Autocracy in the Information Age

November 19, 2024

Pollution Data as a Window Into Modern Autocratic Control

In the age of modern autocracy, brute force has taken a backseat to a more covert tool: information manipulation. Unlike the dictators of the 20th century, who ruled through fear and violence, today's authoritarian leaders maintain their grip on power by projecting an image of democracy while controlling the flow of information. Carefully curated "transparency" bolsters a facade of competence and legitimacy, enough to preserve public trust while avoiding the scrutiny that comes with true accountability.

This evolution isn't just a matter of optics; it's a response to a more interconnected and educated world. Global markets, international monitoring, and increasing domestic education have made overt repression a costly gamble. Instead, autocrats walk a fine line, balancing domestic control with the need to keep foreign investors and allies on their side. Environmental monitoring offers a vivid and urgent case study of this strategy in action; in many authoritarian regimes, government-reported pollution levels often fall suspiciously below independent satellite measurements. It's a reminder that in the modern era, control isn't just about what people fear—it's about what they're allowed to know.

China as a Case Study

Before 2013, China's approach to air pollution data was characterized by limited transparency and potential manipulation. The government controlled the dissemination of environmental information, often underreporting pollution levels to maintain a facade of environmental competence and to avoid public unrest. The lack of reliable public data hindered citizens' ability to make informed decisions about their health and environmental safety. It also impeded international assessments of China's environmental policies and their effectiveness.

In 2013, Beijing's "airpocalypse" drew global attention to China's severe air pollution. In response, the government launched a "War on Pollution" in 2014, implementing stringent measures that led to a 40% reduction in particulate pollution between 2013 and 2021. This shift reveals the paradox of authoritarian control: while it enabled unprecedented progress, it raises critical questions about the trade-offs between efficiency and accountability in governance.

Methodology to Detect Underreporting

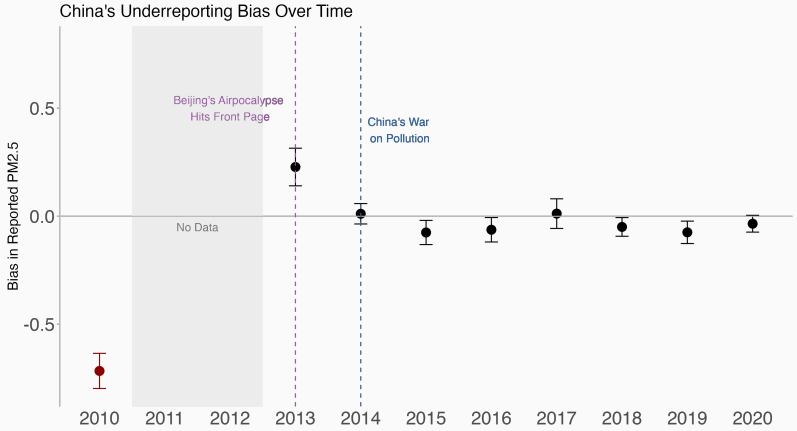
If governments manipulate pollution data, discrepancies should systematically emerge between government-reported data and independent satellite-derived estimates that are free from government control. We can test this by comparing 2 data sources for $PM_{2.5}$

1. Annual city-level average $PM_{2.5}$ reported by national and regional governments. Data is aggregated by the World Health Organization (WHO)
2. Satellite-derived $PM_{2.5}$ data from Donkelaar et al. (2021),

Use regression to compare WHO to satellite-derived data. The regression coefficient shows the estimated change in government-reported $PM_{2.5}$ level for a 1 $\mu g/m^3$ change in the increase in satellite-derived $PM_{2.5}$.

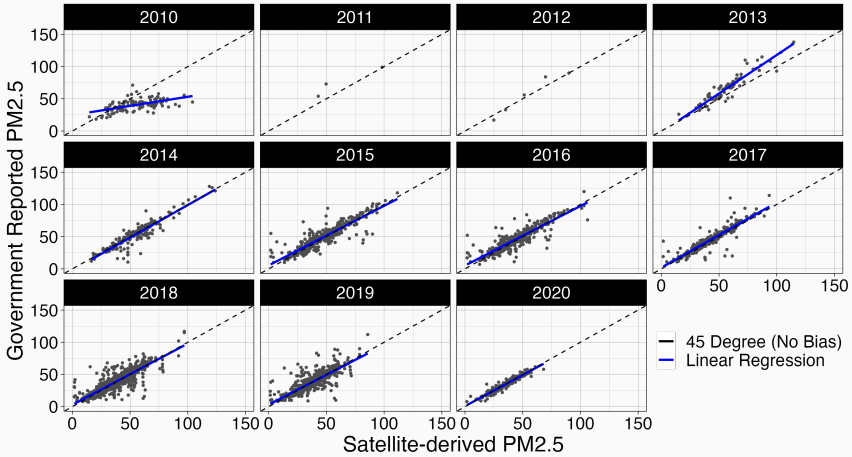
- If there is a close to 1 to 1 relationship between WHO and satellite-derived data, this suggests accurate reporting.
- A regression coefficient that is less than 1 might suggest underreporting.

Timeline of Pollution Reporting Bias in China



Note: To calculate the bias in reported $PM_{2.5}$, I compare government-reported $PM_{2.5}$ against satellite-derived $PM_{2.5}$ estimates using a regression analysis. If the reported data perfectly aligns with the satellite data, the regression coefficient should equal 1. Deviations from 1 indicate bias: values below 1 suggest systematic underreporting. The y-axis shows how far the estimated coefficient deviates from 1, capturing the level of bias. Each data point represents the result of a statistical test to check if the coefficient is significantly different from 1, with the error bars showing the 95% confidence interval (± 1.96 standard errors). If the error bars do not overlap 0, it means the bias is statistically significant.

Raw Data Underlying China's Reporting Bias Estimates



Note: Each dot in the plot represents a city in China. Each panel corresponds to a specific year and shows the relationship between government reported $PM_{2.5}$ levels (y-axis) and satellite-derived $PM_{2.5}$ estimates (x-axis) for these cities. The dashed black line represents the 45-degree line, which indicates perfect agreement between reported and satellite-derived values. The blue line is the linear regression line, summarizing the overall relationship in the data. A blue line with a slope less than 1 (more horizontal) indicates systematic underreporting of pollution levels.

Analysis of Rest of the World

China's Unique Position

China's aggressive approach to combating air pollution is notable for its scale and intensity. Between 2012 and 2020, the country expanded its federal air quality monitoring stations from 661 to 1,800, with additional stations managed by local governments. This extensive monitoring network enabled the implementation of stringent pollution control measures, resulting in a 40% reduction in $PM_{2.5}$ levels from 2013 to 2021. Such rapid progress contrasts with the more gradual improvements observed in countries like the United States, where similar advancements have unfolded over multiple decades.

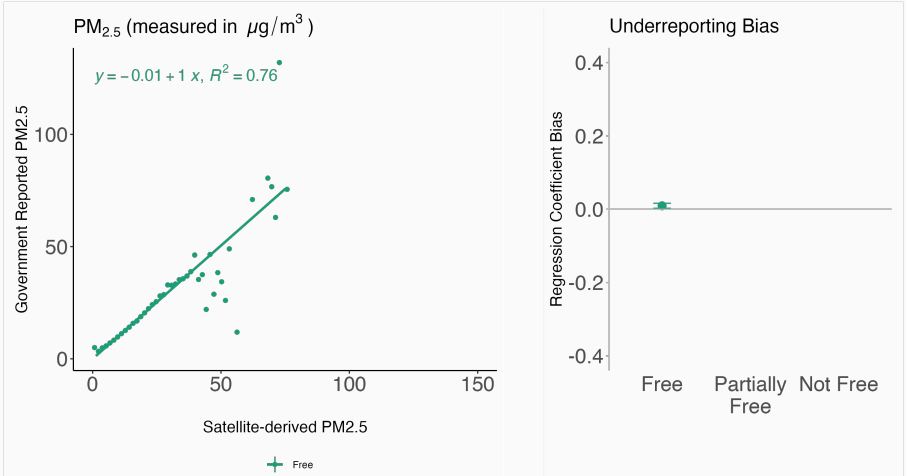
Global analysis of pollution monitoring

China's large-scale efforts to combat pollution have achieved remarkable results, but how do other countries fare in their pollution monitoring and reporting practices? To explore this, we turn to a global analysis, categorizing nations by their democratic freedoms using Freedom House's "Freedom in the World " (FiW) index

- FiW rates countries from most (0) to least democratic (6)
- Freedom House classifies countries as "free" if the FiW index is less than 2, 'partially free' for values $\geq$2 and $\leq$ 4, and 'not free' for values above 4

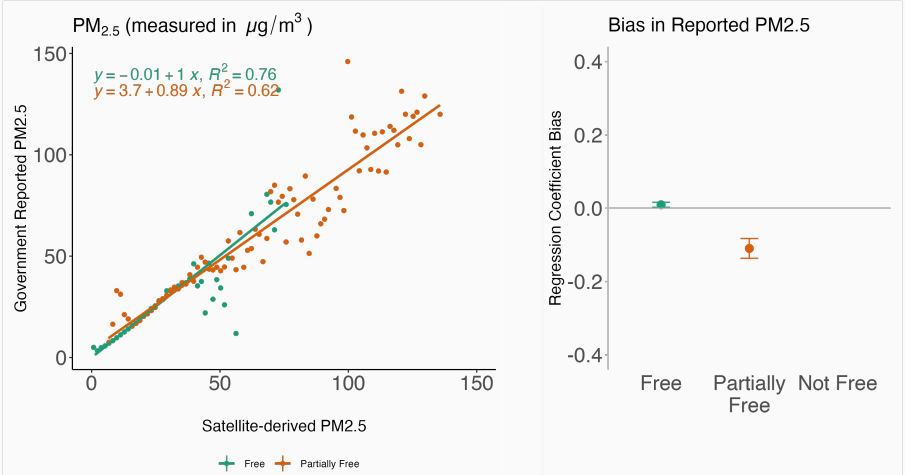
For the rest of the main slides, I use the categorical variable (free, partially free, and not free), excluding Chinese cities from the sample. This approach enables clear visual segmentation, facilitating distinct and straightforward comparisons across groups. A more detailed and formal statistical analysis, included in the appendix, uses both continuous and categorical measures of democracy. Importantly, the findings remain consistent across both approaches.

High accuracy in reporting in “free” countries



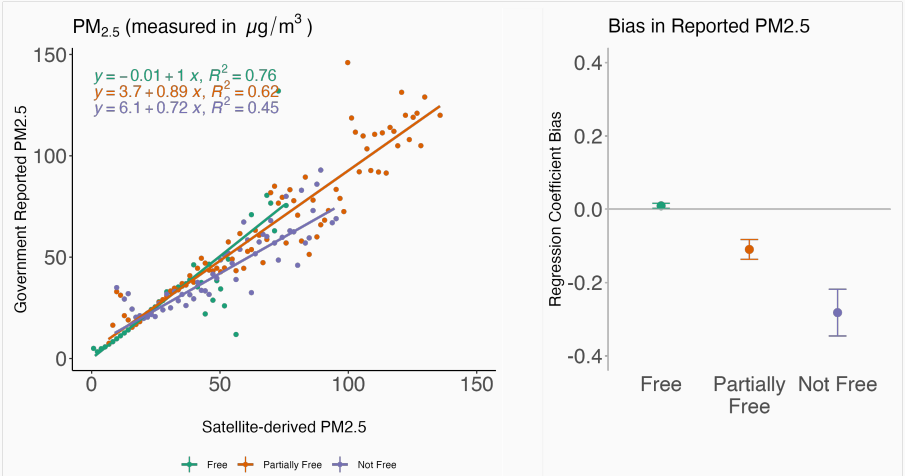
Note: The left panel examines the accuracy of government reported $PM_{2.5}$ levels compared to satellite-derived estimates, with values close to 1 on the y-axis indicating high reporting accuracy. The points show binned averages of city-level measurements. The right panel summarizes these findings by plotting the regression coefficient bias (difference from 1). Error bars represent the 95% confidence intervals.

In “partially free” countries, underreporting bias emerges



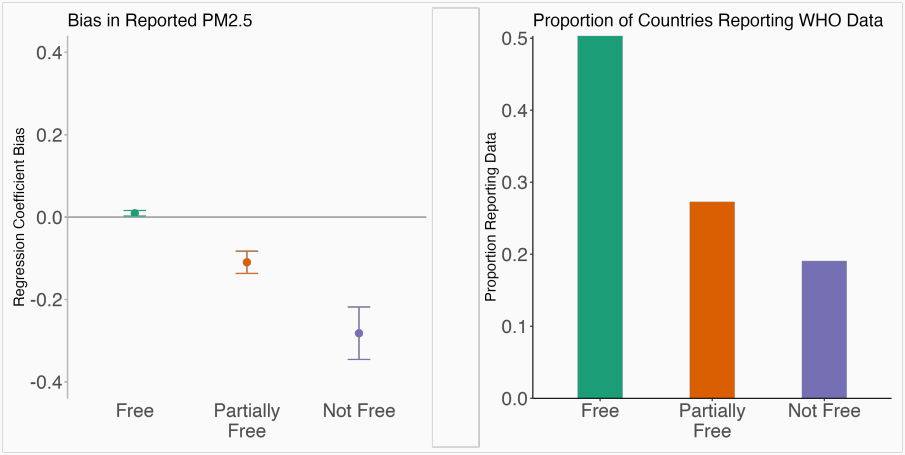
Note: The left panel examines the accuracy of government reported $PM_{2.5}$ levels compared to satellite-derived estimates, with values close to 1 on the y-axis indicating high reporting accuracy. The points show binned averages of city-level measurements. The right panel summarizes these findings by plotting the regression coefficient bias (difference from 1). Error bars represent the 95% confidence intervals.

Severe underreporting in “not free” countries



Note: The left panel examines the accuracy of government reported $PM_{2.5}$ levels compared to satellite-derived estimates, with values close to 1 on the y-axis indicating high reporting accuracy. The points show binned averages of city-level measurements. The right panel summarizes these findings by plotting the regression coefficient bias (difference from 1). Error bars represent the 95% confidence intervals.

“Not free” countries not only underreport, but often do not report data to the WHO



Note: The left panel plots the regression coefficient bias (difference from 1). Error bars represent the 95% confidence intervals. The right panel unit of analysis is at the city-year level; it calculates the proportion of cities reporting to the WHO by democracy category.

Appendix

Data Sources

WHO's database - city-level pollution levels

- Dataset began in 2010
- Sources include (1) reports countries sent to WHO, (2) official national/subnational reports posted publically, (3) regional networks (such as US embassies), (4) values from UN Agencies, Development agencies, and peer-reviewed journals
- For the purposes of this project, only kept local sources, (1) and (2)

Satellite data from Donkelaar et al. (2021), with data going back to 1998

- Calculate average $PM_{2.5}$ levels in 20km circles around the center of cities that were included in the WHO database

Classify political regimes using Freedom House's "Freedom in the World " (FiW) index

- FiW rates countries from most (0) to least democratic (6)
- Freedom House classifies countries as "free" if the FiW index is less than 2, 'partially free' for values ≥ 2 and ≤ 4, and 'not free' for values above 4

Formal Statistical Test of Underreporting from Autocratic Countries

$$Reported_{it} = \delta_t + \gamma_s + \beta_0 Satellite_{it} + \beta_1 Democracy_{it} + \beta_2 (Satellite_{it} \times Democracy_{it}) + \varepsilon_{it} \quad (1)$$

- δ_t is a year fixed effect, which controls for factors that vary across time but are constant across all countries or cities in the sample. These could include global trends or shocks, such as international environmental policies, global economic shifts, or advancements in satellite technology.
- γ_s is a subregion fixed effect. These are 22 subregions of the world defined by the UN. This fixed effect accounts for the fact that countries in a similar region often cultural, geographic, political, or economic similarities.
- i indexes cities

β_2 captures the autocracy effect. A negative value for this coefficient suggests that a same-sized increase in satellite PM2.5 measurements is associate with a smaller increase in reported PM2.5 in more autocratic regimes. In other words, a negative value of β_2 is a test of underreporting.

Results: Evidence of autocratic underreporting

Dependent Variable:	WHO Reported PM2.5	
Model:	(1)	(2)
Variables		
Satellite PM2.5	0.9532***	1.039***
	(0.0433)	(0.0365)
I(Partially Free)	2.596	
	(1.932)	
I(Not Free)	5.180	
	(3.292)	
Satellite PM2.5 × I(Partially Free)	0.0010	
	(0.0688)	
Satellite PM2.5 × I(Not Free)	-0.1753**	
	(0.0750)	
FiW		1.133**
		(0.4957)
Satellite PM2.5 × FiW		-0.0440***
		(0.0131)
Fixed-effects		
UN Sub-Region	Yes	Yes
Year	Yes	Yes
Fit statistics		
Observations	29,943	29,943
R^2	0.74981	0.74862
Within R^2	0.50032	0.49793

Clustered (Country) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Column (1) shows regression using categorical democracy measure

- "Partially free" is for FiW values $\geq$2 and $\leq$ 4, and "not free" is for values above 4. The base excluded category is FiW values less than 2.
- (1) coefficient on $SatellitePM2.5 \times I(NotFree)$ is negative and significant

Column (2) shows regression using continuous FiW democracy measure

- The FiW ranges from 0-6, with lower values corresponding to greater enjoyment of rights.
- Coefficient on $SatellitePM2.5 \times FiW$ is negative and significant

Both columns show not free countries underreport their PM2.5 levels as compared to free countries

Censoring in the WHO data: more autocratic countries are less likely to report

$$I(\text{Reported Data to WHO})_{it} = \gamma_s + \delta_t + \beta Democracy_{it} + \varepsilon_{it}$$

Dependent Variable:	I(Reported Data to WHO)	
Model:	(1)	(2)
Variables		
I(Partially Free)	-0.1127*	
	(0.0621)	
I(Not Free)	-0.2093***	
	(0.0707)	
FiW		-0.0479***
		(0.0147)
Fixed-effects		
UN Sub-Region	Yes	Yes
Year	Yes	Yes
Fit statistics		
Observations	2,304	2,303
Adjusted R^2	0.32809	0.33080

Clustered (Country) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Robustness Tests

To check if the results are robustness, I run four types of checks:

1. Using alternative definitions of satellite-derived $PM_{2.5}$, where I vary the radius size for constructing the measure
2. Different fixed effects specifications
3. Run Equation 1, excluding one sub-region at a time, to test if one sub-region is driving the results
4. Run Equation 1, excluding one country at a time, to test if one country is driving the results

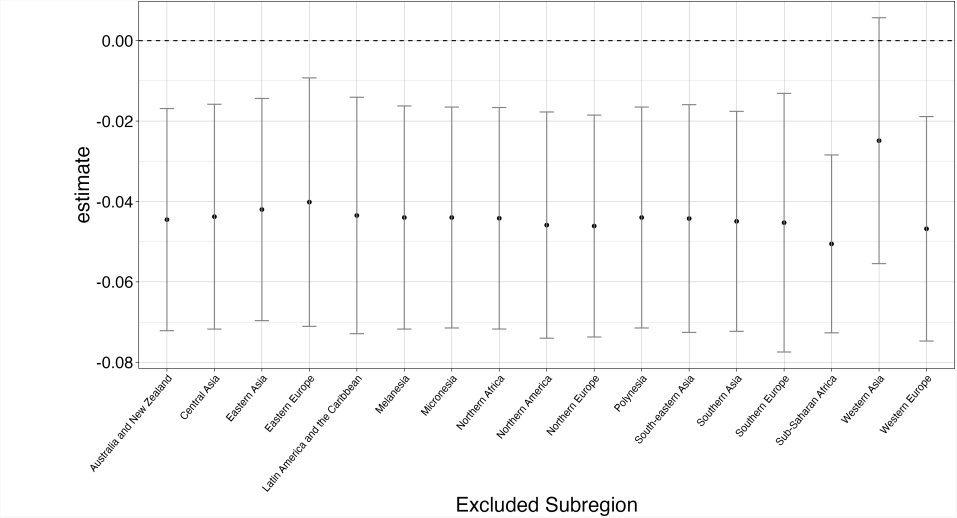
All robustness checks show consistent evidence of underreporting from more autocratic countries

Robustness: Alternative Radius Sizes For Satellite PM2.5 Levels and Fixed Effect Specifications

Model:	Main Spec (1)	10km (2)	50km (3)	FiW Quad (4)	Country FE (5)	Year (6)	Country + Year FE (7)
Variables							
Satellite PM2.5	1.039***	1.048***	1.025***	1.026***	1.023***	1.048***	0.9833***
	(0.0365)	(0.0370)	(0.0391)	(0.0331)	(0.0450)	(0.0256)	(0.0468)
FiW	1.133**	1.081**	1.000*	1.501**	-1.182	1.017***	-0.7476
	(0.4957)	(0.4886)	(0.5116)	(0.5922)	(0.9065)	(0.3675)	(0.9075)
Satellite PM2.5 × FiW	-0.0440***	-0.0449***	-0.0351**	-0.0378**	-0.0328	-0.0439***	-0.0202
	(0.0131)	(0.0130)	(0.0139)	(0.0164)	(0.0210)	(0.0129)	(0.0234)
FiW square				-0.1198			
				(0.1959)			
Fixed-effects							
UN Sub-Region	Yes	Yes	Yes	Yes			
Year	Yes	Yes	Yes	Yes		Yes	Yes
Country					Yes		Yes
Fit statistics							
Observations	29,943	29,938	29,961	29,943	29,943	29,943	29,943
R^2	0.74862	0.75199	0.74208	0.74876	0.77457	0.74728	0.77702
Within R^2	0.49793	0.50472	0.48533	0.49822	0.43107	0.74162	0.40745

Clustered (Country) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Robustness: Excluding One Subregion at a Time



Robustness: Excluding One Country at a Time