Public Provision of Air Quality Forecasts: Access, Accuracy, and Household Avoidance*

Lifeng Ren Qingxiao Li Qingyin Cai†

(Job Market Paper)

Abstract

Do air-quality forecasts help households avoid pollution exposure? We link nationwide household transactions to daily $PM_{2.5}$ and archived EPA AirNow forecasts during 2004–2019, exploiting staggered forecast coverage and wind-driven pollution transport. Coverage, which bundles forecasts with current air-quality information, doubles the response of store visits to $PM_{2.5}$. Households with access shift shopping ahead of high-pollution days and respond more when recent forecasts are accurate. Raising recent accuracy from its mean to the 75th percentile cuts exposure by 26 percent of the access effect. Gross shopping-channel mortality benefits are about $16 million annually, before accounting for exposure from rescheduled trips.

Keywords: air pollution, forecast accuracy, value of information, avoidance behavior

JEL Codes: D83, Q53, D12, Q51

*Disclaimer: Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

†Lifeng Ren, Department of Applied Economics, University of Minnesota. Email: ren00154@umn.edu. Qingyin Cai, Department of Agricultural Economics & Agribusiness, Louisiana State University. Email: qingyincai@agcenter.lsu.edu. Qingxiao Li, Department of Agricultural Economics & Agribusiness, Louisiana State University. Email: qli@agcenter.lsu.edu.

1 Introduction

Governments devote substantial resources to producing and disseminating public forecasts. Weather services cost more than $15 billion annually worldwide (Rogers and Tsirkunov, 2013), while the U.S. National Weather Service alone issues roughly 1.5 million forecasts each year on a budget of about $1 billion.[1] Air-quality forecasts are a more recent addition to this infrastructure and now reach hundreds of millions of U.S. residents through the Environmental Protection Agency's (EPA) AirNow platform, which has expanded gradually across the country since 2004. Yet little is known about whether households act on these forecasts and how forecast accuracy affects the public value of the service.

Despite substantial progress in emissions regulation, households remain exposed to residual pollution and often make routine decisions without knowing the level of air pollution they face. Fine particulate matter ($PM_{2.5}$) is particularly hard to gauge. It is invisible and odorless, so households cannot evaluate their exposure through direct experience or anticipate future concentrations (Graff Zivin and Neidell, 2009; Neidell, 2009). Yet $PM_{2.5}$ exposure raises mortality risk, depresses labor productivity, and generates substantial defensive expenditures (Borgschulte, Molitor, and Zou, 2024; Deryugina et al., 2019; Graff Zivin and Neidell, 2012; Ito and Zhang, 2020). Public forecasts complement emissions regulation by filling this information gap, providing advance notice of pollution risk at precisely the moment when households can postpone purchases, shift trips, and limit outdoor exposure.

This paper evaluates the behavioral value of public air-quality forecasting by examining how forecast access and forecast accuracy affect household avoidance in the United States. Throughout, forecast access refers to entry into AirNow $PM_{2.5}$ forecast coverage, which provides forecasts alongside current air-quality information. The estimated access effect captures the combined effect of both information sources. We identify forecast-specific behavior through anticipatory responses and variation in forecast accuracy. We begin by developing a conceptual framework of household avoidance under imperfect information about pollution. Forecasts create value by reducing uncertainty about future pollution, while their accuracy shapes the credibility of the information provided. The framework distinguishes two policy margins: expanding forecast access and improving forecast accuracy, with accuracy operating through both day-specific errors and accumulated reliability.

We study these margins among U.S. households using NielsenIQ Homescan shopping-

[1] Source: National Oceanic and Atmospheric Administration (NOAA).

transaction data from 2004 to 2019, daily county-level $PM_{2.5}$ concentrations, ZIP-code-level entry dates into AirNow $PM_{2.5}$ forecast coverage matched to household locations, and archived AirNow forecasts. To isolate exogenous variation in pollution, we instrument for local $PM_{2.5}$ using wind-driven pollution transport from upwind counties (Deryugina et al., 2019; Barwick et al., 2024b; Chen et al., 2021; Kim and Trusov, 2025). This variation identifies households' shopping responses to pollution, which we refer to as the shopping–pollution gradients.

Our identification of forecast access relies on the staggered expansion of AirNow forecast coverage across reporting areas. The key identifying assumption is that, absent forecast coverage, the shopping–pollution gradient would have evolved similarly across areas with different rollout timing and that coverage entry did not systematically coincide with residual pollution shocks. The assumption is supported by the institutional setting, in which rollout timing reflected federal program design, funding, and the forecasting capacity of participating state, local, and Tribal air agencies, rather than short-run local demand for forecasts or prior responsiveness to pollution. We provide detailed institutional background in Appendix B.1. We also empirically assess the identifying assumption. An IV event study shows no differential pre-trends in the shopping–pollution gradient in the weeks before coverage entry. Rollout timing is not systematically related to pre-treatment pollution, household characteristics, or the baseline shopping–pollution gradient, and coverage entry is not associated with systematic changes in residual pollution. Additional cohort-specific and placebo exercises provide further support.

We find that higher $PM_{2.5}$ reduces in-person shopping activity. Focusing on trip-based avoidance, a 1 $\mu g/m^3$ increase in $PM_{2.5}$ (about 11 percent of the mean) lowers the probability of making an in-person shopping trip by 0.07 percent and the number of store visits by 0.11 percent of their respective means. In-store spending, an associated shopping outcome, also declines by 0.18 percent of its mean. The magnitudes are slightly smaller than estimates for broader mobility and spending outcomes, consistent with grocery shopping being less discretionary (Barwick et al., 2024b; Chen, Fu, and Lyu, 2026; Sun, 2025).

After coverage entry, these gradients steepen substantially. The trip-probability gradient more than doubles, the store-visit gradient roughly doubles, and the in-store spending gradient rises by 73 percent relative to its pre-coverage level. Because coverage bundles forecasts with current air-quality information, this same-day access effect does not isolate the value of forecasts alone. We identify a forecast-specific margin through households' anticipatory responses to future pollution. After coverage entry, a 1 $\mu g/m^3$ increase in realized next-day $PM_{2.5}$ raises today's shopping probability, store visits, and

spending by about 0.32, 0.39, and 0.76 percent of their respective means more than before coverage entry, with smaller responses to pollution two days ahead. Such forward-looking adjustments cannot be generated by current air-quality information and provide direct evidence that households respond to the forecast component of AirNow.

Forecast-specific behavior is also stronger when forecasts are more accurate. Among areas with forecast access, we exploit within-ZIP variation in two dimensions of forecast quality: day-specific forecast errors and recent forecast accuracy. We find that the response to realized pollution is weaker on days when that day's forecast misses, and stronger when recent forecasts have been more accurate. A key identification concern for the latter result is that recent forecast accuracy may proxy for the underlying predictability of pollution rather than forecast reliability itself. We find little evidence of this alternative explanation, as pre-coverage pollution predictability does not generate similar avoidance responses and the estimates remain stable within realized AQI bins. Moving from the sample mean of recent forecast accuracy to the 75th percentile steepens the shopping–pollution gradient by about 13 percent across outcomes, and the effect is robust to alternative ways of weighting past forecast performance. This pattern is consistent with a credibility channel through which accumulated reliability, beyond any single forecast signal, increases the behavioral value of public information.

A back-of-the-envelope calculation translates our estimates into shopping-channel exposure reductions and monetary benefits using external estimates of the mortality damages associated with $PM_{2.5}$ exposure. Moving forecast accuracy from its mean to the 75th percentile yields an exposure reduction equal to 26 percent of that from forecast access. Among covered households, the access effect corresponds to about $16.1 million in annual benefits, while greater accuracy generates approximately $4.2 million in additional benefits. These estimates provide an upper bound on net shopping-channel benefits because they do not net out exposure from trips shifted across days. Since the calculation excludes other health outcomes and avoidance margins, they represent only a partial valuation of the information provided through AirNow forecast coverage.

Our analysis makes three contributions. First, we provide national-scale revealed-preference evidence that public air-quality forecasts shift household shopping before pollution is realized. This anticipatory response emerges only after forecast entry, distinguishing it from responses to current air-quality information or to the persistence of pollution itself. Second, we show that shopping responses vary with day-specific forecast errors and recent forecast accuracy, indicating that behavioral responsiveness depends on signal quality. Third, we quantify the benefits of expanding access and improving accuracy in common physical and monetary units and compare them with the cost of

providing the service. These contributions connect four strands of literature.

The first concerns the economic value of public forecasts. Forecasts shape expectations and decision-making by households and firms (Carroll, 2003; Coibion and Gorodnichenko, 2015; Coibion et al., 2024), with particular value for environmental risks because advance information enables protective adjustment before exposure occurs. Prior work shows that routinely available weather forecasts affect environmental beliefs and agricultural adaptation (Burlig et al., 2024; Rosenzweig and Udry, 2014) and broader responses to climate risk (Downey, Lind, and Shrader, 2023; Shrader, 2023). A related literature studies forecast-based warnings and alerts issued when predicted risks cross policy-relevant thresholds. Such alerts affect disaster preparedness (Beatty, Shimshack, and Volpe, 2019; Molina and Rudik, 2022), responses to hazardous weather (Anand, 2025), and avoidance during high-pollution episodes (Aguilar-Gomez, 2025; Anderson, Hyun, and Lee, 2026; Graff Zivin and Neidell, 2009; Neidell, 2009; Janke, 2014; Saberian, Heyes, and Rivers, 2017; Ward and Beatty, 2016). Closest to our setting, Ahmad et al. (2026) show experimentally that air-quality forecasts improve individuals' ability to predict pollution and adjust protective behavior in Lahore, Pakistan. We extend this literature by studying routinely issued air-quality forecasts in a large-scale operational system and showing how forecast access and accuracy shape avoidance behavior at scale.

Second, we contribute to the body of work on the value of accuracy and credibility in environmental forecasts. Nascent research shows that forecast accuracy matters for temperature-related mortality (Shrader, Bakkensen, and Lemoine, 2023; Shrader et al., 2026), extreme-weather risk communication (Molina et al., 2021; Molina and Rudik, 2022; Ripberger et al., 2015), and weather-related social and labor responses (Song, 2021, 2024). We bring this signal-quality perspective to public air-quality forecasts, where information is noisy and spatially heterogeneous. By treating forecast accuracy as a margin distinct from access, we highlight accuracy as an important component of public information provision and quantify its behavioral and welfare value relative to expanding access.

The third relevant literature is on information and avoidance behavior. The closest paper is Barwick et al. (2024a), which shows that the rollout of real-time air-quality monitoring in China increased protective spending and reduced outdoor activity. Related work studies the behavioral effects of pollution information disclosure in other high-pollution settings (Gao, Song, and Timmins, 2023; Greenstone et al., 2022; Jha and Nauze, 2022). We extend this literature from contemporaneous information to prospective information by studying how forecasts allow households to adjust behavior before pollution is realized. Building on Barwick et al. (2024a), we model trust in the fore-

casting system as a Bayesian belief about forecast reliability, consistent with economic models of learning from noisy signals (Baley and Veldkamp, 2023; Berentsen, Brueger, and Loertscher, 2008; Eusepi and Preston, 2011; Lorenzoni, 2009; Malmendier and Nagel, 2016). This framework distinguishes avoidance responses arising from current information, forecast-based anticipation, and learning about forecast reliability.

Finally, we contribute to the empirical literature on pollution-induced avoidance behavior. A large body of work documents the effects of pollution on labor supply (Borgschulte, Molitor, and Zou, 2024; Graff Zivin and Neidell, 2012), recreation (Gellman, Walls, and Wibbenmeyer, 2025; Keiser, Lade, and Rudik, 2018), morbidity (Barwick et al., 2024b; Deryugina et al., 2019), and defensive investments (Deschênes, Greenstone, and Shapiro, 2017; Ito and Zhang, 2020). We add to this literature by studying a high-frequency consumption margin that is economically important and closely tied to daily exposure.[2] Unlike studies that infer avoidance from aggregate mobility or foot-traffic patterns (Fitzgerald, 2024; Sun, 2025), our transaction-level data directly reveal household-level shopping decisions, including shopping trips and spending.

The rest of the paper is organized as follows. Section 2 describes the AirNow forecast program, data sources, and sample construction. Section 3 presents the conceptual framework. Section 4 outlines the empirical strategy. Section 5 reports the main results. Section 6 discusses welfare implications, and Section 7 concludes.

2 Background and Data

2.1 AirNow Forecast Program

AirNow is the U.S. EPA's public platform for communicating air-quality information through the Air Quality Index (AQI). It aggregates information from participating air-quality agencies and disseminates it through the AirNow website, email notifications, widgets, and smartphone apps. AirNow provides both current information and forecasts for $PM_{2.5}$ and ozone. Current air quality is summarized using the NowCast, which draws on hourly monitor readings, while forecasts report expected air quality for upcoming days. In our analysis, we focus on the forecast product, which provides advance information before pollution is realized.

AirNow's forecast product can matter for policy, individual behavior, and public

[2] A related study using NielsenIQ data is Sheldon and Zhan (2026), which examines household responses to California wildfire smoke. Our setting differs by focusing on routine day-to-day $PM_{2.5}$ variation and the role of public forecast information, rather than a localized episode-scale wildfire shock.

communication. EPA guidance explains that air quality forecasts are intended to help people plan outdoor activities, and that air quality action days and similar community programs are often based on forecasts when unhealthy conditions are expected. By making future pollution risk salient before it is realized, AirNow forecasts can inform public messaging and household exposure-avoidance decisions. Appendix Figures F.1 and F.2 illustrate the AQI categories associated with different levels of health risk and the AirNow forecast interface, respectively.

Beginning in 2004, AirNow began publishing daily forecasts for fine particulate matter ($PM_{2.5}$).[3] Forecast communication is organized around "reporting areas," the operational geographies used by participating state, local, and Tribal air agencies to report AQI information. These areas do not always coincide with county boundaries: some cover multiple counties, while others divide large metropolitan areas into multiple forecast zones. Forecasts are generally issued before the forecast day and typically cover the next one or two days, although a few reporting areas provide forecasts up to six days ahead. Forecast development draws on weather forecasts, satellite imagery, monitoring data, and pollution transport models.

AirNow forecast coverage expanded gradually over time rather than through a single nationwide rollout. This expansion reflected both federal support for local forecasting programs and differences across participating agencies in the capacity to produce forecasts. Early federal initiatives targeted metropolitan areas according to population and provided support in discrete funding waves, while state, local, and Tribal agencies retained responsibility for developing forecasts for their reporting areas. Forecast adoption therefore depended on slow-moving administrative and technical factors, including agency forecasting capacity, data infrastructure, technical expertise, and communication systems. This institutional background suggests that rollout timing reflected federal program design and heterogeneous agency capacity, rather than being driven primarily by short-run household demand for forecasts or local avoidance behavior. Appendix B.1 provides detailed documentation of these institutional features.

This staggered rollout generates substantial geographic and temporal variation in forecast access, defined by entry into AirNow forecast coverage. Using EPA crosswalk files that link ZIP codes to reporting areas, we identify the first date on which each ZIP code receives AirNow forecast coverage. Figure 1 summarizes this rollout by plotting the calendar year of each ZIP code's first coverage date during 2004–2019.

Entry into AirNow forecast coverage may have increased public awareness of local air quality by making pollution risk more salient and encouraging information acquisition.

[3]Source: EPA Air Quality Research Milestones.

As a suggestive check, we examine Google Trends search activity around the rollout of forecast coverage. Search interest rises after coverage begins, even after residualizing for ambient pollution, consistent with forecast access increasing public attention to local air quality. Details are provided in Appendix D.1.

2.2 Data

Our analysis draws on nationwide datasets of air-pollution measures, meteorological conditions, consumer transactions, and air-quality forecasts in the United States from 2004 to 2019. We link these sources to construct a household-by-day panel containing approximately 176 million observations. Below we describe each data source and present summary statistics.

Air Pollution Readings Data. We use daily monitor-level data on $PM_{2.5}$ and other pollutants from the U.S. EPA's Air Quality System (AQS). We retrieve 24-hour average $PM_{2.5}$ readings from 2004 to 2019 using both Federal Reference Method (FRM) and non-FRM monitors. FRM samplers provide the regulatory gold standard but are spatially sparse and often operate only every third or sixth day. Non-FRM continuous monitors operate daily and cover more sites, though they are not always accepted for regulatory compliance. We combine both monitor types to maximize geographic coverage and reduce missing-data bias on non-reporting days (Mu, Rubin, and Zou, 2026; Zou, 2021). The final dataset includes 4,570 $PM_{2.5}$ monitors.[4]

We construct county-day pollution measures as the population-weighted mean across monitors within each county. We restrict attention to county-days with non-missing pollution measures and exclude counties without any operating monitors. This restriction avoids imputing exposure where local readings are unavailable and reduces potential exposure mismeasurement.[5] Appendix Figure F.5 depicts national trends in county-average $PM_{2.5}$. Panel A shows that average exposure fell by about 35 percent over the sample period from 10.66 $\mu g/m^3$ in 2004 to 6.93 $\mu g/m^3$ in 2019, consistent with stronger pollution controls and technological advances (Shapiro and Walker, 2018). Panel B reports monthly national means, showing recurring winter peaks and short-lived wildfire events. Figure 2 maps 2004–2019 county-level mean $PM_{2.5}$ across the continental United States and uncovers substantial spatial variation, with higher concentrations in

[4]Table F.1 in the appendix reports summary statistics for $PM_{2.5}$ monitors. Figure F.4 illustrates the AQS monitoring network. Panel A plots the annual number of active monitors, Panel B shows the number of FRM monitors, and Panel C maps the spatial distribution of monitors across the continental United States.

[5]This restriction yields a final sample of 974 counties with active air-quality monitoring. Our results generalize most directly to monitored counties rather than to sparsely populated areas with limited monitoring coverage.

the industrial Midwest and Northeast and in wildfire-prone parts of the West.

Meteorological Data. We include daily weather controls to account for environmental conditions that may affect both pollution formation and household behavior. We construct county-level daily maximum temperature, minimum temperature, and precipitation from Parameter-elevation Regressions on Independent Slopes Model (PRISM), and wind speed and direction from 10-meter east–west (U) and north–south (V) wind components in the North American Regional Reanalysis (NARR). Gridded weather variables are aggregated to the county-day level using county boundaries for PRISM and monitor-to-grid-cell matching for NARR. Details on variable construction are provided in Appendix B.

NielsenIQ Homescan Panel Data. We use NielsenIQ Homescan Panel Data from 2004–2019 to measure household shopping trips and purchases.[6] Each year the panel tracks about 40,000–60,000 representative U.S. households, who record consumer packaged goods purchases intended for in-home consumption. Although NielsenIQ records purchases made in stores, online, and through delivery services, transactions are dominated by brick-and-mortar purchases. For each transaction, the data report the purchase date, retailer, household identifier, Universal Product Code (UPC), quantity, unit price, and expenditure. Using these records, we construct household-day measures of in-person trip probability, the number of store visits, and associated total in-store expenditure as the main outcomes in our analysis. We also classify purchases into product groups relevant for avoidance behavior. Defensive goods include protective and mitigation products such as masks, air purifiers and filters, breathing aids, and over-the-counter products for respiratory, cough, throat, and sinus symptoms. Necessities include food and health-related products, while all remaining departments are classified as non-necessities.[7]

NielsenIQ monitors scanning compliance and removes persistently non-reporting households. The average panelist remains in the data for about 54 months. Households report annual demographic and socioeconomic characteristics, including household size and income. We deflate income to constant dollars using the Consumer Price Index (CPI) and place households into three annual-income tiers: low (below $40,000), middle ($40,000–$100,000), and high (above $100,000).[8]

[6]We obtain the data via the Kilts Center at the University of Chicago Booth School of Business.

[7]The United Nations Classification of Individual Consumption According to Purpose (COICOP) designates basic needs that include food and non-alcoholic beverages, housing and utilities, clothing and footwear, and healthcare. We use a narrower definition, classifying only food, non-alcoholic beverages, and health products as necessities, because the dataset does not capture gasoline or utilities and reports clothing and footwear only sparingly.

[8]We reclassify NielsenIQ's categorical income variable by assigning the median of each category's

AirNow Data. We retrieve daily $PM_{2.5}$ AQI forecasts for each reporting area from 2004 to 2019.[9] Using EPA crosswalk files that link ZIP codes to reporting areas, we assign each forecast to the corresponding ZIP code on each day, so that each ZIP code is matched to the official advisory intended for that location. When updated crosswalks are available, we allow the ZIP-to-reporting-area assignment to vary over time. We code ZIP–day forecasts as missing when AirNow does not report a forecast for the associated reporting area. Of the 14,365 unique ZIP codes in our consumer panel, 9,799 received AirNow forecast coverage at some point during the sample period. Appendix Figure F.6 summarizes the cumulative expansion of forecast access among households over 2004–2019, showing that access became increasingly widespread during this period.

Summary Statistics. Panel A of Table 1 summarizes pollution and meteorological variables at the county-day level. The final dataset consists of 2,916,911 county-day observations. Daily $PM_{2.5}$ concentration averages 8.9 $\mu g/m^3$ with a standard deviation of 6.5. Panel B of Table 1 summarizes household shopping behavior across 176,326,439 household-day observations. We deflate expenditures to real 2012 dollars using the consumer price index. On an average day, households have a 0.25 probability of making an in-person shopping trip, visit 0.36 stores, and spend \$12.19 in stores. Panel A of Appendix Table F.2 provides additional summary statistics on spending by product category. Conditional on making an in-person purchase, average spending is about \$48.28. Panel C of Table 1 reports household composition and income. The average household has 2.37 members and the three-category income measure has a mean of 0.84.

3 Conceptual Framework

We develop a stylized model of household shopping behavior under pollution risk and imperfect information. Building on the avoidance framework in Deschênes, Greenstone, and Shapiro (2017) and the value-of-information setup in Barwick et al. (2024a), we study a setting in which households receive both current air-quality information and short-term air-quality forecasts. While existing work largely focuses on current pollution disclosure, AirNow forecasts provide a forward-looking signal that allows households to adjust before pollution is realized. The framework formalizes how forecast access and forecast accuracy shape household avoidance and shopping behavior.

range and then deflating nominal income to 2012 dollars using the CPI. The three-tier grouping mitigates measurement error from coarse categorical reporting.

[9]The 2012–2019 forecasts were downloaded from the AirNow website https://files.airnowtech.org/, whereas archived forecasts for 2004–2011 were obtained from AirNow upon request because they are not publicly available online.

Consider a household that chooses composite consumption, x_t, and avoidance behavior, a_t, on day t, with the price of composite consumption normalized to one. Utility depends on consumption and health, $U_t(x_t, h_t)$, where health, h_t, depends on realized pollution exposure and avoidance behavior. Let c_t denote realized pollution and $\tilde{c}_t$ denote the pollution information available to the household at the time it chooses avoidance behavior. The household has income I_t, wage w_t, health-related earnings function $g(\cdot)$, and avoidance price p_a. We interpret p_a broadly as the resource and flexibility cost of avoidance, including the cost of rescheduling trips, changing shopping timing, or substituting away from in-person shopping. At the time of choice, the household solves:

$$\max_{x_t, a_t} U_t(x_t, h_t) \qquad \text{s.t.} \qquad I_t + w_t g(h_t) \geq x_t + p_a a_t, \qquad h_t = h(\tilde{c}_t, a_t).$$

The information signal $\tilde{c}_t$ determines the household's avoidance choice before exposure is realized.[10] We consider two information regimes corresponding to the empirical coverage comparison. Without forecast access, households rely on a baseline perceived pollution level, c_{0t}, and choose $a_t(c_{0t})$. Forecast access expands the information set. On day t, households observe current pollution c_t and the day $t+1$ forecast released on day t, denoted by $c^F_{t+1|t}$.[11] Under forecast access, the decision-relevant signal is $\tilde{c}_t = \hat{c}_t$, a function of c_t, $c^F_{t+1|t}$, and forecast trust $\mathcal{T}_t$.[12] Avoidance can therefore be written as:

$$a^F_t \equiv a_t(\hat{c}_t) = a_t(c_t, c^F_{t+1|t}, \mathcal{T}_t).$$

We model trust as a Bayesian posterior belief about the reliability of the forecasting system, summarized by a bounded trust stock $\mathcal{T}_t \in [0, 1]$.[13] Let q_t denote the realized quality of the forecast for day t. Given the forecast $c^F_{t|t-1}$ released on day $t-1$ and realized pollution c_t, the forecast error is $e_t = c^F_{t|t-1} - c_t$ and forecast quality is $q_t = Q(|e_t|)$, where $Q'(\cdot) < 0$ and $q_t \in [0, 1]$. Hence, smaller errors imply higher quality. After observing c_t, households learn q_t and update their belief before acting on the newly released fore-

[10] The health function in the optimization problem captures household' ex ante perceived health risk before outcomes are realized. This differs from the ex post health outcome, $h_t = h(c_t, a_t)$, as in Barwick et al. (2024a).

[11] Since the day $t+1$ forecast may incorporate day t pollution conditions, $c^F_{t+1|t}$ is dependent on c_t.

[12] Although the outcome is measured at the daily level, forecast information released during day t can still affect day-t behavior through within-day adjustments or additional avoidance decisions after the forecast becomes available.

[13] This learning process follows Bayesian models of belief updating about latent forecast reliability (Baley and Veldkamp, 2023; Coibion, Gorodnichenko, and Kumar, 2018; Lorenzoni, 2009). Each realized forecast provides a new signal about forecast reliability. $\mathcal{T}_t$ represents accumulated learning from past forecast performance. Details are provided in Appendix A.1.

cast for day $t+1$. We summarize the history of realized forecast quality over the most recent N days by an accumulated quality stock, $\bar{Q}_t^N = \sum_{j=0}^{N-1} \kappa_j^N q_{t-j}$, where $\kappa_j^N \geq 0$ and $\sum_{j=0}^{N-1} \kappa_j^N = 1$.[14] The weight κ_0^N captures the importance of the most recently verified forecast quality. Trust is then given by $\mathcal{T}_t = \Phi(\bar{Q}_t^N)$, with $\Phi'(\cdot) > 0$. This reduced-form mapping absorbs the predetermined prior component of the Bayesian updating process and isolates the marginal role of accumulated realized forecast quality. Because trust enters avoidance through $a_t^F \equiv a_t(c_t, c_{t+1|t}^F, \mathcal{T}_t)$, accumulated forecast quality affects behavior through $\mathcal{T}_t$, with $\partial \mathcal{T}_t / \partial \bar{Q}_t^N = \Phi'(\bar{Q}_t^N)$.

Differentiating the forecast-access decision rule with respect to current pollution decomposes the marginal avoidance response into channels operating through current information, forecast-based anticipation, and credibility-learning.

$$\frac{da_t^F}{dc_t} = \frac{\partial a_t}{\partial c_t} + \frac{\partial a_t}{\partial c_{t+1|t}^F} \frac{\partial c_{t+1|t}^F}{\partial c_t} + \frac{\partial a_t}{\partial \mathcal{T}_t} \left[\frac{\partial \mathcal{T}_t}{\partial \bar{Q}_t^N} \kappa_0^N \frac{\partial q_t}{\partial c_t} \right]. \tag{1}$$

All partial derivatives of a_t are evaluated at $(c_t, c_{t+1|t}^F, \mathcal{T}_t)$. The first term, $\frac{\partial a_t}{\partial c_t}$, captures the direct response to current pollution. The second term, $\frac{\partial a_t}{\partial c_{t+1|t}^F} \frac{\partial c_{t+1|t}^F}{\partial c_t}$, captures forecast-induced intertemporal substitution.[15] The third term, $\frac{\partial a_t}{\partial \mathcal{T}_t} \left[\frac{\partial \mathcal{T}_t}{\partial \bar{Q}_t^N} \kappa_0^N \frac{\partial q_t}{\partial c_t} \right]$, captures Bayesian learning about forecast reliability. Since the newly verified forecast quality q_t enters $\bar{Q}_t^N$ with weight κ_0^N, current pollution affects posterior trust through $\kappa_0^N \frac{\partial q_t}{\partial c_t}$. A larger κ_0^N makes trust more responsive to the most recent quality realization. In the special case where $\kappa_0^N = 1$, households place all weight on the most recent realization, so $\bar{Q}_t^N = q_t$. A smaller κ_0^N implies more persistent trust and a weaker same-day learning response. If households place no weight on forecast information, the avoidance response under forecast access reduces to the direct response to current pollution. The strength of this response may also vary with household adjustment cost p_a: for a common forecast signal, households facing tighter constraints may have less flexibility to act on the information.[16]

[14] Including q_t allows the most recently verified forecast performance to enter households' belief updating as day t pollution conditions become observable. Since households observe current pollution information during day t, they can evaluate the forecast made for day t and use this signal when responding to subsequent forecasts.

[15] The static framework captures this forward-looking channel in reduced form through the avoidance policy $a_t^F = a_t(\hat{c}_t)$. Appendix A.3 extends the framework to a two-period setting in which shopping decisions also depend on the forecast-adjusted signal, $x_t^F = x_t(\hat{c}_t)$. The appendix shows that next-day forecast information can induce intertemporal substitution in shopping behavior, shifting planned shopping across days before future pollution is realized, while also affecting current avoidance behavior.

[16] Appendix A.2 provides the maintained assumptions underlying the model and derives the corresponding predictions for forecast access, adjustment-cost heterogeneity, accumulated forecast accuracy, and intertemporal substitution.

The coverage comparison provides a natural way to define the realized value of information under forecast access. Let $V_t(c_t, \tilde{c}_t)$ denote realized indirect utility, with actual pollution c_t determining ex post exposure and health, and the signal $\tilde{c}_t$ determining ex ante avoidance. Specifically, $V_t(c_t, \tilde{c}_t) = U_t[x_t(c_t, \tilde{c}_t), h(c_t, a_t(\tilde{c}_t))]$. The corresponding values are $V_t(c_t, c_{0t})$, $V_t(c_t, c_t)$, and $V_t(c_t, \hat{c}_t)$ under perceived information, current information, and forecast access, respectively. The value of information under forecast access, VOI_t^F, is:

$$VOI_t^F = V_t(c_t, \hat{c}_t) - V_t(c_t, c_{0t}) = [V_t(c_t, c_t) - V_t(c_t, c_{0t})] + [V_t(c_t, \hat{c}_t) - V_t(c_t, c_t)]\,.$$

The first term is the value of the current-information regime, VOI_t^{RT}, relative to baseline perceived pollution, as in Barwick et al. (2024a). The second term is the incremental realized value of forecast information relative to current-information regime, $VOI_t^{F|RT}$. Our main object is their sum, $VOI_t^F = V_t(c_t, \hat{c}_t) - V_t(c_t, c_{0t})$, which compares the combined current and forecast information under forecast access with the baseline perceived information regime. Realized information value need not be positive on every realization. Because households can costlessly ignore forecasts, however, forecast access has weakly non-negative ex ante value. [17]

Empirically, the central object is the difference between the avoidance–pollution gradient under forecast access, $\frac{da_t^F}{dc_t}$, and the baseline gradient, $\frac{da_t(c_{0t})}{dc_t}$. This difference reflects the behavioral response to current information and forecast information provided through AirNow. Its welfare counterpart is $V_t(c_t, \hat{c}_t) - V_t(c_t, c_{0t})$. In Section 6, we use this decomposition to construct a shopping-channel welfare benchmark that translates forecast access and accumulated accuracy into avoided pollution exposure.

4 Empirical Framework

4.1 Pollution and Behavioral Responses

We begin by estimating the shopping–pollution gradient using the following specification:

$$Y_{ict} = \beta \text{PM2.5}_{ct} + \mathbf{\Omega}_{ct}'\delta_1 + \mathbf{X}_{ict}'\delta_2 + FEs + \epsilon_{ict}. \tag{2}$$

Here, Y_{ict} denotes the outcome of interest for household i in county c on day t. Our main outcomes are in-person trip probability and the number of store visits, with to-

[17] Appendix A.4 discusses the realized value of information under forecast coverage relative to the baseline information benchmark.

tal in-store spending included as an associated shopping outcome.[18] This specification pools household-days with and without AirNow forecast access and estimates the average shopping–pollution gradient over the sample period. The parameter of interest, β, captures the effect of a one-unit increase in $PM_{2.5}$ concentration on the outcome. In line with recent guidance on modeling dependent variables that contain zeros (Chen and Roth, 2024; Mullahy and Norton, 2024), our specification is estimated in levels. For ease of interpretation, we report the estimated effect as a percent of the outcome mean.

We control for daily county-level weather conditions $\mathbf{\Omega}_{ct}$, with 17-bin indicators for maximum and minimum temperature (°C) and decile indicators for precipitation (mm) and wind speed (m/s), as these factors can affect both air pollution and consumer behavior (Barwick et al., 2024b; Deryugina et al., 2019). Additionally, we include annual household-level demographic controls, $\mathbf{X}_{ict}$, such as income and household size, to absorb time-varying demand shifters (Lee and Zheng, 2025). The regression also includes holiday indicators, household fixed effects, day-of-week fixed effects, week-by-year fixed effects, and county-by-year fixed effects, which flexibly absorb time-invariant household characteristics and common temporal and spatial shocks. Unless otherwise stated, all regressions include this full set of controls and fixed effects. We cluster standard errors at the county level to allow for arbitrary within-county correlation over time.

Instrumental Variable. OLS estimates of β may be biased because local $PM_{2.5}$ is correlated with economic activity, time-varying shocks such as wildfires and temperature inversions, and unobserved local conditions that also affect household shopping behavior. Measurement error from sparse monitoring, selective reporting, or strategic monitor placement may further attenuate estimates. We address these concerns with an instrumental variable strategy that shifts local pollution but is plausibly unrelated to household shopping except through pollution. Following Barwick et al. (2024b) and Chen et al. (2021), we construct the IV, UP_{ct}, for county c on day t as:

$$\begin{aligned} UP_{ct} &= \sum_{j \neq c} \frac{P_{jt} \times WS_{jt}}{D_{c,j}^2} \times \cos \psi_{c,j,t} \\ \text{where } \cos \psi_{c,j,t} &= \max \left\{ \cos \left(\rho_{c,j} - \zeta_{jt} \right), 0 \right\}. \end{aligned} \tag{3}$$

The instrument exploits daily variation in the amount of particulate matter that wind can transport toward a focal county. For each focal county c, we treat counties $j \neq c$ whose centroids lie within a 100 to 300 km ring as potential upwind sources.[19] The

[18] On days when a household does not record any transaction, all spending and trip variables are set to zero. NielsenIQ's compliance monitoring reduces the likelihood that missing scans are misclassified as non-shopping days. Households that persistently fail to report are removed from the panel.

[19] We use 300 km as the outer bound of the baseline upwind ring because EPA materials indicate that

contribution of county j on day t increases with its pollution concentration P_{jt} and local wind speed WS_{jt}, and decreases with its squared distance to the focal county $D^2_{c,j}$. The term $\cos \psi_{c,j,t}$ assigns greater weight to counties upwind of c and sets weights to zero for downwind counties. Figure 3 illustrates the construction, with details on implementation provided in Appendix C.1. The instrument is strongly predictive of local pollution, with first-stage Kleibergen-Paap F-statistics well above recent conservative benchmarks for valid IV inference in all specifications (Lee et al., 2022).

For the instrument to be valid, it must affect household shopping only through local air quality. The exclusion restriction would fail only if factors other than transported pollution influenced outcomes and simultaneously covaried with (i) the distance between counties, (ii) neighbors' daily pollution levels, (iii) neighbors' daily wind direction, and (iv) neighbors' wind speeds. We address three potential threats. First, nearby counties may share unobserved economic shocks, so we restrict the donor set to counties 100–300 km away and vary the radius in robustness checks. Second, donor counties may themselves be alternative shopping destinations. We address this by excluding counties that could plausibly attract shoppers from the focal county and control for distance-weighted pollution in the ring. Third, wind patterns may correlate with weather conditions that directly affect shopping, which we address by including a rich set of daily weather controls. In general, these design choices make it unlikely that UP_{ct} affects household outcomes through channels other than local air quality.

The forecast-access, forecast-error, and rolling-accuracy specifications described below interact $PM_{2.5}$ with a moderator and instrument that interaction with the upwind instrument interacted with the same moderator, which requires a stronger exclusion restriction than the one above. Validity now requires that, conditional on the controls, fixed effects, and the moderator, wind-transported pollution affect shopping only through local air quality at every level of the moderator. The strong Sanderson-Windmeijer conditional first-stage statistics establish instrument relevance, but not validity. The main concern arises in the accuracy specification, since forecast accuracy may be higher in more meteorologically stable environments, where the upwind instrument may also predict local pollution more precisely. We assess this concern in the robustness analysis. The forecast-quality results are unchanged after controlling for recent pollution volatility, weather volatility, and wildfire smoke, adding reporting-area-by-season fixed effects, and implementing a pre-coverage placebo that interacts pollution with a predictability

sub-300 km distances are a well-established range for modeling regional fine-particle transport, while longer ranges are possible but more uncertain. Source: Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report and Recommendations for Modeling Long-Range Transport Impacts. Publication No. EPA–454/R–98–019. December 1998.

index built from realized pollution alone, each of which proxies for the meteorological simplicity through which the threat would operate.

4.2 Forecast Access and Behavioral Responses

Using the staggered rollout of AirNow forecast coverage, we next examine whether access to public forecasts changes the extent to which households adjust shopping behavior in response to realized pollution. Motivated by Barwick et al. (2024a) and Gao, Song, and Timmins (2023), we estimate a specification that interacts $PM_{2.5}$ with an indicator for forecast access in the household's ZIP code:

$$\begin{aligned} Y_{ict} = \beta_1^A \text{PM2.5}_{ct} + \beta_2^A \text{PM2.5}_{ct} \times \text{Post}_{\text{zt}} + \beta_3^A \text{Post}_{\text{zt}} \\ + \mathbf{\Omega}'_{ct}\delta_1 + \mathbf{X}'_{ict}\delta_2 + FEs + \epsilon_{ict}. \end{aligned} \tag{4}$$

The indicator variable Post_{zt} equals one once ZIP code z enters AirNow forecast coverage, and zero otherwise.[20] All controls and fixed effects are the same as in the baseline specification, with Post_{zt} additionally included. In this specification, β_1^A captures the shopping–pollution gradient before forecast coverage entry, while the coefficient of interest, β_2^A, measures how that gradient changes after coverage entry. We continue to treat $PM_{2.5}$ as endogenous and instrument for both $PM_{2.5}$ and its interaction with Post_{zt} using the corresponding upwind IV. The full first-stage results for both endogenous variables are reported in columns 1 and 2 of Appendix Table F.3.

The causal interpretation of the interaction coefficient rests on two assumptions (Barwick et al., 2024a). First, conditional on the controls and fixed effects, the shopping–pollution gradient would have evolved similarly across areas with different rollout timing absent forecast access. Second, forecast coverage entry does not systematically coincide with residual pollution shocks after accounting for the same controls.

These assumptions are supported by the institutional design of AirNow and the structure of forecast rollout. As discussed in Section 2.1, forecast participation is organized at the reporting-area level, which must be established before forecasts can be issued, and depends on state, local, and Tribal agencies' ability to produce forecasts. The institutional background indicates that rollout timing was shaped by federal program design, funding, and agency forecasting capacity, rather than short-run household demand for forecasts or pre-existing avoidance behavior. These features imply that roll-

[20] Coverage entry is defined as the first date on which an AirNow forecast was issued for the reporting area covering the ZIP code. $PM_{2.5}$ is measured at the county-by-day level, whereas forecast access is measured at the ZIP-code-by-day level. The NielsenIQ Homescan panel reports both household ZIP code and county identifiers, allowing us to merge county-level pollution with ZIP-level forecast access.

out timing largely reflects slow-moving institutional and operational factors. Such factors influence the timing of coverage entry but are unlikely to coincide with a discrete change in households' pollution responses at entry, which we assess directly with the event-study analysis below.

We complement this institutional argument with several empirical diagnostics. First, we estimate an IV event-study specification that interacts $PM_{2.5}$ with leads and lags relative to the first week of forecast access to evaluate changes in the shopping–pollution gradient around coverage entry. To reduce day-to-day noise, we aggregate observations to event-time weeks centered on coverage entry. Figure 4 shows no differential pre-trends before coverage entry within the baseline $\pm$6-week window, with broadly similar patterns using an extended window (Figure F.7). Second, we test whether pre-period ZIP-code characteristics predict AirNow rollout timing. For each ZIP code, we construct annual lagged measures of local $PM_{2.5}$, household size, income group, and the shopping–pollution gradient, and regress an indicator for coverage entry on these characteristics with ZIP-code fixed effects and either year or county-by-year fixed effects. Columns (1) and (2) of Table F.4 show that these characteristics are neither jointly related to rollout timing nor individually statistically significant. Third, we conduct a complementary cross-sectional test using 2004 baseline ZIP characteristics and regress rollout year directly on these characteristics. This exercise assesses whether earlier rollout was systematically associated with fixed baseline differences across ZIP codes. Columns (3) and (4) of Table F.4 show that baseline characteristics do not jointly predict rollout timing in either the full treated sample or the sample restricted to ZIP codes treated strictly after 2004. Fourth, we test whether residualized $PM_{2.5}$ changes discontinuously at coverage entry, after absorbing ZIP code, week, and county-by-year fixed effects and controlling for weather. Column (5) of Table F.4 shows that the estimated discontinuity is small and statistically indistinguishable from zero, with no evidence that rollout coincided with local pollution shocks. The remaining two diagnostics are reported with the robustness analysis below. Fifth, because early reporting areas may have been selected on metropolitan population, we allow the shopping–pollution gradient to vary by rollout cohort and with national trends. The estimated access effect is essentially unchanged. Finally, we conduct placebo tests that randomly assign fake treatment timing to never-treated ZIP codes and arbitrarily reassign actual rollout timing among treated ZIP codes. These placebo exercises show no spurious changes in the shopping–pollution gradient. These diagnostics provide no evidence that rollout timing was systematically related to pre-existing pollution conditions or differential shopping responses.

Beyond same-day avoidance, forecast access may enable anticipatory behavior by sig-

naling future air quality and allowing households to shift shopping away from expected high-pollution days. To test this intertemporal margin, we augment Equation 4 with one- and two-day leads of $PM_{2.5}$ and their interactions with the post-coverage indicator:

$$\begin{aligned} Y_{ict} = & \sum_{\ell=0}^{2} \beta_{1,\ell}^{F} \text{PM2.5}_{c,t+\ell} + \sum_{\ell=0}^{2} \beta_{2,\ell}^{F} \text{PM2.5}_{c,t+\ell} \times \text{Post}_{zt} \\ & + \beta_{3}^{F} \text{Post}_{zt} + \mathbf{\Omega}_{ct}' \delta_1 + \mathbf{X}_{ict}' \delta_2 + FEs + \epsilon_{ict}. \end{aligned} \tag{5}$$

Here, $\ell = 0$ denotes same-day $PM_{2.5}$, while $\ell = 1$ and $\ell = 2$ denote $PM_{2.5}$ realized one and two days ahead. We focus on a two-day window to capture the near-term forecast horizon most relevant for shifting shopping across days.[21] We use realized lead $PM_{2.5}$ to proxy future air quality rather than the AirNow forecast itself because forecasts are observed only after coverage entry, whereas realized leads are available in both periods. The main effects of the leads absorb the pre-existing relationship between current shopping and future pollution driven by $PM_{2.5}$ autocorrelation. The coefficients $\beta_{2,\ell}^{F}$ for $\ell \in \{1, 2\}$ test whether forecast access induces households to shift shopping toward days preceding future high-pollution events, relative to the pre-coverage period when no forecast was available. These anticipatory responses isolate a forecast-specific behavioral margin because current information alone cannot reveal future pollution.

4.3 Forecast Quality and Behavioral Responses

Forecast access determines whether households receive advance information, but its behavioral value also depends on forecast quality through two complementary margins. Same-day forecast errors weaken the informativeness of the forecast signal for current avoidance decisions, which may attenuate immediate behavioral responses. We measure this margin using the absolute distance between forecast and realized AQI categories. Accumulated forecast accuracy, by contrast, captures the reliability of the forecasting system over time and may strengthen trust, thereby amplifying responses to pollution. We measure accumulated accuracy as the share of exact forecast–realization matches over a recent window, and examine whether forecast quality shapes behavioral responses through both short-run signal informativeness and accumulated reliability.

We begin by estimating whether households with forecast access respond less strongly

[21] As shown in Appendix Figure F.2, the AirNow forecast information available to households commonly covers the next one to two days. Forecasts may extend to longer horizons in some locations, but longer-horizon forecasts are less consistently available.

when the forecast is less accurate on a given day:

$$\begin{aligned} Y_{ict} = \beta_1^e \text{PM2.5}_{ct} + \beta_2^e \text{PM2.5}_{ct} \times \text{Error}_{zt} + \beta_3^e \text{Error}_{zt} \\ + \mathbf{\Omega}'_{ct}\delta_1 + \mathbf{X}'_{ict}\delta_2 + FEs + \epsilon_{ict}, \end{aligned} \tag{6}$$

where $\text{Error}_{zt} \equiv \left|\text{AQI}_{zt}^{\text{forecast}} - \text{AQI}_{zt}^{\text{realized}}\right|$ denotes the absolute forecast error for ZIP code z on day t, corresponding to $|e_t|$ in the theoretical model. Because AirNow consistently reports forecasts in discrete AQI categories, we measure error at the category level by comparing the next-day forecast issued on day $t-1$ for day t with the realized category on day t.[22] Larger values indicate lower accuracy.

Identification relies on short-run variation in errors around a predetermined forecast, conditional on realized pollution and weather conditions. The forecast is issued on day $t-1$, before day t pollution and shopping outcomes are realized, so the resulting error provides an ex post measure of the quality of the signal households received. Conditional on weather controls, short-run variation in forecast error is driven largely by unanticipated atmospheric shocks and model or measurement noise rather than local economic activity (Shrader et al., 2026). Our specification also includes a rich set of region and time fixed effects, which absorb persistent cross-area differences in forecasting environments and common improvements in forecasting technology. Columns 3 and 4 of Appendix Table F.3 report the first-stage results for the endogenous variables. Our coefficient of interest, β_2^e, captures how the shopping–pollution gradient changes with forecast error. If larger forecast errors dampen households' behavioral responses to pollution, we would expect β_2^e to be positive.

Beyond day-specific signal error, households may learn from recent forecast performance and adjust their responses accordingly. This reliability channel corresponds to the conceptual model's accumulated forecast-quality stock. We measure it using recent rolling forecast accuracy and estimate the following specification:

$$\begin{aligned} Y_{ict} = \beta_1^Q \text{PM2.5}_{ct} + \beta_2^Q \text{PM2.5}_{ct} \times \text{Accuracy}_{zt} + \beta_3^Q \text{Accuracy}_{zt} \\ + \mathbf{\Omega}'_{ct}\delta_1 + \mathbf{X}'_{ict}\delta_2 + FEs + \epsilon_{ict}. \end{aligned} \tag{7}$$

Here, Accuracy_{zt} denotes the rolling forecast accuracy rate for ZIP code z. For each ZIP code and date, we calculate the share of forecast days in the previous three months, including day t, on which the one-day-ahead forecast AQI category exactly matches

[22] We convert realized 24-hour average $PM_{2.5}$ concentrations into AQI categories using EPA AQI breakpoints for $PM_{2.5}$. Appendix Figure F.8 summarizes trends in annual mean forecast errors and the distribution of forecast errors in our sample.

the realized category. Missing forecasts are excluded from the denominator, so they do not mechanically lower the accuracy measure. Because rolling accuracy can be noisy when forecast history is sparse, we require at least ten days with forecast information in the previous three months. ZIP-code-days with insufficient forecast history are excluded from the analysis. Finally, we center the rolling accuracy rate at its mean in the estimation sample.[23]

Identification of the accumulated-accuracy margin requires that, conditional on the controls and fixed effects, within-ZIP variation in recent forecast accuracy is not systematically associated with other determinants of the shopping–pollution gradient. The main threat to this assumption is that recent accuracy may proxy for the predictability of pollution, because periods when pollution is stable and easier to forecast mechanically raise measured accuracy and may coincide with different shopping dynamics. We address this concern directly and report the supporting evidence in the appendix. First, in the pre-coverage period, when no forecasts were issued, we construct a predictability index using realized pollution alone and interact it with $PM_{2.5}$. The resulting placebo interactions are small and statistically insignificant across outcomes (Appendix Table F.28, Panels A and B), indicating that predictable pollution dynamics alone do not generate the accuracy gradient. Second, we add AQI-bin fixed effects so that identification comes only from variation in accuracy within narrow realized air-quality bins, and the interaction estimate remains similar (Appendix Table F.27, Panel C). We interpret the accuracy gradient as reflecting forecast reliability rather than the predictability of pollution itself.

All controls and fixed effects follow the baseline specification. The full first-stage results for the endogenous variables are reported in columns 5 and 6 of Appendix Table F.3. Our coefficient of interest is β_2^Q, which captures how the shopping–pollution gradient varies with recent forecast accuracy. If households place more weight on forecasts that have recently been more accurate, we would expect higher forecast accuracy to strengthen avoidance behavior and steepen the shopping–pollution gradient, implying a negative β_2^Q.

5 Empirical Results

In this section, we first estimate how $PM_{2.5}$ exposure affects in-person shopping behavior. We then examine whether forecast access changes households' responses to realized

[23] Figure F.9 in the appendix presents the distribution of the rolling three-month forecast accuracy measure. The mean is 0.775, with an interquartile range from 0.657 to 0.965. The distribution is concentrated near 1 but still exhibits substantial variation across observations.

pollution and enables anticipatory adjustments. Finally, we study how forecast quality, measured by both day-specific forecast error and recent rolling forecast accuracy, shapes these responses.

5.1 Pollution and Consumer Avoidance Behavior

Table 2 presents the contemporaneous effect of air pollution exposure on in-person shopping. We focus on three outcomes that capture both margins of adjustment: the probability of any in-person shopping trip (extensive margin), the number of store visits, and total in-store spending (intensive margin). Throughout, trip probability is reported in percentage points and store visits per 100 household-days for ease of interpretation, while expenditures are reported in dollars per household-day. Panel A of Table 2 reports OLS estimates. The OLS gradients are small, and the spending estimates are statistically insignificant. Panel B reports IV estimates using the upwind instrument. The IV estimates are substantially larger than the OLS estimates, suggesting attenuation in OLS. The estimates indicate that households reduce in-person shopping along both the extensive and intensive margins. A 1 $\mu g/m^3$ increase in $PM_{2.5}$ lowers trip probability by 0.07 percent, store visits by 0.11 percent, and total in-store spending by 0.18 percent relative to their respective means.[24] Conditional on positive spending, spending declines by 0.10 percent per 1 $\mu g/m^3$, indicating smaller basket sizes in addition to fewer trips. The first-stage statistics are strong, with Kleibergen-Paap F-statistics around 380.[25]

Our findings align with the broader literature on pollution-induced avoidance behavior summarized in Appendix Table F.7. For trips, Sun (2025) reports a 0.50 percent decline in overall daily visit rates per $\mu g/m^3$ increase in $PM_{2.5}$, and Fitzgerald (2024) finds a 0.15 percent decline in daily visits per $\mu g/m^3$ increase in smoke $PM_{2.5}$ during very unhealthy air quality episodes. Our estimate of a 0.11 percent decline in daily store visits is smaller, as expected, because grocery shopping is less discretionary than the recreation and dining visits captured in mobility-based measures. For spending, Chen, Fu, and Lyu (2026) find that a 100 $\mu g/m^3$ increase in $PM_{2.5}$ reduces daily offline bank-card spending by 14 percent, while Barwick et al. (2024b) report a negative, though

[24] Although our main analysis focuses on in-person shopping, Appendix Table F.5 also reports effects on online shopping. Online purchases represent a small share of shopping activity in our sample. As shown in Panel B of Appendix Table F.2, the mean online trip probability is 0.005, compared with 0.25 for in-person trip probability, and mean online spending is $0.20, compared with $12.19 for in-store purchases, on an average household-day. Pollution induces a small increase in online trips, but online spending does not offset the decline in in-store spending. This pattern suggests that the expenditure decline reflects a real reduction in purchases rather than compensation through online ordering.

[25] Appendix D.2 extends the analysis by examining heterogeneity across product categories. The results show that households disproportionately cut nonessential goods.

statistically insignificant, effect on spending on necessities. Both findings are consistent with our negative expenditure estimates.

Robustness Checks. These results are robust to a wide range of alternative specifications, as discussed in Appendix E. Appendix Tables F.8–F.14 show similar estimates across alternative functional forms, fixed-effects structures, two-way clustering schemes, controls for past-week shopping activity, controls for co-pollutants, and alternative upwind-ring definitions. An alternative wind-direction IV replicating Deryugina et al. (2019) produces closely similar results. These robustness checks support the baseline specification used in the remainder of the paper, where we turn to the roles of forecast access and forecast quality.

5.2 Forecast Access and Behavioral Responses

Same-day Responses. The baseline estimates document the average relationship between realized pollution and household shopping behavior, pooling observations with and without forecast access. We next examine whether this relationship changes once households gain access to AirNow forecasts. Specifically, we test whether coverage entry amplifies the shopping–pollution gradient.

Table 3 presents estimates based on Equation 4. Before coverage entry, there is little evidence that pollution affects whether households shop at all. Trip probability is not significantly related to $PM_{2.5}$, although store visits and in-store spending decline with pollution. This pattern suggests that, absent forecast access, households adjust mainly by reducing trip frequency and spending rather than avoiding shopping altogether.

After AirNow forecast coverage entry, the shopping–pollution gradient steepens across outcomes. Relative to the pre-coverage gradients, the interaction estimates in Table 3 imply that the post-coverage gradient increases by 127 percent for trip probability, roughly doubles for store visits, and rises by 73 percent for in-store spending. The two intensive-margin interactions are significant at the 5 percent level, while the trip-probability interaction is only marginally significant. The spending response is larger in relative terms for non-necessity purchases. Appendix Table F.15 shows a larger post-coverage response for non-necessities than for necessities when scaled by category-specific mean spending, consistent with households using AirNow information to postpone more discretionary consumption on high-pollution days. We find little evidence of increased spending on defensive goods, suggesting that the access effect primarily operates through trip avoidance rather than same-day protective purchases. In total, after coverage entry, a 1 $\mu g/m^3$ increase in $PM_{2.5}$ reduces trip probability by 0.02 percentage points, store visits by 0.05

per 100 household-days, and in-store spending by \$0.025 per household-day. These effects are equivalent to 0.09, 0.13, and 0.21 percent of their respective sample means. Because coverage bundles forecasts with current air-quality information, the access effect captures the combined value of both, while the anticipatory response documented below isolates the forecast component.

Heterogeneity by Vulnerability. As Prediction 1B in Appendix A.2 suggests, forecast access may generate heterogeneous behavioral responses when households face different adjustment costs. We therefore examine three dimensions of household vulnerability along which such costs may differ: poverty status, age vulnerability, and ailment status.[26] Appendix Table F.16 reports subgroup estimates of Equation 4. The post-coverage amplification of the shopping–pollution gradient is statistically significant for households not in poverty and for non-age-vulnerable households, as well as on the intensive margins for households without pollution-related ailments. For the corresponding vulnerable groups, the amplification is imprecise and insignificant, with some point estimates for households in poverty carrying the opposite sign. We interpret these subgroup patterns with caution, given their imprecision, but they are consistent with an adjustment-cost mechanism in which the behavioral value of forecast access depends partly on households' capacity to adjust shopping in response to AirNow information.

Anticipatory Responses. With forecast access, households may not only strengthen same-day avoidance but also shift shopping before pollution is realized. This anticipatory margin captures a forward-looking response that current information alone cannot generate, since the lead main effects net out the autocorrelation channel and the response emerges only after forecast coverage entry. Table 4 reports the results for Equation 5. The estimates reveal a clear forward-looking substitution pattern. The contemporaneous interaction term is negative and highly significant across all three outcomes, consistent with stronger same-day avoidance. Its larger magnitude relative to Table 3 reflects a conditional same-day response, since the specification nets out variation associated with one- and two-day-ahead pollution and their interactions with forecast access. More importantly, the one-day and two-day lead interactions are positive and significant, indicating that households increase shopping ahead of future high-pollution days. After coverage entry, a 1 $\mu g/m^3$ increase in next-day $PM_{2.5}$ raises today's trip probability, store

[26] Poverty status captures financial constraints among households at or below the federal poverty line. Age vulnerability, defined by a household head aged 65 or older, reflects potential mobility constraints and less flexible routines. Ailment status captures air-pollution-related health conditions that may affect both the benefits and costs of avoidance. Appendix B.4 describes the NielsenIQ ailment data and the construction of ailment status. Panel C of Table F.2 reports summary statistics for the three vulnerability measures.

visits, and in-store spending by 0.32, 0.39, and 0.76 percent of their respective means more than before coverage entry. The corresponding two-day lead effects are smaller, at 0.19, 0.17, and 0.35 percent of the respective means. This attenuation suggests that households place less weight on more distant pollution signals.[27] Summing the contemporaneous and two lead interactions, the three-day cumulative access effect amounts to 0.30, 0.28, and 0.62 percent of the respective outcome means for trip probability, store visits, and expenditure. Thus, when pollution remains elevated over several days, forecast access increases current shopping relative to the pre-coverage response, consistent with households shifting trips across days rather than simply forgoing them.

Robustness Checks. The main findings are robust to a variety of additional checks. First, spatial spillovers may arise if households in untreated areas are exposed to forecast information from nearby treated reporting areas. Such cross-area information flows would attenuate treatment–control differences, biasing the estimated access effect toward zero. To assess this concern, we test whether forecast rollout in one reporting area affects nearby areas that have not yet received coverage. For each treated ZIP code, we identify neighboring ZIP codes that share a border with the treated area but remain untreated at that time, and assign them the same rollout date as their treated counterparts. We then allow the change in the shopping–pollution gradient to differ between treated and neighboring areas. Appendix Table F.18 shows no detectable change in the shopping–pollution gradient in neighboring areas when adjacent treated areas receive forecast access, and the treated-area response relative to neighboring areas is similar in magnitude to the main estimate, although the triple-interaction estimate is less precise.

Second, recent work shows that two-way fixed effects estimators can be biased in staggered-adoption settings when treatment effects are heterogeneous across units or over time (Callaway and Sant'Anna, 2021; de Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2021). The concern arises because already-treated units may serve as controls for later-treated units. Our setting differs from a standard staggered DiD design because the object of interest is the shopping–pollution gradient, rather than a level shift at the onset of forecast coverage. The relevant identifying assumption is accordingly that the shopping–pollution gradient would have evolved similarly across rollout cohorts absent forecast access. In addition, our fixed-effects structure restricts the comparison to higher- and lower-pollution days for the same household within the same broad calendar and geographic environment, further limiting the scope for confounding

[27] To address the concern that future pollution instruments are correlated with weather systems that also affect current shopping, Appendix Table F.17 adds one- and two-day-ahead weather controls. The results remain qualitatively unchanged, though the coefficients are attenuated.

trends.

Even so, similar bias can arise in our setting if the shopping–pollution gradient itself evolves with time since treatment, making already-treated units poor controls for later-treated units. To address this concern, we implement a design-based robustness check inspired by Goodman-Bacon (2021) and compare areas treated in 2004–2008 with areas that never received official AirNow forecast access during our sample period.[28] This design uses the never-treated group as the counterfactual, limiting treated-on-treated comparisons. As a result, any remaining difference in the gradient between the two groups is less likely to reflect differential gradient evolution among already-treated units. The estimates in Appendix Table F.19 are similar in sign and magnitude to the baseline results, suggesting that treated-on-treated comparisons are not driving our main estimates.

Third, our specification imposes a common gradient across rollout cohorts and over time. Because early reporting areas may differ systematically from later cohorts, and pollution sensitivity may also shift nationwide, we relax both restrictions by interacting $PM_{2.5}$ with rollout-cohort indicators and a linear year trend, instrumenting each with the corresponding upwind IV interaction. As shown in Appendix Table F.20, estimated on a 30 percent random sample of households for computational reasons, the coverage interactions remain close to their same-sample baselines, at 0.99 to 1.19 times their magnitude, and are marginally significant for trip probability and store visits but insignificant for in-store spending. Pre-entry gradients do not differ significantly across cohorts at the 5 percent level ($p = 0.25$, 0.08, and 0.74, respectively), and the trend term indicates some flattening for trip probability and store visits but not expenditure. Although the smaller sample yields a somewhat different decomposition between the baseline $PM_{2.5}$ coefficient and the coverage interaction than Table 3, the implied post-coverage gradients are similar. These results suggest that the access effect reflects changes at coverage entry rather than pre-existing cohort differences or shifts in pollution sensitivity.

Fourth, we conduct two placebo tests based on randomly reassigned treatment timing to assess the validity of our identification strategy. In the first placebo, we assign placebo coverage-entry dates, randomly drawn from the empirical distribution of actual AirNow rollout dates, to never-treated ZIP codes. If our baseline estimates capture behavioral responses to forecast access rather than pre-existing differences between treated and untreated areas, the placebo interaction should be insignificant. In the second placebo, we restrict the sample to treated ZIP codes but randomly reassign coverage-entry dates

[28] As shown in Figure 1, forecast coverage expanded most rapidly in the earlier part of the sample, whereas later adoption was more limited and gradual. This makes it difficult to define clean late-treated comparison groups and implies that early-treated versus never-treated areas provide the sharpest contrast while limiting treated-on-treated comparisons.

drawn from the empirical distribution of actual rollout dates. This test examines whether the estimated change in the shopping–pollution gradient is tied to the true timing of forecast rollout rather than to unobserved trends in treated areas. Appendix Table F.21 reports the results. Across all six columns, the interaction coefficients are small and statistically indistinguishable from zero, supporting the interpretation that our main estimates capture responses to AirNow forecast access rather than spurious correlation or pre-existing trends.

Fifth, we examine whether our findings are driven by pollution alerts rather than broader access to air-quality forecasts. We re-estimate the specification after excluding days when forecasted or realized AQI reaches the Unhealthy for Sensitive Groups category or worse (i.e., AQI $\geq$ 101). The estimated effects in Appendix Table F.22 remain similar to our baseline results, suggesting that forecast access influences shopping behavior even on Good or Moderate air-quality days.

Sixth, we examine whether households rely on private cues, rather than public forecasts, to assess pollution risk. Similar to the approach in Barwick et al. (2024a), we use visibility as a salient private signal of local air quality. Although visibility is correlated with ambient pollution, the relationship is noisy, especially at low visibility levels. Before coverage entry, shopping responses to low visibility are modest and imprecisely estimated, and they change little once we control for measured pollution. This evidence suggests that visibility was not an effective substitute for forecast information. Details are provided in Appendix D.3.

Finally, because forecast access is defined at the reporting-area level, we re-estimate the specification with standard errors clustered at the reporting-area level rather than at the county level. Appendix Table F.23 shows that the results are robust to this alternative clustering choice.

5.3 Forecast Accuracy and Behavioral Responses

The Role of Day-Specific Forecast Errors. Forecasts provide households with an additional signal about pollution risk. When this signal deviates from realized pollution, the measured response to realized pollution may be attenuated. Table 5 reports the estimates from Equation 6, which examines whether day-specific forecast errors attenuate households' responses to realized pollution. Forecast error is observed only for covered ZIP-code-days, so the analysis is restricted to the forecast-covered sample. The interaction between $PM_{2.5}$ and the absolute forecast error is positive and statistically significant across outcomes, indicating that larger forecast errors attenuate the shopping–pollution

gradient. Specifically, a 1 $\mu g/m^3$ increase in $PM_{2.5}$ reduces trip probability, store visits, and in-store spending by 0.052, 0.070, and 0.085 percent of their respective means when forecasts are perfectly accurate. A one-AQI-category forecast error attenuates these responses to 0.035, 0.048, and 0.038 percent, reducing the implied avoidance response by roughly 33, 31, and 55 percent relative to the perfect-forecast benchmark. This pattern shows that behavior aligns less closely with realized pollution when the issued category forecast is wrong. Because forecast error is measured in AQI categories, days near a category cutoff are both more likely to register a category error and plausibly where avoidance shifts most, so the error measure and the behavioral response could be linked through the category structure itself. The AQI-bin fixed effects in Panel C of Appendix Table F.26 address this concern, since they restrict identification to variation in forecast error within narrow realized air-quality bins, and the interaction remains positive and statistically significant.

As an alternative strategy to assess the impact of day-specific forecast errors, we incorporate households without forecast access and classify days into three groups: no forecast coverage (G1), correct forecast (G2), and incorrect forecast (G3). Appendix Table F.24 shows that the shopping–pollution gradient is more negative on correct-forecast days than on incorrect-forecast days for all three outcomes, and the differences between G2 and G3 are statistically significant throughout.

The Role of Rolling Forecast Accuracy. Beyond a single forecast miss, recent forecast performance may shape how households interpret and respond to pollution information. If households learn from recent forecast accuracy and update their beliefs, more reliable forecasts would strengthen behavioral responses to pollution. Table 6 reports estimates from Equation 7. The coefficients on $PM_{2.5}$ remain negative and statistically significant for in-person trip probability, store visits, and expenditure, implying declines of 0.088, 0.129, and 0.197 percent of their respective means at the sample mean of forecast accuracy. These mean-accuracy gradients are larger than the perfect-forecast gradients in Table 5, a gap that reflects differences in estimation samples rather than weaker responses on accurately forecast days.[29] More importantly, the interaction between $PM_{2.5}$ and rolling forecast accuracy is negative and statistically significant at the 1 percent level, suggesting that avoidance responses are stronger when recent forecasts have been more accurate. Moving from the sample mean to the 75th percentile of forecast accuracy, a shift

[29]The forecast-error specification in Table 5 is estimated on the forecast-covered sample, in which the baseline $PM_{2.5}$ gradient is smaller and less precisely estimated than in the full sample. The rolling-accuracy specification uses the broader forecast-history sample, in which the baseline gradient is -0.0002 ($p < 0.01$), consistent with the full-sample estimate. The level gap between the two sets of gradients therefore reflects sample composition.

of 19.0 percentage points, steepens these gradients by 0.011, 0.017, and 0.025 percent of the respective outcome means. These changes correspond to increases of approximately 12.7, 12.9, and 12.7 percent relative to the baseline gradients at mean accuracy. These results are consistent with households placing more weight on forecasts with better recent performance. In our framework, this reweighting is rational, since trust is defined as a Bayesian posterior belief about forecast reliability.[30]

Heterogeneity by Vulnerability. Forecast quality need not translate into behavioral responses equally across households. In the subgroup estimates of Appendix Table F.25, the interaction between $PM_{2.5}$ and forecast accuracy is negative and significant at the 1 percent level across all three outcomes for households not in poverty and for non-age-vulnerable households. For households in poverty and age-vulnerable households, the interaction is noisily estimated and mostly insignificant, with one exception, a marginally significant spending response among the age-vulnerable. By ailment status, we find little evidence of differential responses between groups. We interpret these subgroup splits with caution, but their direction is consistent with the adjustment-cost mechanism. More accurate forecasts may improve the decision-relevant signal, but households facing higher adjustment costs may have less scope to act on the same information.

Robustness Checks. We conduct several robustness checks to ensure the forecast-quality results are not driven by persistent differences in forecasting environments, short-run changes in forecastability, broad pollution conditions, or inference choices. Appendix Table F.26 reports these checks for the day-specific forecast-error specification. First, forecast error may reflect persistent differences in forecastability across reporting areas rather than short-run variation in forecast quality. Panel A addresses this concern by adding reporting-area-by-season fixed effects, which absorb slow-moving differences in local forecasting environments. The interaction between $PM_{2.5}$ and forecast error remains similar to the baseline across outcomes.

Second, forecast error may proxy for short-run changes in forecastability rather than the quality of the forecast itself. To address this concern, Panel B adds controls for recent pollution volatility, weather volatility, and wildfire smoke days.[31] Specifically, we control for the lagged 30-day rolling variance of realized $PM_{2.5}$, the lagged 30-day standard de-

[30] Because trust is a Bayesian posterior belief, the credibility channel corresponds to rational updating about the precision of the forecast signal rather than to a separate behavioral force. We accordingly interpret the accuracy gradient as households placing more weight on forecasts whose recent performance indicates a more precise signal. Our design does not separately identify any notion of trust beyond this rational reweighting, and the welfare benchmark in Section 6 does not depend on this interpretation.

[31] Wildfire smoke data come from Childs et al. (2022) and are available beginning in 2006.

viation of mean temperature, the lagged 30-day precipitation frequency, and an indicator for county-days with wildfire smoke $PM_{2.5}$ greater than zero to separate smoke-driven spikes from background pollution. The estimates remain close to the baseline, suggesting that the results are not driven by short-run changes in local forecastability.

Third, forecast error may capture differences across broad pollution environments rather than changes in behavioral responses within comparable air-quality conditions. Panel C adds AQI-bin fixed effects, so identification comes from variation in forecast error within narrow realized air-quality bins. The interaction term remains positive and statistically significant, indicating that larger forecast errors weaken the shopping–pollution gradient even within similar realized pollution conditions. Finally, because forecast quality varies at the reporting-area level, inference may be sensitive to the clustering level. Panel D clusters standard errors at the reporting-area level rather than at the county level. The estimates remain statistically significant and similar in magnitude.

We replicate the same robustness checks for the rolling forecast-accuracy specification. Appendix Table F.27 shows that the results are robust to reporting-area-by-season fixed effects, controls for forecastability, AQI-bin fixed effects, and reporting-area-level clustering. Two placebo tests further distinguish the accumulated-accuracy channel from intrinsic predictability. The first uses a pre-coverage predictability index that measures how closely pollution tracks its previous-day level over a rolling 90-day window, based on either mean absolute error or root mean squared error. Its interaction with pollution is statistically insignificant (Appendix Table F.28, Panels A and B). The second interacts pollution with forecast accuracy realized one year in the future, and this interaction is likewise insignificant (Panel C of Appendix Table F.28), consistent with responses to accumulated past performance rather than persistent locational traits.

We also conduct several robustness checks on the construction of the rolling forecast-accuracy measure. First, because the baseline rolling accuracy rate weights all past forecast realizations equally within the window, we construct an exponentially weighted measure of past forecast correctness (McMahon and Rholes, 2023; Muth, 1960), allowing the influence of older realizations to decay over time.[32] Appendix Figure F.12 shows that the interactions are negative and significant across the full range of half-lives, becom-

[32]The decay rate is summarized by the half-life parameter. A shorter half-life implies that households rely primarily on very recent forecast performance, whereas a longer half-life implies that they incorporate a longer history of past performance when assessing forecast reliability. This specification corresponds to the trust-updating process in the conceptual framework, where accumulated forecast quality is constructed using geometrically declining weights. With weights $\kappa_j^N = (1-r)r^j/(1-r^N)$ and memory persistence $r \in [0,1)$, the memory half-life h satisfies $r^h = 1/2$, so $r = 2^{-1/h}$. Thus, forecast quality h days ago receives half as much weight as current forecast quality.

ing more negative as the half-life lengthens[33] Second, we reconstruct forecast accuracy using only forecast performance verified before the decision day, excluding the realized accuracy of the forecast for the decision day from the rolling measure. The estimates in Appendix Table F.29 are similar to the main results. Third, paralleling the day-specific forecast-error measure, we define forecast quality using the negative absolute gap between forecasted and realized AQI categories. Appendix Table F.30 shows similar results, indicating that the findings are not sensitive to using exact-match rates as the baseline measure. Finally, since rolling accuracy is itself a constructed measure, we test sensitivity to alternative window lengths and minimum-data requirements. Appendix Table F.31 shows that using one-month and six-month rolling windows yields estimates similar to the baseline, and Appendix Table F.32 shows that the findings are also robust to alternative minimum forecast-history thresholds.

6 Welfare Implications

To discipline the magnitude of the shopping-channel welfare value of forecast access and forecast accuracy, we conduct a back-of-the-envelope calculation that maps estimated reductions in store visits into avoided $PM_{2.5}$ exposure. The exercise is best interpreted as a gross, same-day benchmark for the shopping-channel rather than a complete welfare calculation. Because households may shift shopping across days rather than forgo trips, the resulting exposure reductions should be viewed as an upper bound on net shopping-channel exposure avoided. The calculation captures exposure reductions on days when store visits decline but does not offset exposure incurred on days to which shopping is rescheduled, and it abstracts from other behavioral margins and utility costs of rescheduling.

As in the conceptual framework, AirNow forecast coverage changes the household's information environment from baseline information, summarized by c_{0t}, to the combined current and forecast information under coverage, summarized by $\hat{c}_t$. The corresponding realized information-value comparison is $V_t(c_t, \hat{c}_t) - V_t(c_t, c_{0t})$. We measure the associated behavioral access margin using the estimated change in the shopping–pollution gradient when households gain forecast access, without separately identifying the contributions of current information and forecasts. Accumulated forecast accuracy

[33]We do not interpret the downward slope as evidence that households place greater weight on a longer history of past performance. Because daily forecast correctness is noisy, longer half-lives can measure accuracy more precisely and mechanically reduce attenuation. We interpret the figure as showing robustness to alternative weighting schemes for past forecast performance, rather than as identifying households' behavioral memory length.

is a distinct intensive margin. Among households with forecast access, it affects forecast credibility and the forecast-adjusted signal used to choose avoidance. Holding the baseline information environment and other signal inputs fixed, this margin is:

$$\frac{\partial V_t(c_t, \hat{c}_t)}{\partial \bar{Q}_t^N} = \frac{\partial V_t(c_t, \hat{c}_t)}{\partial \hat{c}_t} \frac{\partial \hat{c}_t}{\partial \mathcal{T}_t} \frac{\partial \mathcal{T}_t}{\partial \bar{Q}_t^N}. \tag{8}$$

6.1 Exposure Reductions from Forecast Access and Accuracy

We first place the access and accuracy margins on a common physical scale by translating the estimated shopping responses into incremental $PM_{2.5}$ exposure reductions. We denote these by $\Delta e^{Access}(c_t)$ and $\Delta e^{Accuracy}(c_t, \Delta\bar{Q}^N)$, corresponding respectively to gross exposure reductions associated with forecast access and improvements in accumulated forecast accuracy.

Column (2) of Table 3 reports that forecast access steepens the store-visit–pollution gradient by $\hat{\beta}_2^A = -0.0237$ visits per 100 household-days for each 1 $\mu g/m^3$ increase in $PM_{2.5}$.[34] For ambient concentration c_t, the reduction in store visits associated with the coverage-induced change in the pollution gradient is $-\hat{\beta}_2^A c_t$. We convert each avoided store visit into an outdoor-equivalent incremental exposure share of $\nu \approx 0.021$, which incorporates both time spent in the store and associated travel relative to the counterfactual of remaining indoors.[35] At the sample mean $\bar{c} = 8.9$ $\mu g/m^3$, the access effect implies a shopping-channel exposure reduction of $\Delta e^{Access} = 0.0237 \times 8.9 \times \nu \times 8.9 \approx 0.040$ $\mu g/m^3$ per 100 household-days.[36]

For forecast accuracy, column (2) of Table 6 implies that a one-unit increase in accumulated forecast accuracy steepens the store-visit–pollution gradient by $\hat{\beta}_2^Q = -0.0320$

[34] Because the welfare benchmark concerns pollution-related exposure, we use only this coverage-induced change in the shopping–pollution gradient. The level effect of coverage does not vary with contemporaneous pollution and cannot be interpreted as pollution-induced exposure avoidance. The same logic applies to forecast accuracy.

[35] Scully et al. (2017) report an average supermarket visit duration of 23.2 minutes, while Hamrick and Hopkins (2012) report an average one-way grocery-shopping travel time of 15 minutes using nationally representative American Time Use Survey data. In our sample, households make 1.44 store visits conditional on shopping, based on the ratio of mean store visits to shopping probability. Allowing for travel from home to the first store, between stores, and from the final store home implies approximately 48.6 minutes of shopping-related time per store visit. Following Klepeis et al. (2001), Americans spend about 7.6% of their time outdoors, or 1.82 hours per day. Assuming that exposure during the shopping episode is equal to ambient exposure and that the counterfactual indoor exposure is 60% of ambient $PM_{2.5}$, based on the residential infiltration factor in Allen et al. (2012), yields $\nu = \frac{(48.6/60)\times(1-0.6)}{1.82+(24-1.82)\times 0.6} \approx 0.021$. Calculations below use the unrounded value of ν.

[36] Because both the avoided-visit term and the exposure-per-visit term depend on c_t, avoided exposure is proportional to c_t^2. Evaluating at $\bar{c}^2$ rather than the household-day mean of c_t^2 understates the corresponding average exposure reduction under the maintained assumptions.

visits per 100 household-days for each 1 $\mu g/m^3$ increase in $PM_{2.5}$. We evaluate a policy-relevant improvement from the sample mean to the 75th percentile of the accumulated-accuracy distribution. Because the accuracy measure is demeaned, this corresponds to $\Delta\bar{Q}^N = 0.190$, or 19 percentage points. For a household-day with ambient $PM_{2.5}$ concentration c_t, the additional reduction in store visits is $-\hat{\beta}_2^Q \Delta\bar{Q}^N c_t$. Evaluated at $\bar{c} = 8.9$ $\mu g/m^3$, with $\nu \approx 0.021$ and $\Delta\bar{Q}^N = 0.190$, this implies a shopping-channel exposure reduction of $\Delta e^{Accuracy} = 0.0320 \times 0.190 \times 8.9 \times \nu \times 8.9 \approx 0.010$ $\mu g/m^3$ per 100 household-days. Comparing the two exposure reductions yields

$$\frac{\Delta e^{Accuracy}}{\Delta e^{Access}} \approx 0.26. \tag{9}$$

Moving accumulated forecast accuracy from its mean to the 75th percentile thus yields an exposure reduction equal to about 26 percent of that associated with forecast access. Because both estimates use the same store-visit outcome and exposure conversion, the ratio is independent of population scaling, monetization assumptions, the evaluation point $\bar{c}$, and ν, all of which cancel. The aggregate value of these exposure reductions depends on the frequency of these reductions and the number of households affected, which we consider in the monetization exercise below.

6.2 Monetized Shopping-Channel Benefits and Cost Benchmarks

Mortality Benefits. We next monetize both exposure reductions using the same EPA-based unit value. The EPA framework maps changes in $PM_{2.5}$ exposure into mortality risk using concentration-response functions, baseline mortality rates, and the value of a statistical life (VSL).[37] Combining the EPA's illustrative U.S. mortality rate of 868 deaths per 100,000 people, mortality relative-risk estimates of 1.06–1.07 per 10 $\mu g/m^3$ from the American Cancer Society Cancer Prevention Study II, and a central VSL of $9.25 million in 2018 dollars regardless of population characteristics such as age yields an avoided mortality value of about $466–$542 per person-year for a 1 $\mu g/m^3$ reduction in annual average $PM_{2.5}$.[38]

Because the exposure reductions in Section 6.1 are measured per 100 household-days, we divide them by 100 to obtain average household-day exposure reductions before ap-

[37] Source: U.S. EPA Office of Air and Radiation. June 2024. Estimating $PM_{2.5}$- and Ozone-Attributable Health Benefits: 2024 Update.

[38] The implied unit value is $\lambda = m \times [1 - \exp(-\ln(RR)/10)] \times VSL$, where $m = 868/100{,}000$ is the annual mortality rate, RR is the mortality relative risk associated with a 10 $\mu g/m^3$ increase in $PM_{2.5}$, and $VSL = \$9.25$ million.

plying the EPA annual unit value.[39] To place both benefit estimates on the population covered by the program, we scale them by the period-average number of households with forecast access. We compute this population as $\frac{1}{16}\sum_{t=2004}^{2019} s_t \times N_t$, where s_t is the AirNow forecast coverage share in year t from our sample and N_t is the total number of U.S. households in year t from the U.S. Census Bureau. This yields approximately 79.6 million covered households. Using the midpoint of the EPA-based unit value, the estimated annual mortality benefits are approximately $16.1 million for the access effect and $4.2 million for greater forecast accuracy.[40] Holding the exposure conversion and mortality assumptions fixed, the corresponding 95% confidence intervals based on coefficient uncertainty are about $0.8–$31.4 million and $2.0–$6.3 million, respectively.

These national values should be interpreted as extrapolation benchmarks rather than population totals identified directly from the estimation sample. They assume that the sample-average exposure reductions are representative of the relevant household populations, that $\nu \approx 0.021$ reasonably approximates the incremental shopping-related exposure share, and that the EPA mortality-based unit value is appropriate for valuing shopping-related exposure reductions. Because the EPA unit value is based on ambient annual average $PM_{2.5}$ while our estimates capture outdoor-equivalent personal exposure, applying it assumes a comparable mortality response per unit of avoided personal exposure. We use this mapping as a benchmark in the absence of trip-level dose-response estimates.

Comparison with Cost Benchmarks. To benchmark the magnitude of the estimated benefits, we construct a range of annual costs associated with AirNow and the monitoring infrastructure that supports it. The fixed cost consists of the monitoring equipment that supplies AirNow with pollution data. Regulatory-grade monitors cost approximately $15,000 to $40,000 each (Clements et al., 2022). The 4,570 monitors in our sample were installed or replaced from 2004 to 2019, averaging about 286 monitors per year.

[39] The exposure reductions are household-level measures because avoided store visits are measured per household and ν converts each avoided store visit into the corresponding shopper exposure reduction. We apply the EPA per-person unit value once per household, implicitly assuming one shopper per avoided household visit, without an additional household-size multiplier.

[40] As an alternative monetized benchmark, we use the quasi-experimental mortality estimate from Deryugina et al. (2019), which implies 0.685 additional deaths per million Medicare fee-for-service beneficiaries for a 1 $\mu g/m^3$ increase in daily $PM_{2.5}$. Applying this estimate to our household-day exposure reductions and the period-average covered household population yields an annual value of approximately $74.0 million $\times\phi$ for the access effect and $19.0 million $\times\phi$ for forecast accuracy, where ϕ adjusts the Medicare fee-for-service mortality response to the general population. For illustration, these correspond to about $7.4 million and $1.9 million when $\phi = 0.10$, and $18.5 million and $4.7 million when $\phi = 0.25$. We report this only as a sensitivity benchmark because there is no direct mapping from Medicare fee-for-service beneficiaries to the general population.

This implies an annual capital cost of $4.3 million to $11.4 million.[41] Operating costs add about $1.95 million per year. From fiscal years 2017 to 2021, the EPA obligated an annual average of $0.96 million for the Air Quality System and $0.99 million for AirNow (U.S. Government Accountability Office, 2023).

Combining capital and operating expenses gives an estimated AirNow-related cost of $6.2 million to $13.4 million per year, with a midpoint of $9.8 million. This benchmark is deliberately inclusive because much of the monitoring and data-system infrastructure also supports regulatory functions that would continue in the absence of public forecasts. Direct AirNow operating expenditures are substantially smaller, at about $0.99 million annually. For context, federal, state, local, and Tribal agencies spend approximately $150 million to $170 million per year on the entire national ambient monitoring network (U.S. Government Accountability Office, 2020). We treat this amount as a network-wide reference rather than as an AirNow program cost because the network primarily supports enforcement of the National Ambient Air Quality Standards.

At the point estimate, the $16.1 million annual access-effect benefit is about 1.6 times the $9.8 million midpoint of the AirNow-related cost benchmark and 16 times direct AirNow operating expenditures. Given the statistical uncertainty in the benefit estimate and the fact that the cost benchmark includes infrastructure serving multiple regulatory purposes, these comparisons are intended to indicate scale rather than constitute a formal benefit-cost ratio. For forecast accuracy, the relevant comparison is instead the marginal cost of improving forecast quality. Because this cost is not observed, we do not report a benefit-cost ratio for the accuracy margin. The $4.2 million midpoint mortality-benefit estimate provides a benchmark for the potential gross gains from increasing accumulated forecast accuracy from its sample mean to the 75th percentile, and thus a break-even annual cost for such an improvement. Positive net benefits also depend on exposure on rescheduled shopping days and the utility costs of adjustment. These comparisons capture only mortality benefits generated through changes in shopping behavior. Benefits from other health outcomes and avoidance margins would further increase the value of the information provided through AirNow forecast access.

[41] This calculation assigns the full purchase cost of installed or replacement monitors to the forecast program, applies no salvage value or discounting, and includes monitors primarily used for Clean Air Act compliance.

7 Conclusion

We provide nationwide evidence that public air-quality forecasts shape household avoidance behavior. Combining household transaction records with daily $PM_{2.5}$, AirNow forecast coverage, and forecast performance, we show that pollution reduces in-person shopping and that this response strengthens when forecasts are available and more accurate. Our results underscore the importance of both access and quality in determining the behavioral value of public forecasts.

We find that households reduce in-person trips when $PM_{2.5}$ rises, and that forecast access substantially strengthens this response while inducing households to shift shopping ahead of high-pollution days. Forecast quality further matters, with larger day-specific errors weakening avoidance and higher recent forecast accuracy steepening the shopping–pollution gradient. Heterogeneity analyses yield larger point estimates of both access and accuracy effects among households with greater flexibility to adjust shopping behavior, though these patterns remain suggestive. Our back-of-the-envelope calculations imply sizable shopping-channel benefits. Moving from the mean to the 75th percentile of forecast accuracy yields an exposure reduction equal to 26 percent of that from forecast access. Using the midpoint EPA mortality value, the corresponding annual gross benefit benchmarks among covered households are about $16.1 million for the access effect and $4.2 million for greater accuracy.

These findings have three implications. First, public pollution information affects retail activity along an economically meaningful margin that complements conventional health-based assessments. Forecast provision matters not only for health protection, but also for household consumption behavior and local economic activity. Second, the value of public information depends on both access and accuracy. Expanding forecast access delivers substantial benefits, while improving accuracy within the observed range adds about a quarter as much, providing an additional, forecast-specific source of value. Third, the subgroup estimates, though imprecise, raise the possibility of a distributional dimension. To the extent that the private returns to public information depend on households' adjustment costs, information provision alone may not equalize protective behavior.

References

Addoum, J.M., D.T. Ng, and A. Ortiz-Bobea. 2023. "Temperature Shocks and Industry Earnings News." *Journal of Financial Economics* 150:1–45.

Aguilar-Gomez, S. 2025. "Adaptation and mitigation of pollution: Evidence from air quality warnings." *Journal of Environmental Economics and Management* 134:103215.

Ahmad, H.F., M. Gibson, F. Nadeem, S. Nasim, and A. Rezaee. 2026. "Expectations and Adaptation to Environmental Threats." *The Economic Journal* 136:750–779.

Allen, R.W., et al. 2012. "Modeling the Residential Infiltration of Outdoor PM2.5 in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air)." *Environmental Health Perspectives* 120:824–830.

Anand, V. 2025. "Does Getting Forecasts Earlier Matter? Evidence From Winter Advisories and Vehicle Crashes." *American Economic Journal: Economic Policy* 17:106–134.

Anderson, M.L., M. Hyun, and J. Lee. 2026. "Bounds, Benefits, and Bad Air: Net Benefits of Pollution Alerts." *The Economic Journal*, pp. ueag056.

Baley, I., and L. Veldkamp. 2023. "Bayesian Learning." In *Handbook of Economic Expectations*. Elsevier, pp. 717–748.

Barwick, P.J., S. Li, L. Lin, and E.Y. Zou. 2024a. "From Fog to Smog: The Value of Pollution Information." *American Economic Review* 114:1338–1381.

Barwick, P.J., S. Li, D. Rao, and N.B. Zahur. 2024b. "The Healthcare Cost of Air Pollution: Evidence From the World's Largest Payment Network." *The Review of Economics and Statistics*, pp. 1–52.

Baylis, P., P. Bharadwaj, J.T. Mullins, and N. Obradovich. 2025. "Climate and Migration in the United States." *Journal of Public Economics* 249:105446.

Beatty, T.K.M., J.P. Shimshack, and R.J. Volpe. 2019. "Disaster Preparedness and Disaster Response: Evidence from Sales of Emergency Supplies Before and After Hurricanes." *Journal of the Association of Environmental and Resource Economists* 6:633–668.

Berentsen, A., E. Bruegger, and S. Loertscher. 2008. "Learning, Public Good Provision, and the Information Trap." *Journal of Public Economics* 92:998–1010.

Borgschulte, M., D. Molitor, and E.Y. Zou. 2024. "Air Pollution and the Labor Market: Evidence From Wildfire Smoke." *The Review of Economics and Statistics* 106:1558–1575.

Burlig, F., A. Jina, E.M. Kelley, G. Lane, and H. Sahai. 2024. "Long-range Forecasts as Climate Adaptation: Experimental Evidence from Developing-country Agriculture." Working paper, National Bureau of Economic Research.

Callaway, B., and P.H. Sant'Anna. 2021. "Difference-in-Differences With Multiple Time Periods." *Journal of Econometrics* 225:200–230.

Carroll, C.D. 2003. "Macroeconomic Expectations of Households and Professional Forecasters." *The Quarterly Journal of Economics* 118:269–298.

Chang, T.Y., W. Huang, and Y. Wang. 2018. "Something in the Air: Pollution and the Demand for Health Insurance." *The Review of Economic Studies* 85:1609–1634.

Chen, H., S. Fu, and Z. Lyu. 2026. "Air Pollution and Consumers' Avoidance Behavior: Evidence from High-frequency Payment Data in China." *Journal of Environmental Economics and Management*, pp. 103374.

Chen, J., and J. Roth. 2024. "Logs With Zeros? Some Problems and Solutions." *The Quarterly Journal of Economics* 139:891–936.

Chen, S., Y. Chen, Z. Lei, and J.S. Tan-Soo. 2021. "Chasing Clean Air: Pollution-Induced Travels in China." *Journal of the Association of Environmental and Resource Economists* 8:59–89.

Chen, S., M. Wang, and D. Zhang. 2025. "Air Pollution, Labor Productivity, and Individual Consumption." *Journal of Population Economics* 38:63.

Childs, M.L., J. Li, J. Wen, S. Heft-Neal, A. Driscoll, S. Wang, C.F. Gould, M. Qiu, J. Burney, and M. Burke. 2022. "Daily Local-Level Estimates of Ambient Wildfire Smoke $PM_{2.5}$ for the Contiguous US." *Environmental Science & Technology* 56:13607–13621.

Clements, A., R. Duvall, D. Greene, and Dye, Tim. 2022. "The Enhanced Air Sensor Guidebook." Technical report No. EPA/600/R-22/213, U.S. Environmental Protection Agency, Washington, DC, Sep.

Coibion, O., D. Georgarakos, Y. Gorodnichenko, G. Kenny, and M. Weber. 2024. "The Effect of Macroeconomic Uncertainty on Household Spending." *American Economic Review* 114:645–677.

Coibion, O., and Y. Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review* 105:2644–2678.

Coibion, O., Y. Gorodnichenko, and S. Kumar. 2018. "How Do Firms Form Their Expectations? New Survey Evidence." *American Economic Review* 108:2671–2713.

Colin Cameron, A., and D.L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50:317–372.

de Chaisemartin, C., and X. D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators With Heterogeneous Treatment Effects." *American Economic Review* 110:2964–2996.

Deryugina, T., G. Heutel, N.H. Miller, D. Molitor, and J. Reif. 2019. "The Mortality and Medical Costs of Air Pollution: Evidence From Changes in Wind Direction." *American Economic Review* 109:4178–4219.

Deschênes, O., M. Greenstone, and J.S. Shapiro. 2017. "Defensive Investments and the Demand for Air Quality: Evidence From the NOx Budget Program." *American Economic Review* 107:2958–2989.

Downey, M., N. Lind, and J.G. Shrader. 2023. "Adjusting to Rain Before It Falls." *Management Science* 69:7399–7422.

Eusepi, S., and B. Preston. 2011. "Expectations, Learning, and Business Cycle Fluctuations." *American Economic Review* 101:2844–2872.

Fitzgerald, B.A. 2024. "The Price of Staying In: Estimating Wildfire Smoke Avoidance Costs Using Comprehensive US Spending Data." Working Paper.

Gao, X., R. Song, and C. Timmins. 2023. "Information, Migration, and the Value of Clean Air." *Journal of Development Economics* 163:103079.

Gellman, J., M. Walls, and M. Wibbenmeyer. 2025. "Welfare Losses From Wildfire Smoke: Evidence From Daily Outdoor Recreation Data." *Journal of Environmental Economics and Management* 132:103166.

Goodman-Bacon, A. 2021. "Difference-in-Differences With Variation in Treatment Timing." *Journal of Econometrics* 225:254–277.

Graff Zivin, J., and M. Neidell. 2009. "Days of Haze: Environmental Information Disclosure and Intertemporal Avoidance Behavior." *Journal of Environmental Economics and Management* 58(2):119–128.

—. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review* 102:3652–3673.

Greenstone, M., G. He, R. Jia, and T. Liu. 2022. "Can Technology Solve the Principal-Agent Problem? Evidence From China's War on Air Pollution." *American Economic Review: Insights* 4:54–70.

Hamrick, K.S., and D. Hopkins. 2012. "The time cost of access to food–Distance to the grocery store as measured in minutes." *International Journal of Time Use Research* 9:28–58.

Han, X., W. Li, and H. Wang. 2024. "A Burning Issue: Wildfire Smoke Exposure, Retail Sales, and Demand for Adaptation in Healthcare." *Environmental and Resource Economics* 87:3011–3039.

He, X., Z. Luo, and J. Zhang. 2022. "The Impact of Air Pollution on Movie Theater Admissions." *Journal of Environmental Economics and Management* 112:102626.

Ito, K., and S. Zhang. 2020. "Willingness to Pay for Clean Air: Evidence From Air Purifier Markets in China." *Journal of Political Economy* 128:1627–1672.

Janke, K. 2014. "Air Pollution, Avoidance Behaviour and Children's Respiratory Health: Evidence from England." *Journal of Health Economics* 38:23–42.

Jha, A., and A.L. Nauze. 2022. "US Embassy Air-Quality Tweets Led to Global Health Benefits." *Proceedings of the National Academy of Sciences* 119:e2201092119.

Keiser, D., G. Lade, and I. Rudik. 2018. "Air Pollution and Visitation at U.S. National Parks." *Science Advances* 4:eaat1613.

Kim, S., and M. Trusov. 2025. "The Impact of Air Pollution on Consumer Spending." *Journal of Marketing* 89:78–96.

Klepeis, N.E., W.C. Nelson, W.R. Ott, J.P. Robinson, A.M. Tsang, P. Switzer, J.V. Behar, S.C. Hern, and W.H. Engelmann. 2001. "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants." *Journal of Exposure Science & Environmental Epidemiology* 11:231–252.

Lee, D.S., J. McCrary, M.J. Moreira, and J. Porter. 2022. "Valid t-Ratio Inference for IV." *The American Economic Review* 112:3260–3290.

Lee, S., and S. Zheng. 2025. "Extreme Temperatures, Adaptation Capacity, and Household Retail Consumption." *Journal of the Association of Environmental and Resource Economists* 12:427–457.

Lee, S., et al. 2024. "Ambient Air Pollution Exposure and Adult Asthma Incidence: A Systematic Review and Meta-Analysis." *The Lancet Planetary Health* 8:e1065–e1078.

Li, S., et al. 2022. "Association Between Exposure to Air Pollution and Risk of Allergic Rhinitis: A Systematic Review and Meta-Analysis." *Environmental Research* 205:112472.

Lorenzoni, G. 2009. "A Theory of Demand Shocks." *American Economic Review* 99:2050–2084.

Malmendier, U., and S. Nagel. 2016. "Learning from Inflation Experiences." *The Quarterly Journal of Economics* 131:53–87.

McMahon, M., and R. Rholes. 2023. "Building Central Bank Credibility: The Role of Forecast Performance." SUERF Policy Briefs.

Molina, R., D. Letson, B. McNoldy, P. Mozumder, and M. Varkony. 2021. "Striving for Improvement: The Perceived Value of Improving Hurricane Forecast Accuracy." *Bulletin of the American Meteorological Society* 102:E1408–E1423.

Molina, R., and I. Rudik. 2022. "The Social Value of Predicting Hurricanes." Working paper.

Mu, Y., E. Rubin, and E.Y. Zou. 2026. "What's Missing in Environmental Self-Monitoring: Evidence from Strategic Shutdowns of Pollution Monitors." *The Review of Economics and Statistics*, pp. 1–16.

Mullahy, J., and E.C. Norton. 2024. "Why Transform Y? The Pitfalls of Transformed Regressions With a Mass at Zero." *Oxford Bulletin of Economics and Statistics* 86:417–447.

Muth, J.F. 1960. "Optimal Properties of Exponentially Weighted Forecasts." *Journal of the American Statistical Association* 55:299–306.

Neidell, M. 2009. "Information, Avoidance Behavior, and Health: The Effect of Ozone on Asthma Hospitalizations." *Journal of Human Resources* 44:450–478.

Rajagopalan, S., and P.J. Landrigan. 2021. "Pollution and the Heart." *New England Journal of Medicine* 385:1881–1892.

Ripberger, J.T., C.L. Silva, H.C. Jenkins-Smith, D.E. Carlson, M. James, and K.G. Herron. 2015. "False Alarms and Missed Events: The Impact and Origins of Perceived Inaccuracy in Tornado Warning Systems." *Risk Analysis: An International Journal* 35:44–56.

Rogers, D.P., and V.V. Tsirkunov. 2013. *Weather and Climate Resilience: Effective Preparedness through National Meteorological and Hydrological Services*. Publisher: The World Bank.

Rosenzweig, M.R., and C. Udry. 2014. "Rainfall Forecasts, Weather, and Wages over the Agricultural Production Cycle." *American Economic Review* 104:278–283.

Saberian, S., A. Heyes, and N. Rivers. 2017. "Alerts Work! Air Quality Warnings and Cycling." *Resource and Energy Economics* 49:165–185.

Sanderson, E., and F. Windmeijer. 2016. "A Weak Instrument F-test in Linear IV Models with Multiple Endogenous Variables." *Journal of Econometrics* 190:212–221.

Schlenker, W., and W.R. Walker. 2016. "Airports, Air Pollution, and Contemporaneous Health." *The Review of Economic Studies* 83:768–809.

Scully, J.Y., A. Vernez Moudon, P.M. Hurvitz, A. Aggarwal, and A. Drewnowski. 2017. "GPS or travel diary: Comparing spatial and temporal characteristics of visits to fast food restaurants and supermarkets." *PLoS ONE* 12:e0174859.

Shapiro, J.S., and R. Walker. 2018. "Why Is Pollution from US Manufacturing Declining? The Roles of Environmental Regulation, Productivity, and Trade." *American Economic Review* 108:3814–3854.

Sheldon, T.L., and C. Zhan. 2026. "Wildfires, Smoke Pollution, and Household Purchasing Behaviors." *Journal of Economic Behavior & Organization* 244:107474.

Shrader, J. 2023. "Improving Climate Damage Estimates by Accounting for Adaptation." Working paper.

Shrader, J.G., L. Bakkensen, and D. Lemoine. 2023. "Fatal Errors: The Mortality Value of Accurate Weather Forecasts." Working paper, National Bureau of Economic Research.

Shrader, J.G., S. Thies, L. Bakkensen, M. Linsenmeier, and D. Lemoine. 2026. "Weather Forecasts Become More Important for Reducing Mortality as the Climate Warms." *Proceedings of the National Academy of Sciences* 123:e2523372123.

Song, Y. 2021. "The Value of Accurate Weather Forecasts: Social Sentiment Responses Reflected in Social Media in China." Working paper, SSRN Electronic Journal.

—. 2024. "The Value of Weather Forecasts: Evidence From Labor Responses to Accurate Versus Inaccurate Temperature Forecasts in China." *Journal of Environmental Economics and Management* 125:102970.

Sun, L., and S. Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects." *Journal of Econometrics* 225:175–199.

Sun, X. 2025. "Air Pollution and Economic Activity: Evidence From Foot Traffic Patterns." Working paper.

U.S. Government Accountability Office. 2023. "Air pollution: EPA Needs to Develop a Business Case for Replacing Legacy Air Quality Data Systems." Working paper No. GAO-23-105618, Sep.

—. 2020. "Air pollution: Opportunities to Better Sustain and Modernize the National Air Quality Monitoring System." Report to Congressional Requesters No. GAO-21-38, U.S. Government Accountability Office, Washington, DC, Nov.

Ward, A.L.S., and T.K.M. Beatty. 2016. "Who Responds to Air Quality Alerts?" *Environmental and Resource Economics* 65:487–511.

Williams, A.M., D.J. Phaneuf, M.A. Barrett, and J.G. Su. 2019. "Short-Term Impact of PM2.5 on Contemporaneous Asthma Medication Use: Behavior and the Value of Pollution Reductions." *Proceedings of the National Academy of Sciences* 116:5246–5253.

Zhang, J., and Q. Mu. 2018. "Air Pollution and Defensive Expenditures: Evidence From Particulate-Filtering Facemasks." *Journal of Environmental Economics and Management* 92:517–536.

Zhao, M., et al. 2022. "Association Between Long-Term Exposure to PM2.5 and Hypertension: A Systematic Review and Meta-Analysis of Observational Studies." *Environmental Research* 204:112352.

Zou, E.Y. 2021. "Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality." *American Economic Review* 111:2101–2126.

Tables

Table 1: Summary Statistics, 2004–2019

	Mean	Std. Dev.	N
Panel A. Pollutants and Weather			
$PM_{2.5}$ ($\mu g/m^3$)	8.904	6.486	2,916,911
Ozone (parts per million)	0.031	0.011	1,851,517
CO (parts per million)	0.338	0.235	850,335
T_{min} ($^\circ C$)	6.913	10.164	2,916,911
T_{max} ($^\circ C$)	18.571	11.087	2,916,911
Precipitation (mm)	2.898	7.717	2,916,911
Wind Direction ($^\circ$)	199.657	93.617	2,916,911
Wind Speed (m/s)	3.489	1.901	2,916,911
Panel B. In-person Purchases Statistics			
Trip Probability	0.253	0.434	176,326,439
Store Visits	0.364	0.714	176,326,439
Shopping Expenditure ($)	12.193	35.947	176,326,439
Conditional Expenditure ($)	48.278	58.090	44,531,117
Panel C. Household Characteristics			
Household Size	2.372	1.313	580,181
Income Group	0.835	0.690	580,181

Notes: (1) The reported "N" corresponds to the number of county-day observations in Panel A, household-day observations in Panel B, and household-year observations in Panel C. Conditional expenditure is computed for the subsample of household-days with strictly positive in-person shopping expenditures. (2) Income Group: low (below $40,000), middle (between $40,000 and $100,000), and high (above $100,000).

Table 2: Effects of $PM_{2.5}$ on In-person Trips and Expenditure

Dependent Var:	Trip Probability	Store Visits	Expenditure	Expenditure (Cond.)
	(1)	(2)	(3)	(4)
Panel A. OLS estimates				
$PM_{2.5}$ ($\mu g/m^3$)	-0.0042***	-0.0071***	-0.0012	0.0018
	(0.0013)	(0.0025)	(0.0013)	(0.0027)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Dep. Var. Mean	25.26	36.42	12.19	48.28
Effect relative to mean, %	-0.0166	-0.0195	-0.0098	0.0037
Panel B. IV estimates				
$PM_{2.5}$ ($\mu g/m^3$)	-0.0187***	-0.0409***	-0.0221***	-0.0459***
	(0.0052)	(0.0094)	(0.0047)	(0.0098)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Dep. Var. Mean	25.26	36.42	12.19	48.28
Effect Relative to Mean, %	-0.0740	-0.1123	-0.1813	-0.0951
Kleibergen-Paap F-statistics	388	388	388	380
Observations	176,326,439	176,326,439	176,326,439	44,531,117

Notes: This table reports OLS and IV estimates of the contemporaneous shopping–pollution gradient. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. Controls include holiday indicators, weather controls, and household demographics. Weather controls include bins for minimum and maximum temperature, as well as decile indicators for precipitation and wind speed. Household demographics include household size and income group. All specifications include household fixed effects, day-of-week fixed effects, week-by-year fixed effects, and county-by-year fixed effects. Standard errors are clustered at the county level and reported in parentheses. The Kleibergen-Paap F-statistic reports the weak-instrument diagnostic. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3: Effects of Forecast Coverage on Shopping-Pollution Gradient

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0098	-0.0239**	-0.0145***
	(0.0063)	(0.0104)	(0.0055)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ **1**{After Coverage}	-0.0124*	-0.0237**	-0.0106**
	(0.0067)	(0.0115)	(0.0052)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Total Post-coverage Gradient	-0.0222	-0.0476	-0.0251
	[0.0058]	[0.0105]	[0.0049]
Dep. Var. Mean	25.26	36.42	12.19
Effect Relative to Mean, %			
$PM_{2.5}$ $\times$ **1**{After Coverage}	-0.0491	-0.0651	-0.0870
Total Post-coverage Gradient	-0.0879	-0.1307	-0.2059
Kleibergen-Paap F-statistics	203	203	203
Observations	176,326,439	176,326,439	176,326,439

Notes: This table reports how forecast access changes the shopping–pollution gradient. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. **1**{After Coverage} equals one on and after the first date when AirNow forecasts are available for the household's ZIP code, and zero otherwise. Controls include the **1**{After Coverage} indicator, holiday indicators, weather controls, and household demographics. Weather, demographic, and fixed-effect controls follow the baseline specification. Standard errors are clustered at the county level and reported in parentheses. Standard errors for the total post-coverage gradient are computed using the delta method and reported in brackets. The Kleibergen-Paap F-statistic is the weak-instrument diagnostic. With two endogenous variables, we also report conditional first-stage F-statistics following Sanderson and Windmeijer (2016): $F = 323$ for $PM_{2.5}$ and $F = 391$ for $PM_{2.5} \times$ **1**{After Coverage}, confirming that each instrument is individually strong conditional on the other. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4: Forecast Information and Anticipatory Shopping Responses to Air Pollution

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ **1**{After Coverage}	-0.0533***	-0.1036***	-0.0602***
	(0.0161)	(0.0293)	(0.0141)
$PM_{2.5}$, 1-day lead ($\mu g/m^3$) $\times$ **1**{After Coverage}	0.0820***	0.1426***	0.0930***
	(0.0244)	(0.0458)	(0.0262)
$PM_{2.5}$, 2-day lead ($\mu g/m^3$) $\times$ **1**{After Coverage}	0.0472**	0.0617*	0.0434**
	(0.0183)	(0.0333)	(0.0186)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0055	-0.0110	-0.0061
	(0.0171)	(0.0286)	(0.0150)
$PM_{2.5}$, 1-day lead	-0.0268	-0.0643	-0.0525*
	(0.0211)	(0.0395)	(0.0286)
$PM_{2.5}$, 2-day lead	0.0001	-0.0064	0.0078
	(0.0167)	(0.0306)	(0.0181)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Three-day Cumulative Access Effect	0.0759	0.1007	0.0762
	[0.0079]	[0.0128]	[0.0055]
Dep. Var. Mean	25.25	36.49	12.29
Effect Relative to Mean, %			
$PM_{2.5}$ $\times$ **1**{After Coverage}	-0.2111	-0.2839	-0.4898
$PM_{2.5}$, 1-day lead $\times$ **1**{After Coverage}	0.3248	0.3908	0.7567
$PM_{2.5}$, 2-day lead $\times$ **1**{After Coverage}	0.1869	0.1691	0.3531
Three-day Cumulative Access Effect	0.3006	0.2760	0.6200
Kleibergen-Paap F-statistics	169	169	169
Observations	158,713,248	158,713,248	158,713,248

Notes: This table reports forward-looking shopping responses to air pollution after AirNow forecast access becomes available. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. We instrument for $PM_{2.5}$, its leads, and their interactions with forecast access using the corresponding upwind instruments. **1**{After Coverage} equals one on and after the first date when AirNow forecasts become available for the household's ZIP code, and zero otherwise. Controls include the **1**{After Coverage} indicator, holiday indicators, weather controls, and household demographics. Weather, demographic, and fixed-effect controls follow the baseline specification. Standard errors are clustered at the county level and reported in parentheses. The three-day cumulative access effect is the sum of the contemporaneous, one-day-lead, and two-day-lead interactions with forecast coverage. Standard errors for this sum are computed using the delta method and reported in brackets. The Kleibergen-Paap *F*-statistic is the weak-instrument diagnostic. With multiple endogenous variables, we also report Sanderson-Windmeijer conditional first-stage *F*-statistics (Sanderson and Windmeijer, 2016) for each endogenous regressor: 386 for $PM_{2.5}$, 493 for $PM_{2.5}$ $\times$ **1**{After Coverage}, 261 and 440 for the one- and two-day leads of $PM_{2.5}$, and 346 and 465 for their respective interactions with **1**{After Coverage}. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 5: Effects of Day-specific Forecast Errors on the Shopping-Pollution Gradient

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0129*	-0.0252**	-0.0100*
	(0.0066)	(0.0116)	(0.0056)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Error(abs.)	0.0042**	0.0079***	0.0055***
	(0.0018)	(0.0030)	(0.0017)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Dep. Var. Mean	25.01	35.98	11.84
Effect Relative to Mean, %			
At Zero Forecast Error	-0.0516	-0.0700	-0.0845
With One Additional Category Error	-0.0348	-0.0481	-0.0380
Kleibergen-Paap F-statistics	167	167	167
Observations	82,162,849	82,162,849	82,162,849

Notes: This table reports IV estimates of how contemporaneous AQI forecast errors moderate the shopping–pollution gradient among households in forecast-covered areas. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. Absolute forecast error is measured as the absolute difference between the forecasted and realized AQI categories. Controls include the absolute forecast-error term, holiday indicators, weather controls, and household demographics. Weather, demographic, and fixed-effect controls follow the baseline specification. Effects relative to the dependent variable mean are computed at two levels of absolute forecast error. Zero forecast error corresponds to a perfectly accurate forecast. One additional AQI category of forecast error reflects a forecast off by one category. Standard errors are clustered at the county level and reported in parentheses. The Kleibergen-Paap F-statistic is the weak-instrument diagnostic. With two endogenous variables, we also report Sanderson-Windmeijer conditional first-stage F-statistics (Sanderson and Windmeijer, 2016): $F = 357$ for $PM_{2.5}$ and $F = 384$ for $PM_{2.5} \times$ Forecast Error(abs.), confirming that each instrument is individually strong conditional on the other. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 6: Effects of Rolling Forecast Accuracy on Shopping-Pollution Gradient

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0223***	-0.0470***	-0.0241***
	(0.0060)	(0.0112)	(0.0055)
$PM_{2.5}$ ($\mu g/m^3$) × Forecast Accuracy	-0.0150***	-0.0320***	-0.0161***
	(0.0045)	(0.0083)	(0.0038)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Dep. Var. Mean	25.32	36.50	12.22
Effect Relative to Mean, %			
At Mean Forecast Accuracy	-0.0881	-0.1288	-0.1972
At 75th Percentile Forecast Accuracy	-0.0993	-0.1454	-0.2223
Kleibergen-Paap F-statistics	191	191	191
Observations	106,885,045	106,885,045	106,885,045

Notes: This table reports IV estimates of how rolling forecast accuracy moderates the shopping–pollution gradient. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. The sample is restricted to household-day observations with sufficient forecast history, defined as at least 10 forecast days in the prior 90 days. Unlike Table 5, this specification retains observations for which day t lacks a forecast but the prior 90-day history is sufficient. For each ZIP code, we construct a rolling 90-day forecast accuracy rate as the share of forecast days in the prior 90 days, including the current day, for which the one-day-ahead predicted AQI category matches the realized AQI category. We then center this accuracy rate by subtracting its sample mean. Controls include the forecast-accuracy term, holiday indicators, weather controls, and household demographics. Weather, demographic, and fixed-effect controls follow the baseline specification. Since forecast accuracy is demeaned, the mean accuracy row corresponds to the baseline $PM_{2.5}$ coefficient. The 75th percentile corresponds to an accuracy rate 0.190 above the sample mean. Standard errors are clustered at the county level and reported in parentheses. With two endogenous variables, we also report Sanderson-Windmeijer conditional first-stage F-statistics (Sanderson and Windmeijer, 2016): $F = 417$ for $PM_{2.5}$ and $F = 447$ for $PM_{2.5}$ × Forecast Accuracy, confirming that each instrument is individually strong conditional on the other. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Figures

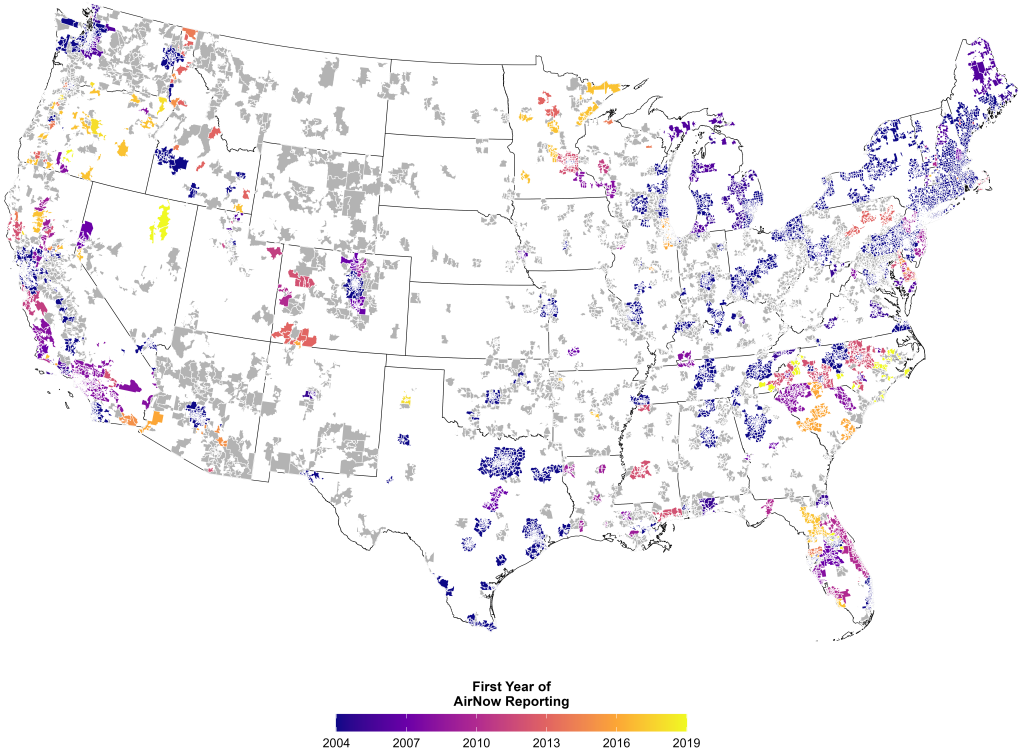


Figure 1: Geographic Coverage and Start Year of AirNow ZIP Code Reporting

Notes: This figure shows ZIP-code-level AirNow forecast coverage across the contiguous United States. Colors indicate the first year in which AirNow forecasts became available for each ZIP code during 2004–2019. ZIP codes included in the NielsenIQ consumer panel but never covered by AirNow during the sample period are shown in gray.

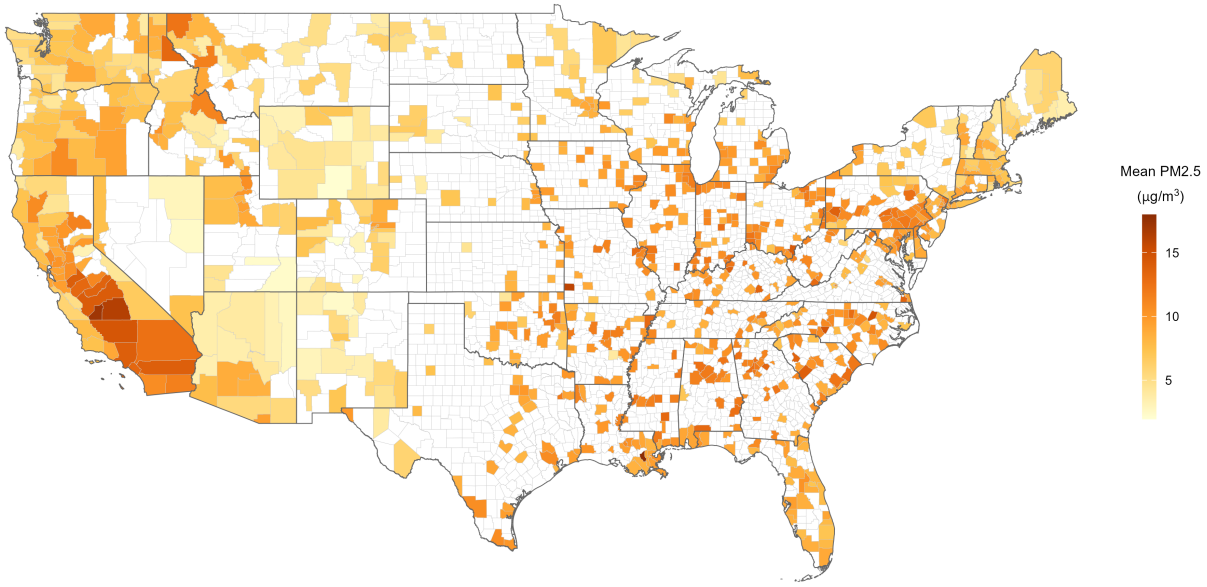


Figure 2: Mean $PM_{2.5}$ Concentrations by County in the Continental United States

Notes: This map shows mean $PM_{2.5}$ concentrations ($\mu g/m^3$) across continental U.S. counties from 2004 to 2019. County means are computed from daily EPA Air Quality System readings, including both FRM and non-FRM monitors. Counties not included in the household analysis sample are shown in white.

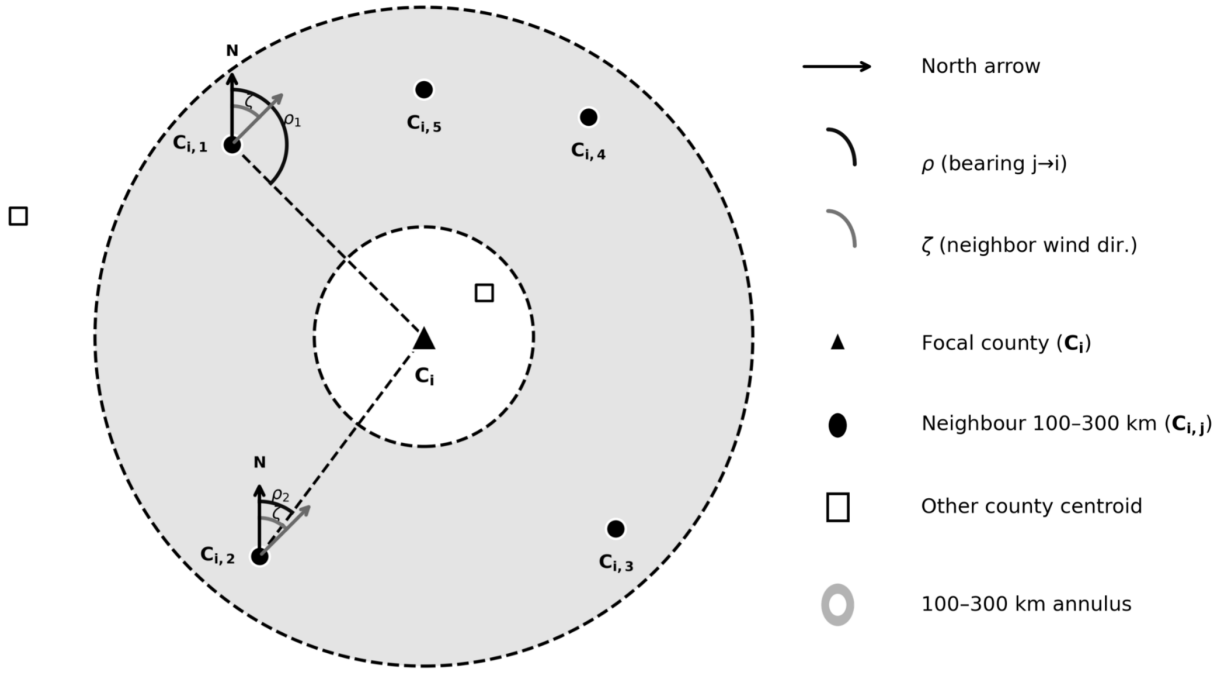


Figure 3: Illustration of Upwind Instrumental Variable Strategy

Notes: This figure illustrates the construction of the upwind pollution instrument. The focal county is denoted by C_i. County $C_{i,2}$ is classified as upwind because the bearing from $C_{i,2}$ to C_i is closely aligned with the wind transport direction, while county $C_{i,1}$ is not because its bearing is perpendicular to the wind direction. In the example, wind blows from southwest to northeast, so the wind transport direction is $\zeta = 45^\circ$ for both counties. The bearing from $C_{i,1}$ to C_i is $\rho_1 = 135^\circ$, yielding $\cos(\rho_1 - \zeta) = \cos(90^\circ) = 0$. The bearing from $C_{i,2}$ to C_i is $\rho_2 = 30^\circ$, yielding $\cos(\rho_2 - \zeta) = \cos(15^\circ) = 0.97$. The instrument weights neighboring-county pollution by directional alignment, wind speed, and inverse squared distance to the focal county.

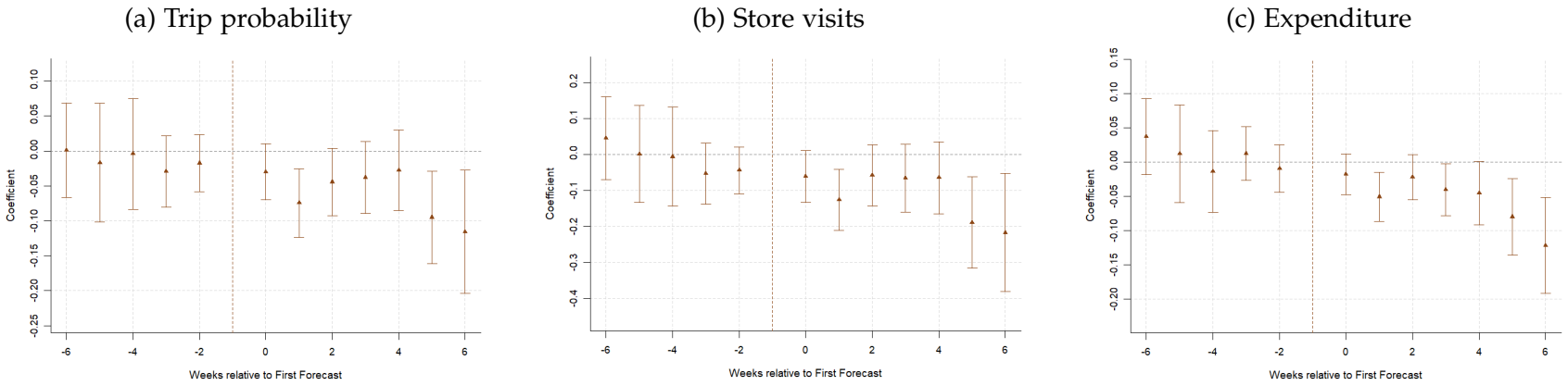


Figure 4: Event Study of Shopping Responses to Forecast Access

Notes: This figure presents IV event-study estimates of how AirNow forecast access changes the shopping–pollution gradient. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. Daily observations are aggregated to event-time weeks relative to the first week of forecast access, with week -1 omitted as the reference period. Each point is a coefficient estimate, and bars represent 95% confidence intervals. Standard errors are clustered at the county level. Joint Wald tests fail to reject the null that all pre-period leads equal zero ($p = 0.842, 0.390, 0.506$ for panels (a)–(c)). An extended ± 12-week specification yields broadly similar patterns shown in Appendix Figure F.7, with joint pre-period p-values of 0.112, 0.096, and 0.120 for panels (a)–(c), respectively.

Online Supplementary Materials

Public Provision of Air Quality Forecasts: Access, Accuracy, and Household Avoidance

Lifeng Ren Qingxiao Li Qingyin Cai

A Conceptual Framework Extension **S1**

A.1 Bayesian Learning and Forecast Trust S1

A.2 Predictions S2

A.3 Two-Period Extension S6

A.4 Value of Information S8

B Institutional Background and Data Construction **S10**

B.1 Institutional Background of AirNow S10

B.2 PRISM Climate Data S11

B.3 NARR Wind Data S12

B.4 NielsenIQ Ailment Data S12

C Instrumental Variable Construction **S14**

C.1 Main Instrumental Variable S14

C.2 Alternative Instrumental Variable S15

D Additional Analyses **S16**

D.1 Public Forecast Access and Awareness S16

D.2 Pollution and Spending Composition S17

D.3 The Role of Private Information S17

E Appendix Robustness Checks **S19**

E.1 Alternative Functional Forms S19

E.2 Different Sets of Fixed Effects S19

E.3 Different Clustering Levels S20

E.4 Controls for Past-week Shopping Activities S20

E.5 Controls for Other Pollutants S20

E.6 Sensitivity of the Upwind IV Construction S21

E.7 Alternative Instrumental Variables S21

F Appendix Tables and Figures **S22**

F.1 Appendix Tables S22

F.2 Appendix Figures S54

A Conceptual Framework Extension

This appendix provides additional details on the trust model, derives the main predictions under the maintained assumptions, and presents a two-period extension.

A.1 Bayesian Learning and Forecast Trust

This appendix formalizes the Bayesian interpretation of the trust stock introduced in the main text. Recall that trust is defined as $\mathcal{T}_t = \Phi(\bar{Q}_t^N)$, where $\bar{Q}_t^N$ summarizes recent realized forecast quality and $\Phi(\cdot)$ is a weakly increasing function that maps accumulated forecast quality into the bounded trust stock $\mathcal{T}_t \in [0,1]$. For the empirical implementation, we use geometrically declining weights:

$$\bar{Q}_t^N \equiv \sum_{j=0}^{N-1} \kappa_j^N q_{t-j}, \qquad \kappa_j^N \equiv \frac{\mu(1-\mu)^j}{1-(1-\mu)^N}, \qquad \sum_{j=0}^{N-1} \kappa_j^N = 1.$$

Here, N is the length of the forecast-history window, j indexes forecast lags, and q_{t-j} is the realized forecast-quality signal for day $t-j$, normalized so that higher values indicate more accurate forecasts. The parameter $\mu \in (0,1]$ is the constant learning gain, or the speed of trust updating. A larger μ places more weight on recent forecast performance, while a smaller μ makes trust more persistent. With this indexing, the most recent verified signal is q_t and its weight is κ_0^N.[1]

This representation follows a simple Bayesian learning process. Let θ denote latent forecast reliability. Before observing the most recent forecast-performance signal, households have prior mean trust $\mathcal{T}_{t-1} = \mathbb{E}[\theta \mid \mathcal{I}_{t-1}]$. $\mathcal{I}_{t-1}$ denotes the household's information set before observing the current realized forecast-quality signal, and $\tau_{T,t-1}$ is the precision, or inverse variance, of the prior belief about θ. The current signal is $q_t = \theta + u_t$, where $\mathbb{E}[u_t] = 0$ and $\mathrm{Var}(u_t) = \tau_q^{-1}$. The noise term u_t captures transitory variation in realized forecast quality, and τ_q denotes the precision of the signal. Bayesian updating gives:

$$\mathcal{T}_t = (1-K_t)\mathcal{T}_{t-1} + K_t q_t, \qquad K_t = \frac{\tau_q}{\tau_{T,t-1} + \tau_q}.$$

The posterior precision is $\tau_{T,t} = \tau_{T,t-1} + \tau_q$, and the learning gain K_t is the share of total posterior precision coming from the current signal. Thus, trust moves more toward q_t when τ_q is large relative to $\tau_{T,t-1}$, and adjusts less when prior trust is more precise. To

[1] Equivalently, if memory persistence is written as $r = 1-\mu$, the weights can be written as $\kappa_j^N = (1-r)r^j/(1-r^N)$.

connect this Bayesian update to the accumulated-quality stock $\bar{Q}_t^N$, suppose the learning gain is approximately constant over the most recent N verified forecast days, with $K_t \approx \mu$. Iterating the update gives

$$\mathcal{T}_t = (1-\mu)^N \mathcal{T}_{t-N} + \left[1 - (1-\mu)^N\right] \bar{Q}_t^N.$$

This expression shows that current trust is a weighted average of the inherited prior trust stock, $\mathcal{T}_{t-N}$, and the accumulated realized forecast quality over the most recent N verified forecast days, $\bar{Q}_t^N$.

A.2 Predictions

For a single time period, the household solves an optimization problem described in the main text:

$$\max_{x_t, a_t} \; U_t(x_t, h_t) \qquad \text{s.t.} \qquad I_t + w_t g(h_t) \geq x_t + p_a a_t, \qquad h_t = h(\tilde{c}_t, a_t).$$

The decision-relevant signal $\tilde{c}_t$ equals the baseline perceived pollution c_{0t} without local AirNow forecast coverage and the forecast-weighted index $\hat{c}_t = \hat{c}(c_t, c^F_{t+1|t}, \mathcal{T}_t)$ with forecast access. The baseline summarizes the information otherwise available to households, while $\hat{c}_t$ incorporates both current information and forecasts. As in the main text, the superscript F denotes the covered information regime. We impose regularity conditions ensuring an interior solution and a positive avoidance response to a worse pollution signal.

Assumption A1. Health production. Health is decreasing in the decision-relevant pollution signal and increasing in avoidance, $h_{\tilde{c}}(\tilde{c}, a) \leq 0$ and $h_a(\tilde{c}, a) \geq 0$. The marginal health benefit of avoidance is weakly decreasing in avoidance, $h_{aa}(\tilde{c}, a) \leq 0$, and weakly larger when the pollution signal is worse, $h_{a\tilde{c}}(\tilde{c}, a) \geq 0$. On the relevant interior margin, $h_{\tilde{c}}(\tilde{c}, a) < 0$ and $h_a(\tilde{c}, a) > 0$.

Assumption A2. Utility, budget regularity, and adjustment-cost attenuation. Utility is increasing in consumption and health, $U_x(x, h) > 0$ and $U_h(x, h) > 0$, and is locally concave in health, $U_{hh}(x, h) < 0$. The budget constraint binds at the optimum, so $\lambda_t = U_{x,t} > 0$. Avoidance is costly, $p_a > 0$, and health increases effective earnings at a weakly decreasing rate, $g_h(h) \geq 0$ and $g_{hh}(h) \leq 0$. For tractability, we assume that the marginal utility of composite consumption is locally constant on the relevant margin. Hence, these conditions imply $U_{h,t} + \lambda_t w_t g_h > 0$ and $U_{hh,t} + \lambda_t w_t g_{hh} < 0$.

Assumption A3. Forecast access weakly raises the decision-relevant pollution signal on the relevant margin. With forecast access, the decision-relevant pollution signal is $\hat{c}_t = \hat{c}(c_t, c^F_{t+1|t}, \mathcal{T}_t)$, which depends on current pollution, the next-day forecast, and forecast trust. We assume that, on the relevant margin where baseline perceptions understate pollution risk, information available under coverage does not lower perceived pollution risk relative to baseline information: $\hat{c}_t \geq c_{0t}$.

Assumption A4. Forecast-quality learning and trust complementarity. Forecast quality is decreasing in the realized forecast error, $q_t = Q(e_t)$, where $e_t \equiv |c^F_{t|t-1} - c_t|$ and $Q'(e_t) < 0$. Trust is increasing in accumulated forecast quality, $\partial \mathcal{T}_t / \partial \bar{Q}^N_t > 0$. If households update trust using the most recently verified forecast quality, then $\mu > 0$ implies $\kappa^N_0 > 0$. On the relevant margin, trust strengthens the behavioral response to forecast information, $\frac{\partial}{\partial \mathcal{T}_t}\left(\frac{da^F_t}{dc_t}\right) > 0$. This complementarity captures the idea that more credible forecasts receive greater behavioral weight, so improvements in realized forecast quality increase the value of forecast access for avoidance decisions.

Assumption A5. Trust persistence and credibility attenuation. The learning weight κ^N_0 determines how strongly the accumulated-quality stock responds to the most recently verified forecast-quality signal. A larger κ^N_0 makes trust more sensitive to short-run forecast-quality fluctuations, while a smaller κ^N_0 makes trust more persistent. Conditional on accumulated forecast quality, the mapping from quality to trust does not directly depend on the short-run learning weight: $\frac{\partial^2 \mathcal{T}_t}{\partial \kappa^N_0 \partial \bar{Q}^N_t} = 0$. Greater trust persistence strengthens the credibility component of the marginal avoidance response to pollution under forecast access. Equivalently, a larger short-run learning weight attenuates the marginal behavioral role of accumulated forecast credibility: $\frac{\partial^2}{\partial \kappa^N_0 \partial \mathcal{T}_t}\left(\frac{da^F_t}{dc_t}\right) \leq 0$.

Prediction 1A. (Decision-relevant signals increase avoidance) The Lagrangian of the single-period household problem is

$$\mathcal{L} = U_t(x_t, h(\tilde{c}_t, a_t)) + \lambda_t \left[I_t + w_t g(h(\tilde{c}_t, a_t)) - x_t - p_a a_t \right],$$

where λ_t is the multiplier on the budget constraint. Let $f(a_t, \tilde{c}_t)$ denote the first-order condition with respect to avoidance:

$$f(a_t, \tilde{c}_t) = \left[U_{h,t} + \lambda_t w_t g_h \right] h_a - \lambda_t p_a = 0.$$

By the implicit function theorem, the optimal avoidance rule $a_t^* = a_t(\tilde{c}_t)$ satisfies

$$\frac{da_t^*}{d\tilde{c}_t} = -\frac{\partial f/\partial \tilde{c}_t}{\partial f/\partial a_t} = -\frac{[U_{hh,t} + \lambda_t w_t g_{hh}]\, h_{\tilde{c}} h_a + [U_{h,t} + \lambda_t w_t g_h]\, h_{a\tilde{c}}}{[U_{hh,t} + \lambda_t w_t g_{hh}]\, h_a^2 + [U_{h,t} + \lambda_t w_t g_h]\, h_{aa}}.$$

All derivatives are evaluated at the optimum. Under Assumptions A1 and A2, the numerator is positive and the denominator is negative. Hence, $\frac{da_t^*}{d\tilde{c}_t} > 0$. Hence, a worse decision-relevant pollution signal increases optimal avoidance.

Forecast access provides an application of this result. After forecasts become available, households form the forecast-adjusted signal $\hat{c}_t$. If forecast information weakly raises perceived pollution risk relative to current information alone, $\hat{c}_t \geq c_t$, then the monotonicity of $a_t(\cdot)$ implies

$$a_t(\hat{c}_t) \geq a_t(c_t).$$

Prediction 1B. (Higher adjustment costs attenuate responses to a common decision-relevant signal) A simple cost-of-action implication follows by allowing the avoidance cost p_a to vary across household types, while holding the decision-relevant signal $\hat{c}_t$ fixed across types. The first-order condition can then be written as $f(a_t, \tilde{c}_t; p_a) = 0$, where p_a is treated as a parameter. By the implicit function theorem,

$$\frac{\partial a_t^*}{\partial p_a} = -\frac{f_{p_a}}{f_a} = \frac{\lambda_t}{f_a} < 0,$$

since $f_{p_a} = -\lambda_t$, $\lambda_t > 0$, and the second-order condition implies $f_a < 0$. Thus, a higher avoidance cost lowers the optimal level of avoidance.

The effect of p_a on the responsiveness to the signal is not signed by concavity alone. As a results, We impose a local monotonicity condition: the signal response, $-\frac{f_{\tilde{c}}}{f_a}$, is weakly increasing in the avoidance level a_t, so that $\partial(-f_{\tilde{c}}/f_a)/\partial a_t \geq 0$, and any income-effect term through λ_t does not overturn this relationship. Because a higher p_a lowers a_t^*, this condition implies

$$\frac{\partial^2 a_t^*}{\partial \tilde{c}_t \partial p_a} = \frac{\partial}{\partial p_a}\left(-\frac{f_{\tilde{c}}}{f_a}\right) = \frac{\partial}{\partial a_t}\left(-\frac{f_{\tilde{c}}}{f_a}\right)\frac{\partial a_t^*}{\partial p_a} \leq 0.$$

Hence, higher adjustment costs attenuate the avoidance response to the same decision-relevant pollution signal. This provides a cost-of-action interpretation for heterogeneous responses to decision-relevant signals across households facing different adjustment constraints.

Prediction 2. (Short-run forecast errors weaken the marginal avoidance response) Un-

der Assumptions A1–A4, larger short-run forecast errors reduce forecast quality, weaken trust, and attenuate the marginal avoidance response to pollution under forecast access. Specifically, since e_t affects da_t^F/dc_t through q_t, $\bar{Q}_t^N$, and $\mathcal{T}_t$, the chain rule gives

$$\frac{\partial}{\partial e_t}\left(\frac{da_t^F}{dc_t}\right) = \frac{\partial}{\partial \mathcal{T}_t}\left(\frac{da_t^F}{dc_t}\right)\frac{\partial \mathcal{T}_t}{\partial \bar{Q}_t^N}\kappa_0^N Q'(e_t) < 0.$$

The inequality follows from trust complementarity, $\frac{\partial}{\partial \mathcal{T}_t}\left(\frac{da_t^F}{dc_t}\right) > 0$, trust increasing in accumulated forecast quality, $\frac{\partial \mathcal{T}_t}{\partial \bar{Q}_t^N} > 0$, positive learning weight, $\kappa_0^N > 0$, and forecast quality decreasing in forecast error, $Q'(e_t) < 0$. Hence, a larger realized forecast error lowers measured forecast quality, weakens trust in the forecast system, and reduces the pollution-induced avoidance response under forecast access.

Prediction 3. (Accumulated forecast accuracy and persistent credibility strengthen avoidance responses) Taking the partial derivative of the marginal avoidance response to pollution under forecast access with respect to accumulated forecast quality yields:

$$\frac{\partial}{\partial \bar{Q}_t^N}\left(\frac{da_t^F}{dc_t}\right) = \frac{\partial}{\partial \mathcal{T}_t}\left(\frac{da_t^F}{dc_t}\right)\frac{\partial \mathcal{T}_t}{\partial \bar{Q}_t^N}.$$

By Assumption A4, trust strengthens the marginal avoidance response under forecast access, and accumulated forecast quality increases trust. Hence, $\frac{\partial}{\partial \bar{Q}_t^N}\left(\frac{da_t^F}{dc_t}\right) > 0$, which implies that higher accumulated forecast quality strengthens the pollution-induced avoidance response under forecast access. The magnitude of this accumulated-accuracy effect depends on the weight placed on the most recently verified forecast-quality signal, κ_0^N. Differentiating with respect to κ_0^N gives:

$$\frac{\partial}{\partial \kappa_0^N}\left[\frac{\partial}{\partial \bar{Q}_t^N}\left(\frac{da_t^F}{dc_t}\right)\right] = \frac{\partial^2}{\partial \kappa_0^N \partial \mathcal{T}_t}\left(\frac{da_t^F}{dc_t}\right)\frac{\partial \mathcal{T}_t}{\partial \bar{Q}_t^N} + \frac{\partial}{\partial \mathcal{T}_t}\left(\frac{da_t^F}{dc_t}\right)\frac{\partial^2 \mathcal{T}_t}{\partial \kappa_0^N \partial \bar{Q}_t^N}.$$

By Assumption A5, κ_0^N does not directly affect the marginal mapping from accumulated forecast quality to trust, so $\frac{\partial^2 \mathcal{T}_t}{\partial \kappa_0^N \partial \bar{Q}_t^N} = 0$. Assumption A5 also implies that greater short-run sensitivity weakly attenuates the marginal behavioral role of accumulated forecast credibility: $\frac{\partial^2}{\partial \kappa_0^N \partial \mathcal{T}_t}\left(\frac{da_t^F}{dc_t}\right) \leq 0$. Together with $\frac{\partial \mathcal{T}_t}{\partial \bar{Q}_t^N} > 0$ by Assumption A4, this implies:

$$\frac{\partial}{\partial \kappa_0^N}\left[\frac{\partial}{\partial \bar{Q}_t^N}\left(\frac{da_t^F}{dc_t}\right)\right] \leq 0.$$

Thus, a larger κ_0^N weakens the marginal effect of accumulated forecast accuracy on the

avoidance response, with strict attenuation when the trust-persistence channel is strict. Equivalently, avoidance responses are stronger when forecast credibility is driven more by persistent accumulated accuracy and less by short-run forecast-quality fluctuations.[2]

A.3 Two-Period Extension

We extend the static model to a two-period setting to show how next-day forecast information affects current avoidance and the timing of household shopping. This extension illustrates a forecast mechanism within the covered information regime rather than separately identifying its contribution to the overall coverage effect. The household allocates a fixed planned shopping amount $\bar{x}$ across adjacent days t and $t+1$. Under forecast access, the household solves:

$$\begin{aligned} \max_{x_t, x_{t+1}, a_t, a_{t+1}} \quad & U_t(x_t, h_t) + \mathbb{E}_t^F \left[U_{t+1}(x_{t+1}, h_{t+1}) \right] \\ \text{s.t.} \quad & I_s + w_s g(h_s) \geq x_s + p_a a_s, \qquad s \in \{t, t+1\}, \\ & h_s = h(\tilde{c}_s, a_s), \qquad s \in \{t, t+1\}, \\ & x_t + x_{t+1} = \bar{x}. \end{aligned}$$

Under forecast access, the day t decision-relevant signal is $\tilde{c}_t = \hat{c}_t$, where $\hat{c}_t = \hat{c}(c_t, c^F_{t+1|t}, \mathcal{T}_t)$ combines current pollution, the next-day forecast, and forecast trust. The expectation reflects that, on day t, households observe the next-day forecast but not realized day $t+1$ pollution. The final constraint introduces the intertemporal margin, allowing planned shopping to shift across adjacent days.[3]

Assumption A6. Two-period timing and next-day forecast information. For the next-day forecast comparative static, current pollution information, the day t trust stock, and other information are held fixed. The two-period problem has an interior local optimum. The next-day forecast signal enters the current forecast-adjusted perceived pollution index with positive weight, $\partial \hat{c}_t / \partial c^F_{t+1|t} > 0$. The marginal net benefit of current avoidance is locally decreasing in current avoidance and locally increasing in the forecast-adjusted perceived pollution index. Planned shopping has locally diminishing marginal utility on both days, $U_{xx,t} < 0$ and $\mathbb{E}_t^F[U_{xx,t+1}] < 0$. Let λ_{t+1} denote the day $t+1$ budget mul-

[2] Empirically, we vary κ_0^N by changing the half-life of the geometric forecast-quality weights. A longer half-life lowers the weight on the most recent forecast-quality realization and increases the persistence of accumulated forecast credibility.

[3] We do not impose a parametric form on $\hat{c}_t$. Instead, we only require that it responds locally to the next-day forecast, $c^F_{t+1|t}$. The index affects choices through perceived health risk in the ex ante decision problem and does not redefine ex post realized health, which depends on realized pollution.

tiplier, evaluated under the day t information set. A higher next-day forecast pollution signal lowers the expected net marginal value of planned shopping on day $t+1$:

$$\frac{\partial}{\partial c^F_{t+1|t}} \mathbb{E}^F_t \left[U_{x,t+1}(x_{t+1}, h_{t+1}) - \lambda_{t+1} \right] < 0.$$

Prediction 4. (High next-day forecasted pollution increases current avoidance and shifts shopping forward) The Lagrangian of the two-period household problem is:

$$\begin{aligned}
\mathcal{L} = {} & U_t\left(x_t, h(\tilde{c}_t, a_t)\right) + \mathbb{E}^F_t \left[U_{t+1}\left(x_{t+1}, h(\tilde{c}_{t+1}, a_{t+1})\right) \right] \\
& + \lambda_t \left[I_t + w_t g\left(h(\tilde{c}_t, a_t)\right) - x_t - p_a a_t \right] \\
& + \mathbb{E}^F_t \left\{ \lambda_{t+1} \left[I_{t+1} + w_{t+1} g\left(h(\tilde{c}_{t+1}, a_{t+1})\right) - x_{t+1} - p_a a_{t+1} \right] \right\} \\
& + \nu \left(\bar{x} - x_t - x_{t+1} \right).
\end{aligned}$$

Under forecast access, the day t decision-relevant signal is $\tilde{c}_t = \hat{c}_t$. Let $f_t(a_t, \hat{c}_t)$ denote the first-order condition with respect to current avoidance:

$$f_t(a_t, \hat{c}_t) = \left[U_{h,t} + \lambda_t w_t g_h \right] h_a(\hat{c}_t, a_t) - \lambda_t p_a = 0.$$

By the implicit function theorem,

$$\frac{\partial a^F_t}{\partial \hat{c}_t} = -\frac{\partial f_t / \partial \hat{c}_t}{\partial f_t / \partial a_t}.$$

By Assumption A6, a higher forecast-adjusted perceived pollution index raises the marginal net benefit of current avoidance, so $\partial f_t / \partial \hat{c}_t > 0$. The second-order condition implies $\partial f_t / \partial a_t < 0$. Hence, $\frac{\partial a^F_t}{\partial \hat{c}_t} > 0$. Since the next-day forecast signal increases the current forecast-adjusted perceived pollution index, $\partial \hat{c}_t / \partial c^F_{t+1|t} > 0$, the chain rule gives

$$\frac{\partial a^F_t}{\partial c^F_{t+1|t}} = \frac{\partial a^F_t}{\partial \hat{c}_t} \frac{\partial \hat{c}_t}{\partial c^F_{t+1|t}} > 0.$$

Thus, a higher forecasted pollution level for day $t+1$ increases current avoidance.

The same next-day forecast signal also shifts planned shopping toward day t. Let λ_s denote the budget multiplier for day $s \in \{t, t+1\}$, as in Assumption A6, and let ν be the multiplier on the intertemporal shopping constraint, $x_t + x_{t+1} = \bar{x}$. Define the net marginal value of planned shopping on day s as $M_s(x_s, h_s, \lambda_s) \equiv U_{x,s}(x_s, h_s) - \lambda_s$. Combining the first-order conditions for x_t and x_{t+1} and using $x_{t+1} = \bar{x} - x_t$, the timing

condition can be written as

$$G(x_t, c^F_{t+1|t}; c_t, \mathcal{T}_t) \equiv M_t(x_t, h_t, \lambda_t) - \mathbb{E}^F_t \left[M_{t+1}(\bar{x} - x_t, h_{t+1}, \lambda_{t+1}) \right] = 0.$$

Differentiating with respect to x_t gives $G_{x_t} = U_{xx,t} + \mathbb{E}^F_t[U_{xx,t+1}] < 0$, where the inequality follows from diminishing marginal utility of planned shopping on both days. Assumption A6 also implies $\frac{\partial}{\partial c^F_{t+1|t}} \mathbb{E}^F_t \left[M_{t+1}(\bar{x} - x_t, h_{t+1}, \lambda_{t+1}) \right] < 0$. Therefore,

$$G_{c^F_{t+1|t}} = -\frac{\partial}{\partial c^F_{t+1|t}} \mathbb{E}^F_t \left[M_{t+1}(\bar{x} - x_t, h_{t+1}, \lambda_{t+1}) \right] > 0.$$

By the implicit function theorem,

$$\frac{\partial x^F_t}{\partial c^F_{t+1|t}} = -\frac{G_{c^F_{t+1|t}}}{G_{x_t}} > 0.$$

Thus, a higher forecasted pollution level for day $t+1$ shifts planned shopping toward day t.

A.4 Value of Information

This subsection defines the realized value of forecast information in the static model. Let $V_t(c_t, \tilde{c}_t)$ denote realized indirect utility when actual pollution is c_t and avoidance is chosen using the decision-relevant pollution signal $\tilde{c}_t$:

$$V_t(c_t, \tilde{c}_t) = U_t \left[x_t(c_t, \tilde{c}_t), h\left(c_t, a_t(\tilde{c}_t)\right) \right].$$

The value of moving from baseline information to the combined current and forecast information under coverage can be decomposed:

$$VOI^F_t = V_t(c_t, \hat{c}_t) - V_t(c_t, c_{0t}) = \underbrace{V_t(c_t, c_t) - V_t(c_t, c_{0t})}_{VOI^{RT}_t} + \underbrace{V_t(c_t, \hat{c}_t) - V_t(c_t, c_t)}_{VOI^{F|RT}_t}.$$

The first bracket, VOI^{RT}_t, is the value of the current information only regime relative to baseline perceived information. The second bracket, $VOI^{F|RT}_t$, is the incremental value of adding forecasts to that regime. The coverage design identifies their combined value, $VOI^F_t = V_t(c_t, \hat{c}_t) - V_t(c_t, c_{0t})$. To connect this decomposition to the standard value-of-information expression, consider a local change in the decision signal $\tilde{c}_t$ holding actual

pollution c_t fixed. Differentiating realized indirect utility with respect to $\tilde{c}_t$ gives:

$$\frac{\partial V_t(c_t, \tilde{c}_t)}{\partial \tilde{c}_t} = \{[U_{h,t} + \lambda_t w_t g_h] h_a - \lambda_t p_a\} a_t'(\tilde{c}_t),$$

where the terms inside the braces are evaluated at $x_t(c_t, \tilde{c}_t)$ and $h(c_t, a_t(\tilde{c}_t))$, with $\lambda_t = U_{x,t}$ at the relevant allocation. The expression in braces is the marginal net benefit of avoidance: the utility value of the health gain, plus the effective-income value of improved health, net of the resource cost of avoidance.

The realized value of information under forecast coverage can be approximated locally by a first-order expansion around the baseline signal c_{0t}:

$$VOI_t^F = V_t(c_t, \hat{c}_t) - V_t(c_t, c_{0t}) = \frac{\partial V_t(c_t, c_{0t})}{\partial \tilde{c}_t}(\hat{c}_t - c_{0t}) + o(|\hat{c}_t - c_{0t}|).$$

Hence, the local approximation depends on the marginal welfare effect of changing the decision-relevant signal evaluated at the baseline and the signal gap $\hat{c}_t - c_{0t}$. Assumption A3 implies a weakly positive signal gap on the relevant margin, but does not by itself imply positive realized welfare: a higher signal can induce costly avoidance that exceeds its realized health benefit. More generally, the sign of realized information value is not determined by the direction of the signal change alone. Under rational use of an expanded information set that can be costlessly ignored, coverage has a weakly nonnegative ex ante value of information relative to the baseline. This ex ante statement does not imply positive realized value on every realization. The two-period extension illustrates the additional forward-looking mechanism through which forecast information can affect shopping timing.

B Institutional Background and Data Construction

B.1 Institutional Background of AirNow

This appendix documents the institutional history of the AirNow forecast program, focusing on federal program design, federal funding cycles, and agency technical capacity that affected the timing of forecast coverage.

First, AirNow's early expansion was supported by federal directive and targeted a pre-specified list of metropolitan areas defined by population, not local demand. This initiative, the Environmental Monitoring for Public Access and Community Tracking (EMPACT) program, was created by presidential directive in 1996 and targeted to the 86 largest U.S. metropolitan areas.[4] EMPACT funding enabled participating areas to invest in the infrastructure needed to collect and disseminate environmental information. Thus, the initial geographic scope of federal support was defined *ex ante* by metropolitan population rather than observed household responses to pollution.[5]

Second, federal support was disbursed in discrete federal award waves, with timing determined by the federal budget cycle. The first EMPACT award wave, announced in October 1998, provided approximately $3.5 million to eight communities.[6] The second wave, announced in December 1999, provided around $3.4 million to seven additional communities, with individual awards continued into 2000.[7] These discrete federal funding waves provide one source of administrative variation in when local agencies acquired the infrastructure needed to participate in public environmental-information programs.

Third, contemporaneous federal reviews attribute the expansion of forecast coverage to national top-down support. The interagency Committee on Environment and Natural Resources identifies three principal federal contributions: EMPACT funding, forecasting guidance to state and local agencies, and EPA-supported dissemination through website development and partnerships with external weather providers. The report explicitly attributes growth in the number of U.S. forecast areas to this national "top-down" sup-

[4]Source: U.S. Environmental Protection Agency. Grants Announced to Local Communities to Access Environmental Information. News release, December 16, 1999.

[5]Consistent with this population-based ordering, reporting areas with larger 2000 populations entered $PM_{2.5}$ forecast coverage earlier. We use 2000 Census population because 5-digit ZIP Code Tabulation Areas (ZCTA5s) were introduced with the 2000 Census and map more consistently to the ZIP codes observed in our 2000s data than the ZIP-code geography available for 1990. Regressing entry year on 2000 Census rank (1 = largest), a 100-position lower rank is associated with entry about 1.1 years later (coefficient = 0.011, robust s.e. = 0.001, adjusted R^2 = 0.57, N = 516.)

[6]Source: U.S. Environmental Protection Agency. Vice President Announces EPA Grants to Local Communities to Access Environmental Information. News release R-149, October 30, 1998.

[7]Source: U.S. Environmental Protection Agency. Grants Announced to Local Communities to Access Environmental Information. News release, December 16, 1999.

port.[8] At the same time, EPA compiled and disseminated forecasts generated by state and local air agencies, with forecasting authority remaining at the agency level. In turn, forecast participation depended on decisions and capabilities of technical agencies.

Fourth, technical capacity was an explicit constraint on local forecasting. The federal review proposed a national forecast model specifically to assist areas that lacked the capability to generate their own forecasts[9], while EPA guidance shows that forecasting methods required different levels of development and operational effort, technical expertise, and computing resources.[10] These capacity requirements provide a direct institutional reason for staggered adoption across agencies that is distinct from local household demand for forecasts.

Overall, the institutional background suggests that variation in forecast timing reflects centralized federal support, which targeted areas by population, and heterogeneous agency capacity rather than local household demand or pre-existing avoidance behavior. Because this evidence alone cannot rule out endogenous rollout, we complement it with direct empirical tests showing that coverage timing is unrelated to pre-treatment pollution, household characteristics, and the pre-existing shopping-pollution gradient, and that treated areas exhibit no differential pre-trends in that gradient prior to forecast entry.

B.2 PRISM Climate Data

Daily climate data for the contiguous United States from 2004 to 2019 come from the PRISM Climate Group at Oregon State University. PRISM is widely used in environmental economics (Addoum, Ng, and Ortiz-Bobea, 2023; Baylis et al., 2025; Lee and Zheng, 2025) and provides high-resolution gridded estimates that combine station observations with local topography to capture spatially heterogeneous climate conditions on a 4 km grid. We use daily precipitation, minimum temperature, and maximum temperature.

To align PRISM with our county-day unit of analysis, we aggregate gridded values to the county level using zonal statistics. For each day and variable, we compute the mean across PRISM grid cells whose centroids fall within county boundaries, excluding cells with PRISM's "no data" value of -9999. The resulting dataset contains county-day precipitation in millimeters and minimum and maximum temperature in degrees

[8]Source: Air Quality Research Subcommittee, Committee on Environment and Natural Resources, Air Quality Forecasting: A Review of Federal Programs and Research Needs, June 2001, p. 9.

[9]Source: Air Quality Research Subcommittee, Committee on Environment and Natural Resources, Air Quality Forecasting: A Review of Federal Programs and Research Needs, June 2001, p. 9.

[10]Source: U.S. Environmental Protection Agency. Guidelines for Developing an Air Quality (Ozone and $PM_{2.5}$) Forecasting Program. June 2003.

Celsius, which serve as weather controls in the main specifications.

In the empirical specifications, we flexibly control for temperature by constructing 17 bins separately for daily maximum and minimum temperature. These consist of one bin for temperature below -15°C, one for temperatures above 30°C, and 15 intermediate bins spanning three degrees each. We also include decile indicators for precipitation.

B.3 NARR Wind Data

Wind data come from the North American Regional Reanalysis dataset, provided by the National Centers for Environmental Prediction (NCEP). NARR provides a consistent, model-assimilated record of atmospheric conditions at approximately 32 km spatial resolution. We use daily eastward (U) and northward (V) wind components at 10 meters above ground level from 2004 to 2019.

We construct county-day wind measures used in our instrumental-variable construction in two steps. First, we assign each AQS monitor to its nearest NARR grid cell and extract daily U and V components at that location to obtain monitor-day wind vectors. Second, we aggregate monitor-day wind vectors to the county level by averaging U and V across monitors within each county.

From county-day U and V components, we compute wind speed and wind direction, where c denotes counties and t denotes days. Wind speed, WS_{ct}, is derived using the Pythagorean theorem:

$$WS_{ct} = \sqrt{U_{ct}^2 + V_{ct}^2}.$$

We also compute meteorological wind direction from the U and V components, following the standard convention that wind direction is measured as the direction from which wind originates, with 0 degrees corresponding to winds from the north. For the upwind instrument, we convert this meteorological wind direction into the wind transport direction, defined as the direction toward which wind blows. This procedure yields a county-by-day panel of wind speed, meteorological wind direction, and the corresponding wind transport direction used in the instrumental-variable construction.

B.4 NielsenIQ Ailment Data

We obtain household health information from the NielsenIQ Annual Ailments, Health, and Wellness Survey, an annual supplement to the Consumer Panel fielded for a subsample of panelists since 2011. We classify households as vulnerable to air pollution if they report any respiratory disease, cardiovascular condition, hypertension, or outdoor

allergy associated with heightened sensitivity to air pollution.[11] Households with any of these conditions may face different benefits and costs of avoiding pollution than households without them (Lee et al., 2024; Li et al., 2022; Rajagopalan and Landrigan, 2021; Zhao et al., 2022).

[11] Ailments include outdoor allergies; asthma; chronic bronchitis, pulmonary disease, COPD, or emphysema; heart disease, heart attack, angina, or heart failure; and high blood pressure or hypertension.

C Instrumental Variable Construction

C.1 Main Instrumental Variable

The instrumental variable, UP_{ct}, captures pollution transported from neighboring counties that are upwind of focal county c on day t. We define two angles to determine whether a neighboring county is upwind. First, $\rho_{c,j}$ is the geographic bearing from the centroid of a neighboring county j to the centroid of the focal county c. Second, ζ_{jt} is the wind transport direction in neighboring county j on day t, measured as the direction toward which the wind blows. For example, wind blowing from southwest to northeast has $\zeta_{jt} = 45°$.

To isolate the upwind component, we apply a cosine filter. The angular difference, $(\rho_{c,j} - \zeta_{jt})$, measures the alignment between the direction from county j to county c and the realized wind transport direction in county j. The filter is defined as:

$$\cos \psi_{c,j,t} = \max\{\cos(\rho_{c,j} - \zeta_{jt}),\, 0\}.$$

This filter is positive only when county j lies upwind of county c given the realized wind direction (i.e., when the angular difference is acute, less than 90° in either direction). If county j is downwind or perfectly crosswind relative to county c, the cosine term is negative or zero and the filter assigns zero weight. This ensures that only pollution from upwind source counties contributes to the instrument.

As illustrated in Figure 3, a neighboring county contributes to the instrument only when the direction from that county to the focal county aligns with the realized wind transport direction. In the example, $C_{i,1}$ is not upwind of the focal county C_i because the bearing from $C_{i,1}$ to C_i is perpendicular to the wind transport direction. Its cosine term is zero, so it contributes nothing to the instrument. By contrast, $C_{i,2}$ lies upwind of C_i, yielding a positive cosine term.

The final instrumental variable is the sum of these filtered, distance-weighted pollution measures from all potential source counties j whose centroids are 100–300 km from the focal county c:

$$\text{UP}_{ct} = \sum_{j \neq c} \frac{P_{jt} \cdot \text{WS}_{jt}}{D_{c,j}^2} \cdot \cos \psi_{c,j,t}.$$

Here, P_{jt} is the pollution level and WS_{jt} is the wind speed in the neighboring county j, and $D_{c,j}^2$ is the squared distance from neighboring county j to focal county c.

C.2 Alternative Instrumental Variable

We construct an alternative instrumental-variable for $PM_{2.5}$ in the spirit of Deryugina et al. (2019). This approach uses daily wind direction as an instrument for pollution and allows the effect of wind direction on $PM_{2.5}$ to vary across space. It provides a robustness check on our main upwind-transport instrument.

The first-stage specification for this alternative IV is as follows:

$$\text{PM2.5}_{ct} = \sum_{g \in G} \sum_{b=0}^{2} \beta_b^g \mathbf{1}[G_c = g] \times \text{WINDDIR}_{ct}^{90b} + \mathbf{\Omega}_{ct}' \delta_1 + \mathbf{X}_{ict}' \delta_2 + FEs + \epsilon_{ict}.$$

In this specification, the excluded instruments are the interaction terms $\mathbf{1}[G_c = g] \times \text{WINDDIR}_{ct}^{90b}$. The term $\text{WINDDIR}_{ct}^{90b}$ is a set of indicator variables where each is equal to 1 if the average daily wind direction in county c falls into the 90-degree interval $[90b, 90b + 90)$, and zero otherwise. The omitted category is the northwesterly quadrant, $[270, 360)$. We classify all counties into one of 10 spatial groups ($g \in G$) using a k-means clustering algorithm based on their geographic location. The interaction with $\mathbf{1}[G_c = g]$ allows the first-stage effect of wind direction to vary across regions, reflecting the idea that winds from the same direction can have different pollution implications depending on local source locations and transport patterns. Given the size of our household-day panel, we limit the analysis to 10 groups for computational feasibility. Appendix Figure F.10 maps these spatial groups. The coefficients of interest, β_b^g, capture region-specific effects of wind direction on $\text{PM}_{2.5}$.

The alternative IV specification uses the same controls and fixed effects as the main IV model, including weather controls, holiday indicators, household demographics, household fixed effects, day-of-week fixed effects, week-by-year fixed effects, and county-by-year fixed effects. The second stage proceeds as in the main analysis, using predicted $\text{PM}_{2.5}$ from the wind-direction first stage. This parallel structure allows us to assess whether the main findings are sensitive to the specific construction of the upwind-transport instrument.

D Additional Analyses

D.1 Public Forecast Access and Awareness

To assess whether AirNow forecast coverage increased public attention to air quality information, we use Google Trends search data as a proxy for public awareness and engagement. Google Trends reports a normalized index of search interest for a given term within a specified geography and period, scaled from 0 to 100 relative to peak search volume in that window. We focus on three air-quality-related terms: "AirNow," "air quality," and "smoke." These searches capture active information acquisition about ambient pollution conditions.

We construct monthly search indices for six DMAs (i.e., Amarillo, TX; Duluth-Superior, MN-WI; Gainesville, FL; Charleston, SC; Augusta, GA; and Columbia, SC). For each DMA, we use a symmetric six-month window centered on its majority-treatment date, defined as the earliest date by which at least 60% of ZIP codes in the DMA had gained AirNow forecast coverage. We use the DMA, rather than the county, as the unit of analysis because Google Trends reports search intensity at the media-market level, making it the finest consistent geography available. The 60% threshold ensures that the treatment date captures a meaningful DMA-wide expansion rather than isolated adoption. We restrict attention to DMAs that received AirNow coverage after 2015, when Google Trends data are more reliable and consistent.

The analysis is descriptive. We compare search interest before and after forecast coverage, using both the raw search index and a pollution-adjusted measure. To construct the adjusted measure, we residualize monthly search interest with respect to DMA-level mean monthly $PM_{2.5}$. The raw comparison provides an initial indication of whether search activity changed around forecast rollout, but it may confound the effect of coverage with changes in underlying air quality. If $PM_{2.5}$ levels were higher after rollout, households may have searched more simply because pollution worsened. To address this concern, we residualize the monthly search index with respect to DMA-level mean monthly $PM_{2.5}$ and use the resulting residual as an adjusted measure of search interest. This approach isolates variation in search activity that is not mechanically explained by contemporaneous pollution levels.

Figure F.3 presents the results. Panel A shows that raw search interest is higher after forecast coverage begins. The median rises from about 17.0 in the pre-coverage period to about 18.5 in the post-coverage period, and the interquartile range shifts upward as well, suggesting that the increase is broad-based rather than driven by a few outlying months. However, $PM_{2.5}$ levels are also somewhat higher after rollout in this sample, so

part of this increase could reflect changes in air quality rather than forecast access itself.

Panel B addresses this issue by plotting the distribution of search residuals after removing the component explained by monthly $PM_{2.5}$. The post-coverage distribution remains shifted upward: the median residual increases from roughly -0.8 before rollout to about 0.2 afterward, and the interquartile range also moves higher. This pattern suggests that search activity increased after AirNow forecast coverage even after accounting for ambient pollution, consistent with forecast coverage raising public awareness and information acquisition.

D.2 Pollution and Spending Composition

To complement the baseline shopping–pollution gradient, we examine which product categories drive the spending response. Table F.6 reports IV estimates by product category, focusing on necessities, non-necessities, and defensive goods. Defensive goods include products that reduce pollution exposure or mitigate short-run health consequences, such as masks, air purifiers and filters, breathing aids, and over-the-counter products for respiratory, cough, throat, and sinus symptoms. Necessities include food and health-related goods, while non-necessities include other non-food purchases. When pollution raises the perceived cost of in-person shopping, households may preserve essential purchases while deferring more discretionary ones (Barwick et al., 2024b; Deschênes, Greenstone, and Shapiro, 2017; Zhang and Mu, 2018).

The estimates show that non-necessity spending declines more than necessity spending. A 1 $\mu g/m^3$ increase in $PM_{2.5}$ reduces non-necessity expenditures by 0.24 percent relative to the mean, compared with 0.17 percent for necessities. This pattern is consistent with households maintaining core purchases while postponing more discretionary consumption. By contrast, the estimate for defensive goods is positive but small and statistically insignificant, suggesting that trip avoidance is the primary defensive margin. If households can reduce exposure by staying home and shifting purchases over a short horizon, they may have limited need to increase same-day spending on protective goods. The imprecise defensive-goods estimate may also reflect incomplete coverage of these purchases in our data, since some relevant products or outlets may not be fully observed.

D.3 The Role of Private Information

Besides public pollution information, households may also rely on private cues to assess local conditions and adjust their behavior. As in Barwick et al. (2024a), we examine

whether air visibility serves as an informative private signal of pollution. Panel A of Figure F.11 in the appendix presents the distribution of visibility. For the vast majority of stations, visibility is top coded at 10 miles (around 16 km).

We have three main findings. First, Panel B of Figure F.11 shows that visibility is negatively related to realized $PM_{2.5}$ on average, but the mapping is noisy, especially at low visibility levels. This dispersion is consistent with the fact that poor visibility can reflect multiple conditions, making it difficult to infer pollution levels from visibility alone.

Second, we re-estimate the baseline specification by replacing realized pollution with visibility-bin indicators. This is a conditional-correlation exercise rather than an instrumental variables specification. We estimate spending responses to indicators for visibility percentile bins, using the 100th-percentile bin, corresponding to the best visibility days, as the omitted category. Panel C shows that, before forecast access begins, the spending–visibility relationship is economically modest, imprecisely estimated, and not strongly monotone across bins.

Third, we add measured pollution controls to the visibility regressions to test whether the spending–visibility relationship reflects pollution avoidance. Specifically, we add contemporaneous realized $PM_{2.5}$ to the visibility-bin specification, instrumenting $PM_{2.5}$ with the upwind IV as in the main analysis. If households use visibility to infer pollution risk, conditioning on $PM_{2.5}$ should attenuate the visibility coefficients. In the pre-coverage period, however, the coefficients change little (Panel C), suggesting that low-visibility responses are not primarily driven by pollution and may instead reflect visibility itself, such as travel or safety concerns. By contrast, in the post-coverage period, the visibility coefficients attenuate once we control for pollution (Panel D). These findings suggest that visibility is an imperfect and noisy private signal. Consistent with Barwick et al. (2024a), we find limited evidence that households used visibility as an effective substitute for public forecast information before AirNow forecasts became available.

E Appendix Robustness Checks

To ensure that our main findings are robust and not driven by particular model choices, we conduct a series of sensitivity checks. We first show that our results are stable across alternative functional forms, fixed-effects structures, and standard-error clustering choices. We next demonstrate that the estimates are not biased by omitting other pollutants, such as ozone and carbon monoxide. Finally, we assess sensitivity to our identification strategy by varying the construction of our primary upwind instrumental variable and, importantly, by replacing it with an alternative wind-direction IV.

E.1 Alternative Functional Forms

We test the robustness of our findings to alternative functional forms in Appendix Table F.8, given that our outcomes have different distributional features that can affect inference under linear specifications. Trip probability is binary, while trip counts and expenditures are nonnegative, right-skewed, and contain many zeros. Accordingly, we estimate a probit model for trip probability and re-estimate the trip-frequency and spending outcomes using the inverse hyperbolic sine transformation, the log-plus-one transformation, and a Poisson specification. Across these alternatives, the estimated effect remains negative and statistically significant. Because transformed-outcome and nonlinear estimates are not directly comparable to the baseline level estimates, we focus on the qualitative pattern rather than exact magnitudes. The results consistently show that higher $PM_{2.5}$ reduces trip-taking and spending, suggesting that the main findings are not driven by the choice of functional form.

E.2 Different Sets of Fixed Effects

We assess the robustness of our results by re-estimating the model with alternative fixed-effects structures that vary the degree to which we absorb unobserved temporal and regional confounders. Appendix Table F.9 shows that the estimated effect of $PM_{2.5}$ remains negative and statistically significant when we either relax the baseline specification by using coarser time or location fixed effects or strengthen it by adding more granular time-by-geography controls. The magnitude is also similar across specifications. These checks indicate that the main estimates are not an artifact of a particular fixed-effects choice and are not driven by overfitting to highly localized variation.

E.3 Different Clustering Levels

In our baseline specification, we cluster standard errors at the county level, a natural choice given that $PM_{2.5}$ is measured at the county-day level. To assess whether inference is sensitive to this choice, we re-estimate standard errors using three two-way clustering schemes following Colin Cameron and Miller (2015): county and year, county and month-by-year, as well as county and week-by-year level. These alternatives allow for correlation within counties over time while also capturing common shocks at annual, monthly, or weekly horizons. The results in Appendix Table F.10 show that the point estimates are unchanged across clustering schemes. Although the trip-probability estimate loses significance under county and week-by-year clustering, the overall pattern remains stable, with somewhat larger standard errors under the more demanding two-way schemes.

E.4 Controls for Past-week Shopping Activities

Since many households shop on a regular weekly cycle, current shopping behavior may depend on recent purchases. Households that stocked up in the previous week may make fewer trips or spend less today, independent of current pollution conditions. To account for this short-run persistence, we add controls for past-week shopping activity. Specifically, we control for past-week shopping days, store visits, and total expenditure in the trip probability, store visits, and spending regressions, respectively. Appendix Table F.11 shows that the estimated effects of $PM_{2.5}$ remain negative and statistically significant across all outcomes, with magnitudes very close to the baseline.

E.5 Controls for Other Pollutants

Since ozone and carbon monoxide can be correlated with fine particulate matter and may be jointly driven by meteorological conditions and emission sources (Deryugina et al., 2019), we include them as additional controls to assess whether co-pollutant variation confounds the estimated $PM_{2.5}$ effects. Appendix Table F.12 reports these checks. Columns (1), (4), and (7) add ozone as a control, columns (2), (5), and (8) add carbon monoxide, and columns (3), (6), and (9) include both pollutants. Across all specifications, the estimated shopping–pollution gradients remain negative and statistically significant, with magnitudes similar to the baseline. Including ozone slightly increases the magnitude of the estimated effects, but the overall conclusions are unchanged. These results indicate that the main findings are not driven by omitted co-pollutant variation.

E.6 Sensitivity of the Upwind IV Construction

In our baseline specification, we define the upwind ring as 100–300 km from the focal county to avoid including neighboring counties that may share similar economic activities. As a robustness check, we vary the ring to alternative distance bands: 100–200 km, 150–300 km, and 300–450 km. Appendix Table F.13 reports the corresponding IV estimates. For the two shorter-distance definitions (100–200 km and 150–300 km), the estimated effects on trip probability and trip frequency remain negative and highly statistically significant across weighting schemes, and the expenditure estimates are also negative and precisely estimated. In contrast, when the ring is expanded to 300–450 km, the trip estimates become unstable and the expenditure effects become small and statistically insignificant. Consistent with the weaker first stage in this range, with Kleibergen-Paap F-statistics around 60–75, the 300–450 km definition provides less predictive power for local $PM_{2.5}$ and yields less stable estimates (Chen et al., 2021).

In addition, we test robustness to alternative weighting schemes in the instrument construction. Panel A of Table F.13 reports an unweighted version that uses the simple average across upwind counties, Panel B uses inverse-distance weights $1/D_{i,j}$, and Panel C uses the inverse-squared-distance weights $1/D_{i,j}^2$, which is our baseline weighting scheme. Across these weighting schemes, the results are very similar for the 100–200 km and 150–300 km distance bands, indicating that our conclusions are not sensitive to the specific distance-decay function used to aggregate upwind pollution.

E.7 Alternative Instrumental Variables

A potential concern is that our estimates may be sensitive to the particular construction of the upwind instrument. To address this, we implement an alternative IV strategy used by Deryugina et al. (2019) and Sun (2025) that instruments for local $PM_{2.5}$ using changes in wind direction (see Appendix C.2 for details). Appendix Table F.14 reports the results. The alternative instrument is strong, with a first-stage Kleibergen-Paap F-statistic of 91.[12] The estimated effects are also highly similar to our baseline IV results: A 1 $\mu g/m^3$ increase in $PM_{2.5}$ reduces trip probability by 0.08 percent, the number of store visits by 0.13 percent, and total expenditure by 0.17 percent. These magnitudes closely match the corresponding baseline estimates in Table 2 (0.07, 0.11, and 0.18 percent, respectively), indicating that our conclusions are robust to the choice of instrument and not driven by any single IV construction.

[12]The relatively lower first-stage F-statistics associated with the wind-direction instruments proposed by Deryugina et al. (2019) are also noted by Barwick et al. (2024b).

F Appendix Tables and Figures

F.1 Appendix Tables

Table F.1: Numbers and Operating Years Statistics for $PM_{2.5}$ Monitors

Monitor Type	N	Mean	Min	Max	Std. Dev.
All Monitors	4570	7.22	1	16	5.02
FRM Monitors	3045	6.85	1	16	5.03
Non-FRM Monitors	1525	7.95	1	16	4.91

Notes: This table reports summary statistics on the number and operating duration of $PM_{2.5}$ monitors used in the analysis. Federal Reference Method (FRM) monitors use gravimetric measurement and serve as the regulatory standard for $PM_{2.5}$ compliance monitoring. Non-FRM monitors employ alternative measurement methods, such as continuous optical or beta attenuation techniques, and are used primarily for supplemental monitoring. *N* denotes the number of monitors. Mean, Min, Max, and Std. Dev. refer to the distribution of operating years across monitors.

Table F.2: Additional Summary Statistics, 2004–2019

	Mean	Std. Dev.	N
Panel A. In-person Purchases Statistics			
Expenditure on Defensive Goods ($)	0.078	1.374	176,326,439
Expenditure on Necessities ($)	8.615	26.543	176,326,439
Expenditure on Non-necessities ($)	3.160	15.9576	176,326,439
Panel B. Online Purchases Statistics			
Trip Probability	0.0045	0.067	176,326,439
No. of Trips	0.0047	0.072	176,326,439
Shopping Expenditure ($)	0.201	4.866	176,326,439
Conditional Expenditure ($, non-zero)	44.605	57.092	796,988
Panel C. Household Vulnerability			
Poverty Status (=1)	0.043	0.204	580,181
Age vulnerability (=1)	0.376	0.484	580,181
Having Ailments (=1)	0.755	0.430	206,363

Notes: (1) "N" values reflect the total number of household-day observations in each category in Panels A and B and the number of household-year observations in Panel C. (2) Alcohol is classified as neither a necessity nor a non-necessity. (3) Conditional expenditure is calculated for the subsample of household-days with strictly positive in-person shopping expenditures. (4) Poverty Status: an indicator equal to one if a household is at or below the federal poverty line based on household income and size. (5) Age vulnerability: an indicator equal to one if a household head is aged 65 or older. (6) Having Ailments: an indicator equal to one if a household reports any ailment closely associated with air pollution.

Table F.3: First-Stage Regression Results

	Forecast Access		Same-day Forecast Error		Rolling Forecast Accuracy	
Dependent Var:	$PM_{2.5}$	$PM_{2.5} \times \mathbf{1}\{\text{Coverage}\}$	$PM_{2.5}$	$PM_{2.5} \times$ Error	$PM_{2.5}$	$PM_{2.5} \times$ Accur.
	(1)	(2)	(3)	(4)	(5)	(6)
Upwind IV	228.208***	-48.846***	190.476***	-145.598***	200.491***	4.526
	(16.234)	(5.334)	(10.484)	(9.906)	(10.712)	(2.874)
Upwind IV $\times \mathbf{1}\{\text{Coverage}\}$	-23.922	265.717***				
	(14.759)	(12.661)				
Upwind IV $\times$ Forecast Error			36.420***	736.365***		
			(8.902)	(42.021)		
Upwind IV $\times$ Forecast Accuracy					-107.139***	701.675***
					(25.825)	(34.960)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Wald *F*-statistic	203	220	167	157	191	202
Observations	176,326,439	176,326,439	82,162,849	82,162,849	106,885,045	106,885,045

Notes: This table reports first-stage estimates for the three specifications with multiple endogenous variables. The "Upwind IV" refers to the main instrument that exploits daily variation in the concentration of particulate matter transported toward a focal county by prevailing winds. Columns (1)–(2) correspond to the forecast access model (Specification 4), Columns (3)–(4) to the same-day forecast error specification (Specification 6), and Columns (5)–(6) to the rolling forecast accuracy model (Specification 7). All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. The Wald *F*-statistic tests the joint significance of the upwind IV and its interaction term. All reported *F*-statistics are statistically significant at the 1 percent level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.4: Diagnostics for AirNow Forecast Rollout Exogeneity

Dependent Var:	Coverage Indicator		Rollout-Year Prediction		Resid. $PM_{2.5}$
	(1)	(2)	(3)	(4)	(5)
Lagged $PM_{2.5}$	0.0038	-0.0006			
	(0.0034)	(0.0007)			
Lagged Household Size	0.0005	-0.0004			
	(0.0014)	(0.0005)			
Lagged Income	0.0005	0.0008			
	(0.0026)	(0.0012)			
Lagged Shopping Gradient	0.0710	-0.0047			
	(0.0680)	(0.0210)			
Baseline $PM_{2.5}$			-0.0598	0.1496	
			(0.2390)	(0.2992)	
Baseline Household Size			0.0280*	0.0070	
			(0.0164)	(0.0212)	
Baseline Income			-0.0008	-0.0333	
			(0.0346)	(0.0412)	
Baseline Shopping Gradient			0.0270	-0.1065	
			(0.0746)	(0.1191)	
Post-Entry Indicator					-0.0011
					(0.0374)
Joint *F*-statistic	0.629	0.393	1.268	0.913	–
ZIP code FE	Yes	Yes	No	No	Yes
Year FE	Yes	No	No	No	No
County FE	No	No	No	No	No
County × Year FE	No	Yes	No	No	No
Observations	129,839	128,564	5,580	1,549	110,997

Notes: This table presents three complementary diagnostics for the exogeneity of AirNow forecast rollout. Columns (1) and (2) report panel balance tests in which the dependent variable is an indicator for AirNow forecast access. All covariates are lagged by one year. Column (1) includes ZIP code and year fixed effects, while column (2) includes ZIP code and county-by-year fixed effects. Columns (3) and (4) report cross-sectional tests predicting rollout year using 2004 baseline ZIP characteristics; column (4) restricts the sample to ZIP codes treated strictly after 2004. Column (5) tests whether residualized $PM_{2.5}$ changes discontinuously at forecast entry, after absorbing ZIP code, week, and county-by-year fixed effects and controlling for weather. Across specifications, we fail to reject the null that rollout timing is unrelated to pre-existing pollution levels, household characteristics, or shopping behavior. The joint F-statistics fail to reject joint insignificance in all four prediction tests, and the estimated discontinuity in column (5) is negligible and statistically indistinguishable from zero. Standard errors are clustered at the ZIP code level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.5: Contemporaneous Effects of $PM_{2.5}$ on Online Trips and Expenditure

Dependent Var:	Trip Probability	No. of Trips	Expenditure	Expenditure (Cond.)
	(1)	(2)	(3)	(4)
$PM_{2.5}$ ($\mu g/m^3$)	0.0011**	0.0011**	0.0004	0.0046
	(0.0005)	(0.0005)	(0.0004)	(0.0530)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Dep. Var. Mean	0.45	0.47	0.20	44.60
Effect relative to mean, %	0.2444	0.2340	0.2000	0.0103
Kleibergen-Paap F-statistics	388	388	388	323
Observations	176,326,439	176,326,439	176,326,439	796,988

Notes: This table reports IV estimates of the contemporaneous effects of $PM_{2.5}$ on online trips and expenditure. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.6: Contemporaneous Effects of $PM_{2.5}$ on Expenditure by Product Category

Dependent Var:	Defensive goods (1)	Necessity (2)	Non-necessity (3)
$PM_{2.5}$ ($\mu g/m^3$)	0.00002	-0.0149***	-0.0075***
	(0.0001)	(0.0033)	(0.0015)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Dep. Var. Mean	0.08	8.62	3.16
Effect relative to mean, %	0.0250	-0.1729	-0.2373
Kleibergen-Paap F-statistics	388	388	388
Observations	176,326,439	176,326,439	176,326,439

Notes: This table reports IV estimates of the contemporaneous effect of $PM_{2.5}$ on household in-store spending by product category. Units are dollars per household-day for expenditure. *Defensive goods* include masks, air purifiers and filters, breathing aids, and over-the-counter products for respiratory, cough, throat, and sinus symptoms. *Necessities* include food and health-related goods. *Non-necessities* include non-food purchases excluding health-related goods. Alcohol is classified as neither a necessity nor a non-necessity. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.7: Summary of the Dose–Response Relationship from the Literature

Source	Additional Dose	Response
Panel A: Trips		
Graff Zivin and Neidell (2012)	10 ppb Ozone	no statistically significant change in the decision to work
Keiser, Lade, and Rudik (2018)	1 ppb maximum ozone	3.9% decrease in monthly park visitation
He, Luo, and Zhang (2022)	pollution day	2.26% decrease in the market share of a movie
Fitzgerald (2024)	1 $\mu g/m^3$ Smoke $PM_{2.5}$ in very unhealthy bins	0.15% decrease in daily visits (all foot traffic)
Sun (2025)	1 $\mu g/m^3$	0.50% decrease in daily visit rates (all foot traffic)
Sheldon and Zhan (2026)	35 $\mu g/m^3$ Smoke $PM_{2.5}$	no statistically significant change in the retail trips
Panel B: Consumption		
Schlenker and Walker (2016)	1 std. dev. of daily pollution	additional $1 million in hospitalization costs
Deschênes, Greenstone, and Shapiro (2017)	10 pp decline in NO_x emissions	0.006 pp decrease in medication spending
Chang, Huang, and Wang (2018)	1 std. dev. of $PM_{2.5}$	7.2% increase in no. of health insurance contracts sold
Zhang and Mu (2018)	100-point AQI	54.5% increase in mask purchases, 70.6% increase in anti-$PM_{2.5}$ mask purchases
Deryugina et al. (2019)	1 $\mu g/m^3$ $PM_{2.5}$	increases three-day inpatient ER spending by over $16,000 per million
Williams et al. (2019)	1 $\mu g/m^3$ $PM_{2.5}$ weekly exposure	0.82% increase in weekly inhaler use
Barwick et al. (2024b)	10 $\mu g/m^3$ $PM_{2.5}$	2.65% increase in healthcare transactions 0.55% decrease in necessities spending though insignificant
Han, Li, and Wang (2024)	day exposed to smoke pollution	0.24-0.77% increase in retail sales of defensive goods
Chen, Wang, and Zhang (2025)	10-unit increase in AQI	3.6 pp decrease in monthly consumption
Chen, Fu, and Lyu (2026)	100 $\mu g/m^3$ $PM_{2.5}$	14% decrease in daily offline spending paid via bankcards
Sheldon and Zhan (2026)	35 $\mu g/m^3$ Smoke $PM_{2.5}$	3-4% more spending on groceries and perishables

Table F.8: Robustness: $PM_{2.5}$ Effects Across Alternative Functional Forms

Dependent Var:	Trip Probability	Store Visits			Shopping Expenditure		
	Probit	log(1+Y)	arcsinh(Y)	Poisson	log(1+Y)	arcsinh(Y)	Poisson
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0007***	-0.0002***	-0.0003***	-0.0011***	-0.0009***	-0.0010***	-0.0019***
	(0.0002)	(0.00005)	(0.00007)	(0.0002)	(0.0002)	(0.0002)	(0.0004)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	176,325,283	176,326,439	176,326,439	176,325,308	176,326,439	176,326,439	176,325,308

Notes: This table reports IV estimates of the contemporaneous effect of $PM_{2.5}$ on household shopping outcomes using alternative functional forms. Because the specifications use different functional forms, coefficients are reported in their native scales rather than converted to common outcome units. Column (1) reports effects from a Probit model for trip probability. The average marginal effect of $PM_{2.5}$ is -0.0002. Columns (2)–(3) and (5)–(6) report linear models for transformed outcomes, using $\log(1+Y)$ and $\text{arcsinh}(Y)$, respectively. Columns (4) and (7) report Poisson models for trip counts and spending. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.9: Robustness: Different Fixed Effects

Dependent Var:	Trip Probability			Store Visits			Expenditure		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0171***	-0.0161***	-0.0101*	-0.0405***	-0.0390***	-0.0289***	-0.0210***	-0.0235***	-0.0149***
	(0.0058)	(0.0059)	(0.0056)	(0.0100)	(0.0107)	(0.0098)	(0.0047)	(0.0056)	(0.0048)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	No	No	Yes	No	No	Yes	No	No
County-by-month FE	No	Yes	No	No	Yes	No	No	Yes	No
County-by-week FE	No	No	Yes	No	No	Yes	No	No	Yes
Week FE	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Month-by-year FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	176,326,446	176,326,428	176,326,302	176,326,446	176,326,428	176,326,302	176,326,446	176,326,428	176,326,302

Notes: This table reports IV estimates of the contemporaneous effect of $PM_{2.5}$ on household shopping outcomes using different sets of fixed effects. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.10: Robustness: Different Clustering Levels

Dependent Var:	Trip Probability			Store Visits			Expenditure		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0187*	-0.0187*	-0.0187	-0.0409**	-0.0409**	-0.0409*	-0.0221*	-0.0221*	-0.0221*
	(0.0113)	(0.0113)	(0.0130)	(0.0201)	(0.0206)	(0.0244)	(0.0117)	(0.0115)	(0.0133)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering level	County & Year	County & Month–year	County & Week–year	County & Year	County & Month–year	County & Week–year	County & Year	County & Month–year	County & Week–year
Observations	176,326,439	176,326,439	176,326,439	176,326,439	176,326,439	176,326,439	176,326,439	176,326,439	176,326,439

Notes: This table reports IV estimates of the contemporaneous effect of $PM_{2.5}$ on household shopping outcomes under alternative clustering schemes for standard errors. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. "Month–Year" and "Week–Year" refer to clustering by month-by-year and week-by-year, respectively. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.11: Robustness: Controlling for Past-week Shopping Activities

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0142***	-0.0357***	-0.0229***
	(0.0054)	(0.0095)	(0.0046)
Past-week shopping days	2.3867***		
	(0.0300)		
Past-week store visits		1.9870***	
		(0.0366)	
Past-week total expenditure			-0.0088***
			(0.0003)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Observations	176,026,619	176,026,619	176,026,619

Notes: This table reports IV estimates of the contemporaneous effect of $PM_{2.5}$ on household shopping outcomes after adding controls for past-week aggregate shopping activities. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.12: Robustness: Controlling for Other Pollutants

Dependent Var:	Trip Probability			Store Visits			Expenditure		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0350***	-0.0203**	-0.0355***	-0.0709***	-0.0391***	-0.0670***	-0.0368***	-0.0124*	-0.0265***
	(0.0060)	(0.0084)	(0.0090)	(0.0108)	(0.0146)	(0.0161)	(0.0060)	(0.0072)	(0.0085)
Ozone (ppm)	19.9884***		23.6475***	41.5804***		48.9201***	24.7557***		28.0577***
	(2.8375)		(3.8627)	(6.0449)		(8.3897)	(3.2122)		(4.4902)
CO (ppm)		-0.8682***	-0.6467***		-1.8301***	-1.3942***		-1.1605***	-0.9177***
		(0.1752)	(0.1857)		(0.3128)	(0.3296)		(0.1598)	(0.1710)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	142,176,417	108,626,014	97,439,370	142,176,417	108,626,014	97,439,370	142,176,417	108,626,014	97,439,370

Notes: This table reports IV estimates of the contemporaneous effect of $PM_{2.5}$ on household shopping outcomes after adding controls for other pollutants. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.13: Robustness: Using Alternative Constructions of Upwind Instrument

Dependent Var:	Trip Probability			Store Visits			Expenditure		
	100-200km (1)	150-300km (2)	300-450km (3)	100-200km (4)	150-300km (5)	300-450km (6)	100-200km (7)	150-300km (8)	300-450km (9)
Panel A. Simple average									
$PM_{2.5}$ ($\mu g/m^3$)	-0.0192***	-0.0159***	-0.0308**	-0.0411***	-0.0390***	-0.0337	-0.0217***	-0.0236***	-0.0016
	(0.0054)	(0.0059)	(0.0148)	(0.0095)	(0.0107)	(0.0239)	(0.0048)	(0.0057)	(0.0104)
Kleibergen-Paap F-statistics	428	400	60	428	400	60	428	400	60
Panel B. Weighted by distance									
$PM_{2.5}$ ($\mu g/m^3$)	-0.0194***	-0.0167***	-0.0269*	-0.0411***	-0.0400***	-0.0277	-0.0216***	-0.0238***	-0.0005
	(0.0054)	(0.0058)	(0.0140)	(0.0095)	(0.0104)	(0.0228)	(0.0047)	(0.0054)	(0.0100)
Kleibergen-Paap F-statistics	358	418	66	358	418	66	358	418	66
Panel C. Weighted by squared distance									
$PM_{2.5}$ ($\mu g/m^3$)	-0.0197***	-0.0173***	-0.0232*	-0.0413***	-0.0406***	-0.0220	-0.0216***	-0.0240***	-0.0026
	(0.0053)	(0.0057)	(0.0133)	(0.0095)	(0.0102)	(0.0219)	(0.0047)	(0.0052)	(0.0097)
Kleibergen-Paap F-statistics	288	415	72	288	415	72	288	415	72
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	174,221,806	176,121,896	177,044,826	174,221,806	176,121,896	177,044,826	174,221,806	176,121,896	177,044,826

Notes: This table reports IV estimates of the contemporaneous effect of $PM_{2.5}$ on household shopping outcomes using alternative constructions of the upwind instrument. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. Columns vary the distance band for potential source counties, and panels vary the weighting scheme used to aggregate upwind pollution. Panel A uses a simple average, Panel B uses inverse-distance weights, and Panel C uses inverse-squared-distance weights, which is the baseline weighting scheme. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.14: Robustness: Alternative Wind-Direction Instrument

Dependent Var:	Trip Probability	Store Visits	Expenditure
	(1)	(2)	(3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0198***	-0.0489***	-0.0205***
	(0.0053)	(0.0096)	(0.0029)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Dep. Var. Mean	25.26	36.42	12.19
Effect relative to mean, %	-0.0784	-0.1343	-0.1682
Kleibergen-Paap F-statistics	91	91	91
Observations	176,326,439	176,326,439	176,326,439

Notes: This table reports IV estimates of the contemporaneous effect of $PM_{2.5}$ on household shopping outcomes using an alternative wind-direction instrument constructed as in Deryugina et al. (2019). The instrument interacts wind-direction bins with spatial-group indicators, allowing the first-stage effect of wind direction on local PM2.5 to vary across regions. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.15: Forecast Coverage and Spending Composition

Dependent Var:	Defensive goods (1)	Necessity (2)	Non-necessity (3)
$PM_{2.5}$ $(\mu g/m^3)$	0.0001	-0.0106***	-0.0052***
	(0.0001)	(0.0040)	(0.0017)
$PM_{2.5}$ $(\mu g/m^3) \times \mathbf{1}\{\text{After Coverage}\}$	0.00003	-0.0060	-0.0033*
	(0.0001)	(0.0038)	(0.0018)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Dep. Var. Mean	0.08	8.62	3.16
Kleibergen-Paap F-statistics	203	203	203
Observations	176,326,439	176,326,439	176,326,439

Notes: This table reports IV estimates of the contemporaneous effect of $PM_{2.5}$ on household shopping spending by product category. Units are dollars per household-day for expenditure. *Defensive goods* include masks, air purifiers and filters, breathing aids, and over-the-counter products for respiratory, cough, throat, and sinus symptoms. *Necessities* include food and health-related goods. *Non-necessities* include non-food purchases excluding health-related goods. Alcohol is classified as neither a necessity nor a non-necessity. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.16: Effects of Forecast Coverage on Shopping–Pollution Gradient by Vulnerability Status

Dependent Var:	Poverty Status		Age vulnerability		Having Ailment	
	No	Yes	No	Yes	No	Yes
Panel A. Trip probability						
$PM_{2.5}$ $(\mu g/m^3)$	-0.0091	-0.0301	-0.0229***	0.0097	0.0155	0.0071
	(0.0067)	(0.0217)	(0.0077)	(0.0079)	(0.0163)	(0.0126)
$PM_{2.5}$ $(\mu g/m^3) \times \mathbf{1}${After Coverage}	-0.0141**	0.0286	-0.0147*	-0.0048	-0.0209	0.0014
	(0.0068)	(0.0235)	(0.0079)	(0.0094)	(0.0184)	(0.0136)
Dep. Var. Mean	25.29	24.39	24.45	26.59	25.54	27.68
Panel B. Store Visits						
$PM_{2.5}$ $(\mu g/m^3)$	-0.0236**	-0.0294	-0.0437***	0.0031	0.0308	0.0154
	(0.0107)	(0.0360)	(0.0118)	(0.0134)	(0.0281)	(0.0222)
$PM_{2.5}$ $(\mu g/m^3) \times \mathbf{1}${After Coverage}	-0.0262**	0.0276	-0.0285**	-0.0064	-0.0516*	-0.0154
	(0.0115)	(0.0409)	(0.0128)	(0.0157)	(0.0312)	(0.0235)
Dep. Var. Mean	36.46	35.52	35.22	38.41	37.35	41.42
Panel C. Shopping Expenditure						
$PM_{2.5}$ $(\mu g/m^3)$	-0.0150***	0.0023	-0.0231***	0.0009	0.0170	-0.0027
	(0.0057)	(0.0141)	(0.0068)	(0.0054)	(0.0119)	(0.0097)
$PM_{2.5}$ $(\mu g/m^3) \times \mathbf{1}${After Coverage}	-0.0108**	-0.0143	-0.0124**	-0.0086	-0.0227*	-0.0043
	(0.0054)	(0.0141)	(0.0062)	(0.0062)	(0.0126)	(0.0095)
Dep. Var. Mean	12.26	10.42	12.08	12.36	11.70	14.28
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap F-statistics	201	194	214	183	163	155
SW F-stat for $PM_{2.5}$	317	342	338	274	255	252
SW F-stat for $PM_{2.5} \times \mathbf{1}${After Coverage}	388	440	441	360	310	285
Observations	168,688,137	7,638,109	109,739,093	66,587,227	16,423,849	50,573,396

Notes: This table reports how the effect of forecast coverage on the shopping–pollution gradient varies by household vulnerability status. Columns split the sample by poverty status, age vulnerability, and ailment status. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. The coefficient on $PM_{2.5}$ captures the pre-coverage shopping–pollution gradient for each group, and the interaction $PM_{2.5} \times \mathbf{1}${After Coverage} captures the change in that gradient after forecast coverage begins. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. The Kleibergen-Paap F-statistic is the weak-instrument diagnostic. SW F-statistics refer to Sanderson-Windmeijer conditional first-stage F-statistics (Sanderson and Windmeijer, 2016), reported for specifications involving multiple endogenous variables. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.17: Forecast Information and Anticipatory Shopping Responses to Air Pollution: Controlling for Future Weather

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ $(\mu g/m^3) \times$ **1**{After Coverage}	-0.0369**	-0.0775***	-0.0511***
	(0.0143)	(0.0254)	(0.0135)
$PM_{2.5}$, 1-day lead $(\mu g/m^3) \times$ **1**{After Coverage}	0.0624***	0.1103***	0.0818***
	(0.0226)	(0.0398)	(0.0237)
$PM_{2.5}$, 2-day lead $(\mu g/m^3) \times$ **1**{After Coverage}	0.0450***	0.0580**	0.0425**
	(0.0166)	(0.0291)	(0.0166)
$PM_{2.5}$ $(\mu g/m^3)$	0.0010	0.0010	-0.0010
	(0.0166)	(0.0275)	(0.0160)
$PM_{2.5}$, 1-day lead	-0.0354*	-0.0824**	-0.0635**
	(0.0196)	(0.0343)	(0.0257)
$PM_{2.5}$, 2-day lead	0.0268	0.0360	0.0270*
	(0.0167)	(0.0256)	(0.0150)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Dep. Var. Mean	25.25	36.49	12.29
Kleibergen-Paap F-statistics	162	162	162
Observations	158,713,162	158,713,162	158,713,162

Notes: This table reports forward-looking shopping responses to air pollution after AirNow forecast access becomes available, controlling additionally for one- and two-day-ahead weather conditions. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. We instrument for $PM_{2.5}$, its leads, and their interactions with forecast access using the corresponding upwind instruments. **1**{After Coverage} equals one on and after the first date when AirNow forecasts become available for the household's ZIP code, and zero otherwise. Controls include the **1**{After Coverage} indicator, holiday indicators, contemporaneous and one- and two-day-ahead weather controls, and household demographics. Demographic and fixed-effect controls follow the baseline specification. Standard errors are clustered at the county level and reported in parentheses. The Kleibergen-Paap F-statistic is the weak-instrument diagnostic. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.18: Spatial Spillovers in Forecast Access Effects

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ $(\mu g/m^3)$ $\times$ **1**{After Coverage}	0.0066	0.0084	-0.0042
	(0.0123)	(0.0198)	(0.0078)
$PM_{2.5}$ $(\mu g/m^3)$ $\times$ **1**{After Coverage} $\times$ **1**{Treated}	-0.0126	-0.0187	-0.0048
	(0.0097)	(0.0162)	(0.0078)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Kleibergen-Paap F-statistics	123	123	123
Observations	158,488,071	158,488,071	158,488,071

Notes: This table tests for potential spatial spillovers from AirNow forecast rollout. We compare ZIP codes that receive forecast coverage with nearby ZIP codes that share a border but have not yet received coverage, allowing the change in the shopping–pollution gradient to differ between the two groups. **1**{Treated} equals one for ZIP codes in the rollout wave and zero for neighboring ZIP codes not yet covered. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.19: Staggered Rollout Design: Early Treated vs. Never Treated Comparison

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0074	-0.0206**	-0.0141**
	(0.0063)	(0.0102)	(0.0062)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ **1**{After Coverage}	-0.0173**	-0.0293**	-0.0118*
	(0.0071)	(0.0117)	(0.0063)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Observations	158,987,110	158,987,110	158,987,110

Notes: This table compares early treated areas (receiving AirNow forecasts during 2004-2008, 84% of treated ZIP codes) to never-treated areas. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. The coefficient on $PM_{2.5} \times$ **1**{After Coverage} captures how forecast access changes the pollution–shopping gradient. **1**{After Coverage} equals one on and after the first date when an AirNow forecast becomes available for the household's ZIP code, and zero otherwise. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.20: Robustness: Cohort-Specific Pollution Gradients and Time Trends

Dependent Var:	Trip Probability		Store Visits		Expenditure	
	(1)	(2)	(3)	(4)	(5)	(6)
$PM_{2.5}$	0.0007	0.0020	-0.0034	0.0049	-0.0146**	-0.0093
	(0.0093)	(0.0105)	(0.0150)	(0.0170)	(0.0066)	(0.0084)
$PM_{2.5} \times \mathbf{1}${After Coverage}	-0.0315***	-0.0374*	-0.0546***	-0.0540*	-0.0168**	-0.0176
	(0.0107)	(0.0226)	(0.0172)	(0.0302)	(0.0073)	(0.0150)
$PM_{2.5} \times$ Trend		0.0022**		0.0041**		0.0010
		(0.0011)		(0.0018)		(0.0009)
Ratio to baseline, $PM_{2.5} \times \mathbf{1}${After Coverage}		1.19		0.99		1.05
Equal pre-entry gradients (p-value)		0.246		0.076		0.740
$PM_{2.5} \times$ Rollout cohort	No	Yes	No	Yes	No	Yes
$PM_{2.5} \times$ Linear trend	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	52,483,413	52,483,413	52,483,413	52,483,413	52,483,413	52,483,413

Notes: This table tests whether the AirNow coverage effect on the pollution–shopping gradient reflects pre-existing differences across rollout cohorts or nationwide changes in pollution sensitivity. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All columns report IV estimates on a 30% random sample of households, drawn because the full specification exceeds available computing memory. $PM_{2.5}$ and each of its interactions are instrumented by upwind $PM_{2.5}$ interacted with the same indicators. Odd columns re-estimate the baseline specification on this sample, while even columns add interactions of $PM_{2.5}$ with rollout-cohort indicators and with a linear year trend centered at 2012. Cohorts are defined by AirNow entry in the first half of 2004, the second half of 2004, 2005–07, 2008–10, 2011–13, or 2014–19, with never-covered ZIP codes as the reference group. The coverage interaction in these columns is therefore identified only from changes in the gradient within a cohort at entry, net of the national trend. The ratio row divides the coverage interaction in each even column by its counterpart in the preceding odd column. The p-value row tests whether pre-entry gradients are equal across cohorts observed both before and after entry. It excludes the first-half-2004 cohort, which is covered from the start of the sample. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.21: Placebo Tests for Forecast Access Effects

	Placebo 1: Never-Treated Zipcodes			Placebo 2: Treated Zipcodes		
Dependent Var:	Trip Probability	Store Visits	Expenditure	Trip Probability	Store Visits	Expenditure
	(1)	(2)	(3)	(4)	(5)	(6)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0042	-0.0154	-0.0069	-0.0266***	-0.0553***	-0.0277***
	(0.0088)	(0.0150)	(0.0065)	(0.0075)	(0.0128)	(0.0048)
$PM_{2.5} \times$ **1**{After Coverage}	-0.0014	0.0046	-0.0026	0.0068	0.0109	0.0032
	(0.0090)	(0.0153)	(0.0066)	(0.0079)	(0.0128)	(0.0031)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,958,623	31,958,623	31,958,623	144,367,694	144,367,694	144,367,694

Notes: This table reports placebo tests for the IV estimates of the effect of forecast access on the shopping-pollution gradient. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. Columns (1)–(3) assign fake adoption dates, randomly drawn from the empirical distribution of actual AirNow rollout dates, to never-treated ZIP codes. Columns (4)–(6) restrict the sample to treated ZIP codes and randomly reassign adoption dates from the same empirical distribution, thereby scrambling the true rollout timing. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table F.22: Effects of Forecast Coverage Excluding Air-Quality Alert Days

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0087	-0.0223*	-0.0148**
	(0.0075)	(0.0122)	(0.0066)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ **1**{After Coverage}	-0.0151**	-0.0292**	-0.0122**
	(0.0075)	(0.0127)	(0.0059)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Dep. Var. Mean	25.26	36.43	12.20
Kleibergen-Paap F-statistics	257	257	257
Observations	175,729,065	175,729,065	175,729,065

Notes: This table reports how forecast access changes the shopping–pollution gradient after excluding days when either forecasted or realized AQI is Unhealthy for Sensitive Groups or worse. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. **1**{After Coverage} equals one on and after the first date when AirNow forecasts are available for the household's ZIP code, and zero otherwise. Controls include the **1**{After Coverage} indicator, holiday indicators, weather controls, and household demographics. Weather, demographic, and fixed-effect controls follow the baseline specification. Standard errors are clustered at the county level and reported in parentheses. The Kleibergen-Paap *F*-statistic is the weak-instrument diagnostic. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.23: Forecast Access Effects with Reporting-area Clustering

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0098**	-0.0239***	-0.0145***
	(0.0047)	(0.0089)	(0.0035)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ **1**{After Coverage}	-0.0124**	-0.0237**	-0.0106**
	(0.0060)	(0.0109)	(0.0041)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Observations	176,326,439	176,326,439	176,326,439

Notes: This table reports how access to AirNow forecast information changes the shopping–pollution gradient, with standard errors clustered at the reporting-area level. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. **1**{After Coverage} equals one on and after the first date when an AirNow forecast becomes available for the household's ZIP code, and zero otherwise. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the ZIP code level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.24: Shopping-Pollution Gradients across Different Forecast Coverage Groups

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5} \times \mathbf{1}\{\text{No Coverage}\}$ (G1)	-0.0056	-0.0200**	-0.0012
	(0.0058)	(0.0100)	(0.0049)
$PM_{2.5} \times \mathbf{1}\{\text{Correct Report}\}$ (G2)	-0.0185***	-0.0352***	-0.0240***
	(0.0059)	(0.0105)	(0.0052)
$PM_{2.5} \times \mathbf{1}\{\text{Wrong Report}\}$ (G3)	-0.0127**	-0.0251***	-0.0162***
	(0.0052)	(0.0093)	(0.0043)
p-value G1 = G2	0.0137	0.0329	0.0000
p-value G1 = G3	0.1260	0.5774	0.0024
p-value G2 = G3	0.0031	0.0021	0.0000
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Observations	150,662,908	150,662,908	150,662,908

Notes: This table reports IV estimates allowing the shopping–pollution gradient to vary across three forecast-coverage groups. Group 1 (G1) consists of ZIP-code-days without AirNow forecast coverage, Group 2 (G2) consists of covered ZIP-code-days with a correct forecast, and Group 3 (G3) consists of covered ZIP-code-days with an incorrect forecast. A correct forecast is defined as a forecasted AQI category that matches the realized AQI category. $PM_{2.5}$ is instrumented using the inverse-distance-squared weighted upwind instrument interacted with group indicators. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.25: Effects of Forecast Accuracy on Shopping–Pollution Gradient by Vulnerability Status

Dependent Var:	Poverty Status		Age vulnerability		Having Ailment	
	No	Yes	No	Yes	No	Yes
Panel A. Trip probability						
$PM_{2.5}$ ($\mu g/m^3$)	-0.0232***	-0.0031	-0.0359***	0.0095	-0.0058	0.0136
	(0.0061)	(0.0143)	(0.0066)	(0.0086)	(0.0117)	(0.0108)
$PM_{2.5}$ × Forecast Accuracy	-0.0160***	0.0099	-0.0168***	-0.0086	0.0088	0.0020
	(0.0046)	(0.0176)	(0.0055)	(0.0070)	(0.0087)	(0.0092)
Dep. Var. Mean	25.35	24.61	24.26	27.66	25.96	28.38
Panel B. Store Visits						
$PM_{2.5}$ ($\mu g/m^3$)	-0.0487***	-0.0089	-0.0673***	0.0002	-0.0183	0.0075
	(0.0113)	(0.0247)	(0.0126)	(0.0143)	(0.0199)	(0.0199)
$PM_{2.5}$ × Forecast Accuracy	-0.0329***	-0.0068	-0.0331***	-0.0157	-0.0025	-0.0121
	(0.0085)	(0.0286)	(0.0095)	(0.0125)	(0.0150)	(0.0162)
Dep. Var. Mean	36.53	35.66	34.83	40.20	37.98	42.61
Panel C. Shopping Expenditure						
$PM_{2.5}$ ($\mu g/m^3$)	-0.0247***	-0.0102	-0.0329***	-0.0032	-0.0062	-0.0043
	(0.0056)	(0.0099)	(0.0061)	(0.0061)	(0.0097)	(0.0103)
$PM_{2.5}$ × Forecast Accuracy	-0.0162***	-0.0085	-0.0185***	-0.0083*	-0.0046	-0.0060
	(0.0039)	(0.0117)	(0.0045)	(0.0047)	(0.0072)	(0.0086)
Dep. Var. Mean	12.31	10.31	12.18	12.33	12.84	14.82
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap F-statistics	190	168	200	165	142	139
SW F-stat for $PM_{2.5}$	415	359	433	364	310	308
SW F-stat for $PM_{2.5}$ × Forecast Accuracy	448	320	452	417	422	407
Observations	102,308,764	4,540,417	73,772,791	33,076,495	18,018,005	23,525,862

Notes: This table reports how rolling forecast accuracy moderates the shopping–pollution gradient by household vulnerability status. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. The sample is restricted to household-day observations with AirNow forecast access and sufficient forecast history, defined as at least 10 forecast days in the past 90 days. Columns split the sample by poverty status, age vulnerability, and ailment status. The coefficient on $PM_{2.5}$ captures the shopping–pollution gradient at the sample mean of rolling forecast accuracy for each group, while the interaction $PM_{2.5}$ × Forecast Accuracy captures how that gradient changes with recent forecast accuracy. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. The Kleibergen-Paap F-statistic is the weak-instrument diagnostic. SW F-statistics refer to Sanderson-Windmeijer conditional first-stage F-statistics (Sanderson and Windmeijer, 2016), reported for specifications involving multiple endogenous variables. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.26: Robustness Checks for Same-Day Forecast Error Effects

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
Panel A: Add reporting-area-by-season fixed effects			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0100	-0.0211*	-0.0077
	(0.0062)	(0.0108)	(0.0053)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Error(abs.)	0.0042**	0.0076**	0.0053***
	(0.0018)	(0.0030)	(0.0017)
Observations	82,162,847	82,162,847	82,162,847
Panel B: Controls for forecastability			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0101	-0.0222*	-0.0089
	(0.0067)	(0.0121)	(0.0062)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Error(abs.)	0.0044**	0.0082**	0.0063***
	(0.0019)	(0.0033)	(0.0018)
Observations	77,500,705	77,500,705	77,500,705
Panel C: Add AQI-bins fixed effects			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0310**	-0.0657***	-0.0306***
	(0.0141)	(0.0243)	(0.0115)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Error(abs.)	0.0055***	0.0103***	0.0068***
	(0.0019)	(0.0033)	(0.0018)
Observations	82,162,849	82,162,849	82,162,849
Panel D: Reporting-area clustering level			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0129*	-0.0252*	-0.0100
	(0.0076)	(0.0139)	(0.0072)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Error(abs.)	0.0042**	0.0079**	0.0055***
	(0.0020)	(0.0032)	(0.0018)
Observations	82,162,849	82,162,849	82,162,849

Notes: This table reports robustness checks for the baseline estimates of how same-day forecast error moderates the effect of $PM_{2.5}$ on household shopping outcomes among households in forecast-covered areas. Panel A adds reporting-area-by-season fixed effects to absorb persistent differences in local forecasting environments. Panel B adds controls for forecastability, including recent $PM_{2.5}$ variance, temperature and precipitation volatility, and an indicator for wildfire smoke days. Panel C adds AQI-bin fixed effects to absorb nonlinear effects of realized air quality. Panel D clusters standard errors at the reporting-area level rather than the county level. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All other controls and fixed effects follow the baseline specification. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.27: Robustness Checks for Rolling Forecast Accuracy Effects

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
Panel A: Add reporting-area-by-season fixed effects			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0211***	-0.0460***	-0.0237***
	(0.0058)	(0.0108)	(0.0054)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Accuracy	-0.0138***	-0.0285***	-0.0146***
	(0.0044)	(0.0081)	(0.0037)
Observations	106,885,045	106,885,045	106,885,045
Panel B: Controls for forecastability			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0216***	-0.0474***	-0.0243***
	(0.0064)	(0.0122)	(0.0061)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Accuracy	-0.0131***	-0.0292***	-0.0157***
	(0.0046)	(0.0084)	(0.0039)
Observations	100,801,058	100,801,058	100,801,058
Panel C: Add AQI-bins fixed effects			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0480***	-0.1059***	-0.0585***
	(0.0130)	(0.0236)	(0.0115)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Accuracy	-0.0142***	-0.0299***	-0.0148***
	(0.0044)	(0.0080)	(0.0036)
Observations	106,885,045	106,885,045	106,885,045
Panel D: Reporting-area clustering level			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0223***	-0.0470***	-0.0241***
	(0.0071)	(0.0139)	(0.0066)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Accuracy	-0.0148***	-0.0316***	-0.0158***
	(0.0047)	(0.0086)	(0.0041)
Observations	106,885,045	106,885,045	106,885,045

Notes: This table reports robustness checks for the baseline estimates of how rolling forecast accuracy moderates the effect of $PM_{2.5}$ on household shopping outcomes among households in forecast-covered areas. Panel A adds reporting-area-by-season fixed effects to absorb persistent differences in local forecasting environments. Panel B adds controls for forecastability, including recent $PM_{2.5}$ variance, temperature and precipitation volatility, and an indicator for wildfire smoke days. Panel C adds AQI-bin fixed effects to absorb nonlinear effects of realized air quality. Panel D clusters standard errors at the reporting-area level rather than the county level. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All other controls and fixed effects follow the baseline specification. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.28: Placebo Tests for Rolling Forecast Accuracy Effects

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
Panel A: Pre-coverage predictability placebo (MAE)			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0047	-0.0149	-0.0112
	(0.0081)	(0.0139)	(0.0072)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Predictability	0.0012	0.0017	-0.0002
	(0.0017)	(0.0029)	(0.0012)
Observations	63,651,817	63,651,817	63,651,817
Panel B: Pre-coverage predictability placebo (RMSE)			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0052	-0.0153	-0.0113
	(0.0082)	(0.0140)	(0.0073)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Predictability	0.0004	0.0008	-0.0002
	(0.0010)	(0.0016)	(0.0007)
Observations	63,651,817	63,651,817	63,651,817
Panel C: Lead-accuracy placebo (accuracy led one year)			
$PM_{2.5}$ ($\mu g/m^3$)	-0.0296***	-0.0582***	-0.0293***
	(0.0063)	(0.0117)	(0.0057)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Lead Forecast Accuracy	0.0006	-0.0008	0.0024
	(0.0052)	(0.0086)	(0.0040)
Observations	87,460,141	87,460,141	87,460,141

Notes: This table reports placebo tests that distinguish accumulated trust in forecasts from intrinsic predictability. Panels A and B use the pre-coverage sample and estimate placebo versions of the rolling accuracy specification. Instead of interacting pollution with forecast accuracy, these specifications interact pollution with a persistence-based predictability index constructed from realized pollution alone. The index measures how closely pollution follows its previous-day level over a rolling 90-day window and is signed so that higher values indicate greater predictability. Panel A uses mean absolute error (MAE), and Panel B uses root mean squared error (RMSE). Panel C uses the forecast-era sample and estimates a lead-accuracy placebo by interacting pollution with forecast accuracy realized exactly one year in the future. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications are estimated by IV using the upwind instrument and its corresponding interaction. They include the full set of baseline controls and fixed effects, and standard errors are clustered at the county level. Standard errors are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.29: Robustness: Timing of the Rolling Forecast Accuracy Measure

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0223***	-0.0470***	-0.0241***
	(0.0060)	(0.0112)	(0.0055)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Accuracy	-0.0149***	-0.0317***	-0.0158***
	(0.0045)	(0.0083)	(0.0038)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Dep. Var. Mean	25.32	36.50	12.22
Kleibergen-Paap F-statistics	191	191	191
Observations	106,849,341	106,849,341	106,849,341

Notes: This table reports IV estimates of how rolling forecast accuracy moderates the shopping–pollution gradient, using an alternative timing of the accuracy measure that excludes day t's realized forecast performance. Specifically, for each ZIP code we reconstruct the rolling 90-day forecast accuracy rate using only forecast outcomes realized through day $t-1$. As in the baseline, the sample is restricted to household-day observations with sufficient forecast history, defined as at least 10 forecast days in the prior 90 days, and the accuracy rate is the share of those forecast days for which the one-day-ahead predicted AQI category matches the realized AQI category. We then center the accuracy rate by subtracting its sample mean. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.30: Robustness: Forecast Accuracy with Alternative Definitions

Dependent Var:	Trip Probability (1)	Store Visits (2)	Expenditure (3)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0245***	-0.0542***	-0.0242***
	(0.0062)	(0.0138)	(0.0055)
$PM_{2.5}$ ($\mu g/m^3$) $\times$ Forecast Accuracy	-0.0143***	-0.0302***	-0.0157***
	(0.0042)	(0.0081)	(0.0037)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Dep. Var. Mean	25.32	36.50	12.22
Kleibergen-Paap F-statistics	189	189	189
Observations	106,885,045	106,885,045	106,885,045

Notes: This table reports a robustness check using an alternative measure of forecast accuracy. Instead of the exact-match indicator, we use the negative absolute gap between the one-day-ahead predicted and realized AQI categories, so that higher values denote greater accuracy. We then construct the rolling 90-day average of this measure and center it as in the baseline specification. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. Sample construction, controls, fixed effects, demeaning, and inference are otherwise identical to Table 6. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.31: Robustness: Forecast Accuracy with Alternative Rolling Windows

	Rolling Window: 1 Month			Rolling Window: 6 Months		
Dependent Var:	Trip Probability	Store Visits	Expenditure	Trip Probability	Store Visits	Expenditure
	(1)	(2)	(3)	(4)	(5)	(6)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0251***	-0.0530***	-0.0282***	-0.0221***	-0.0471***	-0.0238***
	(0.0061)	(0.0113)	(0.0056)	(0.0060)	(0.0112)	(0.0055)
$PM_{2.5}\times$ Accuracy	-0.0070**	-0.0144***	-0.0078***	-0.0242***	-0.0507***	-0.0227***
	(0.0028)	(0.0051)	(0.0027)	(0.0066)	(0.0118)	(0.0056)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap F-statistics	187	187	187	192	192	192
Observations	98,134,025	98,134,025	98,134,025	106,381,765	106,381,765	106,381,765

Notes: This table reports rolling forecast accuracy moderates the shopping–pollution gradient using alternative rolling windows. Columns (1)–(3) define forecast accuracy using a one-month rolling window, and columns (4)–(6) use a six-month rolling window. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table F.32: Robustness: Alternative Minimum Forecast-History Thresholds

	Min. History: 5 Days			Min. History: 20 Days			Min. History: 30 Days		
Dependent Var:	Trip Prob.	Store Visits	Expend.	Trip Prob.	Store Visits	Expend.	Trip Prob.	Store Visits	Expend.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$PM_{2.5}$ ($\mu g/m^3$)	-0.0224***	-0.0471***	-0.0241***	-0.0223***	-0.0470***	-0.0244***	-0.0232***	-0.0489***	-0.0259***
	(0.0060)	(0.0112)	(0.0055)	(0.0060)	(0.0112)	(0.0055)	(0.0061)	(0.0113)	(0.0056)
$PM_{2.5}\times$ Forecast Accuracy	-0.0149***	-0.0314***	-0.0157***	-0.0149***	-0.0313***	-0.0156***	-0.0124***	-0.0259***	-0.0129***
	(0.0045)	(0.0082)	(0.0038)	(0.0046)	(0.0084)	(0.0039)	(0.0048)	(0.0088)	(0.0040)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	107,125,758	107,125,758	107,125,758	105,511,361	105,511,361	105,511,361	100,326,862	100,326,862	100,326,862

Notes: This table reports how rolling forecast accuracy moderates the shopping–pollution gradient under alternative minimum forecast-history thresholds. Columns (1)–(3) require at least 5 forecast days in the previous three months, columns (4)–(6) require at least 20 forecast days, and columns (7)–(9) require at least 30 forecast days. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. All specifications include the full set of controls and fixed effects used in the baseline analysis. Standard errors are clustered at the county level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

F.2 Appendix Figures

AQI Category	Index Value	Description of Air Quality
Good	0 to 50	Air quality is satisfactory, and air pollution poses little or no risk.
Moderate	51 to 100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Unhealthy	151 to 200	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy	201 to 300	Health alert: The risk of health effects is increased for everyone.
Hazardous	301 and higher	Health warning of emergency conditions: everyone is more likely to be affected.

Figure F.1: Air Quality Index Categories and Associated Health Risks

Notes: The Air Quality Index classifies air quality into six color-coded categories based on pollutant concentrations and EPA breakpoints. The categories range from "Good" (0–50), when air pollution poses little risk, to "Hazardous" (301+), when emergency health conditions may affect the general population. By translating pollutant concentrations into standardized risk categories, the AQI provides a common framework for communicating air quality conditions and recommended protective actions.

(a) AirNow website interface



(b) AirNow App interface

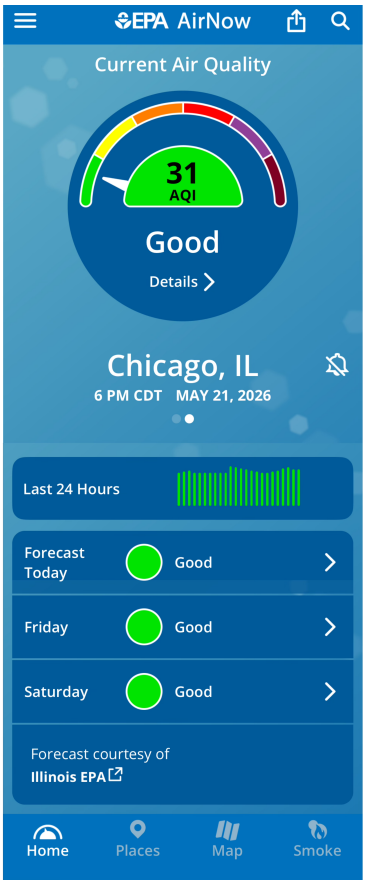


Figure F.2: AirNow portal and App interfaces illustrating AQI communication

Notes: This figure shows the illustrative AirNow public website interface in Panel (a) and the mobile app interface in Panel (b). It illustrates how air-quality information is communicated through AQI categories and short-horizon forecasts for Chicago, Illinois. Source: U.S. EPA AirNow, accessed May 21, 2026.

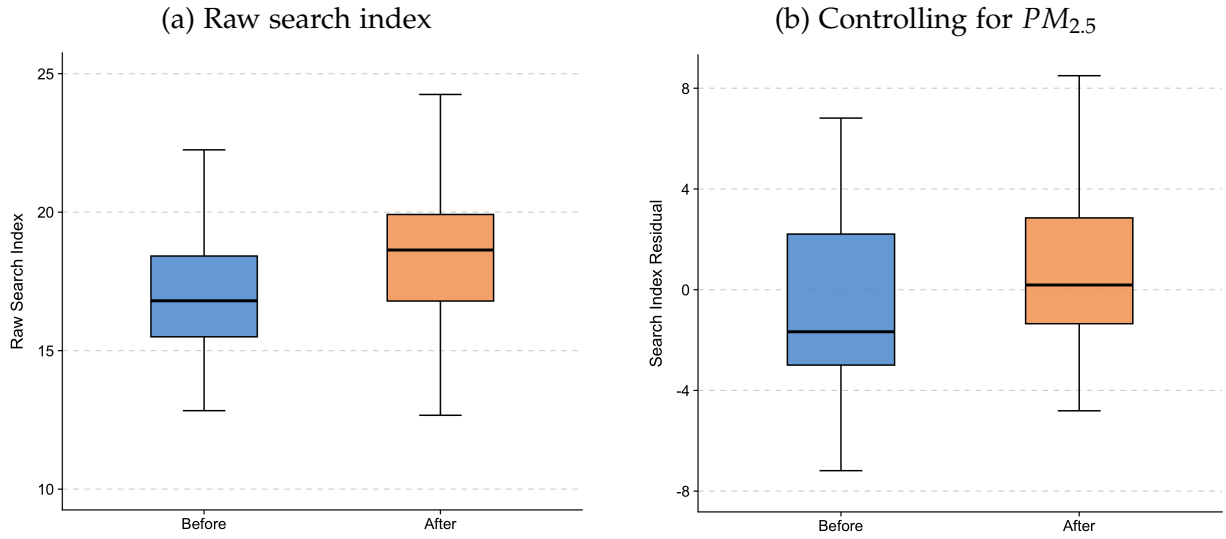
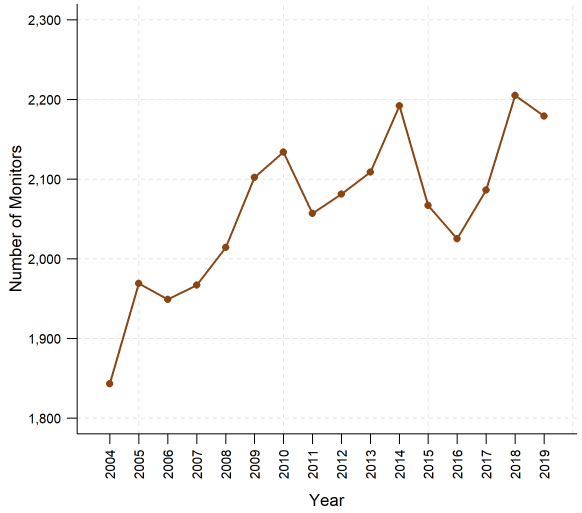


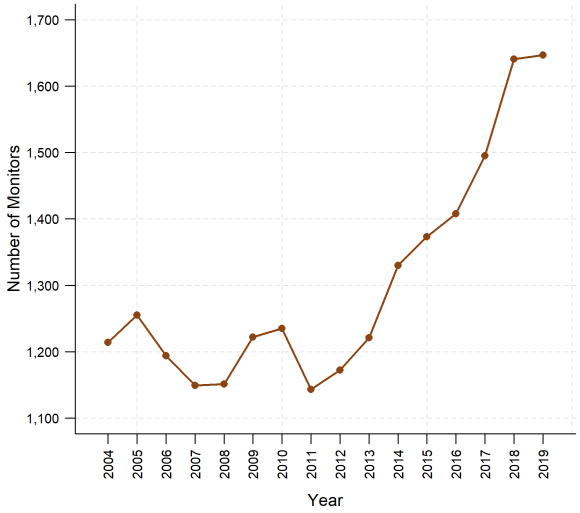
Figure F.3: Search Interest in Air Quality Information Before and After Coverage

Notes: This figure presents the distribution of Google Trends search interest for air quality-related terms across six designated market areas (DMAs) in the six months before and after local AirNow forecast access begins. Each box shows the interquartile range, with the median indicated by the horizontal line; whiskers extend to 1.5 times the interquartile range. "Before" and "After" refer to the six-month windows preceding and following the majority adoption date within each DMA. Panel (a) shows the raw average Google Trends search index for "AirNow", "Air Quality", and "Smoke". Panel (b) shows residuals from a regression of the search index on monthly mean $PM_{2.5}$, isolating variation in search interest not attributable to ambient pollution levels. The sample includes six DMAs that received AirNow forecast access after 2015 and had at least 60% of ZIP codes gain coverage.

(a) All $PM_{2.5}$ monitors



(b) FRM $PM_{2.5}$ monitors



(c) Spatial distribution of monitors by operating years

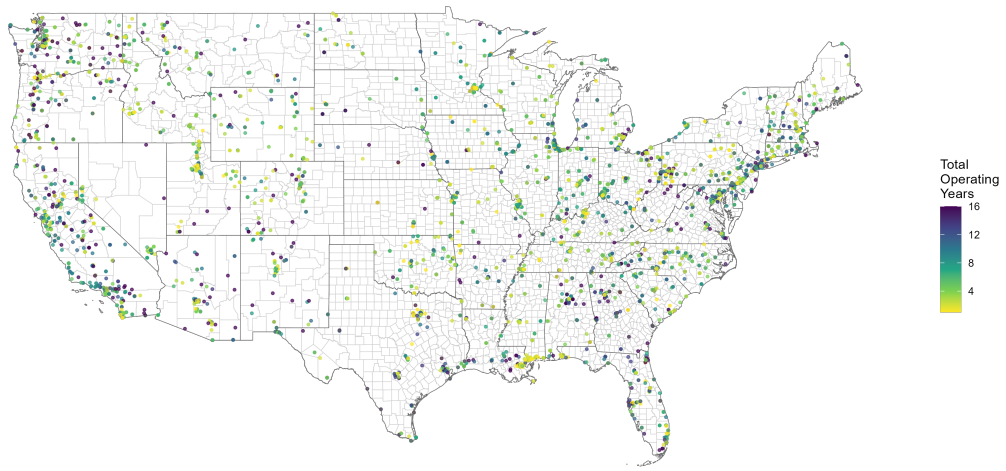
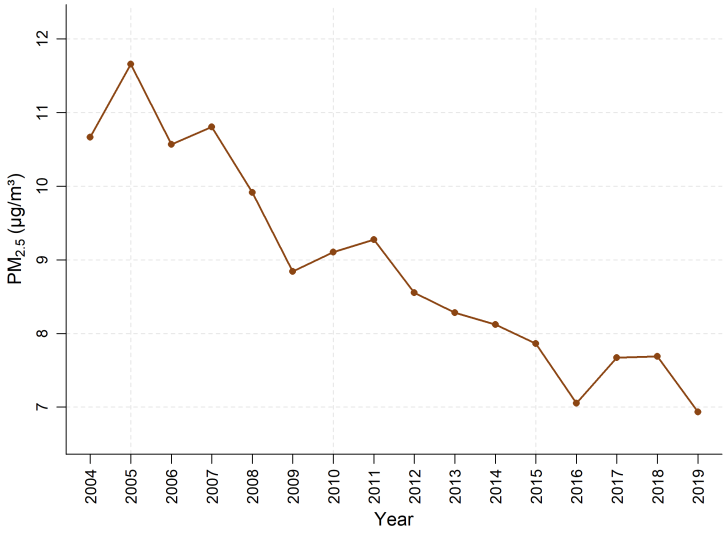


Figure F.4: U.S. $PM_{2.5}$ Monitoring Network (2004-2019)

Notes: The figure details the U.S. $PM_{2.5}$ monitoring network from 2004 to 2019, using EPA Air Quality System data. Panel (a) illustrates the annual number of active $PM_{2.5}$ monitors, including both FRM and non-FRM monitors. Panel (b) presents the annual number of active FRM $PM_{2.5}$ monitors. Panel (c) maps the spatial distribution of individual $PM_{2.5}$ monitors across the continental United States, with colors indicating the number of years each monitor operated during the sample period.

(a) Annual concentration trend



(b) Monthly concentration trend

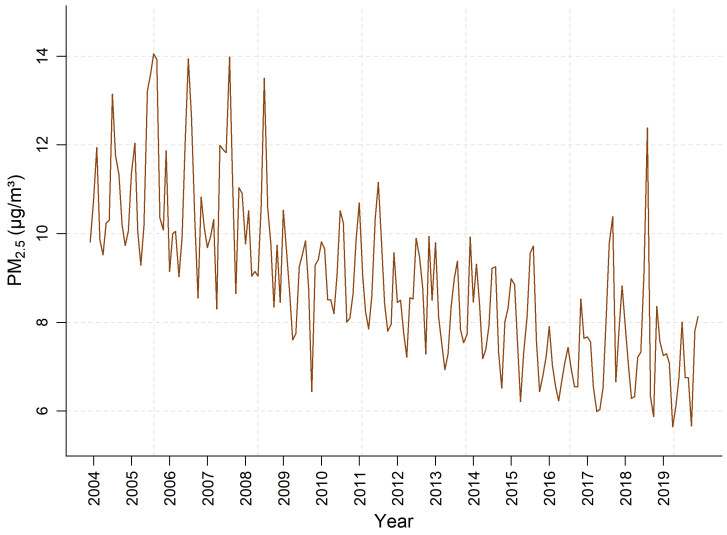


Figure F.5: U.S. National $PM_{2.5}$ Concentration Trends

Notes: This figure plots national average $PM_{2.5}$ concentrations across counties in continental U.S. using EPA Air Quality System data from both FRM and non-FRM monitors. Panel (a) reports annual means and Panel (b) reports monthly means. For each period, we first average daily monitor readings within each county and then average the resulting county-level means across counties.

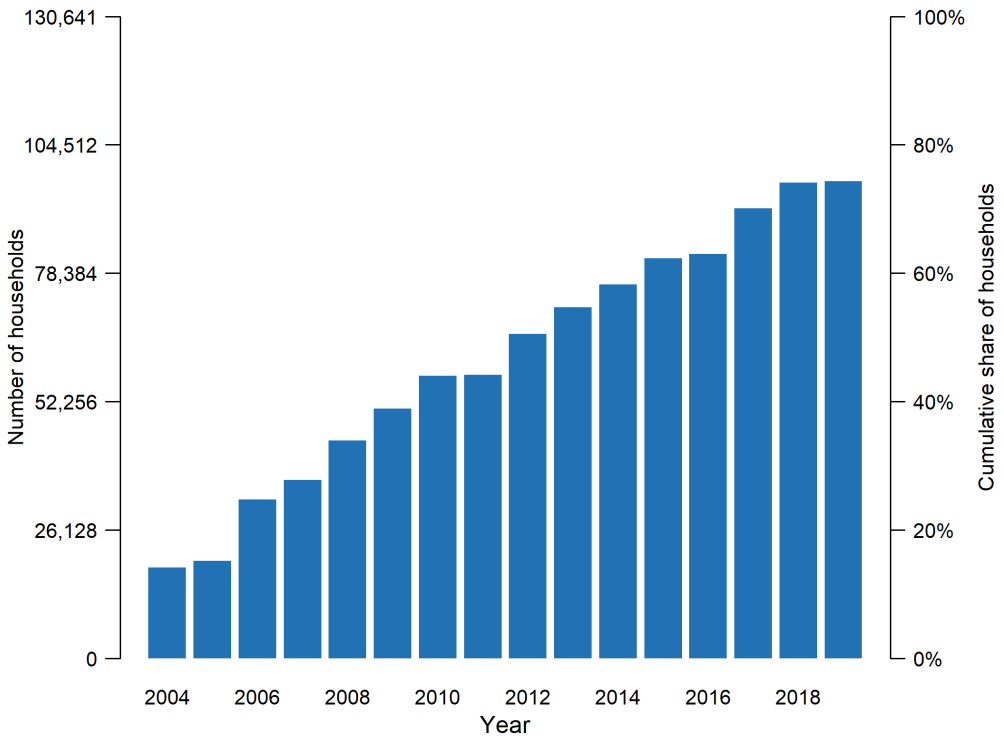


Figure F.6: Cumulative household treatment over time

Notes: This figure displays the cumulative number of households with AirNow forecast coverage over the sample period, from 2004 to 2019. The left y-axis reports the number of households, and the right y-axis reports the cumulative share relative to all households in the sample. A household is classified as treated beginning in the year when AirNow forecasts first become available for its ZIP code.

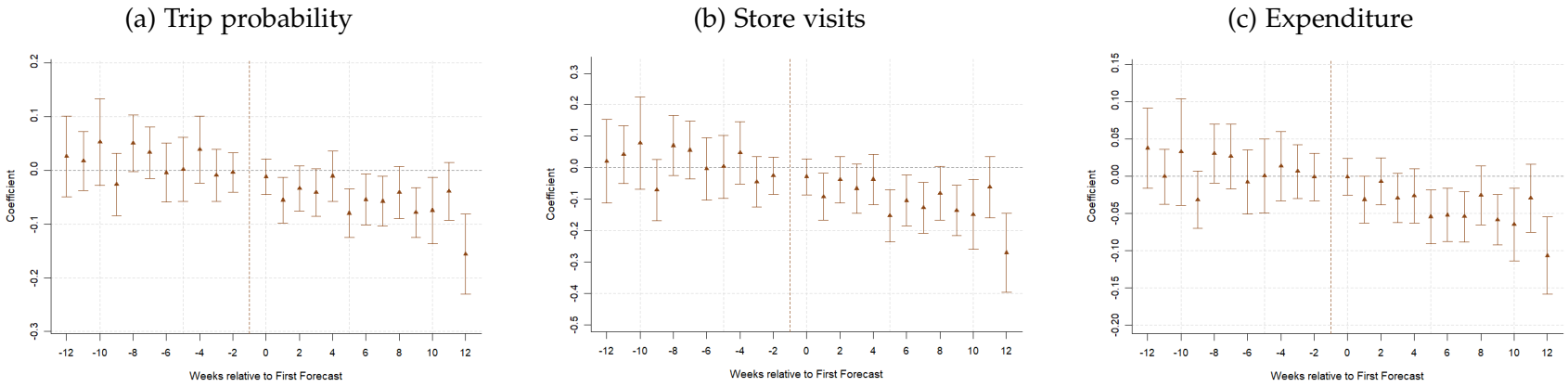
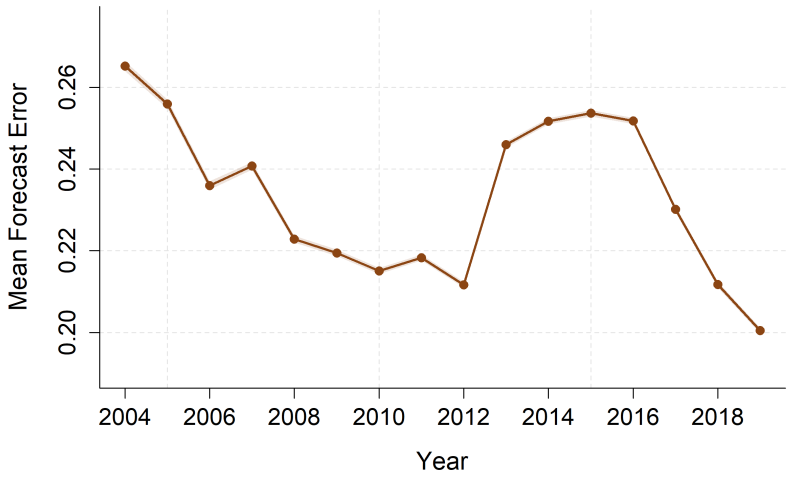


Figure F.7: Event Study of Shopping Responses to Forecast Access Using Alternative Windows

Notes: This figure presents IV event-study estimates of how AirNow forecast access changes the shopping–pollution gradient using an extended $\pm$12-week window. Units are percentage points for trip probability, visits per 100 household-days for store visits, and dollars per household-day for expenditure. Daily observations are aggregated to event-time weeks relative to the first week of forecast access, with week -1 omitted as the reference period. Each point is a coefficient estimate, and bars represent 95% confidence intervals. Standard errors are clustered at the county level. Joint Wald tests fail to reject the null that all pre-period leads equal zero ($p = 0.112, 0.096, 0.120$ for panels (a)–(c)).

(a) National trend of annual mean absolute forecast errors



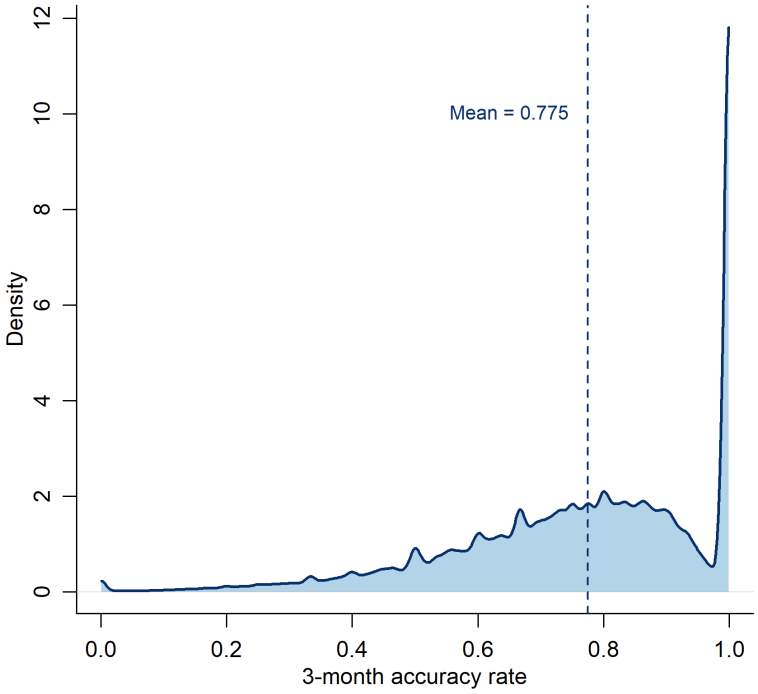
(b) Distribution of forecast errors

Forecast Error (AQI category steps)	Share %
Under-predict by 4 categories	0.0001
Under-predict by 3 categories	0.003
Under-predict by 2 categories	0.034
Under-predict by 1 category	7.444
Exact category match	77.048
Over-predict by 1 category	15.385
Over-predict by 2 categories	0.084
Over-predict by 3 categories	0.003
Over-predict by 4 categories	0.00003

Figure F.8: Trends and Distribution of AirNow Forecast Errors

Notes: Panel (a) plots the annual mean contemporaneous absolute forecast error, defined as the absolute difference between the AirNow forecasted AQI category and the realized AQI category, averaged across observations in each year from 2004 to 2019. The increase in mean forecast error around 2015–2016 may partly reflect unusually high and difficult-to-predict wildfire events during that period. Panel (b) reports the distribution of forecast errors across all observations in the sample. Positive values (over-predict) indicate the forecast assigned a worse air quality category than was realized, while negative values (under-predict) indicate the forecast assigned a better category than was realized.

(a) Distribution of 3-month forecast accuracy rate



(b) Summary statistics of 3-month forecast accuracy rate

Statistic	Value	Demeaned Value
Minimum	0.000	-0.775
1st Quartile	0.657	-0.118
Median	0.803	0.028
Mean	0.775	0.000
3rd Quartile	0.965	0.190
Maximum	1.000	0.225

Figure F.9: Distribution of AirNow 3-Month Forecast Accuracy Rate

Notes: Panel (a) plots the kernel density of the three-month rolling forecast accuracy rate over the Table 6 estimation sample. The accuracy rate is defined as the share of forecast days in the previous three months for which the AirNow forecasted AQI category matches the realized AQI category. The dashed vertical line marks the sample mean (0.775). The spike near 1.0 reflects a large share of observations with near-perfect forecast accuracy. Panel (b) reports summary statistics for the accuracy-rate distribution. The demeaned column reports each statistic relative to the sample mean.

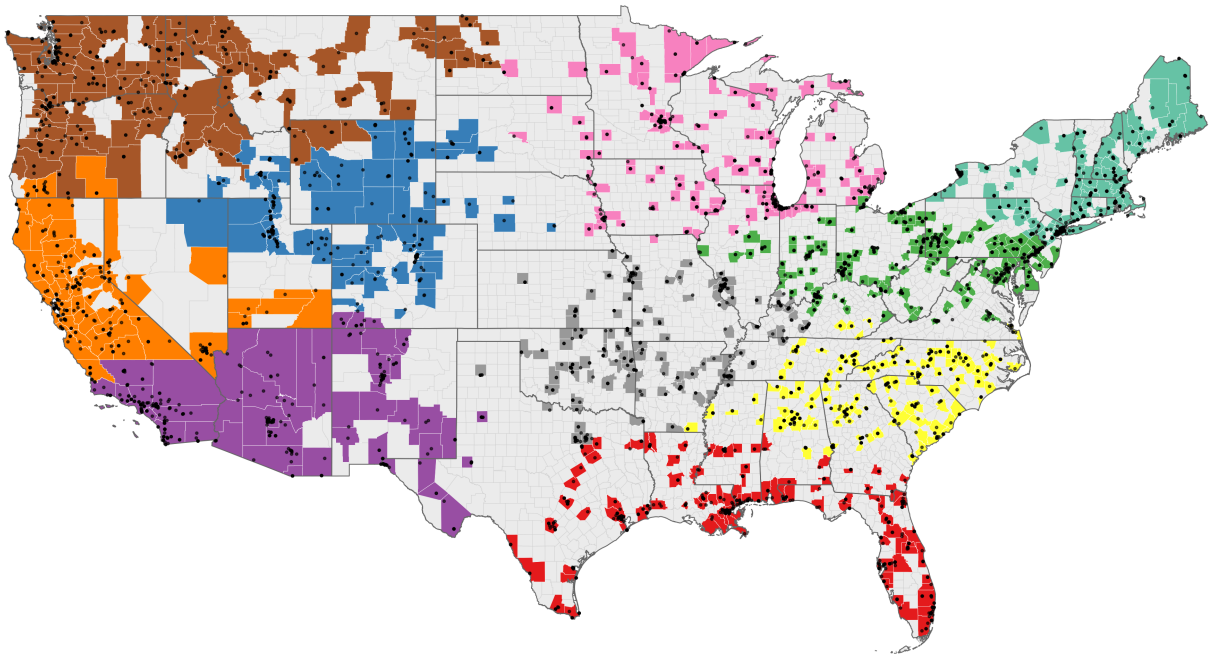


Figure F.10: Spatial Cluster Assignment (k=10) for continental U.S. Counties

Notes: This map shows continental U.S. counties assigned to one of 10 spatial clusters used in the alternative wind-direction IV. Counties are shown only if they contain at least one $PM_{2.5}$ monitor. The clusters are constructed using k-means based on the geographic locations of EPA Air Quality System monitors. Black dots indicate $PM_{2.5}$ monitor locations.

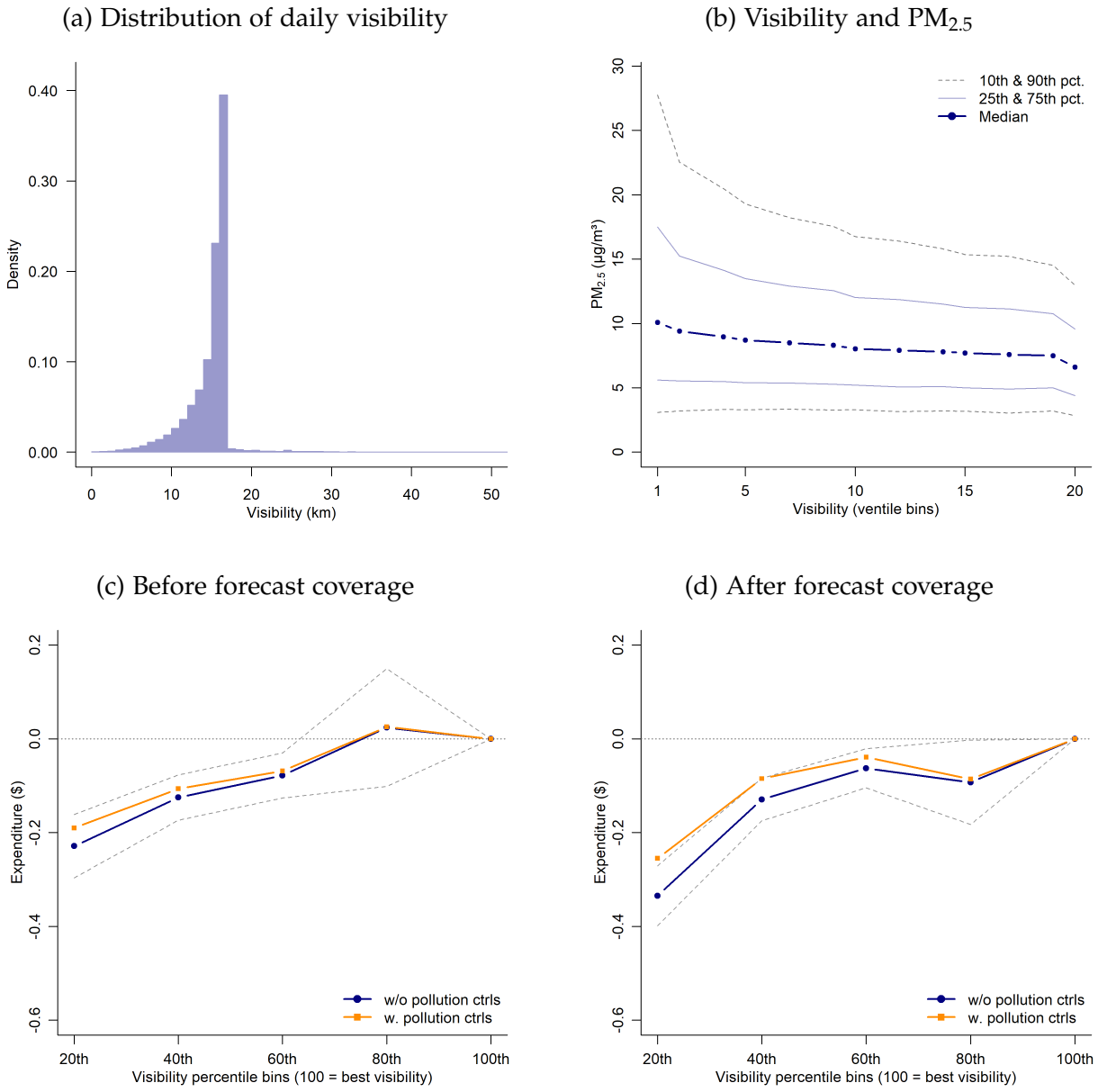


Figure F.11: Visibility as Private Information on Air Pollution

Notes: Panel (a) shows the distribution of raw visibility measure at the county–day level. The spike around 16 km reflects heaping at the commonly reported 10-mile visibility value, rather than a sharp change in underlying atmospheric conditions. Panel (b) shows the median, interquartile range, and 10-90 percentile range of $PM_{2.5}$ by ventile bins of visibility. Panels (c) and (d) plot household expenditures by visibility bins before and after forecast coverage, respectively. The regressions underlying Panels (c)–(d) control for the same set of covariates and fixed effects as our primary specification. The blue lines with solid circles show the effect of visibility without controlling for $PM_{2.5}$, while the orange lines with squares show the effect after controlling for $PM_{2.5}$. Dashed lines show 95% confidence intervals for the specifications without pollution controls, calculated using standard errors clustered at the county level.

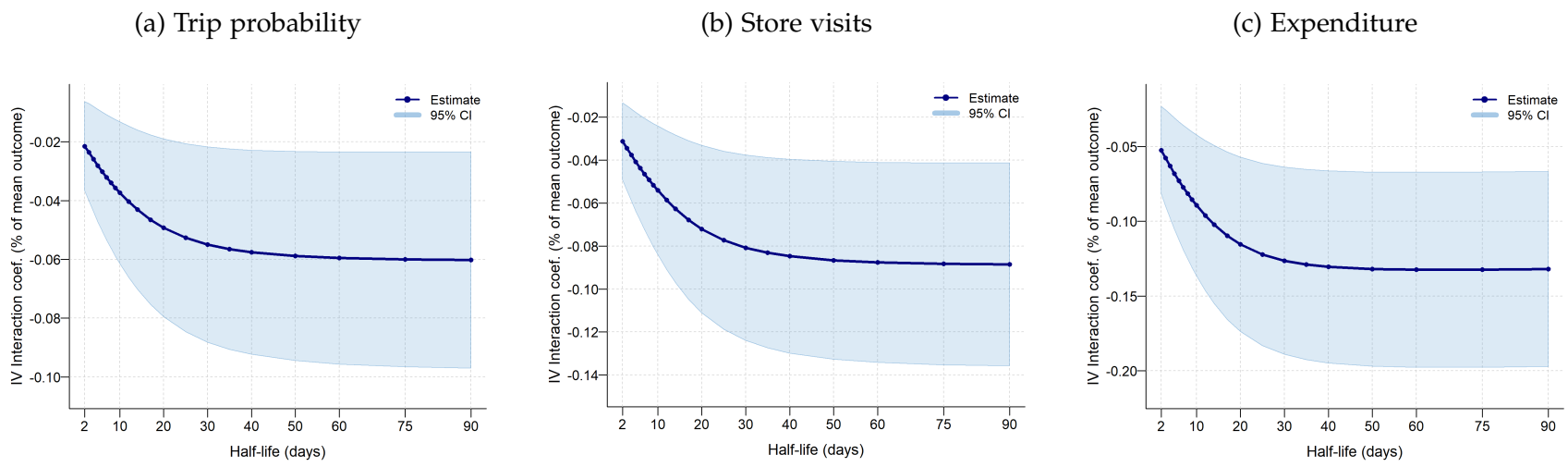


Figure F.12: Robustness: Exponentially Weighted Forecast Accuracy

Notes: Each panel corresponds to a different shopping outcome and plots the IV coefficient of the interaction term, $PM_{2.5} \times$ Forecast Accuracy, as a function of the half-life parameter used to construct the exponentially weighted forecast-accuracy measure. The half-life parameter ranges from 2 to 90 days. A shorter half-life places greater weight on very recent forecast performance, whereas a longer half-life incorporates a longer history of past forecast correctness. Forecast accuracy is constructed as an exponentially weighted average of past binary forecast correctness within a trailing 90-day window. Correctness equals one if the previous day's AQI category forecast matched the realized AQI category and zero otherwise. The measure is centered at its in-sample mean. The shaded band represents the 95% confidence interval based on standard errors clustered at the county level. The sample is restricted to ZIP-code-days with at least 10 forecast reports in the prior 90 days.