1 Introduction

explain why: is regulation site-concentration specific? not obvious to all

I'm not a fan of this phrase, not sure what it means

A critical input to good air quality regulation is good air quality measurements. Economically speaking, the relative efficiency of pollution regulation hinges on our ability to accurately assess air quality across the country. The United States implemented a network of air quality monitors in the 1970s that report air quality to regulators. These air quality monitors are managed by local and state officials who control when the monitors are on or off. These local and state officials are part of the organization that would receive sanctions if air quality in their region falls below federal standards. Given the potential for large fines and costly, forced technology adoption, incentives exist to behave strategically within this reporting system. Additionally, federal regulations have wiggle room in what portion of air quality readings need to be valid. For instance, when measuring particulate matter in the air (one of the most common types of pollution), the EPA allows up to 25% of measurements to be missing or omitted (EPA, 2017). There are currently no restrictions for which 75% of readings must be reported.

see below

these pars can be combined

Indeed, previous research suggests that there is mismeasurement of air quality occurring. Zou (2021) and Mu et al. (2021) provide evidence of strategic behavior in pollution measurement on behalf of local pollution regulators. With allowed gaps in reporting and incentives to behave strategically, is it possible that reported pollution measurements are biased? How could we measure this bias?

see below

see below

This paper examines the questions: What bias exists in reported air quality and how might this bias change recent pollution regulation? To explore these issues, I utilize a new dataset of air quality measurements collected from consumer products. These data ~~help~~ provide an independent groundtruth comparison to air quality reported to the EPA. The most promising new data coming from these consumer products are PM2.5 measurements – the concentration of particles in air that are 2.5 micrometers and below. Specifically, I use PM2.5 measurements from consumer products that surround federally-regulated monitoring stations to estimate the PM2.5 value at the monitoring station. This allows me to construct

promising why?

★ I would like to see this context a little further down in the intro, and instead in the first par give me a more general idea of what this is about:

- "regulators regulating themselves"?
- misaligned incentives between federal & local governments

→ then interpret this list of facts through that lense

★ When you say bias, I immediately think the value of the measurement is wrong, which is not what I think you're getting at. maybe say,

"is it possible that the empirical distribution of reported measurements is a biased estimate of the true distribution of pollution levels?"

★ This may be just me, but you ask a question, ignore that question, then say "this paper asks..."

I don't like this structure; one way to change it could be to restate one or more of these as an imperative. ie,

Given that local regulators have an incentive and the ability to selectively not-monitor local air pollution, we may expect substantial deviations ~~from~~ from the true level of pollution in the reported data. This paper examines...

this is a bit confusing - can it be made clearer or shelved ~~to~~ until more detail is provided?

predicted values of PM2.5 at the station during times when the station was shut down.

I compare these composite estimates of PM2.5 with federal standards authorized by the Clean Air Act (CAA) – the National Ambient Air Quality Standards (NAAQS). Each quarter (3-month period), statistics of the station's daily PM2.5 values are compared to the NAAQS for PM2.5; the EPA refers to these statistics as *design values.*[1] For PM2.5, there is an "annual" and '24-hour" standard and associated design value to calculate. If a station's annual design value is above the annual standard, then the station (and associated region) is determined to be in non-compliance with the standard. Using the full reconstructed dataset of PM2.5 (PM2.5 estimates for all hours that would be reported from a given station), I then construct counterfactual estimates of the design values that determine if a region is in or out of compliance with the NAAQS for PM2.5. I use these counterfactuals to determine which regions would have changed compliance status status if they reported 100% of their PM2.5 measurements – I call these "flipped regions". I also examine how close these flipped regions were to the regulatory threshold and report a measure of the bias related to the station's missing PM2.5 readings.

not really if things are strategic - wouldn't they shut them all down?

Fundamentally, I am analyzing an issue of missing data from sparsely placed monitors – if there were many monitors placed close by, we could partially make up for air quality data missing from one monitor with a nearby monitor. Previous literature has examined the sparse distribution of regulation-grade monitors and the resulting sensitivity of CAA air quality regulation. Grainger et al. (2019) and Grainger and Schreiber (2019) identify a principle-agent problem with the initial spatial placement of sparse pollution monitors; they find evidence that local regulators may be strategically locating their air quality monitors

what does strategic mean for these?

based on pollution, race, and socioeconomic characteristics. To address the issue of sparse data and fill in the gaps, several authors have used satellite data products to provide finer resolution pollution data (Sullivan and Krupnick 2018, Fowlie et al. 2019). Zou (2021) also

[1]For example, the "annual standard" for PM2.5 is compared to a station's "annual design value:" an average of the last three running annual averages of daily averages of hourly PM2.5 measurements. There is also a "24-hour standard;" these statistics are discussed more in the Data section. Specific formulas for these statistics are listed in the appendix.

uses satellite estimates to discuss the issue of strategic behavior in reaction to the timing of pollution monitoring. He provides evidence that some areas have significantly worse air quality on unmonitored days. In related work, Mu et al. (2021) show potential for strategic monitor shutdowns on days of expected high pollution, contributing to air quality data that is missing *in time*.

the beginning of your intro gave me the idea that you were also imputing in time → which is it?

This paper is most similar to the analysis in Fowlie et al. (2019) where they use PM2.5 estimates generated from satellite data to examine counterfactual compliance status. However, they end their analysis noting that the satellite-based data commonly used in these applications has significant prediction error in some areas; this can cause result in incorrect conclusions about design values. This paper compliments their analysis and that of Mu et al., utilizing a different form of ground-truth PM2.5 data to address missing air quality data *in time* rather than *in space*. While satellite-based PM2.5 estimates have potential for large prediction errors, PurpleAir sensors can be fairly accurate measures of their local air quality[2] and can be averaged over multiple nearby sensors. PurpleAir data also have drawbacks however – the sensors are highly non-uniform in coverage across the US and are sensitive to specific placement by the consumer, perhaps leading to hyper-local estimates of air quality.

this is the first time you mention PurpleAir → introduce them first, or here or earlier

For these reasons, this analysis should be seen as a compliment to previous works. As consumer sensors become more widespread, we can augment reliable federal air quality measurements with a growing number of auxiliary data points to better understand the shape of mismeasurement in air quality. In this paper, I explore one way of leveraging these data to test for issues with biased reporting of air quality. After predicting missing observations using PurpleAir measurements, I find that XX of XX monitoring stations flipped their NAAQS compliance status in at least one quarter between 2018 and 2020 and that YY had more than one flipped status. On average, the design values for the monitor-quarters that flipped

[2]PurpleAir sensors have specifically been shown to be less accurate at high levels of PM2.5 concentration. However, the EPA has developed a correction technique that result in PurpleAir readings within 5% of co-located EPA monitors. This correction technique is used here and explained in more detail in the appendix.

out of sample success rate? will this be an outcome?

status had a difference of more than AAμg/m^3 and BBμg/m^3 of PM2.5 for the annual and 24-hour design values, respectively. We know from previous research that pollution in non-attainment areas has been decreasing at significantly faster rates since the introduction of the CAA (Currie et al., 2020); combined with my results, this suggests that changes in reporting standards to decrease allowable omitted observations may result in more non-attainment areas and further increases in air quality.

The remainder of this article is organized as follows. Section 2 briefly reviews the history of air quality standard in the US and some key details of current regulations. Section 3 explains the theoretical framework applied to the problem of measuring pollution at specific points in space. Section 4 then discusses the data used and section 5 describes the empirical framework that will be applied to estimate the missing pollution and resulting policy outcomes. Section 6 reviews the results of the empirical study and discusses the implications. Section 7

this makes me think of Jim's comment about the goal per se not being ↓ pollution → maybe reframe saying improving measurement / decreasing permitted non-monitoring time increases regulatory efficiency by better targeting non attainment?

overall looks great, congratulations! I can't wait to see the cool graphs I know will come out of this.

xoxo gossip girl