Dirty Air and Green Investments: The impact of pollution information on portfolio allocations

Raymond Fisman* Pulak Ghosh† Arkodipta Sarkar‡ Jian Zhang§

January 24, 2024

Abstract

We study exposure to pollution information and investment portfolio allocations, exploiting the rollout of air quality monitoring stations in India. Using a triple-differences framework, we show that retail investors' investments in "brown" stocks are negatively related to local air pollution after a monitoring station appears nearby, with particularly pronounced effects on "alert" dates when air quality is listed as harmful to the general population. The effect of pollution information on investment choices is most prominent amongst tech-savvy investors who are most plausibly "treated" by real-time pollution data, and by younger investors who tend to be more sensitive to environmental concerns. Overall, our results provide micro-level support for the view that salience of environmental conditions affect investors' tastes for green investments, and preferences for environmental amenities more generally.

*Raymond Fisman is at Boston University. email: rfisman@bu.edu

†Pulak Ghosh is at the Indian Institute of Management Bangalore. email: pulak.ghosh@iimb.ac.in

‡Arkodipta Sarkar is at the National University of Singapore. email: asarkar@nus.edu.sg

§Jian Zhang is at the University of Hong Kong email: zhangj1@hku.hk

Conflict-of-interest disclosure statement

Raymond Fisman
I have nothing to disclose

Pulak Ghosh
I have nothing to disclose

Arkodipta Sarkar
I have nothing to disclose

Jian Zhang
I have nothing to disclose

1 Introduction

Globally, investment in so-called ESG (environmental, social, governance) funds has grown by over 500 percent, from US$4.8 trillion in 2010 to US$29.2 trillion in 2021, nearly three times the rate of growth of assets under management more generally.[1] Much of the growth and attention has focused on the "E" in ESG, with sustainable investment seen as one mechanism for disciplining firms that generate negative environmental externalities.

This attention on sustainability in investment appears timed to increasing concerns about pollution generally and climate change specifically. There is, however, little empirical evidence on the forces that have led to the rise in ESG investment (Hong, Karolyi and Scheinkman, 2020). It may be driven by "supply"—a proliferation of ESG funds that provide ready investment opportunities. Alternatively, increased ESG investment may result from a shift in investor preferences due to the greater relevance and salience of environmental issues resulting from, for example, global accords such as the Paris Agreement or increased frequency of extreme climate events (e.g., hurricanes, forest fires, floods, droughts, and heat waves).

Yet identifying a role for investor tastes is a challenge – news about climate change often serves as a common shock which may be confounded by concurrent events that may in turn also impact portfolio allocations. If one wishes to exploit panel variation, one requires shocks that impact investor tastes but do not affect perceived returns of green versus brown investments. Additionally, one requires highly disaggregated investor data—with linkages to geography or some other source of exposure to environmental conditions—in order to take advantage of any panel variation that one might exploit.

In this paper we document that exposure to pollution information leads investors to invest less in "brown" industries, using a triple-differences framework framework applied to geocoded data on the portfolios of Indian retail investors. We take advantage of the introduction of continuous air quality monitoring stations (CAAQMSs), which began in 2006 and accelerated during the 2010s. We posit that the arrival of a monitoring station makes pollution more salient for those located nearby, who

[1] See https://www.unpri.org/about-us/about-the-pri, accessed January 8, 2023.

receive easy access to air quality data. Whereas pollution may have simply been ignored prior to the appearance of CAAQMSs, we assert that investors may have a taste for greener investments in response to more extreme pollution once air quality information becomes available in an easy-to-access (e.g., via smartphone or billboard) format. Helpful for identification, the monitoring stations' readings were advertised as relevant for a range of 20 kilometers, offering a natural way of defining investors who are "treated" with readier access to pollution information, which we compare to the portfolio allocations of "control" investors that are more distant from monitoring stations (but still close enough to the treatment population that they plausibly serve as a relevant benchmark – and we note that in robustness checks we show that our results are robust to using a tighter range for our treatment definition).

We utilize a comprehensive database of trading records from the National Stock Exchange of India (NSE) to construct the portfolios of approximately 19 million investors during the years 2004–2020. We can trace any trades in these portfolio to the individual's (anonymized) permanent account number, and the associated postal index number (PIN) code for the account holder. We can thus calculate, with a high degree of precision, the distance between an investor's address and the nearest CAAQMS, and evaluate how the sensitivity of their portfolio to air quality changed after the monitoring program made this information widely available. (As we explain below, we may observe—albeit imperfectly—air quality prior to the creation of CAAQMSs via satellite data, though importantly, data from these stations were not made available to the public in real-time. This allows us to distinguish the role of salience from shifts in actual pollution that might be correlated with the creation of monitoring stations.)

Our main focus is to evaluate how access to pollution information affects the extent to which household ESG investment respond to short-term pollution fluctuations, specifically the brown-share-pollution-gradient $\frac{\partial \text{BrownShare}}{\partial \text{Pollution}}$.This measurement choice aligns with the theoretical ESG-augmented framework established in earlier work, in which investors derive a warm-glow utility from holding green stocks, making the weighted sum of ESG scores for all equity holdings a critical factor for portfolio allocation (e.g., Pedersen, Fitzgibbons and Pomorski, 2021).[2] Specifically, we examine the

[2]Alternatively, this measurement choice can also be motivated by a simple discrete choice model of green versus brown stocks, which generates the pooled average of choices at the household level, i.e., the brown share (Berry, 1994).

sensitivity of an investors' holdings in "brown" stocks to the extent of air pollution before versus after the opening of a nearby monitoring station. In an event study specification, we show that the brown share of investors in an impacted PIN Code is unrelated to air quality in the two years leading up to the appearance of station, then experiences an immediate and sustained decline. Our point estimates indicates that, for a one standard deviation decline in (within-monitoring-station) air quality, there is a 0.5 percentage point (1.25 percent of the mean) decrease in the share of brown stocks in affected investors' portfolios after a monitoring station appears (whereas the sensitivity is approximately zero beforehand). We use a "donut hole" approach to define a benchmark set of investors, located in PIN Codes in the 40-60 kilometer range around a given monitoring station—these investors are sufficiently distant from the station that its reading had less relevance (as noted above, the range communicated to the public was just 20 kilometers). For this group of "donut hole" investors we observe no change in investment sensitivity to air quality after a monitoring station opens. Since green stocks do not outperform brown stocks over this period, we suggest that our findings are most plausibly driven by investor tastes and pollution salience rather than a shift in expected returns.

To link these patterns more directly to pollution salience, we examine shifts in portfolio allocations around changes in particularly salient shifts in air quality. Specifically, we look at air quality "alert" days when the air quality reading transitions from Moderate (yellow) to Poor (amber), a change that also triggers government warnings about the health risks posed to the broader population. Using a regression discontinuity approach around this air quality cutoff, we show that there is a discrete drop in the share of brown stocks held by retail investors.

In a final set of results, we explore heterogeneity in investors' responses to exposure to air quality information. Given that air quality information from monitoring stations was delivered primarily via a smartphone app, we suggest that salience may have been greater amongst tech savvy investors who were also more apt to trade via mobile phone. Thus, we begin by splitting the sample based on whether a trader most often executed transactions via mobile device, the internet, or by some other means (i.e., through either a trading kiosk or a broker). We find a far greater shift in responsiveness to pollution after a monitoring station appears amongst mobile-based investors; there is also a greater sensitivity for internet-based investors relative to those using more traditional methods.

We next explore heterogeneity by age. Beyond a moderately higher degree of tech savvy, young investors may also have greater environmental awareness given that they disproportionately bear the consequences of climate change and environmental degradation. We thus posit that the young will be most responsive to being confronted with pollution information.[3] We find that air quality information has the greatest effect on investment sensitivity for young investors, and the weakest effect for the elderly. Finally, we split the sample by gender. While the motivation for this heterogeneity test is less straightforward than our other sample splits, we suggest that it links to the broader literature on the "environmental gender gap"—women express greater concern for environmental issues than men (e.g., Xiao and McCright, 2015). We extend this line of research to show that this gender gap applies to male versus female investors: the portfolios of women are more sensitive to air quality than men, once this information becomes readily available.

Our work connects most directly to the body of research that aims to understand investors' non-pecuniary concerns generally, and specifically their interest in ESG investments. Closest to our own work is that of Choi, Gao and Jiang (2020), which studies the link between weather and investment in a cross-stock-exchange framework utilizing largely cross-country variation in temperature. They show that higher temperatures are associated with a decline in prices of "brown" stocks, driven by the trading activity of retail investors in particular. Our work is distinct from Choi, Gao and Jiang (2020) in three main ways, and these distinctions in turn serve to highlight the broader contributions of our paper.

First, our study emphasizes the crucial role of information dissemination on environmental issues in shaping investors' ESG investment behavior. While Choi, Gao and Jiang (2020) focuses on the impact of extreme weather events, we specifically highlight the importance of making information available and salient. In a sense, the analysis of Choi, Gao and Jiang (2020) is analogous to the pollution-investment correlation during the pre-treatment baseline period in our study, so that we show the importance of making information on environmental problems more readily available for

[3] Alternatively, air quality may be more salient for the elderly, given the health consequences, though to the extent that this is the case, our results suggest that it is dominated by the aforementioned effects.

retail investors' trading decisions. This is of direct policy relevance, as it suggests that effective communication of environmental concerns – and social issues more broadly – may encourage more socially conscious investment choices.

Second, our detailed individual trading data allows for more persuasive identification relative to earlier work. Choi, Gao and Jiang (2020) in particular suggest that (retail) investor tastes, driven by increased climate concerns, play an important role in ESG investing and show an effect on stock market returns across the cross-section of firms. However, as we note above, their data consist of relatively coarse information on retail ownership (residual imputed from information on blockholders and institutional ownership) and variation in temperature across cities (located in different countries) where exchanges are located. Consequently, the analysis is subject to potential measurement error and usual critiques of cross-country analyses. Our detailed individual-level trading data allows us to tie trading behavior to much more localized, high-frequency shifts in environmental conditions. Finally, the granularity of the data allows us to exploit important cross-sectional variation across demographic characteristics like age and gender as well as trading method (i.e., mobile versus internet versus broker). Our heterogeneity analysis analysis allows us to provide suggestive evidence on how information on environmental quality influences and alters investor behavior.

More broadly, our work sits at the intersection of several large and growing areas of inquiry: research on ESG investments, salience and investor behavior, and the salience of environmental problems and attitudes toward environmental issues.

As described in Hong, Karolyi and Scheinkman (2020), ESG research may generally be classified in one of several categories. Broadly speaking, one branch of research focuses on the extent to which climate (and resultant environmental) risks are incorporated into stock prices (e.g., Görgen et al., 2020; Bansal, Kiku and Ochoa, 2016) and other assets such as homes (Baldauf, Garlappi and Yannelis, 2020; Murfin and Spiegel, 2020) and agricultural land (Hong, Li and Xu, 2019). To the extent that environmental risks become more important over time in ways that are not fully anticipated, there may be differential returns for green versus brown companies. The question of whether socially responsible companies generated higher returns more generally has been much-studied, but without any clear resolution (e.g., Hong and Kacperczyk, 2009; Berg et al., 2022).

Our own work is much more closely tied to the strand of work that explores the link between the beliefs and preferences of investors, and their green versus brown portfolio allocation decisions. In addition to Choi, Gao and Jiang (2020), which focuses more on retail investors, Alok, Kumar and Wermers (2020) examine institutional investors' responses to climate-related disasters and find that nearby fund managers respond more sharply than more distant ones.

While we focus on pollution salience, there is a much larger literature which examines how inattention and salience impact portfolio allocation decisions, whether driven by the notability of past returns (Bordalo, Gennaioli and Shleifer, 2013; Cosemans and Frehen, 2021); the media (Huberman and Regev, 2001; Jiang et al., 2022); or ESG ratings themselves (Pelizzon, Rzeznik and Hanley, 2021).

Finally, moving beyond finance-focused applications, our work relates to the larger literature in social psychology and economics which explores whether exposure to (idiosyncratic) environmental shocks impact beliefs and attitudes toward climate change (Li, Johnson and Zaval, 2011; Zaval et al., 2014; Lujala, Lein and Rød, 2015). Moving beyond attitudes to actions, Barwick et al. (2019) examine how pollution information—delivered by the same type of real-time monitoring stations that we consider here—impacted online search behavior in China, with more searches related to pollution avoidance behaviors after a monitoring station appears. Our work also documents real behavioral changes, but in a very different domain, and one that has broader social consequences rather than one with private benefits. In line with this earlier work, our findings suggest that policies that give greater visibility to environmental quality may be useful in encouraging "greener" behaviors. While our setting involves investment, it is plausible that one may observe similar shifts in, say, energy use or other consumer behaviors.

2 Institutional Background

This section provides an overview of India's rollout of continuous ambient air quality monitoring stations (CAAQMSs), which we exploit in our empirical design. The introduction of CAAQMSs was a part of a broader National Air Quality Monitoring Programme (NAMP), set up by the Central

Pollution Control Bureau (CPCB) in coordination with state-level control boards. As we discuss in this section, the initiative led to a rapid and significant increase in the availability of pollution data to the general public, resulting in a considerable improvement in awareness of pollution problems.

A Brief History of Air Quality Monitoring in India: The CPCB has been systematically collecting pollution data under its national monitoring programme since 1985.[4] Initially, monitoring stations collected data on four key pollutants: sulfur dioxide (SO2), nitrogen dioxide (NO2), PM10 (particulate matter under 10 microns), and PM2.5 via *manual* monitoring stations scattered around the country. These manual stations involved the collection of ambient air over a period of several days, which was then transported to a central location where the data were analyzed. The resultant (manually generated) report was then archived to the Environment Air Quality Data Entry System. This process of data collection took weeks to complete, so that much of the information was stale by the time the report was finished. Furthermore, even the reports were often inaccessible – it was left to local pollution control authorities to upload and/or make this information available. Finally, the data was, in some cases, constrained or unsuitable for analysis, so no data were available at all (Pant et al., 2019). The CPCB itself released the data at a monthly or annual frequency.

As the preceding description makes clear, the manually collected pollution data gathered under the earlier NAMP were slow to produce, of questionable quality, and hard for the general public to access (if the data were made available at all).

Both the quality and availability of data shifted markedly with the introduction of CAAQMSs, first piloted in 2006 and later expanded in 2016 (see below for more details on the rollout). These newer monitoring stations collect information on a wider range of pollutants than earlier instruments.[5] Furthermore, both the collection and analysis of data has been fully automated via "internet of things" technology that facilitates continuous automated data collection, as well as the transfer of pollution data to a central server in real-time every few minutes. At the center, data analysis is also automated, and is ready for use shortly after collection. Pollution data from the CAAQMSs are used to calculate

[4] Air quality monitoring began earlier, in 1978, in 10 cities – Ahmedabad, Mumbai, Kolkata, Kochi, Delhi, Hyderabad, Jaipur, Kanpur, Chennai, and Nagpur.

[5] Measured pollutants including PM2.5, PM10, nitrogen oxides, sulphuric dioxide, benzene, toluene, ethylbenzene, and xylene.

an air quality index (AQI), a standardized metric that incorporates a wider range of pollutants than the earlier NAMP monitoring program. Real-time AQI readings are publicly available online and via a smartphone app, with historical data archived by the CPCB and downloadable on its website.[6]

A primary purpose of the continuous air quality monitoring program is to create public awareness of environmental conditions. This objective was greatly facilitated with the advent of CAAQMSs, as the earlier manually-driven system could not provide real-time environmental data to the public. This information is now delivered via public displays, web widgets, alerts, and a proliferation of mobile apps that provide localized information on air quality.

Rollout and location of CAAQMSs: Real-time pollution monitoring was first piloted in Delhi in 2006, with the expansion accelerating only in 2016 (see Figure 1). [7]

The decision of where to locate monitoring stations is done by the CPSB in consultation with the state-level pollution control boards. The criteria naturally include consideration of nearby pollution severity and population; however, there is a much longer list of practical concerns that include geographic obstructions, security, cost, and power availability, among many others. While we identify our main results from a difference-in-differences framework, it is still worth noting that within a relatively narrow geographical region, the locations of CAAQMSs (as well as their timing) are dictated in large part by considerations that are largely exogenous to factors that might affect the portfolio allocation decisions of individual traders.[8]

Classification of Treated versus Control PIN Codes: Our empirical approach is to compare individuals located in the immediate vicinity of newly established CAAQMSs (i.e., the treated group) to individuals who are marginally further away, and thus "untreated" with pollution information (i.e., the control group). There is no clear threshold for how the relevance of air quality readings diminish with the distance from a monitoring station. We propose a donut-shaped approach that leverages the

[6]It is difficult to measure exactly how many people keep track of local air quality via smartphone because, in addition to the CPCB's own app, there are dozens more that provide real-time AQI data for India and internationally. Many are listed as having 100,000+ downloads. The CPCB reports that its own app has been downloaded over 500,000 times. For more details on the monitoring program as well as references for AQI measurement, see Pant et al. (2019)

[7]See, e.g., Gulia et al. (2022) on the timing of CAAQMSs. For the current map of stations, see https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing, last accessed June 20,2023.

[8]Greenstone et al. (2022) similarly use the rollout of monitoring stations in China to study the impact of real-time information on local underreporting of pollution, and also online searches for, e.g., face masks and filters.

comparison of average pollution levels within the inner and outer rings of the 'donut,' while leaving out the area in the middle as the buffer region. We use a 20-kilometer radius of a CAAQMS to define traders who are "treated" with AQI information when a monitoring station opens. This range was determined based on discussions with CPCB officials, based on the distance that pollution readings would be perceived as relevant to the public. We use investors located 40 to 60 kilometers from a monitoring station as the benchmark or control group.[9]

There are cases of overlapping treatment and control areas in some densely populated areas, as India's major metropolitan areas are covered by multiple stations. For example, Delhi has 41 monitoring stations to cover its 20 million residents spread across nearly 1500 square kilometers. There are 21 stations for Mumbai (population 21 million, area 603 sq km); and 7 stations for Kolkata (population 15 million, area 205 sq km). We describe below how we deal with the handful of PIN Codes where this issue arises.

Note, however, that once one gets outside of a handful of major metropolitan areas, coverage is much sparser. For example, Jodhpur, a city of 1.6 million, has a single monitoring station, and there are no other stations within 60 kilometers of it. The entire state of Jammu and Kashmir has just a single monitoring station in its largest city of Srinagar (population 1.7 million).

Before proceeding to the next section, we emphasize that our results do not depend on this specific treatment-control definition. We also consider narrower radii of 5, 10, or 15 kilometers in robustness checks below (Appendix Table A4). Furthermore, since India's largest cities have a preponderance of overlapping monitoring stations, we demonstrate that the results still hold when we exclude the small subset of traders located in Mumbai, Delhi, and Kolkata.

[9]These cutoffs are not definitive, and for this reason we offer a number of robustness checks to show that our results are not sensitive to these particular choices. We choose our main cutoffs based on our reading both of Indian government communication as well as prior research on a monitoring station's relevant radius. For example, in a UNICEF report that examines the proportion of children under threat to respiratory infections from pollution, Rees, Wickham and Chandy (2019) highlight the lack of reliable data on ground-level monitoring of pollution through air monitoring stations. In this context, the report uses 10km as the first relevant radius around the monitoring station and 50km as the upper limit. In the context of India, Roychowdhury, Somvanshi and Kaur (2023) highlights that a CAAQMS has a range of 10km, and if there are no topographical barrier the range could be as much as 50km. Further, in most mobile apps the name of the monitoring station is mentioned which makes the pollution number less relevant for a resident situated a significant distance away. We take 0-20km as a distance that would very likely be affected by pollution information from a station (showing also robustness for narrower radii), while in most cases its relevance likely wanes by 40km.

3 Data

We use several data sources in our analysis that allow us to connect pollution, monitoring, and trading behavioral at a granular level. The databases include individual-level stock holdings over time for Indian investors as well as basic geographic and demographic information; geocoded information on the timing of the installation of CAAQ monitoring stations, and information on local pollution inferred from satellite images.

Stock Trading Data: We use comprehensive data on stock trading from the National Stock Exchange (NSE) of India, one of the largest stock exchanges in the world[10]. We obtained the data on the universe of trading records from the NSE for the period of 2004–2020. The data allows us to observe a number of features for each transaction – the anonymized Permanent Account Number (PAN) of the trader, the transaction date, the security ticker, total shares purchased or sold, and the execution price. We limit our analysis to transactions involving stocks included in the Prowess Database (similar to CRSP in the U.S.) maintained by the Centre for Monitoring Indian Economy (CMIE).[11] In addition, we retain only securities that are common shares of domestic stocks and exclude trading activities related to ETFs and foreign stocks.[12] For each retail investor, we further obtain their geographical location at the six-digit PIN Code level, which is essential to match our trading data to information on the location and opening of CAAQMSs. Overall, the initial sample includes equity transactions for 19 million unique investors across India.

We provide the geographic distribution of domestic retail investors in Appendix Figure A1. Unsurprisingly, there is a particularly high concentration of investors in Delhi, Kolkata, and Mumbai. These are India's three most populous cities, and also major financial centers. However, as the figure makes clear, NSE investors are widely distributed across the entire country.[13] We use historical transaction data from as far back as 2004 to construct a holding sample containing an observation

[10]The National Stock Exchange (India) ranks No. 9 in terms of market cap in 2023: https://www.edudwar.com/top-10-stock-exchanges-in-the-world/

[11]Prowess is the standard database employed by researchers studying Indian equity markets. See, e.g.,Khanna and Palepu (2000); Goldberg et al. (2010); Balasubramaniam et al. (2023); Bau and Matray (2023).

[12]None of the ETFs in our sample have an explicit ESG orientation.

[13]Furthermore, as we discuss below, in Appendix Table A5 we show that our baseline results are virtually unchanged when we exclude investors from these three areas.

for each investor-stock-day following standard methods (e.g., Frydman, Hartzmark and Solomon, 2018; Ben-David and Hirshleifer, 2012).[14] We then aggregate investors' portfolio holdings based on their location to match the variation in CAAQMS introduction and conduct our main analysis at the PIN Code-by-month level. Intuitively, each regional account describes the stock holdings of a representative regional investor.

Identifying Brown Stocks: We classify firms as *"brown"* or *"green"* based on their industry classifications, following Choi, Gao and Jiang (2020), which in turn uses the definition of the Intergovernmental Panel on Climate Change (IPCC) to classify five sectors as sources of high emissions – Energy; Transport; Buildings; Industry (such as chemicals and metals); and Agriculture, Forestry, and Other Land Use. Then, following Krey et al. (2014), each sector is classified into sub-categories, each of which is hand matched with industry names provided by Datastream. (Note that emissions of air pollutants are highly correlated with CO2 emissions through fossil-energy consumption, as noted by Lin et al. (2023) and Agee et al. (2014), making this industry-level measure also a plausible proxy for a firm's contribution to lower air quality.)

We categorize green versus brown stocks at the industry- rather than firm-level for two reasons. First, because of our focus on retail investors, it is better-aligned with the likely heuristics used by most retail investors – who are less apt to explore the specific ESG impact of individual firms – when making investment decisions.[15] Second, and more importantly, our choice is driven by the empirical data challenges associated with firm-level ESG coverage in India. Compared to developed markets like Europe and the United States, the availability and quality of firm-level ESG data in India are relatively sparse and less reliable.

We use the industry classifications provided in Prowess database to hand match to Datastream industries, in order to classify each firm (via its industry classification) as green or brown based on the Choi, Gao and Jiang (2020) list. Figure 2 plots the evolution of "brown" and "green" (i.e., non-brown) stocks over our sample period. In Panels A and B we show the number of stocks and

[14] Following Ben-David and Hirshleifer (2012), we remove a trader's investment from the sample if the cumulative number of shares becomes negative at any point (owing to a purchase that occurred prior to the start of the sample period).

[15] Consistent with this assumption, Moss, Naughton and Wang (2023) show that that retail investors exhibit muted responses to firms' ESG disclosures, using data from Robinhood.

market capitalization, and in panel C we show brown stocks' share of market capitalization. Brown stocks' market share varied between about 30 and 40 percent during the period of 2004-2020, peaking in 2010 then dropping steadily thereafter.

Pollution Monitoring Stations – CAAQMSs: We obtained geo-coded data on the location and timing of continous air pollution monitoring stations from India's Central Pollution Control Board. The first CAAQMS appeared in Delhi in 2006, and by 2021 there were 311 stations spread across the country (see Figure 1, Panel B). We classify a PIN Code as in the "treatment" group if its centroid is located within a 20 kilometer radius of a CAAQMS under the assumption that, when a monitoring station appears, investors in the PIN Code gain ready access to real-time air quality data. We leave a buffer of a "donut hole" region comprised of PIN Codes 20-40 kilometers around the station, and then classify PIN Codes that have centroids located 40-60 kilometers from each station as the "control" group.

We classify a PIN Code as control only when its centroid is not within the treated region of any other CAAQMS. Thus, the treatment and control groups are mutually exclusive and the sample is created such that there is no overlap across the groups, with every CAAQMS having unique set of treated and control PIN Codes. Panel A of Appendix Figure A2 illustrates our assignment approach for single monitoring station in Jodhpur, which provides a relative straightforward case, given that the city hosts only a single station. Panel (b) shows an example of two monitoring stations in Delhi which illustrates that India's largest cities have multiple stations that are generally within 60 kilometers of each other; this creates some complications in treatment and control assignments, which we address as described above. However, as also noted earlier, we sidestep these complications in a robustness test in which we exclude Mumbai, Delhi, and Kolkata from our analysis.

Satellite Data on Pollution: We wish to measure local air pollution at the PIN Code level in a consistent manner both before and after the introduction of CAAQMSs. To do so, we take advantage of data generated by NASA's Moderate Resolution Imaging Spectroradiometer on its Terra satellite. These readings are used by NASA to generate data on aerosol optical depth (AOD), which is a widely-used proxy for pollution from outdoor particulate matter and reflects the density of various particles such as sulfates, nitrates, black carbon, and sea salts (see, e.g., Van Donkelaar, Martin and Park, 2006). Following existing studies such as Tsai et al. (2011), we confirm in Appendix Figure A3 that AOD and PM 2.5 (obtained from the CAAQMSs) are highly correlated, based on a comparison following the implementation of the monitoring program, when both pollution proxies are available. The AOD data are available at a frequency of 30 minutes in a 10-by-10 kilometer grid. Our main pollution data are average monthly readings for the grid location that contains the PIN Code's centroid.

One natural concern is the endogenous timing of CAAQMS installations, i.e., monitoring stations may have been placed in areas where pollution is worsening. We discuss this issue in Section 4.2 below: to briefly summarize, we do not observe any shift in pollution (nor investor attributes or local economic conditions) around the opening of a monitoring station.

Weather Data: Air quality is correlated with sunshine and/or precipitation, which may in turn also affect trading behavior. We thus include meteorology controls for weather.These data are collected from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 Land datasets, which combine observations from satellites, weather stations, and weather forecast models as far back as 1963. As part of the ERA5 climate reanalysis, the data also provide high-resolution information about land surface variables. Similar to the AOD data, we use the zonal statistic method to calculate the monthly average and create a panel data at the pincode-month level, for rainfall and for surface temperature, which we include in all main specifications.

Sample Description and Summary Statistics: Our regression analysis is at the PIN Code-month level. Our main sample consists of 1,859 distinct "treated" PIN Codes and 5,254 "control" PIN Codes, over the period January 2004 to June 2020. The panel is unbalanced – in 65 percent of PIN Code-month observations, stock holdings are zero so that our main outcome, share of brown stocks, is undefined. Some monitoring roll-out occurs during 2017-2019, close to the end of our sample period. To ensure that the underlying sample for each station is a relatively balanced panel (i.e., we do not have observations very distant in time from the appearance of a monitoring station), we have additional filtering, and focus on the +/- 4 years time window around the rollout date. The total number of observations for the baseline specification at the PIN Code-month level is 499,036.

Table 1 presents summary statistics for the primary variables we use in our analysis. Aggregated across our entire sample period, we observe that a typical retail investor holds roughly 41.8% of their portfolio in brown stocks, marginally above the overall share of brown stocks in the Indian market (see Figure 2). Of relevance for the heterogeneity analyses we present in Section 5.5, Table 1 also provides portfolio brown shares across investor groups. We observe that female investors hold a marginally lower fraction of assets in brown stocks relative to their male counterparts (41.2 versus 41.9%), while the share of brown stocks is notably lower for young investors (40.4%), relative to mid-aged and old investors (41.9 and 42.8% respectively).

4 Empirical Strategy

4.1 Empirical Specifications

Our objective is to assess how the sensitivity of individual's green investments to pollution changes as the monitoring program improves public access to air quality information. Specifically, we investigate the connection between pollution exposure and investments in brown stocks, which we refer to as the brown-share-pollution gradient, and measure how it differs before versus after the (staggered) rollout of local monitoring station across regions. As described in section 2, our methodology differs slightly from that of Barwick et al. (2019), since we do not rely exclusively on the staggered rollout of monitoring stations for identification. In addition, to control for any time-varying local confounding factors, we leverage detailed information on investors' geographies and add a third layer of differences. Specifically, we compare changes in outcomes for retail investors close (i.e., within 20 kilometers of a station, the *treated ring*) with those located further away from the monitoring station (i.e., 40-60 kilometers, the *control ring*).

We estimate the changes in brown-share-pollution gradient using the following specification:

$$\begin{aligned} Brown\,Share_{p(m),t} &= \beta \times Treat_p \times Pollution_{p,t} \times Post_{m,t} + \sum_k \alpha_k \times Other\ Interactions \\ &\quad + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t} \end{aligned} \tag{1}$$

where p, m, t denote a PIN Code, monitoring station area and month, respectively. $Brown\,Share_{p(m),t}$ represents the average share of brown stocks across all retail investors in PIN Code p in monitoring station area m at time t. A station area m contains PIN Codes with centroids within 20 kilometers of the monitoring station – *treatment PIN Codes* – and PIN Codes 40-60 kilometers from the station – *control PIN Codes*. $Treat_p$ is an indicator variable that equals one if a PIN Code falls into the treated group (i.e., within 20 kilometers of the monitoring station) and zero otherwise. The treatment variable captures exposure to this information in real-time as a result of the arrival of a monitoring station. $Pollution_{p,t}$ is the local ambient air pollution measure from AOD in PIN Code p and month t. It is worth reiterating that AOD measures local residents' exposure to pollution both pre- and post-arrival of a monitoring station, although air quality information (collected by the station) is available only in the post period. The binary variable $Post_{m,t}$ is an indicator variable that takes on a value of one beginning in the first month following the the installation of a local monitoring station m. *Other Interactions* include all lower-order terms, such as $Treat_p \times Pollution_{p,t}$, $Treat_p \times Post_{m,t}$, $Post_{m,t} \times Pollution_{p,t}$, and $Treat_p$, $Pollution_{p,t}$ and $Post_{m,t}$. $X_{p,t}$ is a vector of control variables including local trading activities (number of retail investors, total turnover) and weather conditions (rainfall, temperature) in PIN Code p and month t.

The set of PIN Code fixed effects, γ_p, controls for time-invariant factors of a PIN Code, enabling us to identify differences based on time-series variation in whether a nearby monitoring station has opened. To account for time-varying characteristics in the region around monitoring stations, we include station area $\times$ time fixed effects ($\lambda_{m,t}$) in our most saturated specification. This allows us to identify differences between treatment and control groups for each station at the same point in time. We cluster standard errors at the PIN Code level to account for possible serial correlations among the sample periods. The coefficient of interest, β, captures changes in the relationship between pollution exposure and investment in brown stocks that occur with the appearance of a monitoring station, comparing nearby (treated) versus more distant (control) PIN Codes.

We also implement a event-study version of the specification in Equation (1) and explore dynamics around the treatment date. In particular we substitute for the $Post_{m,t}$ dummy with a collection of indicators measuring the quarters elapsed since the installation of a monitoring station:

$$\begin{aligned} Brown\,Share_{p(m),t} \quad = \quad & \sum_{i=-6}^{6} \beta_i \times Treat_p \times Pollution_{p,t} \times D(t \in i)_{m,t} + \sum_k \alpha_k \times Other\ Interactions \\ & + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t} \end{aligned} \tag{2}$$

where $D(t \in i)_{m,t}$ is a dummy variable that takes a value 1 if the month t belongs to the k^{th} quarter since the monitoring station appeared. This specification allows for the tracking of the evolution of β_i over time in relation to the rollout of monitoring station and provides a graphical representation of the dynamics of brown-share-pollution gradients. The number of pre- and post-treatment periods available varies across different waves of PIN Codes. To ensure a balanced time frame for analysis, we report β_i for a period of 6 quarters prior to and following the monitoring rollout, while binning the remaining more distant periods in the sample.

4.2 Discussion of Possible Threats to Identification

Most previous studies have focused on estimating the pollution-concentration-response function, which quantifies the causal effect of air pollution exposure on health and non-health outcomes. However, identifying this relationship is challenging due to the endogeneity of ambient pollution levels. Endogeneity can arise from omitted variables, such as individual endogenous location choice, or measurement errors. Graff Zivin and Neidell (2013) provide a detailed review of the sources of endogeneity. Equation (1) emphasizes that our focus is distinct from theirs. We are not aiming to estimate the impact of pollution *per se*, but rather to measure how the causal effect (sensitivity of brown share to pollution exposure) varies before versus after the introduction of monitoring stations.

When attempting to estimate the slope of the relationship between brown shares and pollution separately using data from before and after the station rollout, the endogenous nature of pollution could lead to inconsistent estimates for both periods. However, assuming certain conditions hold, the inconsistency in slope estimates could offset each other, hence making the OLS estimate of β in Equation (1) consistent. The crucial identification assumption is that the rollout schedule is not correlated with the difference in time-varying unobservables between treated and control areas that could drive the observed outcome, given the controls.

In this section, we investigate the validity of the identifying assumption and consider the potential estimation biases that arise from a staggered rollout setting, which has garnered increased attention in recent literature.

CAAQMS Rollout: Although the assignment of CAAQMS is not entirely random, it is primarily the result of negotiations between the Central Pollution Control Board and state-level pollution control authorities. As discussed in Section 2, the assignment of CAAQMS is influenced by a long list of exogenous concerns, including population, geographic obstruction, power availability, and so forth. Note that we focus on a relatively narrow geographical region and examine geographically proximate control and treatment PIN Codes around the same monitoring station. Moreover, our regression analyses include station-area $\times$ time fixed effects, and therefore eliminate any potential threats to our identification posed by time-varying differences in local characteristics. Thus, the location and timing of monitoring are largely exogenous to the portfolio allocation decisions of retail investors.

To probe empirically whether endogenous rollout is a substantial concern, we first test whether differences in time-varying local characteristics between control and treatment PIN Codes correlate with the policy rollout, after controlling for fixed effects in Equation (1). We begin by plotting the dynamics of pollution levels (as captured by AOD) between treated and control areas in Figure 3, as a function of time relative to monitoring. We do not observe any significant shift in the local pollution gap between treated and control investors before versus after the introduction of a monitoring station. Second, we examine a set of confounding factors, including local investor composition as captured by growth of investor base and entry of new investors, and economic activities (entry of new firms and night time light images) that likely differentially impact the portfolio allocation of individual investors. We show in these "balance" tests that program rollout is uncorrelated with changes in these attributes (See Appendix Table A2), supporting the identification assumption described at the beginning of Section 4.2. Finally, the event-study analysis detailed in Equation (2) helps identify possible endogeneity issues, to the extent that they are reflected in pre-existing trends in the outcome variable. We return to this issue below in the discussion of Figure 5, which will lend additional support to the identification assumption.

Bias in Staggered DID: A recent literature on difference-in-differences estimation indicates that the two-way-fixed-effects (TWFE) estimator may be thought of as a weighted average of all potential 2×2 DID estimates among three groups, and can be biased when there are heterogeneous treatment effects over time or across units (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). This bias results from some of the 2 DID estimates entering the average with negative weights, thus introducing biased estimates that dilute the true treatment effect. The "negative weight" problem is primarily driven by comparisons between previously treated groups as controls and newly treated groups.

Our results are not affected by the estimation concern mentioned above, as we are interested in the brown-share-pollution gradient rather than the level of brown share itself. Our specification can be viewed as a "stacked" difference-in-differences model, which produces an efficient estimator for uncovering the aggregated treatment effect through OLS (Cengiz et al., 2019; Baker, Larcker and Wang, 2022). In essence, we estimate the average treatment effect on the pollution gradient across multiple "canonical" DiD regressions with two groups and clean pre- and post-periods.

To explore the analogy of TWFE in our setting, below we also estimate the event study specification separately for investors located in the treated and control PIN Codes and plot the β_i estimates. As we discuss further in presenting our main results, we observe a significant shift in the brown-share-pollution gradient, but only for investors in treated PIN Codes, lending support to the validity of the parallel trend assumption.

5 Results

5.1 Visual Evidence

We begin with a visual presentation of how the brown-share-pollution gradient shifts with the introduction of local monitoring stations. We present our results in a canonical 2×2 difference-in-differences setup, which is relatively transparent and easy to interpret, and later also show event-study plots to highlight that these shifts coincide precisely with the arrival of monitoring stations.

In each panel of Figure 4, we show a binned scatterplot that displays the relationship between observed air quality and brown-share-holdings for both control and treated PIN Codes (top and bottom) before and after the arrival of a monitoring station (left and right). We sort PIN-Code-month average pollution into 100 buckets and plot average pollution against brown share. We adjust the local brown share for the time-varying mean in a station area across both control and treated PIN Codes. Consistent with a lack of attention to air quality, we observe no correlation between brown-share and air quality before the appearance of a monitoring station in both treated and control PIN Codes (left two panels). We similarly observe no relationship in control PIN Codes after a monitoring station appears (upper right panel). We observe a distinct pattern in the lower right panel which shows the relationship for nearby (treated) PIN Codes in the post-period – in this case there is a clear negative relationship between air quality and the investors' brown share holdings.

5.2 Baseline Estimation

Both to control for a wider array of geographic and temporal attributes, as well as to study how the pollution-brown-share gradient evolves over time, we present an event-study plot of β_i in Figure 5. The value of the coefficient for the quarter just before the event is set as the benchmark and normalized to zero. The vertical axis provides the estimated change in the brown-share-pollution gradient ($\frac{\partial BrownShare}{\partial Pollution}$), and the horizontal axis provides the quarters relative to the monitoring program. We observe two patterns. First, for the quarters leading up to the introduction of CAAQMs , the β_i estimates are statistically indistinguishable from zero, lending support to the validity of the parallel trend assumption. Second, the brown-share-pollution gradient becomes negative immediately after the monitoring program rollout, with the estimated β values in the range of -2 and -3 in the post-periods.

Table 2 presents our baseline regression results results based on Equation (1), where we test the robustness of the patterns described above using different sets of fixed effects and controls (note that for ease of interpretation, we demean the pollution variable in all specifications). We present our most parsimonious model in Column 1, with PIN Code and year-month fixed effects. Columns 2 and 3 include interactive state-time and district-time fixed effects respectively, to absorb time-varying confounding effects at the regional level (i.e., changes in local economic conditions or employment) and to allow for common trends specific to each region. Column 4 is the most stringent specification, including interactive monitoring-station-time fixed effects and PIN Code fixed effects.[16] Columns 5–8 provide analogous specifications, but include in addition controls for local trading activity and weather. As explained in the preceding section, our primary interest is in the coefficient on the three-way interaction of $Treat_p \times Pollution_{p,t} \times Post_{m,t}$. This reflects the differential response of investors to pollution after a monitoring station appears, in treated relative to control PIN Codes. The point estimate of -2.4 in our favored specification in Column 8 implies that, relative to the pre-period, after a monitoring station appears, an increase in pollution from the 25th percentile to the 75th percentile (an AOD increase of 0.27, which is the within-station interquartile range) is associated with a decline of the brown share in investors' portfolios in treated areas by 0.65 percentage points (0.27*2.4). This represents a 1.6 percent decrease relative to the mean brown share of 41 percent. In other words, the brow-share-pollution elasticity becomes 3.2 percentage points more negative once a monitoring station appears.[17]

Table 2 also demonstrates that retail investors' investments in brown stocks in control PIN Codes are unaffected by pollution both before and after the introduction of a monitoring station – that is, the coefficients on the the $Pollution$ and $Pollution \times Post$ terms are both insignificant when $Treated = 0$. This finding helps to validate our treated and control assignments, and mitigates potential concerns over spatial spillovers.

To further highlight the fact that our findings are driven by shifts in treated PIN Codes around the appearance of monitoring stations, we show event-study plots for treated and control investors separately. Specifically, in Figure 6 we show the point estimates and 95% confidence intervals for the coefficients generated by a variant of specification (2) above, in which we split the sample into investors located in treated and control PIN Codes.

From an identification perspective, it is comforting that our measure of brown share investments

[16] These high-dimensional fixed effects essentially convert the model to a stack of "canonical" DiD regressions with unit fixed effects and time fixed effects.

[17] The estimate change in brown-share-pollution elasticity is $\frac{0.65/41.87}{0.27/0.54} = 3.2\%$

is essentially flat in the two years (eight quarters) preceding the installation of a monitoring station for both treated and control PIN Codes. At that point there is a clearly discernible shift in the sensitivity of brown investment share to AOD in treated PIN Codes—higher AOD (i.e., worse pollution) is associated with a lower brown share after a monitoring station arrives. We observe no such shift in control PIN Codes. These patterns also mitigate a concern suggested by the lower-order terms in the Table 2 – that treated PIN Codes may have had a lower sensitivity to pollution in treated PIN Codes before the appearance of a monitoring station. This may result from the structure imposed by the triple-interaction specification, which generally forces fixed effects and controls to have identical coefficients for treatment and control investors. Our event plot suggests that this difference disappears when we allow for the more flexible specification illustrated in the event plot (and similarly in the alternative approach we introduce at the beginning of the next section, which accounts for a bias resulting from staggered rollout, we do not observe a pre-CAAQMS difference).

5.3 Robustness and Interpretation

Alternate Definition of Control Groups: As discussed in Section 4.2, recent advances in applied econometrics have shown that, when employed in staggered difference-in-differences (DID) settings, two-way fixed effects models may yield biased coefficient estimates due to varying treatment effects across time or units. As noted earlier, we believe that such concerns are less applicable to our setting, as our primary focus is on the brown-share pollution gradient rather than the brown share level. Further, our baseline specification uses a stacked difference-in-difference by using a control group of more distant PIN Codes for every treated (nearby) PIN Code of a monitoring station. Nonetheless, to further assess the robustness of our results, we adopt the stacked method outlined in prior research (Cengiz et al., 2019; Baker, Larcker and Wang, 2022) and make comparisons of treated PIN Codes versus not-yet-treated and/or never-treated PIN Codes, defined based on the timing of their treatment.

The basic idea is to create, for each treated cohort, "clean" datasets that combine the "treated" PIN Codes (those within 20 kilometers of the station) with other nearby PIN Codes that serve as controls. We consider two different approaches to choosing the "clean" controls within the treatment window for each treated cohort: (1) We include only observations for never-treated PIN Codes, i.e., those that were not treated during our sample period, as comparison groups in a given event window;(2) we include "not-yet-treated" PIN Codes, defined as those that do not get treated within 4 years for each treated cohort, in addition to the never-treated units.

We then stack the resultant datasets and estimate the following specification, which is similar to the baseline analysis:

$$\begin{aligned} Brown\,Share_{p,c,t} &= \beta \times Treat_p \times Pollution_{p,t} \times Post_{c,t} + \sum_k \alpha_k \times Other\ Interactions \\ &\quad + X'_{p,t}\theta + \gamma_{p,c} + \lambda_{c,t} + \varepsilon_{p,c,t} \end{aligned} \tag{3}$$

where p, c, t denote a PIN Code, cohort (dataset) and year-month, respectively. $\gamma_{p,c}$ and $\lambda_{c,t}$ represents the PIN Code × Cohort and Cohort × Year-Month fixed effects. When controlling for Cohort × Year-Month fixed effects, we ensure that the coefficients are estimated by comparing treated PIN Codes solely to "clean" controls in their respective dataset. Appendix table A3 presents the results, which show that the brown-share-pollution gradient across specifications ranges from -3.80 to -1.85, which is broadly in line with our baseline findings.

Alternative Specifications: In Appendix Tables A4 and A5, we present a pair of robustness checks for the main results. In Appendix Tables A4 we consider alternative cutoffs to define the "treated' group as those within 15 or 10 or 5 kilometers of a monitoring station; the point estimates in our preferred within-station specification are virtually unchanged. Appendix Table A5 excludes PIN Codes in the largest metropolitan areas—as explained in Section 2, these PIN Codes offer a less straightforward delineation of treatment and control assignments. We observe substantially larger point estimates on the three-way interaction.

Additional Stock Characteristics: A potential concern in interpreting our results is a possible correlation between a stock's "greenness" and other company-level attributes. While it is not obvious ex ante why investor response to pollution should involve other attributes (and as such these variables could be seen as bad controls), we nonetheless consider the possibility that the changes in green investment that we document may reflect shifts in portfolios along other dimensions. To address this concern, we construct portfolio shares based on other stock traits, including size, age, and market-to-book, and estimate the impact of the pollution information program on the elasticity of portfolio shares with respect to pollution exposure. We adopt a similar approach to the one used in constructing our primary variable, "brown share" . For example, we first define value/growth stocks as those with a high/low book-to-market ratio (above/below the median of cross-sectional distribution at year t-1) and then compute the average share of value stocks held by retail investors in PIN Code p at time t. We report the coefficient β for the specification in Equation 1, using these portfolio shares as the dependent variable in place of brown share. The results are presented in Panels A, B, and C of Appendix Table A6, respectively. We do not observe in shifts in the sensitivity of these other portfolio characteristics to pollution after the introduction of CAAQMSs[18].

Preferences or Beliefs: The question naturally arises of whether portfolio adjustments are made as a result of beliefs about the returns of green versus brown stocks, or a distaste for holding brown stocks. While it is beyond the scope of our paper to evaluate whether tastes or beliefs drive investors' portfolio changes, we can at least evaluate whether, during the period that monitoring stations were opening, brown stocks underperformed in India or elsewhere. We show in Appendix Figure A5 the cumulative returns of brown and green portfolios over the period January 2000 to December 2019 (see Section 3 for details on the portfolios' construction). Over any relevant horizon, the two portfolios perform quite similarly, though with the brown portfolio actually generating higher returns. Given this pattern, if beliefs explained our main results, they would necessarily stem from from erroneous expectations of negative brown stock returns. To the extent that investors hold unbiased beliefs, our findings are more readily reconciled with taste-based explanations.

5.4 Response to High-Salience Changes in Air Quality

Thus far, we have posited that the increased salience of pollution via CAAQMSs is responsible for the increased sensitivity of investors' green portfolios to air quality. In this section, we explore a potentially more direct link to salience as an explanation, based on high-frequency changes around

[18] In unreported results, we confirm that there is little shift in the average stock characteristics for the portfolio held by local investors.

particularly salient shifts in air quality. To do so, we look at transitions across the color-coded Air Quality Index (AQI) categories used by the Central Pollution Control Board to communicate pollution severity to the public. We focus on the transition from Moderate (yellow) to Poor (amber) at an AQI cutoff of 200, which is particularly important in terms of government communication: "amber alerts" government messages via social media and elsewhere, warning that prolonged exposure to outside air may cause breathing discomfort and respiratory problems; the government also recommends the use of masks when the AQI is above 200[19].[20].

We study responsiveness around the yellow-amber transition using PIN Code-day observations in the range of 100 to 300 (i.e., 100 units above and below the cutoff). We begin by presenting visual evidence of this relationship in the left panel of Figure 7, which provides a regression discontinuity plot for local brown share, using distance to the "*yellow-amber*" threshold of 200 as the running variable and a local linear trend on each side of the discontinuity. Consistent with our previous findings, we observe a negative relation between local air quality and brown share, at the higher frequency of observation we use in this analysis. In addition, we observe a clear discontinuity at the yellow-amber cutoff, with a discernable drop in portfolio brown share at AQI = 200 (i.e., the level that triggers amber alert warnings).[21] In the right panel of Figure 7, we show that analogous pattern for control Pin Codes and observe no such discontinuous effect (nor do we a general negative relationship between pollution and brown share).

We provide the results of our RD analysis in Table 3, following the procedure outlined by Armstrong and Kolesár (2018) to select the optimal bandwidth and construct confidence intervals. Specifically, we begin with the rule-of-thumb choice of the smoothness constant, M_{rot}, and also confirm robustness using different choices. Across all specifications in Table 3, the coefficient of interest is

[19]For example, as part of the National Programme for Climate Change and Human Health (NPCCHH) initiative, the Indian government issue alerts to the general population to raise awareness about health concerns and risk reduction from exposure to pollution. https://ncdc.gov.in/WriteReadData/l892s/3065716611669017053.pdf

[20]Government messaging continues at higher AQIs, with more dire messages as the AQI transitions to red ("respiratory illness with long exposure"). We do not emphasize this transition in the main text for two reasons. First, we see a sharper distinction between yellow and amber in the sense that it causes a shift from no-warning to warning, as compared to simply a transition to a marginally more extreme warning. Second, pollution in the high-200s and above is arguably so severe that pollution becomes very salient irrespective of any color coding. We do provide results on the amber-to-red transition in the appendix material, and do not find any significant shift around this change. Finally, note that the Central Pollution Control Board of India uses six categories in total, with each level accompanied by different health advisories. AQI statistics and color categories are available at https://app.cpcbccr.com/AQI_India/

[21]We do not observe any change around the amber-red cutoff – see Appendix Figure A4.

negative and statistically significant. In economic terms, Panel A shows that portfolio brown share is 0.65 to 0.91 percentage points (1.66% to 2.22% in relative terms) lower when AQI is just above the amber-alert cutoff.[22]

Overall, we take our RD estimates as bolstering our salience-based interpretation of our main results, given the pronounced shift in brown share portfolio allocations above a threshold that triggers greater visibility of pollution's harmfulness.

5.5 Heterogeneity by investor type

We provide some exploratory analyses based on several attributes of traders that potentially relate to propensity to access local pollution information and/or concern for the environment, including age, gender, and whether investors primarily execute trades via mobile, internet, or more traditional means.

We naturally do not have random assignment or any close approximation to it for these attributes, and as such the results should be interpreted with caution. Still, the patterns are interesting to consider as an extension to our main results given that, based on intuition as well as past research, one may have prior expectations about which groups may have greater sensitivity to pollution salience.

Tech-Savvy Investors: We begin by comparing sensitivity of green investment to pollution based on whether an investor predominantly makes trades via (a) mobile, (b) internet, or (c) broker or other "traditional" method. As observed in the introduction, given that the public predominantly obtained AQI updates via smartphone, it is natural to speculate that more tech savvy investors—who use their mobile devices to execute trades—would be more exposed to pollution updates.

In Figure 8, panel A (the corresponding regression results are in Appendix Tables A8), we show event-study plots for responsiveness to air quality information disaggregated by investors' primary means of trading. As may readily be seen in the graph, the largest effect is observed among investors who use mobile devices to trade. The confidence intervals of the other two groups—internet and physical trading—are largely overlapping, though the effect of monitoring stations is marginally greater for internet-based traders.

[22] As already suggested by our graphical results, we do not find any significant effect when AQI transitions from amber to red, which we show in Appendix Table A7.

Investor's Demographic Profiles: We next turn to explore heterogeneity by age. Before turning to these results, it is worth noting that, while younger investors are more frequently classified as "mobile" traders, the differences are surprisingly modest relative to middle-aged investors: 13 percent of younger investors are classified as mobile, as compared to 10 percent of middle-aged investors (elderly investors are indeed far less likely to use mobile devices for trades—only 3.7 percent of the elderly in our sample are classified as mobile).

More importantly, putting aside technological concerns, there are reasons to expect differential responsiveness to pollution information by age. First, a vast literature documents a very strong negative correlation between environmental concerns and age—unsurprising, given that the young will disproportionately bear the consequences of climate change and environmental degradation.[23] Of more direct relevance, recent survey-based evidence finds higher stated interest in ESG amongst younger investors and lower interest amongst older investors (Haber et al., 2022). An alternative hypothesis is that the *elderly* might be more sensitive to information on air pollution, because they are far more vulnerable to the effects of air pollution (see Gouveia and Fletcher, 2000; Fischer et al., 2003).

In Figure 8, panel B, we revisit our main event plot, disaggregating the sample into young (18-29), middle-aged (30-55), and elderly (above 56) investors; the corresponding regression results are in Appendix Table A9. We observe a substantially greater shift in brown-investment-to-pollution sensitivity amongst the young, relative to the other two age groups. While we cannot put too strong an interpretation around this finding, we see this particular result as reinforcing the above-cited evidence on the age distribution of interest in ESG investing, which may be of direct practical relevance.

Finally, we split the sample by gender. As noted in the introduction, there is an ex ante rationale for a differential response given past work on a "gender environmentalism gap" (e.g., Xiao and McCright, 2015). We illustrate the differential response by gender in the event plots in Figure 8 panel C. The gender difference is striking—women exhibit a responsiveness that is 2-3 times greater than

[23] See, e.g., Liere and Dunlap (1980), for an early and well-cited review which describes age as the "predominant" individual attribute that is correlated with environmental concerns; a more recent review article by Sanchez-Sabate and Sabaté (2019) similarly finds an important role for age, focused specifically on environmental concerns and meat consumption.

that of men. Observe that the male point estimates are more precisely estimated, and also much closer to the full-sample estimates, which reflects the fact that most Indian retail investors are men. For completeness, we also provide the tabular version of the gender split in Appendix Table A10, which provides the same message as the event plots.

Our final "heterogeneity" test explores whether we see a comparable shift in responsiveness for institutional investors as observed in our sample of retail investors, which has been our focus to this point. We present these results in Appendix Figure A6. If anything, institutional investors' response to the availability of air quality information moves in the opposite direction to that of retail investors, though we cannot reject that institutions are simply unaffected by the appearance of real-time pollution data. There are various potential explanations for this non-result. The most natural is that institutional investors may be less sensitive to "taste-based" shifts in investing. However, we may also simply have a less-precise mapping of AQI to relevant location, since the PIN Code for an institutional investor reflects their place of work rather than residence. More broadly, we interpret this non-result with caution, relative to our main findings on retail investors.

6 Conclusion

We document that exposing investors to ready information about air pollution heightens the sensitivity of their "brown" investments to air quality. We interpret these findings through the lens of salience, in the spirit of Bordalo, Gennaioli and Shleifer (2013) among others — ready access to air quality data makes this information a more salient input into investment decisions.

As we noted in our discussion of the results, this shift comes despite the fact that returns for a long-short green-brown portfolio does not generate any excess returns—if investors adjust their portfolios in the expectation of higher returns, the shift is not justified by realized outcomes. That said, we cannot identify whether the shift we document is driven by mistaken beliefs, or shifts in investor tastes as a result of greater attentiveness to environmental problems. Distinguishing between these two explanations is one important future direction, and one that we plan to pursue going forward.

Finally, our findings highlight a role for salience in driving portfolio allocations. To the extent that it is desirable to shift investments away from brown firms (itself a debated point), it may be useful to implement policies that highlight environmental problems in ways that draw investors' attention in particular.

References

Agee, Mark D, Scott E Atkinson, Thomas D Crocker, and Jonathan W Williams. 2014. "Non-separable pollution control: Implications for a CO2 emissions cap and trade system." *Resource and Energy Economics*, 36(1): 64–82.

Alok, Shashwat, Nitin Kumar, and Russ Wermers. 2020. "Do fund managers misestimate climatic disaster risk." *The Review of Financial Studies*, 33(3): 1146–1183.

Armstrong, Timothy B, and Michal Kolesár. 2018. "Optimal inference in a class of regression models." *Econometrica*, 86(2): 655–683.

Baker, Andrew C, David F Larcker, and Charles CY Wang. 2022. "How much should we trust staggered difference-in-differences estimates?" *Journal of Financial Economics*, 144(2): 370–395.

Balasubramaniam, Vimal, John Y Campbell, Tarun Ramadorai, and Benjamin Ranish. 2023. "Who owns what? A factor model for direct stockholding." *The Journal of Finance*, 78(3): 1545–1591.

Baldauf, Markus, Lorenzo Garlappi, and Constantine Yannelis. 2020. "Does climate change affect real estate prices? Only if you believe in it." *The Review of Financial Studies*, 33(3): 1256–1295.

Bansal, Ravi, Dana Kiku, and Marcelo Ochoa. 2016. "Price of long-run temperature shifts in capital markets." National Bureau of Economic Research.

Barwick, Panle Jia, Shanjun Li, Liguo Lin, and Eric Zou. 2019. "From fog to smog: The value of pollution information." National Bureau of Economic Research.

Bau, Natalie, and Adrien Matray. 2023. "Misallocation and capital market integration: Evidence from India." *Econometrica*, 91(1): 67–106.

Ben-David, Itzhak, and David Hirshleifer. 2012. "Are investors really reluctant to realize their losses? Trading responses to past returns and the disposition effect." *The Review of Financial Studies*, 25(8): 2485–2532.

Berg, Florian, Julian F Koelbel, Anna Pavlova, and Roberto Rigobon. 2022. "ESG confusion and stock returns: Tackling the problem of noise." National Bureau of Economic Research.

Berry, Steven T. 1994. "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics*, 242–262.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and asset prices." *American Economic Review*, 103(3): 623–28.

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

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The effect of minimum wages on low-wage jobs." *The Quarterly Journal of Economics*, 134(3): 1405–1454.

Choi, Darwin, Zhenyu Gao, and Wenxi Jiang. 2020. "Attention to global warming." *The Review of Financial Studies*, 33(3): 1112–1145.

Cosemans, Mathijs, and Rik Frehen. 2021. "Salience theory and stock prices: Empirical evidence." *Journal of Financial Economics*, 140(2): 460–483.

Fischer, P, G Hoek, B Brunekreef, A Verhoeff, and J Van Wijnen. 2003. "Air pollution and mortality in The Netherlands: are the elderly more at risk?" *European respiratory journal*, 21(40 suppl): 34s–38s.

Frydman, Cary, Samuel M Hartzmark, and David H Solomon. 2018. "Rolling mental accounts." *The Review of Financial Studies*, 31(1): 362–397.

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova. 2010. "Imported intermediate inputs and domestic product growth: Evidence from India." *The Quarterly journal of economics*, 125(4): 1727–1767.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277.

Görgen, Maximilian, Andrea Jacob, Martin Nerlinger, Ryan Riordan, Martin Rohleder, and Marco Wilkens. 2020. "Carbon risk." *Available at SSRN 2930897*.

Gouveia, Nelson, and Tony Fletcher. 2000. "Time series analysis of air pollution and mortality: effects by cause, age and socioeconomic status." *Journal of Epidemiology & Community Health*, 54(10): 750–755.

Graff Zivin, Joshua, and Matthew Neidell. 2013. "Environment, health, and human capital." *Journal of economic literature*, 51(3): 689–730.

Greenstone, Michael, Guojun He, Ruixue Jia, and Tong Liu. 2022. "Can technology solve the principal-agent problem? Evidence from China's war on air pollution." *American Economic Review: Insights*, 4(1): 54–70.

Gulia, Sunil, Nidhi Shukla, Lavanya Padhi, Parthaa Bosu, SK Goyal, and Rakesh Kumar. 2022. "Evolution of air pollution management policies and related research in India." *Environmental Challenges*, 6: 100431.

Haber, Stephen, John D Kepler, David F Larcker, Amit Seru, and Brian Tayan. 2022. "ESG Investing: What Shareholders Do Fund Managers Represent?" *Rock Center for Corporate Governance at Stanford University Working Paper*.

Hong, Harrison, and Marcin Kacperczyk. 2009. "The price of sin: The effects of social norms on markets." *Journal of financial economics*, 93(1): 15–36.

Hong, Harrison, Frank Weikai Li, and Jiangmin Xu. 2019. "Climate risks and market efficiency." *Journal of econometrics*, 208(1): 265–281.

Hong, Harrison, G Andrew Karolyi, and José A Scheinkman. 2020. "Climate finance." *The Review of Financial Studies*, 33(3): 1011–1023.

Huberman, Gur, and Tomer Regev. 2001. "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar." *The Journal of Finance*, 56(1): 387–396.

Jiang, Han, Le Lexi Kang, Ziye Zoe Nie, and Hui Zhou. 2022. "Can Old Sin Make New Shame? Stock Market Reactions to the Release of Movies Re-Exposing Past Corporate Scandals." *Available at SSRN 884269*.

Khanna, Tarun, and Krishna Palepu. 2000. "Is group affiliation profitable in emerging markets? An analysis of diversified Indian business groups." *The journal of finance*, 55(2): 867–891.

Krey, Volker, O Masera, G Blanford, T Bruckner, R Cooke, K Fisher-Vanden, H Haberl, E Hertwich, E Kriegler, D Mueller, et al. 2014. "Annex 2-metrics and methodology."

Liere, Kent D Van, and Riley E Dunlap. 1980. "The social bases of environmental concern: A review of hypotheses, explanations and empirical evidence." *Public opinion quarterly*, 44(2): 181–197.

Lin, Xiaohui, Ruqi Yang, Wen Zhang, Ning Zeng, Yu Zhao, Guocheng Wang, Tingting Li, and Qixiang Cai. 2023. "An integrated view of correlated emissions of greenhouse gases and air pollutants in China." *Carbon Balance and Management*, 18(1): 9.

Li, Ye, Eric J Johnson, and Lisa Zaval. 2011. "Local warming: Daily temperature change influences belief in global warming." *Psychological science*, 22(4): 454–459.

Lujala, Päivi, Haakon Lein, and Jan Ketil Rød. 2015. "Climate change, natural hazards, and risk perception: the role of proximity and personal experience." *Local Environment*, 20(4): 489–509.

Moss, Austin, James P Naughton, and Clare Wang. 2023. "The Irrelevance of Environmental, Social, and Governance Disclosure to Retail Investors." *Management Science*.

Murfin, Justin, and Matthew Spiegel. 2020. "Is the risk of sea level rise capitalized in residential real estate?" *The Review of Financial Studies*, 33(3): 1217–1255.

Pant, Pallavi, Raj M Lal, Sarath K Guttikunda, Armistead G Russell, Ajay S Nagpure, Anu Ramaswami, and Richard E Peltier. 2019. "Monitoring particulate matter in India: recent trends and future outlook." *Air Quality, Atmosphere & Health*, 12(1): 45–58.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski. 2021. "Responsible investing: The ESG-efficient frontier." *Journal of Financial Economics*, 142(2): 572–597.

Pelizzon, Loriana, Aleksandra Rzeznik, and Kathleen Weiss Hanley. 2021. "The salience of ESG ratings for stock pricing: Evidence from (potentially) confused investors."

Rees, Nicholas, Amy Wickham, and Lawrence Chandy. 2019. "Silent Suffocation in Africa." *UNICEF Report*.

Roychowdhury, Anumita, Avikal Somvanshi, and Sharanjeet Kaur. 2023. "Status of air quality monitoring in India: Spatial spread, population coverage and data completeness." *Urban Lab - Centre for Science and Environment Analysis*.

Sanchez-Sabate, Ruben, and Joan Sabaté. 2019. "Consumer attitudes towards environmental concerns of meat consumption: A systematic review." *International journal of environmental research and public health*, 16(7): 1220.

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

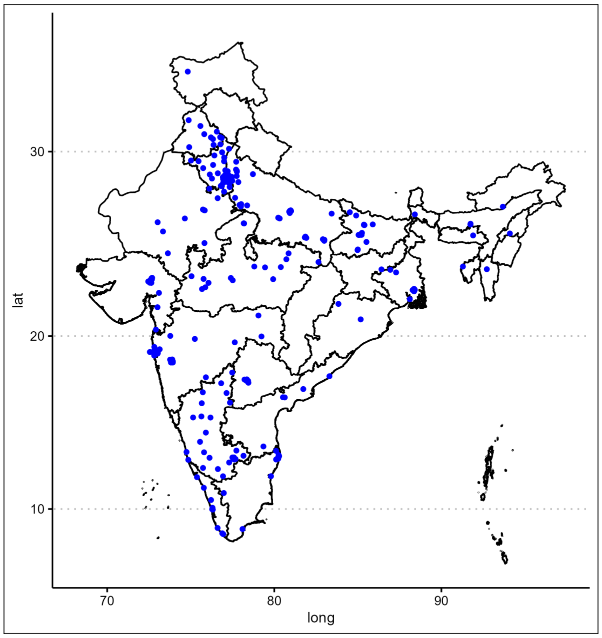
Tsai, Tzu-Chin, Yung-Jyh Jeng, D Allen Chu, Jen-Ping Chen, and Shuenn-Chin Chang. 2011. "Analysis of the relationship between MODIS aerosol optical depth and particulate matter from 2006 to 2008." *Atmospheric Environment*, 45(27): 4777–4788.

Van Donkelaar, Aaron, Randall V Martin, and Rokjin J Park. 2006. "Estimating ground-level PM2. 5 using aerosol optical depth determined from satellite remote sensing." *Journal of Geophysical Research: Atmospheres*, 111(D21).

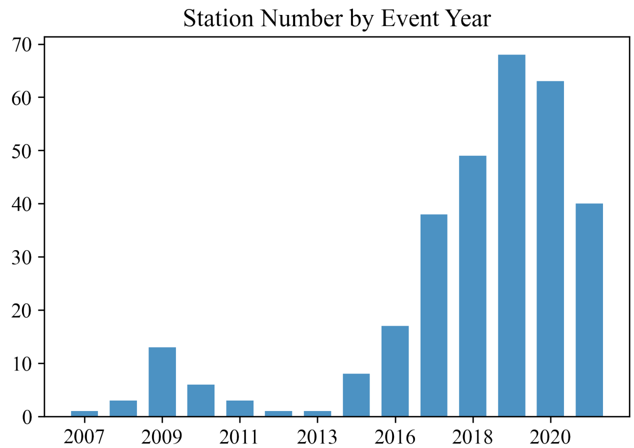
Xiao, Chenyang, and Aaron M McCright. 2015. "Gender differences in environmental concern: Revisiting the institutional trust hypothesis in the USA." *Environment and Behavior*, 47(1): 17–37.

Zaval, Lisa, Elizabeth A Keenan, Eric J Johnson, and Elke U Weber. 2014. "How warm days increase belief in global warming." *Nature Climate Change*, 4(2): 143–147.

Figure 1: Geographic Distribution and Rollout of Continuous Ambient Air Quality Monitoring Stations



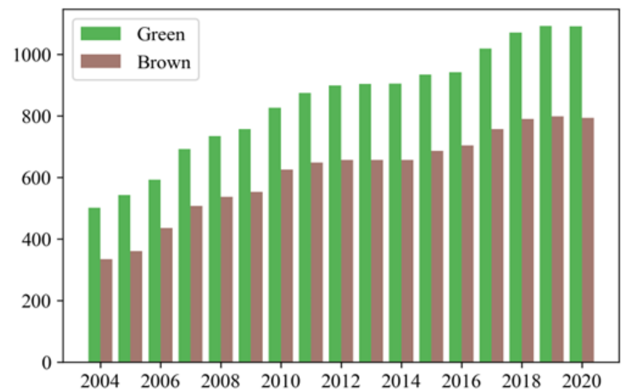
(a) Geographic Distribution of Continuous Ambient Air Quality Monitoring Stations



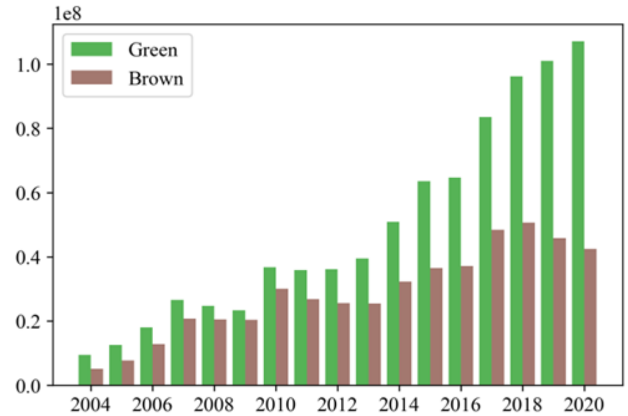
(b) Timing of Rollout of Continuous Ambient Air Quality Monitoring Stations

Note: Panel A plots the geographic locations of the Continuous Ambient Air Quality Monitoring Stations(CAAQMSs) across India. Panel B shows the number of CAAQMSs introduced each year.

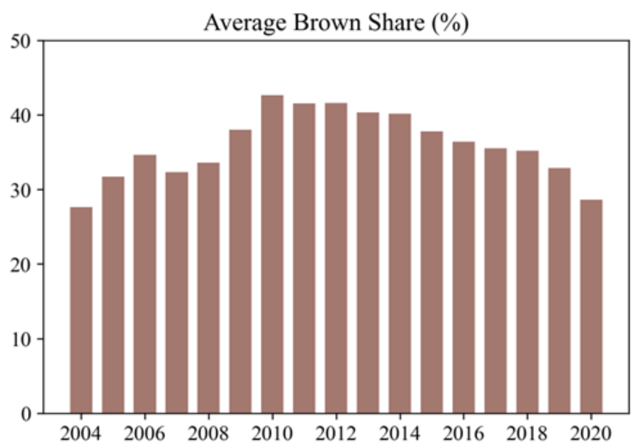
Figure 2: Distribution of Green and Brown Stocks over Time



(a) Brown versus Green Stocks – number of stocks



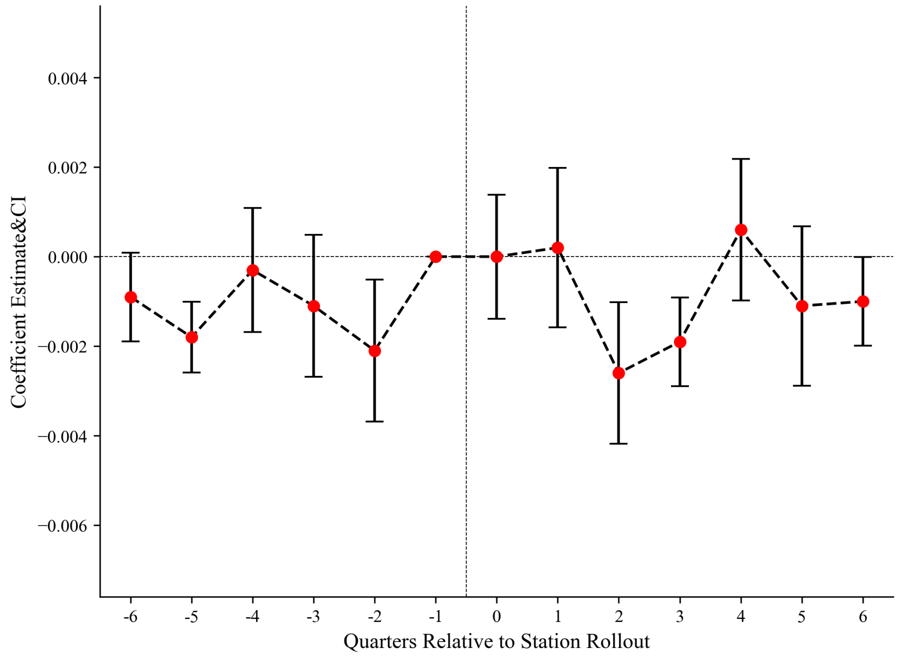
(b) Brown versus Green Stock – market capitalization of stocks



(c) Share of Brown Stocks

Note: This figure plots the distribution of green and brown stock holding over the years in the market. Panel A presents the number of green and brown stocks over time, while Panel B presents the market capitalization of green vs brown stocks. Panel C presents the market share of brown stocks (in terms of market capitalization) among all publicly traded equities in India.

Figure 3: Pollution Levels Around CAAQMS Installation

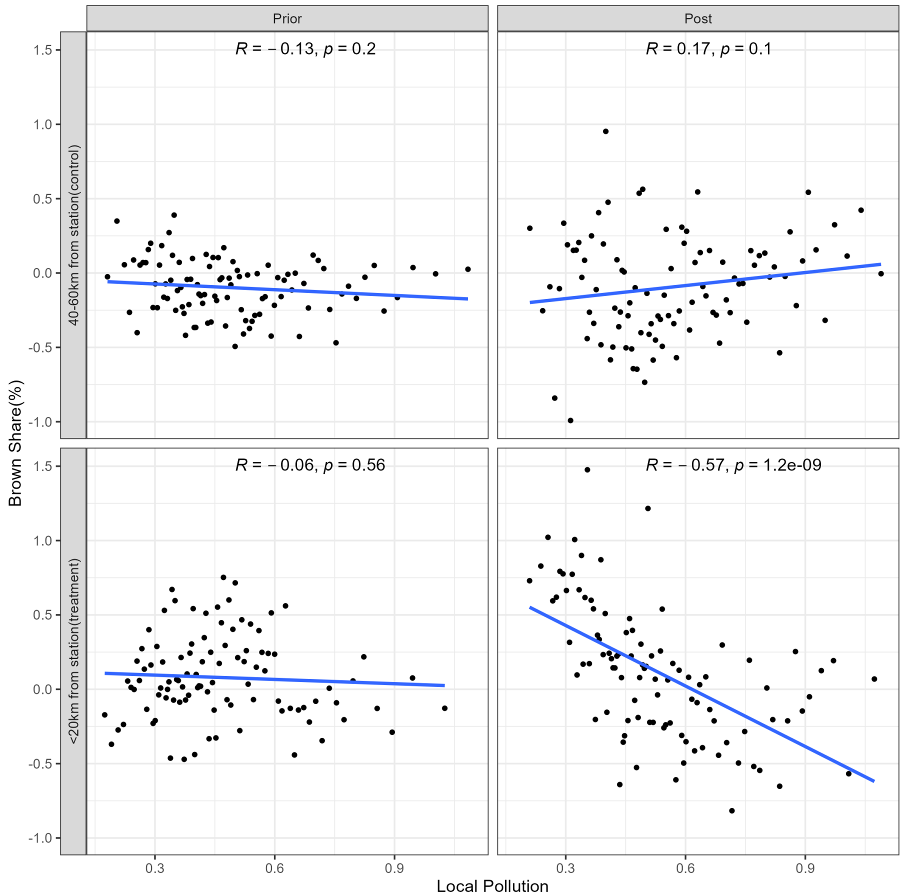


Note: This figure plots the trend in pollution following the installation of Continuous Ambient Air Quality Monitoring Stations. Specifically we plot the coefficient $\{\beta_i\}$ from the following specification:

$$Pollution_{p(m),t} = \sum_{i=-6}^{6} \beta_i \times Treat_p \times 1(t=i) + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t}$$

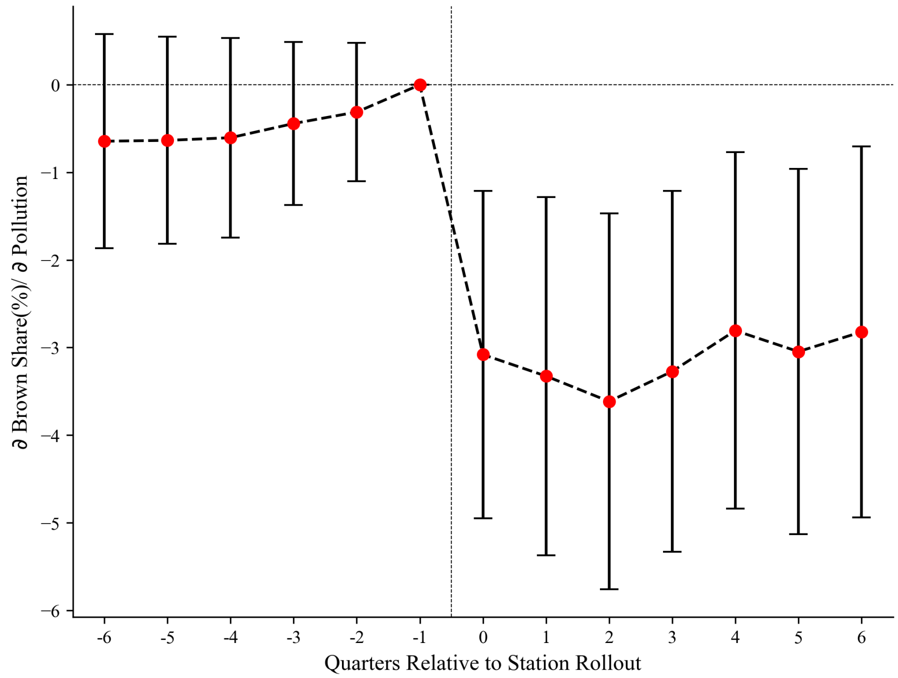
where $Pollution_{p(m),t}$ denotes the average pollution in PIN Code p belonging to a monitoring station area m at time t. A station area m includes treated PIN Codes (those within 20 kilometers of the station) and also control PIN Codes (40-60 kilometers from the station). $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trading activities(no. of retail investors, total turnover by retail investors) and weather conditions(rainfall, temperature). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area $\times$ time fixed effects. The graph shows the estimated coefficients $\{\beta_i\}$ as well as 95% confidence intervals obtained from standard errors clustered at the PIN Code level.

Figure 4: Correlation of Brown Share and Pollution – Control versus Treated PIN Codes



Note: This figure presents scatter plots relating the share of brown stocks and local pollution for both the control and treated groups before and after the installation of Continuous Ambient Air Quality Monitoring Stations. We categorize pollution into 100 buckets and plot the average pollution against the average brown share holdings. The brown share is adjusted for the time-varying mean in a station area for both the control and treatment groups. The top two panels represent the control PIN Codes, while the bottom two panels represent the treated PIN Codes. The left two panels show the relationship prior to the installation of CAAQMS, while the right two panels show the relationship after the installation.

Figure 5: Sensitivity of Brown Share to Pollution

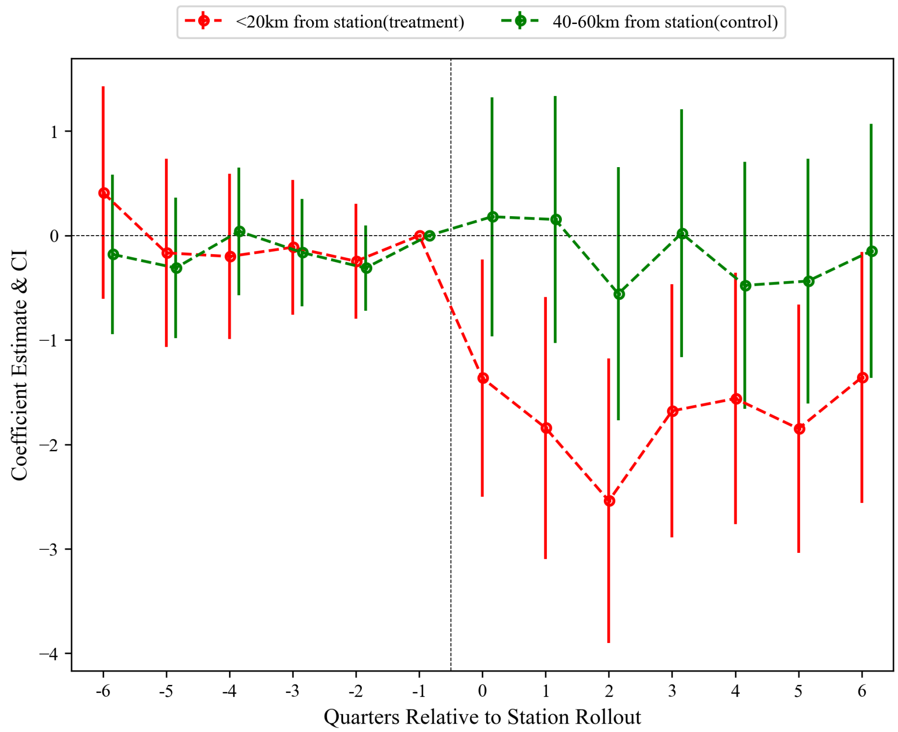


Note: This figure plots the sensitivity of brown investment to local pollution before and after the installation of Continuous Ambient Air Quality Monitoring Stations. Specifically we plot the coefficient $\{\beta_i\}$ from the following specification:

$$\begin{aligned} Brown\,Share_{p(m),t} \quad &= \quad \sum_{i=-6}^{6} \beta_i \times Treat_p \times Pollution_{p,t} \times 1(t \in i)_{m,t} + \sum_k \alpha_k \times Other\ Interactions \\ &\quad + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t} \end{aligned}$$

where $Brown\,Share_{p(m),t}$ denotes the average share of brown stocks held by retail investors in PIN Code p belonging to a station area m at time t. A station area m includes both treated (those within 20 kilometers of the station) and control PIN Codes (40-60 kilometers from the station). $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather characteristics (see text for details). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area $\times$ time fixed effects. The graph shows the estimated coefficients $\{\beta_i\}$ as well as 95% confidence intervals obtained from standard errors clustered at the PIN Code level.

Figure 6: Sensitivity of Brown Share to Pollution – Control versus Treated

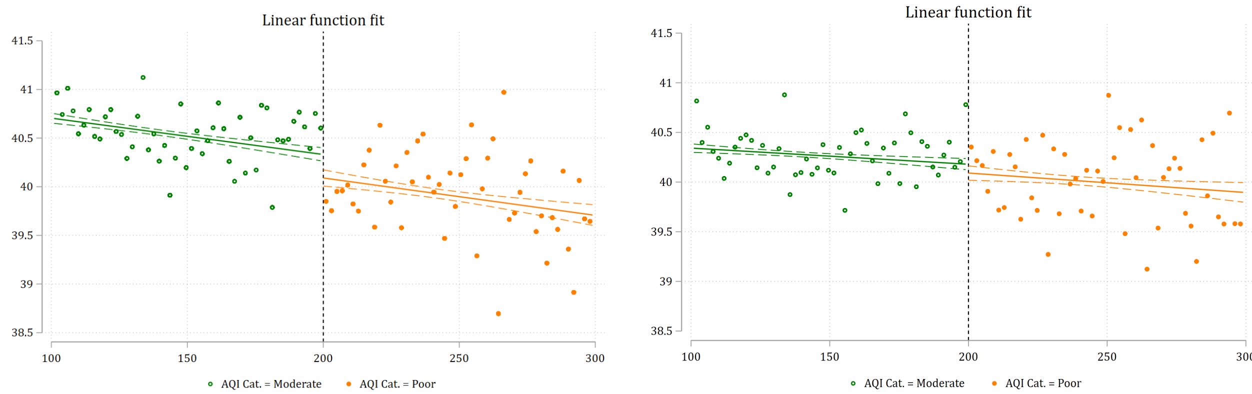


Note: This figure plots the sensitivity of brown investment to pollution before and after the installation of Continuous Ambient Air Quality Monitoring Stations separately for the control and treatment groups. Specifically, we plot the coefficient $\{\beta_i\}$ from the following specification separately for the control and treatment groups:

$$Brown\,Share_{p(m),t} = \sum_{i=-6}^{6} \beta_i \cdot Pollution_{p,t} \times 1(t \in i)_{m,t} + \sum_{k} \alpha_k \times Other\ Interactions + X'_{p,t}\theta + \gamma_p + \lambda_t + \varepsilon_{p,t}$$

where $Brown\,Share_{p(m),t}$ denotes the average share of brown stocks of retail investors in PIN Code p belonging to a station area m at time t. A station area m includes treated PIN Codes (those within 20 kilometers of the station) and also control PIN Codes (40-60 kilometers from the station). $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather (see text for details). γ_p is a set of PIN Code fixed effects and λ_t is time fixed effects. The graph shows the estimated coefficients $\{\beta_i\}$ as well as 95% confidence intervals obtained from standard errors clustered at the PIN Code level.

Figure 7: Regression Discontinuity Plots for Brown Share versus Air Quality Readings

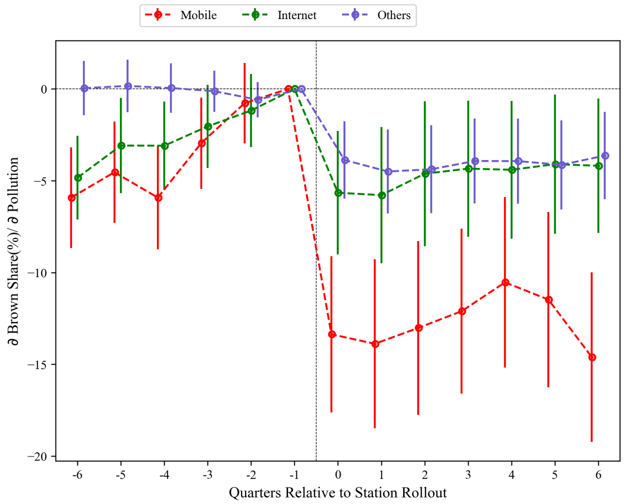


(a) Treated PINCodes

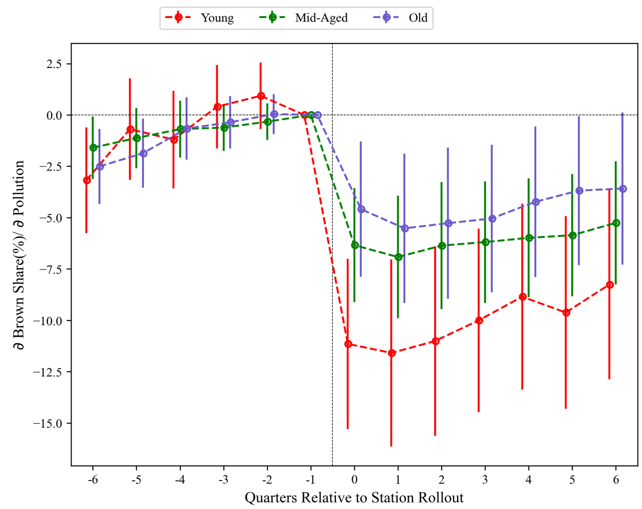
(b) Control PINCodes

Note: The figure presents RD plots for PIN-Code-day level portfolio brown shares among treated (left panel) and control(right panel) investors and local AQI readings. We plot the mean value of portfolio brown shares, binned by AQI, as well as the fitted local linear trend (along with 95% confidence intervals) around the critical AQI cutoffs, 200, the cutoff between “Moderate” (yellow) and “Poor” (amber) pollution.

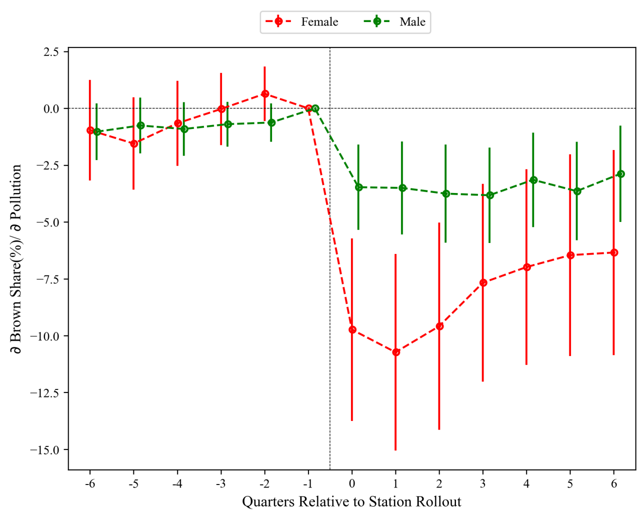
Figure 8: Changes in Brown-Share- Pollution Gradient: Hetergeneity



(a) Heterogeneity by trading technology



(b) Heterogeneity by age



(c) Heterogeneity by gender

Note: This figure shows the sensitivity of brown investment to pollution before and after the installation of Continuous Ambient Air Quality Monitoring Stations for different groups of investors. Specifically we plot the coefficient $\{\beta_i\}$ from the following specification in Table 2, splitting the investor sample by various attributes. Panel A shows heterogeneity by trading technology, Panel B shows heterogeneity by age, and Panel C shows heterogeneity by gender. In each case, the graph shows the estimated coefficients $\{\beta_i\}$ as well as 95% confidence intervals obtained from standard errors clustered at the PIN Code level.

Table 1: Summary Statistics

	mean	sd	min	p5	p10	p25	p50	p75	p90	p95	max
Average Brown Share Holding%	41.876	9.343	0.000	28.040	32.590	37.670	41.840	45.840	51.190	56.010	100.000
Average Brown Share Holding Female%	41.194	19.085	0.000	4.280	18.080	32.980	41.160	48.400	61.580	76.400	100.000
Average Brown Share Holding Male%	41.935	9.644	0.000	27.640	32.340	37.600	41.900	46.010	51.540	56.580	100.000
Average Brown Share Holding Young%	40.414	17.615	0.000	9.500	20.720	32.030	39.970	47.790	59.270	70.760	100.000
Average Brown Share Holding MidAged%	41.943	11.481	0.000	24.440	30.750	37.180	41.910	46.490	53.100	59.250	100.000
Average Brown Share Holding Old%	42.839	16.066	0.000	15.100	26.390	36.540	42.490	48.450	59.390	70.290	100.000
Pollution AOD	0.543	0.217	0.031	0.264	0.305	0.387	0.505	0.656	0.843	0.954	1.880
No. of Investors(Log)	3.861	1.805	0.693	1.099	1.609	2.485	3.638	4.970	6.541	7.288	10.384
Turnover(Log)	16.846	2.891	0.140	11.407	12.925	15.256	17.195	18.801	20.265	21.028	24.848
Rain	3.171	5.322	0.000	0.000	0.000	0.050	0.940	4.170	9.020	13.140	109.590
Temperature	25.764	4.916	3.290	15.750	18.280	23.410	26.620	29.270	31.230	32.500	36.510

Note: This table reports the summary statistics of the key variable used for the empirical analysis in this paper. The summary statistics are for the universe of Indian retail investors who trade on the National Stock Exchange, aggregated at the PIN-Code-month level. We report the average share of brown stocks for retail investors overall, and also disaggregated by gender and age cohort.

Table 2: The Impact of Air Quality Information via CAAQMS on the Brown-Share-Pollution Gradient

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Var. = Brown Share(%)							
Pollution×Treated×Post	-1.5319**	-2.1017**	-1.7251*	-2.5007***	-1.3634*	-2.0087**	-1.5687	-2.4050**
	(0.7715)	(0.8397)	(0.9762)	(0.9617)	(0.7675)	(0.8365)	(0.9739)	(0.9607)
Pollution×Treated	0.7576*	1.1183***	0.9427*	1.2966***	0.6620*	1.0331**	0.8627*	1.2234**
	(0.3914)	(0.4283)	(0.4958)	(0.4835)	(0.3893)	(0.4270)	(0.4945)	(0.4826)
Pollution×Post	0.2348	0.8557	0.7933	2.0042	-0.0026	0.7614	0.7592	1.8593
	(0.5419)	(0.6750)	(0.6974)	(1.6520)	(0.5387)	(0.6729)	(0.6951)	(1.6451)
Treat×Post	0.9842***	0.6365**	0.5494	0.6440**	0.5404**	0.3639	0.3871	0.3974
	(0.2711)	(0.2733)	(0.3393)	(0.3027)	(0.2670)	(0.2704)	(0.3367)	(0.2995)
Pollution	-0.2475	-0.6442*	-0.4243	-1.0167	-0.1799	-0.5311	-0.4084	-0.9371
	(0.2557)	(0.3894)	(0.7409)	(0.7642)	(0.2530)	(0.3883)	(0.7372)	(0.7607)
Post	-0.2702	-0.0112	-0.0175		-0.1455	0.0196	-0.0172	
	(0.1658)	(0.1946)	(0.2513)		(0.1645)	(0.1942)	(0.2509)	
No. of Investors(Log)					-2.1658***	-1.6583***	-1.3128***	-1.5185***
					(0.1905)	(0.1987)	(0.2034)	(0.2034)
Turnover(Log)					-0.0971**	-0.1069***	-0.0955**	-0.1023**
					(0.0406)	(0.0407)	(0.0413)	(0.0409)
Rainfall					-0.0024	-0.0079	-0.0011	-0.0114
					(0.0040)	(0.0054)	(0.0084)	(0.0101)
Temperature					0.0100	-0.0561*	-0.1447**	-0.0566
					(0.0066)	(0.0312)	(0.0572)	(0.0734)
Pincode	Y	Y	Y	Y	Y	Y	Y	Y
Year-month	Y				Y			
State×Year-month		Y				Y		
District×Year-Month			Y				Y	
Station×Year-month				Y				Y
Observations	499,036	499,036	488,681	498,310	499,036	499,036	488,681	498,310
R-squared	0.537	0.546	0.599	0.560	0.540	0.548	0.600	0.561

Note: This table studies the elasticity of the share of brown stocks held in investors' portfolios with respect to information on pollution following the installation of Continuous Ambient Air Quality Monitoring Stations. Specifically we report the coefficient β from the following specification:

$$Brown\,Share_{p(m),t} = \beta \times Treat_p \times Pollution_{p,t} \times Post_{m,t} + \sum_k \alpha_k \times Other\ Interactions + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t}$$

where $Brown\,Share_{p(m),t}$ denotes the average share of brown stocks of retail investors in PIN Code p belonging to a station area m at time t. A station area m includes treated PIN Codes (those within 20 kilometers of the station) and also control PIN Codes (40-60 kilometers from the station. $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group.The binary variable $Post_{m,t}$ represents the information shock and is equal to one after the installation of a local monitoring station. We demean the pollution variable to enhance the interpretability of the coefficient. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather characteristics (see text for details). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area × time fixed effects. Columns 1-4 report results without controls $X_{p,t}$ while columns 5-8 report results with these controls. Standard errors are clustered at the PIN Code level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

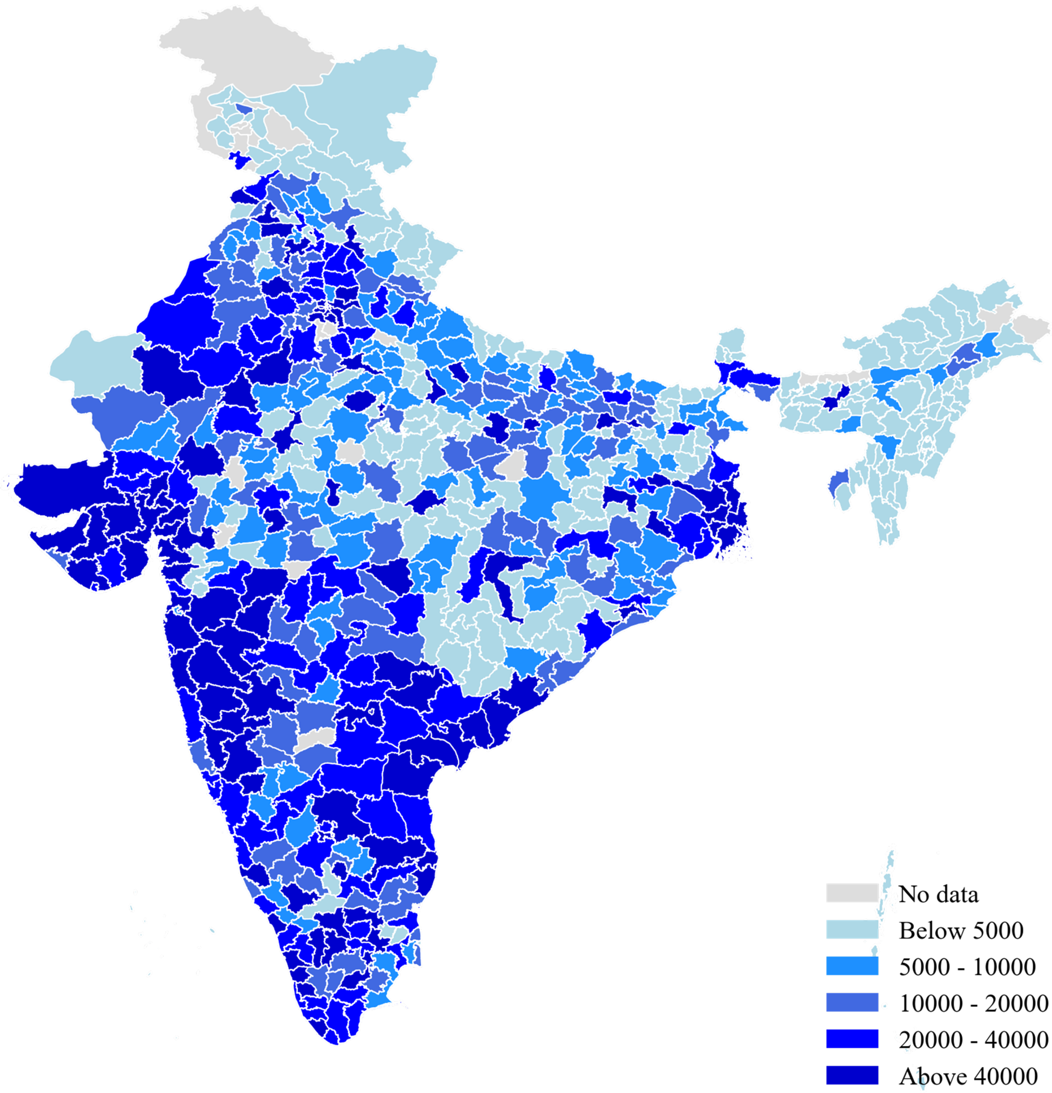
Table 3: Regression Discontinuity Estimates: AQI Transition from "Moderate" to "Poor"

(1)	(2)	(3)	(4)	(5)	(6)
smoothness constant	M	band-width	estimate	Lower CI	Upper CI
Panel A Treated PIN Codes					
Kernel = Uniform					
0.1*M_rot	0.0001	41	-0.65	-0.8297	-0.4703
0.5*M_rot	0.0007	21	-0.9145	-1.1638	-0.6652
M_rot	0.0014	16	-0.8878	-1.1747	-0.6009
5*M_rot	0.0071	8	-0.6721	-1.0825	-0.2616
10*M_rot	0.0143	6	-0.9334	-1.4149	-0.4519
Kernel = Triangle					
0.1*M_rot	0.0001	52.36	-0.6813	-0.8565	-0.5061
0.5*M_rot	0.0007	27.85	-0.8361	-1.0793	-0.593
M_rot	0.0014	20.92	-0.8513	-1.1321	-0.5705
5*M_rot	0.0071	11.20	-0.7361	-1.1386	-0.3337
10*M_rot	0.0143	8.38	-0.7461	-1.2204	-0.2718
Panel B Control PIN Codes					
Kernel = Uniform					
0.1*M_rot	0.0001	40	-0.117	-0.2712	0.0372
0.5*M_rot	0.0007	21	-0.1156	-0.3298	0.0986
M_rot	0.0013	15	-0.0892	-0.3372	0.1588
5*M_rot	0.0067	8	-0.0001	-0.3511	0.3509
10*M_rot	0.0134	6	-0.167	-0.5776	0.2436
Kernel = Triangle					
0.1*M_rot	0.0001	50.64	-0.1372	-0.2874	0.013
0.5*M_rot	0.0007	26.94	-0.0812	-0.2898	0.1275
M_rot	0.0013	20.24	-0.0999	-0.3412	0.1413
5*M_rot	0.0067	10.83	-0.1169	-0.4601	0.2263
10*M_rot	0.0134	8.16	-0.1439	-0.5453	0.2575

Note: The values reported here are RD estimates around the AQI cutoff of 200 at which the air quality index category switch from “Moderate”(yellow) to “Poor”(amber) Panel A and B presents the esimstate for the treated and control PINCodes, respectively. We follow the inference procedure of Armstrong and Kolesár (2018) to select the optimal bandwidth and to construct the confidence interval. M_{rot} indicates the rule-of-thumb choice of the smoothness constant, M. We also vary M as form of sensitivity analysis. The unit of observation is PIN-Code-day.

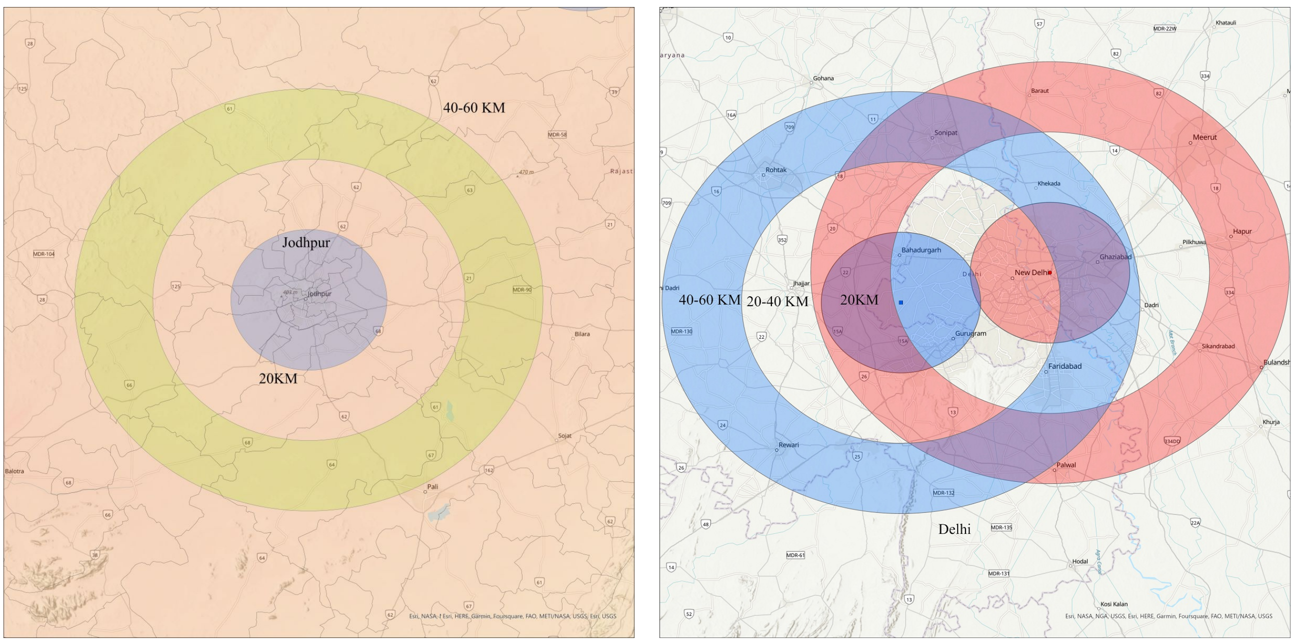
Internet Appendix

Figure A1: **Geography of NSE investors**



Note: This figure shows the geographic distribution of retail investors across districts who trade on the National Stock Exchange from January 2004 to June 2020.

Figure A2: **Illustrative Cases of Treatment and Control Assignment**

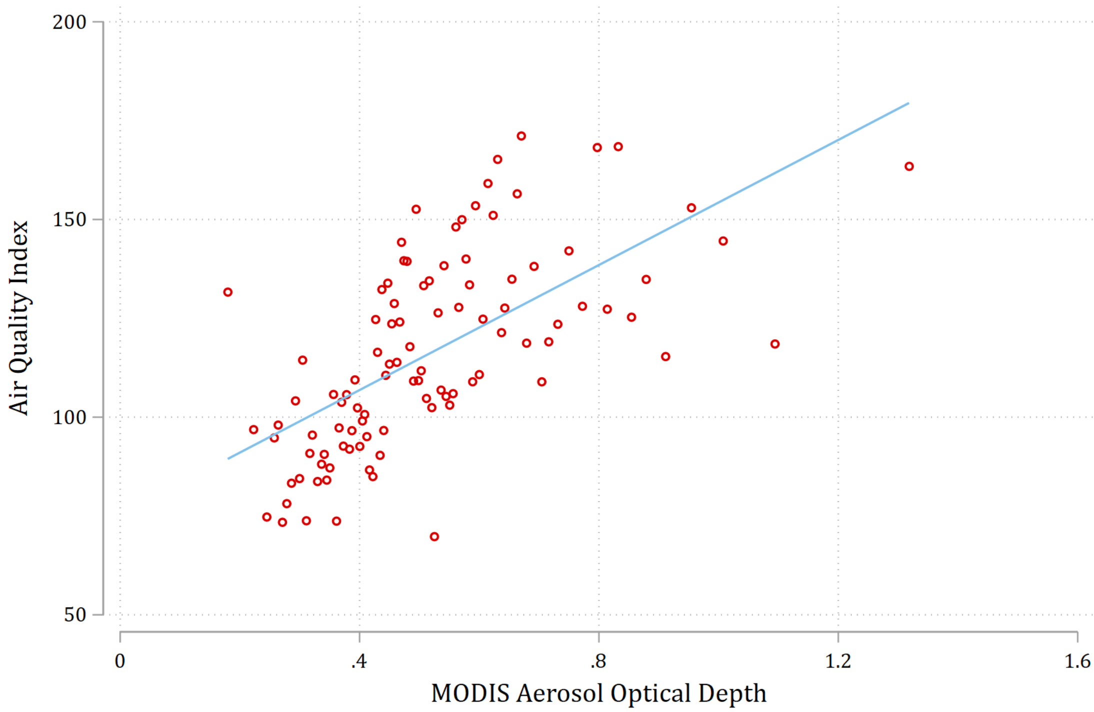


(a) Single-station Example

(b) Two-station Example

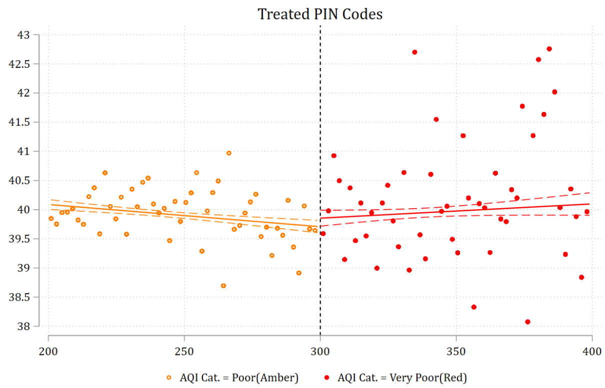
Note: This figure illustrates our treatment and control assignments in the ‘simple’ case of Jodhpur (which has a single monitoring station) and the more complicated case of two overlapping stations in Delhi. PIN Codes within the inner 20 kilometer circles are “treated”, while those between 40 and 60 kilometers are “control” units. In the case of overlapping treatment regions, treatment assignment is to the first monitoring station installed (the right one in this case); we use a similar approach for overlapping control areas. If a treatment and control areas overlap, assignment is to the treatment area.

Figure A3: **Correlation between AQI and AOD**

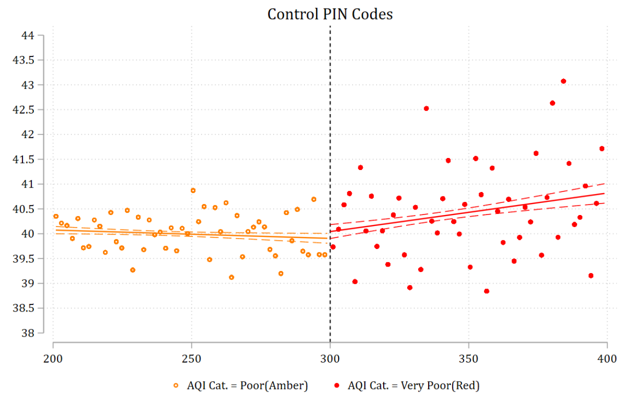


Note: This figure plots the average air quality index(AQI) by 100 equal bins of Aerosol Optical Depth (AOD) after the installation of monitoring stations (when both measures are available).

Figure A4: RD Plots for Brown Share versus AQI: Transition from Amber to Red



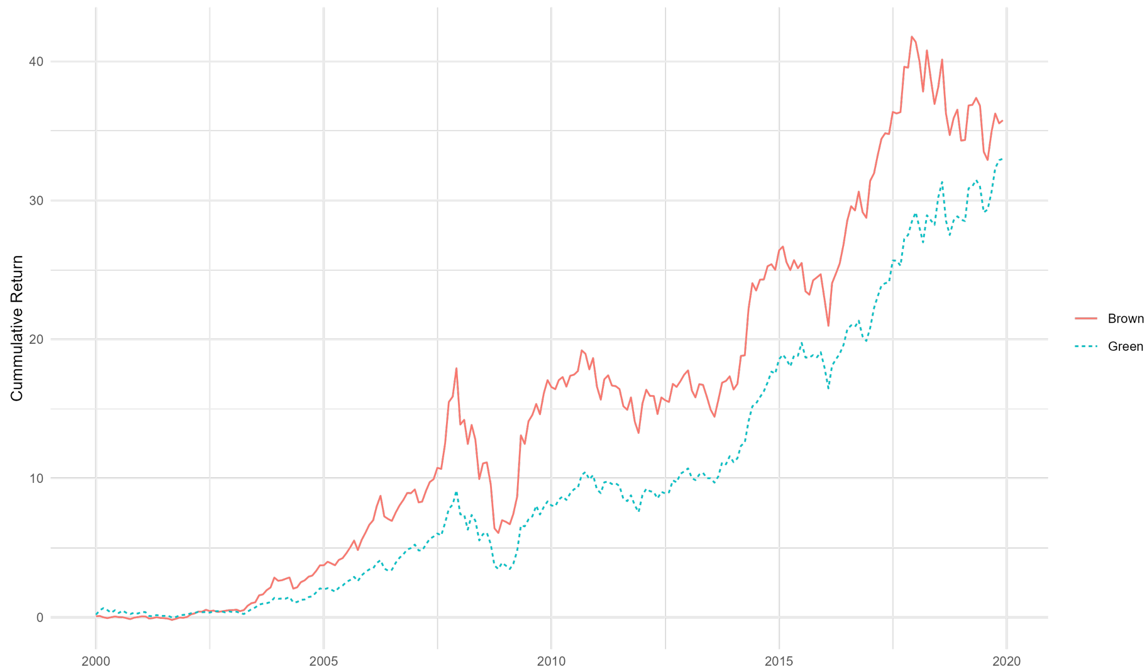
(a) Treated PINCodes



(b) Control PINCodes

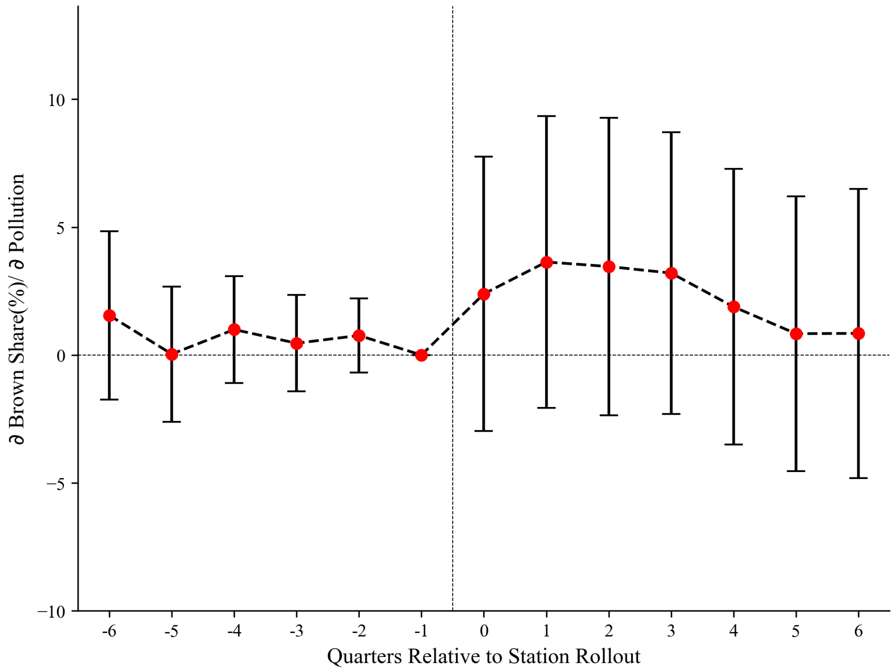
Note: The figure presents RD plots for PIN-Code-day level portfolio brown shares among treated (left panel) and control(right panel) investors and local AQI readings. We plot the mean value of portfolio brown shares, binned by AQI, as well as the fitted local linear trend (along with 95% confidence intervals) around the critical AQI cutoffs, 300, the cutoff between “Poor” (amber) and “Very Poor” (red) pollution.

Figure A5: **Returns on value-weighted green and brown portfolios**



Note: This figure plots the green and brown portfolios' cumulative returns over January 2000 to December 2019.

Figure A6: Changes in Brown-Share-Pollution Gradient: Institutional Investors



Note: This figure plots the sensitivity of trading in brown stocks by institutional investors to pollution before and after the installation of Continuous Ambient Air Quality Monitoring Stations. We plot the coefficient $\{\beta_i\}$ from the following specification:

$$\begin{aligned} Brown\,Share_{p(m),t} &= \sum_{i=-6}^{6} \beta_i \times Treat_p \times Pollution_{p,t} \times D(t \in i)_{m,t} + \sum_k \alpha_k \times Other\ Interactions \\ &\quad + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t} \end{aligned}$$

where $Brown\,Share_{p(m),t}$ denotes the average share of brown stocks of institutional investors in PIN Code p belonging to a station area m at time t. A station area m includes treated PIN Codes (those within 20 kilometers of the station) and also control PIN Codes (40-60 kilometers from the station). $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather characteristics (see text for details). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area × time fixed effects. The graph shows the estimated coefficients $\{\beta_i\}$ as well as 95% confidence intervals obtained from standard errors clustered at the PIN Code level.

Table A1: The Impact of Monitoring Stations on Local Pollution

	(1)	(2)	(3)	(4)	(5)	(6)
	Treated		Control		Full Sample	
	Pollution	Log(Pollution)	Pollution	Log(Pollution)	Pollution	Log(Pollution)
Post	0.0004	0.0004	0.0063	0.0036		
	(0.0056)	(0.0034)	(0.0050)	(0.0029)		
Treat×Post					0.0002	0.0000
					(0.0003)	(0.0002)
Station FE	Y	Y	Y	Y		
PIN Code FE					Y	Y
Year-month FE	Y	Y	Y	Y		
Station×Year-month					Y	Y
Observations	11,489	11,489	14,801	14,801	1,237,258	1,237,258
R-squared	0.619	0.636	0.660	0.686	0.968	0.971

Note: This table presents the results of balance tests for local pollution levels before and after the opening of a Continuous Air Quality Monitoring Station. The outcome variable in Columns 1-4 is the average pollution measured by Aerosol Optical Depth, at the monitoring station on a monthly basis. Columns 5-6 examine the impact on average pollution based on the PIN Code-month pair. Standard errors are clustered at the station level for the first set of analysis and at the PIN Code level for the latter set. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A2: Changes in Local Economic Condition and local investor composition Before and After CAAQMS Rollout

	(1)	(2)	(3)	(4)
Panel A Local Economic Condition				
	Firm Entry		NightLight Density	
Treat×Post	-0.0208	-0.0399	-0.0943	0.1823
	(0.0420)	(0.0484)	(0.1462)	(0.1416)
Post	0.0983***		1.2520***	
	(0.0346)		(0.0851)	
Observations	32,278	31,904	100,429	100,222
R-squared	0.891	0.908	0.978	0.988
Panel B local investor composition				
	Investor Growth Rate(%)		Number of New Traders	
Treat×Post	-0.0174	0.0043	-0.0016	-0.0028
	(0.0398)	(0.0427)	(0.0095)	(0.0099)
Post	-0.1200***		0.0150***	
	(0.0292)		(0.0036)	
Observations	498,624	497,898	499,036	498,310
R-squared	0.129	0.165	0.954	0.957
PIN Code	Y	Y	Y	Y
Year-month	Y		Y	
Station×Year-month		Y		Y

Note: This table presents the balance tests on (a) local economic condition and (b) local investor composition, before and after the opening of a Continuous Air Quality Monitoring Station. Specifically we report the coefficient β from the following specification:

$$Y_{p(m),t} = \beta \times Treat_p \times Post(m,t) + \gamma \times Post(p,t) + \alpha_p + \alpha_t + \varepsilon_{p,t}$$

In Panel A we focus on two proxies for economic conditions: firm entry and nighttime light density (as a high-frequency proxy for local economic activity) and the estimation is done at the PIN Code-year level. In Panel B, the estimation is at the PIN Code by month level, and We test whether the composition of local investors (measured as the growth rate of local equity investors) and the logarithm of the number of new traders correlate with the arrival of a monitoring station. Standard errors are clustered at the PIN Code level.***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A3: Stacked Regression Approaches

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Var. = Brown Share(%)							
	Control = Never-Treated				Control = Not-Yet-Treated + Never-Treated			
Pollution*Treated*Post	-3.3690**	-3.6248*	-3.5628**	-3.8097*	-2.3407**	-2.0066*	-2.1467*	-1.8594*
	(1.6315)	(1.9720)	(1.7316)	(2.1071)	(1.0859)	(1.0633)	(1.1176)	(1.0940)
Pollution*Treated	1.0997	1.0297	1.1804	1.0972	-0.4553	1.0650*	-0.4307	0.9837
	(1.0649)	(1.1089)	(1.0868)	(1.1472)	(0.8996)	(0.6339)	(0.9334)	(0.6466)
Pollution*Post	2.4620	2.6959	2.6609	2.9002	0.3993	0.4959	0.4474	0.5194
	(1.5002)	(1.8771)	(1.6349)	(2.0560)	(0.7494)	(0.8100)	(0.7925)	(0.8579)
Treat*Post	-0.6727*	-0.6058	-0.6986*	-0.6333	0.2737	0.2060	0.2020	0.1308
	(0.3978)	(0.3992)	(0.4106)	(0.4081)	(0.3402)	(0.3615)	(0.3491)	(0.3721)
Pollution	-0.9946	-0.8722	-1.1067	-0.9871	-0.4830	-0.5546	-0.5221	-0.5738
	(1.0251)	(1.0804)	(1.0636)	(1.1434)	(0.5136)	(0.5296)	(0.5344)	(0.5544)
Post	0.1742**	0.1903*			-0.2289*	0.0354		
	(0.0810)	(0.1025)			(0.1235)	(0.0453)		
Treated					-0.6620		-0.8945	
					(0.4554)		(0.5645)	
No. of Investors(Log)	-1.5907**	-1.5165**	-1.5956**	-1.5245**	-0.4951	-0.1936	-0.4891	-0.1893
	(0.7585)	(0.6194)	(0.7628)	(0.6240)	(0.3987)	(0.3542)	(0.3994)	(0.3548)
Turnover	-0.0821	-0.0172	-0.0816	-0.0168	-0.0848	-0.0382	-0.0850	-0.0384
	(0.1822)	(0.1629)	(0.1828)	(0.1634)	(0.0964)	(0.0896)	(0.0965)	(0.0897)
Rainfall	-0.0094	-0.0040	-0.0096	-0.0041	-0.0044	-0.0025	-0.0041	-0.0025
	(0.0120)	(0.0100)	(0.0123)	(0.0102)	(0.0071)	(0.0064)	(0.0071)	(0.0064)
Temperature	-0.0374	-0.0339	-0.0378	-0.0343	-0.0214	-0.0205	-0.0216	-0.0207
	(0.0246)	(0.0236)	(0.0251)	(0.0241)	(0.0138)	(0.0135)	(0.0139)	(0.0137)
Pincode	Y		Y		Y		Y	
Year-month	Y	Y			Y	Y		
Pincode*Policy Vintage		Y		Y		Y		Y
Year-month*Policy Vintage			Y	Y			Y	Y
Observations	2,449,087	2,449,081	2,449,087	2,449,081	832,578	832,576	832,578	832,576
R-squared	0.398	0.451	0.399	0.451	0.463	0.541	0.464	0.542

Note: This table reports the impact of access to pollution information, the installation of Continuous Ambient Air Quality Monitoring Stations, on the elasticity of the brown share of investor's portfolio using the stacked regression approach in prior literature (Cengiz et al., 2019; Baker, Larcker and Wang, 2022). The methodology involves creating cohort-specific "clean 2×2" datasets that combines the nearby PIN Codes (those within 20 kilometers of the station) for each treated cohort, with other nearby PIN Codes serving as controls. Two different approaches are considered to construct the "clean controls" within the treatment window for each treated cohort. In Columns (1) - (4), We only include observations for "never-treated" PIN Codes that were not treated during our sample period (those that are only treated within one year after our sample period ends) as comparison groups in a given event window. In Columns (5) - (8), we further include "not-yet-treated" PIN Codes (those do not get treated within 4 years for each treated cohort), in addition to the "never-treated" units. Then we estimate the coefficient β from the following specification:

$$Brown\,Share_{p,c,t} = \beta \times Treat_p \times Pollution_{p,t} \times Post_{c,t} + \sum_k \alpha_k \times Other\ Interactions + X'_{p,t}\theta + \gamma_{p,c} + \lambda_{c,t} + \varepsilon_{p,c,t}$$

where $\gamma_{p,c}$ and $\lambda_{c,t}$ represent the PIN Code × Cohort and Cohort × Year-Month fixed effects. We demean the pollution variable to enhance the interpretability of the coefficient. Standard errors are clustered at the PIN Code level.***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A4: Pollution Information and the Brown-Share-Pollution Gradient: Alternative Cutoffs for Treated Investors

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep. Var. = Brown Share(%)					
	0-15km		0-10km		0-5km	
Pollution×Treated×Post	-2.1838**	-2.1799**	-2.4685**	-2.4384**	-2.3415*	-2.4247*
	(0.9939)	(0.9913)	(1.0791)	(1.0727)	(1.2780)	(1.2783)
Pollution×Treated	1.0363**	0.9984**	1.2463**	1.2000**	0.8964	0.9264
	(0.5068)	(0.5057)	(0.5723)	(0.5705)	(0.6618)	(0.6636)
Pollution×Post	1.8972	1.7127	1.6945	1.4959	1.4172	1.1641
	(1.6933)	(1.6860)	(1.6886)	(1.6834)	(1.6985)	(1.6941)
Treat×Post	0.8637***	0.5411*	0.9114***	0.5446	1.7612***	1.3103***
	(0.3153)	(0.3115)	(0.3399)	(0.3348)	(0.3749)	(0.3706)
Pollution	-1.0428	-0.9582	-1.0248	-0.9346	-0.8011	-0.6944
	(0.7856)	(0.7816)	(0.7845)	(0.7813)	(0.7843)	(0.7818)
No. of Investors(Log)		-1.5607***		-1.6086***		-1.6410***
		(0.2099)		(0.2143)		(0.2194)
Turnover		-0.1010**		-0.0794*		-0.0794*
		(0.0424)		(0.0429)		(0.0436)
Rainfall		-0.0110		-0.0103		-0.0056
		(0.0104)		(0.0105)		(0.0106)
Temperature		-0.0309		-0.0152		-0.0189
		(0.0756)		(0.0773)		(0.0797)
Pincode	Y	Y	Y	Y	Y	Y
Station×Year-month	Y	Y	Y	Y	Y	Y
Observations	472,453	472,453	449,136	449,136	424,637	424,637
R-squared	0.560	0.562	0.559	0.561	0.559	0.560

Note: This table presents robustness checks that parallel our main analyses, but use different cutoffs to define treated PIN Codes. Specifically we report the coefficient β from the following specification with three different cutoffs of treatment group:

$$Brown\,Share_{p(m),t} = \beta \times Treat_p \times Pollution_{p,t} \times Post_{m,t} + \sum_k \alpha_k \times Other\ Interactions + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t}$$

where $Brown\,Share_{p(m),t}$ denotes the average share of brown stocks of retail investors in PIN Code p belonging to a station area m at time t. A station area m includes treated PIN Codes and control PIN Codes. Treated pincodes are defined in three alternate ways – 0-15 kms of the station (columns 1 and 2), 0-10kms of the station (columns 3 and 4), 0-5kms of the station (columns 5 and 6). Control pincodes are the pincodes located 40-60 kilometers from the station. $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group.The binary variable $Post_{m,t}$ represents the information shock and is equal to one after the installation of a local monitoring station. We demean the pollution variable to enhance the interpretability of the coefficient. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather characteristics (see text for details). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area × time fixed effects. Standard errors are clustered at the PIN Code level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A5: Robustness: Removing Large Metropolitan Cities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Var. = Brown Share(%)							
Pollution×Treated×Post	-2.1120**	-2.4027***	-2.2073**	-2.7009***	-1.9294**	-2.2359**	-2.0348*	-2.5553**
	(0.8453)	(0.8945)	(1.0572)	(1.0153)	(0.8392)	(0.8907)	(1.0551)	(1.0140)
Pollution×Treated	0.9969**	1.2614***	1.0387*	1.3680***	0.9004**	1.1522**	0.9532*	1.2803**
	(0.4221)	(0.4503)	(0.5301)	(0.5049)	(0.4192)	(0.4487)	(0.5286)	(0.5038)
Pollution×Post	0.1721	0.9363	0.7994	2.0449	-0.0588	0.8390	0.7824	1.9175
	(0.5628)	(0.7067)	(0.7514)	(1.7408)	(0.5585)	(0.7041)	(0.7490)	(1.7352)
Treat×Post	0.9223***	0.5433*	0.4528	0.5814*	0.4792*	0.2852	0.2891	0.3448
	(0.2820)	(0.2843)	(0.3537)	(0.3135)	(0.2774)	(0.2812)	(0.3508)	(0.3102)
Pollution	-0.2487	-0.8179**	-0.8067	-1.2876	-0.1796	-0.7043*	-0.7760	-1.2134
	(0.2595)	(0.3988)	(0.7667)	(0.7874)	(0.2566)	(0.3978)	(0.7626)	(0.7845)
Post	-0.2147	0.0265	0.0622		-0.0978	0.0544	0.0609	
	(0.1705)	(0.1999)	(0.2619)		(0.1693)	(0.1994)	(0.2616)	
No. of Investors(Log)					-2.2615***	-1.7023***	-1.3085***	-1.5524***
					(0.1931)	(0.2018)	(0.2066)	(0.2063)
Turnover					-0.0901**	-0.1019**	-0.0923**	-0.0959**
					(0.0413)	(0.0413)	(0.0419)	(0.0416)
Rainfall					-0.0026	-0.0078	-0.0022	-0.0109
					(0.0040)	(0.0054)	(0.0085)	(0.0101)
Temperature					0.0067	-0.0645**	-0.1489***	-0.0947
					(0.0068)	(0.0313)	(0.0577)	(0.0739)
Pincode	Y	Y	Y	Y	Y	Y	Y	Y
Year-month	Y				Y			
State×Year-month		Y				Y		
District×Year-Month			Y				Y	
Station×Year-month				Y				Y
Observations	473,794	473,794	463,091	473,392	473,794	473,794	463,091	473,392
R-squared	0.543	0.552	0.604	0.565	0.545	0.553	0.605	0.566

Note: This table presents robustness checks that parallel our main analyses, but omits India's three largest cities: Delhi, Kolkata, and Mumbai. Specifically, we report the coefficient β from the following specification after removing Delhi, Kolkata, and Mumbai from the sample :

$$Brown\,Share_{p(m),t} = \beta \times Treat_p \times Pollution_{p,t} \times Post_{m,t} + \sum_k \alpha_k \times Other\ Interactions + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t}$$

where $Brown\,Share_{p(m),t}$ denotes the average share of brown stocks of retail investors in PIN Code p belonging to a station area m at time t. A station area m includes treated PIN Codes (those within 20 kilometers of the station) and also control PIN Codes (40-60 kilometers from the station. $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group.The binary variable $Post_{m,t}$ represents the information shock and is equal to one after the installation of a local monitoring station. We demean the pollution variable to enhance the interpretability of the coefficient. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather characteristics (see text for details). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area × time fixed effects. Columns 1-4 report results without controls $X_{p,t}$ while columns 5-8 report results with these controls. Standard errors are clustered at the PIN Code level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A6: Placebo Tests: Other stock characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A Large Stock Share(%)								
Pollution×Treated×Post	-0.2754	-0.5548	0.1986	-0.6441	-0.2968	-0.5647	0.1635	-0.6613
	(0.3567)	(0.4068)	(0.4538)	(0.4631)	(0.3577)	(0.4078)	(0.4553)	(0.4645)
Observations	498,999	498,999	488,648	498,273	498,999	498,999	488,648	498,273
R-squared	0.511	0.518	0.583	0.532	0.511	0.518	0.583	0.532
Panel B Old Stock Share(%)								
Pollution×Treated×Post	0.5144	0.6475	0.5491	0.9227	0.5413	0.6620	0.5850	0.9418
	(0.5017)	(0.5734)	(0.5866)	(0.6503)	(0.5022)	(0.5727)	(0.5864)	(0.6497)
Observations	498,999	498,999	488,648	498,273	498,999	498,999	488,648	498,273
R-squared	0.619	0.623	0.677	0.633	0.619	0.623	0.677	0.633
Panel C Value Stock Share(%)								
Pollution×Treated×Post	0.2156	0.1902	-0.4151	0.1792	0.2333	0.2117	-0.3860	0.1988
	(0.5267)	(0.5873)	(0.6552)	(0.6679)	(0.5267)	(0.5873)	(0.6549)	(0.6673)
Observations	498,999	498,999	488,648	498,273	498,999	498,999	488,648	498,273
R-squared	0.747	0.752	0.788	0.760	0.747	0.752	0.788	0.760
Lower Order Interactions	Y	Y	Y	Y	Y	Y	Y	Y
Local Trading Controls(No. of Investors, Turnover)	N	N	N	N	Y	Y	Y	Y
Local Weather Controls(Rainfall, Temperature)	N	N	N	N	Y	Y	Y	Y
PIN Code	Y	Y	Y	Y	Y	Y	Y	Y
Year-month	Y				Y			
State×Year-month		Y				Y		
District×Year-Month			Y				Y	
Station×Year-month				Y				Y

Note: This table parallels the results we present in our main analysis in Table 2, replacing brown share as the dependent variable with other stock characteristics, including size (Panel A), age (Panel B), and market-to-book (Panel C). We report the coefficient β from the following specification:

$$Y_{p(m),t} = \beta \times Treat_p \times Pollution_{p,t} \times Post_{m,t} + \sum_k \alpha_k \times Other\ Interactions + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t}$$

where $Y_{p(m),t}$ denotes the average share of different characteristics of stocks of retail investors in PIN Code p belonging to a station area m at time t. A station area m includes treated PIN Codes (those within 20 kilometers of the station) and also control PIN Codes (40-60 kilometers from the station. $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group.The binary variable $Post_{m,t}$ represents the information shock and is equal to one after the installation of a local monitoring station. We demean the pollution variable to enhance the interpretability of the coefficient. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather characteristics (see text for details). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area × time fixed effects. Columns 1-4 report results without controls $X_{p,t}$ while columns 5-8 report results with these controls. Standard errors are clustered at the PIN Code level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A7: Regression Discontinuity Estimates: AQI Transition from "Poor" to "Very Poor"

(1)	(2)	(3)	(4)	(5)	(6)
smoothness constant	M	band-width	estimate	Lower CI	Upper CI
Panel A Treated PIN Codes					
Kernel = Uniform					
0.1*M_rot	0.0005	30	0.5145	0.1781	0.8509
0.5*M_rot	0.0026	15	0.3495	-0.118	0.817
M_rot	0.0051	12	0.4253	-0.1006	0.9512
5*M_rot	0.0257	6	-0.0826	-0.8178	0.6525
10*M_rot	0.0514	5	-0.1876	-1.0608	0.6856
Kernel = Triangle					
0.1*M_rot	0.0005	38.83	0.4949	0.17	0.8199
0.5*M_rot	0.0026	20.16	0.2828	-0.1599	0.7256
M_rot	0.0051	15.44	0.1536	-0.3497	0.6569
5*M_rot	0.0257	8.45	-0.4824	-1.1936	0.2287
10*M_rot	0.0514	6.21	-0.8709	-1.7046	-0.0373
Panel B Control PINCodes					
Kernel = Uniform					
0.1*M_rot	0.0003	36	0.2219	-0.0864	0.5302
0.5*M_rot	0.0016	19	0.1794	-0.2417	0.6005
M_rot	0.0032	14	0.3732	-0.1082	0.8547
5*M_rot	0.0161	8	0.1606	-0.5152	0.8364
10*M_rot	0.0321	5	0.2996	-0.4697	1.0688
Kernel = Triangle					
0.1*M_rot	0.0003	46.29	0.3709	0.0732	0.6686
0.5*M_rot	0.0016	24.20	0.2494	-0.1568	0.6556
M_rot	0.0032	18.46	0.2374	-0.2277	0.7024
5*M_rot	0.0161	9.87	-0.0176	-0.6578	0.6226
10*M_rot	0.0321	7.56	-0.1711	-0.9144	0.5722

Note: The values reported here are RD estimates around the AQI cutoff of 300 at which the air quality index category switch from “Poor”(amber) to “Very Poor”(red). Panel A and B presents the estimate for the treated and control PINCodes, respectively. We follow the inference procedure of Armstrong and Kolesár (2018) to select the optimal bandwidth and to construct the confidence interval. M_{rot} indicates the rule-of-thumb choice of the smoothness constant, M. We also vary M as form of sensitivity analysis. The unit of observation is PIN-Code-day.

Table A8: Heterogeneous Response by Trading Technology

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Var. = Brown Share(%)							
Panel A Mobile								
Pollution×Treated×Post	-10.7735***	-9.4363***	-5.1107**	-10.0833***	-9.8264***	-8.1616***	-3.5364*	-8.7934***
	(1.9153)	(1.9185)	(2.1207)	(2.0372)	(1.7966)	(1.8201)	(2.0493)	(1.9375)
Observations	458,195	458,195	449,041	457,475	458,195	458,195	449,041	457,475
R-squared	0.537	0.544	0.584	0.558	0.551	0.556	0.593	0.569
Panel B Internet								
Pollution×Treated×Post	-2.3938*	-2.0753	-0.1046	-2.1225	-1.3632	-0.6745	1.6337	-0.7254
	(1.3431)	(1.3713)	(1.6224)	(1.5657)	(1.2542)	(1.2931)	(1.5453)	(1.4907)
Observations	458,195	458,195	449,041	457,475	458,195	458,195	449,041	457,475
R-squared	0.318	0.326	0.384	0.344	0.346	0.351	0.404	0.367
Panel C Others								
Pollution×Treated×Post	-3.7532***	-3.3125***	-3.1694***	-3.9838***	-3.5697***	-3.1159***	-2.9570***	-3.7997***
	(0.9025)	(0.9118)	(1.0471)	(1.0536)	(0.8917)	(0.9052)	(1.0414)	(1.0505)
Observations	458,195	458,195	449,041	457,475	458,195	458,195	449,041	457,475
R-squared	0.579	0.587	0.632	0.599	0.581	0.588	0.632	0.600
Lower Order Interactions	Y	Y	Y	Y	Y	Y	Y	Y
Local Trading Controls(No. of Investors, Turnover)	N	N	N	N	Y	Y	Y	Y
Local Weather Controls(Rainfall, Temperature)	N	N	N	N	Y	Y	Y	Y
PIN Code	Y	Y	Y	Y	Y	Y	Y	Y
Year-month	Y				Y			
State×Year-month		Y				Y		
District×Year-Month			Y				Y	
Station×Year-month				Y				Y

Note: This table parallels the results we present in our main analysis in Table 2, splitting the sample based on whether they use mobile (Panel A), internet (Panel B), or other means (Panel C) to trade. Specifically, we report the coefficient β from the following specification separately for panels A, B, and C respectively:

$$Brown\ Share_{p(m),t} = \beta \times Treat_p \times Pollution_{p,t} \times Post_{m,t} + \sum_k \alpha_k \times Other\ Interactions + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t}$$

where $Brown\ Share_{p(m),t}$ denotes the average share of brown stocks of retail investors in PIN Code p belonging to a station area m at time t. A station area m includes treated PIN Codes and control PIN Codes. Treated pincodes are defined in three alternate ways – 0-15 kms of the station (columns 1 and 2), 0-10kms of the station (columns 3 and 4), 0-5kms of the station (columns 5 and 6). Control pincodes are the pincodes located 40-60 kilometers from the station. $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group.The binary variable $Post_{m,t}$ represents the information shock and is equal to one after the installation of a local monitoring station. We demean the pollution variable to enhance the interpretability of the coefficient. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather characteristics (see text for details). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area × time fixed effects. Columns 1-4 report results without controls $X_{p,t}$ while columns 5-8 report results with these controls. Standard errors are clustered at the PIN Code level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A9: Heterogeneity by Age

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Var. = Brown Share(%)							
Panel A: Young								
Pollution×Treated×Post	-7.6956***	-7.5119***	-4.0558*	-7.4996***	-7.2385***	-7.2501***	-3.5523*	-7.2053***
	(1.7136)	(1.8791)	(2.0766)	(2.1505)	(1.6967)	(1.8679)	(2.0804)	(2.1402)
Observations	499,036	499,036	488,681	498,310	499,036	499,036	488,681	498,310
R-squared	0.413	0.421	0.475	0.436	0.419	0.425	0.478	0.440
Panel B: Middle-Aged								
Pollution×Treated×Post	-4.3099***	-4.2561***	-3.3478**	-4.6076***	-4.1376***	-4.1517***	-3.1677**	-4.5061***
	(1.0383)	(1.2334)	(1.3268)	(1.4315)	(1.0372)	(1.2323)	(1.3265)	(1.4323)
Observations	499,036	499,036	488,681	498,310	499,036	499,036	488,681	498,310
R-squared	0.517	0.523	0.577	0.537	0.519	0.524	0.578	0.538
Panel C: Elderly								
Pollution×Treated×Post	-3.0747**	-2.8609*	-2.3076	-2.3993	-2.8516**	-2.7450*	-2.1144	-2.2736
	(1.3565)	(1.5133)	(1.7195)	(1.7483)	(1.3486)	(1.5055)	(1.7148)	(1.7401)
Observations	499,036	499,036	488,681	498,310	499,036	499,036	488,681	498,310
R-squared	0.686	0.690	0.726	0.699	0.688	0.691	0.727	0.700
Lower Order Interactions	Y	Y	Y	Y	Y	Y	Y	Y
Local Trading Controls(No. of Investors, Turnover)	N	N	N	N	Y	Y	Y	Y
Local Weather Controls(Rainfall, Temperature)	N	N	N	N	Y	Y	Y	Y
PIN Code	Y	Y	Y	Y	Y	Y	Y	Y
Year-month	Y				Y			
State×Year-month		Y				Y		
District×Year-Month			Y				Y	
Station×Year-month				Y				Y

Note: This table parallels the results we present in our main analysis in Table 2, splitting the sample based on whether they are young (Panel A), middle-aged (Panel B), or elderly (Panel C) to trade. Specifically, we report the coefficient β from the following specification separately for panels A, B, and C respectively:

$$Brown\ Share_{p(m),t} = \beta \times Treat_p \times Pollution_{p,t} \times Post_{m,t} + \sum_k \alpha_k \times Other\ Interactions + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t}$$

where $Brown\ Share_{p(m),t}$ denotes the average share of brown stocks of retail investors in PIN Code p belonging to a station area m at time t. A station area m includes treated PIN Codes and control PIN Codes. Treated pincodes are defined in three alternate ways – 0-15 kms of the station (columns 1 and 2), 0-10kms of the station (columns 3 and 4), 0-5kms of the station (columns 5 and 6). Control pincodes are the pincodes located 40-60 kilometers from the station. $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group.The binary variable $Post_{m,t}$ represents the information shock and is equal to one after the installation of a local monitoring station. We demean the pollution variable to enhance the interpretability of the coefficient. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather characteristics (see text for details). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area × time fixed effects. Columns 1-4 report results without controls $X_{p,t}$ while columns 5-8 report results with these controls. Standard errors are clustered at the PIN Code level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A10: Heterogeneity by Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Var. = Brown Share(%)							
Panel A Female								
Pollution×Treated×Post	-5.5557***	-5.9249***	-5.3078**	-6.8918***	-5.0272***	-5.6203***	-4.6996**	-6.5368***
	(1.6487)	(1.8170)	(2.0709)	(2.1762)	(1.6136)	(1.7939)	(2.0457)	(2.1486)
Observations	499,036	499,036	488,681	498,310	499,036	499,036	488,681	498,310
R-squared	0.664	0.669	0.701	0.679	0.670	0.674	0.703	0.683
Panel B Male								
Pollution×Treated×Post	-1.8036**	-2.1735***	-2.0207**	-2.4068**	-1.6314**	-2.0787**	-1.8625*	-2.3108**
	(0.7722)	(0.8349)	(0.9716)	(0.9539)	(0.7676)	(0.8314)	(0.9695)	(0.9519)
Observations	499,036	499,036	488,681	498,310	499,036	499,036	488,681	498,310
R-squared	0.538	0.547	0.599	0.560	0.541	0.548	0.600	0.562
Lower Order Interactions	Y	Y	Y	Y	Y	Y	Y	Y
Local Trading Controls(No. of Investors, Turnover)	N	N	N	N	Y	Y	Y	Y
Local Weather Controls(Rainfall, Temperature)	N	N	N	N	Y	Y	Y	Y
PIN Code	Y	Y	Y	Y	Y	Y	Y	Y
Year-month	Y				Y			
State×Year-month		Y				Y		
District×Year-Month			Y				Y	
Station×Year-month				Y				Y

Note: This table parallels the results we present in our main analysis in Table 2, splitting the sample based on whether they are male (Panel A) or female (Panel B) Specifically, we report the coefficient β from the following specification separately for panels A and B respectively:

$$Brown\ Share_{p(m),t} = \beta \times Treat_p \times Pollution_{p,t} \times Post_{m,t} + \sum_k \alpha_k \times Other\ Interactions + X'_{p,t}\theta + \gamma_p + \lambda_{m,t} + \varepsilon_{p,t}$$

where $Brown\ Share_{p(m),t}$ denotes the average share of brown stocks of retail investors in PIN Code p belonging to a station area m at time t. A station area m includes treated PIN Codes and control PIN Codes. Treated pincodes are defined in three alternate ways – 0-15 kms of the station (columns 1 and 2), 0-10kms of the station (columns 3 and 4), 0-5kms of the station (columns 5 and 6). Control pincodes are the pincodes located 40-60 kilometers from the station. $Treat_p$ is an indicator variable which denotes PIN Codes that are in the treated group.The binary variable $Post_{m,t}$ represents the information shock and is equal to one after the installation of a local monitoring station. We demean the pollution variable to enhance the interpretability of the coefficient. $X_{p,t}$ is a set of controls for PIN Code p at time t, including local trader and weather characteristics (see text for details). γ_p is a set of PIN Code fixed effects and $\lambda_{m,t}$ is a set of station area × time fixed effects. Columns 1-4 report results without controls $X_{p,t}$ while columns 5-8 report results with these controls. Standard errors are clustered at the PIN Code level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively..