. (2010). Local institutional investors, information asymmetries, and equity returns. *Journal of Financial Economics*, 97(1):81–106.

Bailey, M., Cao, R., Kuchler, T., and Stroebel, J. (2018a). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6):2224–2276.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., and Wong, A. (2018b). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80.

Bailey, M., Dávila, E., Kuchler, T., and Stroebel, J. (2019). House price beliefs and mortgage leverage choice. *Review of Economic Studies*, 86(6):2403–2452.

Bailey, M., Farrell, P., Kuchler, T., and Stroebel, J. (2020). Social connectedness in urban areas. *Journal of Urban Economics*, 118(7):103264.

Bali, T. G., Hirshleifer, D., Peng, L., and Tang, Y. (2020). Attention, social interaction, and investor attraction to lottery stocks. working paper.

Barberis, N., Huang, M., and Thaler, R. H. (2006). Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American Economic Review*, 96(4):1069–1090.

Bernile, G., Kumar, A., and Sulaeman, J. (2015). Home away from home: Geography of information and local investors. *Review of Financial Studies*, 28(7):2009–2049.

Bhamra, H. S. and Uppal, R. (2019). Does household finance matter? small financial errors with large social costs. *American Economic Review*, 109(3):1116–54.

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

Chien, Y. and Morris, P. (2017). Household participation in stock market varies widely by state. *The Regional Economist*, 3:4–5.

Christelis, D., Jappelli, T., and Padula, M. (2010). Cognitive abilities and portfolio choice. *European Economic Review*, 54(1):18–38.

Cole, S., Paulson, A., and Shastry, G. K. (2014). Smart money? the effect of education on financial outcomes. *Review of Financial Studies*, 27(7):2022–2051.

Coval, J. D. and Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance*, 54(6):2045–2073.

Duflo, E. and Saez, E. (2002). Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics*, 85(1):121–148.

Egan, M., Matvos, G., and Seru, A. (2019). The market for financial adviser misconduct. *Journal of Political Economy*, 127(1):233–295.

Garlappi, L., Uppal, R., and Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1):41–81.

Grinblatt, M., Keloharju, M., and Linnainmaa, J. (2011). Iq and stock market participation. *Journal of Finance*, 66(6):2121–2164.

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

Guiso, L. and Sodini, P. (2013). Household finance: An emerging field. In *Handbook of the Economics of Finance*, volume 2, pages 1397–1532. Elsevier.

Gurun, U. G., Stoffman, N., and Yonker, S. E. (2018). Trust busting: The effect of fraud on investor behavior. *Review of Financial Studies*, 31(4):1341–1376.

Haliassos, M. and Bertaut, C. C. (1995). Why do so few hold stocks? *Economic Journal*, 105(432):1110–1129.

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

Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance. *Journal of Finance*, 75(4):1779–1831.

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

Kahneman, D. (1973). *Attention and effort*, volume 1063. Citeseer.

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

Kogan, L., Papanikolaou, D., and Stoffman, N. (2020). Left behind: Creative destruction, inequality, and the stock market. *Journal of Political Economy*, 128(3):855–906.

Kuchler, T., Li, Y., Peng, L., Stroebel, J., and Zhou, D. (2020a). Social proximity to capital: Implications for investors and firms. working paper.

Kuchler, T., Stroebel, J., and Russel, D. (2020b). The geographic spread of covid-19 correlates with structure of social networks as measured by facebook. working paper.

Kuhnen, C. M. (2015). Asymmetric learning from financial information. *Journal of Finance*, 70(5):2029–2062.

Malmendier, U. and Nagel, S. (2011). Depression babies: do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics*, 126(1):373–416.

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

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

Townsend, R. M. (1994). Risk and insurance in village india. *Econometrica*, 62(3):539–591.

Van Rooij, M., Lusardi, A., and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 101(2):449–472.

Vissing-Jorgensen, A. (2003). Perspectives on behavioral finance: Does" irrationality" disappear with wealth? evidence from expectations and actions. *NBER Macroeconomics Annual*, 18:139–194.

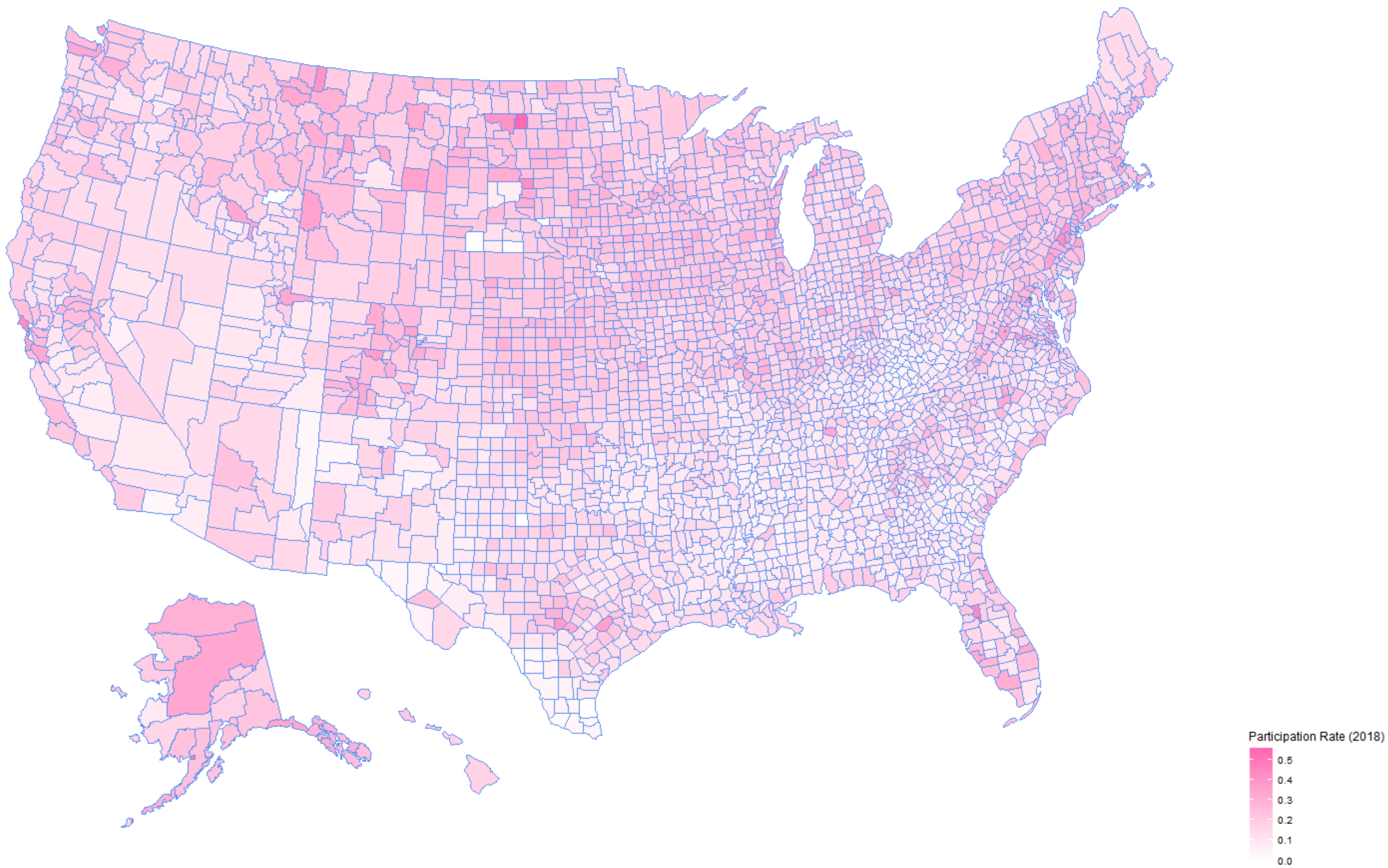

Figure 1: Stock market participation rates of US counties in 2018

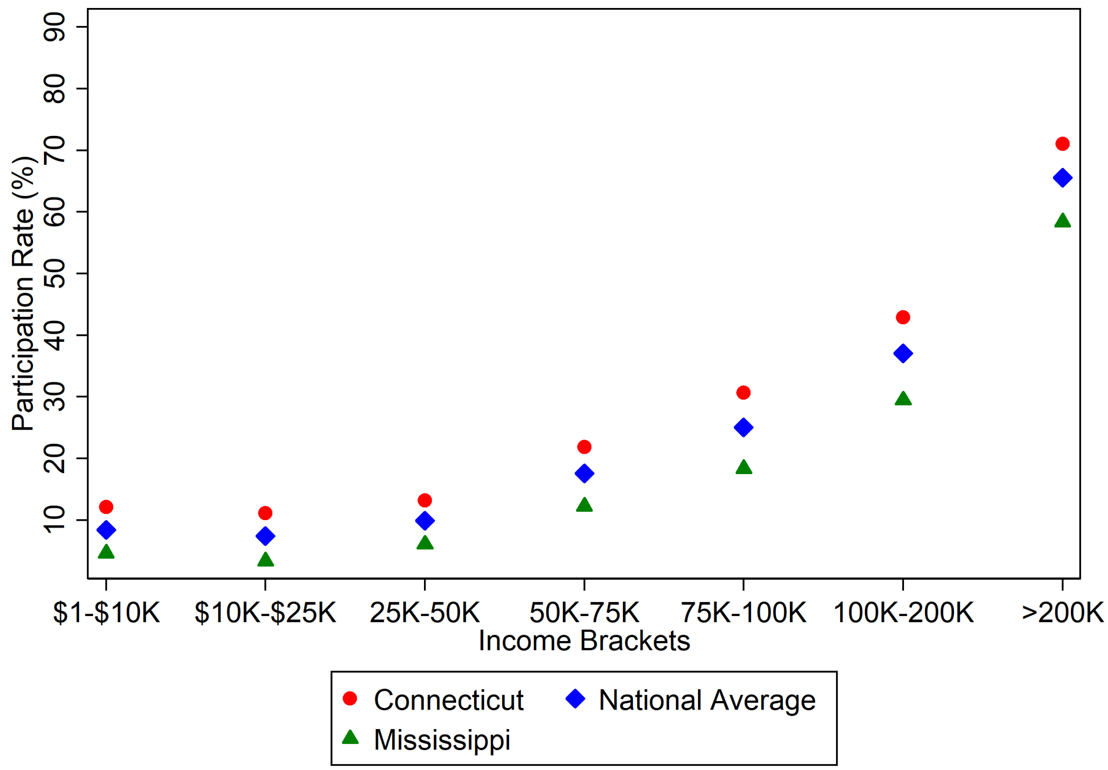

Figure 2: Stock market participation rates by income in Connecticut, Mississippi, and entire nation in 2018

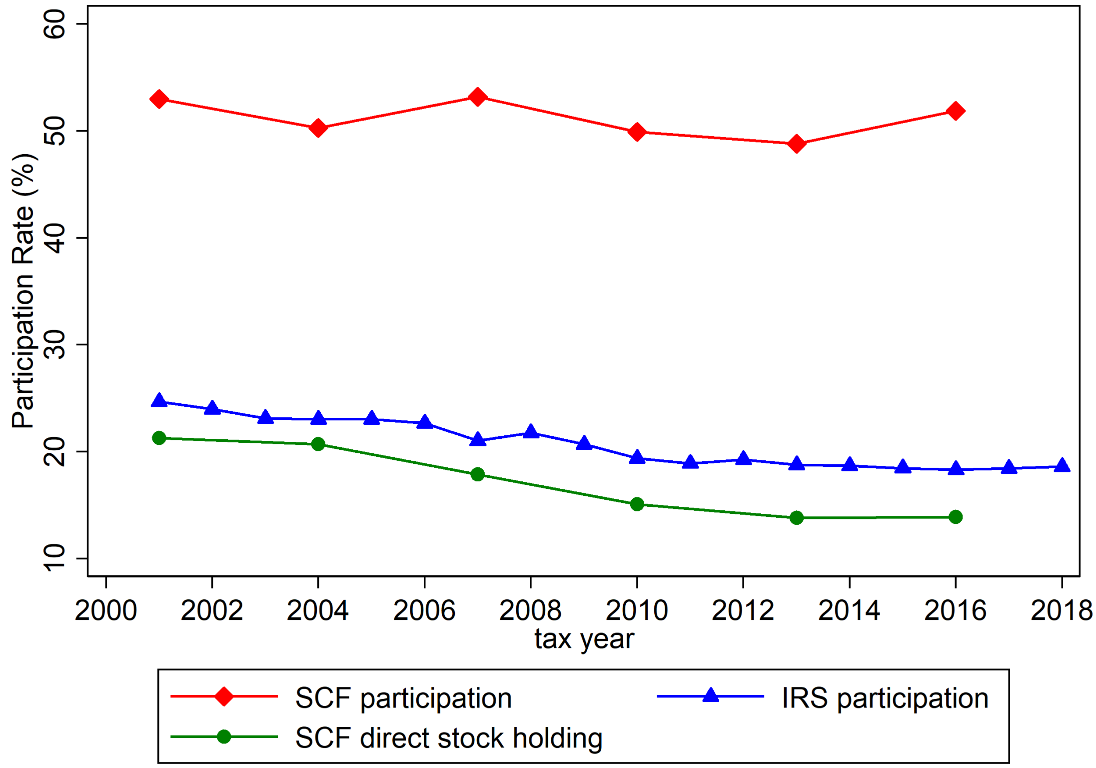

Figure 3: National stock market participation from 2001 to 2018

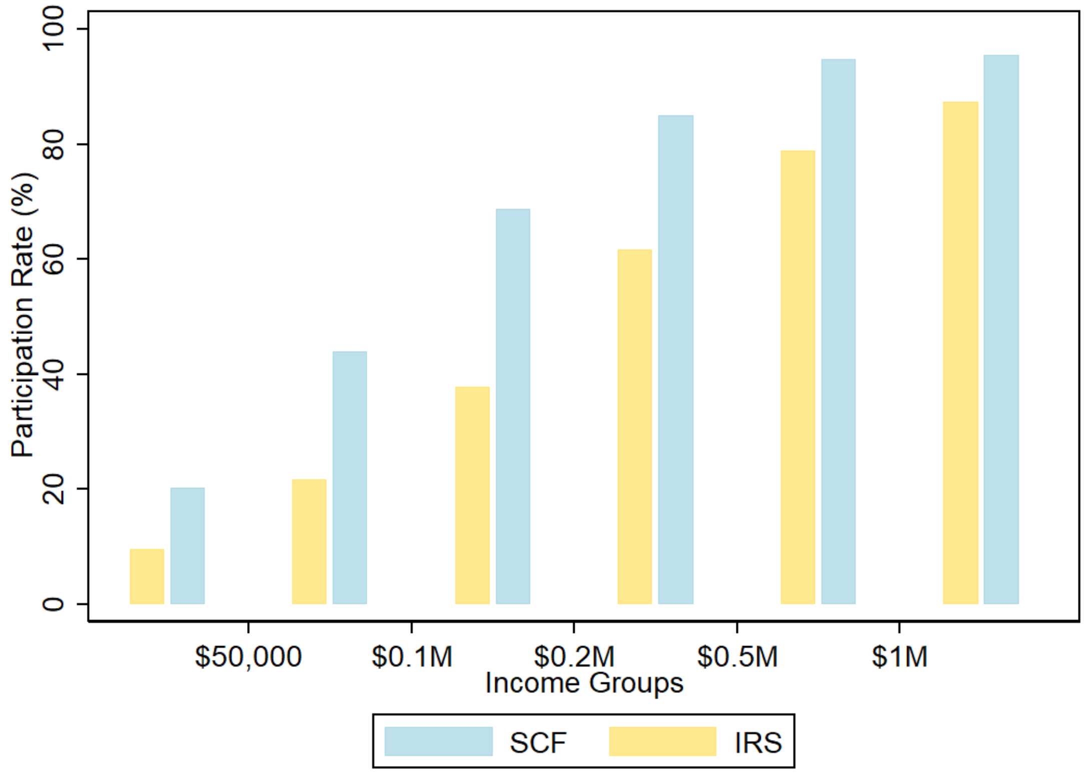

Figure 4: Income cross-section of national stock market participation in 2016

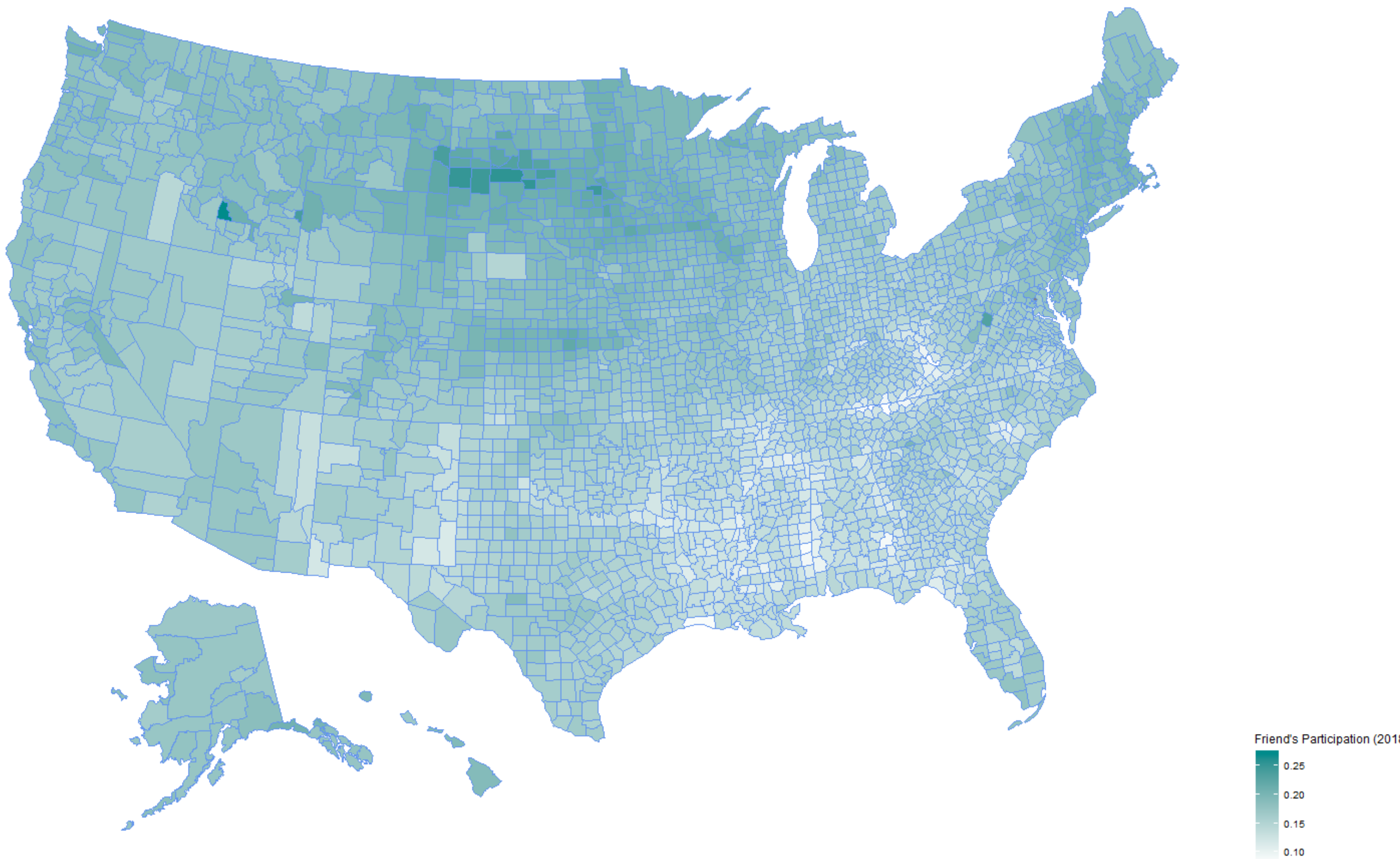

Figure 5: friends' participation of US counties in 2018

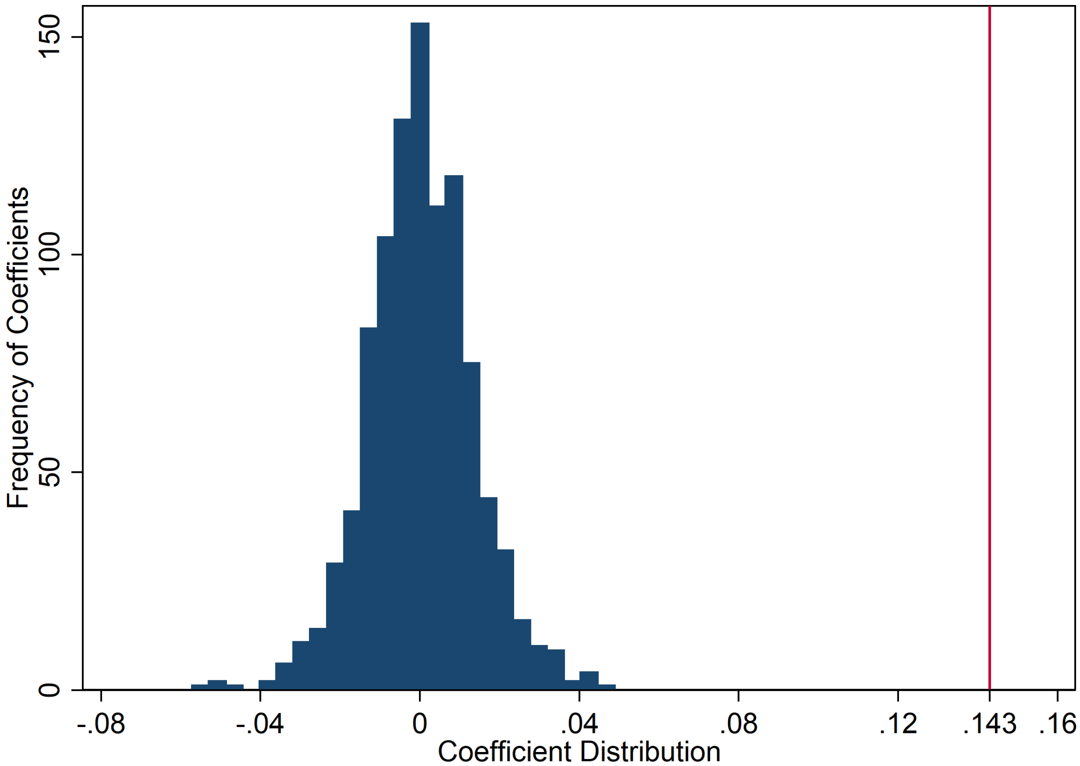

Figure 6: Placebo test: ΔFP coefficient distribution of the 1,000 simulated SCIs

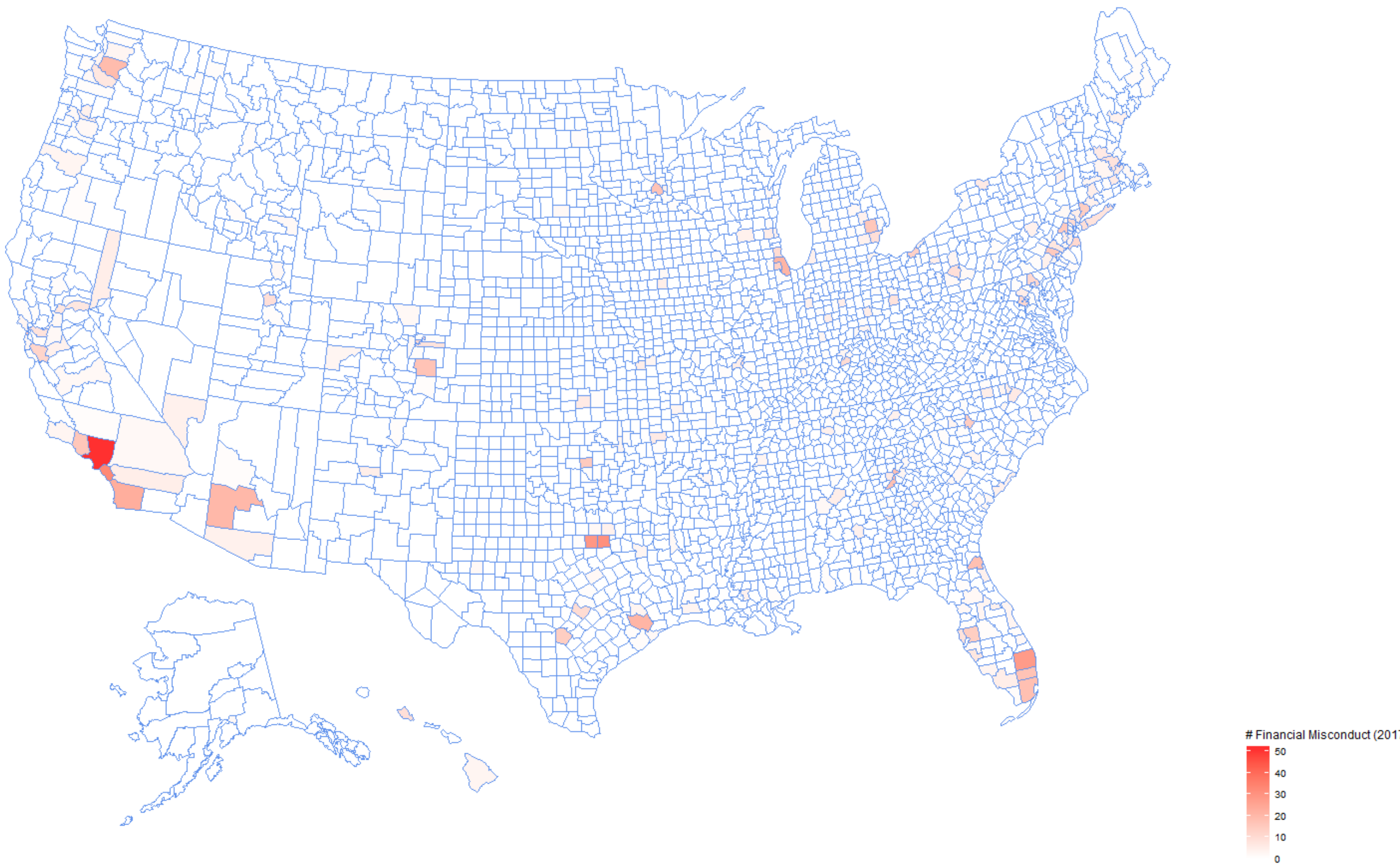

Figure 7: Number of financial misconduct revelation in US counties in 2017

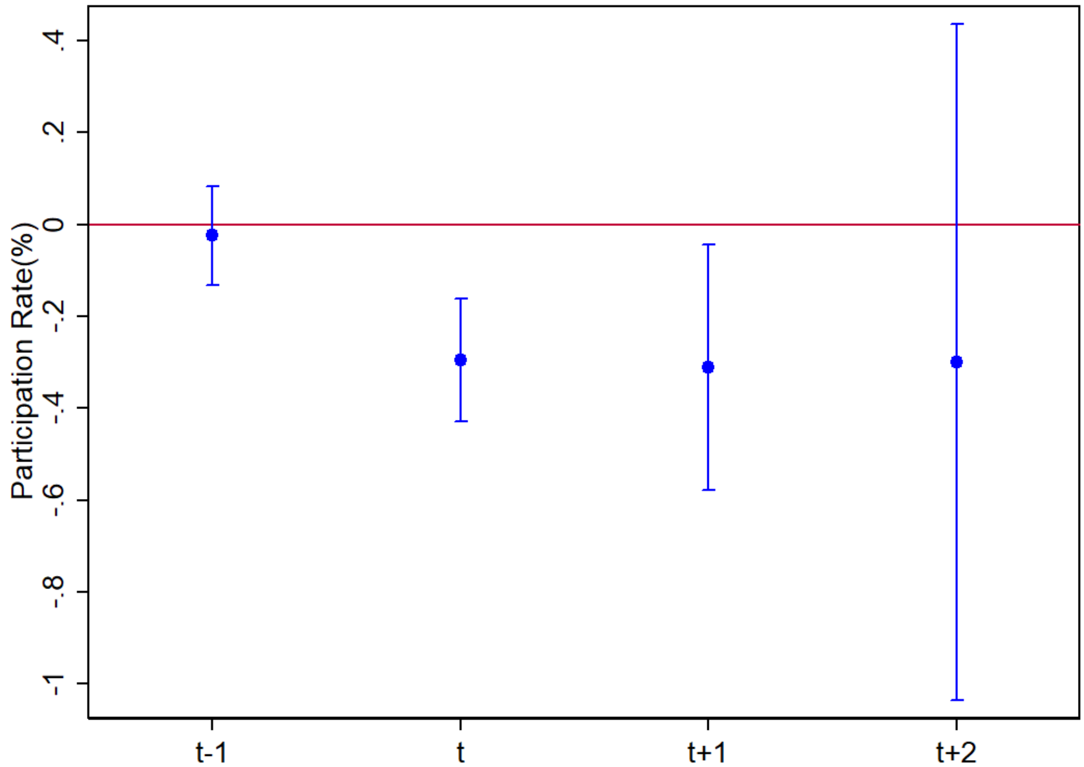

Figure 8: Local participation response to revelation of financial misconduct

Table 1: Summary statistics

Table 1 reports the summary statistics of the main variables. The sample period is from 2010 to 2018. The stock market participation rates are estimated at county level as the fraction of tax filers with dividend tax. County-level participation rates are reported for different income brackets (<$50K, $50K-$100K, >$100K) within each county and metropolitan or non-metropolitan counties. Friends' participation are estimated as the Facebook SCI-weighted average participation rates from outer-state counties. Both the summary statistics for the level and the first difference of friends' participation are reported. Summary statistics for the change in friends' participation in the same state, the instrumented friends' participation, the metro/non-metro friends' participation are also reported. Probability of Misconduct is the number of financial misconducts reported to FINRA in each county each year scaled by thousand population. This variable spans from 2010 to 2017. County-level Gini coefficients are winsorized at 1% and 99% in each year. The number of households, ratio of population with Bachelor degree, unemployment rates, and median household income are from American Community Survey. The local stock return is the value-weighted annual total log return of companies with headquarters located within each county.

Variable	Obs	Mean	Median	Std.	Min	Max
Participation Rate (%)	28,195	16.5	16.2	6.7	0	79.6
Participation Rate (<$50K) (%)	28,195	10.5	10.1	5.2	0	83.2
Participation Rate ($50K-$100K) (%)	28,195	21.6	21.5	7.3	0	83.3
Participation Rate (>$100K) (%)	28,195	40.0	40.6	10.9	0	92.3
Participation Rate (Metropolitan) (%)	10,458	17.4	16.9	6.5	1.5	60
Participation Rate (Non-Metropolitan) (%)	17,737	15.9	15.6	6.8	0	79.6
FP (%)	28,195	16.5	16.5	2.5	7.8	28.1
ΔFP (%)	25,061	-0.10	-0.09	0.29	-6.26	14.73
$\Delta FP^{SameState}$(%)	25,061	-0.11	-0.12	0.90	-49.59	35.51
ΔFP^{IV}(%)	18,796	-0.12	-0.12	0.09	-3.82	1.28
ΔFP^{Metro}(%)	25,062	-0.10	-0.12	0.37	-1.45	19.30
$\Delta FP^{NonMetro}$(%)	25,062	-0.10	-0.08	0.28	-6.68	9.95
Probability of Misconduct	25,063	0.009	0	0.058	0	6.816
Gini Coefficient	28,195	0.403	0.394	0.074	0.195	0.726
# of Households	28,195	37,118.6	9,857	112,421.6	22	3,306,109
Ratio higher than Bachelor Degree (%)	28,195	19.3	17.3	8.8	0	80.2
Unemployment Rate (%)	28,195	7.7	7.3	3.6	0	30.9
Median Household Income	28,195	47,079.4	45,166	12,489.4	18,972	136,268
Local Stock Return (%)	28,195	2.2	0	25.1	-662.7	782.4

Table 2: friends' participation and local stock market participation rate

Table 2 reports the panel regression of county-level stock market participation rates on its friends' participation rate ($FP_{i,t-1} = \frac{\sum SCI_{j\in\Omega\setminus I} * Participation_{j,t-1}}{\sum SCI_{j\in\Omega\setminus I}}$, where $\Omega \setminus I$ is the set of all US counties except counties in state I in which county i resides.) in the previous year along with demographic control variables and fixed effects, including the participation rate in the previous year, population, education attainments, unemployment rates, median household income, local stock returns, county fixed effect, and state*year fixed effects. Population and income are scaled by 1,000. Education attainment is the percentage of population with degrees higher than Bachelor. Unemployment rate and local stock return are also in percentage points. Panel A is the result for level of participation. Panel B is the result for the change (first difference) of participation. In the parentheses are the t-statistics calculated from standard errors clustered at county level.*p<.1; **p<.05; ***p<.01 for the main variables.

Panel A. Level of participation

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent: Participation$_{i,t}$ (%)					
FP$_{i,t-1}$ (%)		1.1848***		1.0324***		0.5484***
		[30.54]		[5.81]		[4.76]
Participation$_{i,t-1}$ (%)					0.4600***	0.4297***
					[5.67]	[5.29]
Population/1,000$_{i,t}$	-0.0052***	-0.0041***	0.0083**	0.0097**	0.0057**	0.0066**
	[-4.17]	[-4.63]	[2.00]	[2.18]	[2.02]	[2.13]
Education$_{i,t}$ (%)	0.3083***	0.2181***	-0.0216***	-0.0229***	-0.0189***	-0.0198***
	[25.35]	[18.95]	[-2.87]	[-3.15]	[-3.24]	[-3.44]
Unemploy$_{i,t}$ (%)	-0.3927***	-0.1424***	0.0106	0.0071	0.0033	0.0019
	[-10.91]	[-4.00]	[1.03]	[0.75]	[0.35]	[0.21]
Income/1,000$_{i,t}$	0.1415***	0.1115***	0.0085	0.0079	0.0089*	0.0085
	[11.77]	[10.52]	[1.19]	[1.14]	[1.70]	[1.63]
Return$_{i,t-1}$ (%)	0.0001	0.0010	-0.0001	-0.0001	0.0001	0.0001
	[0.07]	[1.10]	[-1.09]	[-1.05]	[0.66]	[0.57]
Intercept	6.9311	-11.4557	16.0307	-1.0187	8.5142	-0.0463
	[10.77]	[-12.72]	[37.76]	[-0.35]	[6.07]	[-0.03]
County FE			Y	Y	Y	Y
State*Year FE			Y	Y	Y	Y
Obs	25060	25059	25060	25059	25059	25059
R^2	0.480	0.612	0.975	0.976	0.980	0.980
adj. R^2	0.479	0.612	0.971	0.972	0.977	0.977

Panel B. Change of participation

	(1)	(2)	(3)	(4)
	Dependent: ΔParticipation$_{i,t}$ (%)			
ΔFP$_{i,t-1}$ (%)		0.2472***		0.1417*
		[3.59]		[1.86]
ΔParticipation$_{i,t-1}$ (%)	-0.1356***	-0.1441***	-0.1942***	-0.1988***
	[-3.08]	[-3.11]	[-3.17]	[-3.11]
ΔPopulation/1,000$_{i,t}$	0.0117**	0.0121**	0.0077	0.0080
	[2.47]	[2.50]	[1.02]	[1.07]
ΔEducation$_{i,t}$ (%)	-0.0193***	-0.0190***	-0.0175***	-0.0172***
	[-3.31]	[-3.25]	[-2.79]	[-2.75]
ΔUnemploy$_{i,t}$ (%)	-0.0045	-0.0047	-0.0077	-0.0078
	[-0.47]	[-0.51]	[-0.74]	[-0.76]
ΔIncome/1,000$_{i,t}$	0.0032	0.0039	0.0015	0.0020
	[0.49]	[0.59]	[0.19]	[0.25]
Return$_{i,t-1}$ (%)	0.0002**	0.0002**	0.0002**	0.0002**
	[2.08]	[2.09]	[2.10]	[2.10]
Intercept	-0.1050	-0.0760	-0.1103	-0.0937
	[-8.78]	[-6.35]	[-8.58]	[-9.27]
County FE			Y	Y
State*Year FE	Y	Y	Y	Y
Obs	21924	21924	21924	21924
R^2	0.095	0.097	0.186	0.187
adj. R^2	0.080	0.082	0.032	0.033

Table 3: friends' participation of within- and outer-state friend counties

Table 3 reports the panel regression of county-level stock market participation rates on its friends' participation rates from the same state and/or from outer-state counties in the previous year along with demographic control variables and fixed effects, including the participation rate in the previous year, population, education attainments, unemployment rates, median household income, local stock returns, county fixed effect, and state*year fixed effects. Outer-state friends' participation, $FP_{i,t-1} = \frac{\sum SCI_{j\in\Omega\setminus I} * Participation_{j,t-1}}{\sum SCI_{j\in\Omega\setminus I}}$, where $\Omega \setminus I$ is the set of all US counties except counties in state I in which county i resides; within-state friends' participation, $FP^{SameState}_{i,t-1} = \frac{\sum SCI_{j\in I} * Participation_{j,t-1}}{\sum SCI_{j\in I}}$, where I is the set of all counties in the state county i resides. Population and income are scaled by 1,000. Education attainment is the percentage of population with degrees higher than Bachelor. Unemployment rate and local stock return are also in percentage points. In the parentheses are the t-statistics calculated from standard errors clustered at county level.*p<.1; **p<.05; ***p<.01 for the main variables.

	(1)	(2)	(3)
	Dependent:	ΔParticipation$_{i,t}$ (%)	
ΔFP$^{SameState}_{i,t-1}$ (%)	0.2293***	0.3262*	0.3256*
	[2.71]	[1.75]	[1.73]
ΔFP$_{i,t-1}$ (%)			0.1400*
			[1.93]
ΔParticipation$_{i,t-1}$ (%)	-0.2947***	-0.4221***	-0.4261***
	[-5.32]	[-4.84]	[-4.92]
ΔPopulation/1,000$_{i,t}$	0.0146***	0.0078	0.0081
	[2.93]	[0.96]	[1.00]
ΔEducation$_{i,t}$ (%)	-0.0197***	-0.0178***	-0.0176***
	[-9.37]	[-2.91]	[-2.86]
ΔUnemploy$_{i,t}$ (%)	0.0038	-0.0092	-0.0093
	[0.51]	[-0.86]	[-0.88]
ΔIncome/1,000$_{i,t}$	0.0150**	0.0020	0.0024
	[2.26]	[0.25]	[0.31]
Return$_{i,t-1}$ (%)	0.0000	0.0002**	0.0002**
	[0.12]	[2.05]	[2.05]
Intercept	-0.1039	-0.0983	-0.0819
	[-8.42]	[-5.55]	[-5.94]
County FE		Y	Y
State*Year FE		Y	Y
Obs	21924	21924	21924
adj. R^2	0.024	0.037	0.037

Table 4: Identification strategy: Financial misconduct revelation

Table 4 reports the panel regression of county-level stock market participation rates, using the probability of financial misconduct as the instrument variable for friends' participation. Panel A shows the first-stage results where we regress stock market participation rates on the probability of financial misconduct (number of financial misconduct revelation scaled by population in thousands) along with demographic control variables and fixed effects, including the participation rate in the previous year, population, education attainments, unemployment rates, median household income, local stock returns, county fixed effect, and state*year fixed effects. In Panel B, we estimate local participation rate using the probability of financial misconduct as the excluded variable then value-weight the predicted participation rate to obtain the instrumented friends' participation. We regress county-level stock market participation rates on the instrumented friends' participation along with control variables and fixed effects. In the parentheses are the t-statistics calculated from standard errors clustered at county level.*p<.1; **p<.05; ***p<.01 for the main variables.

Panel A. First stage effect of misconduct on local participation

	(1)	(2)	(3)	(4)
	Dependent: ΔParticipation$_{i,t}$ (%)			
Prob. of Misconduct$_{i,t+1}$			-0.0094	-0.0242
			[-0.17]	[-0.37]
Prob. of Misconduct$_{i,t}$		-0.2589**	-0.2743***	-0.2948***
		[-2.54]	[-3.07]	[-3.63]
Prob. of Misconduct$_{i,t-1}$			-0.2498*	-0.3108*
			[-1.88]	[-1.91]
Prob. of Misconduct$_{i,t-2}$				-0.3003
				[-0.67]
ΔParticipation$_{i,t-1}$ (%)	-0.1942***	-0.2468***	-0.2488***	-0.2489***
	[-3.17]	[-6.57]	[-3.85]	[-3.85]
ΔPopulation/1,000$_{i,t}$	0.0077	0.0048	0.0025	0.0020
	[1.02]	[0.61]	[0.29]	[0.22]
ΔEducation$_{i,t}$ (%)	-0.0175***	-0.0084	-0.0075	-0.0075
	[-2.79]	[-0.71]	[-0.53]	[-0.53]
ΔUnemploy$_{i,t}$ (%)	-0.0077	-0.0097	-0.0056	-0.0057
	[-0.74]	[-0.87]	[-0.44]	[-0.45]
ΔIncome/1,000$_{i,t}$	0.0015	0.0029	0.0089	0.0090
	[0.19]	[0.29]	[0.90]	[0.90]
Return$_{i,t-1}$ (%)	0.0002**	0.0002**	0.0002*	0.0002*
	[2.10]	[1.98]	[1.89]	[1.91]
Intercept	-0.1103	-0.1302	-0.1561	-0.1524
	[-8.58]	[-10.95]	[-10.95]	[-10.16]
County FE	Y	Y	Y	Y
State*Year FE	Y	Y	Y	Y
Obs	21924	18793	15657	15657
adj. R^2	0.032	0.060	0.059	0.059

Panel B. Instrument variable and 2SLS

	(1)	(2)
	Dependent:	ΔParticipation$_{i,t}$ (%)
ΔFP$^{IV}_{i,t-1}$ (%)		0.4553**
		[2.17]
ΔFP$_{i,t-1}$ (%)	0.1417*	
	[1.86]	
ΔParticipation$_{i,t-1}$ (%)	-0.1988***	-0.2553***
	[-3.11]	[-5.21]
ΔPopulation/1,000$_{i,t}$	0.0080	0.0250***
	[1.07]	[2.90]
ΔEducation$_{i,t}$ (%)	-0.0172***	-0.0183***
	[-2.75]	[-2.86]
ΔUnemploy$_{i,t}$ (%)	-0.0078	-0.0098
	[-0.76]	[-0.89]
ΔIncome/1,000$_{i,t}$	0.002	0.0024
	[0.25]	[0.26]
Return$_{i,t-1}$ (%)	0.0002**	0.0002*
	[2.10]	[1.67]
Intercept	-0.0937	-0.0971
	[-9.27]	[-4.64]
County FE	Y	Y
State*Year FE	Y	Y
Obs	21924	18789
adj. R^2	0.033	0.041

Table 5: friends' participation and local stock market participation of different income groups

Table 5 reports the panel regression of county-level stock market participation rates of different income groups on its friends' participation rate in the previous year along with demographic control variables and fixed effects, including the participation rate in the previous year, population, education attainments, unemployment rates, median household income, local stock returns, county fixed effect, and state*year fixed effects. The thresholds for income groups are 50K and 100K. Population and income are scaled by 1,000. Education attainment is the percentage of population with degrees higher than Bachelor. Unemployment rate and local stock return are also in percentage points. In the parentheses are the t-statistics calculated from standard errors clustered at county level. *p<.1; **p<.05; ***p<.01 for the main variables.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent: ΔParticipation$_{i,t}^{IncomeGroup}$ (%)							
	All	Low	Mid	High	All	Low	Mid	High
$\Delta FP_{i,t-1}$ (%)	0.2472***	0.2861***	0.3636	1.0238***	0.1417*	0.1892**	0.1514	0.7534**
	[3.59]	[3.55]	[1.56]	[3.39]	[1.86]	[2.17]	[0.65]	[2.62]
ΔParticipation$_{i,t-1}$ (%)	-0.1441***				-0.1988***			
	[-3.11]				[-3.11]			
ΔParticipation$_{i,t-1}^{low}$ (%)		-0.2257***				-0.2553***		
		[-4.17]				[-3.98]		
ΔParticipation$_{i,t-1}^{mid}$ (%)			-0.3254***				-0.3514***	
			[-13.38]				[-14.27]	
ΔParticipation$_{i,t-1}^{high}$ (%)				-0.3171***				-0.3538***
				[-18.88]				[-21.67]
ΔPopulation/1,000$_{i,t}$	0.0121**	-0.0037	-0.0245***	0.0161**	0.0080	-0.0009	0.0223	0.0423**
	[2.50]	[-1.36]	[-4.05]	[2.29]	[1.07]	[-0.14]	[1.59]	[2.09]
ΔEducation$_{i,t}$ (%)	-0.0190***	0.0064	-0.0152	-0.0172	-0.0172***	0.0008	-0.0255**	-0.0115
	[-3.25]	[0.77]	[-1.12]	[-1.05]	[-2.75]	[0.09]	[-1.71]	[-0.66]
ΔUnemploy$_{i,t}$ (%)	-0.0047	-0.0066	0.0154	-0.0322	-0.0078	-0.0058	0.0000	-0.0493
	[-0.51]	[-0.38]	[0.42]	[-0.81]	[-0.76]	[-0.29]	[-0.00]	[-1.12]
ΔIncome/1,000$_{i,t}$	0.0039	-0.0020	0.0033	0.0025	0.0020	0.0044	-0.0009	0.0108
	[0.59]	[-0.20]	[0.20]	[0.13]	[0.25]	[0.37]	[-0.04]	[0.49]
Return$_{i,t-1}$ (%)	0.0002**	0.0001	-0.0002	0.0002	0.0002**	0.0002	0.0000	0.0000
	[2.09]	[0.66]	[-1.16]	[0.49]	[2.10]	[1.62]	[0.04]	[0.08]
Intercept	-0.0760	-0.2328	-0.5937	-0.6419	-0.0937	-0.2630	-0.6535	-0.7144
	[-6.35]	[-12.25]	[-15.12]	[-12.05]	[-9.27]	[-13.50]	[-17.61]	[-15.99]
County FE					Y	Y	Y	Y
State*Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs	21924	21924	21924	21924	21924	21924	21924	21924
adj. R^2	0.082	0.105	0.122	0.141	0.033	0.024	0.036	0.071

Table 6: friends' participation from metropolitan and non-metropolitan counties

Table 6 reports the panel regression of county-level stock market participation rates of metropolitan and non-metropolitan counties on its friends' participation rate from metropolitan and/or non-metropolitan counties in the previous year along with demographic control variables and fixed effects, including the participation rate in the previous year, population, education attainments, unemployment rates, median household income, local stock returns, county fixed effect, and state*year fixed effects. Panel A reports the results using the participation rate from the entire US counties as the dependent variable. Panel B (FP from entire US) and Panel C (FP from metro- or non-metro-counties) reports the results using the participation rate from metropolitan counties or non-metropolitan counties as the dependent variable. Population and income are scaled by 1,000. Education attainment is the percentage of population with degrees higher than Bachelor. Unemployment rate and local stock return are also in percentage points. In the parentheses are the t-statistics calculated from standard errors clustered at county level. *p<.1; **p<.05; ***p<.01 for the main variables.

Panel A. Participation of all US counties

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent: $\Delta\text{Participation}_{i,t}$ (%)					
$\Delta\text{FP}^{Metro}_{i,t-1}$ (%)	0.1427***		0.1120**	0.1143**		0.0997**
	[2.73]		[2.38]	[2.38]		[2.27]
$\Delta\text{FP}^{NonMetro}_{i,t-1}$ (%)		0.2047***	0.1754**		0.1052	0.0802
		[2.89]	[2.57]		[1.42]	[1.13]
$\Delta\text{Participation}_{i,t-1}$ (%)	-0.1394***	-0.1412***	-0.1423***	-0.1971***	-0.1969***	-0.1988***
	[-3.09]	[-3.11]	[-3.11]	[-3.15]	[-3.13]	[-3.12]
$\Delta\text{Population}/1{,}000_{i,t}$	0.0119**	0.0119**	0.0120**	0.0079	0.0079	0.0080
	[2.49]	[2.48]	[2.49]	[1.05]	[1.05]	[1.07]
$\Delta\text{Education}_{i,t}$ (%)	-0.0193***	-0.0191***	-0.0191***	-0.0173***	-0.0173***	-0.0172***
	[-3.31]	[-3.27]	[-3.28]	[-2.77]	[-2.77]	[-2.76]
$\Delta\text{Unemploy}_{i,t}$ (%)	-0.0042	-0.0047	-0.0045	-0.0075	-0.0078	-0.0077
	[-0.45]	[-0.51]	[-0.48]	[-0.73]	[-0.75]	[-0.74]
$\Delta\text{Income}/1{,}000_{i,t}$	0.0036	0.0036	0.0039	0.0018	0.0018	0.0020
	[0.54]	[0.55]	[0.58]	[0.23]	[0.23]	[0.25]
$\text{Return}_{i,t-1}$ (%)	0.0002**	0.0002**	0.0002**	0.0002**	0.0002**	0.0002**
	[2.07]	[2.07]	[2.06]	[2.09]	[2.09]	[2.08]
Intercept	-0.0865	-0.0815	-0.0704	-0.0954	-0.0983	-0.0882
	[-7.18]	[-6.34]	[-5.59]	[-9.14]	[-8.89]	[-8.59]
County FE				Y	Y	Y
State*Year FE	Y	Y	Y	Y	Y	Y
Obs	21924	21924	21924	21924	21924	21924
adj. R^2	0.081	0.081	0.082	0.033	0.033	0.033

Panel B. Friends' participation and metro/non-metro participation

	(1)	(2)	(3)	(4)
	$\Delta\text{Par}_{i,t}^{Metro}$ (%)		$\Delta\text{Par}_{i,t}^{NonMetro}$ (%)	
$\Delta\text{FP}_{i,t-1}$ (%)	0.3436***	0.3001***	0.2372***	0.1426*
	[3.65]	[3.21]	[3.31]	[1.91]
$\Delta\text{Participation}_{i,t-1}$ (%)	-0.0945***	-0.2606***	-0.1504***	-0.1968***
	[-3.36]	[-10.71]	[-2.94]	[-2.87]
$\Delta\text{Population/1,000}_{i,t}$	0.0086*	0.0045	0.0390	0.0642
	[1.92]	[0.60]	[1.07]	[1.63]
$\Delta\text{Education}_{i,t}$ (%)	-0.0249***	-0.0244***	-0.0092	-0.0067
	[-7.99]	[-7.39]	[-0.82]	[-0.57]
$\Delta\text{Unemploy}_{i,t}$ (%)	0.0073	0.0036	-0.0050	-0.0073
	[0.79]	[0.32]	[-0.45]	[-0.60]
$\Delta\text{Income/1,000}_{i,t}$	0.0175**	0.0112	-0.0033	-0.0040
	[2.32]	[1.10]	[-0.43]	[-0.46]
$\text{Return}_{i,t-1}$ (%)	0.0003*	0.0003*	0.0002	0.0003
	[1.91]	[1.74]	[1.28]	[1.52]
Intercept	-0.0862	-0.1103	-0.0554	-0.0710
	[-5.37]	[-5.99]	[-3.54]	[-6.43]
County FE		Y		Y
State*Year FE	Y	Y	Y	Y
Obs	8134	8134	13783	13783
R^2	0.361	0.490	0.082	0.165
adj. R^2	0.332	0.373	0.059	-0.003

Panel C. Friends' participation from metro-/non-metro counties and metro/non-metro participation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta\text{Par}_{i,t}^{Metro}$ (%)				$\Delta\text{Par}_{i,t}^{NonMetro}$ (%)			
$\Delta\text{FP}_{i,t-1}^{Metro}$ (%)	0.1952**	0.1155			0.1545***	0.1276**		
	[2.05]	[1.21]			[2.71]	[2.41]		
$\Delta\text{FP}_{i,t-1}^{NonMetro}$ (%)			0.2792***	0.2476**			0.1897**	0.1030
			[2.87]	[2.54]			[2.55]	[1.38]
$\Delta\text{Participation}_{i,t-1}$ (%)	-0.0699**	-0.2393***	-0.0900***	-0.2571***	-0.1482***	-0.1965***	-0.1476***	-0.1949***
	[-2.59]	[-10.63]	[-3.41]	[-11.22]	[-2.91]	[-2.88]	[-2.93]	[-2.88]
$\Delta\text{Population}/1{,}000_{i,t}$	0.0081*	0.0040	0.0085	0.0044	0.0411	0.0662*	0.0381	0.0646
	[1.86]	[0.55]	[1.91]	[0.60]	[1.12]	[1.68]	[1.04]	[1.64]
$\Delta\text{Education}_{i,t}$ (%)	-0.0250***	-0.0247***	-0.0251***	-0.0247***	-0.0097	-0.0070	-0.0091	-0.0068
	[-8.04]	[-7.47]	[-8.06]	[-7.45]	[-0.87]	[-0.59]	[-0.82]	[-0.57]
$\Delta\text{Unemploy}_{i,t}$ (%)	0.0063	0.0035	0.0070	0.0033	-0.0042	-0.0070	-0.0049	-0.0073
	[0.68]	[0.31]	[0.75]	[0.30]	[-0.37]	[-0.57]	[-0.44]	[-0.59]
$\Delta\text{Income}/1{,}000_{i,t}$	0.0169**	0.0108	0.0174**	0.0111	-0.0036	-0.0043	-0.0036	-0.0043
	[2.18]	[1.05]	[2.28]	[1.08]	[-0.47]	[-0.49]	[-0.47]	[-0.49]
$\text{Return}_{i,t-1}$ (%)	0.0003**	0.0003*	0.0003*	0.0003*	0.0002	0.0003	0.0002	0.0003
	[2.00]	[1.81]	[1.92]	[1.74]	[1.21]	[1.50]	[1.30]	[1.52]
Intercept	-0.0990	-0.1299	-0.0969	-0.1192	-0.0634	-0.0716	-0.0605	-0.0754
	[-5.83]	[-7.67]	[-5.89]	[-6.43]	[-4.17]	[-6.32]	[-3.63]	[-6.22]
County FE		Y		Y		Y		Y
State*Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs	8134	8134	8134	8134	13783	13783	13783	13783
R^2	0.358	0.488	0.360	0.490	0.081	0.165	0.081	0.164
adj. R^2	0.329	0.371	0.331	0.373	0.059	-0.003	0.059	-0.003

Table 7: Stock market participation and income inequality

Table 7 reports the panel regression of county-level Gini coefficients of metropolitan and/or non-metropolitan counties on its participation rates from different income groups in the previous year along with demographic control variables and fixed effects, including the participation rates of different income groups in the previous year, population, education attainments, unemployment rates, median household income, local stock returns, county fixed effect, and state*year fixed effects. Population and income are scaled by 1,000. Education attainment is the percentage of population with degrees higher than Bachelor. Unemployment rate and local stock return are also in percentage points. The participation rates are decomposed into the predicted values from friends' participation and the residuals. In column (1) to (3), the dependent variables are the change in Gini coefficients from year t-1 to year t; in column (4) to (6), the dependent variables are the change in Gini coefficients from year t to year t+1; in column (7) to (9), the dependent variables are the change in Gini coefficients from year t-1 to year t+1. In the parentheses are the t-statistics calculated from standard errors clustered at county level. *p<.1; **p<.05; ***p<.01 for the main variables.

	(1) Full sample	(2) NonMetro	(3) Metro	(4) Full sample	(5) NonMetro	(6) Metro	(7) Full sample	(8) NonMetro	(9) Metro
	Dependent: $\Delta \text{Gini}_{i,t}$			Dependent: $\Delta \text{Gini}_{i,t+1}$			Dependent: $\Delta \text{Gini}_{i,[t-1:t+1]}$		
$\Delta \widehat{Par}^{low}_{i,t-1}$ (%)	0.0000	0.0004	-0.0062*	-0.0022	-0.0020	-0.0068**	-0.0019	-0.0014	-0.0105***
	[0.01]	[0.39]	[-1.92]	[-1.51]	[-1.34]	[-2.01]	[-1.49]	[-1.08]	[-2.96]
$\Delta Par^{low,resi}_{i,t-1}$ (%)	0.0000	0.0000	-0.0017	0.0001	0.0000	0.0022*	0.0003	0.0003	0.0001
	[-0.01]	[0.07]	[-1.10]	[0.27]	[-0.04]	[1.89]	[0.68]	[0.57]	[0.09]
$\Delta \widehat{Par}^{mid}_{i,t-1}$ (%)	0.0001	0.0001	0.0000	-0.0004	-0.0003	-0.0019*	-0.0004	-0.0003	-0.0012
	[0.15]	[0.11]	[-0.01]	[-0.80]	[-0.49]	[-1.77]	[-0.78]	[-0.53]	[-1.11]
$\Delta Par^{mid,resi}_{i,t-1}$ (%)	0.0001	0.0001	0.0001	0.0000	0.0001	-0.0001	0.0000	0.0000	0.0002
	[0.75]	[0.54]	[0.15]	[0.12]	[0.26]	[-0.23]	[0.21]	[0.13]	[0.22]
$\Delta \widehat{Par}^{high}_{i,t-1}$ (%)	-0.0001	-0.0001	-0.0004	0.0004*	0.0003	0.0017**	0.0000	0.0000	0.0013
	[-0.46]	[-0.38]	[-0.73]	[1.72]	[1.26]	[2.14]	[0.15]	[-0.10]	[1.47]
$\Delta Par^{high,resi}_{i,t-1}$ (%)	0.0000	0.0000	0.0001	-0.0001	0.0000	-0.0001	-0.0001	-0.0001	-0.0001
	[-0.17]	[-0.28]	[0.33]	[-0.49]	[-0.40]	[-0.32]	[-0.62]	[-0.62]	[-0.36]
$\Delta \text{Gini}_{i,t-1}$	-0.2910***	-0.2879***	-0.3166***	-0.0039	-0.0098	0.0236	-0.2585***	-0.2707***	-0.2198***
	[-17.11]	[-15.94]	[-8.05]	[-0.24]	[-0.51]	[1.05]	[-15.90]	[-14.21]	[-8.41]
$\Delta \text{Population}/1{,}000_{i,t}$	-0.0005***	-0.0026**	-0.0003***	-0.0004***	-0.0035**	-0.0002	-0.0008***	-0.0052***	-0.0004**
	[-3.95]	[-2.02]	[-2.75]	[-4.25]	[-2.63]	[-1.64]	[-3.48]	[-2.70]	[-2.30]
$\Delta \text{Education}_{i,t}$ (%)	0.0005***	0.0006**	0.0005***	-0.0008*	-0.0007	-0.0007	-0.0003	0.0000	-0.0007*
	[4.71]	[2.55]	[3.72]	[-1.95]	[-1.44]	[-1.54]	[-0.73]	[-0.05]	[-1.70]
$\Delta \text{Unemploy}_{i,t}$ (%)	0.0006**	0.0006*	0.0008*	-0.0006**	-0.0004	-0.0016***	-0.0001	0.0000	-0.0009
	[2.25]	[1.72]	[1.88]	[-2.14]	[-1.39]	[-2.94]	[-0.21]	[0.05]	[-1.31]
$\Delta \text{Income}/1{,}000_{i,t}$	-0.0002	-0.0002	-0.0004	0.0003	0.0004	0.0000	-0.0001	0.0000	-0.0003
	[-1.41]	[-0.88]	[-1.47]	[0.80]	[0.84]	[0.16]	[-0.36]	[-0.08]	[-1.11]
$\text{Return}_{i,t-1}$ (%)	-0.0000*	-0.0000	-0.0000	-0.0000	0.0000	-0.0000	-0.0000**	-0.0000*	-0.0000
	[-1.72]	[-1.43]	[-1.26]	[-0.18]	[0.94]	[-0.40]	[-2.04]	[-1.89]	[-1.21]
Intercept	0.0026	0.0036	-0.0005	0.0006	0.0015	-0.0023	0.0025	0.0043	-0.0024
	[6.35]	[6.93]	[-0.41]	[1.04]	[2.22]	[-2.26]	[4.74]	[6.99]	[-2.69]
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
State*Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	18789	11811	6972	15656	9841	5810	15656	9841	5810
adj. R^2	0.148	0.140	0.227	0.043	0.043	0.097	0.129	0.135	0.146