

Neighbors Matter: Causal Community Effects and Stock Market Participation*

Jeffrey R. Brown

University of Illinois at Urbana-Champaign and NBER

Zoran Ivković

University of Illinois at Urbana-Champaign

Paul A. Smith

Federal Reserve Board of Governors

Scott Weisbenner

University of Illinois at Urbana-Champaign and NBER

December 2006

Abstract: This paper provides evidence of a causal relation between the average stock market participation decisions of one's community and an individual's decision of whether to own stocks. To establish causality, we instrument for the average ownership of a non-mover's community with the lagged average ownership of the state in which one's non-native neighbors were born (or, more specifically, the state in which a neighbor lived when they applied for their Social Security number). Combining this instrumental variables approach with controls for individual and community fixed effects, a broad set of time-varying individual and community controls, as well as state-by-year effects, enables us to rule out alternative explanations. To further establish that this causal effect is driven by "word-of-mouth," we show that the results are stronger in more sociable communities. The results suggest that a ten percentage-point increase in the stock ownership of one's community increases an individual's own probability of owning stock by four percentage points.

* Acknowledgements: We thank Bo Becker, Josh Coval, Caroline Hoxby, Jeff Kubik, Erzo Luttmer, Ulrike Malmendier, Jennifer Marietta-Westberg, Josh Pollet, Annette Vissing-Jorgensen, and seminar participants at the 2004 Western Finance Association annual meetings, the 2005 People & Money: The Human Factor in Financial Decision-Making conference at DePaul University, Harvard University, Ohio State University, the University of Illinois, the University of Michigan, the University of Virginia, and the University of Wisconsin for their useful comments. The analysis for this paper was completed while Paul Smith was an economist with the Office of Tax Analysis at the U.S. Department of Treasury. We thank Jim Cilke for assistance with the tax data and Jean Roth of the NBER for expert assistance with the Census data.

Standard models of individual portfolio choice typically assume that fully informed rational investors maximize utility by allocating wealth across assets based on the risk and return characteristics of the overall portfolio. As noted by Ellison and Fudenberg (1995), however, “economic agents must often make decisions without knowing the costs and benefits of the possible choices” and thus “it is understandable that agents often choose ... (to) rely on whatever information they have obtained via causal word-of-mouth communication.” Accordingly, given the large body of evidence that average U.S. citizens have limited investment knowledge,¹ might individuals make portfolio choice decisions based on what they have learned from social interactions?

This paper empirically examines the influence of “community effects,” in the form of word-of-mouth communication, on the decision of whether to participate in the stock market. As a result of the equity premium over the past century (Mehra and Prescott (1985), Fama and French (2002)), equity ownership has been an important determinant of individuals’ wealth because, in expectation, those individuals who accepted the risks of owning stock have been able to accumulate significantly more wealth than individuals who had not. Whereas the simplest models of portfolio choice suggest that even extremely risk-averse investors should place at least a small portion of their assets in stocks (e.g., Mas-Colell, Whinston, and Green (1995)), only one-third of U.S. households own stocks or stock mutual funds outside of retirement plans, and the fraction rises only to one-half even with the inclusion of retirement plans.²

Economically, stock market participation rates are important. In aggregate, they affect the level of the equity premium itself (Mankiw and Zeldes (1991), Heaton and Lucas (2000), Brav, Constantinides, and Gezcy (2002)). They are also relevant for a variety of public policy reasons, ranging from the incidence of dividend tax policy to the fact that the presence of equity market non-participants means that investing Social Security surpluses in private investments can have real effects on the economy (Abel (2001), Diamond and Geanakoplos (2003)).

Whether non-participation is driven by entirely rational reasons (e.g., Vissing-Jorgensen (1999) finds that fixed costs of equity market participation as low as \$200 may be sufficient to

¹ For example, John Hancock Financial Services (2004) finds that nearly one-half of defined contribution plan participants report that they have little or no investment knowledge and fewer than 20% consider themselves relatively knowledgeable. In further support of limited investment knowledge, the survey also finds that many respondents think that employer stock is less risky than a domestic or international stock fund and that a money market fund contains stocks.

² These figures are based on authors’ tabulations from the 2001 Survey of Consumer Finances and are consistent with Aizcorbe, Kennickell, and Moore (2003).

explain the observed rates of non-participation) or by behavioral biases, there are numerous reasons to suspect that word-of-mouth communication might lead to a causal relation between one's own decision to invest in stocks and the stock market participation of one's community. For example, the fixed costs of equity market participation in the Vissing-Jorgensen (1999) model may be in part non-pecuniary, such as the need to educate oneself about how markets work or how to invest. Social interaction is a plausible avenue through which individuals might gain the investment knowledge that leads them to start participating in the stock market.

In this paper, we seek to establish a causal relation between individual and community stock market participation. In particular, we study the extent to which an individual is more likely to participate in the stock market when a higher fraction of individuals in the local community are stock market investors. If such a relation exists, and if its causality can be demonstrated, there might be externalities to an individual's decision to invest in the stock market such as lowering the informational or psychological barriers to investment for others. Accordingly, evaluations of public policies that influence participation rates, such as those related to financial education, would need to account for the potential spillover effects that this education might have if it is more broadly disseminated through the community via "word-of-mouth" communication.

Hong, Kubik, and Stein (2004) present a model in which stock market participation may be influenced by social interaction. In this context, social interaction can serve as a mechanism for information exchange via word-of-mouth or "observational learning" (Banerjee (1992), Ellison and Fudenberg (1993, 1995)). Simply put, many individuals may find it easier to learn how to open a mutual fund or brokerage account by talking to their friends than through other mechanisms, thus lowering the psychological "fixed costs" of investing. Individuals may simply enjoy discussing stock market investments with their friends and colleagues and thus may be more apt to participate in the stock market if there is a high participation rate among their friends and colleagues. There may also be a "keeping up with the Joneses" effect. For example, according to Bernheim's (1994) model of conformity, individuals may wish to maintain the same consumption that their social group does, suggesting that participation in the stock market by their social group may have a positive influence on their own decision to do so. This effect could also be generated by external habit formation models (Campbell and Cochrane (1999)) or the scarcity of local resources that lead investors to care about their relative wealth in the

community (DeMarzo, Kaniel, and Kremer (2004)). Thus, there are several reasons to suspect that such a causal relation might exist, each of which relies on some form of word-of-mouth interaction.

It is important to distinguish these *causal* word-of-mouth effects from many other confounding factors, such as similarity in preferences or information sets, which could lead to a spurious correlation between one's stock market participation and that of one's neighbors. In the context of stock market participation, the previous literature is not conclusive. Guiso, Sapienza, and Zingales (2004) show that individuals who currently live or were born in areas with higher levels of social capital are more likely to invest in stocks, but they do not directly test whether higher stock market participation of one's neighbors directly influences one's own decision. Whereas Hong, Kubik, and Stein (2004) show that individuals who visit with neighbors or attend church have higher levels of stock market participation, and that this effect is strongest for individuals who live in more sociable states, concerns linger about the potential for unobserved characteristics to be driving both stock market participation decisions and their measures of social interaction. For example, Gruber (2005) shows that religious participation (one of the measures of sociability used in Hong, Kubik, and Stein (2004)) is correlated with higher levels of income, education, and marriage, as well as reduced levels of welfare receipt, disability, and divorce, all of which, in turn, may be correlated directly with stock market participation. Feng and Seasholes (2004) show, using data from the People's Republic of China, that common reaction to public information, rather than word-of-mouth effects, seems to be a primary determinant of investors' trading behavior, at least in terms of investment decisions conditional on stock market participation. Thus, the presence of an effect of social interaction on stock market participation is far from settled.

One reason the prior literature has not provided conclusive evidence of a causal relation is that, as discussed extensively by Manski (1993, 1995), designing an empirical strategy to overcome the inherent identification problem is challenging. The essence of the problem is that the observed correlation between the stock ownership of an individual's community and the individual's own portfolio choice could be causal, but it could also reflect numerous unobservable influences that induce a correlation even after controlling for observable characteristics. In short, because individuals are not randomly assigned to their communities,

and indeed may choose their communities precisely for reasons that are correlated with stock market participation, it is difficult to establish a causal link.

In this paper, using a large, nationally representative, 10-year panel data set of taxpayers, we provide evidence of an economically and statistically significant causal community effect of stock market participation. We do so by simultaneously implementing an instrumental variables strategy that overcomes the endogeneity problem and controlling for a very broad set of observable and unobservable individual and community characteristics. Our instrument is motivated by Guiso, Sapienza, and Zingales (2004), who suggest that the social capital of one's birth region can have long-lasting effects on one's own financial decisions. We begin by restricting our sample to "native" individuals, that is, those who have lived in the same community over the entire panel and who still reside in their state of birth.³ We then instrument for the average ownership within each native individual's community with the average ownership of the birth states of "non-native" neighbors, that is, those community residents whose Social Security numbers were issued in states that do not overlap with their current community (intuitively, those born in a different community and state). The logic of this instrument is that, whereas the average ownership in the birth states of one's neighbors will be correlated with the ownership of one's neighbors because of the long-lasting effects described above, there is no reason why one's own stock market participation decision should be directly influenced by the ownership rates in these other states except through their effects on one's neighbors.⁴ Further, because we restrict our sample to "native" individuals and instrument for the ownership of their community with the average birth-state ownership of their "non-native" neighbors, we are ensuring that the regression results obtained cannot be contaminated by an individual and his community members sharing the same background (i.e., birth state).

For example, suppose an individual lives in the St. Louis Metropolitan Statistical Area (henceforth MSA). A naïve (non-instrumental variables) regression analysis might simply relate the individual's ownership to the ownership of the others who live in the St. Louis MSA. The coefficient on such a regression has little meaning, as numerous unobserved factors may be spuriously driving this correlation.

³ Throughout the paper, in the interest of brevity, we will often use terms such as "state of birth" or "birth state" to refer to the state in which an individual resided when they applied for a Social Security number.

⁴ We are also careful to rule out simultaneous correlated responses across communities to the release of new information. To mitigate such concerns, we use a one-year lagged value of the instrument.

In contrast, in our instrumental variables regression, we first limit our sample to individuals that have lived in the St. Louis MSA over the entire sample, and who were also born in Missouri or Illinois (states that contain part of the St. Louis MSA). We then instrument for the average community ownership of these St. Louis natives with the average ownership across the birth states in which the non-native St. Louis residents had received their Social Security numbers. The instrument is constructed using only those community residents whose birth state does not overlap with the current MSA. Thus, in the case of St. Louis, neighbors who received their Social Security number in Missouri or Illinois would be excluded from the instrument. This construction is possible because the tax panel data contains each taxpayer's Social Security number (SSN), and the first three digits of each person's SSN provide information about their state of residence at the time of their SSN application.⁵

In conjunction with this instrumental variables approach, we include individual fixed effects to control for observable and unobservable individual characteristics that are fixed over time, such as gender, race, or average risk tolerance over the sample period. Because our sample is restricted to non-movers, the individual fixed effects also control for time-invariant community characteristics such as the average level of financial sophistication of a community over the sample period. Thus, we are relating *changes* in individual ownership to *changes* in community ownership, as identified by *changes* in ownership in the SSN states of one's neighbors. Because not all individual and community covariates are time-invariant, we also include a very broad set of time-varying individual controls (e.g., income, age, marital and tax filing status) and time-varying community controls (e.g., composition of industry and occupation, community-level income, median home prices, and characteristics of locally-headquartered publicly-traded firms). We also include state-by-year effects to allow for unobservable statewide characteristics that might be changing over time, thus ruling out any spurious correlation that might occur because of state-specific trends in average ownership rates over the sample.

The results provide evidence of significant causal community effects. Specifically, we find that a ten percentage-point increase in the average ownership in one's community leads to a four percentage-point increase in the likelihood that an individual will own stocks. To further

⁵ We are grateful to the Office of Tax Analysis at the U.S. Department of Treasury for assistance with this confidential data.

verify that our results are driven by word-of-mouth, in our final specification we interact our instrumented community effects variable with an independent measure of community “sociability” (specifically, whether households are likely to be asked by neighbors for advice). These results, which suggest that the community ownership effect is strongest in more sociable communities, strongly suggest that word-of-mouth is the avenue through which our community effects operate.

This paper proceeds as follows. In Section I we introduce the data set, along with our definitions of stock market participation and community. Section II discusses the challenge of establishing that a correlation between individual and community stock ownership is causal in nature. Section III outlines our instrumental variables strategy. Section IV presents the instrumental variables regression results. Section V concludes.

I. The Data

A. The Panel of U.S. Taxpayers

Conducting a nationally representative study of how individual equity market participation decisions are related to community equity market participation requires a large, nationally representative dataset that contains many observations for each community. To the extent that the community effect is based on an individual’s interactions with friends, neighbors, and co-workers, it is important to choose a geographic size large enough to capture most of the individuals in a person’s social group, but not so large as to completely dilute these effects. Most standard micro data sets are not well suited for this task. For example, the Survey of Consumer Finances (henceforth SCF), arguably the best available nationally representative study for examining household financial decision making, is a cross-sectional study with only approximately 5,000 observations nationwide. The Health and Retirement Study, which offers the potential advantage of being a panel study, is only slightly larger than the SCF and focuses exclusively on cohorts nearing or in retirement. Similar problems are associated with most other standard household data sets.

Our primary source of data is a large panel of tax returns covering the years 1987 to 1996.⁶ The panel is based on the IRS's annual cross-sectional sample of tax returns, which are large samples chosen to represent the population of tax-return filers. The population of tax filers is similar to the population of households, except that a household may comprise more than one tax-filing unit (e.g., married couples may file separately), and some households do not file any tax returns.⁷ For example, our data for 1994 represent about 130 million tax returns, while the 1995 SCF represents about 100 million households. Throughout the paper we use the terms individuals, households, and taxpayers interchangeably, but, strictly speaking, the unit of observation for this data set is the taxpayer.

The annual IRS cross-sections are stratified samples, with the probability of inclusion rising steeply with total income.⁸ Variables include most of the values from the 1040 tax form and associated schedules, as well as data from Social Security records, such as the age and gender of each individual represented on the return.⁹ The 1987 cross-sectional sample of about 88,000 returns is the base year of the panel. The panel was constructed by matching, in each subsequent year, the full population of tax returns that appeared in the 1987 sample. This method allows individuals to be tracked over time regardless of income changes and changes in marital status.¹⁰ We exclude married couples filing separate returns.

An advantage of the tax data is that it contains precise geographical identifiers of the taxpayers, that is, their zip codes, which allow us to assign them to the appropriate community. A second advantage is that Treasury employees, who have access to the underlying Social Security numbers for the individuals in our data, can identify the individuals' "birth states," that is, the states in which individuals had resided when they applied for their respective Social

⁶ The panel was developed as a joint effort between the Treasury Department's Office of Tax Analysis and the IRS's Statistics of Income division. The panel uses confidential records and is not publicly available.

⁷ Low-income households are not required to file returns, but many do so to claim a refund of over-withholding or a refundable tax credit such as the EITC. In 1987, the filing threshold ranged from \$4,400 to \$10,000, depending on age and filing status.

⁸ We use appropriate sampling weights when reporting all tabulations and descriptive statistics.

⁹ Further details on the construction of the tax data are available in Amromin and Smith (2003).

¹⁰ As a result of this sampling methodology, our sample is not subject to sample attrition bias that sometimes afflicts panel studies using survey data. However, there is attrition from the *population* of tax filers. It occurs when a filer's income drops below the filing threshold. This source of attrition is non-random—older and lower-income workers are more likely to drop out of the filing population than middle-aged, higher-income workers. Using data from the Survey of Consumer Finances, though, we find that the fraction of non-filers who own equities in a non-retirement account is only 2-4% over the period from 1989 to 1998.

Security numbers.¹¹ Ultimately, the sample contains 753,521 observations, covering 85,888 distinct taxpayers.

This overall sample can then be divided into “natives,” “non-natives” and “others.” “Natives” are those individuals: (i) who reside in the same MSA over the sample period, and (ii) whose current state of residence is their birth state.¹² “Non-natives” are individuals whose current MSA does not overlap with their birth state (e.g., for people currently living in St. Louis, non-natives are those born in any state but Missouri or Illinois). “Others” are individuals who do not fall into the previous two categories (e.g., people whose current residence is in their birth state but who switched MSAs over the sample period, or those who remained in the same MSA over the sample period but who no longer live in their birth state).

To implement our empirical strategy, we will limit our core sample to the 398,585 taxpayer-by-year observations that are natives (i.e., we will estimate regressions over this subsample). When constructing the average community ownership for each observation (as well as other community-level variables constructed using the tax data), we are able to use all 753,521 observations. When constructing our instrument for community ownership, we use the 220,783 “non-native” observations.

Whereas tax data has many advantages, its traditional disadvantage for research purposes is that it contains less information than household surveys about non-financial demographic characteristics such as educational attainment, occupation, and race. To address this concern, we utilize the panel aspect of the data and include individual fixed effects that control for these factors, as well as numerous other unobservable time-invariant individual traits. We also use a broad set of income data from the tax returns to control for time-varying individual characteristics. Because our sample is restricted to non-movers, individual fixed effects simultaneously act as both community and state fixed effects, and thus they control for all time-invariant community characteristics. To control for time-varying community characteristics, we include the average community levels of each of the tax-data variables for each year and also construct, using Census data from 1980, 1990, and 2000, annual time trends for key community measures such as the distribution of race, ethnicity, education, occupation, employment sector,

¹¹ One of the authors of this study, Paul Smith, was on the staff of the Office of Tax Analysis (OTA) during most of this project. We are grateful to other OTA staff members for their assistance.

¹² One result of this definition, which is useful in our empirical work, is that individual fixed effects for this non-mover sample will simultaneously serve as MSA fixed effects.

and median home prices. We use data from Compustat and CRSP to control for time-varying characteristics of the firms headquartered within a community. In later specifications that utilize an independent measure of a community's "sociability," we also use survey data from the DDB Life Style survey. Among other items, that survey collects information that facilitates assessment of a community's stock of social capital (specifically, whether one's neighbors are likely to approach one for advice about products and brands).

B. Measuring Equity Ownership

For tax purposes, it is not necessary for the IRS to know whether an individual owns stock, only whether they receive income from this stock. Thus, our measure of an individual's stock ownership in a given year is based on whether an individual reported dividends on their federal tax return that year. This dividend measure of stock ownership captures all dividends from stocks and taxable mutual funds. It does not include individuals who invest in stocks or equity mutual funds only through non-taxable (i.e., retirement) accounts. Thus, when we refer to equity market participation throughout the paper, we are specifically referring to stock or mutual fund ownership in one's taxable account (i.e., outside of retirement plans).

Although using dividends as a proxy for equity ownership has a very long history in the public finance literature,¹³ there are two reasons why this proxy is imperfect. First, dividend income, as reported on a tax return, includes payouts from any mutual fund, even if the fund is exclusively invested in fixed-income assets. Second, whereas dividends will capture equity ownership for those who invest in dividend-paying firms, it may miss those who invest primarily in non-dividend paying firms, such as technology firms, a sector that became increasingly important during the 1990s. The potential for both type 1 and type 2 errors in defining equity ownership prompts the question whether our dividend measure is a satisfactory proxy for actual equity ownership. We use the SCF to investigate this concern and find that reporting dividend income indeed is a good proxy for stock ownership. First, we compare the trend in actual ownership (as reported in the SCF) to the trend in taxpayers reporting dividends to the IRS from 1987 to 1996. Figure 1 illustrates the growth rates in actual equity ownership as well as the dividend proxy. Over the overlapping period (1989–1996), our proxy for equity ownership

¹³ This practice dates back at least as far as the Means (1930) study of diffusion of stock ownership. For a much more recent example, see Desai, Dharmapala, and Fung (2005).

increased by 24%, comparable to the 20% increase in actual equity ownership. Second, we examine the correlation between these measures. We find that, pooling the four SCF cross-sections from 1989-1998, the correlation at the individual level between actual and proxied equity market participation is 0.62.¹⁴

C. Defining “Community”

We define an individual’s community as the Metropolitan Statistical Area (MSA) in which the individual resides.¹⁵ This community definition, widely used in empirical studies in economics and finance, has several advantages. First, MSAs are small enough to constitute a reasonable definition of a “community” while still capturing the vast majority of one’s social interactions, including work-related interactions. Second, MSAs are large enough to contain sufficiently many individual observations within each community to obtain a reasonable estimate of the stock market participation in that community. Third, because MSAs are well defined and non-overlapping, we can correct for the correlation of error terms within a community. Finally, MSAs provide a convenient community definition for merging our tax panel with external data, such as census data and survey data (used to determine a community’s social capital).

D. Summary Statistics

Table I reports summary statistics for several key variables of interest for the pooled 1987 – 1996 cross-sections. Panel A focuses on equity ownership, as defined by reporting dividends on tax returns. The first row of Panel A shows that 25.6 percent of our sample owns equity. The second row provides the distribution across households of community-level equity ownership. The median household lives in an MSA in which 26 percent of the household’s community owns stock. At the 10th percentile of the community ownership distribution, nearly 17 percent of the household’s community members own stock, whereas this doubles to 34 percent at the 90th percentile. In the third row of Panel A, we report the ownership rate in individuals’ “birth”

¹⁴ We also find that including capital gains realizations along with dividends in our proxy for stock ownership does not significantly alter this correlation (correlation of actual equity ownership is 0.623 with our “report dividends on tax return” measure and 0.628 with our “report dividends or capital gains” measure). Thus, adding capital gains does not improve the correlation of the proxy with actual equity ownership and adds measurement error, as realizing capital gains on several assets (such as real estate) does not imply equity ownership.

¹⁵ There are 337 MSAs in the United States. We match individuals to their MSA based on the zip code from their tax return. For the small fraction of the population, primarily those in rural areas, who live outside of a formal MSA, we assign the individual to the closest MSA as measured by the mileage from the center of the individual’s zip code to the nearest MSA (calculated using the latitude and longitude of each).

states, that is, the states in which individuals lived when they applied for a Social Security number. Because this is a statewide measure, it is not surprising that the distribution shows slightly less dispersion than the distribution at the MSA level reported in the previous row. The final row of Panel A reports the distribution across households of the average equity ownership of their “non-native” community members’ “birth” state (“non-native” community members are those whose Social Security numbers were issued in states that do not overlap with their current MSA). This variable will be essential for our instrumental variables strategy (Sections III and IV).

Panel B focuses on individuals’ characteristics. Its first two rows report the distribution of total non-capital income (in 1996 dollars) and age over our sample, weighted by population. The next four rows feature indicator variables for gender interacted with filing status (single male, single female, male head of household, and female head of household. Married couples filing jointly is the omitted category, accounting for 52.4 percent of all observations, whereas married couples filing separately have been excluded from the analyses. The last six rows of Panel B report an indicator variable for whether the taxpayer claims mortgage interest deduction (a proxy for home ownership), the number of dependents, whether the taxpayer is self-employed, whether he reports supplemental income on schedule E, whether he is subject to the alternative minimum tax, and whether he contributes to a tax-qualified defined contribution plan.

II. The Challenge of Establishing Causality

Our hypothesis is that an individual’s equity market participation decision is influenced by the equity market participation of other individuals in the community. Thus, the null hypothesis of “no community effects” is that, conditional on a complete set of individual and community covariates, a regression of an individual’s ownership against that of the other members of the individual’s community would produce a coefficient of zero. The challenge, to which we will return shortly, is that it is impossible to control for a *complete* set of covariates.

At the outset, consider the simplest case in which there is a single community and every individual i , $i = 1 \dots N$, has a probability of participating in the stock market that is drawn from the same distribution with mean p . A regression of individual participation against the average participation of the remaining community members will produce a zero coefficient because such a specification essentially amounts to regressing one draw from a distribution against the average

of the other, independent draws from the same distribution. The outcome remains unchanged if there are multiple communities $m, m = 1 \dots M$, and individuals are randomly sorted into those communities. Indeed, even though there will be some random variation in the average community ownership rates (centered around p), a regression of individual ownership against mean community ownership will produce a coefficient that is not significantly different from zero. The outcome is still unchanged even if individual stock ownership is determined by a set of covariates $\{x_1, x_2, \dots, x_k\}$, as long as the econometrician can perfectly observe all of the relevant x 's (and thus control for these covariates in the regression).

Of course, in reality, the econometrician can never hope to perfectly observe all the relevant determinants of individual stock market participation. As a result, it is virtually impossible to learn anything about community effects from examining a simple correlation between individual and community ownership. For example, suppose that individuals sort themselves into communities based on a vector of observed (X) and unobserved (U) characteristics, and that these same characteristics are correlated with stock market participation. For example, X might include education and income, and U might include risk preferences, ideology, or simply non-linear functions of the observed Xs.

The existence of U poses a serious problem. Consider, for instance, the extreme case in which there is perfect sorting into communities based on U, and that U is also a primary determinant of individual stock market participation. By sorting on U, individuals have implicitly also sorted into communities based on their stock market participation. Accordingly, it is as if each individual i in community m has the same probability p_m of participating in the stock market, and p_m differs across communities $1 \dots M$, where each community m has I_m individuals. We denote equity ownership of individual i in MSA m by $y_{i,m} \in \{0,1\}$, and define the average ownership of the other $I_m - 1$ individuals in the community as $\bar{y}_{i,m} = \frac{1}{I_m - 1} \sum_{j \neq i} y_{j,m}$. With sufficiently many observations in each MSA, each average $\bar{y}_{i,m}$ will be close to p_m , the probability of participation prevailing in its respective community m . Therefore, a simple regression of y_i on $\bar{y}_{i,m}$ would be roughly equivalent to a regression y_i against p_m , which will yield a regression coefficient of one because $\text{Cov}(y_{i,m}, p_m) = \text{Var}(p_m)$.¹⁶

¹⁶ We thank the referee for highlighting this example.

As a result of such spurious correlation, induced by the omitted U that is correlated with both community choice and our outcome of interest, the coefficient on average community ownership in a simple OLS regression provides no information regarding community effects. For example, in our full sample of over 750,000 observations, a regression of individual ownership on average community ownership, even after controlling for age, income and additional covariates, is approximately 34 (on a scale in which our binary indicator variable has been rescaled to 0 to 100), and this effect is highly statistically significant. This suggests that an individual living in a community with a ten-percentage point higher level of average ownership is 3.4 percentage points more likely to own stocks. However, we cannot place any causal interpretation on this coefficient, and, indeed, given the problems just outlined, it is even difficult to characterize whether this coefficient should be considered large (as it would be relative to the null hypothesis of no community effects), or small (as it would be if one believed that individuals implicitly sorted into communities based on their stock market participation and this sorting mechanism is not controlled for by the covariates included in the regression).

Ultimately, although panel data allow us to control for individual and community fixed effects, and although we can further control for a very broad set of time-varying individual-level and community-level variables, even such detailed controls do not allow us to render a causal interpretation to the coefficient. This is because the correlation between community and individual ownership could still be driven by *unobserved* time-varying factors such as changes in the information set available to a community. For example, an aggressive brokerage firm could open up in a community and convince thousands of individuals to enter into the stock market for the first time. Alternatively, a major shift in the community's economic environment could make it seem like a better or worse time to participate in the market. In each case, we would observe a correlation between community ownership and individual ownership that is not causal.

Thus, the basic identification problem, which has not been satisfactorily addressed in any prior study of stock market participation, is that it is difficult to find factors that would influence the stock market participation decisions of a person's neighbors but not have any direct effect on the person's own participation decision. Consequently, it is relatively easy to construct plausible stories explaining how the correlation between individual and community ownership shown in the prior literature could be driven by some omitted time-varying factor. To identify the correlation as being causal in nature (i.e., reflecting word-of-mouth communication), it is

necessary to find a source of exogenous variation in the stock market participation of one's neighbors. Put differently, if we can find such a source of exogenous variation in community equity ownership, then we can construct regression specifications that test for causal community effects, i.e., regressions in which a zero coefficient on community ownership suggests no word-of-mouth effects upon an individual's ownership decision and a significant positive coefficient indicates causal word-of-mouth are present. We now turn to a description of how we address these difficulties.

III. Establishing Causality: Methodology

A. Overview of Identification Strategy

There are two ways to overcome the identification problem, neither of which has been previously used to examine the question of stock market participation. First, in some contexts, it is possible to randomly assign individuals to a peer group. Examples include Sacerdote's (2001) study of peer effects on grades using randomly assigned college roommates, or Hoxby's (2002) use of random cross-cohort variation in the gender and racial mix of school children to study peer effects and educational outcomes. With such random assignment, one can also use analysis of variance approaches to establish the presence of community effects (Graham (2005)). To our knowledge, however, no data are available to study stock market participation among a group of individuals who were driven by exogenous factors to move to a new community.

The second approach, which we follow in this paper, is to find an instrument correlated with the stock market participation of an individual's neighbors, but not correlated with the stock market participation of the individual, except through the individual's interactions with the community. Motivated by the evidence of Guiso, Sapienza, and Zingales (2004) that there may be long-lasting effects of one's place of birth on financial outcomes, we utilize the unique nature of our data to develop an instrument that relies upon the stock market participation of the birth states of one's "non-native" neighbors (i.e., those born in a state that does not overlap with their current MSA).

We begin by limiting our sample to those individuals who are "native" to their current community (i.e., those whose "birth state" is their current state of residence and who did not

change MSA during the sample period), and then construct our instrument over their “non-native” neighbors. As we will show shortly, ownership in the birth state of one’s “non-native” neighbors will be highly correlated with ownership of one’s neighbors, but, given our sample restriction to natives of the current community, there is little reason to suspect that it will be correlated with one’s own ownership except through its effect on one’s neighbors. In short, if one has lived in St. Louis one’s entire life and one’s neighbor is from Boston, it is reasonable to think that the level of equity ownership in Massachusetts may be correlated with that neighbor’s ownership, but there is no reason to think that it should affect one’s own stock market participation decision unless word-of-mouth effects are at play. Thus, because we restrict our sample to “native” individuals and instrument for the ownership of their community with the average birth-state ownership of their “non-native” neighbors, we are ensuring that the regression results obtained cannot be contaminated by an individual and his community members sharing the same background (i.e., birth state). Further, we use lagged values of the instrument in our regressions to reduce the potential for picking up a spurious correlation that might arise if, for example, there were a simultaneous correlated response across states to the release of new information.

To provide a highly simplified example, consider a “native” individual A, who was born (i.e., had the Social Security number issued) in Missouri and who has lived in the St. Louis MSA over the entire sample period. Suppose that individual A has neighbors B, C, and D who currently live in the St. Louis MSA, and their Social Security numbers were issued in Missouri, California, and Delaware, respectively. We define “community ownership” for individual A to be the average stock market participation of individuals B, C, and D. However, because individual B’s Social Security number was issued in Missouri, we do not include individual B’s birth-state ownership as part of our instrument.¹⁷ Individuals C and D, by contrast, are included in our instrument because they are “non-natives” (their Social Security numbers had been issued in California and Delaware, respectively). Thus, in the first stage of our regression, we regress the average ownership of the community in which individual A lives (here, consisting of the average ownership of individuals B, C, and D) upon the lagged average ownership in the states

¹⁷ The same restriction would apply if B’s Social Security number had been issued in Illinois (rather than Missouri) because Illinois also overlaps with the St. Louis MSA.

from which the “non-local” neighbors came (in this case, the average of the ownership in California and Delaware).

B. The First-Stage Regression

Table II shows that the instrument meets the first criteria—that it is highly correlated with the equity ownership of one’s community. The table reports the coefficients from our first-stage regression of the average ownership of one’s MSA upon the one-year lagged average equity ownership of one’s “non-native” community members’ “birth” state.¹⁸

In Column 1, we report the coefficient from a simple regression with no other controls, and find a highly significant relation. In Column 2, we control the complete set of variables that will ultimately be included in our second-stage regression. That is, we include individual and community fixed effects, state-by-year fixed effects, time-varying household-level controls, and time-varying community-level controls, all of which will be discussed in more detail below. In both specifications, there is a strong statistical relation between equity ownership in a community and our instrument. Moreover, the F-statistic (which is 20.4 in column 1 and 10.6 in column 2) indicates that the instrument is sufficiently powerful according to the Staiger-Stock (1997) test.

We also explore the two sources of exogenous variation in our instrument. First, holding fixed the level of ownership of the birth states, the instrument can vary as the composition of one’s neighbors changes (e.g., somebody born in California leaves the community, and somebody from Ohio joins it). Second, holding one’s neighbors constant, the instrument can vary when there are changes in the levels of ownership in the birth states of the neighbors who are not native to the community.

To test whether one or both sources of variation matter, we construct two separate instruments. To construct the first measure, which we loosely call the “mover” instrument, we hold the ownership in each SSN state fixed at its average level over the entire sample. Thus, for a given individual, this instrument varies only with the community composition. To construct the second measure, which we call the “stayer” instrument, we only include in the instrument the SSN state ownership of those non-native neighbors who have been in the community over the entire length of the panel (but, by definition, whose birth state does not overlap with their current

¹⁸ Because our instrument is lagged by one year, observations from 1987, the first year of our sample, are dropped from the regression and thus the number of person-year observations of “native” individuals is 353,281 (rather than 398,585 over the full sample).

MSA). Thus, for a given individual, this instrument varies only with the changes in the SSN state ownership of the individual's "non-native" (but "stayer") neighbors. We then re-estimate our first-stage regression including both instruments in the regression, thus allowing these two measures to enter separately. Using the specification from Table II, Column 2, we find that the coefficient on the "mover" instrument is 24.2 and the coefficient on the "stayer" instrument is 15.0. Both are individually statistically significant (at the 1% level for the "mover" instrument and at the 10% level for the "stayer" instrument). More important than their individual significances, however, are the following two facts: (i) we cannot reject that the coefficients on the two components of the instrument are the same, and (ii) the two instruments are jointly significant (p-value of 0.0034). Because both instruments matter and have effects that are statistically indistinguishable from each other, and the fact that the combined instrument provides more power because it makes use of additional variation that arises from interactions between these two sources of variation,¹⁹ we use the full instrument in all of our second-stage regressions.

IV. Instrumental Variable Results

A. Core Results

Table III reports the core set of results using our instrumental variables strategy. The specification is a linear probability model in which the left hand side is rescaled to a {0,100} scale to facilitate the interpretation of the regression coefficient estimates as percentage point changes. The key variable of interest is the average equity ownership rate in the community.

In column 1, we begin with a specification that includes household and community fixed effects, year effects, as well as a wide range of household-level time-varying controls. The household-level covariates include: (i) an indicator variable for each percentile of the non-capital income distribution as measured each year, for a total of 99 variables (the 50th percentile is the omitted category), (ii) an indicator variable for each possible age through age 106 (the oldest in our sample), with age 40 as the omitted category, (iii) indicator variables for gender interacted with filing status (single male, single female, male head of household, female head of household), and (iv) numerous other items from the tax return, including dummies for whether

¹⁹ A regression of the combined instrument on the two separate instruments has an R-squared of 0.92, suggesting that the interactions account for approximately 8 percent of the total variation.

the household claims a mortgage interest deduction, self-employment, reporting supplemental income on schedule E, contributions made to tax-qualified defined contribution plans, as well as the number of dependents. Although not the central focus of the paper, stock ownership is related to these household level variables in ways consistent with one's intuition (e.g., the probability of ownership is rising with income, age, and with the various proxies for household wealth that are available on the tax return).

Of central importance to this paper, the coefficient on the equity ownership within the community is 35, which is significant at the 1 percent level. This suggests that a 10-percentage point increase in community ownership leads to a 3.5 percentage point increase in individual ownership.

In Column 2, we augment the covariates with a broad set of time-varying community-level controls. These include community-level averages of all of the individual-level controls for income, age, filing status, and so on.²⁰ In addition, we use 1980, 1990, and 2000 Census data to compute community-specific time trends for a host of variables, including the distribution of completed education in the community (e.g., fraction of community that completed grades 0-8, 9-11, high school, some college, and college graduates), the distribution of community residents by race and ethnicity (e.g., the fraction of Hispanics, Blacks, Indians, Asians, and other groups in the MSA), the distribution across 13 occupation classifications, 11 industry classifications, and 6 sector classifications (e.g., state and local government, private sector, and so on). We also include trends in the median home value in a community to help capture community specific economic conditions. We also include non-linear controls for the population of the MSA in each year.

In recognition of the empirical evidence that both professional money managers and individual investors tend disproportionately to invest in stocks of the companies headquartered in their local communities (Coval and Moskowitz (1999, 2001), Ivković and Weisbenner (2005), and Massa and Simonov (2006)), the specification presented in column 2 also includes a large set of covariates characterizing locally-headquartered publicly-traded firms. Specifically, we include a dummy variable that indicates whether there are publicly-traded firms headquartered within the community, the fraction of total U.S. market value represented by these local firms,

²⁰ For age and income, rather than including community level dummy variables for each point in the distribution, we include a quadratic specification of the community averages.

the value weighted rate of return of local firms, the fraction of local firms that were in the top ten percent of nation-wide performers that year, the fraction of local firms that pay dividends, and the fraction of the population employed by these local firms. With the inclusion of these controls, the coefficient on community ownership drops slightly to 26.3, but remains significant at the 10 percent level.

Over the sample period, from 1987 to 1996, the participation rate in the stock market has increased. Although the inclusion of year fixed effects controls for this national trend, we can go even further by replacing year fixed effects with state-by-year effects in the regression. This ensures that our results are not driven by differential rates of change in stock-market participation across states, or, for that matter, any other unobservable time-varying factor that differs across states. In other words, the inclusion of state-by-year effects means that we are no longer using any state-by-year variation in any variable to identify our key effect of interest. Thus, if there are state-level differential trends or shocks in economic variables, such as a state being hit particularly hard by a recession or a changing tax policy, or even in social norms, such as a state becoming increasingly conservative, these factors will now be controlled for.

As seen in Column 3, the coefficient of 41.8 in this comprehensive specification is highly statistically significant. This coefficient suggests that a ten percentage-point increase in the ownership of one's community, which roughly corresponds to a move from the 25th to the 75th percentile of the community ownership distribution, increases the probability that one will invest in stocks by 4.2 percentage points.

The results from this core specification are identified from those lagged changes in the average ownership of the birth states of one's non-native neighbors that are correlated with one's own changes in ownership through changes in the ownership of one's community. They are not identified from *all* lagged changes in the instrument, but only from those that are orthogonal to the individual and community controls which account for any time variation in the economic status and composition of the community, including any unobservable factors that vary over time at the state level. Also, as discussed earlier, because we restrict our sample to "native" individuals and instrument for the ownership of their community with the average birth-state ownership of their "non-native" neighbors, we are ensuring that the regression results obtained cannot be contaminated by an individual and his community members sharing the same background (i.e., birth state).

In this comprehensive specification, the remaining correlation between community ownership and individual ownership must be caused by one of only two remaining explanations. The first is that there truly are causal word-of-mouth community effects. The second is that, even after our very broad set of controls, there are still some unobserved time-varying factors that are correlated with changes in the lagged instrument as well as changes in individual ownership. For example, suppose that there are two types of communities—red and blue—and that people from red communities tend to move to other red communities, whereas people from blue communities tend to move to other blue communities. Now suppose that there is some change in the information set (e.g., the election of a new President, appointment of a new Federal Reserve Chairman, or some macroeconomic trend) that leads people from blue communities to decide that this is a good time to start investing in equities, whereas people in red communities interpret the change as suggesting that it is not a good time to start investing in equities. Because a person residing in a blue community is more likely to have neighbors who have moved there from other blue communities, it is possible that the correlation could be driven by a common reaction to this new information, rather than word-of-mouth effects.

Most of these alternative explanations are already ruled out by the specification presented in column 3 of Table III. For example, the correlated response must have a trend component to it (rather than be a simultaneous change in behavior) because our instrument is lagged by one year. Furthermore, the correlated response must occur at the community level, and not the state level, because our state-by-year indicators already absorb the latter effects. Given this, we consider the simple word-of-mouth effect to be much more likely to explain our results than the “correlated-response” hypothesis. Nonetheless, to rule out this final competing hypothesis even more completely, we consider one additional specification.

B. Interaction of Community Effect with “Sociability”

To provide additional evidence that our finding is driven by word-of-mouth effects, we use the insight that a word-of-mouth channel is likely to be stronger in those communities in which individuals are more likely to interact socially with their neighbors. In contrast, other explanations for the correlation between community and individual equity ownership—including the correlated response hypothesis—should not vary with the sociability of a community. Thus,

a positive coefficient on the interaction of sociability and community ownership would provide further evidence that the community effect is caused by word-of-mouth effects.

To create measures of community sociability, we take advantage of the DDB Life Style data, which covers the period from 1975 to 1998 and has been used most notably by Putnam (2000) in his well-known study of the decline in social capital in the U.S.²¹ One of the questions asked in this survey is particularly suited for our purposes of understanding whether individuals are likely to share information about consumption and investment decisions with their neighbors. Specifically, respondents are asked to rank their agreement with the statement “My friends and neighbors often come to me for advice about products and brands” on a 6-point scale ranging from 1 = “definitely disagree” to 6 = “definitely agree.” Each response in the survey is characterized by the community to which it belongs, the year in which the survey was completed, and its sampling weight.

This enables us to create for each community and each year the weighted average of respondents’ perceptions of their inclination to seek out advice regarding brands and products. We normalize the variable by rescaling it to the unit interval, thus assigning a 0 to “definitely disagree” and 1 to “definitely agree” for each observation. By averaging across observations in a given community for a given year, we develop a time-varying measure of “sociability” for each community. The key feature of this measure of “sociability” is that it is not systematically correlated with other observable characteristics of the community. Specifically, regression analysis indicates that community-level economic and demographic characteristics (the same community controls employed in our regressions of stock ownership) do not explain community-level sociability (the p-value of this regression is only 0.24).

In Column 4 of Table III, we present the results of a regression specification in which we interact our instrumented measure of community ownership with the sociability measure. The sample size drops to 304,107 because this question was not collected in 1990, and was missing for a small number of MSAs in other years. The key result from this regression specification is that the interaction term is large and statistically significant (47.3, significant at the 10% level), suggesting that in more sociable communities a higher level of community ownership leads to greater individual ownership. These results are strongly consistent with a “word-of-mouth”

²¹ The DDB Life Style data are made available through the generosity of DDB Worldwide of Chicago, Illinois, who retains appropriate rights, including copyright, on these data, while allowing fair use for scholarly and academic research.

community effect and are not easily explained by alternative stories about correlated shocks to information sets.

These results provide stronger identification of community effects in stock market participation than the previous literature. The results are consistent with causal word-of-mouth effects and inconsistent with every major alternative hypothesis. To summarize, the inclusion of individual fixed effects rules out alternative hypotheses relying on time-invariant characteristics, be they observable, such as gender, or unobservable, such as individuals' average risk aversions over the sample period. By restricting the sample to non-movers, the individual fixed effects also operate as community effects, thus controlling for time-invariant community differences such as fixed social norms or the presence of educational or financial institutions. The instrumental variables approach ensures that the correlation is not driven simply by the community sharing a common response to the arrival of new information or institutions in the community such as the arrival of a new broker in town or increased local news coverage of the stock market. The use of a lagged value of the instrument further rules out contemporaneous correlations across similar communities that might be driven by events such as national coverage of a financial event or other relevant news. Further, because we restrict our sample to "native" individuals and instrument for the ownership of their community with the average birth-state ownership of their "non-native" neighbors, we are ensuring that the regression results obtained cannot be contaminated by an individual and his community members sharing the same background (i.e., birth state).

Thus, any alternative explanation for the results must explain not only the correlation in levels of ownership between individuals and their community, as well as the correlation in *changes* in individual and community ownership, but also the correlation in the $t-1$ change in the ownership in the neighbors' birth states with the correlation in time t ownership in the current community. One such story that might pass these hurdles is that communities are changing over time, and that a city with a rapidly growing technology industry, for example, might draw new residents from communities in other states that also have rapidly growing technology industries, and that our regression results are merely picking up these similarities. However, we rule out this and similar stories by controlling, among others, for a large number of time-varying individual and community-level characteristics that capture the changes in the economic and social environment, such as levels of education, racial and ethnic mix, the composition of

occupation, employment sector, a broad set of characteristics of publicly-traded firms in the community, and median home prices. We also include state-by-year fixed effects to control for both observable and unobservable time-varying characteristics across states. Finally, to provide evidence that the effect is driven by social interaction, rather than by some unobserved correlated shock, we show that this community effect operates primarily through its interaction with an independent measure of the community's sociability (in particular, the community's willingness to ask neighbors for advice).

In short, our results provide robust evidence of an economically and statistically significant relation between community and individual ownership that cannot be easily explained by plausible alternative hypotheses. Thus, the most likely explanation for the observed result is a causal one: word-of-mouth communication with friends, co-workers, and neighbors directly influence individuals' correlated decisions to participate in the stock market.

C. Economic Significance

Our key results suggest that a ten percentage-point increase in the share of an individual's community that owns stock, which roughly corresponds to a move from the 25th to the 75th percentile of the community ownership distribution, makes the individual four percentage points more likely to own stock. To assess whether this change is economically meaningful, we compare it to the effect of one particularly influential set of individuals—one's parents. Because this tax panel data tracks dependents, it is possible to construct a "parent-child" data set for a very limited subset of our data. Of course, this sub-sample of the data is far from representative. For example, because the children had to be claimed as a dependent on their parent's return within our sample before starting to file their own return, the sample of children consists primarily of very young adults (typically in their twenties), and we typically observe them filing their own return only for 2-3 years on average. Whereas we hesitate to use this parent-child sub-sample for any primary analysis, it does provide a valuable benchmark for comparing our "community effect" results. Specifically, we run a regression of children's stock market participation against the participation of their parents. The inclusion of individual fixed effects in the regression implies that we are identifying the effect of parents solely from *changes* in the stock market participation rate of one's parents. We follow this approach to be able to make a meaningful comparison to the effect of a *change* in the average participation rate of one's

community. We find that, in a given year, if one's parents switch from not participating to participating in the stock market, one's own stock market participation probability increases by 3.6 percentage points. Thus, one way to interpret the magnitude of the community effect is that a 10 percentage-point increase in the fraction of one's community that owns stock has an effect that is comparable to the effect of a change in the participation decision of one's own parents.

V. Conclusion

A growing literature has been focusing on trying to establish causal community effects, or peer effects, in a wide variety of settings. In general, econometrically identifying such effects is challenging because individuals are not randomly assigned to communities and it is difficult to adequately account for unobservable characteristics that may cause a spurious correlation between individual and community behavior.

This paper provides evidence of causal community effects in the context of stock market participation, an economic behavior that has very important implications for individual welfare and numerous public policy outcomes. Using an instrumental variables strategy that employs variation in the stock ownership in the states in which one's "non-native" neighbors lived at the time they applied for their Social Security numbers (which is generally the state in which they were born), combined with individual and community fixed effects, time-varying individual and community characteristics, and state-by-year effects, we demonstrate that a ten percentage-point increase in community stock ownership makes an individual approximately four percentage points more likely to participate in the stock market. For perspective, this magnitude is comparable to the effect of one's parents switching from not participating in the stock market to becoming stock-market participants upon one's own equity ownership decision. Consistent with a "word-of-mouth" interpretation, we find that this community effect is stronger in the more "sociable" communities, as measured by an independent survey on the likelihood that friends and neighbors ask for advice.

One ramification of these findings is that there may be "externalities" to policies that promote stock ownership, either directly (e.g., the introduction of personal accounts in Social Security) or indirectly (e.g., financial education). Our results suggest that an initial increase in the average stock ownership level in a community may have a "multiplier" effect, making it more likely that other individuals will also begin participating.

References

- Abel, Andrew B., "The Effects of Investing Social Security Funds in the Stock Market When Fixed Costs Prevent Some Households from Holding Stocks," *American Economic Review*, XCI (2001), 128-48.
- Aizcorbe, Ana, Arthur Kennickell, and Kevin Moore, "Recent Changes in U.S. Family Finances: Evidence from the 1998 and 2001 Survey of Consumer Finances," *Federal Reserve Bulletin*, IXC (2003), 1-32.
- Amromin, Gene, and Paul Smith, "What Explains Early Withdrawals from Retirement Accounts? Evidence from a Panel of Taxpayers," *National Tax Journal*, LVI (2003), 595-612.
- Banerjee, A., "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, CVII (1992), 797-817.
- Bernheim, B. Douglas, "A Theory of Conformity," *Journal of Political Economy*, CII (1994), 841-77.
- Brav, Alon, George M. Constantinides, and Christopher C. Geczy, "Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence," *Journal of Political Economy*, CX (2002), 793-824.
- Campbell, John Y., and John H. Cochrane, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, CVII (1999), 205-51.
- Coval, J. D., and T. J. Moskowitz, "Home Bias at Home: Local Equity Preference in Domestic Portfolios," *The Journal of Finance*, LIV (1999), 1-39.
- Coval, J. D., and T.J. Moskowitz, "The Geography of Investment: Informed Trading and Asset Prices," *Journal of Political Economy*, CIX (2001), 811-841.
- DeMarzo, Peter M., Ron Kaniel, and Ilan Kremer, "Diversification as a Public Good: Community Effects in Portfolio Choice," *The Journal of Finance*, LIX (2004), 1677-1715.
- Desai, Mihir A., Dhammika Dharmapala, and Winnie Fung, "Taxation and the Evolution of Aggregate Corporate Ownership Concentration." Forthcoming in Alan Auerbach, James R. Hines Jr. and Joel Slemrod (eds.), *Taxing Corporate Income in the 21st Century*, Cambridge University Press (This paper is a revised version of NBER Working Paper No. w11469, July 2005), 2006.
- Diamond, Peter, and John Geanakoplos, "Social Security Investment in Equities," *The American Economic Review*, XCIII (2003), 1047-1974.
- Ellison, G., and D. Fudenberg, "Rules of Thumb for Social Learning," *Journal of Political Economy*, Vol. 101, Iss. 4 (Aug. 1993), 612-43.

- Ellison, G., and D. Fudenberg, "Word of Mouth Communication and Social Learning," *Quarterly Journal of Economics*, Vol. 110, Iss. 1 (1995), 93-125.
- Fama, Eugene F., and Kenneth R. French, "The Equity Premium," *The Journal of Finance*, LVII (2002), 637-59.
- Feng, F., and M. Seasholes, "Correlated Trading and Location," *The Journal of Finance*, LIX (2004), 2117-2144.
- Graham, Bryan S., "Identifying Social Interactions through Excess Variance Contrasts," University of California – Berkeley, 2005.
- Gruber, Jonathan, "Religious Market Structure, Religious Participation, and Outcomes: Is Religion Good for You?" NBER Working Paper No. 11377, May 2005.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "The Role of Social Capital in Financial Development," *American Economic Review*, Vol. 94, No. 3, June 2004, pp. 526-556.
- Heaton, John, and Deborah Lucas, "Stock Prices and Fundamentals," *NBER Macroeconomics Annual 1999* (Volume 14): Cambridge and London: MIT Press (2000), 213-42.
- Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, "Social Interaction and Stock-Market Participation," *The Journal of Finance*, LIX (2004), 137-163.
- Hoxby, Caroline, "The Power of Peers: How Does the Makeup of a Classroom Influence Achievement," *Education Next*, Summer 2002.
- Ivković, Zoran, and Scott Weisbenner, "Local Does as Local is: Information Content of the Geography of Individual Investors' Common Stock Investments," *The Journal of Finance*, LX (2005), 267-306.
- John Hancock Financial Services, The Ninth Defined Contribution Plan Survey, 2004.
- Mankiw, N. Gregory, and Stephen P. Zeldes, "The Consumption of Stockholders and Nonstockholders," *Journal of Financial Economics*, IXXX (1991), 97-112.
- Manski, Charles F., "Identification of Exogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, LX (1993), 531-542
- Manski, Charles F., *Identification Problems in the Social Sciences*, Harvard University Press, 1995.
- Mas-Colell, Andreu, Michael D. Whinston, and Gerry R. Green, *Microeconomic Theory*, Oxford University Press, 1995.
- Massa, M., and A. Simonov, "Hedging, Familiarity and Portfolio Choice," *Review of Financial Studies*, 19 (2) (2005), 633-685.
- Means, G.C., "The Diffusion of Stock Ownership in the United States," *Quarterly Journal of Economics*, 44 (1930), 561-600.

- Mehra, Rajnish, and Edward C. Prescott, "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, XV (1985), 145-61.
- Putnam, Robert D., *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon & Schuster, 2000.
- Sacerdote, Bruce, "Peer Effects with Random Assignments: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, CXVI (2001), 681-704.
- Staiger, D., and J. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, LXV (1997), 557-586.
- Vissing-Jorgensen, Annette, "Limited Stock Market Participation and the Equity Premium Puzzle," University of Chicago Working Paper, 1999.

Table I
Summary Statistics for the 1987-1996 Pooled Sample

Statistics are based on a panel of tax returns 1987-1996 and are weighted by population weights. Community is defined as Metropolitan Statistical Area (MSA). “Birth” state refers to the state in which individuals lived when they applied for their Social Security numbers (generally the state in which they were born). “Non-native” community members are those whose Social Security numbers were issued in states that do not overlap with their current community. Panel A focuses on equity ownership (as proxied by reporting dividends on tax returns for the year). Panel B focuses on the remaining individual characteristics our specifications employ in subsequent analyses.

Panel A: Equity Ownership Summary Statistics								
	Mean	Percentiles						
	(Std. Dev.)	1st	10 th	25 th	50 th	75 th	90 th	99 th
Does individual own equity?	25.6 (43.6)							
Average equity ownership of community	25.6 (6.7)	9.7	16.6	21.2	26.0	29.9	34.0	41.8
Ownership of individual’s “birth” state	25.6 (5.4)	13.8	18.1	22.1	26.2	29.4	32.7	37.7
Average equity ownership of “non-native” community members’ “birth” states	25.5 (3.2)	17.8	21.9	23.2	25.1	27.8	29.9	32.7

Table I (Continued)

Panel B: Individual Characteristics								
	Mean (Std. Dev.)	Percentiles						
		1 st	10 th	25 th	50 th	75 th	90 th	99 th
Income (non-capital), 1996\$ (in thousands)	43.7 (195.2)	-4.5	1.7	12.8	31.4	58.1	89.5	243.9
Age	46.3 (16.2)	20	27	33	43	58	70	86
Single filing status—male* (in %)	18.6 (38.9)							
Single filing status—female* (in %)	16.0 (36.7)							
Head of HH filing status—male* (in %)	4.0 (19.7)							
Head of HH filing status—female* (in %)	7.0 (25.6)							
Claim mortgage interest deduction (in %)	31.4 (46.4)							
Number of dependents	0.8 (1.1)	0	0	0	0	1	2	4
Self-employed (in %)	6.2 (24.2)							
Schedule E—supplemental income (in %)	15.9 (36.6)							
Pay Alternative Minimum Tax (in %)	0.3 (5.1)							
Contribute to DC plan (in %)	34.7 (47.6)							

* Married couples filing jointly is the omitted category, accounting for 52.4 percent of all observations. Married couples filing separately have been excluded from the sample.

Table II
First-stage Regression of Community Stock Ownership, 1988 – 1996 pooled^a

	(1)	(2)
	No other controls	+ Other controls (including Household fixed effects and State-by-year fixed effects)
Average equity ownership of “non-native” community members’ “birth” states (lagged one year)	49.5 ^{***} (11.0)	24.6 ^{***} (7.6)
F-statistic of instrument in first-stage	20.4	10.6
Adjusted R ²	0.051	0.928
Number of Observations	353,281	353,281

^a Notes: The dependent variable, the ownership of individuals’ communities, is measured in percentage points, whereas all explanatory variables are expressed as raw ratios (i.e., not in percentage points). The sample is restricted to “native” individuals (i.e., those whose current state of residence is their birth state and who have lived in the same MSA for the entire sample period). “Birth state” refers to the state in which individuals lived when they applied for their Social Security numbers (generally the state in which they were born). “Non-natives” are those whose Social Security numbers were issued in states that do not overlap with their current community. The key independent variable is the average equity ownership of “non-native” community members’ birth states. It is lagged by one year to reduce the potential for picking up a spurious correlation that might arise if, for example, there were a simultaneous correlated response across states to the release of new information. Other controls include time-varying household-level controls (specifically, (i) an indicator variable for each percentile of the non-capital income distribution as measured each year, for a total of 99 variables (the 50th percentile is the omitted category, (ii) an indicator variable for each possible age through age 106 (the oldest in our sample), with age 40 as the omitted category, (iii) indicator variables for gender interacted with filing status (single male, single female, male head of household, female head of household), and (iv) numerous other items from the tax return, including dummies for whether the household claims a mortgage interest deduction, self-employment, reporting supplemental income on schedule E, contributions made to tax-qualified defined contribution plans, as well as the number of dependents), household fixed effects (and since the sample is restricted to non-movers, the household fixed effects also control for community fixed effects), state-by-year fixed effects, and a number of time-varying community-level controls. The latter include community-level averages of all of the individual-level controls for income, age, filing status, and so on (for age and income, rather than including community level dummy variables for each point in the distribution, we include a quadratic specification of the community averages). In addition, we use 1980, 1990, and 2000 Census data to compute community-specific time trends for a host of variables, including the distribution of completed education in the community (e.g., fraction of community that completed grades 0-8, 9-11, high school, some college, and college graduates), the distribution of community residents by race and ethnicity (e.g., the fraction of Hispanics, Blacks, Indians, Asians, and other groups in the MSA), the distribution across 13 occupation classifications, 11 industry classifications, and 6 sector classifications (e.g., state and local government, private sector, and so on). We also include trends in the median home value in a community to help capture community specific economic conditions, as well as non-linear controls for the population of the MSA in each year. Community is defined as MSA. Standard errors, shown in parentheses, allow for correlation within each community each year as well as for heteroskedasticity.

***, **, * Significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table III
Instrumental Variables Regression, 1988 – 1996 pooled^a

	(1)	(2)	(3)	(4)
	HH-level controls only	+ MSA-level controls	+ State-by-year effects	+ Sociability of MSA
Equity Ownership within Community	35.0 ^{***} (14.0)	26.3 [*] (14.2)	41.8 ^{***} (15.6)	21.5 (23.2)
Community Asks for Advice				-13.1 (8.6)
Equity Ownership within Community * Community Asks for Advice				47.3 [*] (27.3)
Adjusted R ²	0.781	0.781	0.781	0.778
Number of Observations	353,281	353,281	353,281	304,107

^a Notes: The dependent variable, individuals' stock ownership, is measured in percentage points, whereas all explanatory variables are expressed as raw ratios (i.e., not in percentage points). The sample is restricted to "native" individuals (i.e., those whose current state of residence is their birth state and who have lived in the same MSA for the entire sample period). Equity ownership within the community is instrumented for with the one-year lagged average equity ownership of "non-native" community members' "birth states" (for details concerning the first-stage regression, see Section III.B and Table II). "Birth" state refers to the state in which individuals lived when they applied for a Social Security number (generally the state in which they were born). "Non-native" community members are community members who were issued Social Security numbers in states that do not overlap with their current community. All specifications include household fixed effects (and since the sample is restricted to non-movers, the household fixed effects also control for community fixed effects). Time-varying household-level controls, also included in all specifications, are the same as those included in Table II. See Section IV.A for a description of the time-varying community-level controls added in Column 2. Column 3 also features state-by-year fixed effects (instead of only year effects, as in Columns 1 and 2). The "Community Asks for Advice" variable, utilized in Column 4, is constructed on a community-by-year basis using the DDB Life Style survey respondents' perceptions of how often they are asked for advice regarding products and brands (see Section IV.B for further details). This variable is normalized to range from zero to one, with zero being least likely to be asked by neighbors for advice and one being most likely to be asked by neighbors for advice. Community is defined as MSA. Standard errors, shown in parentheses, allow for correlation within each community each year as well as for heteroskedasticity.

***, **, * Significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Figure 1: Cumulative Percentage Growth in Equity Ownership and Reporting of Dividends on Tax Returns, 1987-1998 (relative to 1989). The SCF provides equity ownership estimates for 1989, 1992, 1995, and 1998. Equity ownership estimates for years between surveys are obtained by linear interpolation.

