Power Plants, Air Pollution, and Regulatory Rebound

Meredith Fowlie
UC Berkeley

Edward Rubin
University of Oregon

Catherine Wright
Reddit

March 2022
Some motivating observations

• EPA Clean Air Act regulations consistently deliver the greatest monetized benefits of any federal regulatory agency actions (OMB, 2019).

• Reduced exposure to PM$_{2.5}$ is the most economically significant source of assessed benefits from EPA regulations.

• A significant share of these benefits are achieved *indirectly* via rules that limit precursors to particulate matter formation (*NOx* and *SO$_2$*)
Research questions

**Q₁** When the power sector reduces emissions, do we see the air quality benefits that were projected/expected

- Predict the impacts of significant power plant emissions reductions on downwind PM$_{2.5}$ concentrations using alternative modeling approaches.
- Systematically compare these predictions against realized PM$_{2.5}$ outcomes.

**Q₂** Do local regulators ‘take back’ air quality improvements in constrained areas?

- In areas constrained by NAA AQs, reductions in regional pollution could be offset by permitting of local sources.

**Q₃** If local regulations do respond, what are the efficiency and equity implications?
Existing Research and Contributions

• **Strategic behavior of local regulators**: (e.g. Zou, 2021; Gossett and Schlenker, 2021)
  ▶ Recent literature focuses on short-run response to monitoring rules/meteorological shocks.

• **Reduced complexity modeling of PM$_{2.5}$ formation and exposure** (e.g. Di et al. 2019; Fann et al. 2019; Henneman et al. 2020)
  ▶ We integrate particle transport modeling and machine learning prediction into an empirical model of PM$_{2.5}$ exposure.

• **Air pollution, air quality, and health**:
  ▶ **Short(er) run impacts**: IV strategies have been used to identify health-related impacts of air quality improvements (e.g. Deryugina et al. 2019; Deschenes et al. 2017).
  ▶ **Long run impacts drive benefits assessments**: Workhorse integrated assessment models are used to estimate longer run health impacts of pollution reductions (e.g. Muller, Mendelsohn and Nordhaus, 2011; Holland et al. 2018; Jenkins et al, 2021).
Outline

- Institutional context
- Theoretical framework
- Empirical strategy and data
- Predicting impacts of power plant emissions on downwind PM$_{2.5}$
- An empirical test for regulatory rebound
- Implications and next steps
Overlapping regulations

There are multiple provisions of the Clean Air Act that are designed to reduce levels of ambient PM$_{2.5}$ exposure:

1. **Direct regulation of ambient PM$_{2.5}$**:
   - Threshold-based **NAAQs** are designed to prevent ambient concentrations from exceeding a prescribed level.
   - Failure to comply imposes significant economic costs.

2. **Regulation of PM$_{2.5}$ precursors**:
   - Ambient PM$_{2.5}$ in the US is mostly ‘secondary’. Precursor emissions (SO$_2$ and NO$_x$) increase PM$_{2.5}$ levels downwind.
   - Rules targeting power plant emissions of SO$_2$ and NO$_x$ have become increasingly stringent.
Power plant SO2 emissions: 2005-2019

SO2 (Daily Tons)

Baseline

Regime 1

Regime 2

SO2 Regulation

NAAQS 2006

NAAQS 2012

2005

2010

2015

2019

ARP Designation

CAIR SIP Designation

CSPAR + MATS SIP
Regulatory rebound?

**Hypothesis:** When air-quality standards bind locally, sustained reductions in upwind/background emissions may induce a rebound in local emissions (e.g., permitting).

**Implication:** Effects of power plant emissions will differentially affect regions that are more versus less constrained by NAAQS.

Existing literature documents significant decreases in pollution concentrations following non-attainment designation. Ex: Chay and Greenstone (2003); Auffhammer, Bento, Lowe (2009); Bishop, Ketcham, Kuminoff (2019)

Why worry about local regulatory rebound?
NAAQS constrained county

$\lambda_0$
NAAQS constrained county sees a reduction in background PM$_{2.5}$
Local emissions ‘rebound’ in response

\[
\text{MAC}_0 \rightarrow \text{MAC}_1
\]

PM$_0$ − $\Delta_1$ → PM$_1$ → PM$_0$ → PM$_2.5$

$\lambda_0$$\lambda_1$

Rebound
Why worry about local regulatory rebound?
1. **Air quality modeling:** Analyze relationships between power plant emissions of NOx and SO$_2$ and downwind PM$_{2.5}$ concentrations during a base period (2005-2009).

2. **Groundtruthing:** Predict changes in PM$_{2.5}$ concentrations post-2010. Compare predictions with observed PM$_{2.5}$ monitor readings.

3. **Empirical test for regulatory rebound:** Do model projections systematically overstate air quality improvements in areas where NAAQS bind?

4. **Work in progress:** Assess the health impacts of power sector emissions reductions in constrained versus unconstrained areas.
Data sources

**EPA Continuous Emissions Monitoring System (CEMS) data:**
- Hourly emissions of $\text{NO}_x$, $\text{SO}_2$, $\text{PM}$ from all thermal electricity generating units >25 MW.
- Boiler level information: location, stack height, etc.

**EPA Air Quality System (AQS) data:**
- Daily monitor readings from 88101, 88502, IMPROVE monitors.

**Meteorological data:**
- NOAA’s North American Regional Reanalysis (NARR) daily reanalysis data Mesinger2006, NARR.
- PRISM’s data on daily temperature, precipitation, and vapor pressure deficit.

**Wildfire smoke**
- Smoke-plume data from the National Oceanic and Atmospheric Administration’s (NOAA) Hazard Mapping System (HMS).
Two modeling approaches

1. **Off-the-shelf:** *Intervention Model for Air Pollution (InMAP)* is a reduced-complexity model of pollution transport and PM formation:
   - Coarsely approximates pollutant transport and downwind PM$_{2.5}$ formation
   - Easy, fast, cheap (low computational requirements)

2. **DIY:** *HYSPLIT + ML prediction*
   - HYSPLIT models 3-D particle transport w/ high-frequency (hourly) meteorological data
   - We use these simulated trajectories as inputs into a data-rich prediction exercise.
Our preferred approach (overview)

- We leverage recent advances in high-resolution atmospheric particle-trajectory simulation (HYSPLIT), rich meteorological data, and machine-learning prediction.
- We predict the downwind impacts of reductions in emissions from power plants, historically the most significant point source of precursor emissions in the United States.
- We use residualized PM$_{2.5}$ and predictors. This focuses the model on predicting the within-variation in that remains after conditioning on rich fixed effects.
Simulated particle transport and dispersion from three sample power plant stacks over two days
Features we use to predict PM$_{2.5}$ concentrations:

- Pollutant-specific $E_{ijtp}$ bucketed by distance from source (measured in kms): [0-100), [100-500), [500-100), [1000-2000), and [2000,inf)
- Prism meteorological data: Temperature, precipitation, minimum/maximum vapor pressure deficit (mean/min/max).
- Wildfire smoke data.
- Monitor-specific average PM$_{2.5}$ in 2005

Random forest algorithm (standard)

- Fix number of trees at 200.
- Tuning parameters are (1) the number of variables to try at each node and (2) the minimum number of observations in a leaf to be considered for further splitting.
- We choose the tuning parameters using cross validation (13 rolling folds)
- Pick the combination of tuning parameters that results in the lowest average RMSE across all 13 assessment sets.
Feature importance - Random Forest

Variable importance (mean decrease in impurity)

Predictor

- SO$_2$ 100–500 km
- # Low smoke days
- SO$_2$ 500–1,000 km
- # Medium smoke days
- # High smoke days
- NO$_x$ 500–1,000 km
- Precipitation (avg)
- SO$_2$ 0–100 km
- Precipitation 75th p
- SO$_2$ 1,000 km
- SO$_2$ >2,000 km
- NO$_x$ 100–500 km
- Temperature 75th p
- NO$_x$ 1,000–2,000 km
- Vapormax (avg)
- Temperature 90th p
- NO$_x$ >2,000 km
- Precipitation 50th p
- Temperature 10th p
- Month
Regression-based comparison of fit: $PM_{it} = \alpha_i + \delta_{rt} + \beta \hat{PM}_{it} + \epsilon_{it}$

Table: Evaluating models’ out-of-sample predictive performances

<table>
<thead>
<tr>
<th>Dep. variable: Monthly avg. PM$_{2.5}$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{PM}_{2.5}$ Random forest</td>
<td>0.93**</td>
<td>0.77**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td>(0.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{PM}_{2.5}$ LASSO</td>
<td>1.05**</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{PM}_{2.5}$ InMAP</td>
<td>1.02**</td>
<td>0.26*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.11)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model fit

- $R^2$          | 0.63 | 0.60 | 0.55 | 0.63 |
- $R^2$ within  | 0.24 | 0.19 | 0.08 | 0.25 |

The two ML models trained on data from 2005–2009 (and InMAP is based upon 2005 meteorology). This regression only uses ‘testing’ data the models have not seen, 2010–2017. Standard errors clustered at the state-month level. *$p < 0.05$; **$p < 0.01$. 
The training period is 2005-2009; the testing period is 2010-2017.

Time period:
Pre (training)  Post (testing)

Model:
InMAP  LASSO  RF
Evidence of rebound? An error decomposition exercise

$$ERROR_{irt} = \sum_j \beta_j \Delta \overline{EGUPM}_{irt} \cdot D_{ijt} + \alpha_i + \delta_{rt} + \sum_y \sum_j \gamma_{yj} D_{ijt} + \epsilon_{it}$$

- $ERROR_{irt}$: Prediction error: $PM_{2.5 \ it} - \overline{PM}_{2.5 \ it}$
- $\alpha_i$: Monitor fixed effects
- $\delta_{rt}$: Region-by-month-of-sample fixed effects
- $\Delta \overline{EGUPM}_{irt}$: Predicted change in power plant PM$_{2.5}$ contribution (post-2009)
- $D_{ijt}$: Indicators for whether monitor $i$ is constrained/unconstrained by NAAQS standard
Table: Regression-based test for regulatory rebound using HYSPLIT+ML predictions

<table>
<thead>
<tr>
<th>Dep. variable: Prediction error (observed – predicted PM$_{2.5}$)</th>
<th>(1) Annual</th>
<th>(2) Annual</th>
<th>(3) Annual + 24hr</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \text{EGU} \times D_{j=\text{constrained}}$</td>
<td>1.01**</td>
<td>1.05**</td>
<td>1.02**</td>
<td>0.92**</td>
<td>0.92**</td>
<td>1.12**</td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td>(0.33)</td>
<td>(0.35)</td>
<td>(0.30)</td>
<td>(0.30)</td>
<td>(0.36)</td>
</tr>
<tr>
<td>$\Delta \text{EGU} \times D_{j=\text{unconstrained}}$</td>
<td>0.39**</td>
<td>0.44**</td>
<td>0.37**</td>
<td>0.42**</td>
<td>0.41**</td>
<td>0.47**</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
</tbody>
</table>

Include extreme events ✓ ✓ ✓ ✓ ✓ ✓

$\Delta \text{EGU} \times \text{Baseline}$ 06-09 06-09 06-09 06-09 06-09 08-09

EPA Region × month-of-sample ✓ ✓ ✓ ✓ ✓ ✓

State × month-of-sample ✓ ✓ ✓ ✓ ✓ ✓ ✓

Observations 121,067 121,067 121,030 121,217 121,217 121,217

Monitor sites 1,180 1,180 1,184 1,184 1,184 1,184

$R^2$ 0.22 0.35 0.22 0.35 0.35 0.35

$F$ statistic 2.67 3.37 3.71 2.90 2.90 3.28
A random-forest prediction model using hourly particle trajectory simulations and rich meteorological data outperforms standard reduced-complexity approaches in the prediction of monthly average PM$_{2.5}$.

Prediction errors are larger (on average) in areas with high PM$_{2.5}$ concentrations – consistent with our rebound story.

An error decomposition exercise further finds that prediction errors are increasing with simulated reductions in power plant contributions. This is consistent with our local regulatory rebound hypothesis.

However, we also find a (weaker) relationship between simulated power plant emissions reductions and prediction errors in unconstrained locations.
Work in progress...

- Additional robustness checks.
- Explore downwind spillovers as a possible explanation for why prediction errors increasing with dose reductions in unconstrained areas.
- Incorporate additional data sources possible mechanisms driving local emissions rebound (e.g. ECHO database, SIPs).
- Use our RF-HYSPLIT model to isolate reductions in local PM$_{2.5}$ concentrations associated with reduced power plant emissions. Strong first stage could support IV estimates of mortality impacts.
- Bound the efficiency implications of local regulatory response