Road to Stock Market Participation

Sumit Agarwal[*], *Meghana Ayyagari*[†], *Yuxi Cheng*[‡] and *Pulak Ghosh*[§¶]

This Version: February 2022

Abstract

We study the impact of a national road construction program that brought access to previously unconnected pincodes in India, on stock market participation. Using a unique dataset on the trading behavior of over 13 million individuals, we find that construction of new feeder roads to a pincode increases the number of new investors by 28.7% and the number of trades by 64.5% and the effects are larger for rural vs. urban areas and for pincodes at intermediate levels of development. The stock market participation effects are largely driven by new bank branch openings within three years of the road construction suggesting a financial inclusion channel. We also see greater effects for pincodes more distant from the nearest big city, greater portfolio diversification, and increased trading in companies located farther away, all suggesting an information channel.

JEL Classification: 012, O16, 018, D8, G21

[*]National University of Singapore, Email: ushakri@yahoo.com

[†]School of Business, George Washington University, Email: ayyagari@gwu.edu

[‡]PhD Student, School of Business, George Washington University, Email: chengyx@gwmail.gwu.edu

[§]Indian Institute of Management, Email: pulak.ghosh@iimb.ac.in.

[¶]We are grateful to Viral Acharya, Isha Agarwal, Giorgia Barboni, Indraneel Chakraborty, Dimas Fazio, Ross Levine, Andrea Presbitero, Srinivas Rangan, Andre Silva and conference/seminar participants at the Australasian Finance and Banking Conference, MOFIR Virtual Seminars on Banking for helpful comments and suggestions.

Introduction

An extensive literature has documented the role of financial markets in economic development (Levine [2005] provides a review).[1] And yet stock market participation rates vary widely with the income level across countries (e.g. Guiso et al. [2003]; Giannetti and Koskinen [2010]). The low participation rates in developing countries is particularly puzzling given that developing country households are known to consume less and save more than those in developed countries (Modigliani and Cao [2004]). Understanding the underlying determinants of stock market participation has implications for both the consumption and risk behaviors of households and for aggregate macroeconomic effects.[2]

In this paper, we examine how stock market participation is influenced by infrastructure development, specifically by the construction of feeder roads that connect rural areas. We use comprehensive data on stock trading activity on the National Stock Exchange (NSE) of India by over 13 million individuals/households. The data includes information on the exact pincode location of each trader (or investor), trader demographics such as age and gender, and the specific stocks traded from 2004 to 2015. Using this data, we examine the change in stock market participation around a shock to the rural road network over this period arising from a national rural road-building program (the Pradhan Mantri Gram Sadak Yojana or Prime Minister's Village Road Program, or PMGSY) that connected previously unconnected roads.

A-priori it is not clear that infrastructure improvements should lead to increased stock market activity. The evidence on benefits of improved road connectivity in India is mixed with some studies showing large financing responses to road connectivity (e.g. Agarwal et al. [2021]), while others pointing to limited economic development from improved connectivity and inefficiencies in rural financial markets (e.g. Agarwal et al. [2017]). There are at least two broad channels through which road construction might influence stock market participation. The first is a financial inclusion channel whereby roads bring in greater access to financing. Specifically, in the Indian context,

[1] Also see Levine and Zervos [1998], Demirgüç-Kunt and Levine [1996], Beck and Levine [2004], and Brown et al. [2009].

[2] Limited stock market participation has been shown to have a direct effect on the equity premium (Mankiw and Zeldes [1991], Heaton and Lucas [1999], Vissing-Jørgensen [1999], Brav et al. [2002]); diversification discount, market liquidity, and market crashes (see Basak and Cuoco [1998] and Huang and Wang [2009]). Also see Guiso and Sodini [2013] for a review of this literature.

new roads allow for opening up of new bank branches that allow for new investors to open trading accounts (also called "Demateralized or Demat" accounts). The second is an information channel whereby with increased connectivity, there is greater awareness of investing in the stock market (e.g. Hong et al. [2004], Guiso and Jappelli [2005]).[3]

Our empirical setting offers us several advantages to test the role of infrastructure on stock market participation. First, unlike previous studies where stock market participation is determined from surveys, we can accurately measure the changes in stock market participation over time. Second, we are able to address concerns about the non-random nature of road placement by using the differential timing of road completion in each pincode, and comparing stock market participation in pincodes before and after a road is built, controlling for differences between pincodes that receive roads in different years. Finally, comprehensive data on the universe of bank branch openings from the Reserve Bank of India allows us to test the financial inclusion channel.

We see a significant increase in stock market participation associated with building of new feeder roads. In pincodes that received a new road under the PMGSY program, there is a 18.2% increase in the number of investors and a 55.7% increase in the total number of trades in that pincode-month. Most of these effects are driven by new investors (i.e. traders with less than three years trading experience) (28.7% increase in new investors and 64.5% increase in trades by new investors) and are economically significant, translating into 6 new investors and 214 additional trades per month. While these effects are largest within the first year of road construction, we still see positive and significant (though diminished) long-run effects of roads on stock market participation up to five years. These results are robust to a number of robustness checks including using state-year fixed effects, balanced sample regressions, and placebo tests where on randomizing the road completion date, we find no impact of road construction on stock market participation. We also find consistent results when we use the semi-parametric estimator proposed by Callaway and Sant'Anna [2021] that avoids the concerns with two-way fixed effect estimations with differential timing associated with negative weights and using already treated groups as controls in the estimation.

Next, we find substantial heterogeneity in the impact of road construction across different sub-

[3] In a survey of households in rural and urban India, NCAER [2011] report inadequate information to be a key factor for low rate of stock market participation. The survey also points out that households in villages that are close to urban centers have higher participation rates than households in remote villages.

samples of investors. First, we find greater stock market participation among male investors and mature investors (over the ages of 55) after improved road connectivity than female investors and younger investors respectively. This is consistent with prior literature that documents that men trade more than women (e.g. Barber and Odean [2001]) and stock market participation peaks in the middle-age group (e.g. Van Rooij et al. [2011]). Second, across geographies, we see that new road construction mainly benefits investors in rural pincodes. In rural pincodes, increased road connectivity leads to 71.5% increase in trades and 28.1% increase in investors compared to urban pincodes. When we look at pincodes at different levels of economic development, we find that its not investors in the poorest regions (as measured by consumption per capita) but those in areas at intermediate levels of economic development that participate more actively in the stock market after road construction. These results are robust to using poverty rates as an alternate measure of economic development.

To better understand the mechanisms through which road construction affects stock market participation, we explore heterogeneity in the treatment effects. First, we find support for the information channel by showing that improved connectivity leads to increased stock market participation in more distant pincodes. The number of trades by new investors increases by 0.9% and the number of new investors increases by 0.4% for each additional 10kms between a pincode and the nearest metropolitan area. We also see that following new road construction, there is increased portfolio diversification with an increase of 51.1% in the number of unique tickers in the average portfolio. In addition, when we look at the type of stocks traded, we see that after new road construction in a pincode, investors are more likely to trade stocks of firms whose headquarters are located farther away from the pincode.

Second, consistent with the financial inclusion channel, we see that improved road connectivity in a pincode leads to greater stock market participation when there is a new bank branch opened in that pincode within three years after the road completion. Specifically, with the opening of a bank branch within a 3-year window after the road construction, there is a 9.6% increase in the number of trades by new investors and 8.9% increase in the number of new investors. We find larger effects for bank branch openings of state-owned banks compared to private banks. Given the greater outreach of public banks in India (e.g. Berger et al. [2008]), their role in financial stability and inspiring investor confidence, and lower fee-based incentives to encourage stock market trading, our results on public banks are suggestive of greater demand effects than supply effects.

To explore if the financial inclusion channel is benefiting only the marginal investor who experiences a positive wealth shock, we use rainfall shocks as a proxy for positive income shocks. We find no evidence that a positive wealth shock is moderating the impact of road construction on stock market participation in the pincode suggesting that the benefits of road construction is experienced widely.

To understand the source of the investment money, we look at the consumption and savings behavior of these households using more aggregate data. First, using a survey dataset of households at the district level (the Consumer Pyramids Household Survey data), we find a decline in consumption and increase in income in districts with new road construction. This provides suggestive evidence of capital re-allocation in households post new road construction from reduced consumption to increased investment. Second, using aggregate data on bank deposits, we find that after a pincode is connected by a PMGSY road,there is decline in deposits in the savings accounts in bank branches in that pincode suggesting again capital re-allocation from savings to the equity market.

Finally, we find that after road completion, new investors pick stocks that carry short-term profits but generate negative monetary returns over longer term. We also find investors become more risk-taking as they start to trade on more volatile stocks.

Overall, our paper shows a causal impact of improved road connectivity on stock market participation by Indian households via the information and financial inclusion channel.

Our findings contribute to the large literature documenting different factors responsible for stock market participation including fixed costs (e.g. Vissing-Jørgensen [2003]), financial literacy and household education (e.g.Bernheim and Garrett [1996], Bayer et al. [2009]), Campbell [2006], Calvet et al. [2007], Christiansen et al. [2008], Van Rooij et al. [2011]), social interaction and peer effects (e.g. Madrian and Shea [2001], Duflo and Saez [2002] and Hong et al. [2005]), lack of stock market awareness (Hong et al. [2004], Guiso and Jappelli [2005], Brown et al. [2008]), IQ stanine (e.g. Grinblatt et al. [2011]) and lack of trust (Guiso et al. [2008]). Our paper advances a new mechanism - road infrastructure development - as a way of overcoming many of the factors that lead to low stock market participation. In addition, ours is one of the few papers to explore factors that can affect stock market participation in a developing country.

Previous studies have studied the effect of the PMGSY program on rural households examining the impact on schooling and educational attainment (e.g. Mukherjee [2012]), labor reallocation out of agriculture and productivity (e.g. Shamdasani [2021] and Aggarwal [2018]). Our paper is closely related to Agarwal et al. [2021] but is different from their paper both in the focus and the data used. The focus in their paper is to examine if rural road connectivity increases the probability of getting a loan. They study financing responses to road connectivity in just two of the states in India (Odisha and Uttarakhand) and use loan level data from a single lender. Our focus on the other hand is to examine if rural road connectivity leads to increased stock market participation and if one of the channels through which this occurs is through road connectivity leading to the opening of more bank branches. Our data is also more comprehensive since we are looking at rural road construction under PMGSY in the whole country and have detailed information on the opening of all bank branches in the country.

1 Data and Summary Statistics

1.1 NSE Trading Data

We use comprehensive data on the universe of individual investors' daily trading activity compiled by the National Stock Exchange of India (NSE)[4] over the period 2004 to 2017. For each trade, we have the key elements of a stock transaction including the date of the transaction, account type, tickers traded, the number of shares purchased or sold, and the execution price. Based on the account identifiers, 97% of the accounts are individual investor accounts and the remaining 3% are institutional investors. We restrict our sample to individual investor accounts and to trading of domestic stocks. Thus, we exclude all trading activity on ETFs and stocks issued by Non-Indian public firms.

[4] The National Stock Exchange of India is the leading stock exchange of India with 86% market share. According to the 2018 World Federation of Exchanges Report, it is the world's 11$^{\text{th}}$ largest stock exchange.

The NSE data also contains demographic details on the investors including each investors gender, age, and their residential pincodes. We match the pincodes in the NSE dataset to the official list of 19,252 pincodes published by the Indian government.[5] For observations which do not have a pincode directly listed in the NSE data, we replace their pincode with the most common pincode in the city they located.[6]

Overall, we have trading information for 13,510,473 unique investors located in 19,192 unique pincodes. Table A2 in the Appendix shows the geographical distribution of investors across all the states in India. We see that nearly one fifth of the investors in our sample are located in the state of Maharashtra, whose capital city Mumbai is home to the National Stock Exchange (NSE) and Bombay Stock Exchange (BSE) and since historically been the center of economic activity. Five other states including Gujarat, Tamil Nadu, West Bengal, Karnataka, and Uttar Pradesh account for another 42% of the sample.

All our analysis is at the pincode-month level. Thus, to measure stock market participation, we construct the following two variables: *Number of Trades* defined as the natural log of 1 plus the total number of trades in that pincode in a particular year-month; *Number of investors* defined as the natural log of 1 plus the total number of active investors in that pincode in a particular year-month.

To measure the extent of portfolio diversification in a pincode, we construct the variable *Number of Tickers* defined as the natural log of 1 plus the total number of tickers traded in a pincode in a particular month. We also calculate the monthly volatility of each stock and recognize a stock as a *High Volatility Stock* in a particular year-month if its price volatility is above the sample median for all stocks in the same year-month. We drop the top and bottom 1% outliers to eliminate the influence of outliers. Table 1 presents the summary statistics on each of the variables described above. The average number of trades in a pincode in a month is 602 trades though the range is

[5] The official list of pincodes corresponding to post offices published by the Government of India is available at `https://data.gov.in/catalog/all-india-pincode-directory`. The NSE dataset provides data at the sub-post office level and has 78,647 unique 6-digit pincodes which we then match up to the 19,252 pincodes in the official pincode list.

[6] For instance, the most common pincode for the city of Bangalore in our sample is 560078. We therefore assign the pincode 560078 to all investors in Bangalore whose exact pincode is unknown. We are able to assign pincode to 113,230 investors using this method (about 1% of the entire sample). All our results hold if we were to instead drop the observations with missing pincode.

from 1 trade in rural pincodes such as 212206 (a rural area in Kaushambi District, Uttar Pradesh) and 816118 (a rural area in Dumka District, Jharkhand) to 16,669 trades in metro areas such as 400055 (Mumbai, Maharashtra) and 388009 (Ahmedabad, Gujarat). There are an average of 36 active investors and 79 stock tickers being traded in a pincode-month. Table A1 in the Appendix provides a detailed description of all the variables and Table A2 show the state-wise breakdown of all investors.

To explore the heterogeneity among the investors, we construct the following variables, again aggregated at the pincode year-month level: Using account opening date, we define *New Investors* as investors who opened trading accounts within the last three years and and *Experienced Investors* as those who have had a trading account for more than 3 years. On the basis of age, we categorize the investors into three groups: *Young* (18-30 years), *Middle age*(30-55 years), and *Mature*(55+ years). We also separate the investors by gender into *Male* and *Female* investors. As seen in Table 1, across our sample, 63.1% of the investors are new investors. The majority are middle-age investors (53.3%) followed by young (24.6%) and mature investors (22.1%). Most of the investors are male (91.2%).[7] We then calculate our trading variables, *Number of Trades* and *Number of Investors* in each of the identified demographic groups.

To obtain information on the firms being traded, we merge the tickers in the NSE data to those in Prowess, a database of provided by the Center for Monitoring the Indian Economy (CMIE). Prowess provides detailed financial information on public and private companies including the address of the headquarter location. For each of the public companies in an investor's portfolio, we define *Distance* to be the distance between the geocode location (or latitude/longitude coordinates)[8] of the investor and the public company.[9]

[7] For 5.1% of the investors in our sample the gender is not specified.

[8] The geocode coordinates are calculated for the center of the pincode. To obtain geocodes, we use *LocationIQ*, a commercial Application Programming Interface (API) that uses location identifiers such as state, district, and city to provide latitude/longitude coordinates. We supplement this with *GeoNames* (`https://www.geonames.org/about.html`), an open source database that provides latitude/longitude coordinates for pincodes around the world. While GeoNames is continually updated, the bulk of our data download from GeoNames was done in October 2020.

[9] The distance is computed using the standard formula used in several studies including Ivković and Weisbenner [2005] and Tian [2011]:

$$\begin{aligned} d(location_i, location_j) = \arccos\{&\cos(lat_i)\cos(lat_j)\cos(lon_i)\cos(lon_j) \\ &+ \cos(lat_i)\cos(lat_j)\sin(lon_i)\sin(lon_j) + \sin(lat_i)\sin(lon_i)\}R \end{aligned} \tag{1}$$

where R denotes the radius of the Earth (3,963 statutory miles approximately).

Figure 1 plots the geocodes of all the investors in the NSE sample (blue dots), and the publicly listed companies that they invest in (black dots). Based on the *Distance* measure, we look at investment in companies that are less than 50 kilometers away, 50-100 kilometers away, 100-500 kilometers away, and over kilometers away. Specifically, the average monthly number of trades among the representative investors ranges from 364 (for companies located over 500 kilometers away) to 7 (for companies less than 50 kilometers). The monthly average number of trades for companies 50-100 kilometers and 100-500 kilometers away are 10 and 70 respectively.

In addition to the variation in geographic location of the public companies, we also explore variation in the distance between investors and the nearest cities. For each pincode, based on the 2011 Census we calculate the distance to the nearest cities with population larger than 100,000 people (referred to as *Tier-1* towns by the Census) and 1 million people (referred to as *Metros* by the Census). All distances are computed between geocodes. As seen in Table 1, the average distance between investors' pincodes and a city with population more than 100,000 (1 million) is 45.7 Kms (140.5 Kms) respectively.

1.2 PMGSY Data

While India has one of the largest and densest road networks in the world, till the year 2000, around 30% of its population, or 300 million people, lacked access to all-weather roads. In 2000, the Government of India launched the Prime Minister's Village Road Program (Pradhan Mantri Gram Sadak Yojana or PMGSY) to build all-weather roads and improve rural connectivity to unconnected villages across India. By the end of 2014, the PMGSY program had successfully built or upgraded a road for over 115,000 villages connecting over 30 million households in nearby towns. The focus of the program was to build feeder roads that provide terminal connections between the broader transportation network and previously unconnected villages.

We obtain data on all new roads built under the PMGSY program from the SHRUG (Socioeconomic High-resolution Rural-Urban Geographic) platform created by Asher et al. [2020]. For each new road built under the program, we have the exact completion date (year-month), which is crucial for our identification strategy, and the village and geocode location.[10]

Since the trading data is at the pincode level, we assign the villages to pincodes based on their geocodes. Specifically, all the villages within a 5 kilometer radius from the center of a pincode are assigned to the pincode. By construction, one village can be assigned to multiple pincodes if it is located within the 5 kilometer boundary of several pincodes. Thus, we assume that a paved road built in these boundary villages will have equal effects on all the pincodes it is assigned to. Across all the pincodes of the NSE trading data, we find 8,390 pincodes with at least one paved road built within a 5 kilometer radius.

We further restrict our sample to all pincodes that did not receive a new PMGSY road prior to October 2004 but received a new road between October 2004 and February 2015. Thus, all the pincodes in our sample receive a new road at some point during the sample period. In the final sample for analysis, we have trading variables across 4,154 unique pincodes in India.[11] Figure 2 provides a map for the 4,154 pincodes (blue dots) and the corresponding villages (red dots) in our sample and shows that most of the new roads under this program was constructed in the northern part of India. We define dummy variable *Connect* to take the value one in the month of road completion and thereafter. *Connect* is defined for the earliest date when a new paved road is finished under the PMGSY program in any of the villages associated with a pincode. [12] As shown in Table 1, 55.3% of the sample pincode-month observations have the *Connect* variable equal to 1.

1.3 Bank Branch Data

To analyze if the effect of road construction on stock market participation is through the financial inclusion channel, we use data on the opening of bank branches across the country. Specifically, we use confidential and proprietary bank branch level data from the Reserve Bank of India (RBI).

[10]The geographic unit targeted for road construction in the program was a habitation (cluster of population), which is the smallest rural administrative unit in India. As discussed in Adukia et al. [2020], many villages have only one habitation and many habitations were pooled to the village level for the purposes of the program.

[11]Out of the 8,390 pincodes, 3201 already had a new road under the PMGSY program before October 2004; 1022 pincodes have always been connected by a PMGSY road, and 13 pincodes were never connected by a PMGSY road during our sample period. All these pincodes are removed from the sample.

[12]Thus, one pincode could be associated with multiple villages with new road completion, either in the same year or over different years. For instance, the pincode 721634 is associated with 43 villages within the 5 kilometer threshold, with new roads constructed between the years 2007 and 2014. On average, over our sample period, one pincode is associated with 8 villages with a new road.

For each bank branch location, we have information on the pincode and the opening date of the branch/office. For the 102,968 bank branches that opened across India between 2004 and 2018, we use the pincode information of each branch and match them to the pincodes in our NSE trading sample. Over our sample period, we have 23,919 bank branches that opened in 3,557 pincodes in our sample. For pincodes in the NSE trading data that can not be matched with the bank branch opening data, we assume there is no new branch opening for these pincodes during our sample period.

Next, we focus on bank branches that open within a 36-month window after the road construction to isolate the effect of bank branch openings. *New Branch Dummy* is an indicator variable which takes the value 1 if a new bank branch is opened in the pincode within a 36-month window after the road construction under the PMGSY program in that pincode. *Number of New Branches* is the total number of new bank branches that open within the same 36-month window after the road construction under the PMGSY program within the pincode. Table 1 shows that 43.2% of the pincodes in our sample have at least one new bank branch open in the 36-month window after receiving the new road, and the average number of new branches is around one.[13] Our analysis is robust to using both 12-month and 24-month windows instead of 36-months.

We are also able to classify the bank branches into *State-owned branch* and *Private branch* based on ownership type.[14] Each of these are dummy variables take the value 1 for the corresponding bank type and 0 otherwise. We also use the *Number of State-owned branches* and *Number of Private branches*.

1.4 Measures of Economic Development

To explore the heterogeneity across regions with different levels of economy development, we first separate the pincodes into rural and urban using the post office type of a pincode.[15] As seen in

[13] 1,811 pincodes out of 4,154 pincodes have at least one new bank branch opened within the 36-month window after the road construction.

[14] Public banks include the State Bank of India, all other nationalized banks and regional rural banks while private banks include all non-state owned banks including foreign banks.

[15] Post offices in India are classified into five categories by the Department of Posts: DO(Divisional office), GPO(General Post Office), HO(Head office), SO(Sub Office) and BO(Branch office). We consider a pincode as belonging to a rural area only if its post office type is BO. As a robustness check, we also use an alternate ranking of the pincodes based on population as used by the Indian Space Research Organization (ISRO) - Metro, Tier-1, Tier-2,

Table 1, of the 4,154 pincodes in our sample, 3,293 pincodes (79.3% of the sample) belong to rural areas, while the rest of pincodes are in the urban areas.

In addition to identifying rural and urban areas, we also use *Consumption per capita* data obtained from the SHRUG as our measure on the economic development level across different geographic areas.[16] Since the *Consumption per capita* is defined at the village level in the original data, we use the weighted average across all the villages associated with the pincode, weighting by the inverse distance between the village and the pincode to obtain measures at the pincode level. We further split the pincodes in our sample into three major categories: *Developed*, where the values of the *Consumption per capita* are above the sample median; *Intermediate development* where the values of *Consumption per capita* fall between median and 10th percentile values of the sample, and *Least developed* where the values*Consumption per capita* are below 10th percentile value of the entire sample. Table 1 shows that the average pincode in our sample has consumption/capita of 16478.11 INRs (equivalent to 331.42 USD according to the average annual exchange rate of 49.72 INR/USD over our sample period 2004 to 2015).

1.5 Rainfall Data

Following a large literature that uses rainfall shocks to proxy for income shocks, we collect rainfall data from the the Center for Environmental Data Analysis (CEDA) Archive.[17] The rainfall data is available at the monthly 0.5×0.5 geo-grid level. We assign the geo-grid to a pincode if the distance between the pincode and the center of the geo-grid is less than 30 kilometers.[18] We then use the inverse of distance as weights and calculate the pincode level weighted average monthly rainfall.[19]

Following Jayachandran [2006], Kaur [2019] and Gupta [2020], we define a pincode-month ob-

and Tier-3 pincodes - and classify Tier-3 pincodes as rural. Our results are materially the same under this alternate ranking.

[16] A large literature in development economics has used Consumption per capita as a reasonable proxy for long-term income/wealth especially in low income countries (see Ravallion [1995], Hamilton [2003]). We also use *Poverty Rate* data from the Census as an alternate measure for regional economic development and find similar results.

[17] The rainfall data is publicly available from CEDA's website `http://archive.ceda.ac.uk/`

[18] Several studies suggest a radius of 30 Kms is optimal when estimating the regional rainfall using the Inverse Distance Weighting (IDW) method. See Chen and Liu [2012] and Noori et al. [2014] for instance.

[19] For instance, if geo-grids A, B, and C have been assigned to pincode 200022, and the distance from each geo-grid to pincode 200022 is 10 kilometers, 5 kilometers, and 25 kilometers respectively, the corresponding weights for geo-grids A, B, and C are $\frac{1/10}{1/10+1/5+1/25}$, $\frac{1/5}{1/10+1/5+1/25}$, and $\frac{1/25}{1/10+1/5+1/25}$ respectively.

servation to have a positive rainfall shock if the rainfall is above the 80^{th} percentile for that pincode and calendar month across all sample years and a negative rainfall shock if the rainfall measure is below the 20^{th} percentile for that pincode and calendar month across all sample years.[20] Thus, the variable $Rainfall\ Shock$ takes the value 1 for positive shocks, -1 for negative shocks, and 0 otherwise. Over our sample period, 13.9% of the pincode-month observations are classified as positive rainfall shocks, and 11.2% of the pincode-month observations are classified as negative rainfall shocks.

2 Empirical Strategy

To study the causal impact of roads on stock market participation, we follow the empirical strategy in Adukia et al. [2020]. Our main empirical specification is a panel fixed effects regression that exploits the timing of road construction, within the set of all pincodes that received new roads between October 2004 and February 2015 under the PMGSY program. The specific equation we estimate is:

$$Log(Y_{i,t}) = \beta \cdot Connect_{i,t} + \xi_i + \kappa_t + \Theta_{D,Y} + \varepsilon_{i,t} \tag{2}$$

where Y are the two main measures of stock market participation in pincode i in year-month t: *Number of Trades* and *Number of Investors*.

The main coefficient of interest is β, which measures the impact of a new road on pincode-level stock market participation. Every pincode in our sample has $Connect_{i,t} = 0$ in the first month of our sample and $Connect_{i,t} = 1$ in the last month of our sample. Pincode fixed effects, ξ_i, are used to capture other time invariant factors within the pincode, and to control for systematic differences between pincodes which received roads at different times. Month fixed effects, κ_t, captures any variation along time that might impact stock market trading. In addition, we also add the district-year fixed effects $\Theta_{D,Y}$ to control flexibly for differential growth in stock market participation across districts. This alleviates concern that districts located in states with more effective governments simultaneously built roads and also provided other government services; it also controls for any broader regional trends in stock market participation that might be correlated

[20] We already remove extreme cases such as hurricane before we identify positive/negative rainfall shocks.

with road construction. The standard errors are clustered at the pincode level to account for serial correlation in the dependent variable. In our baseline estimation, we use a balanced panel where trading variables for pincode-months with no trading activities are coded as zero.

The crucial assumption in our setting is that in the absence of PMGSY, trading activities would have followed the same trajectory in pincodes that received a paved road in different years, after partialling out the various types of fixed effects. We test the time trends of the road effects following a similar setting to equation (1):

$$Log(Y_{i,t}) = \sum_{k \in (-5,5)} \gamma_k \cdot D_k + \xi_i + \kappa_t + \varepsilon_{i,t} \tag{3}$$

where D_k are a series of indicator variables that take the value 1 in the k^{th} year relative to the construction of road for that pincode. For instance, $D_2 = 1$ indicates 2 years after the road construction for the pincode. Unlike the setting in equation (1) where all the variables are defined at the monthly level, we define the variables annually for analyzing the time trends. Since all the pincodes in our sample are eventually "treated", as suggested by Borusyak and Jaravel [2017] and Adukia et al. [2020], (3) can only be estimated with two coefficients being omitted. We therefore, use the coefficients in the year before road construction ($k = -1$) and in the first year of this setting ($k = -5$) as our reference group.

To explore the heterogeneity among investors, we replace Y in the above equation with the same measure constructed for different sub-groups: *New/Experienced investors, Male/Female investors,* and *Young/Middle-age/Mature investors.* Next, we explore whether the effects vary across regions with different levels of economic development. We first interact the *Connect* dummy with a dummy for Rural/Urban (*Rural*) pincodes and estimate the following regression:

$$Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} + \beta \cdot Connect_{i,t} \times Rural_i + \xi_i + \kappa_t + \Theta_{D,Y} + \varepsilon_{i,t} \tag{4}$$

where $Y_{i,t}$ are the different measures of stock market participation as before. We then interact the *Connect* dummy with an indicator for different levels of economic development (*Developed* (omitted category), *Intermediate development*, and *Least Developed*) for the pincode as proxied by

Consumption per capita and estimate the following regression:

$$
\begin{aligned}
Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} &+ \beta_1 \cdot Connect_{i,t} \times Intermediate\ development_i \\
&+ \beta_2 \cdot Connect_{i,t} \times Least\ developed_i + \xi_i + \kappa_t + \Theta_{D,Y} + \varepsilon_{i,t} \quad (5)
\end{aligned}
$$

where *Intermediate developed* and *Least developed* are dummy variables that take the value 1 if the *Consumption per capita* values for the pincode is between sample median and 10th percentile, and below 10th percentile of the sample respectively.

We then explore three channels through which road construction might affect stock market participation. First is the information channel wherein new road construction can lead to greater stock market investing by reducing the information asymmetry. A large literature has suggested that non-participation can be explained by moderate fixed costs related to entry and participation (e.g. Haliassos and Bertaut [1995], Vissing-Jørgensen [2002] Vissing-Jørgensen [2003]) which include the costs related to learning about stocks, and acquiring information. With increased connectivity via new road construction, investors in rural areas have easier and better access to "hard" information (e.g. better and more accurate information on companies' operating conditions, profitability, and growth opportunities) which could increase local stock market participation. In addition, newly connected roads also enable the transmission of "soft" information (e.g. rumors, or stock trading activities from remote friends or family members) via increased social interactions, peer effects that have been shown to increase stock market participation (e.g. (Hong et al. [2004], Guiso and Jappelli [2005], and Brown et al. [2008]). To test the information channel, we perform three types of tests: First, we investigate if the effect of new road construction is larger for investors in pincodes that are more remote from larger metropolitan areas. We estimate a model similar to our baseline specification but change the $Y_{i,t}$ to be specific to companies located a certain distance away from the local pincode. Specifically, we use the following specification:

$$
Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} + \beta \cdot Connect_{i,t} \times Distance_i + \xi_i + \kappa_t + \Theta_{D,Y} + \varepsilon_{i,t} \quad (6)
$$

where $Distance_i$ is a continuous variable (in 10km units) calculated as the shortest distance between the pincode location of the investor to the nearest city with population above 100,000 people (*Tier*

1 cities) and the nearest city with population over 1 million people (*Metros*).

Second, we examine if there is a greater degree of portfolio diversification upon new road construction. We use the main baseline specification but with *Number of Tickers* traded in the pincode i in month t for the dependent variable $Y_{i,t}$. Third, we explore if there is a larger number of trades in companies that are more distant. We use the same baseline specification as before but replace the dependent variable, *Number of trades*, with Number of trades in companies that are located less than 50 kms away, 50-100 kms away, 100-500kms away, and over 500kms away.

An alternate channel through which roads allow for greater financial infrastructure in the local region and, thus, enabling greater stock market participation. Several recent studies including Agarwal et al. [2021] and Das et al. [2019] show that construction of new roads and road upgradation leads to increased financing to households, especially in rural areas, allowing them to make the best use of new productive opportunities. To isolate this, we examine the effect of new bank branch openings in the pincode within 36 months of the road completion date in the pincode. Specifically, we estimate the following equation:

$$Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} + \beta \cdot Connect_{i,t} \times New\ Branch\ Dummy\ (\text{or } Number\ of\ New\ Branches)_i + \xi_i + \kappa_t + \Theta_{D,Y} + \varepsilon_{i,t} \quad (7)$$

where $Y_{i,t}$ are the different measures of stock market participation as before.

There are several explanations for why we might expect bank branch openings after road construction to aid in stock market participation - first, it could be a simple enabling story where villagers are now able to open trading accounts with the bank;[21] second it could be a demand side story where villagers reallocate capital from consumption to more productive uses via savings accounts in banks; or third it could be a supply side story where banks actively recommend stock market investments to newly connected villagers.

While we are open to all of these interpretations, we try to investigate more closely if we can separate out these mechanisms. Using data on bank ownership we explore if our results are driven

[21] The Securities and Exchange Board of India (SEBI) requires everyone who wants to trade Indian securities to have a demat account (or dematerialised account), which is an electronic record tracking ownership of tradable assets.

by the opening of state-owned (or public) bank branches versus private bank branches. First, on the enabling mechanism, state-owned banks in India have been shown to have greater outreach (e.g. Berger et al. [2008]) than private banks. Second, a large literature on bank ownership shows that lending by state-owned banks is less pro-cyclical and hence they play a critical role in financial stability (see for example, Brei and Schclarek [2013],Cull and Peria [2013], Bertay et al. [2015], and Bosshardt and Cerutti [2020]) which in turn is suggestive of greater investor confidence in public banks. Both these explanations (enabling and greater demand) would suggest that we should find stronger effects for state-owned banks than private banks. On the other hand, our sample is restricted to trading on individual stocks, while the investment products offered by banks are usually mutual funds, credit deposits and exchanged traded funds (ETFs). Therefore, it is less likely that bank financing advisors would recommend investors to buy individual stocks directly. Furthermore, there is some evidence suggesting that public banks earn less fee-based income and thus have less incentives in interest income from stock brokerage (e.g. Pennathur et al. [2012]). Hence if it is only the supply side channel, we should expect weaker effects for the opening of state-owned bank branches compared to private bank branches. Finally, in section 4.4 we use the household-level consumption data in several districts to analyze whether the completion of new roads leads to different consumption behaviors within households, which could help validate the demand side story.

Finally, we test whether the new investors post new road construction are marginal investors by looking at the effect of income shocks. While most studies predict a positive association between wealth and stock market participation, interpreting it as a financial participation cost (e.g. Vissing-Jørgensen [2003], Calvet et al. [2009], Calvet and Sodini [2014]), Brunnermeier and Nagel [2008] and Andersen and Nielsen [2011] find that sudden windfalls do not increase stock market participation. Households in rural India face high income volatility due to agricultural income being their main source of income, which in turn, is dependent on rainfall. Studies such as Jayachandran [2006], Kaur [2019] and Gupta [2020] have shown rainfall shocks to be a good proxy for local income shocks. Following this setting, we test the income shock channel using the equation:

$$Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} + \beta Rainfall\ Shock_{i,t} + \gamma Connect_{i,t} \times Rainfall\ Shock_{i,t} + \xi_i + \kappa_t + \Theta_{D,Y} + \varepsilon_{i,t} \quad (8)$$

where *Rainfall Shock*$_{i,t}$ is the rainfall shock defined at the pincode-month level. If we find that the effects of new road construction are driven entirely by investors experiencing positive income shocks (positive and significant interaction term γ), then it would suggest that it is the fixed financing costs of participation that is driving stock market participation.

3 Results

3.1 Does Road Construction lead to Increased Stock Market Participation?

Table 2 reports the main results of the impact of roads on stock market participation. Columns 1 and 2 report the estimation results for all investors, columns 3 and 4 for new investors, and columns 5 and 6 for experienced investors. In the sample of all investors, the construction of a newly paved feeder road to a pincode leads to a 55.7% increase in the Number of Trades and 18.2% increase in the *Number of Investors* in that pincode-month. Economically, these effects translate to 334 additional trades and 7 more investors in each pincode for each month on average. These results are almost entirely driven by new investors. Construction of a newly paved new road leads to a 64.5% increase in the number of new investors and an 28.7% increase in stock trading by new investors while the number of trades by experienced investors goes up only marginally by 39.6% and the number of experienced investors increase by 15.4%. Economically, the new road leads to 6 more new investors, 1 more experienced investors and 214 additional trades by new investors.

Figure 3 shows the dynamic effects by plotting the coefficient estimates of the indicator variables described in equation (3). For each of the dependent variables, the coefficients become positive and significant only in the year of the road construction and thereafter, with the maximum effect in the first year after road construction. While the magnitude of the coefficients slightly diminish after the first year, they are still positive and significant suggesting long-run effects of roads on stock trading activities. The coefficients from this specification test whether there is a trend break in treated pincodes relative to their own linear trend, however, they could not test either pre-trends or averages trends. We then do a *F-test* on the coefficients for all pre-treatment coefficients and our estimation results reject the existence of non-linear pre-trends ($p = 0.58$).

As a robustness check, we run placebo regressions to rule out the possibility that it is some other type regional shocks rather than the road constructions are driving the main effects. First, we randomize the road completion date for all pincodes across the entire sample period in panel A of Table A3 of the Appendix. Second, we randomize the road completion date for all pincodes within the same month in panel B of Table A3 of the Appendix. For each setting, we run 500 samples and average the regression coefficients and standard errors across these 500 regressions. The placebo tests allow us to maintain the distribution of completed roads across pincodes while disrupting the completion date for road construction. Table A3 of the Online Appendix shows that the coefficients for *Connect* dummy are *not* economically and statistically different from zero. Overall, our robustness checks support the hypothesis that construction of new feeder roads has a positive impact on on stock market participation, especially for new investors.[22]

A recent literature in econometrics has raised concerns on DiD estimations with two-way fixed effects (TWFE) specifications with heterogeneous treatment times across different groups (see Borusyak et al. [2021], Callaway and Sant'Anna [2021], Goodman-Bacon [2021], De Chaisemartin and d'Haultfoeuille [2020], and Sun and Abraham [2021]). Specifically, the literature shows that the TWFE estimator is biased because it uses the already-treated groups as controls. Furthermore, De Chaisemartin and d'Haultfoeuille [2020] demonstrate that the negative weights associated with the heterogeneous treatment time could potentially bias the estimation results. To address this, we estimate the staggered DiD regression following the statistical package `csdid` developed by Callaway and Sant'Anna [2021] on our entire sample. The semi-parametric DiD estimator proposed by Callaway and Sant'Anna [2021] avoids the weighting problems associated with the TWFE estimations and never uses the previously treated groups as the control group in the estimation. The online appendix results in Table A4 and Figure A1 confirm the significantly positive effects of road on all stock trading variables that we have in the baseline setting.

[22]While the main specifications are based on the trading activities of new investors, in the unreported results, we have the similar results when we focus on trading activities of all types of investors.

3.1.1 Heterogeneous Effects

Tables 3 and 4 report the heterogeneous effects on stock market participation among different demographic groups and across different regions respectively. Panel A of Table 3 shows that while construction of a new road is associated with increased stock market participation for both male and female investors, the economic effects are much larger for male investors. Economically, a new road construction in a pincode leads to 188 additional trades by male investors (compared to 26 by female investors) and 5 additional male investors (compared to 1 female investor) participating in the market. Panel B of Table 3 shows that a new paved road increases the number of trades by 46.2%, 67.2%, and 22.4% respectively among the young, mid-age and mature investors, which is economically equivalent to 35, 160, and 19 new trades by young, mid-age and mature investors respectively. We also see that largest increase in number of investors is among the middle-age investors (15.2%) followed by 14.2% for young and 9.7% for mature investors, which are economically one more investors for each age group. These results show that building of new roads impacts male and older investors, two groups who previous literature has shown to participate more in the stock market (e.g. Barber and Odean [2001], Van Rooij et al. [2011], Constantinides et al. [2002]).

In columns 1 and 3, the coefficients for all the interaction terms are positive and significant, suggesting that the positive effects of roads on stock market participation are larger in rural areas compared to urban areas. When we look across different levels of economic development, we see that road construction has the greatest impact on stock market participation in areas with intermediate development - columns 2 and 4 show that the completion of the new road leads to 14.6%(12.6%) increase in trades and investors respectively in areas with intermediate development. The interaction of *Connect* and *Least Developed* is not statistically significant suggesting that compared to the developed regions, roads stimulate stock market participation in areas that are relatively less developed but not necessarily the least developed ones.[23] In addition, we also plot the time trends of roads impacts across areas with different levels of development in Figure 4. Consistent with our regression results, the positive effects of roads on the different measures of

[23] In Appendix Table A5, we undertake robustness checks where we use an alternative definition of *Rural* variable (Tier-3 pincodes as identified by the Census), and *Poverty Rate* as the regional economic developed measure. The results are similar to the ones in Table 4.

stock market participation are largest for areas of intermediate economic development.[24]

4 Mechanisms

In this section, we explore the possible underlying mechanisms through which road construction leads to increased stock market participation. In section 4.1, we test whether additional connectivity brought about by the new road construction both reduces information asymmetry and generates peer effects, leading to increased stock market participation, which we refer to as the *Information Channel*. In section 4.2, we explore whether completion of new roads leads to the improvement of financial infrastructure in the local region allowing for greater stock market participation, which we refer to as the *Financial Inclusion* channel.

4.1 The Information Channel

We report our results for the information channel in Table 5. Panel A of Table 5 examines if the effects of roads on stock market participation are larger for pincodes located further away from large metropolitan cities. The *Distance* measure (in 10km units) captures the shortest distance between the pincode with the new road construction and the nearest city with more than 100,000 people (Tier 1 city) and 1,000,000 people (Metro city) respectively. Columns 1 and 2 show the results for number of stock trades by new investors and columns 3 and 4 report results for the number of new investors. The results show that the benefits of new road construction on stock market participation are particularly strong for more distant pincodes, especially those that are farther away from metropolitan cities. The number of trades by new investors increases by 1.1% and the number of new investors increases by 0.4% for each additional 10kms between a pincode and the nearest metropolitan city with population more than 100,000 (1 million) people respectively.

An alternate test of the information channel is whether investors hold more diversified portfolios after the construction of the new road. Column 1 of panel B of Table 5 uses *Number of tickers* as an alternate dependent variable and shows that following the construction of a new feeder road

[24] A similar set of results using *Poverty Rates* as the economic development measure is shown in Table A5 and Figure A2.

to the pincode, the number of unique traded tickers increases by 51.1%, which is equivalent to an increase of 43 tickers among all new investors. In unreported results, we find these results to hold across different samples of investors, including male, female, young, mid-age, and mature investors.

In columns 2 to 5, we examine the effects of road construction on the number of distant stocks traded by examining the number of trades in public companies that are located less 50 Kms, 50-100 Kms, 100-500 Kms, and more than 500 Kms away respectively from the pincode. The coefficients of the *Connect* dummy are positive and significant for all columns. The number of trades on public companies located less than 50 Kms do not increase, while for companies locate further away, the stock tradings on companies located 50-100 Kms, 100-500 Kms and over 500 Kms away increases by 1.9%, 21.2% and 60.6% respectively after the completion of the new road.

4.2 The Financial Inclusion Channel

The results of the financial inclusion channel are reported in Table 6. Columns 1 and 2 in panel A of Table 6 report the results for *Number of Trades* by new investors and columns 3 and 4 report the results for *Number of Investors* (new investors). The coefficients for the interaction term on *Connect* and *New Branch dummy* are positive and significant in columns 1 and 3 suggesting that pincodes that had new bank branches open within three years of the road completion have 9.6% increase in stock trades, and 8.9% increase in the number of new investors participating in the stock market. Columns 2 and 4 also report significant results for the intensity of new branch openings. Each additional bank branch opened within three years of the road construction leads to 5.3% increase in stock trades and 1.3% increase in the number of new investors.

As discussed before, the opening of bank branches could result in decreased participation costs since banks allow for establishment of demat and trading accounts, making investing easy for the casual and beginner investor. It could also be a demand side story where clients who open bank accounts are able to reallocate capital from consumption to investing or a supply side story where banks actively recommend investing in individual stocks to clients. While we are unable to differentiate between these three channels, we try to make some progress by using information on the bank ownership type of the new branch opening. State-owned banks have greater outreach and are less likely to promote trading in individual stocks because of lower fee based income compared to private banks. So even if there are supply side effects at play, these should be less important for state-owned banks.

In panel B of Table 6, we interact *Connect* dummy with *State-owned branch* and *Number of state-owned branches* Columns 1 to 4 report results for *Number of New Trades* and columns 5 to 8 report the results for *Number of New Investors*. We find the interaction of State ownership of the new bank branch and *Connect* dummy to be positive and significant suggesting that it is more likely the demand side effects rather than supply side effects driving stock market participation.

4.3 Is there a reallocation from consumption and saving to investment?

In this section, we explore if the financial inclusion channel is consistent with households having greater opportunities to reallocate capital from consumption and savings to more productive uses such as investment.

4.3.1 Allocation from consumption

While we don't have consumption data for individual investors, we try to shed some light on this using aggregate statistics at the district level from the CMIE Consumer Pyramid Household Survey (CPHS) data. CPHS are household surveys designed to capture household well-being in India with detailed information on income and expenditure patterns each month since January 2014. Thus, this data is available only for the last year of our sample period since the PMGSY data ends in February 2015. In addition, the CPHS data is at the district level, while the PMGSY data is at the village level. To take into account these two issues, we alter our empirical design in this section in the following ways: First, we only focus on districts that had PMGSY roads completed between January 2014 and December 2015 (so we allow for at least 10 months of consumption data after the last road completion). If a district had multiple roads completed during this time period, we use the earliest road completion date to identify pre-treated and post-treated groups. Second, we only include a district in our sample if the district does not have any roads in a continuous five-year window before the new road completion date. Third, similar to our NSE sample, we only include the household in our sample if they have records in both the pre-treated and post-treated period. These criteria provide a long enough window for the effects of any prior road construction to become steady, thereby allowing us to isolate the real effects of current roads on household income and consumption. The final CPHS sample contains 145,958 household-month observations over 7,178 households across 16 districts over the 14 month period. For each household-month, we calculate the log transformation of the total income and total expenditure after removing outliers below the 1^{st} percentile and above the 99^{th} percentiles. On average, the representative household in the sample has income level of 8165.3 Indian Rupees (164.26 USD), and consumption level of 6789.17 Indian Rupees (136.58 USD).[25]

To analyze the effects of roads on consumption behavior, we estimate the following equation:

$$Y_{h,t} = \beta \cdot Connect_{d,t}^{CPHS} + \eta_h + \kappa_t + \varepsilon_{h,t} \tag{9}$$

where Y are *Log(Consumption)* and *Log(Income)* which are the log values of household consumption and income for household h in month t. $Connect_{d,t}^{CPHS}$ is an indicator variable which equals to 1 if a district received a new PMGSY road between 2014 and 2015, and didn't receive any road during the past 5 years. η_h represents the household fixed effects. To identify heterogeneous effects of roads on consumption across households with different income levels, we interact the $Connect^{CPHS}$ dummy with a dummy for *High Income* households where High Income households are defined based on whether the average household income is above the median, 75^{th}, or 90^{th} percentile of monthly household income in our sample.

Panel A of Table 7 shows that the completion of new roads leads to a decrease in average consumption while increasing the average household level income. These results provide suggestive evidence of capital re-allocation in households after the construction of a new paved road to that pincode. Panel B of Table 7 shows that high income households see greater declines in consumption and increases in income after the road construction. Overall, these results provide suggestive evidence of capital re-allocation in households after the construction of a new paved road to that pincode, especially in high income households.

[25] We use an average annual exchange rate of 49.71 INR/USD over our sample period 2004 to 2015.

4.3.2 Allocation from saving

To analyze whether there is an allocation of capital from savings in the bank to the equity market, we use the bank deposit data from the RBI-Basic Statistical Return (BSR) dataset. The RBI-BSR data reports annual deposit information in every branch of every scheduled commercial bank in India. For each branch-year, we have information on deposits in the current account (business account), savings account and term account, as well as the aggregate amount of deposit. We focus our analysis on the deposits by individuals rather than other type of depositors.

We restrict the RBI-BSR sample to 919,034 branches across India whose location information is available during 2004-2015. Further, we narrow our sample to branches that are located in the 4,154 pincodes that are part of our main analysis. This yields 161,422 branch-year observations across 23,580 bank branches over 3512 pincodes between 2004 and 2015.[26] After removing outliers, the sample average deposit in the current, saving, and term account is 10634.85, 118609.2, and 135471.1 Indian Rupees respectively.

We use the following equation to estimate the effects of roads on savings:

$$Deposit_{b,i,t} = \beta \cdot Connect_{i,t} + \phi_b + \kappa_t + \varepsilon_{b,t} \tag{10}$$

where $Deposit_{b,t}$ is the log value of total amount of deposit in each type of account for individual depositors in branch b. $Connect_{i,t}$ is an indicator variable which equals 1 if the pincode i, where branch b is located, received a new PMGSY road in year t and thereafter. ϕ_b and κ_t represents the branch and year fixed effects. If investors were to indeed move capital from the savings account to the equity market after the road construction, we should observe declines in bank deposits, especially in the savings account, after the pincodes are connected by new roads.

The effects of improved road connectivity on bank deposits are reported in Table 8. Cols 1-3 report results for each of the three accounts and col 4 reports results for aggregate deposits. The coefficients are statistical insignificant for current account and term account, but negative and significant for savings deposits and aggregate deposits. These results suggest the re-allocation

[26] There are 642 pincodes do not have any branches located in.

of capital from bank deposits to the stock market, providing another supporting evidence on the financial inclusion channel.

4.4 Do roads benefit only the marginal investor?

In this section, we examine if the effects of roads on stock market participation through the financial inclusion channel is only for the marginal investor experiencing a positive wealth shock, as proxied by a positive rainfall shock. Before we conduct our regression analysis, we first investigate if there is any seasonality in terms of trading activities across one year. For each month, we plot the average distribution of *Number of New Trades* and *Number of New Investors* across all the pincodes in our sample in Figure 5. The monthly average values for both *Number of New Trades* and *Number of New Investors* are consistent across all the months in a year, indicating there is no seasonality for the trading activities among investors.

Table 9 reports the regression results where we use the concurrent rainfall shock and 1-month lagged rainfall shock interacted with *Connect* dummy. The coefficients of the interaction terms for the *Number of New Trades* are negative but statistically insignificant. When we look at *Number of New Investors*, the interaction of *Connect* x *Rainfall Shock* is negative and significant, suggesting that with improved road connectivity a positive income shock is not associated with greater number of new investors participating in the market though the overall effect of *Connect* is still positive and significant. Overall, these results show that it is not the case that the financial inclusion channel is driven only by investors who experience a positive income shock.

5 Monetary Profits & Risk-Taking

In this section, we investigate whether new road construction has implications for the risk-taking by investors and monetary profits. Following Barber et al. [2009], we calculate the monetary profits (and losses) from trades by calculating the net change on stock trades by investors. For instance, if an investor buys 1000 shares of company A on Jan 1st, 2012 and sells 600 shares on company A on the same day, we calculate the investor's net holding change of company A as 400 shares, and code it as a *net buy* for the day. If the number of shares sold is larger than the number of shares bought, we code it as a *net sell*. The purchase price of the stock is then calculated as the difference between the buy value and sell value divided by net shares bought. We assume that investors hold the stocks for a fixed period of time and consider different time windows - namely 1, 10, 25, 140 trading days. For each stock, for *net buy* activity on day t, we calculate the monetary profits for holding period X (where $X = 1$, 10, 25, and 140) days using the formula:

$$\text{Profits}_t^{\text{Net Buy}} = \text{Net Buy Value}_t \times \frac{\text{Price}(t+X) - \text{Price}(t)}{\text{Price}(t)}$$

Analogously, the monetary profits from a *net sell* can be calculated using the formula:

$$\text{Profits}_t^{\text{Net Sell}} = \text{Net Sell Value}_t \times \frac{\text{Price}(t) - \text{Price}(t+X)}{\text{Price}(t)}$$

After calculating the monetary profits for each stock traded, we aggregate the data from investor-stock-day level to investor-month level and calculate the aggregate investors' monetary profits across all trading activities in a month. Finally, we take average monetary profits for all investors in a pincode-month.

To examine if new road construction changes the risk taking profile of investors, we use the number of trades in high volatility stocks as a proxy for risk-taking activities. For each stock, we calculate its monthly price volatility and recognize it as a high volatility stock if its price volatility is above the sample median for all stocks in the same month. As a proxy for risk-taking, for each pincode-month we count the total number of trades in high volatility stocks in the previous month.

We use a similar set of specifications as our baseline setting replacing the dependent variable with *Monetary Profits* and *Number of Trades in High Volatility Stocks* in Table 10. Columns 1-4 report the results for monetary profits for different holding periods, and column 5 shows the result for stock trades in high volatility stocks. The coefficients of *Connect* dummy are positive and significant over 1 and 10-day trading windows, insignificant over 25-day trading windows and negative and significant for 140-day trading windows. These results indicate that after the completion of a new road, investors pick stocks that carry short-term profits but generate negative monetary profits over longer term. In addition, investors also start to hold more volatile stocks after the road construction, suggesting greater risk-taking.

6 Conclusion

A long standing finance and growth literature has argued for the case of a well developed stock market with broad participation as one of the channels of ushering economic growth. However, low stock market participation has been a long standing puzzle especially for a fast growing developing country like India with a high savings rate. In this paper, we advance a novel mechanism that affects stock market participation rates - improvement of physical infrastructure. Using a unique data set on stock trading activities of over 13 million households in India between 2004 and 2015, and a national rural road building program, we find that construction of new feeder roads increases stock market participation, especially for new, inexperienced investors in rural areas of intermediate levels of economic development.

We find support for two channels underlying the positive effects of roads on stock marker participation: the information channel and the financial inclusion channel. First, we show that construction of new roads leads to greater stock market participation for more remote pincodes. There is also greater portfolio diversification, and a reduction in home bias (greater trading in stocks that are more distant), all suggesting that infrastructure improvements lead to a reduction in information asymmetry, allowing for greater stock market participation. Second, we show that increased stock market participation following new road construction is particularly high in areas where there is a new bank branch opening, especially new state-owned bank branches, within three years of the road construction. We also have suggestive evidence from household consumption survey data and bank deposits data that there is a decrease in consumption and savings in districts with new road constructions. These findings are consistent with a financial inclusion channel where new bank branches following road construction allow for opening of trading accounts as well as better saving decisions, allowing for greater stock market participation.

We also find that the total stock portfolios of investors in the pincodes earn short-term profits but lose money longer-term after the new road construction. Investors also seem to become more risk-taking after the new road construction.

Overall, our paper has important implications for household welfare through stock market participation. If wider stock ownership is a policy reform agenda for developing countries, our paper suggests that improving physical infrastructure addresses the twin problems of financial access and inadequate information, allowing for better integration of investors in remote and rural areas.

References

Anjali Adukia, Sam Asher, and Paul Novosad. Educational investment responses to economic opportunity: evidence from indian road construction. *American Economic Journal: Applied Economics*, 12(1):348–76, 2020.

Sumit Agarwal, Shashwat Alok, Pulak Ghosh, Soumya Ghosh, Tomasz Piskorski, and Amit Seru. Banking the unbanked: What do 255 million new bank accounts reveal about financial access? *Columbia Business School Research Paper*, (17-12), 2017.

Sumit Agarwal, Abhiroop Mukherjee, and S Lakshmi Naaraayanan. Roads and loans. *Review of Financial Studies forthcoming*, 2021.

Shilpa Aggarwal. Do rural roads create pathways out of poverty? evidence from india. *Journal of Development Economics*, 133:375–395, 2018.

Steffen Andersen and Kasper Meisner Nielsen. Participation constraints in the stock market: Evidence from unexpected inheritance due to sudden death. *The Review of Financial Studies*, 24 (5):1667–1697, 2011.

Sam Asher, Tobias Lunt, Ryu Matsuura, and Paul Novosad. The Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG). World Bank Economic Review (Revise and Resubmit), 2020.

Brad M Barber and Terrance Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1):261–292, 2001.

Brad M Barber, Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean. Just how much do individual investors lose by trading? *The Review of Financial Studies*, 22(2):609–632, 2009.

Suleyman Basak and Domenico Cuoco. An equilibrium model with restricted stock market participation. *The Review of Financial Studies*, 11(2):309–341, 1998.

Patrick J Bayer, B Douglas Bernheim, and John Karl Scholz. The effects of financial education in the workplace: Evidence from a survey of employers. *Economic Inquiry*, 47(4):605–624, 2009.

Thorsten Beck and Ross Levine. Stock markets, banks, and growth: Panel evidence. *Journal of Banking & Finance*, 28(3):423–442, 2004.

Allen N Berger, Leora F Klapper, Maria Soledad Martinez Peria, and Rida Zaidi. Bank ownership type and banking relationships. *Journal of Financial Intermediation*, 17(1):37–62, 2008.

B Douglas Bernheim and Daniel M Garrett. The determinants and consequences of financial education in the workplace: Evidence from a survey of households. Technical report, National Bureau of Economic Research, 1996.

Ata Can Bertay, Asli Demirgüç-Kunt, and Harry Huizinga. Bank ownership and credit over the business cycle: Is lending by state banks less procyclical? *Journal of Banking & Finance*, 50: 326–339, 2015.

Kirill Borusyak and Xavier Jaravel. Revisiting event study designs. *SSRN Working Paper Series*, 2017.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*, 2021.

Joshua Bosshardt and Eugenio Cerutti. Why did public banks lend more during the global financial crisis? *IMF Working paper*, 2020.

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

Michael Brei and Alfredo Schclarek. Public bank lending in times of crisis. *Journal of Financial Stability*, 9(4):820–830, 2013.

James R Brown, Steven M Fazzari, and Bruce C Petersen. Financing innovation and growth: Cash flow, external equity, and the 1990s r&d boom. *The Journal of Finance*, 64(1):151–185, 2009.

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

Markus K Brunnermeier and Stefan Nagel. Do wealth fluctuations generate time-varying risk aversion? micro-evidence on individuals. *American Economic Review*, 98(3):713–36, 2008.

Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

Laurent E Calvet and Paolo Sodini. Twin picks: Disentangling the determinants of risk-taking in household portfolios. *The Journal of Finance*, 69(2):867–906, 2014.

Laurent E Calvet, John Y Campbell, and Paolo Sodini. Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, 115(5):707–747, 2007.

Laurent E Calvet, John Y Campbell, and Paolo Sodini. Fight or flight? portfolio rebalancing by individual investors. *The Quarterly Journal of Economics*, 124(1):301–348, 2009.

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

Feng-Wen Chen and Chen-Wuing Liu. Estimation of the spatial rainfall distribution using inverse distance weighting (idw) in the middle of taiwan. *Paddy and Water Environment*, 10(3):209–222, 2012.

Charlotte Christiansen, Juanna Schröter Joensen, and Jesper Rangvid. Are economists more likely to hold stocks? *Review of Finance*, 12(3):465–496, 2008.

George M Constantinides, John B Donaldson, and Rajnish Mehra. Junior can't borrow: A new perspective on the equity premium puzzle. *The Quarterly Journal of Economics*, 117(1):269–296, 2002.

Robert Cull and Maria Soledad Martinez Peria. Bank ownership and lending patterns during the 2008–2009 financial crisis: Evidence from latin america and eastern europe. *Journal of Banking & Finance*, 37(12):4861–4878, 2013.

Abhiman Das, Syed Ejaz Ghani, Arti Grover Goswami, William Kerr, and Ramana Nanda. Infrastructure and finance: Evidence from india's gq highway network. *Harvard Business School Entrepreneurial Management Working Paper*, (19-121), 2019.

Clément De Chaisemartin and Xavier d'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96, 2020.

Asli Demirgüç-Kunt and Ross Levine. Stock markets, corporate finance, and economic growth: an overview. *The World Bank Economic Review*, 10(2):223–239, 1996.

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

Mariassunta Giannetti and Yrjö Koskinen. Investor protection, equity returns, and financial globalization. *Journal of Financial and Quantitative Analysis*, pages 135–168, 2010.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 2021.

Mark Grinblatt, Matti Keloharju, and Juhani Linnainmaa. Iq and stock market participation. *The Journal of Finance*, 66(6):2121–2164, 2011.

Luigi Guiso and Tullio Jappelli. Awareness and stock market participation. *Review of Finance*, 9 (4):537–567, 2005.

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

Luigi Guiso, Michael Haliassos, and Tullio Jappelli. Household stockholding in europe: where do we stand and where do we go? *Economic Policy*, 18(36):123–170, 2003.

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

Apoorv Gupta. Firm heterogeneity, demand for quality and prices: Evidence from india. *Job Market Paper*, 2020.

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

Kirk Hamilton. Sustaining economic welfare: estimating changes in per capita wealth. *Environment, Development and Sustainability*, 5:419–436, 2003.

John Heaton and Deborah Lucas. Stock prices and fundamentals. *NBER Macroeconomics Annual*, 14:213–242, 1999.

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

Harrison Hong, Jeffrey D Kubik, and Jeremy C Stein. Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *The Journal of Finance*, 60(6):2801–2824, 2005.

Jennifer Huang and Jiang Wang. Liquidity and market crashes. *The Review of Financial Studies*, 22(7):2607–2643, 2009.

Zoran Ivković and Scott Weisbenner. Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance*, 60(1):267–306, 2005.

Seema Jayachandran. Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of Political Economy*, 114(3):538–575, 2006.

Supreet Kaur. Nominal wage rigidity in village labor markets. *American Economic Review*, 109 (10):3585–3616, 2019.

Ross Levine. Finance and growth: theory and evidence. *Handbook of economic growth*, 1:865–934, 2005.

Ross Levine and Sara Zervos. Stock markets, banks, and economic growth. *American economic review*, pages 537–558, 1998.

Brigitte C Madrian and Dennis F Shea. The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly Journal of Economics*, 116(4):1149–1187, 2001.

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

Franco Modigliani and Shi Larry Cao. The chinese saving puzzle and the life-cycle hypothesis. *Journal of Economic Literature*, 42(1):145–170, 2004.

Mukta Mukherjee. Do better roads increase school enrollment? evidence from a unique road policy in india. *SSRN Working Paper Series*, 2012.

NCAER. How households save and invest: Evidence from ncaer household survey. *National Council of Applied Economic Research Main Report*, 2011.

Mohamad J Noori, Hussein H Hassan, and Yaseen T Mustafa. Spatial estimation of rainfall distribution and its classification in duhok governorate using gis. *Journal of Water Resource and Protection*, 2014, 2014.

Anita K Pennathur, Vijaya Subrahmanyam, and Sharmila Vishwasrao. Income diversification and risk: Does ownership matter? an empirical examination of indian banks. *Journal of Banking & Finance*, 36(8):2203–2215, 2012.

Martin Ravallion. Growth and poverty: Evidence for developing countries in the 1980s. *Economics letters*, 48(3-4):411–417, 1995.

Yogita Shamdasani. Rural road infrastructure & agricultural production: Evidence from india. *Journal of Development Economics*, 2021.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.

Xuan Tian. The causes and consequences of venture capital stage financing. *Journal of Financial Economics*, 101(1):132–159, 2011.

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

Annette Vissing-Jørgensen. Stock prices and fundamentals: Comment. *NBER Macroeconomics Annual*, 14:242–253, 1999.

Annette Vissing-Jørgensen. Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures. Technical report, National Bureau of Economic Research, 2002.

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

Figure 1: **Geographic Distribution of Pincodes and Indian Public Listed Companies**
This map plot the geographic distribution of investors in our NSE sample and all public Indian companies from Prowess. The blue dots represents the pincode for investors in the NSE data, while black dots represents the location of Indian public companies.

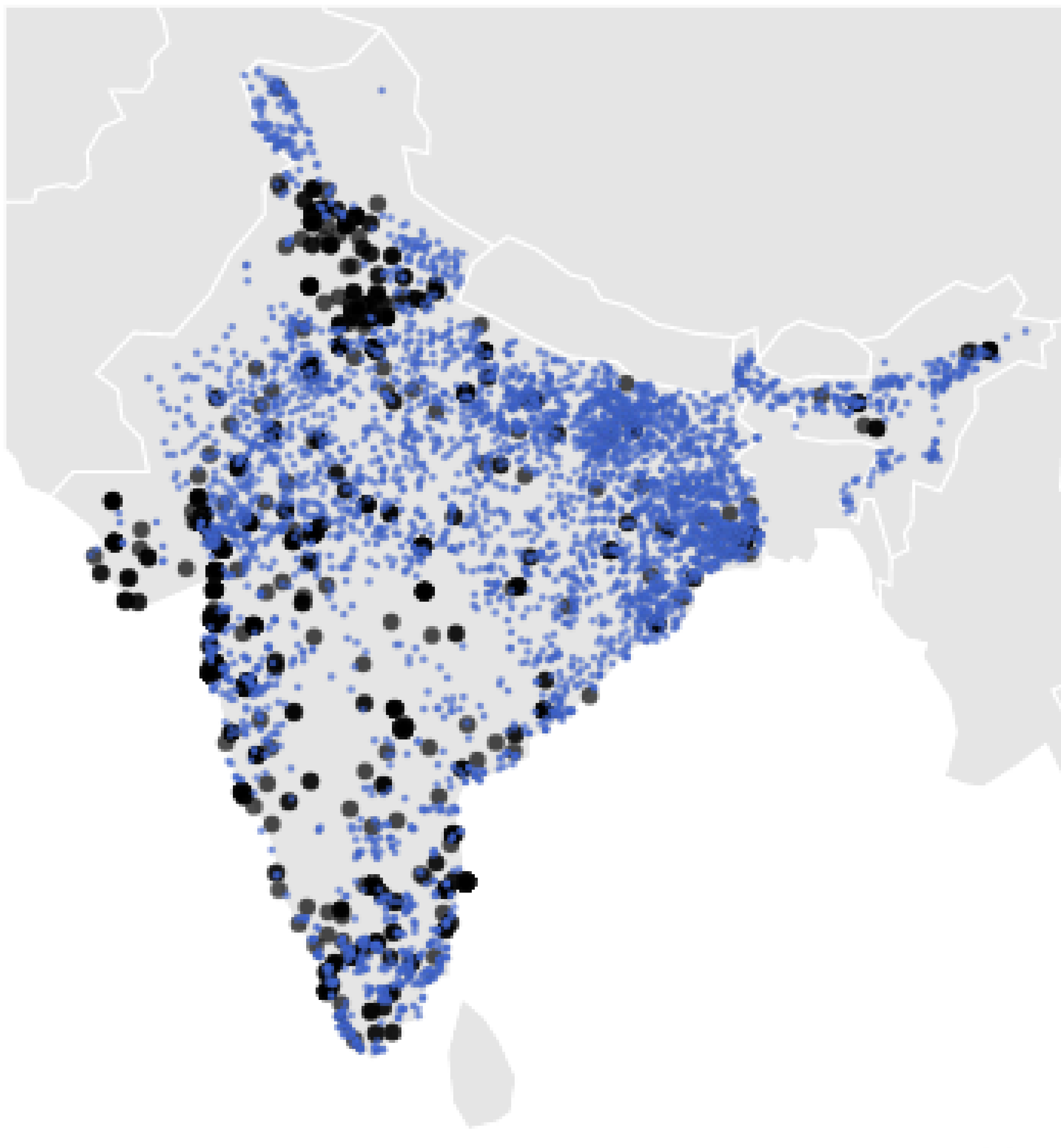

Figure 2: **Geographic Distribution of Villages and Pincodes in our Sample**
This map plot the geographic distribution of investors in our NSE sample and all villages in the PYGSY data. The blue dots represents the pincode for investors in the NSE data, while red dots represents the village within 5 Kms radius for each pincode.

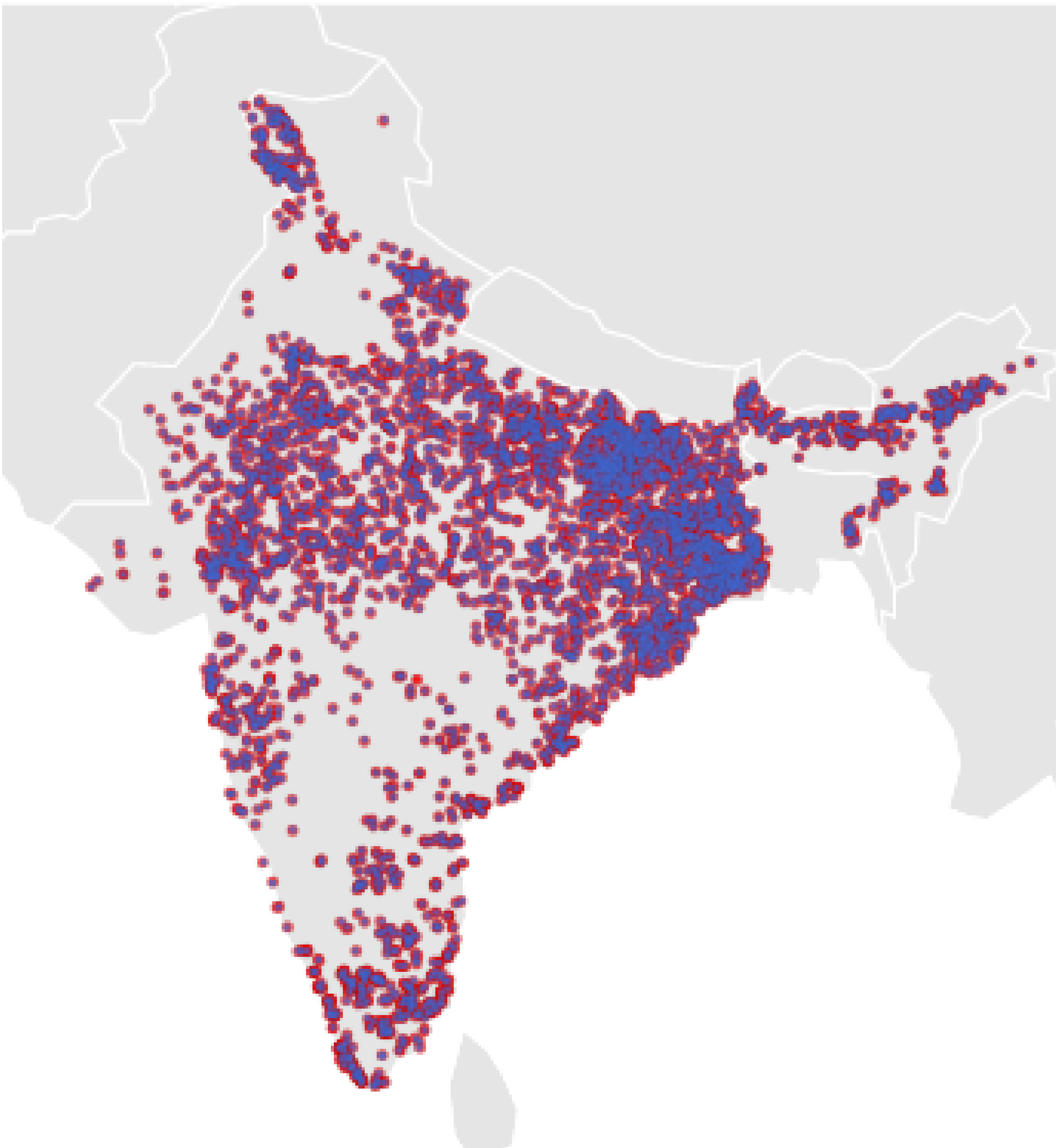

Figure 3: **Time Trends of Road Impacts on Stock Tradings**

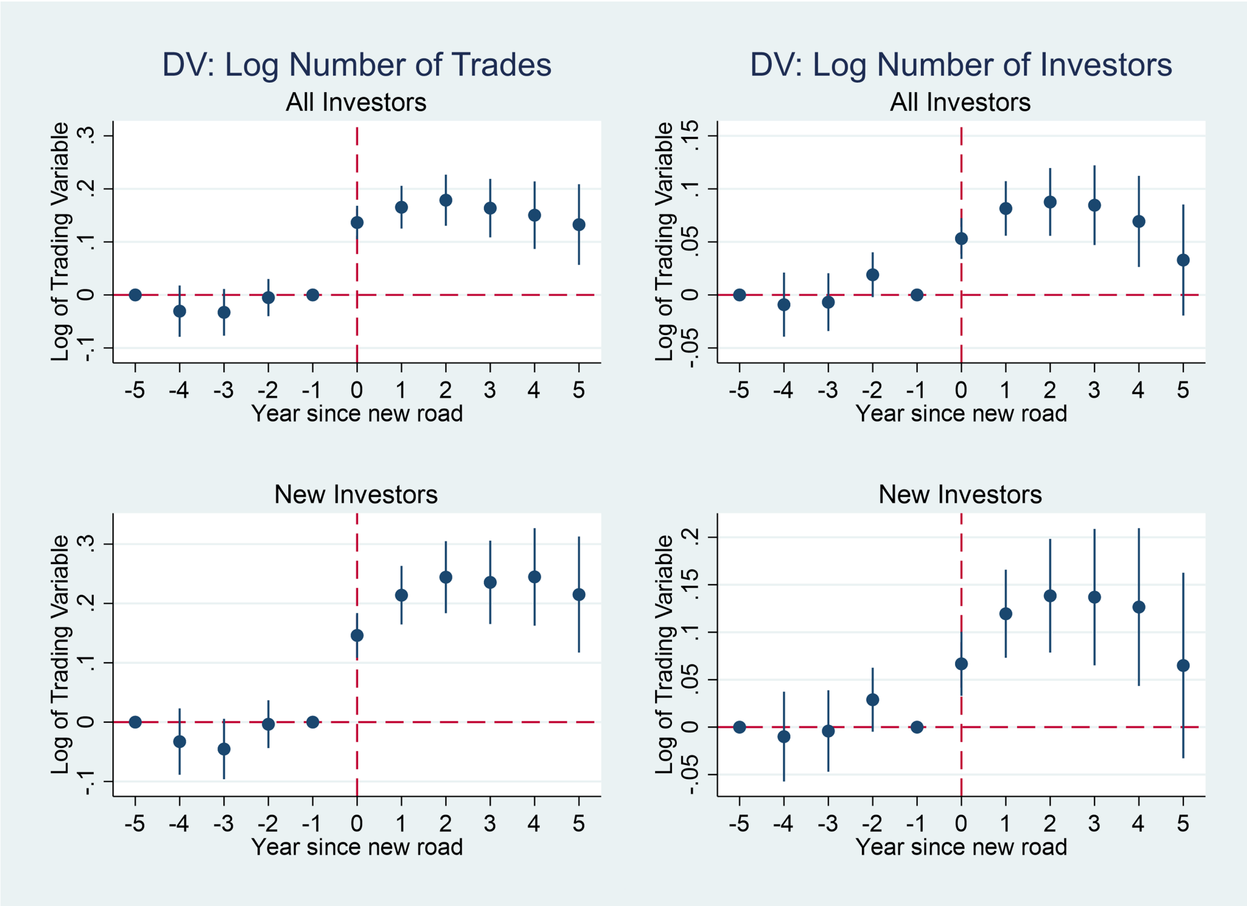

Figure 4: **Time Trends of Road Impacts on Stock Tradings: By Regional Economic Development**

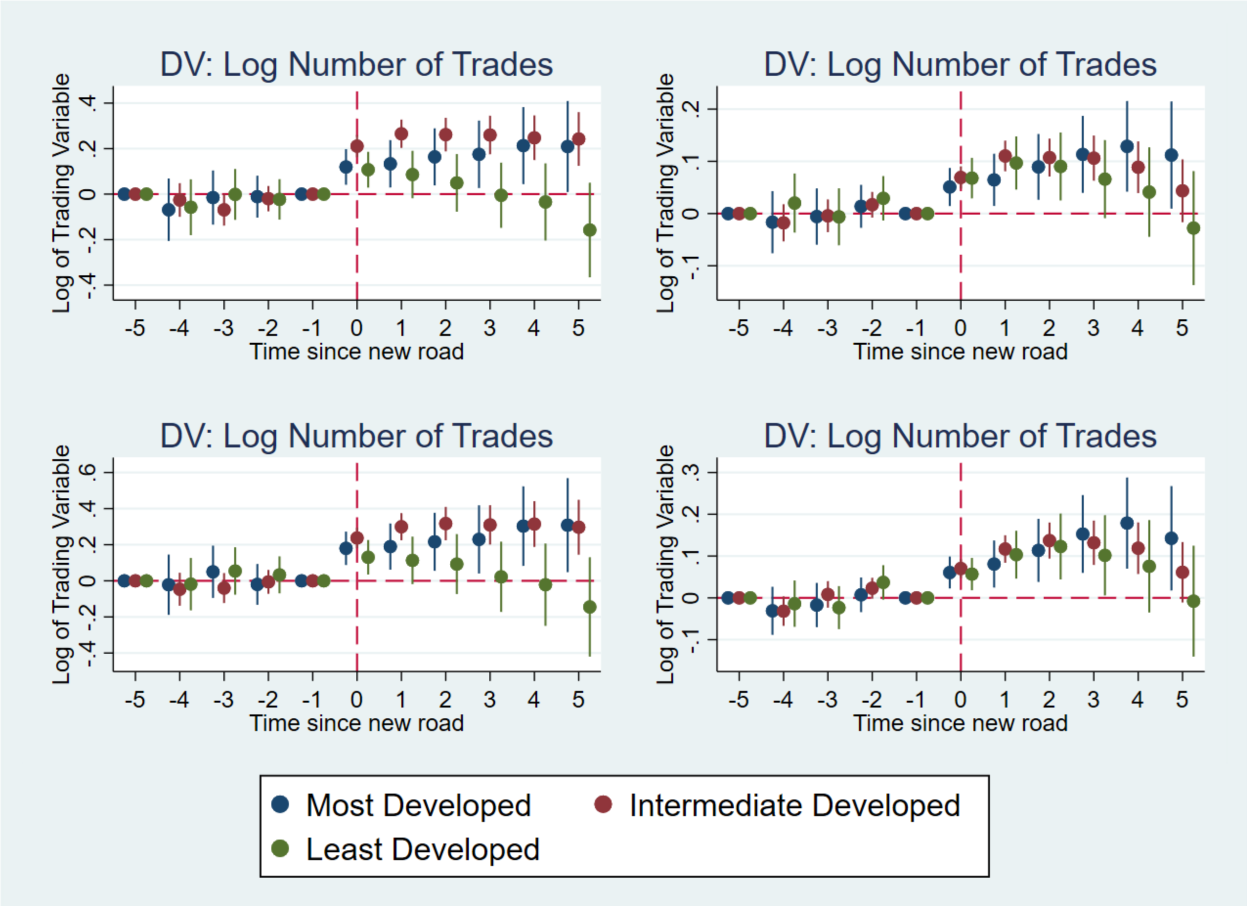

Figure 5: **Monthly Average of Stock Tradings**

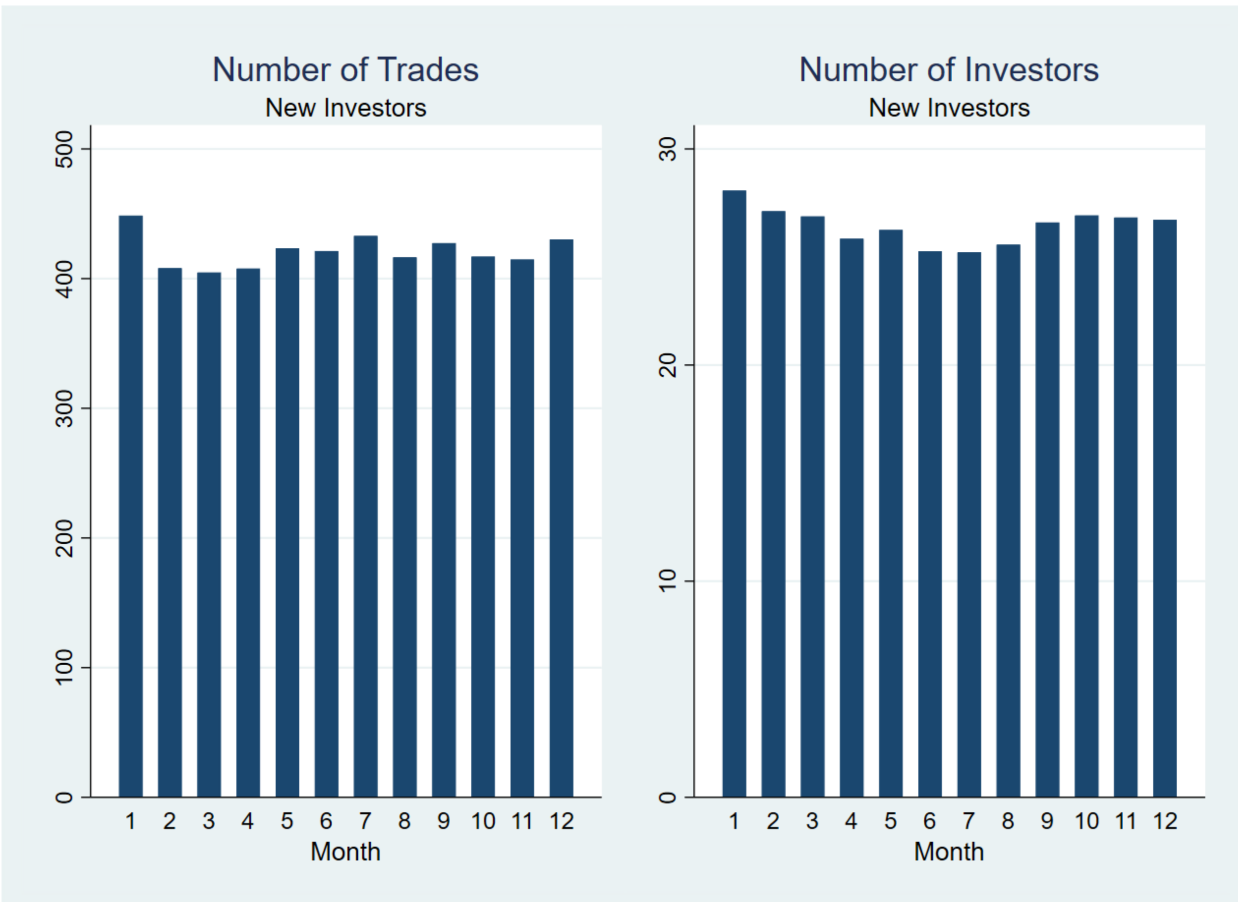

Table 1: **Summary Statistics**

This table reports the summary statistics for the key variables in our analysis. All variable definitions are in the Variable Appendix.

	N	**Mean**	**S.D.**	**Min**	**P25**	**P50**	**P75**	**Max**
Pincode-month descriptors								
Number of Trades	592195	601.657	1637.826	0	5	66	336	14831
Number of Investors	591371	35.873	94.562	0	1	5	21	917
Number of Tickers	590504	78.809	119.913	0	4	28	96	685
Share of New Investors	592195	0.369	0.316	0	0.045	0.333	0.615	1
Share of Experienced Investors	592195	0.246	0.243	0	0.071	0.200	0.333	1
Share of Young Investors	592195	0.533	0.266	0	0.402	0.532	0.667	1
Share of Middle Investors'	592195	0.221	0.221	0	0.000	0.200	0.333	1
Share of Mature Investors	592195	0.912	0.135	0	0.858	0.957	1.000	1
Share of Male Investors	592195	0.031	0.078	0	0.000	0.000	0.038	1
Share of Female Investors	592195	0.025	0.072	0	0	0	0.024	1
Connect	592195	0.553	0.497	0	0	1	1	1
Rainfall Shock	592195	0.024	0.500	-1	0	0	0	1
Pincode descriptors								
Distance to city with population > 100K	4154	45.717	30.392	0.089	22.970	39.190	62.087	275.879
Distance to city with population > 1M	4154	140.493	129.523	0.672	66.658	109.403	161.860	1025.957
New Branch Dummy	4154	0.432	0.495	0	0	0	1	1
Number of New Branches	4154	1.119	2.106	0	0	0	1	30
Rural	4154	0.793	0.405	0	1	1	1	1
Consumption Per Capita	4154	16478.11	5267.186	0	13506.41	15539.7	18765.04	66378.92

Table 2: **Roads and Stock Market Participation: New vs. Experienced Investors**
This table reports estimates from the following regression:

$$Log(Y_{i,t}) = \beta \cdot Connect_{i,t} + \xi_i + \kappa_t + \varepsilon_{i,t}$$

where Y are the different measures of stock market participation in each pincode i in year-month t: *Number of Trades* and *Number of Investors*. $Connect_{i,t}$ is an indicator variable which equals 1 in the year-month (and thereafter) when a pincode is connected by a paved road under the PMGSY program and 0 otherwise. Cols. 1 and 2 report results for the full sample, cols. 3 and 4 report results for new investors and cols. 5 and 6 report results for experienced investors. New Investors are defined as investors whose trading account opening date is $\leq$ 3 years old and Experienced Investors are those whose trading account opening date is >3 years. All regressions are estimated using pincode and year-month fixed effects. Standard errors clustered by pincode are reported in parentheses. All variables are defined in the Variable Appendix. (***); (**); (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

	Number of Trades	Number of Investors	Number of Trades	Number of Investors	Number of Trades	Number of Investors
	All Investors		New Investors		Experienced Investors	
Connect	0.557***	0.182***	0.645***	0.287***	0.396***	0.154***
	(0.022)	(0.010)	(0.021)	(0.010)	(0.020)	(0.010)
Pincode FE	Y	Y	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y	Y	Y
N	592195	592204	592195	592204	592195	592204
Adj. R-sq	0.863	0.939	0.802	0.899	0.832	0.904

Table 3: **Investor Heterogeneity by Gender and Age**
This table reports estimates from the following regression:

$$Log(Y_{i,t}) = \beta \cdot Connect_{i,t} + \xi_i + \kappa_t + \varepsilon_{i,t}$$

where Y are different measures of stock market participation defined in each pincode i in year-month t: *Number of Trades* and *Number of Investors*. $Connect_{i,t}$ is an indicator variable which equals 1 in the year-month (and thereafter) when a pincode is connected by a paved road under the PMGSY program and 0 otherwise. All estimates are for sample of New Investors, defined as investors whose trading account opening date is ≤ 3 years old. Panel A reports results for male and female investors, panel B reports results for young (18-30 years), mid-age (30-55 years) and mature investors (55+ years). All regressions are estimated using pincode and year-month fixed effects. Standard errors clustered by pincode are reported in parentheses. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

Panel A: Heterogeneity by gender

	1	2	3	4
	Number of Trades	Number of Investors	Number of Trades	Number of Investors
	Male Investors (New)		Female Investors (New)	
Connect	0.622***	0.275***	0.068***	0.025***
	(0.021)	(0.010)	(0.016)	(0.007)
Pincode FE	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y
N	592195	592204	592195	592204
Adj. R-sq	0.770	0.876	0.610	0.693

Table 3: **Investor Heterogeneity by Gender and Age (Continued...)**

Panel B: Heterogeneity by age-group

	1	2	4	5	7	8
	Number of Trades	Number of Investors	Number of Trades	Number of Investors	Number of Trades	Number of Investors
	Young Investors (New)		Mid-Age Investors (New)		Mature Investors (New)	
Connect	0.462***	0.149***	0.672***	0.152***	0.224***	0.097***
	(0.019)	(0.008)	(0.020)	(0.009)	(0.019)	(0.009)
Pincode FE	Y	Y	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y	Y	Y
N	592193	592202	592193	592202	592193	592202
Adj. R-sq	0.718	0.834	0.773	0.876	0.754	0.838

Table 4: **Roads and Stock Market Participation: Local Economic Development**
This table estimates the following regression:

$$Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} + \beta \cdot Connect_{i,t} \times Rural_i + \beta_1 \cdot Connect_{i,t} \times Intermediate\ development_i + \beta_2 \cdot Connect_{i,t} \times Least\ developed_i + \xi_i + \kappa_t + \varepsilon_{i,t}$$

where Y are different measures of stock market participation defined in each pincode i in year-month t: *Number of Trades* and *Number of Investors*. All estimates are for sample of New Investors, defined as investors whose trading account opening date is ≤ 3 years old. $Connect_{i,t}$ is an indicator variable which equals 1 in the year-month (and thereafter) when a pincode is connected by a paved road under the PMGSY program and 0 otherwise. In cols. 1 and 3, we use $Rural_i$ as a measure of development where $Rural_i$ is an indicator variable which equals to 1 if the pincode is in a rural area (as identified by the post office) and 0 otherwise. In cols. 2 and 4, we classify pincodes into $Developed(top50\%)$; $IntermediateDevelopment(10-50\%)$ and $Least\ developed(bottom10\%)$ based on the values of *Consumption per Capita*. All regressions are estimated using pincode and year-month fixed effects. Standard errors clustered by pincode are reported in parentheses. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

	1	2	3	4
	Number of Trades (New Investors)		Number of Investors (New Investors)	
Connect	0.060	0.562***	0.058***	0.198***
	(0.041)	(0.028)	(0.022)	(0.013)
Connect X Rural	0.715***		0.281***	
	(0.044)		(0.023)	
Connect X Intermediate Development		0.146***		0.126***
		(0.037)		(0.017)
Connect X Least Development		0.024		-0.005
		(0.063)		(0.030)
Pincode FE	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y
N	592195	592195	592204	592204
Adj. R-sq	0.804	0.802	0.900	0.899

Table 5: **Information Channel**
This table reports results from the following regression:

$$Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} + \beta \cdot Connect_{i,t} \times Distance_i + \xi_i + \kappa_t + \varepsilon_{i,t}$$

In Panel A, Y are different measures of stock market participation defined in each pincode i in year-month t: *Number of Trades* and *Number of Investors*. All estimates are for sample of New Investors, defined as investors whose trading account opening date is ≤ 3 years old. *Distance* is the distance between pincode i and the nearest city. In Panel B, in col. 1 Y is *Number of Tickers* traded by new investors in each pincode i in year-month t, and in cols. 2-5 Y is the *Number of Trades* in year-month t on companies that are located at the following distances from pincode i: less than 50km, 50-100kms, 100-500km, and over 500km away. $Connect_{i,t}$ is an indicator variable which equals 1 in the year-month (and thereafter) when a pincode is connected by a paved road under the PMGSY program and 0 otherwise. All regressions are estimated using pincode and year-month fixed effects. Standard errors clustered by pincode are reported in parentheses. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

Panel A: Distance from nearest city

	1	2	3	4
Trading Variable	Number of Trades (New Investors)		Number of Investors (New Investors)	
City Type	Tier 1	Metro	Tier 1	Metro
Connect	0.468***	0.515***	0.234***	0.225***
	(0.036)	(0.031)	(0.017)	(0.015)
Connect X Distance	0.037***	0.009***	0.011***	0.004***
	(0.006)	(0.002)	(0.003)	(0.001)
Pincode FE	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y
N	592195	592195	592204	592204
Adj. R-sq	0.802	0.802	0.899	0.899

Table 5: **Information Channel (Continued...)**

Panel B: Portfolio Diversification

	1	2	3	4	5
	Number of Unique Tickers	Number of Trades on Distant Companies			
		< 50 Kms	50-100 Kms	100-500 Kms	> 500 Kms
Connect	0.511***	0.002	0.019***	0.212***	0.606***
	(0.017)	(0.004)	(0.005)	(0.013)	(0.021)
Pincode FE	Y	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y	Y
N	592218	592195	592195	592195	592195
Adj. R-sq	0.794	0.852	0.818	0.809	0.801

Table 6: **Financial Inclusion Channel**
This table reports estimates from the following regression:

$$Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} + \beta \cdot Connect_{i,t} \times NewBranchDummy_i(orNumberofNewBranches_i) + \xi_i + \kappa_t + \varepsilon_{i,t}$$

where Y are different measures of stock market participation defined in each pincode i in year-month t: *Number of Trades* and *Number of Investors*. All estimates are presented for sample of New Investors, defined as investors whose trading account opening date is ≤ 3 years old. $Connect_{i,t}$ is an indicator variable which equals 1 in the year-month (and thereafter) when a pincode is connected by a paved road under the PMGSY program and 0 otherwise. *New Branch dummy* is an indicator variable that takes the value 1 when a bank branch is opened within a 3-year window after the PMGSY road completion in pincode i and 0 otherwise. *Number of New Branches* is the total number of bank branches opened within a 3-year window after the PMGSY road completion in pincode i. All regressions are estimated using pincode and year-month fixed effects. Standard errors clustered by pincode are reported in parentheses. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

Panel A: All Banks

	1	2	3	4
	Number of Trades (New Investors)		Number of Investors (New Investors)	
Connect	0.584***	0.599***	0.250***	0.274***
	(0.026)	(0.023)	(0.012)	(0.011)
Connect X New Branch Dummy	0.096***		0.089***	
	(0.034)		(0.016)	
Connect X Number of New Branches		0.053***		0.013***
		(0.007)		(0.004)
Pincode FE	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y
N	592195	592195	592204	592204
Adj. R-sq	0.802	0.802	0.900	0.899

Table 6: **Financial Inclusion Channel (Continued...)**

Panel B: State-owned Banks vs Private Banks

	1	2	3	4	5	6	7	8
	Number of Trades (New Investors)		Number of Investors (New Investors)		Number of Trades (New Investors)		Number of Investors (New Investors)	
Reference Group	Pincodes with either no branch opening or branch openings of private banks				Pincodes with branch openings of private banks			
Connect	0.482***	0.592***	0.263***	0.277***	0.482***	0.511***	0.235***	0.252***
	(0.024)	(0.022)	(0.011)	(0.010)	(0.049)	(0.035)	(0.022)	(0.018)
Connect X New Branch Dummy	0.132***		0.084***		0.177***		0.004	
	(0.036)		(0.017)		(0.054)		(0.024)	
Connect X Number of New Branches		0.088***		0.019***		0.117***		-0.011
		(0.012)		(0.007)		(0.013)		(0.007)
Pincode FE	Y	Y	Y	Y	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y	Y	Y	Y	Y
N	592195	592195	592204	592204	252358	252358	252302	252302
Adj. R-sq	0.802	0.802	0.899	0.899	0.854	0.854	0.931	0.931

Table 7: **Household Consumption in pincodes connected by PMGSY Roads**
This table reports estimates from the following regression:

$$Y_{h,t} = \alpha \cdot Connect_{d,t}^{CPHS} + \beta \cdot HighIncome_{h,t} + \gamma \cdot Connect_{d,t}^{CPHS} \times HighIncome_{h,t} + \eta_h + \kappa_t + \varepsilon_{h,t}$$

In panel A, Y is either Log(Consumption) or Log(Income) for each household h in year-month t. $Connect_{d,t}^{CPHS}$ is an indicator variable which equals 1 if a district, d, received a new PMGSY road between 2014 and 2015 and had not received any road under PMGSY during the previous 5 years. It takes the value 0 if the district d received no new roads under PMGSY in 2014 and 2015. $High\ Income_{h,t}$ is an indicator variable that takes the value 1 if household i's income is above the following cutoffs and 0 otherwise: 50^{th}, 75^{th}, and 90^{th} percentile of household income in that month. Standard errors clustered by household are reported in parentheses. All regressions are estimated using household and year-month fixed effects. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

Panel A: Household Expenditure & Income

	1	2
	Log(Consumption)	Log(Income)
$Connect^{CPHS}$	-0.051***	0.115***
	(0.051)	(0.017)
Household FE	Y	Y
Month FE	Y	Y
N	125376	125376
Adj. R-sq	0.656	0.358

Table 7: **Household Consumption in pincodes connected by PMGSY Roads (Continued...)**

Panel B: Household Expenditure & Income across Income Group

	1	2	3	4	5	6
	Log(Consumption)			Log(Income)		
Income Level	Above 50th	Above 75th	Above 90th	Above 50th	Above 75th	Above 90th
$Connect^{CPHS}$	-0.029***	-0.047***	-0.054***	0.093***	0.096***	0.095***
	(0.007)	(0.006)	(0.006)	(0.020)	(0.018)	(0.018)
High Income	0.118***	0.109***	0.191***	0.242***	0.403***	0.515***
	(0.008)	(0.011)	(0.018)	(0.023)	(0.025)	(0.032)
$Connect^{CPHS}$ X High Income	-0.034***	-0.020**	-0.024*	0.054***	0.038**	0.026
	(0.007)	(0.009)	(0.013)	(0.017)	(0.016)	(0.018)
Household FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
N	125376	125376	125376	125376	125376	125376
Adj. R-sq	0.658	0.657	0.659	0.362	0.363	0.362

Table 8: **Bank Deposits in pincodes connected by PMGSY Roads**
This table reports estimates from the following regression:

$$Deposit_{b,t} = \beta \cdot Connect_{i,t} + \phi_b + \kappa_t + \varepsilon_{b,t}$$

where *Deposit* are the log values of total amount of deposit in *Current/Saving/Term* account, and total value of deposit in branch b in year t. $Connect_{b,t}$ is an indicator variable which equals 1 in the year (and thereafter) when the pincode i, where the branch located, in is connected by a paved road under the PMGSY program and 0 otherwise. All regressions are estimated using branch and year fixed effects. Standard errors clustered by pincode are reported in parentheses. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

	1	2	3	4
	Current Account	Saving Account	Term Account	Total Deposit
Connect	-0.035	-0.031**	-0.018	-0.041***
	(0.024)	(0.014)	(0.019)	(0.014)
Branch FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
N	157476	157483	157479	157485
Adj. R-sq	0.645	0.834	0.807	0.849

Table 9: **Roads and Local Income Shocks**
This table reports estimates from the following regression:

$$Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} + \beta \cdot Connect_{i,t} \times RainfallShock_{i,t} + \xi_i + \kappa_t + \varepsilon_{i,t}$$

where Y are different measures of stock market participation defined in each pincode i in year-month t: *Number of Trades* and *Number of Investors*. All estimates are presented for sample of New Investors, defined as investors whose trading account opening date is ≤ 3 years old. $Connect_{i,t}$ is an indicator variable which equals 1 in the year-month (and thereafter) when a pincode is connected by a paved road under the PMGSY program and 0 otherwise. $Rainfall\ Shock_{i,t}$ takes the value 1 if rainfall is above 80^{th} percentile for the same pincode-month over the entire sample period; -1 if rainfall is below 20^{th} percentile for the same pincode-month over the entire sample period; and 0 otherwise. All regressions are estimated using pincode and year-month fixed effects. Standard errors clustered by pincode are reported in parentheses. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

	1	2	3	4
	Number of Trades (New Investors)		Number of Investors (New Investors)	
Timing of rainfall	Concurrent	One month-lag	Concurrent	One month-lag
Connect	0.645***	0.641***	0.287***	0.284***
	(0.021)	(0.021)	(0.010)	(0.010)
Rainfall Shock	-0.003	-0.009**	-0.000	-0.000
	(0.004)	(0.004)	(0.002)	(0.002)
Connect X Rainfall Shock	-0.003	0.009	-0.004	-0.001
	(0.006)	(0.006)	(0.002)	(0.002)
Pincode FE	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y
N	592195	588044	592204	588051
Adj. R-sq	0.802	0.802	0.899	0.900

Table 10: **Monetary Profits & Risk Taking**
This table reports estimates from the following regression:

$$Y_{i,t} = \beta \cdot Connect_{i,t} + \xi_i + \kappa_t + \varepsilon_{i,t}$$

where Y are Buy-and-hold profits in Indian Rupees for buy-and-hold periods of 1/10/25/140 trading days in cols 1-4 and the monthly *Number of Trades on High Volatility Stocks* in column 5, where a stock is recognized as high volatility if its price volatility is above the sample median for all stocks in the same month. $Connect_{i,t}$ is an indicator variable which equals 1 in the year-month (and thereafter) when a pincode is connected by a paved road under the PMGSY program and 0 otherwise. Standard errors clustered by household are reported in parentheses. All regressions are estimated using household and year-month fixed effects. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

	1	2	3	4	5
	Average Profits (in Indian Rupees) per investor-month				Number of Trades on High Volatility Stocks
Holding Period	1 trading day	10 trading days	25 trading days	140 trading days	
Connect	72.341***	135.868***	13.538	-1644.754***	0.398***
	(9.563)	(29.715)	(55.995)	(186.261)	(0.012)
Pincode FE	Y	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y	Y
N	586214	586214	586214	586214	592197
Adj. R-sq	0.171	0.296	0.321	0.319	0.572

Appendix

Figure A1: **Time Trends of Road Impacts on Stock Tradings**
All Type of Investors

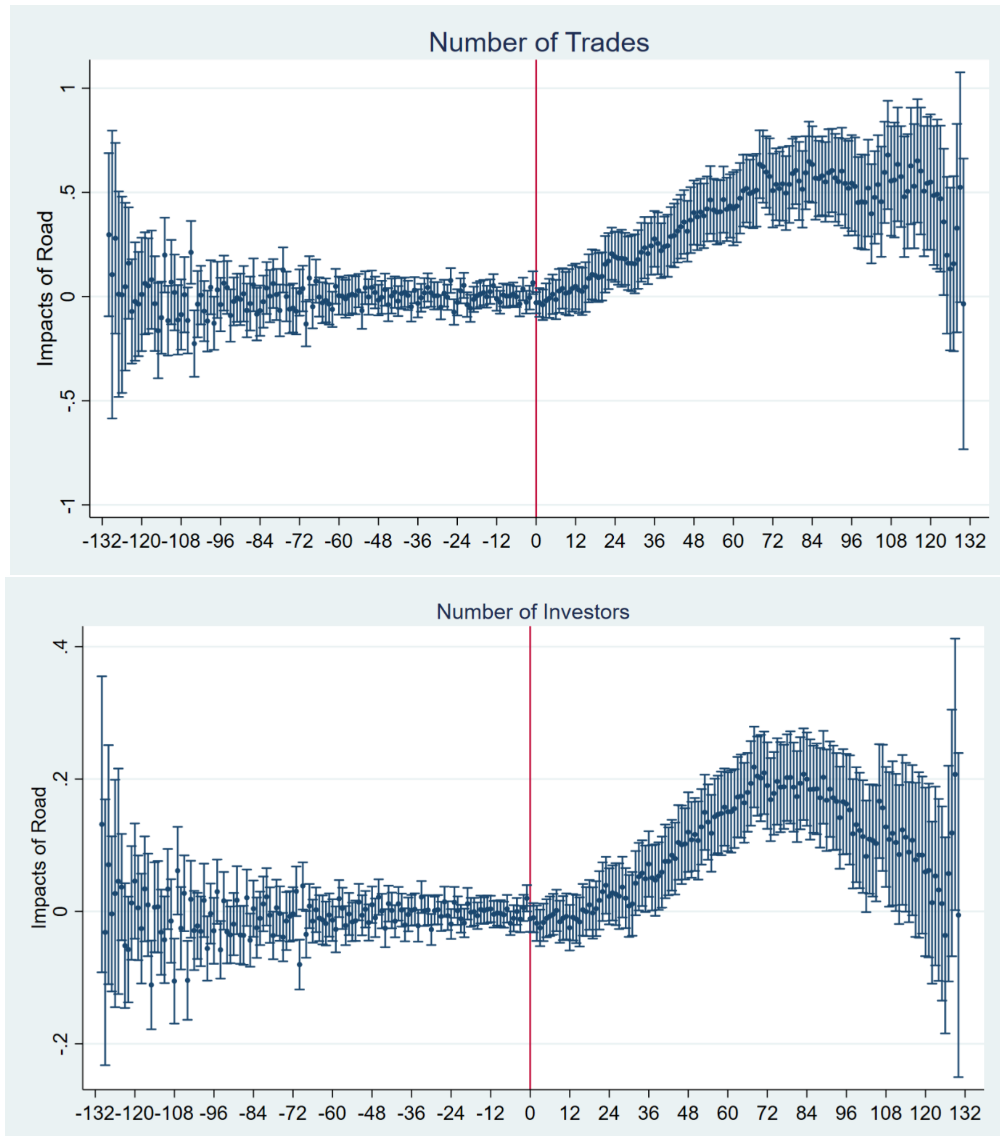

Figure A1: **Time Trends of Road Impacts on Stock Tradings (Continue)**
New Investors

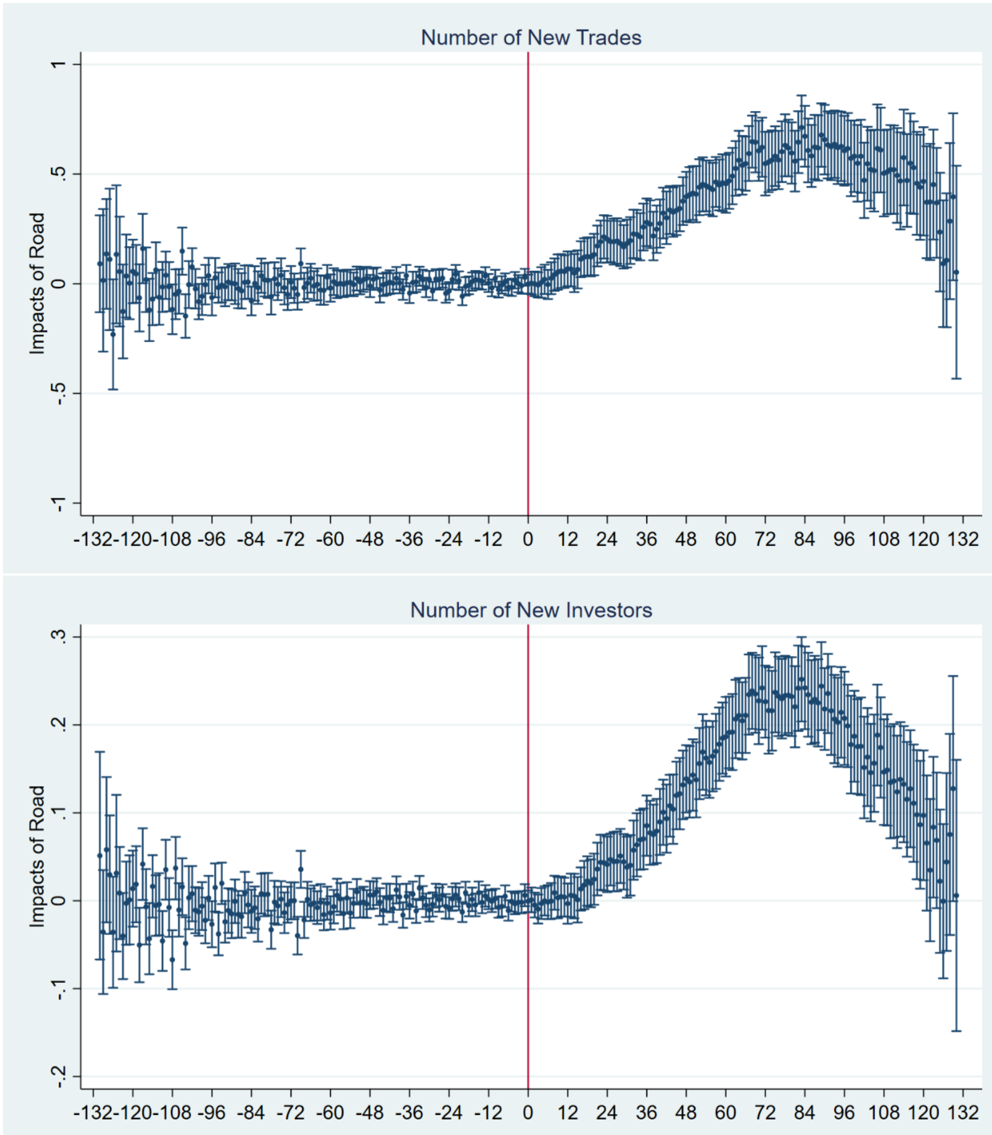

Figure A2: **Time Trends of Road Impacts on Stock Tradings: By Regional Economic Development**

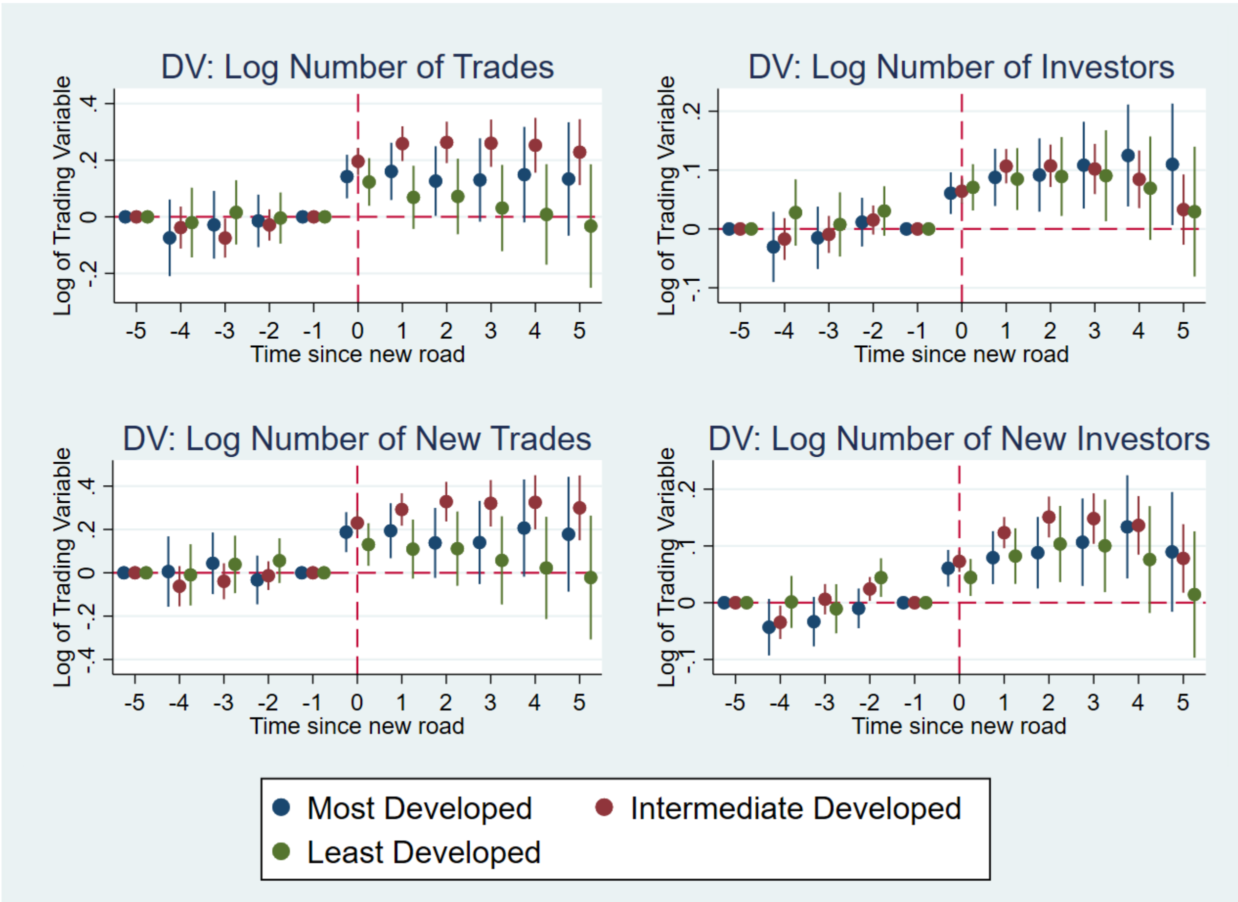

Table A1: **Variable Definition**
This table reports definition of each variable used in this paper.

Variable	Definition Source	Source
Number of Trades	Total number of stock trading activities at pincode i in month t.	NSE
Number of Traders	Total number of stock traders activities at pincode i in month t.	NSE
Number of Tickers	Total number of unique tickers being traded at pincode i in month t.	NSE
High Volatility Stock	Stocks whose monthly price volatility is above the sample median for all stocks in the same month	NSE
New Investor	Investors whose overall trading experiences are less than or equal to 3 years	NSE
Experienced Investor	Investors who have had a trading account for more than 3 years i	NSE
Young Investor	Trader whose age is between 18-30	NSE
Middle-Age Investor	Trade whose age is between 30-55	NSE
Mature Investor	Trade whose age is beyond 55	NSE
Connect	Indicator variable which equals to 1 if the pincode i is connected by a paved road at month t.	PMGSY
Rural	Indicator variable which equals to 1 if the post office type of that pincode if BO.	Indian Post Office
Developed	Indicator variable which equals to 1 if the consumption per capita(poverty rate) value of the pincode is above median value across all the pincodes in the sample.	SHRUG
Intermediate Developed	Indicator variable which equals to 1 if the consumption per capita(poverty rate) value of the pincode fall between the 50th and 10th percentile value across all the pincodes in the sample.	SHRUG
Least Developed	Indicator variable which equals to 1 if the consumption per capita(poverty rate) of the pincode is below the 10th percentile value across all the pincodes in the sample.	Census
New Bank	Indicator variable which equals to 1 if any new bank branches/offices open within a 3-year window after the completion of road at pincode i.	RBI
Number of New Banks	Total number of new bank branches/offices that open within a 3-year window after the completion of road at pincode i.	RBI
Distance	Distance between pincode and the nearest cities.	Census
Consumption	Monthly total consumption of the household.	CMIE
Income	Monthly total income of the household.	CMIE
Deposit	Year amount of deposit in bank branch.	RBI-BSR

Table A2: **State-level distribution of investors**
This table reports the number and percentage of investors in each state who trade on the National Stock Exchange of India over the period 2004 to 2015.

States	**Number of Investors**	**%**	**States**	**Number of Investors**	**%**
Maharashtra	2685160	19.875%	Assam	82744	0.612%
Gujarat	1954062	14.463%	Uttarakhand	78702	0.583%
Tamil Nadu	1046034	7.742%	Jammu and Kashmir	43869	0.325%
West Bengal	995263	7.367%	Himachal Pradesh	42182	0.312%
Karnataka	907781	6.719%	Goa	38684	0.286%
Uttar Pradesh	893399	6.613%	Pondicherry	10734	0.079%
Delhi	835927	6.187%	Tripura	10383	0.077%
Rajasthan	551151	4.079%	Dadra and Nagar Hav.	5176	0.038%
Telangana	532050	3.938%	Megalaya	4729	0.035%
Kerala	490173	3.628%	Chandigarh	4574	0.034%
Andhra Pradesh	432272	3.200%	Daman and Diu	3619	0.027%
Haryana	406657	3.010%	Manipur	2740	0.020%
Madhya Pradesh	398663	2.951%	Sikkim	2674	0.020%
Punjab	328561	2.432%	Nagaland	1998	0.015%
Bihar	228644	1.692%	Andaman and Nico.In.	1526	0.011%
Jharkhand	200192	1.482%	Arunachal Pradesh	1355	0.010%
Odisha	191229	1.415%	Mizoram	604	0.004%
Chattisgarh	96898	0.717%	Lakshadweep	64	0.000%
Total			**13510473**		

Table A3: **Placebo Test: Roads and Stock Market Participation**
This table reports the placebo test of the baseline model

Panel A: Randomization for Entire Sample

	Number of Trades	Number of Traders	Number of Trades	Number of Traders	Number of Trades	Number of Traders
	All Investor		New Investor		Experienced Investor	
Connect	0.000102	-0.000004	0.000296	0.000029	-0.000043	0.000019
	(0.0027857)	(0.001339)	(0.003719)	(0.001549)	(0.003441)	(0.001518)
Pincode FE	Y	Y	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y	Y	Y

Table A3: **Placebo Test: Roads and Stock Market Participation (Continued...)**

Panel B: Randomization within Each Month

	Number of Trades	Number of Traders	Number of Trades	Number of Traders	Number of Trades	Number of Traders
	All Investor		New Investor		Experienced Investor	
Connect	0.000042	-0.000066	0.000221	0.000012	0.000032	0.000062
	(0.002727)	(0.001344)	(0.003694)	(0.001615)	(0.003306)	(0.001405)
Pincode FE	Y	Y	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y	Y	Y

Table A4: **Roads and Stock Market Participation: Robustness using Callaway and Sant'Anna [2021] specification**

This table reports results from the staggered DiD setting using the approach developed by Callaway and Sant'Anna [2021] and estimated using the Stata package csdid. All regressions are estimated using pincode and year-month fixed effects. Standard errors clustered by pincode are reported in parentheses. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

	1 Number of Trades	2 Number of Investors	3 Number of Trades	4 Number of Investors
	All Investors		New Investors	
Connect	0.481***	0.154***	0.538***	0.214***
	(0.103)	(0.045)	(0.152)	(0.012)
Pincode FE	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y

Table A5: **Roads and Stock Market Participation: Local Economic Development**
This table estimates the following regression:

$$Log(Y_{i,t}) = \alpha \cdot Connect_{i,t} + \beta \cdot Connect_{i,t} \times Rural_i + \beta_1 \cdot Connect_{i,t} \times Intermediate\ development_i + \beta_2 \cdot Connect_{i,t} \times Least\ developed_i + \xi_i + \kappa_t + \varepsilon_{i,t}$$

where Y are different measures of stock market participation defined in each pincode i in year-month t: *Number of Trades* and *Number of Investors*. All estimates are presented for sample of New Investors, defined as investors whose trading account opening date is ≤ 3 years old. $Connect_{i,t}$ is an indicator variable which equals to 1 in the year-month (and thereafter) when a pincode is connected by a paved road under the PMGSY program and 0 otherwise. In cols. 1 and 3, we use $Rural_i$ as a measure of development where $Rural_i$ is an indicator variable which equals to 1 if the pincode ranked as the tier-3 pincode and 0 otherwise. In cols. 2 and 4, we classify pincode into $Developed(top50\%)$; $IntermediateDevelopment(10-50\%)$ and $Least\ developed(bottom10\%)$ based on the values of *Poverty Rate*. All regressions are estimated using pincode and year-month fixed effects. Standard errors clustered by pincode are reported in parentheses. All variables are defined in the Variable Appendix. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

	1	2	3	4
	Number of Trades (New Investors)		Number of Investors (New Investors)	
Connect	-0.143**	0.549***	0.064**	0.195***
	(0.056)	(0.028)	(0.030)	(0.013)
Connect X Rural	0.870***		0.280***	
	(0.058)		(0.031)	
Connect X Intermediate Development		0.114***		0.016***
		(0.038)		(0.017)
Connect X Least Development		0.015		-0.017
		(0.061)		(0.030)
Pincode FE	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y
District-Year FE	Y	Y	Y	Y
N	592195	592193	592204	592202
Adj. R-sq	0.773	0.802	0.875	0.899