Beliefs, Information Trust, and Air Pollution

Yuting Wang*

Abstract

This paper examines how the quality of disclosures shapes public trust in official data. Using daily emission records from Waste-to-Energy plants in China (2020 - 2022) and housing data from 92 cities, I estimate pollution beliefs through a residential sorting model. Households perceive higher and more uncertain emissions when records are flawed, despite technical explanations of the flaws. There is limited evidence of manipulation by plants, suggesting fundamental distrust. The estimated household's annual willingness to pay for a 1% improvement in information quality is $1.23. The results show that low-quality disclosure discounts the value of information and erodes information trust.

JEL Classifications: H41, Q58, R21

*Postdoctoral Researcher, UMR GAEL, CNRS, INRAE, Université Grenoble Alpes, Grenoble, France, yuting.wang@univ-grenoble-alpes.fr.

I thank Sam Altmann, Sebastian Axbard, Simon Franklin, Lucie Gadenne, Cristina Gualdani, Emmanuel Guerre, Jan Knoepfle, and Ralf Martin for helpful thoughts and discussions. I am also grateful to participants at Queen Mary Internal Seminar, CESAER-INRAE Internal Seminar, SMUE-China Workshop, EAERE 2025, and ESWC 2025 for comments and suggestions. All errors are my own.

I. Introduction

Information disclosure is a central regulatory instrument used to promote accountability and influence behavior, in context such as environmental quality, product safety, financial reporting, and public health. Administrative records, however, are often incomplete, irregular, or selectively reported. These credibility frictions raise fundamental questions about how individuals form beliefs under imperfect disclosure, and how the value of disclosure depends on trust in information. Empirical evidence on this process remains limited, largely because beliefs are not directly observed. This paper explores these questions through the lens of pollution disclosure. By developing an empirical framework that recovers belief distributions from revealed location preferences, it examines how individuals trust and respond to official disclosure and how they value information quality.

The empirical setting is a nationwide disclosure system for daily emissions from municipal solid Waste-to-Energy (WtE) plants in China. Information quality fluctuates exogenously across plants and over time due to flaws in the records. I use this variation to estimate household beliefs about pollution through a residential sorting model. To address growing public concerns about waste incineration, China launched the first nationwide emission disclosure platform for WtE plants on January 2, 2020.[1] It provides public access to daily incinerator-level concentrations of major pollutants and five-minute furnace temperature readings. However, WtE plants do not always provide valid emission records. Some records are flawed, marked by missing entries or abrupt fluctuations in temperature readings. Although regulations recognize technical reasons for flawed records, such as maintenance or equipment malfunction, some plants have produced flawed data to conceal excess emissions or misclassified violations as technical problems.

Public beliefs about pollution in response to flawed records therefore depend on the extent to which people trust the stated technical explanations. When the flaws are viewed as genuine technical issues, they are unlikely to alter beliefs. When perceived as deliberate manipulation, they increase public skepticism and raise both the level and uncertainty of perceived pollution, as

[1] The platform is accessible at https://ljgk.envsc.cn.

households may discount official disclosures and seek external information sources. Assuming that beliefs are represented by a distribution and formed separately under valid and flawed records, the level of information trust can be inferred by comparing these distributions with official data. Convergence between beliefs and the official benchmark, irrespective of flaws, signals stronger trust in official data. It shows that households consistently rely on the stated information when forming their perceptions.

I estimate households' pollution beliefs using a discrete choice model of residential sorting. Within this framework, households maximize utility by jointly choosing consumption of composite goods, housing services, and perceived pollution exposure from the nearest operating WtE plant. The utility specification includes fixed effects for fine geographic units to control for unobserved local attributes that affect location attractiveness. Pollution exposure depends on perceived emissions, distance to the plant, and local wind conditions. These factors enter the utility function through an exposure function derived from the Gaussian Plume Model, a typical chemical dispersion model. I assume households form their perceptions of emissions based on the records observed in the 30 days preceding their transactions. To approximate overall emission levels, I construct a daily pollution index that aggregates multiple pollutant concentrations relative to their regulatory limits.

As the pollution index is not always observed because of flaws, I allow beliefs to differ between valid and flawed records. Specifically, I model beliefs as a mixture of two independent lognormal distributions, with lognormality capturing both the non-negativity and skewness of emission levels. The first component centers on the average (log) pollution index for days with valid records. This specification assumes no systematic bias in beliefs, justified by evidence that large-scale data manipulation is difficult (Greenstone et al., 2022). I further include a variance term to capture noise in beliefs, which may arise from cognitive limitations, alternative information sources, or inattention. A larger variance indicates greater belief dispersion, suggesting that even with valid records, households do not fully interpret or rely on the disclosed information. The second component represents beliefs in the absence of valid data, modeled as a separate lognormal distribution with an unknown mean and variance.

Overall beliefs are a weighted combination of the two components, with the share of days with valid records determining the weight on the first. This weight varies exogenously across households because of differences in choice sets and transaction timing. The simple specification provides a general representation of belief formation under mixed credible and questionable signals. Its unrestricted mean and variance parameters allow beliefs to shift in either direction and to vary in dispersion as information quality varies. This flexibility makes the framework applicable to a wide range of settings where credibility frictions influence belief formation and the value of information. Further details are provided in Section IV.D.

I estimate the model using a methodology extended from Bayer, Keohane and Timmins (2009). I compile housing data for 92 cities from Beike, a leading real estate platform in China that provides comprehensive coverage of the second-hand housing market. The dataset includes detailed information on location, price, transaction and listing dates, and other housing attributes. To capture city-level heterogeneity in perceptions, I estimate the model separately for each city. The results show that in most cities, households perceive emissions to be higher when they observe flawed records than when they observe valid records. The variance is significantly positive in both components, indicating a persistent gap between perceptions and disclosures, even when the data are valid. Nevertheless, the variance under flawed records is larger and more dispersed across cities. These findings suggest that households inflate their perceptions of pollution and feel greater uncertainty when emission data contain flaws.

Further analysis examines why households perceive higher emissions under flawed records in these cities (hereafter *over-perception* cities). I first test whether plants in over-perception cities manipulate flawed records to conceal excessive emissions. Using external $PM_{2.5}$ data as a proxy for pollution at plant locations, I compare pollution levels on flawed-record days with those on adjacent days. The results provide limited evidence of higher emissions on flawed-record days in over-perception cities, suggesting that residents may hold a fundamental distrust of plants when they encounter flawed records. Second, I compare city characteristics between over-perception cities and others. Over-perception cities tend to have larger waste incineration scales, higher industrial

pollution, weaker enforcement, and stronger public opposition to WtE projects before construction. In contrast, over-perception is less likely when the mayor or party secretary has local origins. Third, weaker institutional information quality, proxied by average information quality over the sample period, is associated with greater over-perception under flawed records and higher uncertainty under valid records. This suggests that a poor disclosure system heightens household mistrust and reduces reliance on official data, regardless of its validity.

Finally, I quantify the monetary value of information quality. First, I estimate the marginal willingness to pay (MWTP) for information quality. With the current average information quality at 0.85, the estimated annual MWTP for a 1% improvement is \$1.23. The annual total WTP for raising the current level to one is \$19.54, while the loss from a decline to 0.5 is \$37.29. Second, I examine the substitution effect between information quality and distance. Raising information quality from 0.85 to 1 reduces the preferred distance by 16.82 meters for households located five kilometers downwind from the plant. Assuming wind blows steadily toward true north, this shift corresponds to an aggregate inward adjustment of 86,648 m^2 around the plant. Conversely, a decline in information quality can lead to a sizable outward shift, signaling broad economic costs in land-use efficiency and housing surplus.

This study contributes to the literature in the following three dimensions. First, it aligns with a substantial body of research on public responses to pollution information (Barwick et al., 2024; Bresnahan, Dickie and Gerking, 1997; Chang, Huang and Wang, 2016; Chen, Oliva and Zhang, 2022; Gao, Song and Timmins, 2023; Greenstone et al., 2022; Hanna et al., 2021; Ito and Zhang, 2020; Neidell, 2009; Xie, Yuan and Zhang, 2023; Zhang and Mu, 2018), while distinguishing itself by shifting the focus from how people react to available information to how they interpret and trust it. The paper develops a simple belief framework and an empirical identification strategy that comprehensively capture both preferences and aversions toward information. By showing that households interpret identical emission data differently and often discount records that appear flawed, it uncovers a fundamental information friction that arises before behavioral responses. This perspective advances understanding of how information *quality*, rather than the mere *availability*

of data, influences environmental and economic outcomes.

Second, this paper adds to the existing studies on pollution information disclosure and housing markets (Agarwal et al., 2023; Mastromonaco, 2015; Moulton, Sanders and Wentland, 2024; Wang and Yang, 2024), where most studies document a negative impact of pollution disclosure on nearby property values. A closely related study is Nie et al. (2025), which exploits China's 2017 "Installing, Erecting, and Networking" policy (discussed in Section II) as a quasi-natural experiment. It finds that the launch of real-time disclosure of WtE emissions flattens the housing price gradient by mitigating risk perceptions and strengthening public trust. My analysis differs by emphasizing the importance of information quality. It shows that low-quality disclosure can undermine transparency by fostering skepticism, amplifying perceived risks, and weakening information trust.

Third, this study contributes to the emerging literature on data quality in pollution information disclosure (Mu, Rubin and Zou, 2024; Zhang et al., 2019). A related study by Wang et al. (2022) analyzes continuous emission monitoring system (CEMS) data from key polluting firms in Henan Province, China, between 2017 and 2019. They find that about one-third of firms failed to meet official standards for data completeness, with some deliberately suspending CEMS operations to hide excessive emissions. My analysis focuses on how information receivers, rather than producers, respond to flawed data. The results reveal that households distrust emission records with flaws even when evidence of manipulation is limited. These findings suggest that poor data quality alone can undermine fundamental trust. By quantifying households' willingness to pay for information quality, the paper demonstrates that credible disclosure has substantial monetary value. Improving data reliability, therefore, generates welfare gains that extend beyond regulatory compliance.

The paper proceeds as follows. Section II reviews the background of pollution information disclosure for WtE plants. Section III describes the main data sources. Section IV presents the empirical framework. Section V reports the estimation results and additional analyses. Section VI concludes.

II. Background

A. *Pollution Disclosure in WtE Plants*

Since 2012, waste incineration in China has expanded rapidly. Appendix Figure A2 presents the annual number of WtE plants. The expansion accelerated following the 2012 *Notice on Improving the Pricing Policy for Waste-to-Energy Incineration*, which raised subsidies for new plant construction. However, the rapid expansion of WtE plants has triggered strong public concern and opposition.[2] A major issue is the mismatch between the regulated safe distance of 300 meters and the larger distances perceived by the public as necessary. Appendix Section A illustrates this gap with examples from public complaints. Many residents remain dissatisfied even when plants are located several kilometers away, far beyond the official standard.

A lack of transparency is a primary source of public concern. Consequently, the government introduced the "*Install, Display, Connect*" policy in 2017. The policy required WtE plants to *install* continuous emissions monitoring systems (CEMS), *display* the data on electronic screens outside the plants, and *connect* the data online with the Environmental Protection Agency (EPA). However, implementation was limited. By 2018, only 191 of 359 operational plants disclosed environmental information, and 62% of these did not update emission data in a timely manner.[3]

To improve transparency, the Ministry of Ecology and Environment (MEE) launched the *Automatic Monitoring Data Disclosure Platform for Municipal Solid Waste-to-Energy Plants* (hereinafter "the platform") on January 2, 2020. The platform discloses information for all plants, including address, capacity, start date, and contact details. It provides daily emission data for each incinerator, including average concentrations of particulate matter (PM), nitrogen oxides (NO_x), sulfur dioxide (SO_2), hydrogen chloride (HCl), and carbon monoxide (CO). It also records daily furnace temperature curves constructed from five-minute averages. Furnace temperature serves as a proxy for dioxin emissions, as real-time monitoring is technologically infeasible. Since most

[2] Notable incidents include the 2014 protest in Hangzhou and the 2016 protest in Zhaoqing.

[3] *Source*: Wuhu Ecology Centre (in Chinese).

dioxins decompose above 850 °C, regulations require plants to maintain furnace temperatures above this threshold during normal operations.

The platform has significantly improved public access to pollution monitoring. Appendix Figure A3 shows a sharp increase in searches on waste incineration after its launch, especially in cities with WtE plants. By 2021, the platform had recorded over 60 million visits,[4] indicating strong public awareness and engagement.

B. Flawed Emission Records

High-frequency emission data, however, face several technical challenges. A common issue is the regular maintenance of CEMS, which is necessary for accuracy but temporarily halts data collection. Technical failures, such as communication disruptions or thermocouple malfunctions, also occur. These problems lead to missing or faulty records, which undermine the reliability of pollution information.

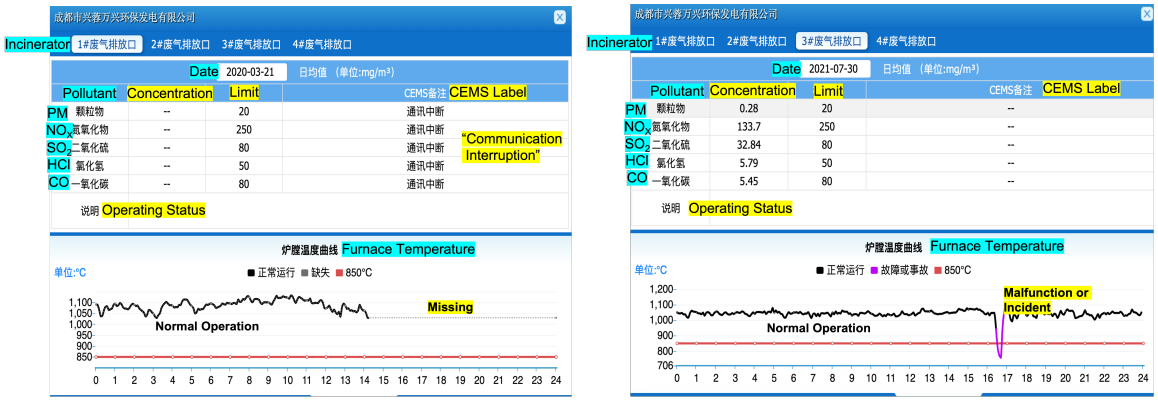
To ensure completeness and validity of monitoring data, the MEE issued a guideline in 2019 that defines four labels indicating the technical reasons for flawed emission records: (1) *CEMS Maintenance* for periods of calibration, malfunction, or repair of CEMS; (2) *Communication Interruption* for disruptions in data transmission due to network or equipment issues; (3) *Furnace Temperature Anomaly* for cases where furnace temperatures fall below 850 °C due to force majeure; (4) *Thermocouple Malfunction* for instances where faulty thermocouples fail to reflect temperatures accurately because of fouling or damage. Unlabeled data are treated as valid and regulated under normal emission standards.[5]

Figure 1 displays sample platform screenshots with two flawed emission data types: missing and faulty. Each screenshot contains two panels. The upper panel reports concentrations of five

[4] *Source*: the China Environmental Protection Association (in Chinese).

[5] The guideline also defines seven operational statuses for the furnace. When temperatures fall below 850 °C (e.g., during preheating or cooling), waste burning is prohibited. Plants are required to label these statuses on the platform. Failure to do so is considered a violation. Unlabeled operations are assumed to satisfy the 850 °C requirement for normal operation. *Source*: Guidelines for Automatic Monitoring Data Labelling of Municipal Waste-to-Energy Plants.

pollutants, and the lower panel shows the furnace temperature curve. Figure 1a shows missing records, where pollutant concentrations and part of the temperature curve are absent, labeled as "communication interruption." Figure 1b shows faulty records, where the temperature curve exhibits a sudden drop followed by rapid recovery, labeled as "malfunction or incident."



(a) Missing (b) Faulty

Figure 1. Flawed Emission Records

Note: The figures show sample platform screenshots with flawed emission data. Each screenshot corresponds to a specific incinerator on a given date. The upper panel reports concentrations of five pollutants (PM, NO_x, SO_2, CO, and HCl) relative to regulatory limits. The lower panel presents the furnace temperature at five-minute intervals. Both panels display labels where flawed records exist.

Regulations on Emission Data Manipulation In 2019, the MEE introduced new regulations on the management of automatic monitoring data for WtE plants.[6] Plants engaged in data manipulation or mislabeling face severe penalties. These include the withdrawal of renewable energy subsidies, monetary fines, production suspension, plant closure, and legal prosecution. For example, a plant in Taizhou labeled excess particulate matter emissions as "CEMS Maintenance" and was fined around $22,200. A plant in Zhengzhou labeled periods of substandard furnace temperature as "Malfunction" and was fined around $20,500.[7] Such violations suggest that, despite strict regulations, manipulation of emission data remains a persistent practice.

[6]The regulation is titled *Regulations on the Application and Management of Automatic Monitoring Data for Municipal Waste-to-Energy Plants*.

[7]*Source*: violations records published by the MEE (in Chinese): Taizhou, Zhengzhou.

III. Data

The analysis draws on multiple datasets. I compile information on emissions and locations of WtE plants, housing transactions and listings, household income, public complaints, and other spatial attributes from various sources to construct the final dataset.

Emissions from WtE Plants I collect incinerator-level emission data from 2020 to 2022 from the platform. The dataset contains each plant's geolocation and basic information, including designed waste incineration capacity, operational start date, and power generation capacity. It also reports daily average concentrations and emission limits for five pollutants (PM, NO_x, SO_2, CO, and HCl), as well as five-minute average furnace temperatures. As the number of WtE plants increased over the sample period, the dataset covers 585 plants with 1,388 incinerators in 2020, 809 plants with 1,840 incinerators in 2021, and 918 plants with 2,056 incinerators in 2022.

Housing Transactions and Listings I collect data on second hand housing transactions and listings from Beike, the largest platform for integrated online and offline real estate services in China. The dataset covers 92 cities from 2020 to 2022. It contains detailed information on list prices, listing dates, transaction prices and dates (if sold), locations, and housing characteristics. These characteristics include size, total number of building floors, and number of rooms. For properties without geo-coordinates, I use their addresses together with the Baidu Map API to obtain geographic coordinates. Additional details on the housing data are provided in Appendix Section B.

Wind I obtain wind data from the ERA5 Land dataset with a horizontal resolution of $0.1^\circ \times 0.1^\circ$ (approximately 9km $\times$ 9km). Wind speed and direction at each plant are derived using the Inverse Distance Weighting (IDW) method. Further details are provided in Appendix Section C.

Household Income Household disposable income is calculated as per capita disposable income multiplied by the average household size at the county level. The county is a standard administrative unit in China, positioned below the prefecture-level city. It encompasses counties, county-level cities, and urban districts. Per capita income data are taken from city level statistical yearbooks for 2020 to 2022, and averaged over this period. The average household size is obtained from the 2020 Population Census.

Supplementary Pollution Measurement To measure pollution at plants regardless of the availability of valid emission records in Section V.B, I use two proxy sources based on fine particulate matter smaller than 2.5 micrometers in diameter (PM2.5). The first source is station-based PM2.5 data from daily records of national monitoring stations in China.[8] The second source is machine-learning predictions of PM2.5 from He et al. (2023), reported at $0.01^\circ \times 0.01^\circ$ (approximately $1\text{km} \times 1\text{km}$). This dataset combines ground monitoring with daily MAIAC AOD data from NASA and additional predictors such as weather, surface features, and PM2.5 components. A random forest algorithm is used to impute missing values. For the first source, I interpolate monitoring data using the same method applied to wind data (see Appendix Section C). For the second source, I assign to each plant the value from the nearest grid point.

Public Complaints I collect 30,580 posts from Weibo, China's social media platform equivalent to X, for the period 2017–2022. The posts are identified using the Chinese keyword "waste incineration." From this dataset, I manually extract 1,314 posts that refer to specific WtE plants.

Other variables Information on local amenities is constructed from 2019 Point-of-Interest data from Baidu Map. Nighttime light data are obtained from Chen et al. (2020). Administrative punishment records related to WtE plants are collected from local EPA official websites for 2020–2022. Personal information on mayors and party secretaries is obtained from Baidu Baike. Grid-level

[8] Real-time monitoring data are available from the China National Environmental Monitoring Center, but no historical records are provided. I obtain historical data from Xiaolei Wang's website.

population density, measured as the number of people per square kilometer, is obtained from WorldPop. Other variables are drawn from the City Statistical Yearbook for 2020–2022.

Table 1. Descriptive Statistics

Variable	Mean	SD	Observations
Panel A: Housing - Traded			
Transaction price (log)	14.017	0.787	1,926,884
Listed price (log)	14.055	0.784	1,927,107
Area (m^2)	90.309	33.695	1,997,926
# Rooms	3.887	1.232	2,019,523
Distance to a WtE plant (km)	16.299	7.695	2,039,779
Panel B: Housing - Listed			
Listed price (log)	14.019	0.795	3,936,559
Area (m^2)	100.745	36.726	3,859,322
# Rooms	4.200	1.252	3,889,834
Distance to a WtE plant (km)	16.408	7.875	3,936,561
Panel C: Household Income (County Level)			
Per capita income (USD)	6,625.483	1,918.764	1,900
Household size	2.612	0.309	1,968
Panel D: Characteristics of WtE Plants			
Years of operation (by 2022)	5.123	4.314	918
Incineration capacity (t/d)	1,146.282	917.448	918
Power generation capacity (MW)	24.506	26.311	915
Panel E: Emissions			
Particulate matter	3.615	4.773	631,340
NO_X	140.170	53.561	631,337
SO_2	23.647	21.320	631,296
CO	8.588	10.551	631,319
HCl	13.027	9.866	631,321
P(temperature $> 850\,^{\circ}$C)	0.974	0.125	635,618
$\mathbb{I}$(flawed records)	0.146	0.353	663,619

Note: This table presents the descriptive statistics for key variables used in the analysis. Key housing characteristics include transaction price (in log), listed price (in log), property area (in square meters), number of rooms, and the distance to the nearest operating WtE plant (in kilometers). County-level per capita income is reported in USD. The characteristics of WtE plants include years of operation as of 2022, incineration capacity (in tons per day), and power generation capacity (in megawatts). P(temperature $> 850\,^{\circ}$C) denotes the proportion of five-minute intervals in a day during which the furnace temperature exceeds $850\,^{\circ}$C. $\mathbb{I}$(flawed records) is an indicator variable that equals one when daily emission records exhibit flaws.

Summary Statistics Table 1 reports the mean and standard deviation of the main variables. *Housing-Traded* refers to houses transacted between 2020 and 2022, while *Housing-Listed* denotes the remaining houses available for sale during the sample period. Both groups exhibit comparable housing characteristics and proximity to a WtE plant. Panel C illustrates that household size is stable across counties. In contrast, county-level per capita income varies considerably, which reflects substantial income disparities across regions.

Panel D reveals that WtE plants differ substantially in years of operation and capacity. Panel E shows substantial variation in concentration levels across pollutants, primarily due to differences in regulatory limits. To address this variation, I construct a comprehensive pollution index, which provides a unified measure of exposure. More details are provided in Section IV.A. Under valid records, the mean of P(temperature $>$ 850 °C) is 0.974, indicating that furnace temperatures are generally stable and compliant with the standard. On average, the probability of flawed records is 14.6 percent.

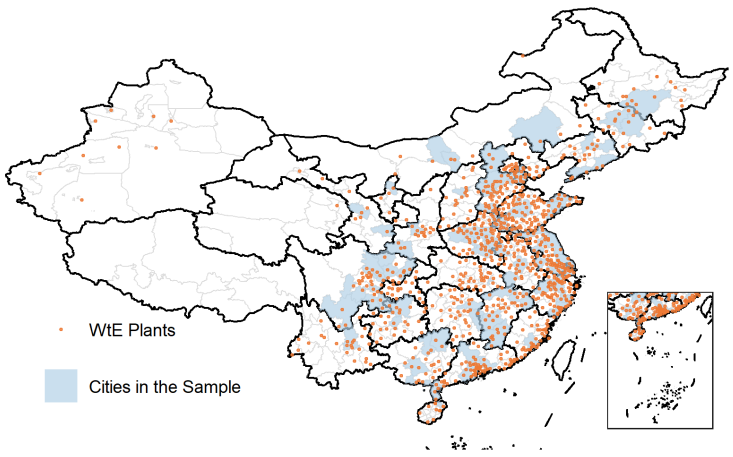


Figure 2. Geographical Distribution of WtE Plants and Selected Sample Cities in China

Note: The map shows the locations of WtE plants in China, represented by the points. The shaded areas indicate the 92 cities used in the sample for analysis.

Figure 2 shows the geographical distribution of WtE plants (points) and the 92 sample cities (shaded areas). Most plants are located in eastern China, where population density is high. The

sample cities are distributed relatively evenly across this region. Overall, the sample provides broad coverage of WtE industry and supports the representativeness of the analysis.

IV. Model

I develop a structural framework that explicitly models beliefs, information, and residential location choices. Section IV.A introduces the exposure function from the Gaussian Plume Model, which extrapolates emissions from the plant to each residence. Section IV.B specifies households' beliefs about emissions. Section IV.C presents the residential sorting model that incorporates exposure and beliefs into household utility. Section IV.D discusses the willingness to pay for information quality derived from the model.

A. Emissions from WtE Plants

To measure pollution exposure from the nearest operating WtE plant at each property location, I assume that emissions decay with distance following the Gaussian Plume Model. I then introduce two modifications to adapt the model to this study.

The Gaussian Plume Model (GPM), developed in the early 1960s by Pasquill (1961), provides an analytical solution to the transport-diffusion equations for continuous air pollutant releases. The GPM is widely used by regulatory agencies to predict the dispersion of pollutants from point sources, such as smokestacks or industrial plants (Snoun, Krichen and Chérif, 2023). The model assumes that pollutants disperse in a Gaussian distribution both horizontally and vertically as they move downwind from the source. Appendix Figure A4 illustrates the three-dimensional distribution of the dispersion.

According to the GPM, pollutant concentration decreases with distance from the source. It also depends on wind speed, atmospheric stability, and the effective stack height. Specifically, for a pollutant emitted from a plant with emission rate Q at an effective stack height H[9] under wind speed

[9]The effective stack height combines the physical stack height with the plume rise: $H = h_s + V_s \cdot \text{diameter} \left(1.5 + 2.68 \times 10^{-3} \cdot P \cdot \text{diameter} \cdot \delta T / T_s\right) / W$. Following Boré et al. (2022), h_s is the physical stack height

W, the ground-level concentration at point $(x, y, 0)$, where x is the distance along the downwind direction, can be expressed as

$$\kappa(x, y, 0) = \frac{Q}{\pi \sigma_y \sigma_z W} \exp\left(-\frac{y^2}{2\sigma_y^2} - \frac{H^2}{2\sigma_z^2}\right), \tag{1}$$

where σ_y and σ_z are the horizontal and vertical dispersion coefficients, expressed parametrically as $\gamma_1 x^{\alpha_1}$ and $\gamma_2 x^{\alpha_2}$. The coefficient values depend on the atmospheric stability category and are taken as given from established technical guidelines.[10] Let θ be the angle between the property–plant line and the downwind direction. Then $x = d \cdot \cos(\theta)$ and $y = d \cdot \sin(\theta)$, where d is the distance between the property and the plant.

Coping with Wind Direction The first modification of the GPM accounts for variation in wind direction. I compute frequency-weighted averages of the wind angle (θ) and wind speed (W). The calculation is restricted to wind directions that fall within 90° of the plant-to-property vector, where the property is considered downwind. I then construct r^{dw}, the total frequency of downwind occurrences, which scales pollution exposure in equation (2). Further details are provided in Appendix Section D.

A Comprehensive Pollution Index The second modification addresses the limitation that the GPM captures the dispersion of only a single pollutant. Health outcomes, however, reflect exposure to multiple pollutants. To address this, I construct a pollution index that serves as a proxy for Q.

(80 m). V_s is the stack gas velocity (16 m/s). The stack inner diameter is set to 2 m. P is the atmospheric pressure (1013.25 mb). T_s is the stack exit temperature (413.27 K). δT is the temperature difference between the stack and the ambient air (298.15 K).

[10] Following Boré et al. (2022), I assume stability class B (moderately unstable) for all sites. This class is associated with moderate vertical mixing and moderate wind speeds. Coefficient estimates are drawn from the *Technical Methods for Establishing Local Emission Standards of Air Pollutants (1992)*. For downwind distances below one kilometer, $\alpha_1 = 0.9144$ with $\gamma_1 = 0.2818$. For distances above one kilometer, $\alpha_1 = 0.8650$ with $\gamma_1 = 0.3964$. The vertical parameters follow the same structure: $\alpha_2 = 0.9614$ and $\gamma_2 = 0.1272$ for distances below one kilometer, and $\alpha_2 = 1.0934$ with $\gamma_2 = 0.0570$ for distances above one kilometer.

This index is defined as the geometric mean of two components, $Q_{\text{pollutants}}$ and Q_{dioxin}.

The first component, $Q_{\text{pollutants}}$, aggregates five pollutants: particulate matter, NO_x, SO_2, CO, and HCl. For each pollutant, the observed concentration c_i is divided by its emission limit s_i. This standardizes across pollutants and highlights severity relative to regulatory benchmarks.[11] These ratios are then averaged to obtain $Q_{\text{pollutants}}$:

$$Q_{\text{pollutants}} = \frac{1}{5} \sum \frac{c_i}{s_i}.$$

The second component captures dioxin emissions based on furnace temperature. As aforementioned, effective control requires operation at or above 850 °C. Accordingly, I define Q_{dioxin} as the daily share of five-minute intervals in which the temperature falls below 850 °C:

$$Q_{\text{dioxin}} = \frac{\text{number of 5-minute intervals with temperature} < 850\,^{\circ}\text{C}}{\text{total number of 5-minute intervals}}.$$

Updated GPM After incorporating the modifications for wind and multiple pollutants, the ambient pollution level at a household is expressed as a function of the plant–property distance (d), the angle (θ), the wind speed (W), the downwind probability (r^{dw}), and the pollution index (Q). This yields the following specification:

$$\kappa(d, \theta, W, Q) = \exp\left[r^{dw} \left(\log \frac{Q}{\pi W} + D(d, \theta) \right) \right], \tag{2}$$

where

$$Q = \sqrt{Q_{\text{pollutants}} \times Q_{\text{dioxin}}} \quad ,$$

$$D(d, \theta) = -\left[d^{2-2\alpha_1} \frac{(\sin\theta)^2}{2\gamma_1^2 (\cos\theta)^{2\alpha_1}} + d^{-2\alpha_2} \frac{H^2}{2\gamma_2^2 (\cos\theta)^{2\alpha_2}} + (\alpha_1 + \alpha_2) \log(d \cos\theta) + \log(\gamma_1 \gamma_2) \right].$$

[11] For plants with multiple incinerators, c_i is the maximum concentration across incinerators. If no incinerator is in operation, c_i is set to zero.

B. Household Emissions Beliefs

The belief specification follows from a general signal extraction problem under uncertain information quality. Households observe a mix of valid and flawed records but cannot determine whether a flawed record arises from a benign technical issue or from an attempt to obscure excessive emissions. The resulting uncertainty creates two distinct perceived regimes. One regime is anchored on the disclosed measurements when records are valid. The other is anchored on beliefs formed when records are flawed. To capture these regimes and allow direct comparison, I estimate separate belief distributions associated with each type of record.

I assume households form beliefs about emissions from records disclosed in the 30 days preceding a transaction. This window captures the information most relevant for household decisions and provides a representative measure of recent emission levels.[12] I then represent household beliefs as a distribution of the pollution index Q. The distribution consists of one component for observed valid records and another for unobserved valid records, that is, flawed records:

$$\log Q = \text{Info} \cdot [\log Q^o + \varepsilon^o] + (1 - \text{Info}) \cdot \log Q^u, \tag{3}$$

where Info denotes the level of information quality, which serves as the weight for the observed component. It is measured as the proportion of days with valid records during the reference period. $\log Q^o$ is calculated as the average of the log-transformed pollution index on days with valid records, following the method in Section IV.A. The additional term ε^o captures idiosyncratic deviations in perception from the disclosed data. It is assumed to follow a normal distribution with mean zero and variance σ_o^2. The unobserved component, $\log Q^u$, is also assumed to be normally distributed, with mean μ_u and variance σ_u^2. Both ε^o and Q^u are treated as independent random variables.

The beliefs specification assumes no systematic log bias when records are valid. This assumption

[12]The plausibility of the chosen information window is supported by Appendix Table A1 and Figure A1, which compare the 30-day moving averages of information quality and emission levels with 60- and 90-day windows. Both show no statistically significant differences in their distributions.

is plausible for two reasons. First, systematic manipulation of data is costly and difficult (Greenstone et al., 2022). Second, the government incorporates public reporting of violations from the real-time emission platform and responds to them effectively (Buntaine et al., 2024). The model nevertheless allows for heterogeneity in how households interpret pollution levels. This heterogeneity is captured by a noise term, which measures the dispersion of beliefs around the observed data. Such dispersion may arise from limited attention to official reports or cognitive constraints in interpreting data. It may also result from alternative information sources, such as local news, social networks, personal observation of air quality, or other unofficial signals.

Belief Parameters and Information Trust Under this framework, information trust is inferred from belief parameters. When trust is high, households interpret emissions in line with official disclosure, and both belief components converge toward the official benchmark. First, the mean μ_u of the unobserved component should approach the mean of $\log Q^o$. This convergence implies that households view flawed records as technical issues rather than as evidence of systematic manipulation. Second, the variances σ_o^2 and σ_u^2 should approximate the variance of the disclosed series. Perfect convergence may not hold because of household heterogeneity in attention or cognitive capacity, but relative proximity remains informative. A weaker alignment of σ_u^2 relative to σ_o^2 indicates that households supplement flawed records with external signals and rely less strongly on official disclosure.

C. Residential Sorting Model

I extend the model of Bayer, Keohane and Timmins (2009) to investigate how information disclosure shapes pollution perceptions and housing choices. In the discrete choice framework, households simultaneously choose a location, housing services, and a composite commodity, with the location choice conditional on amenities such as air quality. I adapt this approach to examine housing choices near a localized pollution source (i.e., a WtE plant), rather than city-wide location decisions. I assume that households consider properties listed for sale within the same city as the purchase.

The choice set is restricted to properties within 30 square meters of the size of the chosen property. This restriction avoids noise from alternatives with limited substitutability caused by large size differences.

I introduce two additional elements into the household utility function. The first is a pollution exposure function, based on the GPM and defined in equation (2), which measures exposure from WtE plants. The second is a belief specification, given in equation (3), which recovers households' beliefs about emission levels from their location choices.

The utility of household i for property k, $U_{i,k}$, follows a Cobb–Douglas form. It depends on three major factors: the consumption of a composite good C_i, housing services H_i, and pollution from the nearest operating WtE plant. Housing services include both the physical attributes of a property (e.g., size, number of rooms) and its micro-location characteristics, such as access to public facilities, nearby commercial activity, and exposure to local externalities. The utility function is expressed as follows:

$$\widetilde{U}_{i,k} = C_{i,k}^{\beta_C} \cdot H_{i,k}^{\beta_H} \cdot \left(\frac{W_k}{Q_k}\right)^{\beta_Q \cdot r_k^{dw}} \cdot (\exp\left[-D_k\right])^{\beta_d \cdot r_k^{dw}} \cdot \exp\left(\xi_k + \eta_{i,k}\right), \tag{4}$$

where $\eta_{i,k}$ is an idiosyncratic error term that captures household-specific preferences. ξ_k denotes a fixed effect capturing unobserved location characteristics, such as local public service quality, environmental amenities, and other persistent factors that influence area attractiveness.

The variables $D_k = D(d_k, \theta_k)$, Q_k, W_k, and r_k^{dw} are derived from the GPM, where Q_k denotes the perceived emissions specified in equation (3). As per equation (2), when $\beta_d = \beta_Q$, the utility function corresponds to the inverse of pollution at location k as specified by the GPM. In practice, however, households cannot calculate pollution dispersion precisely as in the theoretical benchmark. They rely instead on imperfect signals such as distance and other observable proxies. Allowing $\beta_d \neq \beta_Q$ relaxes the benchmark and captures deviations from the GPM. Such deviations may arise from local topography, precautionary measures such as insulation or air filters, or heterogeneous household cognitive skills and preferences regarding distance.

Integrating over ε^o and Q^u from equation (3) yields the household expected utility function:[13]

$$\begin{aligned} U_{i,k} = C_{i,k}^{\beta_C} \cdot H_{i,k}^{\beta_H} \cdot W_k^{\beta_Q \cdot r_k^{dw}} &\times \exp\left[\beta_Q \cdot r_k^{dw} \cdot \left(-\mu_u + \text{Info}_k \cdot \left(\mu_u - \log Q_k^o\right)\right)\right] \\ &\times \exp\left[\left(\beta_Q \cdot r_k^{dw}\right)^2 \cdot \left(\sigma_o^2 \frac{\text{Info}_k^2}{2} + \sigma_u^2 \frac{(1-\text{Info}_k)^2}{2}\right)\right] \\ &\times \exp\left(-\beta_d \cdot r_k^{dw} \cdot D_k\right) \cdot \exp\left(\xi_k + \eta_{i,k}\right). \end{aligned} \tag{5}$$

Assume that the composite commodity price is normalized to 1. Let p_k denote the price of housing services for property k, and let $M_{i,k}$ represent the income of household i associated with choosing property k. The selected property K then satisfies the following condition:

$$(K, C_{i,K}, H_{i,K}) = \arg\max_{k, C_i, H_i} \left\{U_{i,k};\ C_{i,k} + p_k H_{i,k} = M_{i,k}\right\}. \tag{6}$$

Maximizing the utility for each k with respect to $C_{i,k}$ and $H_{i,k}$ under the budget constraint gives

$$C_{i,k} = \frac{\beta_C}{\beta_C + \beta_H} \cdot M_{i,k}, \quad H_{i,k} = \frac{\beta_H}{\beta_C + \beta_H} \cdot \frac{M_{i,k}}{p_k}.$$

Substituting these into equation (6) gives the indirect utility which is proportional to

$$\begin{aligned} V_{i,k} = \exp&\left[\beta_M \cdot \log M_{i,k} - \beta_H \cdot \log p_k - \beta_Q \cdot r_k^{dw} \cdot (\text{Info}_k \cdot \log Q_k^o - \log W_k) - \beta_d \cdot r_k^{dw} \cdot D_k\right] \\ &\times \exp\left[-\beta_Q \cdot \mu_u \cdot r_k^{dw} \cdot (1 - \text{Info}_k) + \left(\beta_Q \cdot r_k^{dw}\right)^2 \cdot \left(\sigma_o^2 \frac{\text{Info}_k^2}{2} + \sigma_u^2 \frac{(1-\text{Info}_k)^2}{2}\right)\right] \\ &\times \exp\left(\xi_k + \eta_{i,k}\right). \end{aligned} \tag{7}$$

where $\beta_M = \beta_C + \beta_H$. This equation formalizes how the belief parameters are identified in the discrete choice model. First, variation in $r_k^{dw} \cdot (\text{Info}_k \cdot \log Q_k^o - \log W_k)$ across locations k identifies β_Q. Further, variation in $r_k^{dw} \cdot (1 - \text{Info}_k)$, its square, and $\left(r_k^{dw} \cdot \text{Info}_k\right)^2$, which are not collinear,

[13] This expression of expected utility follows from the normality of the independent terms $\log Q_k^u$ and ε^o, and from the property $\mathbb{E}\left[\exp\left(\mathcal{N}(\mu, \sigma^2)\right)\right] = \exp(\mu + \sigma^2/2)$.

identifies $\beta_Q \cdot \mu_u$, $(\beta_Q \cdot \sigma_u)^2$, and $(\beta_Q \cdot \sigma_o)^2$, respectively. Consequently, the belief parameters μ_u, σ_u^2, and σ_o^2 are identified.

Equation (7) includes variables that are not directly observed in the data. First, I decompose $M_{i,k}$ into a county-level mean $M_{j(k)}$, defined as the average household disposable income over the three sample years in county j where property k is located, and an individual-specific error term $\varepsilon_{i,k}^M$:

$$M_{i,k} = M_{j(k)} + \varepsilon_{i,k}^M.$$

Second, I assume that the unit price of housing services, p_k, is constant within a local area. Specifically, all properties located in the same county j share a common unit price, such that $p_k = p_{j(k)}$. The procedure for estimating $p_{j(k)}$ is described in Appendix Section E.1. Assume the fixed effect ξ_k in equation (7) is determined at the county level. That is, $\xi_k = \xi_{j(k)}$. Substituting $M_{i,k}$ and p_k, the indirect utility can be rewritten as follows. For simplicity, j denotes $j(k)$:

$$\begin{aligned} V_{i,k} = & \exp\left[\alpha_j - \beta_Q \cdot r_k^{dw} \cdot (\text{Info}_k \cdot \log Q_k^o - \log W_k) - \beta_d \cdot r_k^{dw} \cdot D_k\right] \\ & \times \exp\left[-\beta_Q \cdot \mu_u \cdot r_k^{dw} \cdot (1 - \text{Info}_k) + \left(\beta_Q \cdot r_k^{dw}\right)^2 \cdot \left(\sigma_o^2 \frac{\text{Info}_k^2}{2} + \sigma_u^2 \frac{(1-\text{Info}_k)^2}{2}\right)\right] \\ & \times \exp\left(v_{i,k}\right). \end{aligned} \tag{8}$$

where the county-level fixed effect α_j and individual error term $v_{i,k}$ are given by:

$$\begin{aligned} \alpha_j &= \xi_j + \beta_M \cdot \log M_j - \beta_H \cdot \log p_j, \\ v_{i,k} &= \eta_{i,k} + \beta_M \varepsilon_{i,k}^M. \end{aligned} \tag{9}$$

Exogeneity of Information Quality One concern is that residents with higher income potential, captured by the idiosyncratic income shock $\varepsilon_{i,k}^M$, prefer to locate in areas where their earnings can be realized, typically in more developed or better served neighborhoods. These households may pay higher taxes and demand better local public services, including higher-quality disclosures. Since $\varepsilon_{i,k}^M$ is part of the composite error term in the indirect utility function, this sorting behavior can

induce correlation between Info_k and $\nu_{i,k}$, resulting in endogeneity bias in the model estimation. Although county fixed effects absorb broad socioeconomic differences, correlation may still arise from micro-level variation around each plant.

To assess this possibility, I test whether average information quality at each plant is systematically related to predetermined neighborhood characteristics. Using 2019 POI data, I construct three sets of measures: (i) counts of restaurants, banks, and hotels within five kilometers as proxies for local economic activity and income levels; (ii) distances to the nearest school, hospital, and park to capture access to public services; and (iii) population density to reflect the scope for monitoring pressure. I then regress each plant's mean information quality, measured as the overall share of valid records in the data, on these local attributes.

Appendix Table A2 reports the regression results. Column (1) presents estimates based on amenity measures in all wind directions, and column (2) restricts the measures to the downwind region.[14] In both specifications, the coefficients are small in magnitude and statistically insignificant. These results indicate that information quality does not systematically increase in neighborhoods with stronger economic conditions or better amenities. This pattern supports the view that information quality is not driven by unobserved micro level advantages and is plausibly exogenous.

D. Willingness-to-Pay for Information Quality

The MWTP for information quality is the income a household would give up for a marginal improvement in information quality, holding indirect utility constant. That is,

$$\text{MWTP}_{i,k} = \frac{\frac{\partial \log V_{i,k}}{\partial \text{Info}_k}}{\frac{\partial \log V_{i,k}}{\partial M_{i,j(k)}}}.$$

[14]The downwind sector is defined as a wedge that extends from each plant along its dominant wind direction and covers plus or minus one sector within a radius of five kilometers. The division of wind sectors is detailed in Appendix Section D.

It follows from the expression of indirect utility in equation (7):

$$\text{MWTP}_{i,k} = r_k^{dw} \frac{\beta_Q \cdot \left(\mu_u - \log Q_k^o\right) - r_k^{dw} \cdot \beta_Q^2 \cdot \sigma_u^2 + r_k^{dw} \cdot \beta_Q^2 \cdot \left(\sigma_o^2 + \sigma_u^2\right) \cdot \text{Info}_k}{\beta_M} \cdot M_{i,j(k)}. \quad (10)$$

Equation (10) indicates that the MWTP for information quality increases linearly with Info_k. It vanishes as r_k^{dw} approaches zero, as households located always upwind place no value on pollution disclosure.

MWTP under a General Belief Framework Equation (10) links $\text{MWTP}_{i,k}$ directly to households' beliefs about pollution. $\text{MWTP}_{i,k}$ increases with μ_u and σ_o^2, but decreases with σ_u^2. Different parameter combinations yield heterogeneous responses to information disclosure.

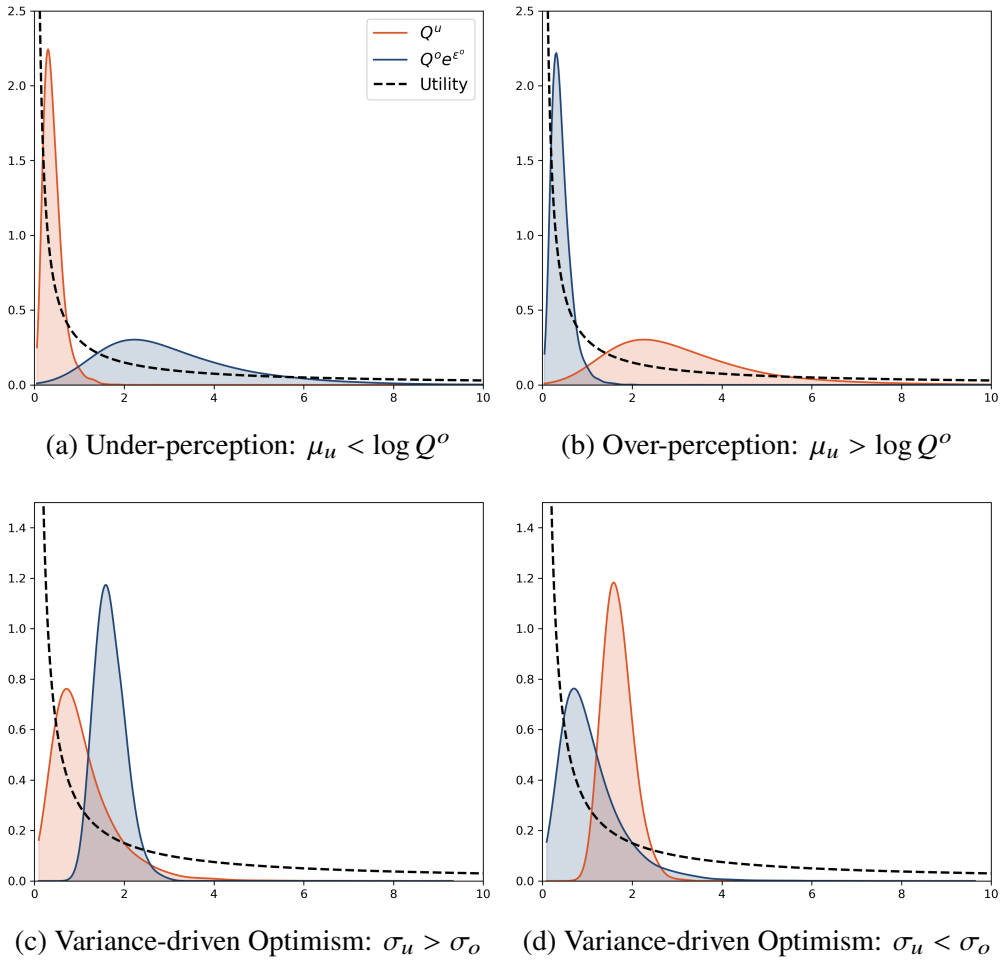
Figure 3 illustrates how these parameters determine the probability distribution function (pdf) of perceived pollution and its implications for expected utility. Consistent with equation (3), the dark gray curve shows the pdf when households are informed by valid records ($Q^o \exp(\varepsilon^o)$), while the light gray curve shows the pdf when they are uninformed (Q^u). The black dashed curve illustrates the Cobb–Douglas utility from equation (4) as a function of perceived pollution, holding other variables constant. Utility declines as perceived pollution increases.[15]

Two mechanisms can be distinguished. The first is *mean-driven*. The relative position of μ_u and $\log Q^o$ determines whether uninformed households under- or over-perceive pollution relative to when they are informed. If $\mu_u < \log Q^o$, uninformed households perceive lower pollution. Disclosure shifts beliefs upward, raises perceived pollution, and reduces indirect utility (Figure 3a). If $\mu_u > \log Q^o$, uninformed households perceive higher pollution. Disclosure shifts beliefs downward, lowers perceived pollution, and increases indirect utility (Figure 3b).

The second mechanism is *variance-driven*. It depends on the relative dispersion of σ_u and σ_o. When $\sigma_u > \sigma_o$, the uninformed distribution is more concentrated around low pollution outcomes. Remaining uninformed can then yield higher indirect utility. This reflects variance-driven optimism: greater unobserved variance raises the likelihood of perceiving pollution as lower

[15] Utility depends on Q through $Q^{-\beta_Q}$ with $\beta_Q > 0$. It rises as Q approaches zero and declines as Q increases.

than it is. It resembles "wishful thinking" in behavioral economics, where uncertainty permits optimistic interpretations of adverse situations (Engelmann et al., 2024) (Figure 3c). Conversely, when $\sigma_o > \sigma_u$, access to valid information can raise indirect utility, not by improving accuracy but by widening the distribution and increasing the chance of relatively low perceived pollution (Figure 3d).



(a) Under-perception: $\mu_u < \log Q^o$ (b) Over-perception: $\mu_u > \log Q^o$

(c) Variance-driven Optimism: $\sigma_u > \sigma_o$ (d) Variance-driven Optimism: $\sigma_u < \sigma_o$

Figure 3. Belief Parameters and Expected Utility

Note: This graph illustrates four scenarios determined by the magnitudes of the belief parameters. The dark gray and light gray shaded areas represent the pdf of perceived pollution when households are informed by valid data ($Q^o \exp(\varepsilon^o)$) and when they are uninformed (Q^u), respectively. The black line depicts the inverse relationship between the utility in equation (4) and perceived pollution. Multiplying this curve by the density of $Q^o \exp(\varepsilon^o)$ or Q^u and integrating yields the mean and variance effects discussed in this section.

Taken together, these mean- and variance-based effects imply that households may differ in their

preferences for information disclosure. The magnitude of MWTP reflects these differences. When MWTP is positive, households prefer to be better informed. This can arise under over-perception or when the informed distribution is more dispersed than the uninformed one. The latter, however, does not mean that disclosure is preferred because it helps households assess pollution more precisely. Instead, it increases their uncertainty, which runs counter to the policy objective of information disclosure.

By contrast, when MWTP is close to zero or negative, disclosure provides little or no benefit. This outcome can arise when strong under-perception and high unobserved variance dominate, particularly under low information quality. In such cases, households may prefer to avoid distressing pollution signals, or uncertainty itself may sustain more optimistic beliefs about emissions. In the extreme case of negative MWTP, households would accept disclosure only if compensated. This potentially reflects recent evidence that pollution information imposes psychological costs when it reveals conditions worse than previously believed (Wang and Cao, 2024; Xie, Yuan and Zhang, 2023).

V. Estimation Results

This section discusses the estimation of parameters related to beliefs, information, and residential location choices. A full exposition of the estimation procedure is provided in Appendix Section E.

A. Divergence in Emissions Beliefs

To capture cross-city heterogeneity, the belief parameters (μ_u, σ_u, and σ_o) are estimated for each city using maximum likelihood estimation (MLE) in the discrete choice framework. As discussed in Section IV.B, comparing the belief distributions under valid and flawed records reveals the extent to which households trust that flaws are purely technical.

First, I compare the means of the two distributions by calculating $\Delta_{\mu_u} = \mu_u - \log Q^o$. Δ_{μ_u} is estimated using

$$\widehat{\Delta}_{\mu_u} = \widehat{\mu}_u - \overline{\log Q^o},$$

where $\overline{\log Q^o}$ denotes the city mean of valid pollution index.

Figure 4 presents the distribution of $\widehat{\Delta}_{\mu_u}$ across cities. It is positive in 67 of 92 cities, indicating that 72.83% of the sample systematically perceives higher emissions when records are flawed than when records are valid. The sample mean of $\widehat{\Delta}_{\mu_u}$ is 0.94, highly significant with a standard deviation of 0.01. This indicates that the distribution of Δ_{μ_u} assigns greater weight to positive values than to negative ones.

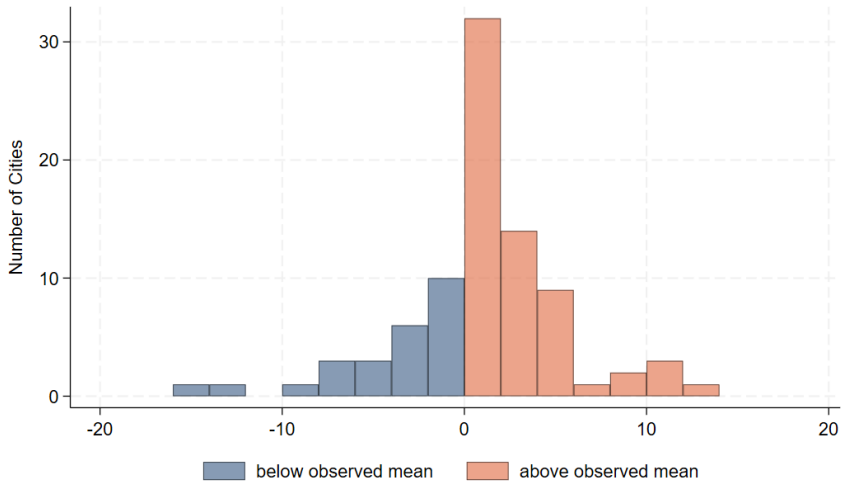


Figure 4. City-level Distribution of $\widehat{\Delta}_{\mu_u}$

Note: This graph displays the histogram of city-level $\widehat{\Delta}_{\mu_u}$. A positive $\widehat{\Delta}_{\mu_u}$ indicates that households perceive higher pollution under flawed records than under valid records. Dark gray bars denote positive values, while light gray bars denote negative values. In total, 67 cities show a positive $\widehat{\Delta}_{\mu_u}$ and 25 cities show a negative value.

To account for potential sampling error in positive point estimates, I classify a city as an *over-perception city* if $\widehat{\Delta}_{\mu_u}$ is significantly greater than zero at the 5% level in a one-sided t-test.[16] Appendix Figure A5 reports the distribution of the city-level t-statistics. Based on this criterion, 43 of 92 cities are classified as over-perception cities. These cities are then compared with the remaining ones in the next section to investigate potential sources of belief bias.

[16]The standard deviation of $\widehat{\mu}_u$ is employed to approximate that of $\widehat{\Delta}_{\mu_u}$. This approximation is justified because Q is defined relative to emission limits and exhibits little variation, as plants operate steadily in most periods. Consequently, $\widehat{\mu}_u$ constitutes the dominant source of variance.

The remaining two belief parameters, σ_o and σ_u, measure the dispersion of beliefs under valid and flawed records. Figure 5 plots the distribution of the city-level estimates. The distribution of $\widehat{\sigma}_o$ (light gray line) is concentrated near small values, suggesting that emission perceptions cluster closely around the official disclosure under valid records. In contrast, the distribution of $\widehat{\sigma}_u$ (dark gray line) is wider and has a larger mean, indicating greater uncertainty under flawed records. However, since both distributions deviate considerably from zero, this indicates substantial variation across households in their interpretation and reliance on official data, even under valid records.

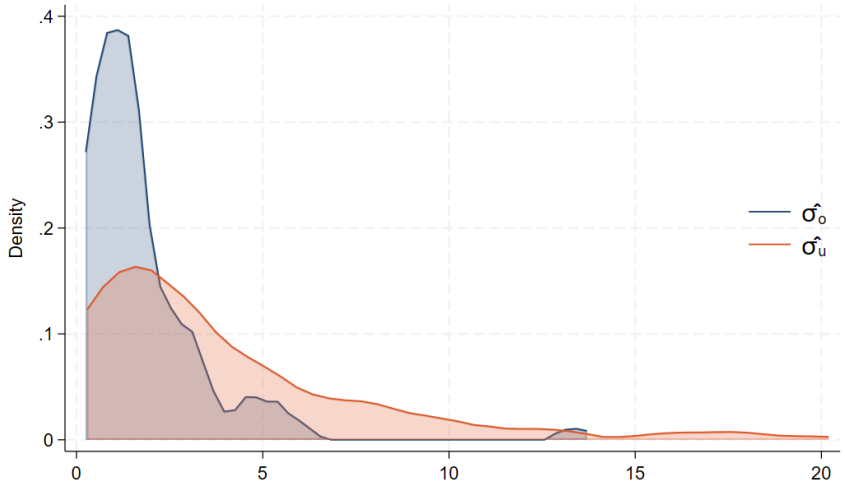


Figure 5. City-level Distributions of $\widehat{\sigma}_o$ and $\widehat{\sigma}_u$

Note: This graph shows the distributions of city-level $\widehat{\sigma}_o$ (light gray line) and $\widehat{\sigma}_u$ (dark gray line) across 92 cities. A larger value indicates greater household uncertainty in emission perceptions.

Overall, compared with beliefs under valid records, flawed records systematically lead households to perceive higher levels of emissions, with greater uncertainty. These findings suggest the critical role of complete and consistent emission data in shaping household assessments of environmental risks. Enhancing information quality, as formalized in equation (3), reduces exposure to flawed records and thereby diminishes exaggerated perceptions of pollution.

B. Sources of Pollution Over-perception

This section explores potential explanations for *over-perception*. First, I investigate whether plants in over-perception cities use flawed records to hide excessive emissions. Second, I assess whether over-perception reflects city-specific factors that lead households to fundamentally distrust flawed records. Third, I examine whether the institutional quality of information within the disclosure system correlates with the level of information trust.

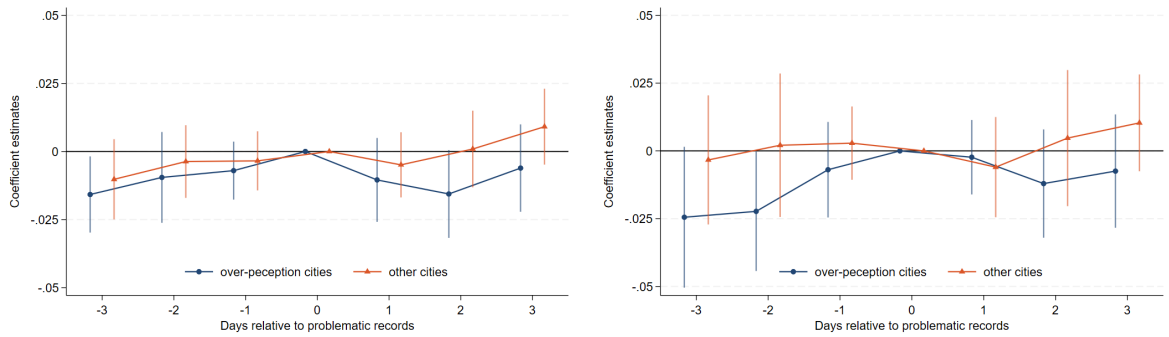
Do Plants Use Flawed Records to Conceal Excessive Emissions? A plausible source of over-perception is public skepticism that plants strategically manipulate emission records. To evade detection, plants may produce flawed records by shutting down monitoring devices or by attributing exceedances to technical issues. When data gaps align with visible signs of heavy pollution, such as thick and dirty emissions from chimneys, the public may infer concealment. These inferences can generate systematic over-perception. If such perceptions reflect actual manipulation, plants in over-perception cities should exhibit a higher incidence of flawed records during periods of excessive emissions.

To investigate this possibility, I follow the approach of Zou (2021) and compare pollution levels on days before and after those with flawed records. If plants use flawed data to conceal excessive emissions, actual pollution on those days should be higher than on surrounding days.

To obtain complete daily coverage of emissions independent of plant disclosure, I use two proxies based on PM2.5. As a key byproduct of waste incineration (Kicińska, Caba and Barria-Parra, 2024), PM2.5 is widely recognized as an indicator of air quality and thus provides a suitable proxy. As described in Section III, the first proxy is constructed by interpolating from national monitoring stations. The second is drawn from He et al. (2023), who predict PM2.5 using a machine learning model that combines station-monitored and satellite data. I then estimate the following equation:

$$\text{PM2.5}_{wt} = \sum_{\tau \in [-3,3], \tau \neq 0} \beta_\tau \cdot \mathbb{I}(t = \tau) + X_{wt}\beta + \delta_{wy} + \rho_m + \varepsilon_{wt},$$

where the dependent variable is the log of PM2.5 at plant w on day t. The indicator $\mathbb{I}(t = \tau)$ equals one for day τ relative to the date with flawed records. The set $\tau \in [-3, 3]$ compares pollution levels within three days before and after that date. X_{wt} is a vector of weather controls, including temperature, precipitation, wind direction, and wind speed. I include plant-by-year fixed effects, δ_{wy}, to capture annual variations in plant-location characteristics, and calendar-month dummies, ρ_m, to control for seasonality. Standard errors are clustered at the city level to allow for correlation in shocks across plants within the same city.



(a) ML-predicted PM2.5 levels

(b) Station-based PM2.5 levels

Figure 6. Estimates of Pollution Around Flawed-record Dates

Note: This graph plots coefficient estimates for two proxies based on PM2.5. It compares over-perception cities (dark gray lines with circles) with other cities (light gray lines with triangles) across days relative to the date of flawed records. The estimates include 95 percent confidence intervals based on standard errors clustered at the city level.

The results are plotted in Figure 6. The dark gray lines represents the coefficients for over-perception cities, while the light gray line shows those for other cities. The patterns suggest that pollution levels around the flawed-records date are lower for plants in over-perception cities. However, since the coefficients are not consistently significant, the evidence for manipulation is limited. This implies that public skepticism toward flawed records can be unsubstantiated.[17]

[17] One concern is that PM2.5 measurements can be affected by nearby pollution sources. To address this, I exclude plants located within five kilometers of other high-polluting industrial sites (e.g., chemical plants, paper mills, smelters, coal power plants) and re-estimate the model. As shown in Appendix Figure A6, the patterns are robust, with some coefficients remaining statistically insignificant.

Fundamental Distrust and City-specific Characteristics The limited evidence of manipulation suggests that over-perception may arise from fundamental distrust, where the public attributes flawed records to concealment even in the absence of manipulation. To test whether certain cities exhibit greater distrust, I conduct t-tests comparing potential factors between over-perception and other cities. The factors include industry scale, regulatory environment, public attention, political leadership, and socioeconomic conditions. Figure 7 presents the estimated differences.

Industry Scale and Regulatory Environment. Over-perception cities host larger waste incineration and power generation capacities and exhibit longer operational histories. Larger plants handle greater volumes of waste, and older facilities more often depend on outdated technologies, both of which may intensify public suspicion of elevated violation risks. Additionally, plants in these cities exhibit modestly higher violation rates yet face fewer penalties and lower fines, which may be perceived by the public as a signal of weak enforcement.

Public Attention. I collect Weibo data as described in Section III. Plants in over-perception cities attract more pre-operation complaints and greater public engagement (posts, likes, shares, and comments), although the latter difference is not statistically significant. These early complaints reflect concerns about potential risks, whereas post-operation complaints related to actual performance do not differ across cities. This pattern suggests that residents who are most vocal before construction are also more sensitive to flaws in official records.

Political Leadership. Over-perception is less likely when top leaders, such as mayors or party secretaries, are native to the city. This is consistent with the view that hometown ties strengthen incentives to safeguard local environmental quality (Ma, Zou and Qin, 2024). Such ties may reassure households that local governments hold stronger commitments to credible environmental enforcement.

Socioeconomic Conditions. Over-perception cities are larger, wealthier, and more industrialized, with higher SO_2 emissions and larger shares of vulnerable populations such as children and the elderly. Residents in these cities may exhibit greater environmental awareness and stronger expectations for pollution control.

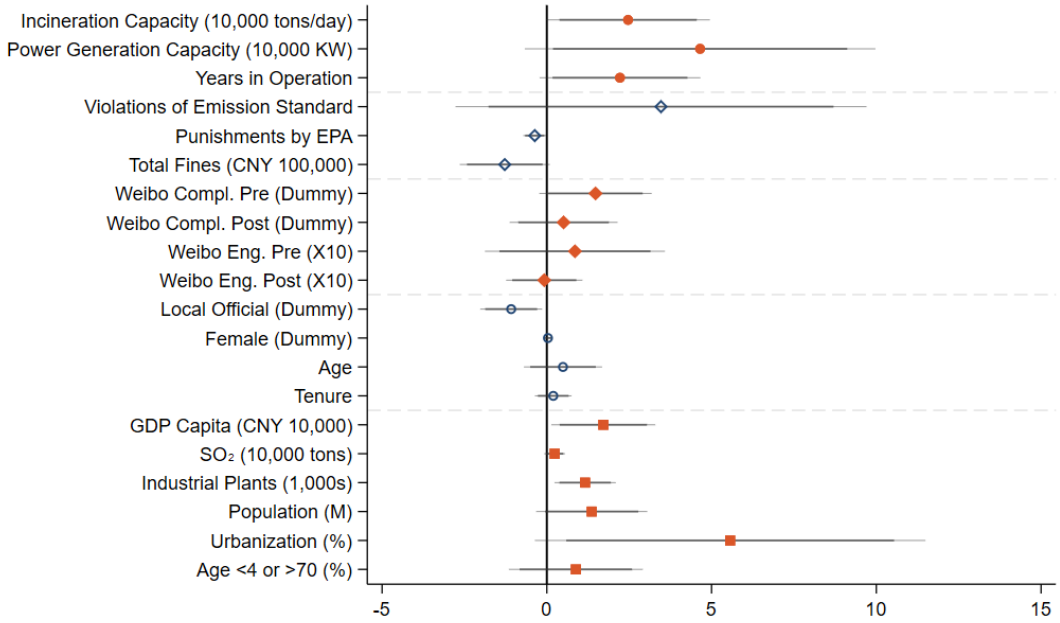


Figure 7. Comparison of City Characteristics: Over-Perception vs. Other Cities

Note: This figure shows the average differences in characteristics between over-perception cities and other cities across five domains: industry scale of waste incineration (circle), regulatory environment (hollow diamond), public attention (diamond), political leadership (hollow circle), and socioeconomic factors (square). Positive coefficients indicate that the characteristic is higher in over-perception cities. "Local Official" equals 1 if either the mayor or party secretary was born in the same city. "Female" equals 1 if either leader is female. Age and Tenure are averages of the mayor and party secretary. Estimates for all dummy variables are rescaled by multiplying by 10 for clarity. Any additional rescaling methods are specified in parentheses for each variable, where applicable. "Urban Population" is the urban-to-total population ratio. "Age Below 4 or Above 70" is the share of individuals under 4 or over 70, multiplied by 100. Range bars display 90% (dark grey) and 95% (light grey) confidence intervals based on the t-test.

Institutional Information Quality and Information Trust Information trust can also depend on the consistency of disclosure practices by WtE plants. For example, a sustained record of valid reporting leads households to interpret occasional flaws as minor technical issues rather than deliberate manipulation. By contrast, frequent missing or faulty records erode trust, as they signal weaknesses in the information system and raise doubts about the credibility of disclosure.

To capture the capacity and integrity of local institutions to produce emission records, I construct a measure of institutional information quality, denoted $\overline{\text{Info}}$. It is defined as the sample mean of Info in each city over the period 2020 to 2022. Averaging over multiple years reduces the influence

of short-term shocks and therefore better reflects the underlying institutional environment. As discussed in Section IV.B, information trust is reflected in the discrepancy between the distribution of official disclosures and household beliefs under valid or flawed records. Accordingly, I define

$$\widehat{\Delta}_{\mu_u} = \widehat{\mu}_u - \overline{\log Q^o}, \quad \widehat{\Delta}_{\sigma_u} = \log \widehat{\sigma}_u - \log \widehat{\mathrm{Var}}^{1/2} (\log Q^o), \quad \widehat{\Delta}_{\sigma_o} = \log \widehat{\sigma}_o - \log \widehat{\mathrm{Var}}^{1/2} (\log Q^o).$$

where the sample mean $\overline{\log Q^o}$ and standard deviation $\widehat{\mathrm{Var}}^{1/2} (\log Q^o)$ are computed for each cities, as $\widehat{\mu}_u$, $\widehat{\sigma}_u$ and $\widehat{\sigma}_o$.

I run three separate regressions of the spread measures on the complement of institutional information quality, $1 - \overline{\mathrm{Info}}$:

$$\widehat{\Delta}_y = \gamma_0 + \gamma_1 \cdot \left(1 - \overline{\mathrm{Info}}\right) + \varepsilon^y,$$

where $y \in \{\mu_u, \sigma_u, \sigma_o\}$. Since the dependent variables are obtained from estimation, inverse-covariance weights are applied to account for the additional estimation error.[18]

The intercept γ_0 and slope γ_1 capture different dimensions of institutional information quality. The intercept γ_0 reflects systematic biases that persist even under perfect disclosure. In the mean equation, a positive γ_0 suggests that households suspect underreporting, while a negative value indicates persistent underestimation of emissions. In the variance equation, a positive γ_0 implies that households systematically perceive greater uncertainty than official records convey, whereas a negative value suggests reliance on alternative sources regarded as more credible. The slope γ_1 captures responsiveness to institutional information quality. A positive γ_1 indicates that improved disclosure narrows belief gaps, while a negative value suggests that households interpret consistently valid records as manipulated or too perfect to be genuine.

Table 2 presents the regression results. Column (1) shows a positive and significant slope γ_1 with an insignificant intercept, indicating that higher institutional information quality reduces the upward

[18] As the variances of $\widehat{\mu}_u$, $\log \widehat{\sigma}_u$, and $\log \widehat{\sigma}_o$ dominate, their inverses are used as weights. The asymptotic variances of $\log \widehat{\sigma}_u$ and $\log \widehat{\sigma}_o$ are derived by the Delta Method.

bias in household beliefs about the mean and strengthens information trust. Column (2) shows that both the intercept and slope are large but statistically insignificant, indicating limited influence of institutional information quality on $\widehat{\Delta}_{\sigma_u}$. Column (3) shows a significantly positive intercept and slope, indicating that household perceptions are more dispersed than official disclosures, but improved institutional information quality enhances the alignment between perceptions under valid records and official data.

Table 2. Belief Parameters and Institutional Information Quality

	(1)	(2)	(3)
	$\widehat{\Delta}_{\mu_u}$	$\widehat{\Delta}_{\sigma_u}$	$\widehat{\Delta}_{\sigma_o}$
γ_0	0.008	2.164	2.961***
	(0.474)	(2.342)	(0.180)
γ_1	8.678*	12.155	3.691**
	(4.431)	(17.261)	(1.682)
R-squared	0.51	0.11	0.01
Observations	92	92	92

Note: This table reports inverse-variance weighted regressions of belief gaps on the complement of institutional information quality, $1 - \overline{\text{Info}}$. The dependent variables are the estimated spreads in belief distributions: $\widehat{\Delta}_{\mu_u}$, $\widehat{\Delta}_{\sigma_u}$, and $\widehat{\Delta}_{\sigma_o}$. Each regression is weighted by the variance of the corresponding dependent variable. Standard errors are in parentheses.
* p < 0.10, ** p < 0.05, *** p < 0.01.

C. The Value of Information Quality

This section assesses the value of information quality. I begin by estimating households' willingness to pay for improvements in information quality. I then analyze the extent to which enhanced information quality mitigates the disutility associated with proximity to the plant.

Willingness-to-Pay for Information Quality To obtain the parameters for calculating MWTP in equation (10), I estimate β_Q, β_d, and county fixed effects for each city using the same MLE procedure as for the belief parameters. The city-level estimates are then aggregated using inverse-variance weighting. The remaining parameters, β_H and β_M, are identified by regressing the county

fixed effects estimated from the MLE on county-level average household income and housing service prices, controlling for other county attributes and year fixed effects. Further details on the estimation process are provided in Appendix Section E.2.

Table 3 reports the results of the aggregated estimates. The significant β_d confirms that distance is a key determinant of utility, consistent with real-world evidence where public complaints emphasize maintaining a "safe distance," and property prices decline with proximity to the plant (Nie et al., 2025).

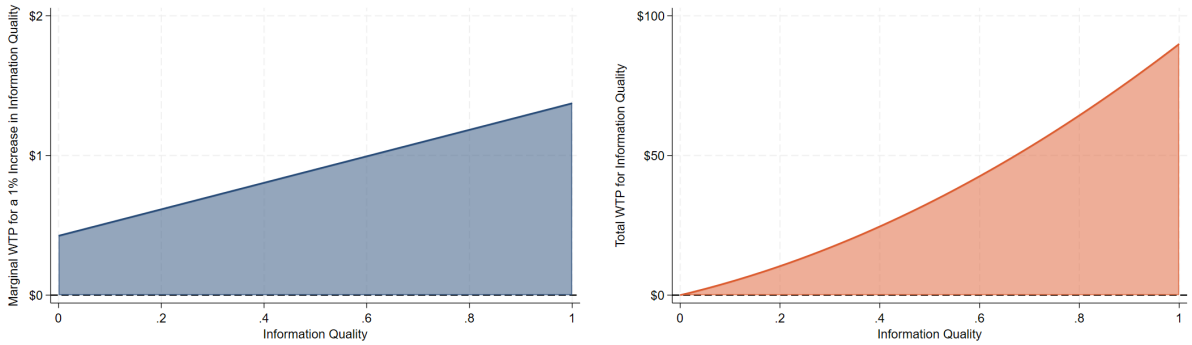
Table 3. Estimates of MWTP and Preference Parameters

Parameters	Estimation	Std. Err.
β_Q	0.123***	0.042
β_{I1}	0.335**	0.183
β_{I2}	0.492**	0.223
β_{I3}	0.080***	0.023
β_d	9.491***	2.14e-07
β_H	0.690***	0.073
β_M	4.141 ***	0.116

Note: The first block reports MWTP parameters estimated by MLE at the city level and then aggregated. The second block reports parameters estimated by regressing county-level fixed effects obtained from the MLE on county-average household income and housing service prices, controlling for other county attributes and year fixed effects.

* p < 0.10, ** p < 0.05, *** p < 0.01.

The parameter estimates are applied to equation (10) to calculate the MWTP for improvements in information quality. The calculation relies on three sample values: the average share of properties located downwind from a plant ($r^{dw} = 0.51$), the median annual household income ($M = \$18{,}752$), and the average observed pollution level ($\log Q^o = -0.39$). Using these inputs, equation (10) implies that MWTP is a linear function of Info.



(a) MWTP for 1% Increase in Info

(b) TWTP

Figure 8. Willingness-to-Pay for Information Quality

Note: Figure 8a plots the MWTP for a 1% increase in Info at different initial levels. Figure 8b displays the TWTP, calculated as the cumulative sum of MWTP at each level of Info.

Figure 8a reports the MWTP for a 1% improvement in information quality at different initial levels. MWTP increases from slightly above \$0.43 to \$1.37 as Info rises from zero to one. At the current average quality of 0.85, the MWTP for a 1% improvement is \$1.23. Figure 8b displays the total willingness to pay (TWTP), defined as the cumulative sum of MWTP across levels of information quality. As shown, household TWTP for raising information quality from 0.85 to 1 is \$19.54, while a decline to 0.5 would impose a loss of \$37.29.

Figure 8 shows that individuals place limited value on improvements when current disclosure provides a weak signal of pollution levels. The marginal benefit of further improvements increases with baseline quality, indicating complementarity between existing quality and additional improvements rather than diminishing returns. However, even at the lowest quality, MWTP remains positive. This reveals that the over-perception bias dominates other belief factors. As discussed in Section IV.D, improvements in information quality enhance household welfare by down-weighting flawed records and mitigating the effects of upward mean-driven bias.

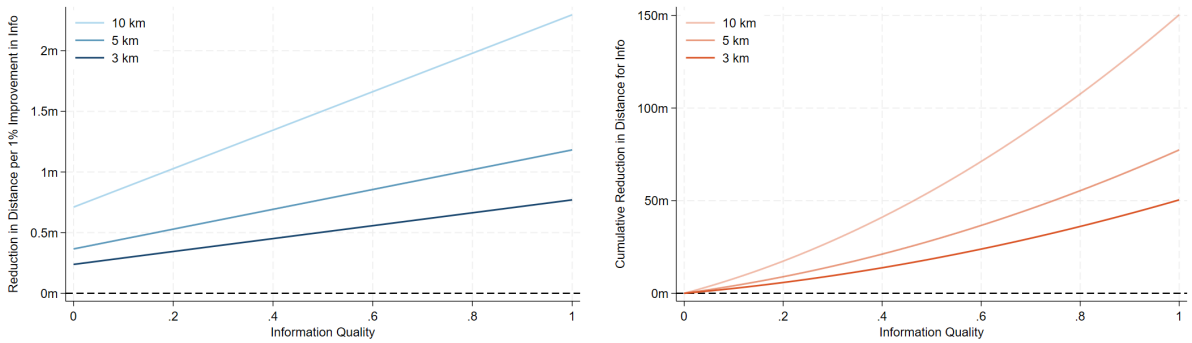
Substitution Effect Between Information Quality and Distance I then examine whether higher information quality increases the acceptability of residences near the plant by quantifying the extent to which it offsets the disutility of proximity to a WtE plant. Specifically, I derive the marginal rate

of substitution, $\text{MRS}_{\text{d,Info}}$, from equation (7):

$$\left|\text{MRS}_{\text{d,Info}}\right| = \frac{\frac{\partial V}{\partial \text{Info}}}{\frac{\partial V}{\partial d}} = \frac{\beta_Q \cdot (\log Q^o - \mu_u) + r^{dw} \cdot \beta_Q^2 \cdot \left[\sigma_u^2 - \left(\sigma_o^2 + \sigma_u^2\right) \cdot \text{Info}\right]}{\beta_d \cdot D^{(1)}}, \text{ where}$$

$$D^{(1)} = -\left((2 - 2\alpha_1)\, d^{1-2\alpha_1} \frac{(\sin\theta)^2}{2\gamma_1^2 (\cos\theta)^{2\alpha_1}} - 2\alpha_2 d^{-2\alpha_2 - 1} \frac{H^2}{2\gamma_2^2 (\cos\theta)^{2\alpha_2}} + \frac{\alpha_1 + \alpha_2}{d} \right). \tag{11}$$

Figure 9 illustrates the interpretation of equation (11), with parameters substituted using the same values as in the WTP analysis. Distance is fixed to isolate the effect of information quality. For simplicity, I consider properties located exactly downwind of a WtE plant, where $\theta = 0$.



(a) MRS for 1% Increase in Info

(b) Cumulative Substitution Effects

Figure 9. Substitution Effects Between Information Quality and Distance

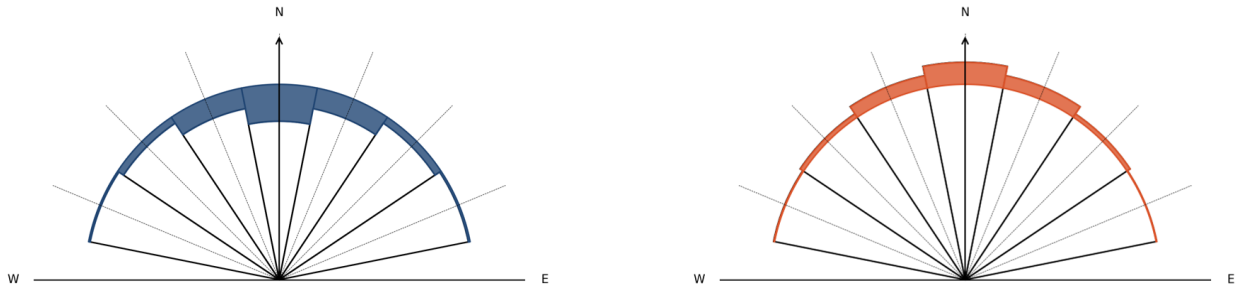
Note: Figure 9a reports the MRS for a 1% improvement in information quality with respect to distance. Figure 9b presents the cumulative substitution effect, defined as the total reduction in distance at each level of information quality. Both figures report results for houses located exactly downwind at distances of three, five, and ten kilometers from a WtE plant.

Figure 9a presents the MRS at different initial levels of information quality. At the current average quality of 0.85, the MRS equals 0.69, 1.06, and 2.06 meters for properties located three, five, and ten kilometers downwind, respectively. Figure 9b reports the cumulative distance reduction at a given information quality. Increasing quality from 0.85 to 1 leads to additional reductions of 10.95, 16.82, and 32.66 meters, respectively, while decreasing to 0.5 leads to increases in distance of 20.90, 32.10, and 62.33 meters, respectively.

Although the reduction in distance is modest at the individual level, the aggregate effect across

the surrounding area can be substantial. Figure 10 illustrates directional substitution effects under changes in information quality. Properties are assumed to be initially located five kilometers from the plant, with wind consistently blowing toward true north. The surrounding area is divided into directional sectors. The angle between each sector's central line and the wind direction, θ, enters equation (11) to determine the MRS. As θ increases, MRS decreases, producing smaller shifts in sectors farther from the downwind direction. The shaded areas aggregate these directional shifts and represent the implied regional change in surface area.

Figure 10a shows the inward shift in distance across directional sectors when information quality improves from the current level of 0.85 to 1, with a total inward adjustment of approximately 86,648 m^2. Figure 10b presents the corresponding outward shift when information quality declines to 0.5, with a total outward adjustment of approximately 165,902 m^2. This outward shift reflects re-sorting in residential demand, which signals a loss in land-use efficiency and housing surplus.



(a) Inward Shift Under Improved Information

(b) Outward Shift Under Degraded Information

Figure 10. Substitution Effects in Entire Areas

Note: This figure compares substitution effects under changes in information quality. Figure 10a illustrates the inward shift in distance across directional sectors when information quality increases from the current level of 0.85 to 1, assuming an initial radius of five kilometers and a constant wind blowing toward true north. Figure 10b presents the corresponding outward shift when information quality declines to 0.5. Shaded areas aggregate sector-level adjustments to quantify the implied change in surface area.

VI. Conclusion

This paper examines how information quality shapes public beliefs about pollution and trust in official disclosures. I focus on a high-frequency disclosure platform for Waste-to-Energy plants

in China and estimate household beliefs of emissions using a residential sorting model. In most cities, flawed records increase perceived pollution and uncertainty, even with technical explanations provided. My analysis uncovers limited evidence of deliberate manipulation by plants; instead, distrust appears rooted in city-specific factors such as regulatory stringency, political leadership, socioeconomic conditions, and public attention. Information trust declines with lower institutional quality, even under valid records. Finally, my estimates highlight the sizable economic cost of degrading information quality.

This study provides empirical evidence that households do not take information disclosures at face value. When they do not sufficiently rely on disclosed information, disclosure becomes inefficient and risks wasting resources. Evaluating the value of information without accounting for trust can therefore lead to biased conclusions. Poor-quality records can further weaken the informativeness of disclosures and erode trust, potentially limiting or even reversing the intended policy targets. Although this study focuses on emission disclosure, the argument applies more broadly, as public communications often rely on official disclosures. When authorities suspend or selectively disclose information, credibility declines and political trust diminishes, undermining the effectiveness of transparency as a governance tool.

References

Agarwal, Sumit, Yanhao Ding, Weida Kuang, and Xiao Zhu. 2023. "Are environmental punishments good news or bad news? Evidence from China." *Journal of Environmental Economics and Management*, 120: 102847.

Barwick, Panle Jia, Shanjun Li, Liguo Lin, and Eric Yongchen Zou. 2024. "From fog to smog: The value of pollution information." *American Economic Review*, 114(5): 1338–1381.

Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins. 2009. "Migration and hedonic valuation: The case of air quality." *Journal of Environmental Economics and Management*, 58(1): 1–14.

Boré, Abdoulaye, Jicui Cui, Zhuoshi Huang, Qing Huang, Johann Fellner, and Wenchao Ma. 2022. "Monitored air pollutants from waste-to-energy facilities in China: Human health risk, and buffer distance assessment." *Atmospheric Pollution Research*, 13(7): 101484.

Bresnahan, Brian W, Mark Dickie, and Shelby Gerking. 1997. "Averting behavior and urban air pollution." *Land Economics*, 340–357.

Buntaine, Mark T, Michael Greenstone, Guojun He, Mengdi Liu, Shaoda Wang, and Bing Zhang. 2024. "Does the squeaky wheel get more grease? The direct and indirect effects of citizen participation on environmental governance in China." *American Economic Review*, 114(3): 815–850.

Chang, Tom Y, Wei Huang, and Yongxiang Wang. 2016. "Something in the air: Projection bias and the demand for health insurance." *Review of Economic Studies*.

Chen, Shuai, Paulina Oliva, and Peng Zhang. 2022. "The effect of air pollution on migration: Evidence from China." *Journal of Development Economics*, 156: 102833.

Chen, Zuoqi, Bailang Yu, Chengshu Yang, Yuyu Zhou, Shenjun Yao, Xingjian Qian, Congxiao Wang, Bin Wu, and Jianping Wu. 2020. "An extended time-series (2000-2023) of global NPP-VIIRS-like nighttime light data." https://doi.org/10.7910/DVN/YGIVCD Accessed: 2024-08-30.

Engelmann, Jan B, Maël Lebreton, Nahuel A Salem-Garcia, Peter Schwardmann, and Joël J van der Weele. 2024. "Anticipatory anxiety and wishful thinking." *American Economic Review*, 114(4): 926–960.

Gao, Xuwen, Ran Song, and Christopher Timmins. 2023. "Information, migration, and the value of clean air." *Journal of Development Economics*, 163: 103079.

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

Hanna, Rema, Bridget Hoffmann, Paulina Oliva, and Jake Schneider. 2021. "The power of perception limitations of information in reducing air pollution exposure." IDB Working Paper Series.

He, Qingqing, Tong Ye, Weihang Wang, Ming Luo, Yimeng Song, and Ming Zhang. 2023. "Spatiotemporally continuous estimates of daily 1-km PM2. 5 concentrations and their long-term exposure in China from 2000 to 2020." *Journal of Environmental Management*, 342: 118145.

Ito, Koichiro, and Shuang Zhang. 2020. "Willingness to pay for clean air: Evidence from air purifier markets in China." *Journal of Political Economy*, 128(5): 1627–1672.

Kicińska, Alicja, Grzegorz Caba, and Fernando Barria-Parra. 2024. "Burning of municipal waste in household furnaces and the health of their owners." *Scientific Reports*, 14(1): 32011.

Ma, Hongqi, Jingxian Zou, and Cong Qin. 2024. "Environmental regulations implemented by local officials in China: Is there a hometown effect?" *Energy Economics*, 130: 107340.

Mastromonaco, Ralph. 2015. "Do environmental right-to-know laws affect markets? Capitalization of information in the toxic release inventory." *Journal of Environmental Economics and Management*, 71: 54–70.

Ministry of Ecology and Environment of the People's Republic of China. 1992. "Technical methods for making local emission standards of air pollutants." http://english.mee.gov.cn/Resources/standards/others1/others3/200808/t20080828_127786.shtml Accessed: 2024-08-30.

Moulton, Jeremy G, Nicholas J Sanders, and Scott A Wentland. 2024. "Toxic assets: How the housing market responds to environmental information shocks." *Land Economics*, 100(1): 66–88.

Mu, Yingfei, Edward Rubin, and Eric Yongchen Zou. 2024. "What's Missing in Environmental Self-Monitoring: Evidence from Strategic Shutdowns of Pollution Monitors." *Review of Economics and Statistics*, 1–45.

Neidell, Matthew. 2009. "Information, avoidance behavior, and health: the effect of ozone on asthma hospitalizations." *Journal of Human Resources*, 44(2): 450–478.

Nie, Rong, Juliana Cunha Carneiro Pinto, Jinbo Song, and Yueming Qiu. 2025. "Real-time emissions data disclosure of Waste-to-Energy incineration plants and public risk perceptions: Evidence from the housing market." *Journal of Environmental Economics and Management*, 103207.

Pasquill, Frank. 1961. "The estimation of the dispersion of windborne material." *Meteoro. Mag.*, 90: 20–49.

Shepard, Donald. 1968. "A two-dimensional interpolation function for irregularly-spaced data." *Proceedings of the 1968 23rd ACM national conference*, 517–524.

Snoun, Hosni, Moez Krichen, and Hatem Chérif. 2023. "A comprehensive review of Gaussian atmospheric dispersion models: current usage and future perspectives." *Euro-Mediterranean Journal for Environmental Integration*, 8(1): 219–242.

Stockie, John M. 2011. "The mathematics of atmospheric dispersion modeling." *Siam Review*, 53(2): 349–372.

Taufiq, Luthfi Chaliqi, Aqlima Putri, and Fadhlullah Apriandy. 2024. "Simplified Spatial Wind Vector Interpolation Method for Airport Runway Orientation Analysis." Vol. 476, 01057, EDP Sciences.

Wang, Chen, and Juanjuan Cao. 2024. "Air pollution, health status and public awareness of environmental problems in China." *Scientific Reports*, 14(1): 19861.

Wang, Xiaohan, and Mengjun Yang. 2024. "The effect of soil pollution information disclosure on housing prices." *China Economic Review*, 83: 102112.

Wang, Xinhao, Lulin Xu, Qin Zhang, Da Zhang, and Xiliang Zhang. 2022. "Evaluating the data quality of continuous emissions monitoring systems in China." *Journal of Environmental Management*, 314: 115081.

WorldPop. 2023. "WorldPop Population Density for China." https://data.humdata.org/dataset/worldpop-population-density-for-china Accessed: 2024-08-30.

Xie, Tingting, Ye Yuan, and Hui Zhang. 2023. "Information, awareness, and mental health: Evidence from air pollution disclosure in China." *Journal of Environmental Economics and Management*, 120: 102827.

Zhang, Da, Qin Zhang, Shaozhou Qi, Jinpeng Huang, Valerie J Karplus, and Xiliang Zhang. 2019. "Integrity of firms' emissions reporting in China's early carbon markets." *Nature Climate Change*, 9(2): 164–169.

Zhang, Junjie, and Quan Mu. 2018. "Air pollution and defensive expenditures: Evidence from particulate-filtering facemasks." *Journal of Environmental Economics and Management*, 92: 517–536.

Zou, Eric Yongchen. 2021. "Unwatched pollution: The effect of intermittent monitoring on air quality." *American Economic Review*, 111(7): 2101–2126.

Appendix

This appendix is organized as follows. Section A provides examples of public complaints related to the proximity of WtE plants. Section B describes the data sources used to construct housing choices. Section C explains the imputation method used to calculate wind and $PM_{2.5}$ at specific locations. Section D details the calculation of weighted-average wind angles and speeds from daily data. Section E presents the detailed estimation procedure for the residential sorting model. Finally, Section F includes supplementary tables and figures referenced in the main text.

A Household Perceptions of Safe Distance

This section presents examples from 2019 of perceived safe distances. The cases are drawn from public complaints on the leadership message board, an online platform hosted by People's Daily that allows citizens to submit concerns and suggestions directly to government officials.

Example 1 Message left to Yunnan Provincial Party Secretary Chen Hao on September 4, 2019:

> 云南省昆明市经开区垃圾焚烧发电厂目前处于生活区附近，距离最近的村子不足500米，并且周边规划为居住区，所以建议垃圾焚烧发电厂外迁离生活区10公里左右。请书记同志及省委领导考虑考虑，谢谢。
>
> *Translation: The WtE plant in the Economic Development Zone of Kunming, Yunnan Province, is currently near residential areas, less than 500 meters from the nearest village, and the surrounding area is planned for housing. Therefore, I suggest relocating the plant about 10 kilometers away from residential areas. Please consider this request. Thank you.*

Example 2 Message left to Hainan Governor Shen Xiaoming on January 12, 2019:

老城垃圾焚烧厂在居民区不到20公里之内建，35万居民如何生活，难道长寿村就这样毁于一旦。

Translation: The WtE plant in Laocheng was built within 20 kilometers of residential areas. How can 350,000 residents live under such conditions? Will Long-live Village be destroyed in this way?

Example 3 Message left to Hainan Provincial Party Secretary Liu Cigui on January 26, 2019:

琼海市将垃圾焚烧厂建在离居民区3公里范围内，居民怨声载道，望书记关注将该厂迁往20公里以外无人区。

Translation: In Qionghai City, a WtE plant was built within 3 kilometers of residential areas. Residents are strongly opposed. We ask the Secretary to consider relocating the plant to an uninhabited area at least 20 kilometers away.

B Housing Transaction and Listing Data

The housing data used in this study come from Beike, the largest integrated platform for online and offline housing transactions in China.

Beike's data are well suited for this study. First, Beike holds a large share of the second-hand housing market, which makes its transaction data broadly representative. By 2020, it operated about 47,000 real estate offices and employed more than 493,000 agents in over 100 cities. Its gross transaction value for existing homes reached $297.3 billion, accounting for 26.2% of the national total. Second, Beike provides extensive coverage of reliable housing alternatives. It guarantees authentic listings, with a verified accuracy rate above 95%. It also creates the Housing Dictionary, the largest housing information database in China. This database integrates brokerage resources and covers most properties on the market. By 2020, it included about 240 million properties across 587,000 communities in 332 cities.[19] Each property is assigned a unique ID, which eliminates duplication and ensures data accuracy.

[19] *Source*: KE Holdings Inc. 2020 Annual Report.

Overall, the demand to supply ratio on Beike is about 0.3 across the 92 cities, showing that supply far exceeds demand. Such weak competition allows buyers to face a wide and credible set of alternatives, rather than being constrained by competing bidders. Observed housing choices therefore provide a close approximation to underlying preferences rather than reflecting market scarcity. This condition satisfies a key requirement for the validity of the discrete choice model described in Section IV.C.

One limitation of the dataset is that it excludes unsold properties that were delisted prior to 2023. These delistings may have occurred because the properties were of poor quality and difficult to sell, or because sellers withdrew them for personal reasons. In either case, such properties represent "unqualified" alternatives with limited attractiveness to buyers. Their absence therefore does not substantially compromise the completeness of the choice set. Given the large number of listings in the dataset, the marginal loss of alternatives is unlikely to materially affect the estimation results.

C Imputation Method for Wind and PM2.5

This section outlines the application of the Inverse Distance Weighting (IDW) method to interpolate wind and PM2.5 concentration levels at the locations of the plants. IDW is a widely used spatial interpolation technique that estimates values at unsampled locations using nearby measured points, with closer points exerting greater influence on the estimation (Shepard, 1968). The interpolated value at the location of plant x is formally expressed as follows:

$$Z(x) = \frac{\sum_{i=1}^{n} \frac{Z(x_i)}{d(x,x_i)^p}}{\sum_{i=1}^{n} \frac{1}{d(x,x_i)^p}}$$

where $Z(x)$ is the estimated value of plant x, $Z(x_i)$ is the known value at location x_i, and $d(x, x_i)$ is the distance between the known point x_i and the estimated point x. The parameter p controls the weight assigned to each data point based on its distance, which is set to 2, indicating that the weight is inversely proportional to the square of the distance. To capture local variations while avoiding overfitting due to noise from very close points, n is set to 3, meaning that only the three nearest

data points are considered for estimation.

Wind To interpolate wind direction and speed, I use the three nearest grid points to each plant from the ERA5-Land dataset. The data provide the u and v wind components measured ten meters above the Earth's surface, where the u component represents eastward wind speed and the v component represents northward wind speed. Each component is interpolated separately using the IDW method (Taufiq, Putri and Apriandy, 2024). By treating the u and v components independently, this approach preserves the directional integrity of the wind vector. The interpolated components, $u_{\text{interpolated}}$ and $v_{\text{interpolated}}$, are subsequently combined to determine the wind direction and speed at the plant location. Specifically, the wind speed at plant x is calculated using the Pythagorean theorem as follows:

$$W(x) = \sqrt{u^2_{\text{interpolated}} + v^2_{\text{interpolated}}}$$

The wind direction is given by:

$$\theta_w(x) = \text{mod}\left(180 + \frac{180}{\pi}\ \text{atan2}(u_{\text{interpolated}}, v_{\text{interpolated}}),\ 360\right)$$

The $\theta_w(x)$ measures the direction from which the wind is blowing from. Therefore, the wind direction toward a specific location is given by:

$$\theta(x) = (\theta_w(x) + 180) \mod 360$$

PM2.5 To estimate PM2.5 levels at the plant locations, I use ground-based data from national air quality monitoring stations in China. For each plant, I identify the three nearest stations by geographic coordinates and apply the IDW method to interpolate the daily average PM2.5 concentrations.

D Wind-Based Exposure Variables

This section outlines the construction of θ, W, and r^{dw}, which are used in equation (2).

To capture daily variation in wind conditions, I divide wind direction into 16 sectors of 22.5° each, following the standard compass rose. Using ERA5-Land data interpolated to plant locations, I record the number of days each sector was dominant between 2020 and 2022.

For each property–plant pair, exposure is considered only when the property lies downwind of the plant. The downwind region is defined as the set of wind sectors where the angle between the wind direction and the plant–property vector is less than 90°. Within this region, I compute two frequency-weighted averages, wind direction θ and wind speed W, given by

$$\theta = \sum_{s \in \text{downwind}} \theta_s \cdot \frac{f_s}{\sum_{s \in \text{downwind}} f_s}, \quad W = \sum_{s \in \text{downwind}} w_s \cdot \frac{f_s}{\sum_{s \in \text{downwind}} f_s}.$$

where f_s is the number of days with wind from sector s, θ_s is the angle between the center of sector s and the plant–property vector, and w_s is the average wind speed in that sector.

I then define the total frequency of downwind days as

$$r^{dw} = \sum_{s \in \text{downwind}} f_s.$$

This variable summarizes how often a property lies downwind of a plant and enters equation (2) as a scaling factor of exposure.

Figure A7 illustrates this procedure. The plant is located at the central point, and wind directions (blowing from the plant) are divided into 16 sectors, with the frequency of each sector represented by the length of the wind rose. The black arrow shows the direction from the plant to the property, and the solid black line perpendicular to it divides the rose into downwind and non-downwind halves. For each sector on the "downwind" side, the angle θ_s (shown by the gray lines) is calculated based on the sector's central angle (shown by the dashed lines) relative to the plant-property axis. These sector-level angles, along with their frequencies and average wind speeds, are then used to

compute these wind-based exposure variables.

E Estimation Details

This Section describes how I estimate the unit price of housing services, $p_{j(k)}$, and the parameters of the residential sorting model. The notation follows that in Section IV.C of the main text.

E.1 Estimating Unit Prices of Housing Services

I assume that the unit price of a bundle of housing services, p_k, is constant within a local area. Specifically, all properties within the same county j are assumed to share the same unit price. This price is estimated using an OLS regression of the log transaction price $P_{k,t}$ on a vector of housing characteristics $h_{k,t}$, year fixed effects τ_t, and county fixed effects $\log p_{j(k)}$:

$$\log P_{k,t} = \log p_{j(k)} + \tau_t + h'_{k,t}\beta_h + \varepsilon^H_{k,t},$$

where $p_{j(k)}$ is unit price of county j where property k locates. $h_{k,t}$ includes (i) housing characteristics such as size, number of rooms, and total floors of the building; (ii) convenience factors, including distances to primary and secondary schools, parks, hospitals, and coach stations, as well as the number of restaurants within 5 kilometers; and (iii) other nearby pollution sources, measured by the number of other industrial plants within 5 kilometers.[20]

Convenience and external pollution variables are constructed from 2019 POI data provided by Baidu Map. These measures are predetermined relative to the housing transactions in the sample, which mitigates potential endogeneity from contemporaneous local economic developments. At the same time, the spatial pattern of amenities changes slowly in general. This slow evolution makes the 2019 data a reasonable proxy for micro-location socioeconomic and environmental conditions during the sample period.

[20]To account for potential COVID-19 effects on housing transactions, I also include the monthly number of newly confirmed COVID-19 cases at the city level, obtained from the Centers for Disease Control and Prevention.

E.2 Three-Stage Estimation

Stage I - Residential Sorting Model via Maximum Likelihood Assuming $\nu_{i,k}$ follows a type I extreme value distribution yields a conditional logit specification. The probability of choosing property k is then given by the following expression, where CS_i denotes the consideration set of household i.[21]

$$\begin{aligned}\mathbb{P}(K=k) &= \mathbb{P}\left(V_{i,k} = \max_{\ell \in CS_i} V_{i,\ell}\right) \\ &= \frac{\exp\left[\alpha_{j(k)} - \beta_Q \cdot \left(Iq_k - r_k^{dw} \cdot \log W_k\right) - \beta_d \cdot r_k^{dw} \cdot D_k - \beta_{I1} \cdot I_{1k} + \beta_{I2} I_{2k} + \beta_{I3} I_{3k}\right]}{\sum_{\ell \in CS_i} \exp\left[\alpha_{j(\ell)} - \beta_Q \cdot \left(Iq_\ell - r_\ell^{dw} \cdot \log W_\ell\right) - \beta_d \cdot r_\ell^{dw} \cdot D_\ell - \beta_{I1} \cdot I_{1\ell} + \beta_{I2} I_{2\ell} + \beta_{I3} I_{3\ell}\right]}.\end{aligned}$$

where $Iq_k = r_k^{dw} \cdot \text{Info}_k \cdot \log Q_k^o$, $I_{1k} = r_k^{dw} \cdot (1 - \text{Info}_k)$, $I_{2k} = [r_k^{dw} \cdot (1 - \text{Info}_k)]^2/2$, $I_{3k} = [r_k^{dw} \cdot \text{Info}_k]^2/2$, $\beta_{I1} = \beta_Q \cdot \mu_u$, $\beta_{I2} = (\beta_Q \cdot \sigma_u)^2$, and $\beta_{I3} = (\beta_Q \cdot \sigma_o)^2$.

With appropriate fixed-effect normalization, all parameters in equation (8) can be estimated by Maximum Likelihood (MLE). The mean μ_u and variance σ_u^2 of $\log Q^u$, together with the perception divergence σ_o^2 from $\log Q^o$, can then be recovered as follows:

$$\mu_u = \frac{\beta_{I1}}{\beta_Q}, \sigma_u = \frac{\sqrt{\beta_{I2}}}{\beta_Q} \text{ and } \sigma_o = \frac{\sqrt{\beta_{I3}}}{\beta_Q}. \tag{E.1}$$

The housing dataset contains six million properties, generating 11.4 billion observations across all choice sets. To reduce the computational burden, the sample is split by city and parameters are estimated at the city level. This procedure yields city-specific pollution belief parameters, as defined in equation (E.1). It also enables analysis of potential sources of belief divergence related to city-specific social and economic characteristics, as discussed in Section V.B.

For the county fixed effects $\alpha_{j(k)}$, since the logit model identifies preferences through log-utility differences, one county fixed effect is set to zero in the MLE for each city to prevent over-parameterization and maintain identifiability. The omitted county serves as the benchmark.

[21] The consideration set includes all properties within the same city as the purchased one, with size variation up to 30 square meters.

Additional details are provided in the Stage III Estimation section.

Stage II **- Aggregating the Coefficients** To account for differences in estimation precision across cities, I aggregate the city-level coefficients using an inverse-variance weighting approach.

Consider a generic parameter λ_c, estimated separately for each city c in Stage I. Its estimate can be decomposed as $\widehat{\lambda}_c = \lambda_c + \widehat{e}_c$, where λ_c is the true city-specific value of the parameter and $\widehat{e}_c$ is an estimation error. λ_c can be further decomposed as $\lambda + \eta_c$, where λ denotes the cross-city average to be recovered and η_c is a centered random variable capturing city-level heterogeneity. Substituting this representation, the estimator can be written as

$$\widehat{\lambda}_c = \lambda + \eta_c + \widehat{e}_c,$$

As the considered $\widehat{\lambda}_c$ are obtained through the MLE in Stage I, an estimator $\widehat{s}^2(\lambda_c)$ of the asymptotic variance, which equals the variance of $\widehat{e}_c$, can be computed. To downweight imprecise estimates, λ is estimated using the weighted city mean:

$$\widehat{\lambda} = \frac{\sum_c \frac{\widehat{\lambda}_c}{\widehat{s}^2(\lambda_c)}}{\sum_c \frac{1}{\widehat{s}^2(\lambda_c)}}.$$

In the absence of heterogeneity, that is when $\eta_c = 0$, $\widehat{\lambda}$ becomes the optimal Generalized Least Squares estimator.

Constructing the confidence interval for λ requires an estimator of the variance of $\widehat{\lambda}$. A simple estimator that accounts for potential city heterogeneity is:

$$\widehat{s}^2(\lambda) = \sum_c \omega_c^2 \left(\widehat{\lambda}_c - \widehat{\lambda}\right)^2, \text{ where } \omega_c = \frac{\frac{1}{\widehat{s}^2(\lambda_c)}}{\sum_\ell \frac{1}{\widehat{s}^2(\lambda_\ell)}}.$$

Stage III **- Estimating $\boldsymbol{\beta_M}$ and $\boldsymbol{\beta_H}$** As mentioned earlier, one county fixed effect is set to zero in the MLE for each city to avoid over-parameterization. Therefore, instead of estimating α_j, the Stage I estimation recovers the fixed effects as $\Delta\widehat{\alpha}_j = \widehat{\alpha}_j - \widehat{\alpha}_{j_0}$, where $\widehat{\alpha}_{j_0}$ denotes the fixed effect

of the benchmark county. According to equation (9), $\Delta\widehat{\alpha}_j$ can be expressed as:

$$\Delta\widehat{\alpha}_j = \beta_M \cdot \Delta \log M_j - \beta_H \cdot \Delta \log p_j + \Delta\xi_j + \widehat{\delta}_j,$$

where the estimation error $\widehat{\delta}_j = \Delta\widehat{\alpha}_j - \Delta\alpha_j$ asymptotically vanishes.

Assume that $\xi_j = Z_j\gamma + \varepsilon_j$, where ε_j captures unobserved county-specific characteristics. The vector Z_j consists of observable county-level characteristics capturing county attractiveness. It includes the total number of schools (primary and secondary) and hospitals (in log), the average $PM_{2.5}$ concentration, population density (in log), and average nighttime light intensity, which serves as a proxy for GDP.[22] Neglecting $\widehat{\delta}_j$ and assuming that ε_j is independent of Z_j, β_M and β_H are estimated using weighted least squares, with weights given by the inverse of the estimated variance of $\Delta\widehat{\alpha}_j$:

$$\Delta\widehat{\alpha}_j = \beta_M \cdot \Delta \log M_j - \beta_H \cdot \Delta \log p_j + \Delta Z_j\gamma + \Delta\epsilon_j. \tag{E.2}$$

This precision-weighted approach assigns greater importance to counties whose fixed effects are estimated more precisely in Stage I, resulting in more efficient and reliable estimates of β_M and β_H.

[22] The numbers of schools and hospitals are calculated by summing the corresponding Points-of-Interest within each county. The county-average PM2.5 concentration is calculated by averaging satellite-derived PM2.5 data within each county over 2020–2022. Nighttime light data are averaged over 2020–2022.

F Additional Tables and Figures

Table A1. Comparison for Different Information Windows - Mean & SD

	INFO		log Q	
Difference in Moving Average	Mean	SD	Mean	SD
30 vs. 60 days	-0.0022	0.0623	0.0003	0.0097
30 vs. 90 days	-0.0047	0.0795	0.0006	0.0133

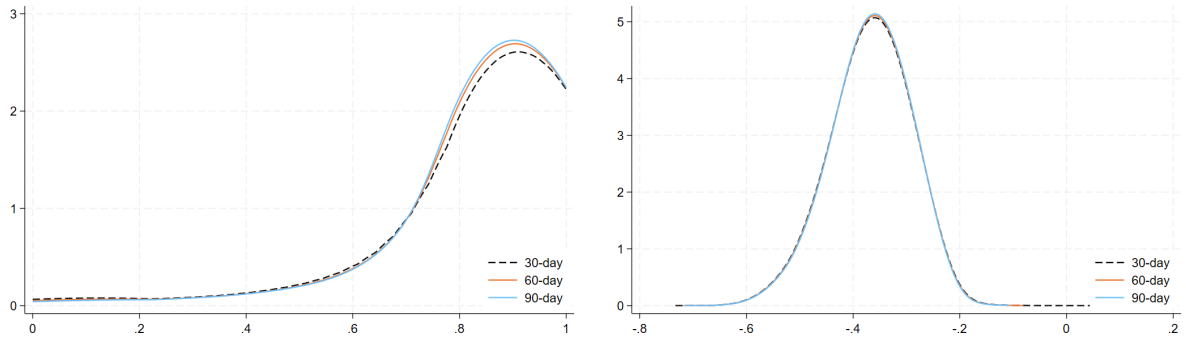
Note: This table compares differences in the moving averages of information quality (INFO) and emission levels ($\log Q$) when calculated using 30-, 60-, and 90-day windows. The reported statistics show the mean and standard deviation (SD) of pairwise differences relative to the 30-day window.

Table A2. Correlation of INFO with Micro-Location Amenities

	(1)	(2)
	All wind directions	Downwind only
# Restaurants	0.014	-1.100
	(0.397)	(0.908)
# Banks	0.174	-1.816
	(1.140)	(2.251)
# Hotels	-0.822	-0.570
	(1.059)	(2.069)
Population density	-0.015	-0.007
	(0.010)	(0.005)
Distance to primary/secondary school	-0.002	-0.000
	(0.002)	(0.001)
Distance to hospital	0.003	0.001
	(0.002)	(0.000)
Distance to park	0.001	-0.000
	(0.000)	(0.000)
Observations	575	565
R squared	0.053	0.060

Note: This table reports OLS regressions of INFO on micro-location amenity measures. Each column presents a separate specification that uses the same set of proxies for neighborhood economic activity and the provision of public facilities. The first column uses amenity measures computed in all wind directions. The numbers of restaurants, banks, and hotels are counted within five kilometers of each WtE plant. Population density is measured as the number of people per square kilometer and averaged within the same radius. All distances are measured in kilometers. The second column uses amenity measures restricted to the downwind sector, which is defined as a wedge that extends from each plant along its dominant wind direction and covers ±1 sector within a 5-kilometer radius. The division of wind sectors is detailed in Appendix Section D. Robust standard errors are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.



(a) INFO (b) log Q

Figure A1. Comparison for Different Information Windows - Kernel Density Distribution

Note: This figure presents kernel density distributions of the moving averages of information quality (INFO) and emission levels (log Q) using 30-, 60-, and 90-day windows. The curves largely overlap, indicating that the choice of window size has little effect on the distribution of either variable.

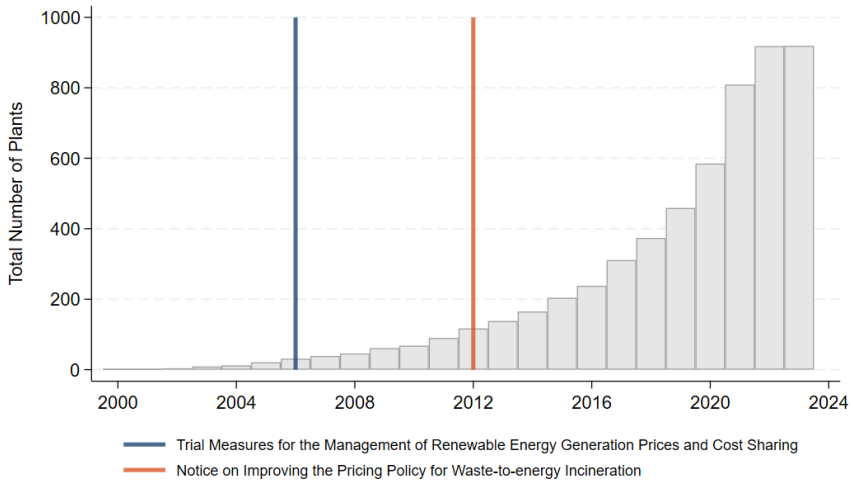


Figure A2. Number of WtE Plants in China

Note: This figure shows the total number of WtE plants in China from 2000 to 2023. The vertical lines indicate two major policies that contributed to the rapid expansion of the WtE sector: the *Trial Measures for the Management of Renewable Energy Generation Prices and Cost Sharing* (2006) and the *Notice on Improving the Pricing Policy for Waste-to-Energy Incineration* (2012).

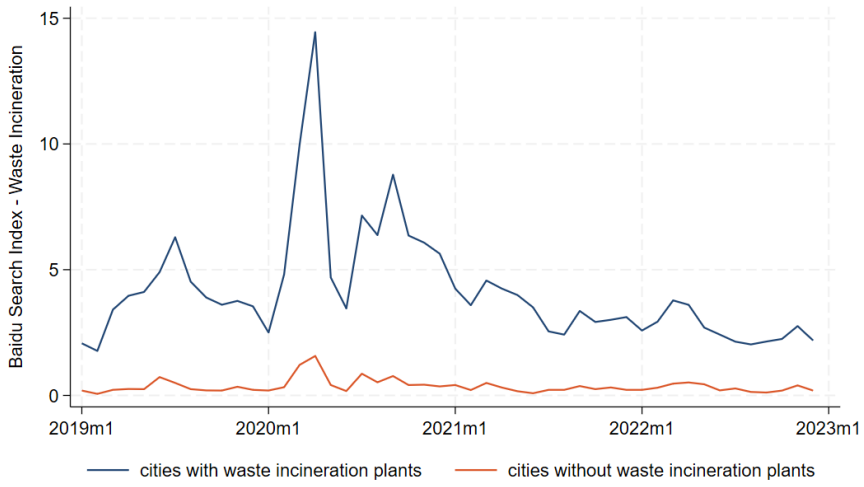


Figure A3. Baidu Search Index for "Waste Incineration"

Note: This figure plots the monthly Baidu Search Index for the keyword "waste incineration" ("垃圾焚烧" in Chinese) from 2019 to 2022. The dark gray line represents cities with WtE plants, and the light gray line represents cities without such facilities. A marked increase in search activity appears after the platform was launched in January 2020, particularly in cities with WtE plants.

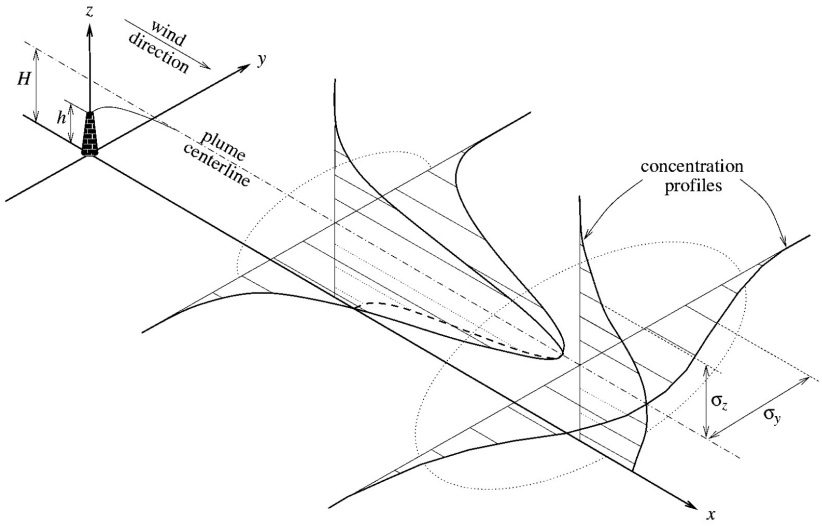


Figure A4. The Gaussian Plume Model in the Downwind Coordinate System

Note: This figure illustrates the Gaussian Plume Model in the downwind coordinate system. It shows how pollutant concentrations disperse horizontally (σ_y) and vertically (σ_z) from the emission source under prevailing wind conditions. The graph is reprinted from Stockie (2011).

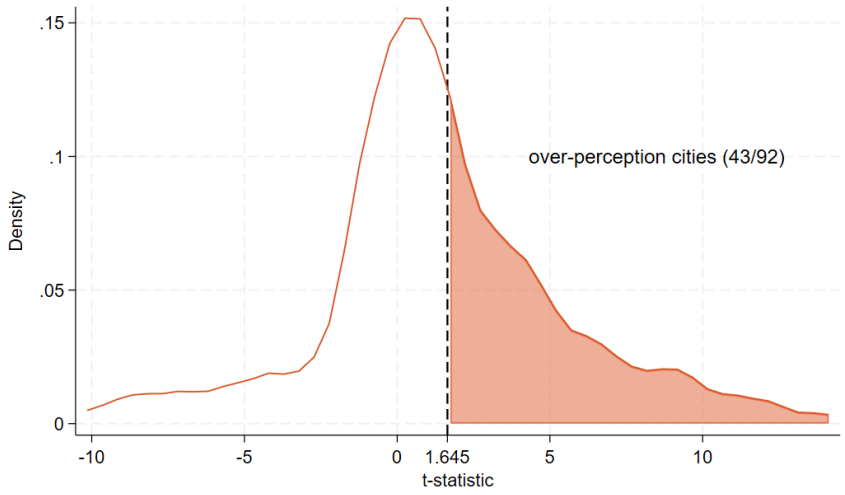
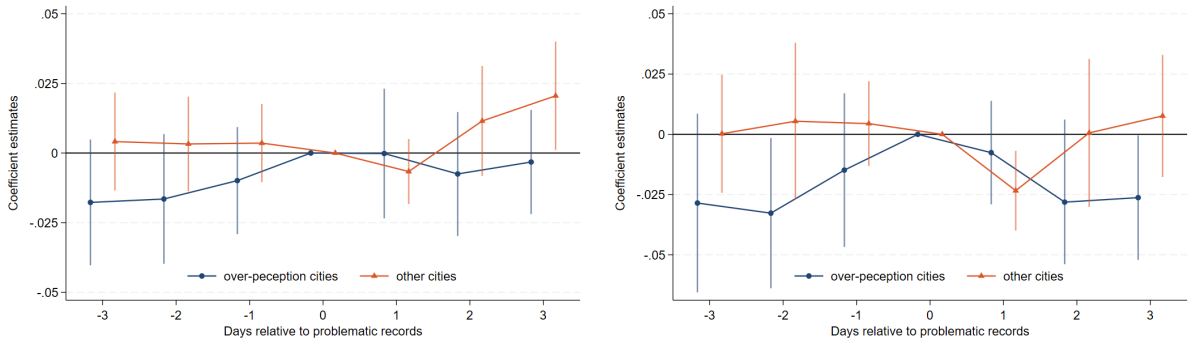


Figure A5. Distribution of T-statistic for $\widehat{\Delta}_{\mu_u}$

Note: This figure presents the distribution of the t-statistic for $\widehat{\Delta}_{\mu_u}$, using the estimated standard deviation of $\widehat{\mu}_u$ to approximate the standard deviation of $\widehat{\Delta}_{\mu_u}$. In 43 out of 92 cities, the t-statistic exceeds 1.645, indicating statistical significance at the 5% level for a one-sided t-test. Therefore, I classify these 43 cities as "over-perception" cities.



(a) ML-predicted PM2.5 levels

(b) Station-based PM2.5 levels

Figure A6. Pollution Around Flawed-Record Dates - *Net of Other Sources*

Note: This figure plots coefficient estimates for PM2.5 after excluding plants within 5 kilometers of other high-polluting sites. Dark gray lines with circles represent over-perception cities, and light gray lines with triangles represent other cities. The estimates include 95 percent confidence intervals with standard errors clustered at the city level.

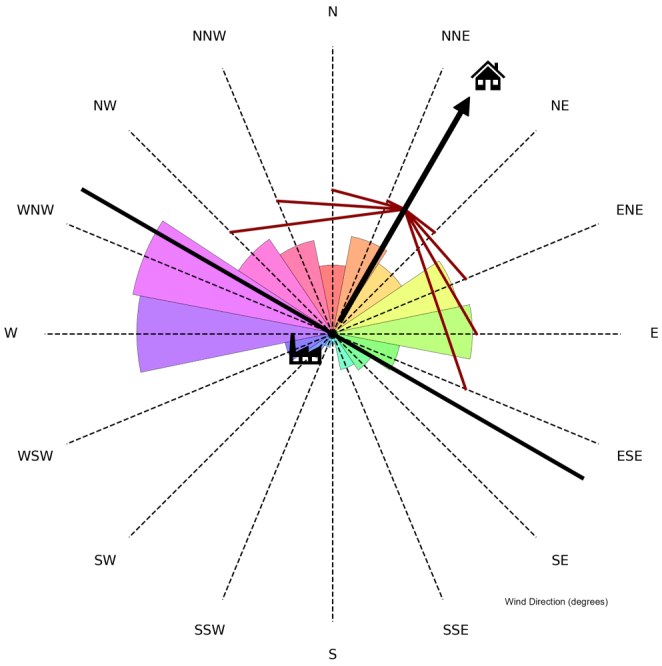


Figure A7. Example for Weighted Wind