

Combining Remote Sensing and Cell Phone Users' Mobility Data to Monitor the Impact of Transportation on NO₂ Concentrations in India

Corbett Grainger

Adam Theising

Fan Zhang



WORLD BANK GROUP

Infrastructure Chief Economist Office

December 2022

Abstract

Estimating the extent to which transportation contributes to air pollution levels has been hampered by the difficulty in separating the relative degree of ambient nitrogen dioxide generated by transportation, power generation, and industrial activity—all of which play roles. This paper addresses this gap by isolating the impact of ground-level mobility on air pollution in India through a combination of remotely sensed tropospheric nitrogen dioxide measures and data from mobile phone users' locations. The paper constructs vectors of ground-level movement of cell phones to estimate

the impact of daily changes in mobility within a given district, controlling for both daily thermal electricity generation from upwind power plants and trends in ambient pollution concentrations over time and space. The findings show that tropospheric nitrogen dioxide concentrations are very responsive to changes in mobility, and that the effect varies with population density. The findings show that a 1 percent increase in mobility increases nitrogen dioxide concentrations by more than 2 percent, suggesting that traffic congestion plays a significant role in air pollution.

This paper is a product of the Infrastructure Chief Economist Office. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The author may be contacted at fzhang1@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Combining Remote Sensing and Cell Phone Users' Mobility Data to Monitor the Impact of Transportation on NO₂ Concentrations in India

Corbett Grainger^a Adam Theising^b Fan Zhang^c

JEL: Q53, R41, K32

^a Associate Professor, Department of Agriculture and Applied Economics, University of Wisconsin; CESifo; and Centre for Climate and Energy Transformation, Universitetet i Bergen.

^b US Environmental Protection Agency.

^c Lead Economist, Water Global Practice, World Bank.

1. Introduction

The ability to accurately monitor emissions from transport-related activities with a high degree of spatial and temporal resolution is important for understanding the impact of transportation on air quality and human health. However, measuring the impact of traffic on ambient air pollution concentrations is difficult, particularly over large areas. The impact of traffic on pollution varies with vehicle characteristics, traffic congestion, and environmental conditions. There is a dearth of information on the number of vehicles on the road at any time. Even indicators of fuel use at the regional level are typically unavailable at the daily level. Traditional environmental indicators also have drawbacks; they are generally backward looking and often only available at coarse levels of spatial resolution. In developing country contexts, in particular, the collection of key environmental indicators relies on administrative procedures that take time and may be subject to biases resulting from existing political incentive structures (Xiong, 2018; Sandefur and Glassman, 2015).

In this paper, we use data from mobile-phone users' locations to construct vectors of ground-level movement, and combine these data with remotely sensed nitrogen dioxide (NO₂) measures to better understand how transportation infrastructure usage impacts air pollution in India. To better measure the relative contribution of traffic to emissions, we exploit changes in mobility within a special unit to estimate elasticities of NO₂ with respect to on-the-ground mobility. To understand the impact of ground-level transportation on pollution, our specifications leverage variations in mobility within a given district and by the day of the week. We use combinations of fixed effects to control for daily thermal electricity generation by upwind power plants and for trends in ambient pollution concentrations over time and space. Our study period leverages the big, abrupt changes in mobility that were a result of the COVID-19

pandemic in India. Other studies have pointed to the large changes in pollution that resulted from shutdowns during the pandemic, but to our knowledge we are the first to use this variation to study the relative impact of transportation on NO₂ concentrations.

We examine the impact of movement between 16x16km Bing Map tiles by leveraging within-pixel-by-day-of-the-week variation in mobility. We find an average elasticity for “movement” of greater than two, meaning that a 1 percent increase in movement between these tiles translates to more than a 2 percent increase in tropospheric NO₂. We also find that the elasticity with respect to *non-movement* is slightly larger in magnitude and negative. Our estimates suggest that a 1 percent increase in the number of people *who do not move* between tiles is associated a decrease in in NO₂ concentrations of 3 to 4 percent. In theory, we would expect an increase in the number of people *not moving* between tiles to offset the effect of people moving, but we highlight two measurement issues: 1) censoring, by which we mean the underreporting of movements; and 2) omitted variables bias, by which we mean that non-movement during the period we study was correlated with pandemic-related economic shutdowns, which affected other pollution sources. To minimize the impact of these issues, we use subsamples of the data from regions where censoring was less prominent, and from periods when pandemic-related lockdown policies were not in effect in India.

This paper contributes to several strands of literature. First, the mechanisms through which traffic impacts emissions and pollution concentrations are of broad interest. Studies in this line of the literature have thus far tended to focus on single urban environments, such as Seattle (Xiang et al., 2020) and the Danish cities of Copenhagen and Roskilde (Khan et al., 2020); Fu et al.,(2020) simulate the dispersion of NO_x from traffic emissions in Baoding, Hebei, China. Misra

et al. (2013) provide an example of an integrated-modeling approach to understand the impact of urban traffic emissions in Canada. (See Colvile et al. (2001) for an overview of this literature.)

Second, this paper contributes to the literature that studies the impacts of various policy changes on commuting behavior. Infrastructure and public transportation affect commuter behavior and, thus, pollution. Recent empirical work shows that the density of subway-network coverage has a significant impact on pollution in China (Li et al., 2019), and that expansions in regional train service in Germany reduced carbon monoxide and nitrogen dioxide (Lalive et al., 2018). Tolls and congestion charges offer alternative policies to combat pollution from traffic because an increase in the cost of driving should induce substitution toward public transit or change the timing of trips during the day. The literature documents that temporary bans on car trips reduced pollution in Santiago, Chile (Rivera, 2021), and that temporary driving restrictions in China led to a significant reduction in key criteria pollutants (Han et al., 2020).¹ Leveraging the waiver of tolls during a National Day holiday in 2012 in China, Fu and Gu (2017) find that the waiver led to pronounced increases in pollution.

Third, the paper contributes to recent studies in economics that have leveraged changes in pollution from transportation to study health and mortality outcomes. For example, Anderson (2020) shows that changes in wind direction from highway traffic impact mortality rates in Los Angeles. Knittel et al. (2016) use an instrumental variable strategy to study the effect of particulate matter from traffic congestion on infant mortality and health in California, finding large effects on infant mortality. Currie and Walker (2011) leverage changes in congestion in New Jersey and Pennsylvania from the rollout of an electronic toll collection system (EZ-Pass) to study the impact of transportation emissions on infant health, also finding significant impacts.

¹ There are also analytical models of urban environments that integrate traffic and pollution from mobile and point sources (e.g., Kyriakopoulou and Picard, 2021).

Zhou et al. (2010) study the impact of driving restrictions on pollution in the lead-up to the 2008 Olympics in Beijing, documenting steep reductions of CO and NO_x emissions that result. Recent studies have also leveraged the COVID-19 lockdown policies to examine effects on ambient pollution (e.g., Cole et al., 2020), but these studies measure the net impacts of a lockdown on air pollution, not the relationship between mobility and pollution.

Fourth, the paper adds to the increasing number of studies that use remote-sensing data to estimate ambient pollution concentrations with high spatial resolution (Shen et al., 2021; Demetillo et al., 2020; Hopkins et al., 2016; Jiang et al., 2016), or to show the long-term patterns of air pollution over a large geographic area (Li et al., 2020). Akbar et al. (2018) provide a useful descriptive example of *city-level* patterns of mobility and congestion from India. While there are descriptive studies of pollution patterns over time and space, few studies incorporate mobility or traffic data on a large scale. To the best of our knowledge, we are the first to isolate the impact of ground-level mobility on air pollution.

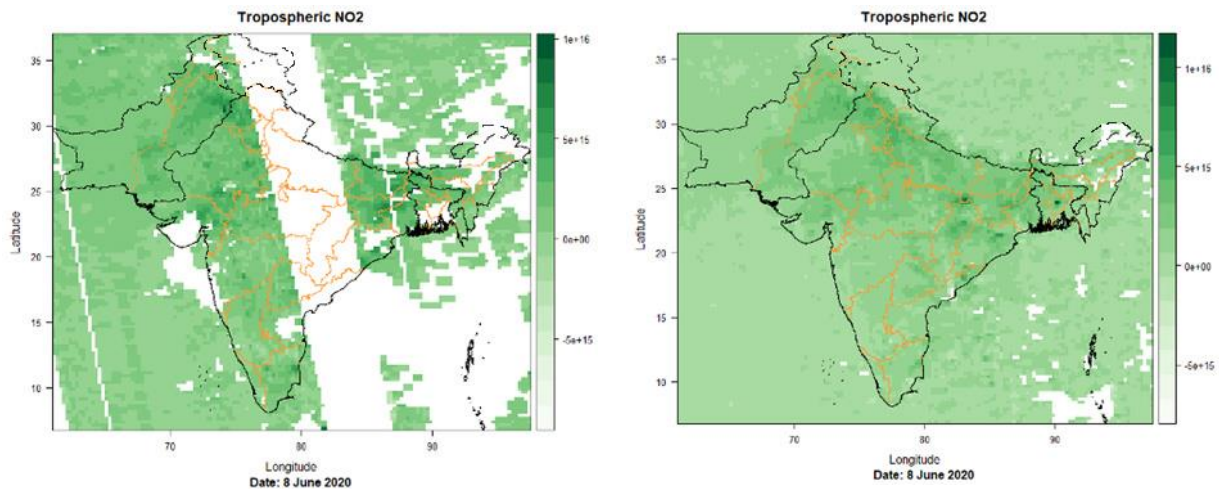
We proceed as follows: Section 2 describes the data used in the analysis. Section 3 introduces our empirical approach and discusses the results. Section 4 offers conclusions and suggestions for future work.

2. Data Description

NO₂ is an emissions byproduct of combustion from cars, trucks, power plants, and industrial facilities. We are interested in tropospheric NO₂, which is sensed remotely by several satellites maintained by the National Aeronautics and Space Administration (NASA) and the European Space Agency. We rely on NASA's AURA Ozone Monitoring Instrument (OMI) Level-3 product; this has a spatial resolution of roughly 25x25 km, and it records daily observations of tropospheric column densities (Krotkov et al, 2019). This product was selected

primarily for computational ease at the geographic scale we study; pre-processing has been completed by NASA and the resulting pixel-cell size is relatively compact. The satellite passes over the Indian subcontinent in the early afternoon, giving daytime estimates of NO₂ molecules in each tropospheric column.

Figure 1. Comparisons of Raw and Cleaned Daily NO₂ Estimates



Notes: Left panel: raw NO₂ satellite data from June 8, 2020. Right panel: pixel-level moving average taken over the preceding 14-day period to smooth missing observations.

Pixel-level satellite observations are often missing, or they are flagged as unusable due to cloud cover, mechanical obstructions, or variations in satellite orbital paths (Duncan et al., 2014). Following best practices and in an appeal to parsimony, we address the missing observations by relying on a backwards-looking, 14-day moving average taken at the pixel level, as shown in Figure 1. This adaptation smooths noise due to missing observations, while allowing us to maintain an analysis at the daily level.

For our mobility data, we rely on Movement Maps (MM) provided by Meta (formerly Facebook). These are particularly well suited to our analysis. The data include the number of users of Meta’s suite of mobile-phone apps (Facebook, Messenger, Instagram, WhatsApp).

Meta’s raw MM data compile aggregate counts of cell-phone movements between 16x16km Bing tiles. We restrict the data sample to include only Bing tiles that fall within India’s national borders. For a given starting tile, we observe the daily count of movement flows to every other Bing tile that occurred in the eight-hour window between 5:30am (0:00 UTC) and 1:30pm (8:00 UTC). This afternoon timing coincides with the OMI satellite’s daily NO₂ readings. Thus, we can link morning inter-tile population movements with contemporaneous tropospheric nitrogen dioxide levels.

Moreover, the relatively fine spatial dimensions of the MM data permit us to map mobility statistics onto individual NO₂ pixel cells and to conduct analysis at that level. Most publicly available mobility data sets (e.g., Google Community Mobility, Apple Mobility Trends) provide only summary data at far larger spatial scales, such as the district or metropolitan-area levels.

The primary mobility measures we generate for our analysis are simple in nature, and we calculate the measurements at the NO₂ “pixel” level ($N \approx 4,000$ for each day in the analysis). The first measure is the sum of all *non-movers*, which we define as those people in a given NO₂ pixel cell whose mobile phones remained in the same Bing tile at the beginning and ending times of day that we examine. The second measure is a count of all *movers* associated with a pixel; this includes any phone that starts in one Bing tile but ends in another. This measure is a rough daily proxy for total population mobility. To conduct heterogeneity analyses, we also decompose this movement measure by distance. We separately sum mover counts by distance traveled: 0-20km, 20-50km, 50-100km, or greater than 100km.

The mobility data provide counts of movement between tile pairs for all tile pairs (i,j). The starting and ending tiles are determined by the location of an individual’s mobile phone at

the beginning and ending windows of time. The resulting movement matrix is analogous to a source-receptor matrix, with counts of people moving between all pairs (i,j) .

The diagonal elements (j,j) of the matrix represent individuals we call non-movers. That is, any individual appearing in some tile j in the beginning and ending period will be counted as a non-mover for tile j . Due to the construction of the data set, we can only count individuals moving *between* tiles; an individual may be moving about *within* a tile during the day, but that individual would be counted as a non-mover. We note that this movement would be at most roughly 23 kilometers. Our specifications include spatial-unit fixed effects (or interactions with day-of-week indicators), so, to the extent that this movement is constant or cyclical within a week, it will be absorbed by our fixed effects.

In some cases, the data are censored. That is, if fewer than ten people move between a tile pair (i,j) , that pair is coded as zero. This censoring leads to an underestimate of movement in certain cases, such as those tile pairs that are over long distances and/or are in very rural areas. Though this bias is likely smaller in urban areas where population densities are higher, there are more people moving between tile pairs than our estimates would indicate.

Another caveat is that the data are only available from late March 2020; this is because the data were compiled and maintained in response to the COVID-19 epidemic. The pandemic induced significant variation in mobility and electricity demand during this period, which we leverage to study the impact on NO_2 concentrations.

To map the mobility pixels to the pollution pixels, we first calculate the total movement between any two tile pairs i and j . To capture the total movement associated with tile i , we sum

over j . In cases in which the pixels do not line up, we assign movement proportionally based on overlapping area.²

Finally, we also use a rich set of electricity generation data for the entirety of India. Because NO₂ emissions are produced both *directly* by coal- and gas-powered electricity generation and *indirectly* through industrial processes that also use this generation, we use these data primarily to control for non-transportation-related nitrogen oxides in the air.³ In their raw form, our electricity data provide daily total megawatts generated from 2013 to 2021 by 533 facilities across the country. We merge the daily generation numbers from each facility with its geographic coordinates and fuel type, and then spatially join facilities to the NO₂ -pixel-cell boundaries.⁴

We control for changes in wind patterns for each thermal generator. We calculate a downwind buffer area with radius 100 km for each stack location. For example, Figure 2a shows the average prevailing wind for January 2020, and Figure 2b illustrates the direction for June 2020. Following Barrows, Garg and Jha (2019), we make the simplifying assumption that “downwind” areas can be characterized as a quadrant (i.e., a 90-degree slice) centered on the predominant monthly wind direction away from the plant.

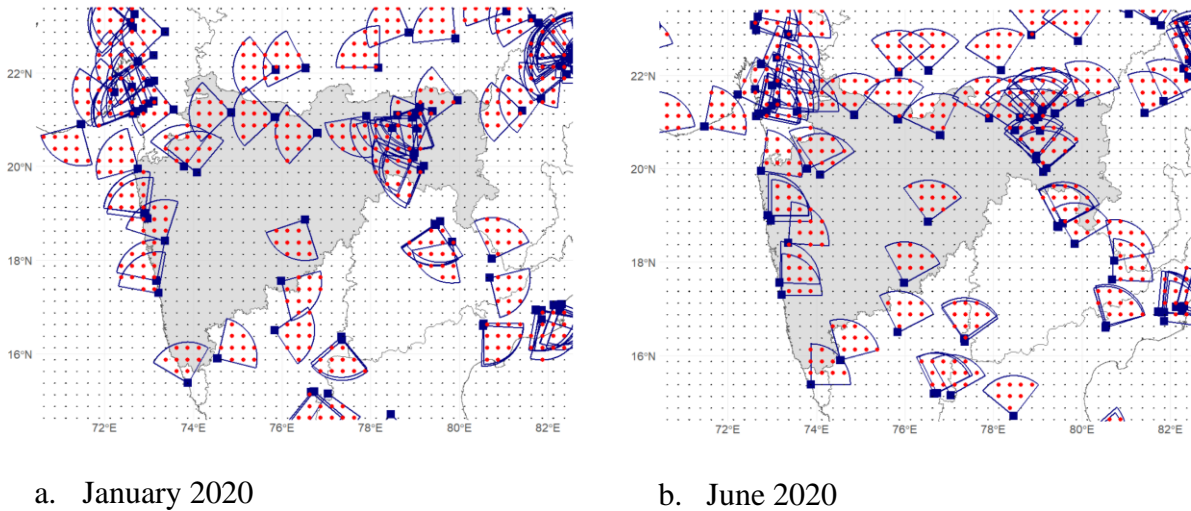
² As a simplified example, suppose that two pollution pixels overlap with tile i . We assign the movement associated with tile i to the two pollution pixels based on the how much overlap there is with the two pollution pixels. If 75 percent of the area of tile i overlaps with the first pollution pixel, we assign 75 percent of the total movement of i to that pollution pixel and 25 percent to the other.

³ Many industrial facilities in India also have onsite captive generators to ease production constraints when there are blackouts or supply shortages. We have no simple way to measure use of these backups.

⁴ Captive generation is not reported, and so it is missing from our analysis. Because buying electricity from the grid is less expensive than captive generation, it is likely the case that there was little captive generation during the period of study because reported shortages were rare.

Wind direction data come from the fifth generation (ERA5) atmospheric reanalysis of climate data produced by the Copernicus Climate Change Service (3CS) at the European Centre for Medium-Range Weather Forecasts (ECMWF).⁵ (See Hersbach et al. (2019) for further details.)

Figure 2. Defining Downwind Regions for Thermal Power Plants



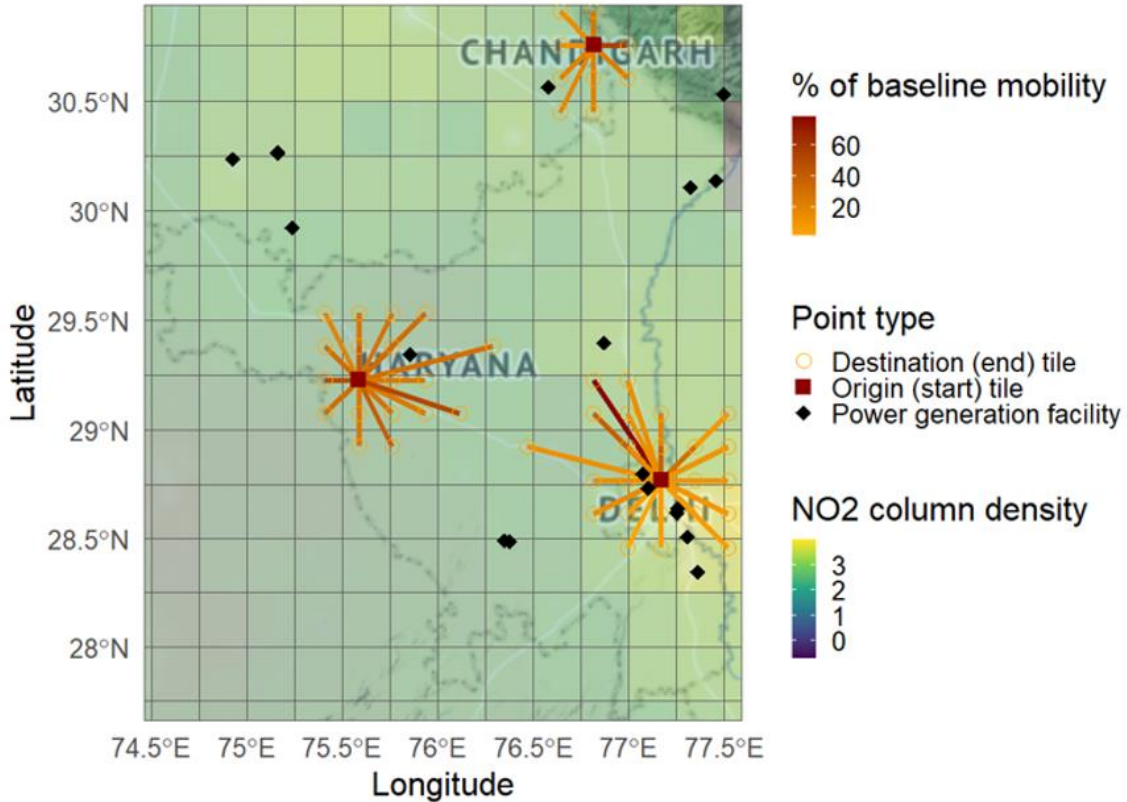
Notes: Thermal electric-generating units in and around Maharashtra are shown as blue squares. For each stack location we then calculate a downwind buffer area with radius 100 km. Figure a. shows the average prevailing wind for January 2020, and b. illustrates the direction for June 2020. Following Barrows, Garg and Jha (2019), we make the simplifying assumption that “downwind” areas can be characterized as a quadrant (i.e.. a 90-degree slice) centered on the predominant monthly wind direction away from the plant. Wind direction data come from the fifth generation (ERA5) atmospheric reanalysis of climate data produced by the Copernicus Climate Change Service (3CS) at the European Centre for Medium-Range Weather Forecasts (ECMWF) (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview>); see Hersbach et al. (2019) for further details.

⁵ <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview>

We conclude by overlaying these data sets in a single visualization. In Figure 3, we illustrate each component of our data on April 16, 2020. The map is situated over northern India; it includes the cities of New Delhi and Chandigarh, as well the neighboring states of Haryana and eastern Punjab. The gridded background coloring illustrates the NO₂ pixel-cell raster. Relatively higher NO₂ concentrations are shown in yellow, while lower concentrations are shown in purple/red. The black diamond points are power plant locations. The orange-to-red points and lines visualize a selected subset of the movement data. To minimize visual clutter, we illustrate these data for three start tiles: one each in New Delhi, Chandigarh, and Haryana. These are marked with the red squares.

For each of these three origin tiles, flowlines are drawn to all local tiles (<100km) towards which we observe at least ten movements during morning hours. The color of the lines between start and end tiles differentiates mobility declines relative to a baseline established in January and February 2020. Darker red lines show a relatively soft decline during this early period of the COVID-19 pandemic, while the brightest orange lines denote that mobility fell to nearly 20 percent of baseline levels on these date. We do this for illustration purposes, but our empirical specifications do not define movement (or non-movement) relative to a baseline; instead these specifications rely on within-pixel deviations from the mean for that sample period.

Figure 3. Overlaying NO₂ Data with Plant-level Generation and Granular Mobility Data



3. Empirical Analysis

Our goal is to identify the impact of ground-level transportation on ambient NO₂ concentrations. We leverage variation in electricity demand and mobility due to COVID-19-related restrictions during 2020-21 and use fixed effects to obtain “within-pixel” or “within-pixel-by-day-of-week” variation over time.

We define an observation at the pixel-by-day level, where a pixel is defined by the gridded NO₂ data. In addition to the fixed-effects approach described above, we also use a two-step approach, in which we first estimate a model of NO₂ concentrations as a function of power

generation and a suite of fixed effects and local trends for the pre-2020 period. We use the model to predict (out of sample) NO₂ concentrations in 2020 and 2021, and we then use the residual as a new dependent variable that is regressed on measures of mobility. This allows us to leverage the longer panel structure of the power generation and NO₂ data.

3.1 Empirical Approach

We are interested in estimating the impact of transportation on atmospheric concentrations of NO₂ using tile-to-tile movements as a proxy for the level of transportation activities. We also control for the number of individuals *not* moving between tiles in a day. Our baseline regression specification is represented by

$$\begin{aligned} \text{arsinh}(\text{NO2}_{idt}) = & \alpha_1 \text{arsinh}(\text{Generation}_{it}) + \beta_1 \text{arsinh}(\text{MM}_{it}) + \\ & \beta_2 \text{arsinh}(\text{NM}_{it}) + \gamma_{id} + \varepsilon_{ist}, \end{aligned}$$

where $\text{arsinh}(\text{NO2}_{ist})$ is the inverse hyperbolic sine of remotely sensed NO₂ in pixel i , day of the week d and date t . We choose this transformation because it has similar properties as a natural logarithm but is defined at zero. This has the advantage that the coefficients can be interpreted directly as elasticities (see Bellemare and Wichman (2020) for a discussion). On the right-hand side of the equation, Generation_{it} is electricity generation upwind of pixel i on date t , while MM represents the numbers of people moving and NM represents those not moving across tiles in the data. Finally, γ_{id} represents a vector of pixel-by-day-of-week fixed effects, and ε_{ist} is an idiosyncratic error term.

Our preferred specifications utilize day-of-week-by-pixel fixed effects. This allows us to control nonparametrically for pixel-specific fluctuations in pollution throughout the week; it also ensures that we are isolating within-pixel variation in mobility, thermal generation, and NO₂.

Descriptive statistics for the key variables are shown in Table 1. To highlight the role of censoring in the mobility data we also show the proportion of observations that are censored and the conditional mean (i.e., the mean conditional on there being at least 10 individuals observed).

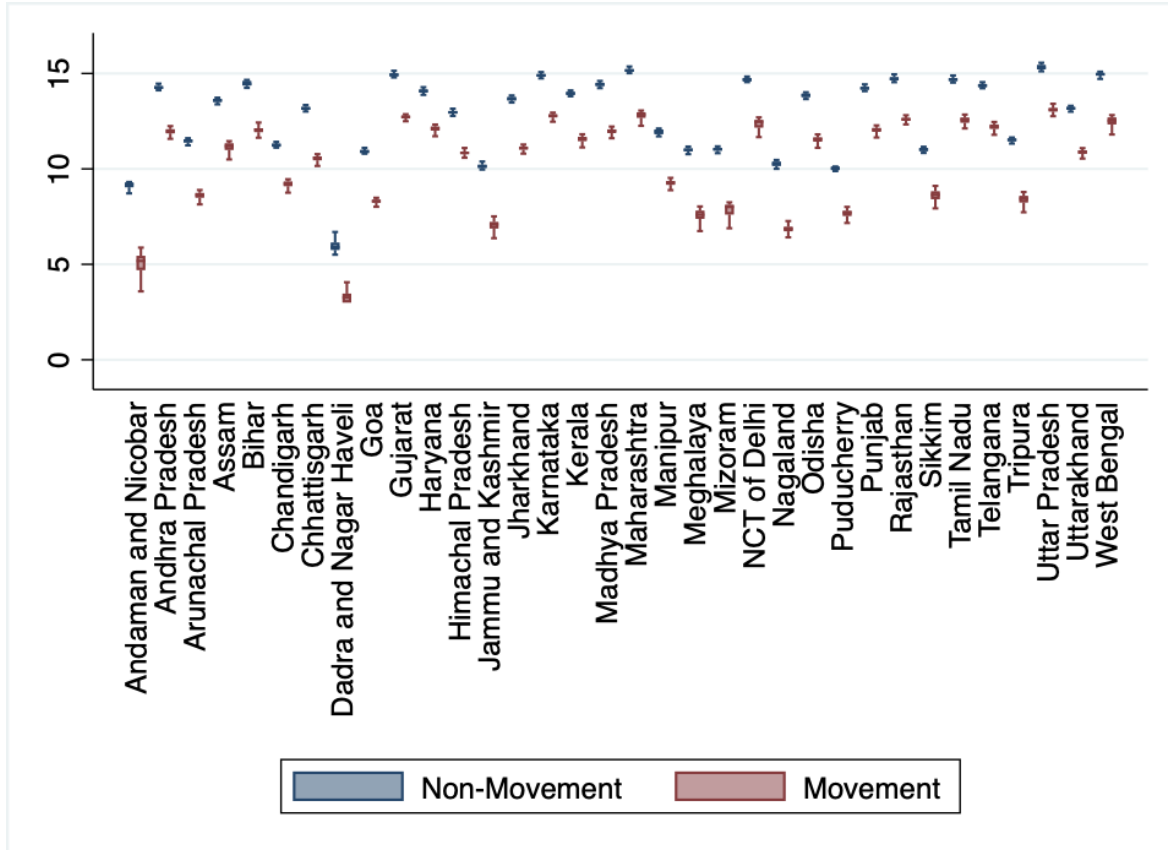
Table 1. NO₂ Pixel-level Means and Standard Deviations

	Mean (Std Dev)	Proportion Censored	Conditional Mean (Uncensored Observations)
NO ₂ (Remotely Sensed)	2.11e+15 (1.37e+15)		
Daily Electricity Generation	0.6412 (5.6998)		
Upwind Elec. Gen. within 100 km	18.7066 (36.1052)		
Individuals Staying in Same Tile	5167.513 (21026.08)	0.00000135 (0.0012)	5167.513 (21026.08)
Individual Movement btw Tiles 0-20 km	361.7216 (1391.589)	0.1408 (0.3478)	421.0031 (1492.961)
Individual Mvmt btw Tiles 20-50 km	120.8068 (477.0327)	0.3370 (0.4727)	182.2031 (576.2153)
Individual Mvmt btw Tiles 50-100 km	24.4723 (102.5978)	0.7002 (0.4582)	81.6179 (174.4774)
Individual Mvmt btw Tiles >100 km	10.1020 (73.5048)	0.8863 (0.3174)	88.8743 (201.3282)

Notes: Authors' calculations using the datasets described in this paper and including all pixel-level data from 2020-2021 (March 2020 – April 2021). There are 1,483,558 pixel-by-day observations.

The average number of individuals in a pixel who do not leave the Bing tile on a given day is 5,168, with a standard deviation nearly 4 times the mean. Figure 4 illustrates these differences in mobility across the states of India.

Figure 4. Variation in Mobility and Non-mobility Counts Across States



Notes: Variability in total daily counts of movement and nonmovement (transformed by the inverse hyperbolic sine, *arsinh*) by state. The figure illustrates each state’s median, interquartile range and adjacent values. Adjacent values for a variable x are defined by the following. Define U as $x_{[75]} + 2^3(x_{[75]} - x_{[25]})$. The upper adjacent value is defined as x_i , such that $x_{(i)} \leq U$ and $x_{(i+1)} > U$. Define L as $x_{[25]} - 2^3(x_{[75]} - x_{[25]})$. The lower adjacent value is defined as x_i , such that $x_{(i)} \geq L$ and $x_{(i-1)} < L$.

The baseline fixed-effects regression results are shown in tables 2 and 3. As expected, the estimated elasticities vary with the level of fixed effects included in the estimation. In Table 2 results are shown for the impact of mobility on NO_2 . In column (2) of Table 2, a 10 percent increase in mobility leads to an increase in NO_2 of roughly 1.3 percent. By contrast, a 10 percent

increase in *non-movement*, holding all else constant, leads to a *decrease* in NO₂ concentrations of 0.9 percent, but the effect is insignificant at the 5 percent level.

Table 2. Fixed-effects Regressions, Pixel Level

	(1)	(2)	(3)	(4)
arsinh(Total Mvmt btw Cells)	0.114** (0.0107)	0.134** (0.0116)		
arsinh(Mvmt btw Cells, 0-20 km)			0.105** (0.0110)	0.126** (0.0121)
arsinh(Mvmt btw Cells, 20-50 km)			0.0636** (0.00648)	0.0688** (0.00683)
arsinh(Mvmt btw Cells, 50-100 km)			0.0354** (0.00601)	0.0372** (0.00625)
arsinh(Mvmt btw Cells, >100 km)			0.0114 (0.00809)	0.0124 (0.00838)
arsinh(Upwind Gen < 100km)	0.0259** (0.00491)	0.0258** (0.00495)	0.0260** (0.00491)	0.0258** (0.00495)
Pixel FE	X		X	
DOW FE	X		X	
Month-by-Year FE	X	X	X	X
Pixel-by-DOW FE		X		X
Adjusted R-Squared	0.093	0.081	0.093	0.081

The dependent variable is arsinh(NO₂). There are 1,469,328 (pixel-by-day) observations. * and ** represent significance at the 5 percent and 1 percent levels, respectively. The within-pixel standard deviation of “Total Movement” is 0.534.

Table 3. Fixed-effects Regressions, Full Sample, Pixel Level

	(1)	(2)	(3)	(4)
arsinh(Total “Non-Movers”)	-2.371** (0.0330)	-0.0899 (0.0520)	-0.0913 (0.0523)	-0.0947 (0.0534)
arsinh(Upwind Thermal Gen < 100km)	-0.0544** (0.00484)	0.0263** (0.00491)	0.0263** (0.00491)	0.0262** (0.00495)
Pixel FE	X	X	X	
DOW FE	X		X	
Month-by-Year FE		X	X	X
Pixel-by-DOW FE				X
Adjusted R-Squared	0.073	0.093	0.093	0.081

The dependent variable is arsinh(NO₂). There are 1,469,328 (pixel-by-day) observations. * and ** represent significance at the 5 percent and 1 percent levels, respectively. The within-pixel standard deviation of “Non-movers” is 0.169.

As shown in Table 1, censoring is a concern in the mobility measures. The movement between tiles *i* and *j* is not reported if there are fewer than 10 individuals observed. We are also

concerned with mobility being correlated with changes in economic activity driven by the pandemic and related policies. To address these concerns, we now estimate an analogous model for a subsample of the observations without zeros (i.e., where some movement is detected, but where movement between tile pairs may still be unreported). We also restrict the analysis to include observations from June to November 2020, a period when economic activity had resumed and the Delta variant of COVID-19 had not been identified as dominant in India.

Tables 4 and 5 show the results for *movement* and *non-movement*, respectively. The preferred specifications for mobility indicate that the impact of movement on NO₂ is more than unit elastic; a 1 percent increase in movement leads to roughly a 2.5 percent increase in NO₂. *Non-movement* has a larger impact in absolute value; a 1 percent increase in the number of individuals staying home, relative to the mean for that pixel, leads to a 4 percent decrease in NO₂.

Table 4. Pixel-level Regressions: June-Nov. 2020, Dropping Censored Observations

	(1)	(2)	(3)	(4)
arsinh(Total Mvmt btw Cells)	2.411** (0.204)	2.775** (0.223)		
arsinh(Mvmt btw Cells, 0-20 km)			0.199 (0.294)	0.244 (0.334)
arsinh(Mvmt btw Cells, 20-50 km)			0.364* (0.182)	0.323 (0.204)
arsinh(Mvmt btw Cells, 50-100 km)			0.548** (0.0907)	0.672** (0.0999)
arsinh(Mvmt btw Cells, >100 km)			0.365** (0.0714)	0.367** (0.0766)
arsinh(Upwind Gen < 100km)	0.0246 (0.0275)	0.0213 (0.0283)	0.0236 (0.0274)	0.0198 (0.0283)
Pixel FE	X		X	
DOW FE	X		X	
Pixel-by-DOW FE		X		X
Adjusted R-Squared	0.080	0.051	0.081	0.053

The dependent variable is arsinh(NO₂). There are 63,475 (pixel-by-day) observations. * and ** represent significance at the 5 percent and 1 percent levels, respectively. The sample is restricted to June-November 2020, and to observations without censoring.

Table 5. Pixel-level Regressions: June-Nov. 2020, Dropping Censored Observations

	(1)	(2)
arsinh(Total “Non-movers”)	-3.950** (0.0895)	-4.043** (0.0919)
arsinh(Upwind Thermal Gen < 100km)	0.0228* (0.00949)	0.0225* (0.00962)
Pixel FE	X	
DOW FE	X	
Month-by-State		
Pixel-by-DOW FE		X
Adjusted R-Squared	0.106	0.083

The dependent variable is $\text{arsinh}(\text{NO}_2)$. There are 732,066 (pixel-by-day) observations. * and ** represent significance at the 5 percent and 1 percent levels, respectively. The sample is restricted to June-November 2020 and observations without censoring.

3.2 Heterogeneity between States

There is significant heterogeneity in NO_2 levels, movement, and power generation between states. The specifications so far have restricted the elasticities of NO_2 with respect to mobility to be the same across states. To test for heterogeneous responses of pollution concentrations to mobility we estimate the regressions separately by state.

Table 6 shows the results for the three most populous states in India: Uttar Pradesh, Maharashtra and Bihar. As before, we separately estimate the elasticities with respect to total movement and total *non-movement*. The results are largest in magnitude for Maharashtra, with estimated elasticities of 7.5 for movement and -9.7 for non-movement. These estimates suggest that congestion in this highly urbanized state contributes significantly to NO_2 .

Table 6. State-specific Estimates for Three Most Populous States (Pixel-level Specification)

	(1)	(2)	(3)	(4)	(5)	(6)
	U. Pradesh	U. Pradesh	Maharashtra	Maharashtra	Bihar	Bihar
arsinh(Total Mvmt)	1.162** (0.214)		7.511** (0.860)		0.562 (0.313)	
arsinh(Total “Non-movers”)		-0.633** (0.106)		-9.6823** (0.3659)		-0.550** (0.163)
arsinh(Upwind Gen)	-0.00491 (0.0272)	-0.0615** (0.0121)	-0.384** (0.140)	0.0389** (0.0406)	0.0265 (0.0414)	0.0443** (0.0152)
Pixel-by-DOW FE	X	X	X	X	X	X
N	8464	64394	5418	72,532	2893	24839
Adjusted R-squared	-0.006	0.016	0.022	0.0589	0.018	0.018

The dependent variable is asinh(NO₂). Observations are at the pixel-by-day level. * and ** represent significance at the 5 percent and 1 percent levels, respectively. The sample is restricted to June-November 2020, and to observations without censoring.

4. Conclusions

Fossil fuel combustion from transportation is known to cause air pollution, but separately identifying the impact of transportation on NO₂ concentrations has been difficult due to data limitations and confounding factors. Measuring the impact of traffic on ambient air pollution concentrations is difficult, particularly over large areas, because the impact of traffic on pollution varies with vehicle characteristics, traffic congestion, and environmental conditions. There is also a dearth of information on the number of vehicles on the road at any time, and for daily indicators of fuel use at regional levels. We overcome these issues by combining remote-sensing estimates of tropospheric NO₂ in India with data on electricity generation and on the movement of mobile phones. This combination allows us to estimate the responsiveness of tropospheric NO₂ to population movements.

We find that ground-level mobility has a large, significant impact on NO₂ concentrations. The average elasticity for “movement” estimated nationwide is roughly 2.5, meaning that a 1 percent increase in movement translates to a 2.5 percent increase in tropospheric NO₂. We also

demonstrate that there is significant heterogeneity over space. In Maharashtra, for example, the estimated elasticities can be two to three times the average elasticity. Maharashtra is relatively industrialized and has higher incomes than most of India, which could be a driver in this difference.

We also find that NO₂ concentrations are very responsive to within-pixel changes in the number of individuals *not moving long distances during the day*. Because this measure is less contaminated by censoring, the elasticity is larger in magnitude. Our preferred estimates suggest that a 1 percent increase in the number of individuals *not traveling long distances* decreases NO₂ concentrations by about 4 percent.

There are several areas for future research. First, machine learning could help researchers use the rich data available to better understand how the effects vary with other covariates, such as weather, topography, and land use. One could also estimate the effects separately for smaller spatial units, such as neighborhoods or cities, as well as determining how the effects vary with demographic characteristics. Finally, we note that our findings are for NO₂ concentrations, but further work should be done to determine the impact on other pollutants as well as greenhouse gas emissions.

References

- Akbar, P.A., Couture, V., Duranton, G. and Storeygard, A., 2018. *Mobility and congestion in urban India* (No. w25218). National Bureau of Economic Research.
- Anderson, M.L., 2020. As the wind blows: The effects of long-term exposure to air pollution on mortality. *Journal of the European Economic Association*, 18(4), pp.1886-1927.
- Barrows, Geoffrey, Teevrat Garg and Akshaya Jha. 2019. "The Health Costs of Coal-Fired Power Plants in India." IZA Discussion Paper 12838.
- Bellemare, M.F. and Wichman, C.J., 2020. Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics* 82(1), pp.50-61.

- Cole, M.A., Elliott, R.J. and Liu, B., 2020. The impact of the Wuhan Covid-19 lockdown on air pollution and health: a machine learning and augmented synthetic control approach. *Environmental and Resource Economics*, 76(4), pp.553-580.
- Colville, R.N., Hutchinson, E.J., Mindell, J.S. and Warren, R.F., 2001. The transport sector as a source of air pollution. *Atmospheric environment*, 35(9), pp.1537-1565.
- Currie, J. and Walker, R., 2011. Traffic congestion and infant health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, 3(1), pp.65-90.
- Demetillo, M.A.G., Navarro, A., Knowles, K.K., Fields, K.P., Geddes, J.A., Nowlan, C.R., Janz, S.J., Judd, L.M., Al-Saadi, J., Sun, K. and McDonald, B.C., 2020. Observing nitrogen dioxide air pollution inequality using High-Spatial-Resolution remote sensing measurements in Houston, Texas. *Environmental Science & Technology*, 54(16), pp.9882-9895.
- Duncan, B. N., Prados, A. I., Lamsal, L. N., Liu, Y., Streets, D. G., Gupta, P., ... & Ziemba, L. D. (2014). Satellite data of atmospheric pollution for US air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid. *Atmospheric environment*, 94, 647-662.
- Fu, S. and Gu, Y., 2017. Highway toll and air pollution: Evidence from Chinese cities. *Journal of Environmental Economics and Management*, 83, pp.32-49.
- Fu, X., Xiang, S., Liu, Y., Liu, J., Yu, J., Mauzerall, D.L. and Tao, S., 2020. High-resolution simulation of local traffic-related NO_x dispersion and distribution in a complex urban terrain. *Environmental Pollution*, 263, p.114390.
- Guo, J.X., Zeng, Y., Zhu, K. and Tan, X., 2021. Vehicle mix evaluation in Beijing's passenger-car sector: From air pollution control perspective. *Science of the Total Environment*, 785, p.147264.
- Han, Q., Liu, Y. and Lu, Z., 2020. Temporary driving restrictions, air pollution, and contemporaneous health: Evidence from China. *Regional Science and Urban Economics*, 84, p.103572.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N. (2019): ERA5 monthly averaged data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store. 10.24381/cds.f17050d7
- Heyes, A. and Zhu, M., 2019. Air pollution as a cause of sleeplessness: Social media evidence from a panel of Chinese cities. *Journal of Environmental Economics and Management*, 98, p.102247.
- Hopkins, F.M., Kort, E.A., Bush, S.E., Ehleringer, J.R., Lai, C.T., Blake, D.R. and Randerson, J.T., 2016. Spatial patterns and source attribution of urban methane in the Los Angeles Basin. *Journal of Geophysical Research: Atmospheres*, 121(5), pp.2490-2507.

- Khan, J., Kakosimos, K., Jensen, S.S., Hertel, O., Sørensen, M., Gulliver, J. and Ketzel, M., 2020. The spatial relationship between traffic-related air pollution and noise in two Danish cities: Implications for health-related studies. *Science of the Total Environment*, 726, p.138577.
- Krotkov, N.A., Lamsal, L.N., Marchenko, S.V., Celarier, E.A., Bucsela, E.J., Swartz, W.H., Joiner, J. and the OMI core team (2019). OMI/Aura NO₂ Cloud-Screened Total and Tropospheric Column L3 Global Gridded 0.25 degree x 0.25 degree V3, NASA Goddard Space Flight Center, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 2020-2021. [10.5067/Aura/OMI/DATA3007](https://doi.org/10.5067/Aura/OMI/DATA3007)
- Kyriakopoulou, E. and Picard, P.M., 2021. On the design of sustainable cities: Local traffic pollution and urban structure. *Journal of Environmental Economics and Management*, 107, p.102443.
- Lalive, R., Luechinger, S. and Schmutzler, A., 2018. Does expanding regional train service reduce air pollution?. *Journal of Environmental Economics and Management*, 92, pp.744-764.
- Li, S., Liu, Y., Purevjav, A.O. and Yang, L., 2019. Does subway expansion improve air quality?. *Journal of Environmental Economics and Management*, 96, pp.213-235.
- Li, J., 2020. Pollution trends in China from 2000 to 2017: A multi-sensor view from space. *Remote Sensing*, 12(2), p.208.
- Jiang, J., Zhang, J., Zhang, Y., Zhang, C. and Tian, G., 2016. Estimating nitrogen oxides emissions at city scale in China with a nightlight remote sensing model. *Science of the Total Environment*, 544, pp.1119-1127.
- Knittel, C.R., Miller, D.L. and Sanders, N.J., 2016. Caution, drivers! Children present: Traffic, pollution, and infant health. *Review of Economics and Statistics*, 98(2), pp.350-366.
- Misra, A., Roorda, M.J. and MacLean, H.L., 2013. An integrated modelling approach to estimate urban traffic emissions. *Atmospheric Environment*, 73, pp.81-91.
- Rigobon, Roberto, and Thomas M. Stoker. "Estimation with censored regressors: Basic issues." *International Economic Review* 48, no. 4 (2007): 1441-1467.
- Rivera, N.M., 2021. Air quality warnings and temporary driving bans: Evidence from air pollution, car trips, and mass-transit ridership in Santiago. *Journal of Environmental Economics and Management*, 108, p.102454.
- Shen, Y., Jiang, C., Chan, K.L., Hu, C. and Yao, L., 2021. Estimation of Field-Level NO_x Emissions from Crop Residue Burning Using Remote Sensing Data: A Case Study in Hubei, China. *Remote Sensing*, 13(3), p.404.

- Silva, R.A., Adelman, Z., Fry, M.M. and West, J.J., 2016. The impact of individual anthropogenic emissions sectors on the global burden of human mortality due to ambient air pollution. *Environmental health perspectives*, 124(11), pp.1776-1784.
- Streets, D.G., Canty, T., Carmichael, G.R., de Foy, B., Dickerson, R.R., Duncan, B.N., Edwards, D.P., Haynes, J.A., Henze, D.K., Houyoux, M.R. and Jacob, D.J., 2013. Emissions estimation from satellite retrievals: A review of current capability. *Atmospheric Environment*, 77, pp.1011-1042.
- Xiang, J., Austin, E., Gould, T., Larson, T., Shirai, J., Liu, Y., Marshall, J. and Seto, E., 2020. Impacts of the COVID-19 responses on traffic-related air pollution in a Northwestern US city. *Science of the Total Environment*, 747, p.141325.
- Zhou, Y., Wu, Y., Yang, L., Fu, L., He, K., Wang, S., Hao, J., Chen, J. and Li, C., 2010. The impact of transportation control measures on emission reductions during the 2008 Olympic Games in Beijing, China. *Atmospheric Environment*, 44(3), pp.285-293.

Appendix: Leveraging a Longer Data Series

The results so far have been based on data beginning in 2020, when the mobility data first became available. In this subsection we leverage the longer panel of electricity generation and NO₂ concentrations. We first estimate a high-dimensional fixed-effects model using data from 2013 to 2019. We include upwind power generation, day-of-week-by-pixel fixed effects, state-by-month fixed effects, and district-specific linear trends. Table 7 shows the results. The model explains a large share of the variation in NO₂. We then use the model to predict out-of-sample NO₂ concentrations in over the 2020-21 period, and the predictions are subtracted from the actual NO₂ concentrations to generate residuals. The average residuals for each state are plotted over time in Figure 1A.

We then use the residuals at the pixel-by-day level as the dependent variable in a regression model, again restricting the sample to observations with reduced censoring as in Tables 5 and 6. The results are shown in Table 8. The results are similar in spirit, though with slightly smaller magnitudes.

Table 7. Predictive **Model, Pixel-Level**

Electric Generation	0.0217** (0.000616)
Gen. within 100 km	0.00976** (0.000185)
Gen. 100 - 200 km	0.00569** (0.000157)
Pixel-by-DOW FE	X
State-by-Month FE	X
District-Specific Linear Trend	X
Adj. R-sq	0.350

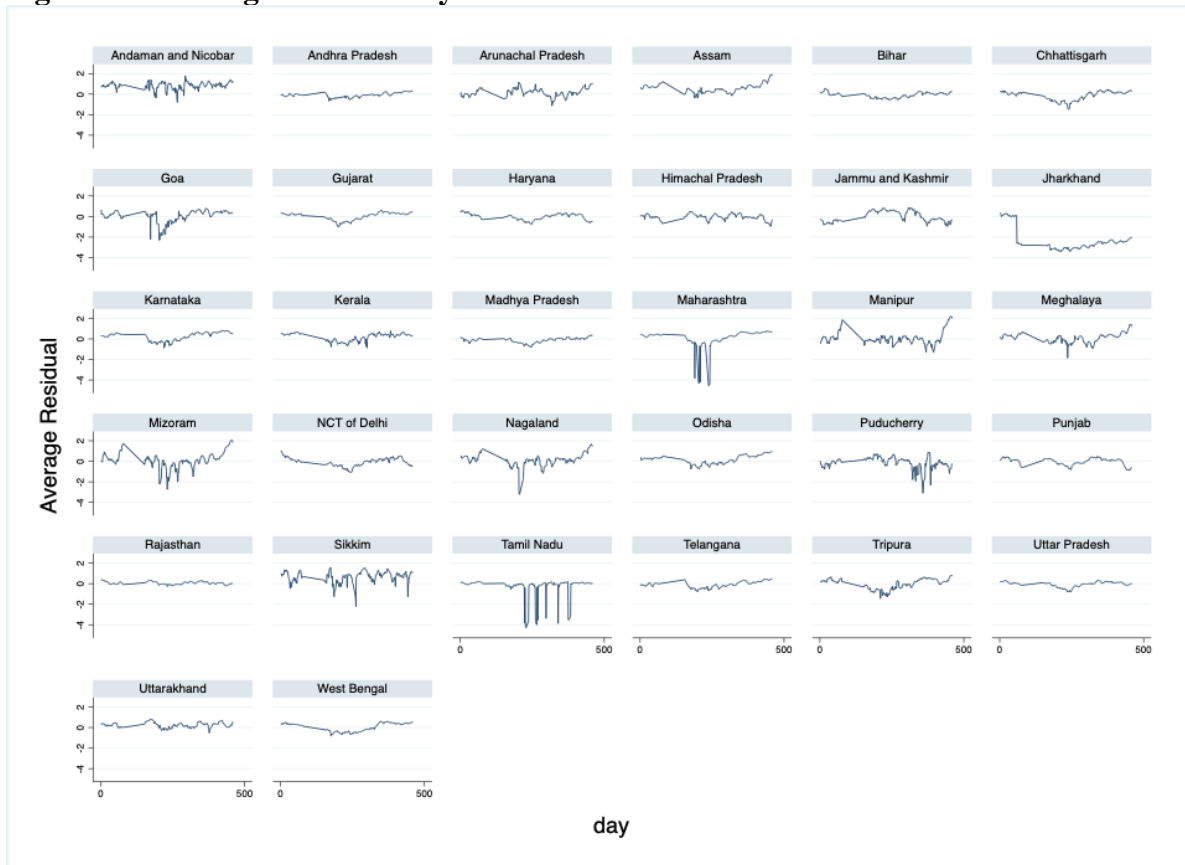
The dependent variable is $\text{arsinh}(\text{NO}_2)$. There are 9,780,392 (pixel-by-day) observations. * and ** represent significance at the 5 percent and 1 percent levels, respectively.

Table 8. Explaining Unexpected NO₂ Fluctuations: Out-of-Sample Residuals (Pixel Level)

	(1)	(2)			(3)	(4)
arsinh("Non-movers")	-1.684** (0.585)	-1.722** (0.612)				
arsinh(Total Mvmt)			0.147** (0.0478)	0.189** (0.0602)		
arsinh(Mvmt 0-20 km)					0.0927* (0.0399)	0.126* (0.0507)
arsinh(Mvmt 20-50 km)					0.0794** (0.0286)	0.0979** (0.0326)
arsinh(Mvmt 50-100 km)					0.0828 (0.0465)	0.0924 (0.0509)
arsinh(Mvmt >100 km)					0.0674 (0.0557)	0.0723 (0.0578)
Pixel FE	X		X		X	
Pixel-by-DOW FE		X		X		X
Adj. R-sq	0.997	0.997	0.997	0.997	0.997	0.997

The dependent variable is the residual from the predictive model described above. There are 732,067 (pixel-by-day) observations. * and ** represent significance at the 5 percent and 1 percent levels, respectively.

Figure 1A. Average Residuals by State and Date



Note: The plots show average residuals at the state-by-day level. Residuals were calculated by taking the actual $\text{arsinh}(\text{NO}_2)$ levels minus out-of-sample predictions from model (3) of Table 5. For graphical exposition some outliers were dropped from this plot (those with an absolute value of greater than five).