Kim, Seung Min; Gillingham, Kenneth

Working Paper

Air Pollution and Solar Energy: Evidence from Wildfires

CESifo Working Paper, No. 10948

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Kim, Seung Min; Gillingham, Kenneth (2024) : Air Pollution and Solar Energy: Evidence from Wildfires, CESifo Working Paper, No. 10948, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/296037

CESifo WORKING PAPERS

10948
2024

February 2024

Air Pollution and Solar Energy: Evidence from Wildfires

Seung Min Kim, Kenneth Gillingham

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10948*

Air Pollution and Solar Energy: Evidence from Wildfires

Abstract

There is strong evidence on the mortality and morbidity external costs of air pollution. This paper focuses on another source of air pollution externality: the loss in solar electricity generation from increased atmospheric opacity due to air pollution. We use data from residential rooftop solar panels and exploit quasi-random variation in air opacity from wildfire smoke and wind conditions to estimate that PM2.5 pollution reduced annual solar generation in California by over 450GWh, or 4% of total potential generation. This is equivalent to an annual externality of $177 per ton of PM2.5 emitted. Our findings suggest that there is a positive feedback from decarbonisation policies, where reduced dependence on fossil fuels improves air quality, leading to an even greater productivity of solar generation, further reducing emissions.

JEL-Codes: Q480.

Keywords: solar energy, air pollution, equity.

Seung Min Kim
Columbia University
420 West 118th Street
USA – New York, NY 10027
sk5316@columbia.edu

Kenneth Gillingham
Yale University
195 Prospect Street
USA – New Haven, CT 06511
kenneth.gillingham@yale.edu

February 6, 2024
The authors would like to thank Kai Chen, Luming Chen, Benjamin Krebs, Robert Mendelsohn, Oswald Schmitz, Jasjeet Sekhon, Eugene Tan, and participants of Yale School of the Environment's Confluence Seminar and Northeast Workshop on Energy Policy and Environmental Economics for their helpful comments. We are also grateful to the two anonymous referees and the editor of this article for insightful comments that greatly improved the paper. All remaining errors are ours.

1 Introduction

From the Los Angeles smog to the post-*Diwali* haze in New Delhi, acute pollution events often highlight the impact of air pollution on visibility. While the salience of such incidents has led to empirical studies on their health impacts (e.g., Garaga and Kota, 2018), little attention has been given to the impact of this air quality deterioration on solar generation. In stark contrast to this absence, small-scale field experiments and engineering simulations have consistently found evidence of impacts from air pollution on solar photovoltaic (PV) panel productivity (Bergin et al., 2017; Peters et al., 2018). These studies have found that air pollution, such as fine particulate matter (PM2.5) pollution, increases atmospheric opacity, which in turn reduces the amount of solar energy that reaches the surface.

The implications of such an externality of air pollution may be significant, as many federal and state-level energy transition strategies in the United States involve the expansion of solar generation. Such increases in solar energy adoption are expected to yield large reductions in ambient air pollution (Deetjen and Azevedo, 2020; Sergi et al., 2020; Gillingham and Huang, 2020). The possible link between air pollution and solar generation hence suggests a positive feedback from decarbonization policies. Lower dependence on fossil fuels could improve atmospheric visibility, leading to higher solar productivity, thus resulting in an additional increase of renewable energy share in the fuel mix and further reductions in emissions.

This study estimates the causal impact of air pollution on residential rooftop solar generation. We assess the contemporaneous effect of PM2.5 pollution on metered rooftop solar generation using day-to-day variations in wind conditions (Knittel et al., 2016; Schlenker and Walker, 2016) and wildfire smoke influx. The paper takes a two-part approach. We first establish that atmospheric opacity, measured from spaceborne satellites, is causally linked with rooftop solar generation in Californian households over a seven-year period. We then show that this air opacity is also linked with PM2.5 pollution, using plausibly exogenous day-to-day variations in PM2.5 conditions. This two-part approach is motivated by the fact that PM2.5 is but one of several pollutants that is affected by wildfire smoke that can then affect solar generation[1], so directly linking PM2.5 to solar generation could miss an

[1] Borgschulte et al. (2022) also provides insightful discussion on the existence of other pollutants correlated with smoke exposure.

important part of the effect of air pollution on solar generation.

We find that a one standard deviation increase in air opacity decreases the average solar generation in that day and the following two days by 9.4%.[2] Our empirical approach uses quasi-experimental variation in wind conditions and smoke influx as instruments for air opacity and PM2.5 air pollution in the two stages of our estimation respectively, addressing the possibility of endogeneity from reverse causality and measurement error. The two-stage least squares (2SLS) estimation results are consistently larger than OLS estimates, suggesting that endogeneity is a real concern in this setting. Using PM2.5 concentration as the endogenous variable, rather than air opacity, underestimates the effect. This is consistent with the hypothesis that non-PM2.5 aerosols also reduce solar generation, strongly suggesting the violation of exclusion restriction if we directly model the relationship between PM2.5 and solar generation.

We also find that PM2.5 pollution is causally linked with contemporaneous air opacity. A $1\mu g/m^3$ increase in PM2.5 pollution leads to a 2% increase in air opacity on average, while summertime pollution is particularly nefarious, with a 3% increase in opacity. This empirical finding has significant implications, especially in the context of California's recent growth in the adoption of residential rooftop solar. Our back-of-the-envelope calculations reveal that as of 2021, air opacity and PM2.5 reduced potential solar generation from residential rooftops by 1,616GWh and 459GWh, or 14.8% and 4.2% of potential generation respectively. We estimate this previously unaccounted externality of PM2.5 pollution is \$177/ton/year.

This paper contributes to the existing empirical literature in three useful ways. First, we add to a rich literature that documents the effect of air pollution on such diverse outcomes as health (Chay and Greenstone, 2003; Moretti and Neidell, 2011; Schlenker and Walker, 2016), crop yield (Metaxoglou and Smith, 2020; Proctor, 2021), and human capital formation (Chen et al., 2018). While our analysis is confined to ZIP codes within California, the impact of air pollution on solar generation will likely be significantly larger elsewhere in the world with higher pollution levels (e.g., developing countries, as in Landrigan et al., 2018). Our findings identify an added benefit of air pollution mitigation policies, an externality whose size will

[2]We use three-day solar generation to account for lagged effects due to the possibility of deposition on the solar panels (Deryugina et al., 2019).

only grow with the expansion of solar energy.

Second, we make methodological contributions to the empirical literature that uses quasi-random variation in air pollution levels to assess the impact of air pollution on economic outcomes (Moretti and Neidell, 2011; Schlenker and Walker, 2016; Knittel et al., 2016). The empirical strategy in these studies relies on atmospheric conditions (e.g., wind conditions) and the interaction of these conditions with pollution-generating local economic activities (e.g, local traffic level). In contrast to such local sources of pollution, wildfire smoke in California originates from distant sources (McClure and Jaffe, 2018). Such spatial exogeneity has the potential to improve our identification strategy, which relies on the interaction of smoke influx with wind direction. To the best of our knowledge, this is the first empirical study to use wildfire smoke as a quasi-experimental source of heterogeneity in wind condition impact on air quality.[3]

Third, we add to a very limited literature that studies the impact of air pollution on solar generation. Studies on the subject have been dominated by small-scale field experiments and modeling studies. For example, Bergin et al. (2017) studies the global impact of PM2.5 on solar generation using outdoor experiments in Ahmedabad, India, lasting less than six months in total. Existing efforts to confirm such results in realistic environments were largely limited to anecdotal comparisons around a single high-pollution event (e.g., prescribed burns, as in Gómez-Amo et al., 2019) or OLS regressions without consideration of endogeneity (Son et al., 2020). Our analysis covers seven years of data from more than 1,800 ZIP codes in California and takes potential endogeneity concerns seriously.

A recent study using data from Arizona directly examines the effect of PM2.5 and PM10 pollution on residential and commercial rooftop solar generation using wind direction as the instrumental variable for PM2.5 (He et al., 2020). Our study is similar to He et al. (2020) in its use of quasi-experimental variation in PM2.5 to study solar generation. Our study, however, differs from He et al. (2020) in three important ways. First, we use satellite-measured air opacity rather than the PM2.5 (or PM10) concentration as the endogenous

[3]In the usage of wildfire smoke as a driver of local air conditions, we build on a recent study on the impact of wildfire smoke on labor market outcomes (Borgschulte et al., 2022). We differ from the study, however, in that we are not interested in the effect of smoke *per se* on our outcomes. We also do not use wildfire smoke as an instrument for PM2.5 concentrations.

variable of interest to better model the causal channel by which air pollution affects solar generation. As alluded to above, the main issue with a 2SLS estimation directly linking PM2.5 to solar generation is that it requires that PM2.5 be the only causal pathway through which wind affects solar generation; this is likely untrue, considering that a plethora of other pollutants can be influenced by wind and also affect atmospheric opacity and reduce solar generation. We avoid this violation of the exclusion restriction assumption through our two-stage approach that also estimates the relationship between air opacity and solar generation.

A second way that our study differs from He et al. (2020) is that the previous study uses a linear model to explain the solar generation-air quality relationship. In our study, we present fundamental principles in atmospheric optic theory (Lambert, 1760; Beer, 1852) that cast doubt on this linearity assumption. The fundamental principles suggest a log-linear relationship between solar generation and air pollution. Indeed, this relationship is supported by evidence from small-scale field experiments (Peters et al., 2018). Put simply, the physics suggest that air pollution would affect the *efficiency* of solar cells instead of the *level* of solar production. It is possible that the misspecification can explain part or all of the remarkable difference in the He et al. (2020) estimate of air pollution impact's on residential versus commercial solar units (30.6% versus 0.15% decrease for a standard deviation increase in PM2.5 concentration).[4] Lastly, while He et al. (2020) relies strictly on wind conditions to identify air pollution impacts, we interact wind condition with wildfire smoke influx to introduce a more robust identification strategy that provides additional statistical power. We find that all three of these differences in methodology have important implications for the estimates of the effect of PM2.5 on solar generation.

The remainder of this article is organized as follows. In the following section, we introduce the definition of aerosol optical depth and its relationship with wind directions, which is central to our empirical strategy and identification assumptions. We then present our data and empirical strategy in Sections 3 and 4. Section 5 summarizes our empirical results. Section 6 discusses the policy implications of our results, and Section 7 concludes.

[4]The average solar generation in residential units in He et al. (2020) was 7.4kWh/day/consumer, compared to 85.5kWh/day/consumer for commercial units. Both point estimates (30.6% and 0.15%) are quite far off estimates from previous engineering-based studies in similar environments (e.g., 6–18%, Bergin et al., 2017).

2 Background

In this section, we highlight key findings from the engineering and physics literature that motivate our research design. We then introduce the intuition behind our identification strategy.

2.1 PM2.5 pollution and solar generation

After entering the atmosphere, radiative energy from the sun is subjected to a process called "optical extinction," where the energy is either scattered or absorbed as it collides with ambient molecules.[5] In the lower atmosphere, the two main drivers of such extinction are: solid- or liquid-state particulates (i.e., aerosols), such as water droplets and PM2.5 molecules, and gas-state molecules with high absorption capacities in select wavelengths, such as nitrogen dioxide (NO_2).[6] An increased concentration of such molecules hence reduces the penetration of solar radiation, leading to atmospheric opacity. Between the two types of opacity-inducing molecules, aerosol particulates (i.e., particulates in liquid and solid states) are the main drivers of visibility reduction, accounting for $50-95\%$ of extinction (Seinfeld and Pandis, 2016).[7]

Such interaction of solar radiation with the atmosphere is dictated by the Beer-Lambert Law, which posits that the rate of optical extinction is a linear function of the incoming energy itself (Lambert, 1760; Beer, 1852). Appendix Section A.1 demonstrates how a series of simple manipulations of this well-known law leads to the following theoretical definition of "optical depth" τ:[8]

$$F_0 = F_{TOA} \exp(-\tau) \tag{1}$$

[5] Strictly speaking, scattering does not precisely lead to reductions in ground-level irradiance; it is the *backscattering* of radiation, which is the strongest in a type of scattering called elastic scattering, that reduces the net amount of radiation power arriving on the ground surface.

[6] In the upper atmosphere, the main driver of optical extinction is the stratospheric ozone layer in the ultraviolet spectrum. But the fluctuation of ozone concentration, conditional on a fixed position, shows very strong seasonality. The effect of ozone layer absorption is hence subsumed in the time-fixed effect of our model.

[7] "Scattering by particles accounts for 50–95% of extinction, depending on location, with urban sites in the 50–80% range and non-urban sites in the 80–95% range (Seinfeld and Pandis, 2016, p.646.)"

[8] In principle, the equations here are applied to *spectral*, or spectrum-specific, measurements of irradiance. We drop the spectrum λ notation here for simplicity, but Appendix Section A.1 provides additional explanations as to the spectral bands used for our analysis.

where F_0 is the surface-level irradiance (radiative power per area)[9]. F_{TOA} is the irradiance at the top of the atmosphere. The units for both F_0 and F_{TOA} are W/m^2, or power per area. The optical depth τ is hence a real number that measures the degree to which the incoming sunlight is extinguished from the atmosphere. For example, an optical depth of 0.1 suggests that approximately 10% of irradiance at the top of the atmosphere will be lost as it travels through the atmosphere. The advent of spaceborne sensors, such as the Moderate Resolution Imaging Spectroradiometer (MODIS), have provided continuous measurements of the optical depth (Appendix Section A.1).

It is this concept of optical depth τ that we use as our endogenous variable of interest. We are particularly interested in "aerosol optical depth" (AOD), which refers to the value of τ attributable to aerosol particles. AOD is measured from satellite sensors, such as the MODIS sensor that we use in our analysis. Equation (1) strongly suggests that it is the *log* of rooftop solar generation that the AOD reduces. Previous engineering studies involving small-scale air sampling experiments have confirmed this theoretical log-linear relationship (Li et al., 2017; Donaldson et al., 2021; Zhang et al., 2021).[10] Further, many engineering studies (e.g., Styszko et al., 2019) have found a lag of a few days in AOD impacts solar generation, as the deposition of molecules on panel surfaces can temporally alter the panels' optical properties.

While engineering studies have effectively confirmed the relationship between AOD and solar generation as posited in (1), one cannot directly translate the engineering results to empirical estimates because the engineering studies themselves (e.g., Perry and Troccoli, 2015) have found that air opacity's impact on solar generation can be heterogeneous, depending on the material and manufacturing technique used for the panel in question. For example, Li et al. (2017) estimates the impact of aerosols in East China to be 21% and 34% of solar generation for fixed-array and two-axis tracking solar systems, respectively.

The heterogeneous impact of air pollution, as identified by engineering studies, limits the applicability of engineering estimates in empirical settings. This is because the engineering

[9]Throughout this paper, we use the terms "irradiance" and "radiation" interchangeably.

[10]Such studies have involved small-scale experiments that, in a set outdoor environment, simultaneously measure both the ambient aerosol composition and the performance of an experimental solar PV panel. Based on the collected air sample, the studies have backtracked the AOD of the atmosphere using *a priori* knowledge of the optical properties of individual aerosol molecules.

studies are based on short data collection campaigns at specific sites under circumstances that may not be representative of real-world environments faced by rooftop solar systems in California. For example, while Peters et al. (2018) provides global estimates of the impact of PM2.5 on solar generation, their estimates are based on a 13-month measurement record with two PV systems and a single pyranometer (irradiance measurement equipment) in Delhi, India. Solar generation in California very likely differ from the experimental conditions in the maintenance, installation, and composition of panels.

We also note that equation (1) pertains to the AOD, not PM2.5. Recent studies in the remote sensing literature have shown a robust link between day-to-day variations in AOD and those in PM2.5. This has motivated studies that use satellite AOD imagery to estimate PM2.5 concentration with unprecedented resolution (Ma et al., 2022).[11] However, AOD is a measurement of air opacity attributable to *all* aerosol matter in the atmosphere. There is no guarantee that an increase in AOD is the result of PM2.5 pollution. Increases in ambient water droplet concentration from high humidity, for example, may increase AOD even without changing the PM2.5 concentrations.

These relationships between PM2.5, AOD, and solar generation have two implications for our identification strategy. First, to the extent that it is AOD, not PM2.5, that explains fluctuations in solar generation, the use of PM2.5 concentrations as the covariate of interest in an IV framework will violate the exclusion restriction whenever the instrument is correlated with other factors that contribute to AOD. Given the many factors that can influence AOD, it may be difficult to find an instrument that does not violate the exclusion restriction. Second, it is useful to empirically disentangle the share of PM2.5 in AOD fluctuations to better understand how this important pollutant influences opacity.

2.2 Air quality, wildfire, and wind

Wildfire smoke is one of the strongest drivers of day-to-day variations in air quality in the western United States (McClure and Jaffe, 2018). While some wildfire events lead to statewide impacts on air quality, most smoke episodes are local in nature (Appendix Figure

[11] Such fine-scale PM2.5 data have recently been used for more detailed empirical analysis on the distribution of pollution exposure (Currie et al., 2023).

F1). The spatial extent of wildfire smoke is strongly influenced by daily weather conditions, in particular wind direction and wind speed. These properties make wildfire smoke especially useful for empirical analyses by providing a source of quasi-experimental variation in air quality driven by wind conditions.

Upon wildfire ignition, the position of a ZIP code with regards to wildfire smoke (i.e., downwind or upwind from smoke origin) is determined by two factors: the angle between the smoke plume and the ZIP code, and the wind direction at each ZIP code. For example, if the smoke is placed east (90°) of a ZIP code, then this ZIP code will be downwind from the smoke when its local wind direction is easterly (90°). The ZIP code will be upwind under westerlies (270°). For the two cases, the cosine of difference in the two angles is $\cos(0°) = 1$ and $\cos(180°) = -1$. The difference between these two angles thus becomes a useful instrument for wildfire smoke exposure. The sign of the cosine of this angle summarizes whether a ZIP code is downwind from smoke or not, where a positive "cosine of angle difference" implies that the ZIP code is downwind of the smoke. Surface radiation in downwind regions (with positive cosine angle differentials) will be strongly affected by wildfire smoke than the upwind regions (with negative cosine angle differentials).

Figure 1 shows this relationship in action. We focus on a local wildfire smoke event near Los Angeles on October 19, 2016. The smoke's impact was local, as is the case for most wildfire smoke events. The blue arrows show wind direction in each ZIP code, while ZIP codes with wildfire smoke as of are demarked in red borders. Point "A" shows the origin of the smoke plume. The smoke spread across the Los Angeles metropolitan area following the wind direction, resulting in a smoke plume with centroid at point "D." The smoke plume is shown in red box in the top left subfigure of Figure 1a. Figure 1a shows the demeaned surface-level radiation, where radiation averages are taken in each ZIP code.

Point "B" in Figure 1 was dominated by easterly wind (90°), with the smoke plume's centroid to the west (270°). Point B was hence upwind of the wildfire smoke. This favorable wind condition led to B and its adjacent regions to experience a less severe drop in surface radiation than the rest of the ZIP codes in the area (Figure 1a). Figure 1b shows that the cosine of the angle difference at point B (180°) was negative.

On the contrary, point "C" was dominated by southeasterly winds (135°), with the smoke

polygon to the southeast (135°). Point C, being downwind, experienced a stronger drop in radiation than point B (Figure 1a). The cosine of angle differential was positive (Figure 1b).

The importance of local wind conditions in modulating smoke impact is a key part of our identification strategy. For each wildfire smoke episode, we measure the angle between the smoke's centroid and each ZIP code and calculate its difference from local wind direction, as in Figure 1b. We also include wind speed as an instrument, as low wind speed can accentuate smoke exposure.

Our approach builds on previous studies that have used wind direction as drivers of pollutant transport (e.g., Schlenker and Walker, 2016). Contrary to such studies, which compared wind direction to a stationary pollution source (e.g., airports), our method has the added benefit of allowing the position of the pollution source to vary. The random nature of wildfire ignition, combined with day-to-day shifts in wind direction, provide a plausibly exogenous driver of local air conditions.

3 Data and Descriptive Statistics

3.1 Rooftop solar generation

We obtain data on our primary outcome variable, rooftop solar generation, from California Solar Initiative's (CSI) "15-minute interval" data set (California Public Utilities Comission, 2016). The data set compiles production meter records from 504 rooftop PV systems registered in the CSI program over a seven-year period from 2010 to 2016, at a fifteen-minute interval. Each PV system in the data set can be identified using their unique CSI ID codes, and we pair these with the "CSI Application" data set, which includes data on the characteristics of each PV system (e.g., installed ZIP code).

As the CSI data set is obtained from automated production meters, removing PV systems with anomalous records becomes a key component of our data preparation process. We apply three major criteria to filter out spurious data inputs. First, to assess the degree of data omission at each production meter, we aggregate the fifteen-minute interval data to hourly levels, and count the number of missing entries. We remove 26 PV systems which either (i) had more than 20% of their entries missing or (ii) had more than 10% of entries consecutively

unrecorded.[12] Second, we identified seven PV systems which had excessively large negative production records.[13] While negative production is theoretically possible after sunset (e.g., some power used for the inverter), the negative production should be negligible compared to the amount of solar production. We hence drop these systems from the data set. Third, for remaining observations, we treat negative entries as missing values; this constitutes 0.4% of the entire data set.[14]

The data cleaning process presented here is analogous to that taken by a CSI-commissioned report that used the same data to analyze the solar generation potentials of ZIP codes in California (Clean Power Research, 2018). We conduct robustness checks on these data cleaning procedures and find that they do not change our results.

We aggregate the PV system-level data to the ZIP code level by calculating the average of solar energy generated in each ZIP code. For compatibility with other data, we aggregate the 15-minute data set to total daily production. The resulting data set contains 470 PV systems over 288 ZIP codes.

3.2 Simulated surface-level irradiance

We supplement our CSI production meter data set with a simulated solar irradiance data from the National Solar Radiation Database (NSRDB) (Sengupta et al., 2018). NSRDB provides a gridded estimate of surface-level solar radiation, based on an algorithm combining physical models of the atmosphere, cloud and aerosol measurements, and *in situ* weather observations. NSRDB data have shown superb performance when compared with on-the-ground precision measurements of irradiance, especially when they were aggregated to daily intervals (Habte et al., 2017).

We obtain three different types of surface-level radiation estimates from NSRDB. First, global horizontal irradiance (GHI) is a measure of the intensity of solar energy reaching the surface per unit area from all directions, and is a proxy of solar PV cell productivity. Second,

[12] The 20% and 10% thresholds were set by conducting unsupervised k-means clustering on the missing entry ratio of the PV systems. k was set at 3.

[13] PVs with "excessive negative production record" are defined as those whose sum of negative entries throughout the seven-year period exceeded −10,000kWh.

[14] 2,159 rows out of 505,960 observations. The mean of these records is −1.13kWh/day, compared to an average of 38kWh/day for the entire data set.

direct normal irradiance (DNI) isolates the share of GHI that originates from the direction of the sun. Diffuse horizontal irradiance (DHI) measures the level of radiance that does not come directly from the sun, but has been diffused *via* collision with atmospheric particles. Theories in atmospheric physics dictate that increases in atmospheric opacity will reduce DNI, while potentially increasing DHI due to intensified scattering.

We use the GHI data in NSRDB to evaluate the external validity of our findings. Specifically, we assess whether the GHI of our CSI-included ZIP codes show a systematically different response to those across the entire state of California. The DHI and DNI data are used to understand the mechanisms of air opacity impact on surface-level irradiance. Appendix Figure F2 illustrates the coverage of our CSI data, and the expansion of data coverage conferred by NSRDB.

3.3 Satellite-measured air opacity and downscaled PM2.5

The variable of interest in our first analysis is satellite-measured air opacity, which is instrumented for with daily wind conditions and wildfire smoke influx. We obtain estimates of AOD from the MODIS sensor on board the Terra and Aqua satellites. We specifically refer to the Multi-angle Implementation of Atmospheric Correction (MAIAC) data set's AOD estimates, which is accessed *via* Google Earth Engine (Gorelick et al., 2017), a cloud computing platform that facilitates acquisition of gridded geospatial data.

While the MODIS AOD data that we access is already quality-controlled for spurious variations (e.g., those from "water glints," or production of overly bright retrievals from sunlight reflections from adjacent water bodies), the data set does flag best-quality observations. We use the full AOD data set for our main analysis to maximize statistical power, but we conduct robustness check with only "best" AOD observations, as flagged by the data provider.

While AOD is our variable of interest, we also use downscaled PM2.5 for comparison purposes to show the implications of directly modeling the effect of PM2.5 on solar generation. To this end, we use a recently developed data set (Reid et al., 2021) that estimates PM2.5 concentration at the census tract level in California.[15] These data have shown good

[15]These data are obtained from Machine Learning algorithms, applied to a composite data set of satellite

downscaling performance in the state of California ($R^2 > 0.76$, RMSE around $4\mu g/m^3$), with a particular strength in its urban regions, where the great majority of our ZIP codes are located.

3.4 Wildfire smoke and measured PM2.5

We collect wildfire smoke data from NOAA's Hazard Mapping System (HMS, Brey et al., 2018), which uses satellite imagery to capture the spatiotemporal extent of wildfire smoke *à la* Figure 1. We declare that a ZIP code has been "exposed" to wildfire smoke if its boundary intersects with the smokes' boundary. A given day is declared as "wildfire smoke day" if at least one of the ZIP codes in California was exposed to smoke. Appendix Figure F3 shows that the smoke days are concentrated in California's "wildfire season," from June to October.

On each smoke day, we identify the centroid of each smoke polygon.[16] We then calculate the difference between two angles: (1) the angle between each ZIP code and its smoke polygon, and (2) the azimuth angle of wind direction (direction measured in degrees from the north) at each ZIP code. As shown in Figure 1b, the cosine of this difference shows whether a region is downwind from the closest smoke polygon.

For our analysis of AOD-PM2.5 relationship, it is important to use *measured*, not downscaled, PM2.5 data. This is because downscaling algorithms of PM2.5 often include AOD among the variables used for the downscaling. Any results from regressing downscaled PM2.5 on AOD thus risk being a simple artifact of the original model used. We thus collect PM2.5 concentrations from years 2010 to 2016 in Environmental Protection Agency's (EPA) air quality measurement sites across California.

One possible concern with measured PM2.5 data is that the distribution of EPA monitors within each county and ZIP code may not be random (Grainger and Schreiber, 2019). Appendix Figures F4 to F7 address this possibility. Using downscaled data, Appendix Figure F4 illustrates that the census tracts that host EPA monitors do not have substantially higher PM2.5 pollution than other tracts in the same ZIP code. Appendix Figure F5 finds

imagery, land cover, chemical transport model, and weather observations.

[16] If the smoke is such that there are multiple polygons overlapped on one another (representing the spreading of smoke across space), we select the smallest polygon as the "original" polygon and identify its centroid.

strong alignment between monitor-observed PM2.5 levels and ZIP code-wide PM2.5 levels from downscaled data (with complete spatial coverage). Both figures suggest the PM2.5 monitors reasonably represent their ZIP codes' average PM2.5 levels in our setting.

Appendix Figure F6 compares each census tracts' average PM2.5 levels with their counties' average, and find that tracts with and without monitors show similar distribution in this deviation. Lastly, in Appendix Figure F7 we rank the PM2.5 level at each tract within its county, and find no evidence consistent with systematic sensor concentration in low-pollution areas. However, we find a weakly bimodal concentration of monitors on either low- or high-pollution areas. We thus test the robustness of our results to dropping sensors outside the 10% to 90% percentile range.

3.5 Daily weather and nitrogen dioxide (NO_2)

We obtain data on daily weather conditions from a 4km-resolution gridded daily weather data set called gridMET (Abatzoglou, 2013). The data set merges a temporally rich interpolated weather reanalysis product (NLDAS-2) with a spatially rich data set (PRISM) to produce a spatially continuous data set of daily temperature, relative humidity, wind speed, and wind direction. We take the cosine of the wind direction to assure continuity. We access the gridMET data *via* Google Earth Engine and retrieve information from each ZIP code centroid over the seven-year period.

To the extent that wind could be representative of large-scale weather conditions, the use of wind conditions as an instrument for PM2.5 conditions is vulnerable to a violation of the exclusion restriction. We hence obtain two proxies of large-scale weather[17], 700hPa vertical velocity ("omega" measure) and mean sea level pressure from the NCEP-DOE Reanalysis II product (Kanamitsu et al., 2002).

Lastly, we note that AOD by definition focuses on atmospheric opacity from aerosols. Thus, AOD only measures opacity from pollutants in liquid and solid states, not in gaseous states, for aerosols are defined to be liquid or solid particles or droplets. As NO_2 is the pri-

[17]Weather events are categorized into synoptic scale, mesoscale, or microscale depending on their horizontal size. Here we define "large-scale weather" as "synoptic scale weather," which refers to large-scale properties of the weather system that has a spatial scale of around 1,000 km or more; common examples of synoptic scale weather include high and low pressure systems and weather fronts.

mary source of gas radiation absorption (Seinfeld and Pandis, 2016), we collect tropospheric NO_2 concentration estimates from the Ozone Monitoring Instrument (OMI) sensor on board the Aura satellite (Krotkov et al., 2019).

3.6 Socioeconomic characteristics

To assess the possibility of disproportionate exposure to adverse air quality conditions and potential heterogeneity, we collect data on socioeconomic characteristics of each ZIP code in California. The data are accessed through EPA's EJScreen data set (US EPA, 2022). We aggregate EJScreen's census tract-level data to the ZIP code level. The collected socioeconomic characteristics are % of people of color, % of population with less than high school education, % of low income households. Appendix Table F1 summarizes the data sources and definitions used for this research. While we use EJScreen data from 2021[18], we examine other years and find no notable shifts over time in the distribution (Appendix Figures F8 to F10). We also find that within each ZIP code, the socioeconomic characteristics do not show strong changes from 2016 to 2021 (Appendix Figures F11 to F13).

3.7 Summary statistics

Three characteristics stand out from the summary statistics of our data, shown in Table 1. First, there are significant day-to-day fluctuations in PM2.5 and AOD, with coefficients of variation (CV) 0.8 and 0.6 for the two measures. Similarly, rooftop solar generation shows considerable variation (CV around 0.7), while GHI estimates are relatively stable (CV around 0.3). NO_2 and wind direction also show very strong day-to-day fluctuations.

Second, the scales of the four outcome variables (panel A) are all different. The mean of DNI in particular is almost eight times as large as that of DHI. Although DNI and DHI are two subcomponents of GHI, reductions in the two outcomes by the same percentage should not be given the same weight. This further justifies the log-transformation of the outcome variable. Third, due to the nature of retrieving algorithms, the PM2.5 and NO_2 estimates

[18] 2016 is the earliest year from which we have access to EJScreen data. We do not use EJScreen data set from 2022 as it has a different spatial unit of analysis compared to those in previous years (2020 Census blocks versus 2010 Census blocks).

in the data set show some negative values. While our primary results drop these entries, we run robustness checks with these observations included.

4 Empirical Strategy

This study aims to analyze the effect of air pollution on rooftop solar generation to quantify its externality. As discussed above, we develop a two-stage approach motivated by the fact that PM2.5 is but one source of air opacity, and that other sources can be correlated with PM2.5 pollution. Under this circumstance, wind conditions will likely be an invalid instrument for PM2.5 when directly examining the relationship between PM2.5 and solar generation because the exclusion restriction would require that PM2.5 is the *only* way through which the instrument affects solar generation. Wind conditions can affect other pollutants and thus are not likely to satisfy this condition. Our approach is to first establish the casual effect between air *opacity* and solar generation, and then quantify the causal effect of air *pollution* on this air opacity measure. Combined, these two analyses allow us to quantify the causal relationship between PM2.5 pollution and solar generation.

4.1 Air opacity and solar generation

The first part of our analysis aims to estimate the contemporaneous effect of air opacity on rooftop solar generation and on other related outcomes (e.g., DHI). The reduced-form equation of interest is:

$$\log(y_{iswt}) = \beta \text{AOD}_{iswt} + X'_{iswt}\lambda + \alpha_{is} + \gamma_w + \varepsilon_{iswt}, \tag{2}$$

where y_{iswt} is the outcome variable, rooftop solar generation, in ZIP code i, year s, week w, day t. β is the coefficient of interest, which estimates the marginal effect of AOD, a measure of air *opacity*, on the log of outcome variable. In light of fundamental theories in atmospheric physics (Lambert, 1760; Beer, 1852) discussed in Section 2 and encapsulated in equation (1), we adopt a log-linear model specification.

While our primary outcome variable of interest is the rooftop solar generation, we also assess the effect of AOD on simulated ground-level irradiance, a strong predictor of solar

generation with statewide data coverage. The simulated irradiance data are particularly useful in testing the external validity of the results from production meters. If we find similar β estimates from the 288 ZIP codes (with rooftop solar data) and from the entire state of California using the simulated ground-level irradiance, this suggests that the results based on the subsample with production meters are likely to have external validity to all of California.

Aerosol conditions on a specific day could have lagged effects, as aerosol particles can be deposited on the surface of solar PV panels, temporarily altering their optical properties. Following previous empirical literature (e.g., Deryugina et al., 2019), we account for this lagged effect by calculating y_{iswt} using the three-day sum of rooftop solar generation. The β hence measures the effect of AOD on day t on the sum of solar generation on days t, $t+1$, and $t+2$. For both OLS and 2SLS estimation, we include leaded covariates and instrumental variables. This prevents the time-varying covariates and AODs of days $t+1$ and $t+2$ from affecting the three-day outcome. We also estimate separate regressions with different lags (up to five-day sums) to identify the lag structure of (2). For placebo tests, we substitute the y_{iswt} by the outcome variable thirty days into the past.

We include a vector of time-varying covariates X_{iswt} to control for important observables. Six covariates are used: temperature minimum & maximum, relative humidity minimum & maximum, precipitation, and tropospheric NO_2 concentration. The last time-varying covariate is of particular importance, as AOD does not account for non-aerosol (i.e., gaseous) molecules that extinguish solar radiation. To the extent that NO_2 levels may be correlated with AOD, omission of this variable exposes the model to endogeneity. We account for nonlinear effects by including quadratic and cubic transformations of these covariates.

We add nonparametric controls for time-invariant characteristics in each ZIP code and ZIP code-specific year shocks by the fixed effect term α_{is}. We also control for within-year fluctuations of AOD and solar generation (notably from day length) using the week-fixed effect γ_w. We run robustness checks with month-fixed effects. All standard errors are clustered by ZIP code and year-week; the former accounts for temporal autocorrelation, the latter accounts for spatial correlation of errors within each year-week.

OLS estimation of the β term in equation (2) will be unbiased if the AOD variable is not

exposed to measurement error, is not correlated with any other factors that affect rooftop solar generation, and if there is no simultaneity. Such assumptions are likely untrue for two reasons. First, to the extent that we obtain AOD estimates at the centroid of each ZIP code,[19] there is a risk of measurement error, which could place a downward bias on the coefficient estimates (Moretti and Neidell, 2011).

More importantly, OLS estimates of (2) are vulnerable to simultaneity. Empirical literature has found that the fuel mix for electricity generation affects local air pollution conditions (Johnsen et al., 2019; Sexton et al., 2021). Shocks in solar generation productivity thus likely lead to changes in AOD. The sign of bias is not obvious; increases in solar generation may reduce the overall share of fossil fuel in the fuel mix, but could also increase demand for dispatchable energy sources upon sunset (following what is often called the "duck curve," from the shape of the dispatchable energy demand curve under strong solar share in the fuel mix). On top of this electricity generation-related source of reverse causality, there can be natural sources of reverse causality. Well-known principles in atmospheric chemistry suggest that high solar irradiance can facilitate the formation of atmospheric pollutants, *independently* from temperature and humidity conditions.[20]

To support this claim, Appendix Section B and Appendix Table B1 present a series of tests that provide suggestive evidence of reverse causality. Intuitively, the tests show that ZIP codes with natural gas power plants have lower OLS β estimates than those without gas plants. Such findings are consistent with favorable radiation conditions (i.e., positive solar anomaly) causing a decrease in gas generation by altering the daily fuel mix, leading to reductions in AOD levels. The heterogeneity grows with the increased adoption of solar in California over the 2010-2016 period.

Faced with this possibility, we adopt a two-stage least squares (2SLS) method exploiting quasi-random variations in atmospheric opacity from wildfire smoke influx and wind conditions. As shown in Figure 1, daily wind conditions play an important role in the transport and dispersion of wildfire smoke. The interaction of daily wind conditions and wildfire smoke

[19]This is due to computational limits; assigning multiple sampling points within each ZIP code leads to exponential increases in Google Earth Engine query time.

[20]A prime example is the facilitation of photolysis under high radiation conditions.

hence constitute a plausibly exogenous shock to AOD. The first-stage regression is:

$$\begin{aligned} AOD_{iswt} = &\text{SMOKEDAY}_t \times \text{DOWNWIND}_{iswt} \times f(WS_{iswt})'\delta_1 \\ &+ \text{SMOKEDAY}_t \times \text{DOWNWIND}_{iswt} \times f(WD_{iswt})'\delta_2 \\ &+ X'_{iswt}\lambda + \alpha_{is} + \gamma_w + \epsilon_{iswt}, \end{aligned} \tag{3}$$

where SMOKEDAY_t is a dummy variable indicating smoke days, defined as a day in which at least one ZIP code in California was affected by wildfire.

The definitions of DOWNWIND_{iswt} and WD_{iswt} differ depending on SMOKEDAY_t. When $\text{SMOKEDAY}_t = 1$, we first take the difference of two angles: the wind direction, and the angle between ZIP code i and its closest smoke polygon. We then define WD_{iswt} as the cosine of this angle differential. As shown in Figure 1b, this measure effectively captures the position of each ZIP code relative to wildfire smoke origin. DOWNWIND_{iswt} is an indicator variable with $= 1$ when this cosine measure is above zero, hence being downwind (blue ZIP codes in Figure 1b). If $\text{SMOKEDAY}_t = 0$, then WD_{iswt} is simply the cosine of wind direction, and DOWNWIND_{iswt} the sign of WD_{iswt}. We run robustness checks on alternative definitions of SMOKEDAY_t and WD_{iswt}, where $\text{SMOKEDAY}_t = 1$ only for ZIP codes within the 100km perimeter of each smoke polygon.

WS_{iswt} is the wind speed observed for ZIP code i on day t. $f(\cdot)$ is a third-order polynomial that accounts for nonlinear effects of wind direction and wind speed. The fixed effects are identical to the structural equation (3).

The first-stage regression (3) exploits plausibly random variation in the occurrence of wildfire smoke exposure (SMOKEDAY_t) and the position of each ZIP code with regards to the smoke itself (DOWNWIND_{iswt}) to identify the effect of AOD on solar generation. The exclusion restriction assumption is that AOD on day t is the only channel through which same-day wildfire smoke conditions, interacted with wind conditions on day $t, t+1$, and $t+2$, can affect solar generation. This assumption is reasonable, especially in light of the "catch-all" characteristic of AOD measurement and the rich set of covariates and fixed effects included. The 2SLS estimate of β will be biased if day-to-day wind changes and wildfire smoke influx are correlated with the measurement error or the errors from reverse causality.

This is unlikely, considering the random nature of daily wind conditions and the position of each smoke polygon with regards to its neighboring ZIP codes.

4.2 Air pollution and air opacity

AOD is a catch-all measure of aerosol concentration in the atmosphere. Fluctuations in AOD thus cannot be attributed solely to changes in PM2.5 air pollution. Thus, linking PM2.5 air pollution to reductions in solar generation requires an additional step to estimate the casual effect of PM2.5 on AOD.

One distinct characteristic of the PM2.5-AOD relationship is its seasonal heterogeneity. PM2.5 pollution in California has different sources depending on the season (Hu et al., 2014), which also have different optical properties (i.e., potential to increase opacity). We thus run separate regressions for each season. The structural equations are, for seasons $a = 1, 2, 3, 4$,

$$AOD_{isawt} = \tilde{\beta}_a \text{PM2.5}_{isawt} + \tilde{X}'_{isawt}\tilde{\lambda} + \tilde{\alpha}_{is} + \tilde{\gamma}_w + \tilde{\varepsilon}_{isawt}, \tag{4}$$

where AOD_{isawt} is the AOD at ZIP code i, year s, season a, week w, day t. $\tilde{\beta}_a$ is the coefficient of interest. The fixed effects are identical to the first part (2), and the standard errors are once again clustered by ZIP code and year-week. We also estimate the model for the entire year (without seasonal coefficients) to test seasonality.

Equation (4) differs from the solar-AOD structural equation (2) in three important ways. First, rather than the log-linear specification, we adopt the linear specification in light of previous studies on the theoretical basis of AOD-PM2.5 relationship (Appendix Section A.2). Second, as the variable of interest in the AOD-solar generation model was single-day air opacity, we also focus on single-day PM2.5 concentration $PM2.5_{isawt}$.

Lastly, the covariate vector $\tilde{X}_{isawt}$ has three additional components. We control for large-scale ("synoptic") weather conditions: mean sea level pressure (MSLP) and 700hPa vertical velocity. This controls for the potential correlation between PM2.5 pollution levels and weather conditions (e.g., temperature inversion), which may be correlated with air opacity (Sager, 2019). We also control for PM10 levels for the annual model, to account for possible

correlations in PM2.5 and PM10 shocks.[21]

While using single-day PM2.5 concentration eases the linking of our previous results (AOD-solar generation) from (2) to those from (4), the approach could be problematic for identification. PM2.5 pollution could easily have a lagged/leaded effect on AOD (due to atmospheric chemistry or anticipatory behaviors), so that PM2.5 pollution is correlated across days. If this is the case, a regression of AOD on same-day PM2.5 may end up confounding the true effect of same-day PM2.5 with the effect from previous days.

We address this concern by two methods. First, we instrument for PM2.5 using daily wind conditions, using the first-stage regression

$$PM2.5_{isawt} = f(WS_{isamwt})'\tilde{\delta}_{1m} + f(WD_{isamwt})'\tilde{\delta}_{2m} + \tilde{X}'_{isawt}\tilde{\lambda} + \tilde{\alpha}_{is} + \tilde{\gamma}_w + \tilde{\epsilon}_{it}, \quad (5)$$

where all variables are identical to (3), except for WD_{isamwt}. Wind direction is now simply the cosine of daily wind direction. Assuming that day-to-day wind conditions are not serially correlated, 2SLS estimates with (5) will reveal the contemporaneous impact of same-day PM2.5 concentration.

Second, we perform robustness checks that simply add leads and lags of PM2.5_{isawt} and covariates to test changes in the coefficients. The additional regression is:

$$AOD_{isawt} = \tilde{\beta}_a^0 \text{PM2.5}_{isawt} + \tilde{\beta}_a^1 \text{PM2.5}_{isaw,t-1} + \tilde{\beta}_a^2 \text{PM2.5}_{isaw,t+1} + \tilde{X}'_{isawt}\tilde{\lambda} + \tilde{\alpha}_{is} + \tilde{\gamma}_w + \tilde{\varepsilon}_{isawt} \quad (6)$$

In Appendix Section C.1, we show that this augmented model will lead to significant differences in estimates of $\tilde{\beta}_a^0$ when (a) the true Data Generation Process (DGP) includes the lag and lead terms, *and* (b) PM2.5 concentration is serially correlated. The contrapositive tells us that, the absence of changes in $\tilde{\beta}_a^0$ coefficient estimates (versus (4) and (6)) would imply either that (a) the true DGP does not include leads and lags or (b) there is no serial correlation in PM2.5_{isawt}. We present evidence against (b) in Appendix Section C.2. The comparison of original model (4) to its augmented version (6) thus provides a useful (although still suggestive) test for whether the DGP includes lagged and leaded effects.

[21] We do not fit PM10 regressions for seasonal models due to lack of observations; for most EPA sites, PM10 samples are not collected at a daily interval.

5 Results

5.1 Air opacity and solar generation

Table 2 presents the main results of our analysis. Panel A shows the result of the OLS regression examining the impact of air opacity (AOD on solar generation (equation (2)). Panel B shows results from the panel IV regression, instrumenting for AOD by wind conditions (equation (3)). Each column represents separate regressions; all columns use AOD as the endogenous variable of interest, except for column 2.

Column 1 presents the estimates examining AOD's impact on rooftop solar generation, the first part of our two-part analysis. While OLS estimates demonstrate that a unit increase in AOD leads to a 22.6% drop ($(e^{-0.256}-1)\times 100$) in three-day rooftop solar generation, panel IV estimates a significantly larger impact of 79.0% ($(e^{-1.559}-1)\times 100$). This translates to a 1.6% and 9.4% drop under one standard deviation (0.063) increase in air opacity, with 2.6% and 15.1% reductions in solar generation from average AOD levels (0.105). The latter figure is in line with Bergin et al. (2017), which used short-run experimental observations from India to estimate an aerosol impact of 6-18% in California solar production. The remarkable difference between OLS and IV estimates is consistent with previous studies involving IV methods (Schlenker and Walker, 2016), and suggests that measurement error and/or reverse causality may be a major issue. The large F-statistics of our first-stage regressions attest to the strength of our instruments.

Column 2 presents results from an alternative approach, which directly links PM2.5 with rooftop solar generation. The estimation hence involves substituting AOD by PM2.5 in equations (2) and (3). The OLS and IV estimates show that a $1\mu g/m^3$ increase in PM2.5 concentration leads to a 0.5% and 1.2% decrease in three-day rooftop solar generation, respectively. A standard deviation increase of PM2.5 concentration ($4.5\mu g/m^3$) hence leads to a 5.03% drop in rooftop solar generation, according to IV estimates. The smaller size of the PM2.5 impact compared to that of AOD, which includes all sources of atmospheric opacity, is consistent with PM2.5 not being the only pollutant that reduces solar generation.

Column 3 shows that for ZIP codes included in our CSI data, the negative relationship between AOD and solar generation is also shown in NSRDB surface-level irradiance mea-

surements. Column 4 confirms that such trends are also true when expanded to the entire state of California. The coefficient estimate of -0.95 is striking, as this is very close to the theoretical value of the coefficient suggested by well-known principles in atmospheric physics (-1, Appendix A.1). The remarkable similarity of coefficient estimates in columns 3 and 4 also suggest that the ZIP codes included in our data set do not show a substantially different response to air opacity, when compared to the remaining California ZIP codes. Columns 3 and 4 hence indirectly establish the external validity of our findings in column 1. Both columns also show the same trend of IV estimates being stronger in magnitude compared to OLS estimates.

Columns 5 and 6 perform a placebo test. Solar generation and surface-level irradiance thirty days into the past (our placebos) should not have any functional relationship with contemporaneous AOD. Both OLS and IV estimations fail to find any meaningful estimates for such outcomes.

Table 3 suggests an explanation for the mechanisms driving the AOD impact on solar generation. The table presents regression results for model (2), but with irradiance subcomponents (direct and diffused irradiance) as the outcome in separate regressions. Columns 1 and 2 show that it is the direct component of irradiance (i.e., solar radiation that is coming from the direction of the sun) that is adversely affected by AOD. Three-day direct irradiance is reduced by 10.9% under one SD increase in AOD. On the contrary, higher AOD *increases* diffused irradiance (i.e., radiation that reaches the surface after scattering) by 14.7%. The significantly larger dimensions of the direct radiation (Table 1) result in a net decrease in the surface-level irradiance from AOD. Such findings are congruous with fundamental theories of atmospheric physics, which find that ambient particulates reduce irradiance directly originating from the sun, but increase solar ray interference mid-air. Again, the coefficient estimates are similar for the rooftop data-sample and the augmented state-wide sample (column 1 versus 2, 3 versus 4). IV models consistently find a stronger effect compared to their OLS counterparts. Coefficients are not significant for placebo outcomes.

5.2 Air pollution and air opacity

We now show that there is a clear link between daily shifts in PM2.5 air *pollution* and daily changes in air *opacity*. This constitutes the second part of our analysis and allows us to causally link PM2.5 pollution to solar generation losses.

Panel A of Table 4 estimates model (5) using fixed effects OLS methods. Panel B uses wind conditions to instrument for PM2.5 conditions using first-stage regression (5). Panel C estimates an OLS model with lagged and leaded terms (equation (6)) to see if controlling for lagged or leaded effect influences our coefficient of interest. Columns 1 to 4 are results from seasonal subsamples, while columns 5 and 6 use the entire data set. Column 6 includes PM10 data to test for potential omitted variable bias from co-emission of PM10 with PM2.5, which could affect atmospheric opacity.

Columns 1 to 4 suggest that PM2.5 concentration in the summer has a larger impact than in other seasons. An increase in PM2.5 by $1\mu g/m^3$ leads to an increase of AOD by 0.0019 in the winter versus 0.0041 in the summer, equivalent to 2.9% and 6.3% of the standard deviation of AOD in the sample, respectively. The similar impacts of PM2.5 in fall and winter are consistent with atmospheric sampling studies which found similar sources of PM2.5 pollution (e.g., residential wood stoves) in these two seasons (Hu et al., 2014).

Across all columns, the three approaches in Panels A, B, and C show similar coefficient estimates. All leaded and lagged term coefficients in panel C are not significant and the coefficients on the same-day PM2.5 are nearly identical to those in Panel A. As previously mentioned, given serial correlation in PM2.5 levels, comparisons between estimates in panels A to C indirectly test the violation of strict exogeneity (Appendix C). The very little difference in coefficients is consistent with the true DGP not including leaded and lagged terms of PM2.5. The addition of contemporaneous PM10 concentrations, in column 6, also do not shift the results. Overall, these results suggest quite robust results.

Combining Table 4 results with Table 2 column 1, panel B results allows us to estimate the effect of PM2.5 on solar generation. Using column 5, panel B of Table 4 estimates, the average PM2.5 level in California ($10.3\mu g/m^3$) leads to an increase in AOD by 0.0258, which translates into 4% drop in solar generation ($e^{-0.0258\times 1.559}-1$). This is a larger effect than in

Peters et al. (2018), which used experimental estimates from Delhi, India, to estimate the impact of PM2.5 in Los Angeles to be 2.2±0.2%.

One can also compare the results from Tables 2 and 4 with those from a recent empirical study on the impact of PM2.5 on solar generation in Arizona (He et al., 2020). As previously discussed, this study differs from He et al. (2020) in the usage of AOD as the endogenous variable, the use of wildfire in first-stage regression, and in the functional specification. Appendix Section E explores the implications of relaxing these differences. The results from using PM2.5 (not AOD) as the endogenous variable support the hypothesis that there are other pollutants correlated with PM2.5 that affect solar generation (i.e., violation of exclusion restriction). We also find that adopting the He et al. (2020) first-stage regression fails to identify the AOD impact on solar generation. The linear specification does not lead to large differences in estimates of AOD effects, but changes the interpretation of the results in a way that is not consistent with well-known theories in atmospheric physics.

It is also notable that OLS and 2SLS models for the AOD-PM2.5 relationship show similar coefficient estimates (Table 4), contrary to the large differences in the solar-AOD analysis (Tables 2 and 3). One potential explanation for this difference is the vulnerability of solar-AOD analysis to measurement errors. The independent variable for the latter analysis is measured at the centroid of the ZIP code, while the former analysis uses PM2.5 measured at EPA sensors. Under classical measurement error, OLS coefficient estimates will be biased towards zero. This is consistent with the estimate differences found in Tables 2 to 4.

5.3 Lagged and heterogeneous effects of AOD on solar generation

We now explore the lagged effect of AOD on solar generation. To this end, we run IV regressions as in (2) and (3), but with same-day to five-day solar generation as outcome variables (versus the three-day solar generation shown in Table 2). Table 5 shows that while a SD increase in AOD leads to a sharp decrease in same-day solar generation, the effects of this increase persist even after four days (11.62% for same-day, 4.9% for five-day generation). Similar trends are observed for surface-level irradiance. The lasting impacts of AOD are likely signs of atmospheric particulates being deposited on PV panels, a phenomenon observed in field experiments (Styszko et al., 2019).

Appendix Table F2 explores the heterogeneity in AOD impact on rooftop solar generation. To assess heterogeneity, we divide the ZIP codes into two groups depending on their socioeconomic characteristics, and run separate 2SLS regressions with (2) as the structural equation. We find very little difference in the effect of air opacity across subgroups.

5.4 Robustness checks

Table 6 shows the robustness of our IV estimates to different conditions. The results in columns 1, 2, and 3 are obtained from various changes to the data set. Column 1 expands the number of solar panels in the data to include those with more than 10% of entries consecutively unrecorded. Column 2 only keeps AOD measurements with "best" quality that did not have any neighboring clouds.[22] Column 3 retains negative NO_2 entries. None of these estimates are statistically different from the preferred estimate in column 1 of Table 2.

Columns 4 to 6 address concerns of our model specification and estimation methods. Column 4 uses wind direction bins, with bins set at 45° interval, instead of cosine of wind direction and its polynomials. Column 5 takes an alternative definition of smoke day, where the SMOKEDAY variable is set at 1 (and angle differences taken) only if there is a smoke polygon within 100km radius of each ZIP code. Column 6 relaxes the homoskedasticity assumption and use ZIP code-average rooftop solar generation as weights. The confidence intervals for all these alternative specifications overlap with those of the preferred specification, column 1 of Table 2.

Appendix Section D presents additional robustness checks of our results. We find that our Table 2 results are robust to alternative fixed effects and different structural equation specifications. We also find that Table 4 results are also robust to dropping EPA monitors placed in high- and low-pollution areas within each county.

[22]The AOD retrieval mechanism tends to be more noisy under neighboring clouds. It should be noted, however, that the "non-best" AODs that we keep in our data set (for the purposes of statistical power) have also passed quality-control tests for water glint and anomalously low or high retrievals.

6 Policy Implications

6.1 Reduction in solar generation from air pollution

We now assess the policy implications of our empirical findings. We first ask: How much rooftop solar productivity loss in California is incurred by PM2.5 pollution? To answer this question, we refer to the estimates from column 1 of Table 2 and columns 1 to 4 of Table 4. We calculate the average loss of solar generation attributable to PM2.5 at each ZIP code from daily PM2.5 data. We also estimate the loss attributable to AOD writ large.

The results show that on average, AOD accounted for a 14.5% loss in three-day rooftop solar generation (Appendix Figure F14), while some ZIP codes lose close to 30% of productivity from AOD.[23] On the contrary, PM2.5 levels account for 3.91% reduction in solar generation on average, with up to 6% loss in certain high-pollution areas (Appendix Figure F15).

We then combine these ZIP-code level reduction estimates with the CSI application data set (California Public Utilities Comission, 2022) to estimate the potential energy generated from rooftop solar panels using a capacity factor of 18%.[24] Multiplying this installed capacity to the PM2.5-lost productivity, we find that the PM2.5-attributed loss in rooftop solar capacity has steadily increased with the expansion of distributed generation (Figure 2). As of December 2021, PM2.5 pollution has reduced annual residential rooftop solar generation by 459 GWh. AOD writ large has led to a loss of 1,616 GWh in solar generation. As residential CSI installations only take up a small share of the entire solar capacity, this figure is very likely a lower bound of air quality-attributed rooftop solar generation loss in California. As of 2021, CSI residential rooftop solar capacity stood at 6,983 MW, compared to California's utility-scale solar capacity of 15,115 MW.

We can also use our empirical findings to assess the externality of PM2.5 emissions caused by loss in solar generation. This previously unaccounted-for externality of PM2.5

[23] The ZIP code with worst compromises in solar generation is placed in Treasure Island, in the San Francisco Bay Area.

[24] We refer to National Renewable Energy Laboratory (NREL)'s annual technology baseline (ATB) estimates for capacity factor values in California (National Renewable Energy Laboratory, 2021). The 18% capacity factor is applicable to regions with GHI in the range $[5.25, 5.5]$; most of California is above this range.

abatement originates from the fact that solar generation loss from PM2.5 conditions will lead to increased generation of electricity. If the electric utility generation is from fossil fuels, this will increase emissions further.

We find that a PM2.5 concentration increase of $1\mu g/m^3$ will reduce CSI solar generation by 8,071 MWh, 7647 MWh, 17,388 MWh, and 11,463 MWh in winter, spring, summer, and fall. Using recent estimates of the social marginal cost (SMC) of electricity generation in California of 7¢/kWh (Borenstein and Bushnell, 2022), this amounts to a negative externality of $3.12 million due to lost rooftop solar generation ($1,768 per ZIP code). Previous atmospheric modeling studies in California have linked this marginal increase in PM2.5 concentration to approximately 10 ton/year emission of surface-level primary PM2.5 (Kelly et al., 2015).[25] This results in the marginal externality of primary PM2.5 emission of $177/ton/year, which is smaller than the $3,220/ton/year (in 2000 USD, $5,073/ton/year in 2021 USD) estimate of mortality damages from from PM2.5 pollution (Muller and Mendelsohn, 2009), but it is still an economically significant externality.

6.2 Disproportionate exposure to solar generation loss

Another policy question relevant to our empirical finding is whether the loss of rooftop solar generation from air opacity, whose distribution is shown in Appendix Figure F14, is correlated with the socioeconomic characteristics of each ZIP code.

Out of the three socioeconomic characteristics included in this study, we find "% less than high school education" to be most strongly correlated with the AOD impact on solar generation. Over the ZIP codes included in our analysis, a 10% increase in the variable is associated with a 1% increase in AOD-attributable generation loss (Appendix Table F3). Similar trends are found for other two variables, % people of color and % low-income households, a trend which is also graphically ascertained (Figure 3, Appendix Figures F16 to F18).

[25] Appendix Figure S22 of Kelly et al. (2015) demonstrates the effect of 100 ton/year of primary PM2.5 emission from a single-source chemical transport model in California in July 2007. We take the very conservative assumption that the impact of PM2.5 emission is limited to the original ZIP code. We do not estimate the value of SO_2 emission (which can also increase PM2.5 concentration), as the impact of SO_2 emission reaches far beyond the ZIP code level (Baker et al., 2016); accurate estimation of SO_2 emission impact will require the use of integrated assessment models with chemical transport models, *à la* Muller and Mendelsohn (2009).

One possible driver of such correlation is the robust relationship between socioeconomic characteristics and air pollution exposure in United States, established by previous empirical studies. (e.g., Currie et al., 2023) Nevertheless, the socioeconomic characteristics still retain their significance even after controlling for the average PM2.5 level at each ZIP code (Appendix Table F3). This suggests that there is another factor correlated with socioeconomic characteristics that also affect the optical properties of the atmosphere, even after controlling for the PM2.5 conditions.

Appendix Table F4 identifies some possible drivers of this heterogeneity. The coefficients suggest that increased population density could lead to increased AOD-attributed solar generation loss keeping PM2.5 levels constant. This may be due to other urbanization-driven changes in air quality conditions not measured by PM2.5. Relatedly, lower wind speed and smaller day-to-day wind variability are correlated with higher losses from AOD, while average NO_2 concentration is not significant. It should be noted that these trends are associational, and the identification of its casual drivers is saved for future work.

7 Conclusion

This paper studies the impact of air pollution on residential rooftop solar generation in California. To this end, we take a two-part approach, where we first establish the causal relationship between satellite-measured air *opacity* and rooftop solar generation, and then link this air opacity with PM2.5 air *pollution*. The two-part approach is important for developing a causal estimate. We exploit day-to-day variations in wind and wildfire smoke to instrument for air opacity variations in order to mitigate concerns from omitted variables, measurement error, and reverse causality.

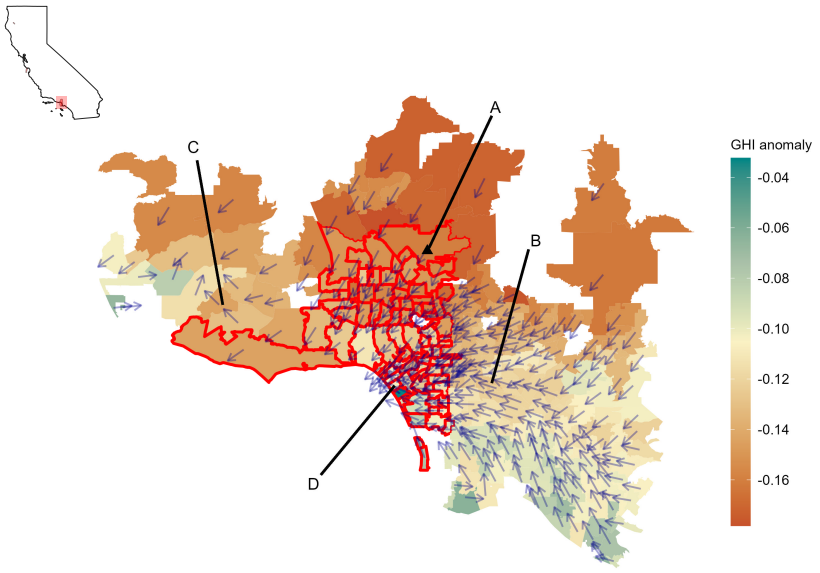
The results show that a single standard deviation increase in AOD is causally linked with an 9.4% reduction in solar generation. A $1\mu g/m^3$ increase in PM2.5 concentration has a different impact on the air opacity depending on the season. Expanding the analysis to the entire state of California, PM2.5 pollution causes a 459 GWh loss in residential rooftop solar generation, which amounts to 4% of total potential generation. The losses from PM2.5 appear to be disproportionately focused on disadvantaged communities.

This study has assessed the impact of air pollution at the producing unit level. Although this approach significantly simplifies the identification strategy, an analysis at the electric grid level could also provide important insights into the implications of air pollution impact on solar generation. For example, future studies could assess how large-scale events of air quality perturbation affect utility-scale outcomes, such as the fuel mix and wholesale electricity price.

In a similar vein, the impact of air pollution on generation of distributed solar generation across the broader electric grid is likely larger than our study results. Previous studies on residential electricity consumption have pointed to an elevated use of electricity under high air pollution conditions (He et al., 2020). The marginal effect of air pollution on the *net* generation of solar energy from distributed sources will thus be a sum of the effect of reduced solar generation from increased opacity (as identified in our study) and the increased electricity consumption effect.

Another avenue of future research could be to examine the long-run implications of air quality on solar generation. This study has shown that decarbonization policies can engender a positive feedback, where improved air quality from reduced fossil fuel use leads to an even larger share of solar energy in the fuel mix. Further research involving energy system models could analyze the long-run implications of this feedback loop, perhaps showing an example of an important path dependency that could affect social welfare.

(a) Changes in surface-level radiation



(b) Measure of downwind status

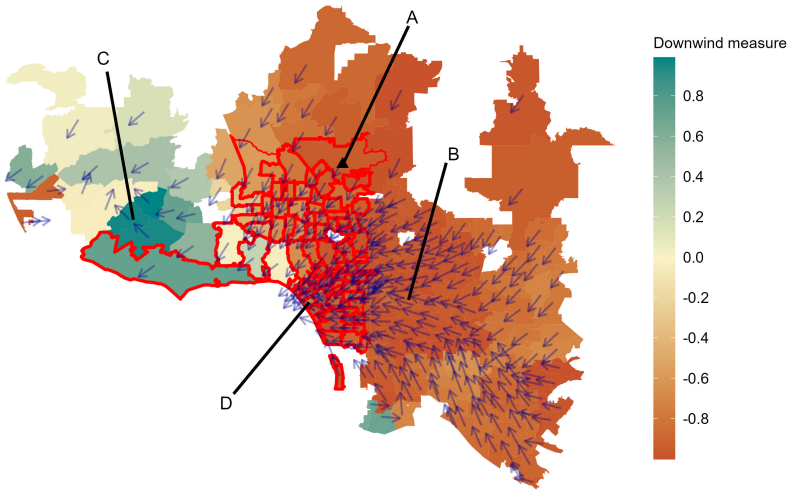


Figure 1. Wildfire smoke near Los Angeles on October 19, 2016

Notes: Figure shows the impact of wind conditions on wildfire smoke impact. Blue arrows show the wind direction, while wildfire-impacted ZIP codes are demarked in red. (a) shows reductions in surface level irradiance (GHI, Global Horizontal Irradiance) anomaly, calculated as the deviation of log-transformed GHI from ZIP-code average. (b) illustrates our cosine measure of downwind status (see text for definition). Positive value (blue) means downwind, negative value (red) means upwind from the smoke source. Wind data are from a gridded daily weather data set (Abatzoglou, 2013), and solar generation data from simulated surface-level irradiance data set (Sengupta et al., 2018). Wildfire data from Hazard Mapping System (Brey et al., 2018).

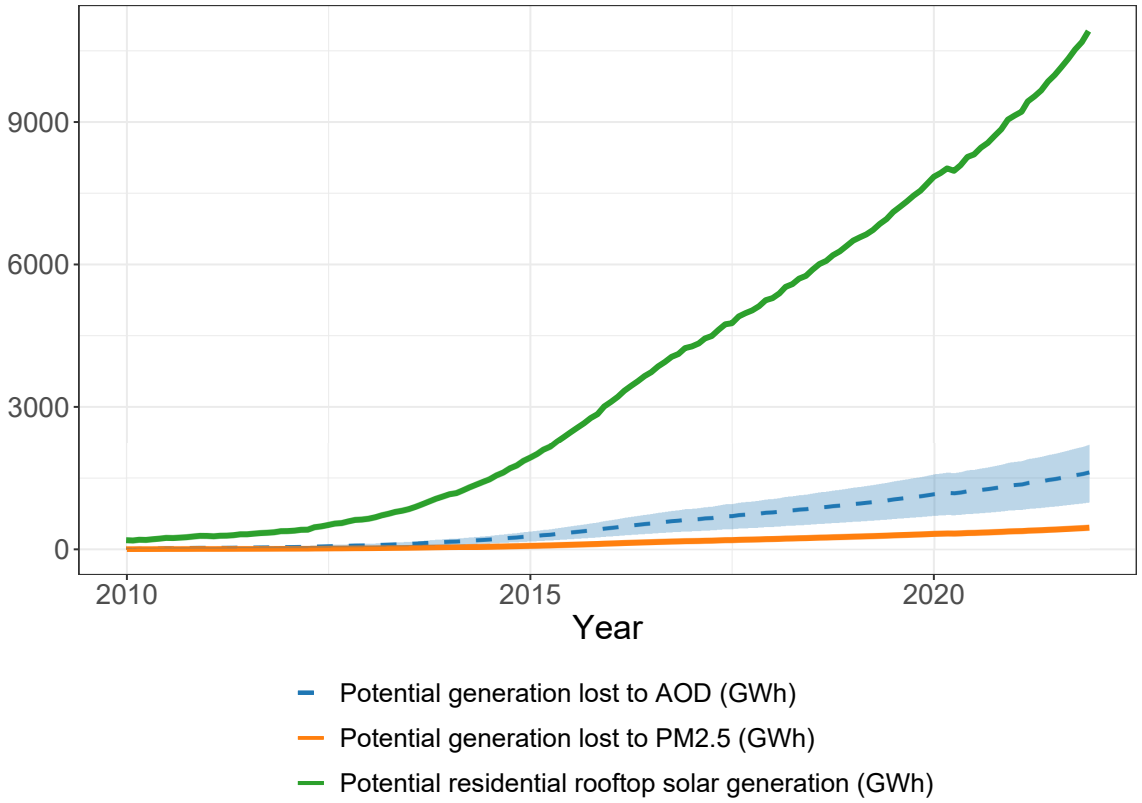


Figure 2. Loss of residential rooftop solar generation from AOD and PM2.5

Notes: Figure shows the trend of installed rooftop solar generation capacity in California residential units, and the loss of this capacity from AOD and PM2.5. Note that our analysis is confined to the 2010-2016 period.

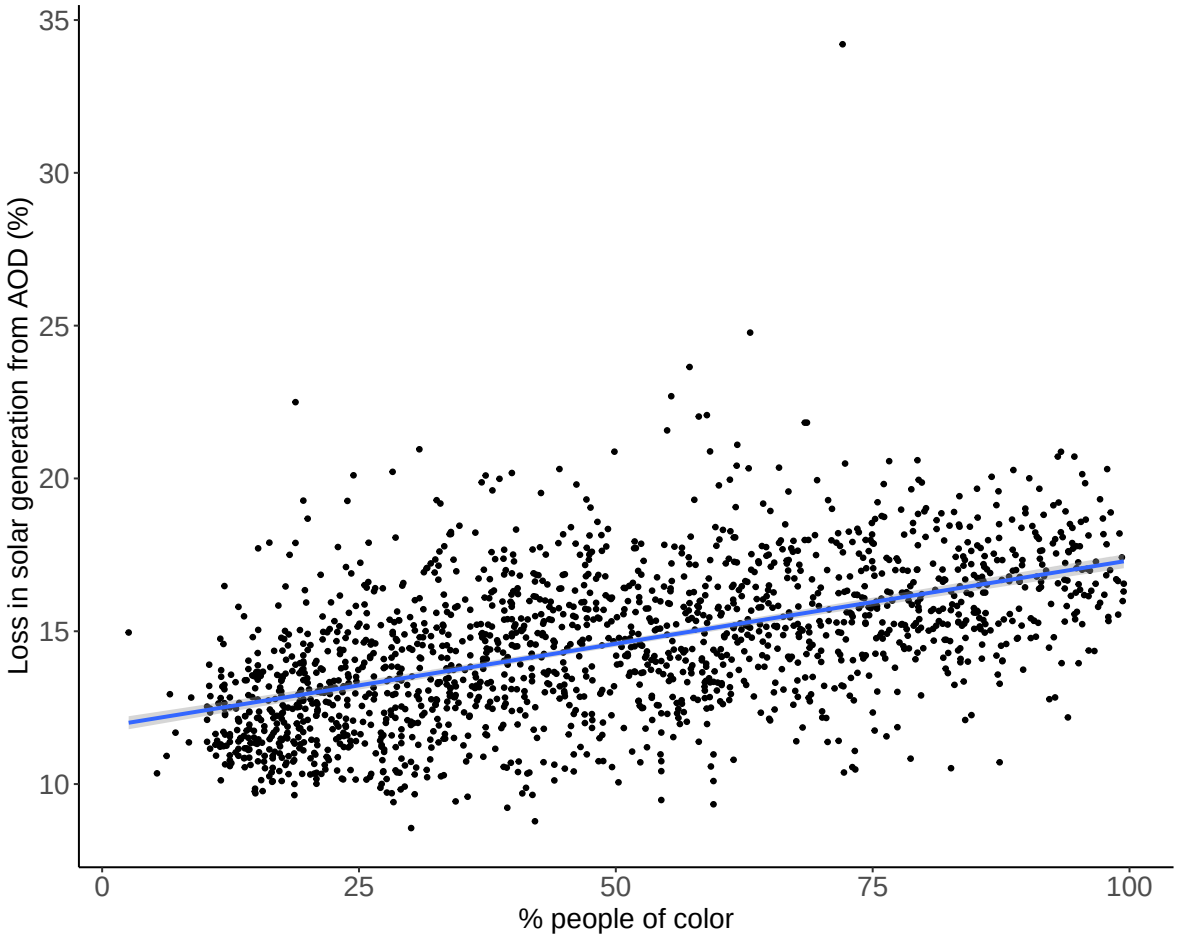


Figure 3. Unequal distribution of AOD-induced solar productivity loss

Notes: Figure places the % loss of solar generation from AOD on the y-axis, and the % of people of color in the ZIP code population in the x-axis. The blue line traces linear regression line. Similar figures for % low income household and % less than high school education included in the Appendix.

Table 1. Summary statistics

Variable (Unit)	N	Mean	St. Dev.	Min	Max
A. Solar generation & irradiance					
Three-day solar generation (kWh)	481,872	54.904	36.458	0.006	475.303
Three-day surface-level radiation (GHI, kW/m^2)	946,485	19,965.960	6,060.386	791.000	28,956.000
Three-day direct radiation (DNI, kW/m^2)	946,485	26,756.310	5,730.912	0.000	39,657.000
Three-day diffused radiation (DHI, kW/m^2)	946,485	3,430.767	1,176.502	786.000	11,750.000
B. Air opacity & air pollution					
PM2.5, downscaled (μg/m^3)	4,499,967	9.650	5.450	−21.140	322.450
PM2.5, measured (μg/m^3)	252,693	9.711	7.732	−6.000	270.100
PM10, measured (μg/m^3)	174,461	25.541	55.719	−53.000	8,187.000
AOD†	3,065,221	0.105	0.070	0.009	2.960
C. Wind & wildfire					
Wind direction†	4,608,759	3.228	1.580	0.292	19.854
Wind speed (m/s)	4,608,759	−0.126	0.684	−1.000	1.000
Smoke day†	4,608,759	0.612	0.487	0	1
D. Covariates					
NO_2 (mol/km^2)	3,698,997	2.603	4.257	−63.719	67.254
Precipitation (mm/day)	4,608,759	1.377	6.105	0.000	348.938
Maximum relative humidity (%)	4,608,759	78.336	19.759	6.348	100.000
Minimum relative humidity (%)	4,608,759	34.654	18.027	1.000	100.000
Minimum temperature (°C)	4,608,759	9.696	6.173	−23.775	36.547
Maximum temperature (°C)	4,608,759	22.935	8.014	−9.753	51.429
E. Socioeconomic characteristics					
% Low-income households	1,808	29.991	13.635	0.000	78.973
% People of color	1,808	49.319	24.631	2.580	99.444
% Less than high school education	1,808	9.868	6.854	0.000	40.458

Notes: †: Unitless variables. Table summarizes the data used for analysis. For wind direction, we take the cosine of degree measure to assure continuity.

Table 2. Effect of air opacity on solar generation

	Dependent variable:					
	Rooftop solar generation		Surface irradiance		Placebo	
	AOD	PM2.5	Limited	Expanded	Solar	Irradiance
	(1)	(2)	(3)	(4)	(5)	(6)
A. OLS estimates						
AOD	-0.256^{***}		-0.192^{***}	-0.129^{**}	0.140	0.090
	(0.040)		(0.027)	(0.035)	(0.112)	(0.053)
PM2.5		-0.008^{***}				
		(0.001)				
Adj. R^2	0.864	0.837	0.978	0.980	0.828	0.916
B. IV estimates						
AOD	-1.559^{***}		-0.961^{***}	-0.950^{***}	-0.969	0.449
	(0.324)		(0.202)	(0.135)	(0.558)	(0.273)
PM2.5		-0.016^{***}				
		(0.003)				
F-statistic	89.059	359.018	32.629	167.404	86.924	20.204
Adj. R^2	0.855	0.854	0.958	0.957	0.681	0.932
Dep. var. mean	3.897	3.795	9.849	9.883	2.712	8.706
AOD/PM2.5 mean	0.105	10.042	0.097	0.096	0.105	0.108
AOD/PM2.5 SD	0.063	4.448	0.055	0.060	0.063	0.064
N	202,079	282,278	61,589	535,078	195,566	51,722

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Table reports OLS and IV estimation results of equation (2). All columns use three-day measures of the dependent variable to account for lagged effects. Column 1 estimates the equation (2). Column 2 substitutes AOD by PM2.5 as the endogenous variable of interest. Placebo tests use dependent variables 30 days into the past. Columns 3 and 4 share the same dependent variable, but column 3 is limited to ZIP codes included in the rooftop solar data; column 4 covers the entire state of California (Appendix Figure F2). Columns 5 and 6 perform placebo tests: we use rooftop solar generation and surface irradiance thirty days into the past. All results include ZIP code-year and week fixed effects. Three-day diurnal temperature, precipitation, relative humidity, NO_2 concentration, and their polynomials, are used as covariates. Standard errors two-way clustered by year-week and ZIP code. For panel B, Kleibergen-Paap cluster-robust F-statistics for the first-stage regression are presented.

Table 3. Effect of air opacity on surface radiation subcomponents

	Dependent variable:					
	Direct irradiance		Diffused irradiance		Placebo	
	Limited	Expanded	Limited	Expanded	Direct	Diffused
	(1)	(2)	(3)	(4)	(5)	(6)
A. OLS estimates						
AOD	-0.371^{***}	-0.265^{***}	0.600^{***}	0.486^{***}	0.063	-0.001
	(0.037)	(0.051)	(0.032)	(0.058)	(0.048)	(0.084)
Adj. R^2	0.858	0.870	0.929	0.903	0.579	0.775
B. IV estimates						
AOD	-1.872^{***}	-1.934^{***}	2.011^{***}	2.281^{***}	1.050	-0.212
	(0.122)	(0.126)	(0.454)	(0.433)	(0.542)	(0.336)
F-statistic	32.629	167.404	32.629	167.404	41.875	41.884
Adj. R^2	0.684	0.652	0.873	0.818	0.534	0.793
Dep. var. mean	10.177	10.209	8.068	8.071	8.994	7.004
AOD mean	0.097	0.096	0.097	0.096	0.107	0.107
AOD SD	0.055	0.060	0.055	0.060	0.064	0.064
N	61,589	535,078	61,589	535,078	97,186	97,203

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Table reports OLS and IV estimation results of equation (2), with diffused and direct solar radiation as dependent variable. All columns use three-day measures of the dependent variables to account for lagged effects. "Limited" and "Expanded" refer to rooftop solar data-included ZIP codes and California-wide ZIP codes (Appendix Figure F2). Placebo tests use dependent variables 30 days into the future. All results include ZIP code, year, and week fixed effects. Three-day diurnal temperature, precipitation, relative humidity, and their polynomials, are used as covariates. Standard errors two-way clustered by year-month and ZIP code. For panel B, Kleibergen-Paap cluster-robust F-statistics for the first-stage regression are presented.

Table 4. Effect of PM2.5 pollution on atmospheric opacity

	(1)	(2)	(3)	(4)	(5)	(6)
A. OLS estimates						
PM2.5	0.0019***	0.0018***	0.0041***	0.0027***	0.0021***	0.0021***
	(0.0002)	(0.0004)	(0.0009)	(0.0003)	(0.0002)	(0.0002)
Adj. R^2	0.3731	0.3589	0.2921	0.3211	0.3103	0.3165
B. IV estimates						
PM2.5	0.0027***	0.0023*	0.0061*	0.0028***	0.0025***	0.0025***
	(0.0002)	(0.0011)	(0.0029)	(0.0005)	(0.0005)	(0.0006)
F-statistic	416.8860	224.5905	45.6672	396.8212	1364.1863	558.9419
Adj. R^2	0.3599	0.3580	0.2778	0.3209	0.3090	0.1998
C. Lagged & leaded OLS estimates						
PM2.5	0.0019***	0.0019***	0.0038	0.0025***	0.0023***	0.0022***
	(0.0003)	(0.0005)	(0.0020)	(0.0004)	(0.0003)	(0.0003)
						(0.0000)
PM2.5, t-1	−0.0000	−0.0001	0.0003	−0.0002	−0.0002	−0.0001
	(0.0001)	(0.0005)	(0.0009)	(0.0004)	(0.0001)	(0.0001)
PM2.5, t+1	−0.0000	−0.0002	0.0002	0.0005	−0.0001	−0.0001
	(0.0001)	(0.0004)	(0.0007)	(0.0005)	(0.0002)	(0.0003)
Adj. R^2	0.3801	0.3712	0.3036	0.3351	0.3200	0.3271
Sample subsets	Winter	Spring	Summer	Fall	All	All
PM10 included	No	No	No	No	No	Yes
AOD mean	0.0823	0.1235	0.1284	0.1022	0.1113	0.1178
PM2.5 mean	12.5131	8.8217	9.9011	10.4735	10.3114	10.8499
PM2.5 SD	10.9397	5.3244	5.9181	7.3374	7.5113	8.0218
N	28,998	34,665	46,388	40,233	150,920	58,312

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Table reports OLS and IV estimation results of equation (4). All results include ZIP code, year, and week fixed effects. Panel B uses wind conditions to instrument for PM2.5. Panel C includes leaded and lagged PM2.5. Diurnal temperature, precipitation, relative humidity, and their polynomials, are used as covariates. Columns 1 to 4 use seasonal subsamples, while columns 5 and 6 the entire sample. Column 6 includes PM10 levels as covariate. Standard errors two-way clustered by year-month and ZIP code. For panel B, Kleibergen-Paap cluster-robust F-statistics for the first-stage regression are presented.

Table 5. Lagged effects of air opacity

	(1)	(2)	(3)	(4)	(5)
AOD	-1.961^{***}	-1.520^{***}	-1.559^{***}	-0.777^{**}	-0.694^{*}
	(0.255)	(0.237)	(0.324)	(0.271)	(0.340)
F-statistic	335.659	106.345	89.059	37.560	27.463
Adj. R^2	0.742	0.840	0.840	0.873	0.887
Solar generation	Same-day	2-day	3-day	4-day	5-day
Dep.var. mean	2.832	3.502	3.897	4.172	4.385
AOD mean	0.104	0.103	0.105	0.102	0.101
AOD SD	0.063	0.063	0.063	0.063	0.064
N	264,408	198,940	202,079	126,376	99,968

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Table reports IV estimation results of equation (4) over different summation horizons. All results include ZIP code, year, and week fixed effects. Standard errors two-way clustered by year-month and ZIP code. Kleibergen-Paap cluster-robust F-statistics for the first-stage regression are presented.

Table 6. Robustness check: Key results

	Alternative data set			First-stage		Weighted
	(1)	(2)	(3)	(4)	(5)	(6)
AOD	-1.55^{***}	-2.11^{***}	-1.42^{***}	-1.57^{***}	-1.55^{***}	-1.51^{***}
	(0.32)	(0.40)	(0.29)	(0.29)	(0.31)	(0.32)
F-statistic	91.05	59.16	104.06	3.49	95.23	62.14
Adj. R^2	0.84	0.84	0.84	0.84	0.84	0.86
Dep. var. mean	3.91	3.92	3.89	3.90	3.90	3.90
AOD mean	0.10	0.09	0.10	0.10	0.10	0.10
AOD SD	0.06	0.05	0.06	0.06	0.06	0.06
N	206,002	158,197	263,090	202,079	202,095	202,079

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Column 1 expands the number of solar panels included. Column 2 limits the data to "best" AOD retrievals. Column 3 removes negative NO_2 retrievals. Column 4 uses wind direction bins, rather than polynomials, in the first-stage regression (3). Column 5 defines "smoke day" for ZIP codes within 100km radius of each polygon. Column 6 uses average generation at each ZIP code as weights. See Appendix for additional robustness checks.

References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology 33*(1), 121–131.

Baker, K. R., R. A. Kotchenruther, and R. C. Hudman (2016). Estimating ozone and secondary pm2.5 impacts from hypothetical single source emissions in the central and eastern united states. *Atmospheric Pollution Research 7*(1), 122–133.

Beer, A. (1852). Bestimmung der absorption des rothen lichts in farbigen flussigkeiten. *Annalen der Physik und Chemie 162*, 78–88.

Bergin, M. H., C. Ghoroi, D. Dixit, J. J. Schauer, and D. T. Shindell (2017). Large reductions in solar energy production due to dust and particulate air pollution. *Environmental Science & Technology Letters 4*(8), 339–344.

Borenstein, S. and J. B. Bushnell (2022). Do two electricity pricing wrongs make a right? cost recovery, externalities, and efficiency. *American Economic Journal: Economic Policy 14*(4), 80–110.

Borgschulte, M., D. Molitor, and E. Y. Zou (2022). Air pollution and the labor market: Evidence from wildfire smoke. *Review of Economics and Statistics*, 1–46.

Brey, S. J., M. Ruminski, S. A. Atwood, and E. V. Fischer (2018). Connecting smoke plumes to sources using hazard mapping system (hms) smoke and fire location data over north america. *Atmospheric Chemistry and Physics 18*(3), 1745–1761.

California Energy Commission (2022). California power plants.

California Public Utilities Comission (2016). Csi 15-minute interval data.

California Public Utilities Comission (2022). Csi working data set.

Chay, K. Y. and M. Greenstone (2003). The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *The quarterly journal of economics 118*(3), 1121–1167.

Chen, S., C. Guo, and X. Huang (2018). Air pollution, student health, and school absences: Evidence from china. *Journal of Environmental Economics and Management 92*, 465–497.

Clean Power Research (2018). *Developing a Comprehensive, System-Wide Forecast to Support High-Penetration Solar*.

Currie, J., J. Voorheis, and R. Walker (2023). What caused racial disparities in particulate exposure to fall? new evidence from the clean air act and satellite-based measures of air quality. *American Economic Review 113*(1), 71–97.

Deetjen, T. A. and I. L. Azevedo (2020). Climate and health benefits of rapid coal-to-gas fuel switching in the us power sector offset methane leakage and production cost increases. *Environmental Science & Technology 54*(18), 11494–11505.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review 109*(12), 4178–4219.

Donaldson, D. L., D. M. Piper, and D. Jayaweera (2021). Temporal solar photovoltaic generation capacity reduction from wildfire smoke. *IEEE Access 9*, 79841–79852.

Garaga, R. and S. H. Kota (2018). Characterization of pm10 and impact on human health during the annual festival of lights (diwali). *Journal of Health and Pollution 8*(20).

Gillingham, K. T. and P. Huang (2020). Long-run environmental and economic impacts of electrifying waterborne shipping in the united states. *Environmental Science & Technology 54*(16), 9824–9833.

Gómez-Amo, J., M. Freile-Aranda, J. Camarasa, V. Estellés, M. Utrillas, and J. Martínez-Lozano (2019). Empirical estimates of the radiative impact of an unusually extreme dust and wildfire episode on the performance of a photovoltaic plant in western mediterranean. *Applied Energy 235*, 1226–1234.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment 202*, 18–27.

Grainger, C. and A. Schreiber (2019). Discrimination in ambient air pollution monitoring? In *AEA Papers and Proceedings*, Volume 109, pp. 277–282. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Habte, A., M. Sengupta, and A. Lopez (2017). Evaluation of the national solar radiation database (nsrdb): 1998-2015. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States).

He, P., J. Liang, Y. L. Qiu, Q. Li, and B. Xing (2020). Increase in domestic electricity consumption from particulate air pollution. *Nature Energy 5*(12), 985–995.

Holben, B. N., D. Tanré, A. Smirnov, T. Eck, I. Slutsker, N. Abuhassan, W. Newcomb, J. Schafer, B. Chatenet, F. Lavenu, et al. (2001). An emerging ground-based aerosol climatology: Aerosol optical depth from aeronet. *Journal of Geophysical Research: Atmospheres 106*(D11), 12067–12097.

Hu, J., H. Zhang, S. Chen, Q. Ying, C. Wiedinmyer, F. Vandenberghe, and M. J. Kleeman (2014). Identifying pm2. 5 and pm0. 1 sources for epidemiological studies in california. *Environmental science & technology 48*(9), 4980–4990.

Johnsen, R., J. LaRiviere, and H. Wolff (2019). Fracking, coal, and air quality. *Journal of the Association of Environmental and Resource Economists 6*(5), 1001–1037.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. Hnilo, M. Fiorino, and G. Potter (2002). Ncep-doe amip-ii reanalysis (r-2). *Bulletin of the American Meteorological Society 83*(11), 1631–1644.

Kelly, J. T., K. R. Baker, S. L. Napelenok, and S. J. Roselle (2015). Examining single-source secondary impacts estimated from brute-force, decoupled direct method, and advanced plume treatment approaches. *Atmospheric Environment 111*, 10–19.

Knittel, C. R., D. L. Miller, and N. J. Sanders (2016). Caution, drivers! children present: Traffic, pollution, and infant health. *Review of Economics and Statistics 98*(2), 350–366.

Koelemeijer, R., C. Homan, and J. Matthijsen (2006). Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over europe. *Atmospheric Environment 40*(27), 5304–5315.

Krotkov, N. A., L. N. Lamsal, S. V. Marchenko, E. J. Bucsela, W. H. Swartz, and J. Joiner (2019). Omi/aura nitrogen dioxide (no2) total and tropospheric column 1-orbit l2 swath 13x24 km v003.

Lambert, J. H. (1760). *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Eberhardt Klett.

Landrigan, P. J., R. Fuller, N. J. Acosta, O. Adeyi, R. Arnold, A. B. Baldé, R. Bertollini, S. Bose-O'Reilly, J. I. Boufford, P. N. Breysse, et al. (2018). The lancet commission on pollution and health. *The Lancet 391*(10119), 462–512.

Li, X., F. Wagner, W. Peng, J. Yang, and D. L. Mauzerall (2017). Reduction of solar photovoltaic resources due to air pollution in china. *Proceedings of the National Academy of Sciences 114*(45), 11867–11872.

Liu, Y., J. A. Sarnat, V. Kilaru, D. J. Jacob, and P. Koutrakis (2005). Estimating ground-level pm2. 5 in the eastern united states using satellite remote sensing. *Environmental science & technology 39*(9), 3269–3278.

Ma, Z., S. Dey, S. Christopher, R. Liu, J. Bi, P. Balyan, and Y. Liu (2022). A review of statistical methods used for developing large-scale and long-term pm2. 5 models from satellite data. *Remote Sensing of Environment 269*, 112827.

McClure, C. D. and D. A. Jaffe (2018). Us particulate matter air quality improves except in wildfire-prone areas. *Proceedings of the National Academy of Sciences 115*(31), 7901–7906.

Metaxoglou, K. and A. Smith (2020). Productivity spillovers from pollution reduction: reducing coal use increases crop yields. *American Journal of Agricultural Economics 102*(1), 259–280.

Moretti, E. and M. Neidell (2011). Pollution, health, and avoidance behavior evidence from the ports of los angeles. *Journal of Human Resources 46*(1), 154–175.

Muller, N. Z. and R. Mendelsohn (2009). Efficient pollution regulation: getting the prices right. *American Economic Review 99*(5), 1714–39.

National Renewable Energy Laboratory (2021). Annual technology baseline.

Perry, M. and A. Troccoli (2015). Impact of a fire burn on solar irradiance and pv power. *Solar Energy 114*, 167–173.

Peters, I., S. Karthik, H. Liu, T. Buonassisi, and A. Nobre (2018). Urban haze and photovoltaics. *Energy & Environmental Science 11*(10), 3043–3054.

Proctor, J. (2021). Atmospheric opacity has a nonlinear effect on global crop yields. *Nature Food 2*(3), 166–173.

Reid, C. E., E. M. Considine, M. M. Maestas, and G. Li (2021). Daily pm2. 5 concentration estimates by county, zip code, and census tract in 11 western states 2008–2018. *Scientific data 8*(1), 1–15.

Remer, L. A., D. Tanré, and K. Y. J (2008). Algorithm for remote sensing of tropospheric aerosol from modis: Collection 5. Technical report, National Aeronautics and Space Administration.

Sager, L. (2019). Estimating the effect of air pollution on road safety using atmospheric temperature inversions. *Journal of Environmental Economics and Management 98*, 102250.

Schlenker, W. and W. R. Walker (2016). Airports, air pollution, and contemporaneous health. *The Review of Economic Studies 83*(2), 768–809.

Seinfeld, J. and S. Pandis (2016). *Atmospheric chemistry and physics*. John Wiley & Sons.

Sengupta, M., Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby (2018). The national solar radiation data base (nsrdb). *Renewable and sustainable energy reviews 89*, 51–60.

Sergi, B. J., P. J. Adams, N. Z. Muller, A. L. Robinson, S. J. Davis, J. D. Marshall, and I. L. Azevedo (2020). Optimizing emissions reductions from the us power sector for climate and health benefits. *Environmental Science & Technology 54*(12), 7513–7523.

Sexton, S., A. J. Kirkpatrick, R. I. Harris, and N. Z. Muller (2021). Heterogeneous solar capacity benefits, appropriability, and the costs of suboptimal siting. *Journal of the Association of Environmental and Resource Economists 8*(6), 1209–1244.

Son, J., S. Jeong, H. Park, and C.-E. Park (2020). The effect of particulate matter on solar photovoltaic power generation over the republic of korea. *Environmental Research Letters 15*(8), 084004.

Styszko, K., M. Jaszczur, J. Teneta, Q. Hassan, P. Burzyńska, E. Marcinek, N. Lopian, and L. Samek (2019). An analysis of the dust deposition on solar photovoltaic modules. *Environmental Science and Pollution Research 26*(9), 8393–8401.

US EPA (2022). Ejscreen 2021 version.

Wei, J., Z. Li, Y. Peng, and L. Sun (2019). Modis collection 6.1 aerosol optical depth products over land and ocean: validation and comparison. *Atmospheric Environment 201*, 428–440.

Zhang, L., X. Yi, M. Zhao, and Z. Gu (2021). Reduction in solar photovoltaic generation due to aerosol pollution in megacities in western china during 2014 to 2018. *Indoor and Built Environment 30*(8), 1286–1294.

Appendix

A Physical theories of air opacity, PM2.5, and solar generation

This section introduces the fundamental theories of (atmospheric) physics that underlie the analysis introduced in our main text. We first show a simple derivation of equation (1) from Beer-Lambert Law. We then briefly summarize the physical theories of AOD-PM2.5 relationship, with a specific focus on the origins of specification.

A.1 Atmospheric optical density and solar generation

We here show how the fundamental principles of optical extinction can be translated into the theoretical results shown in equation (1) in main text, and how the optical depth τ is measured by satellites. According to the Beer-Lambert Law, the rate of optical extinction of a radiation traversing the atmosphere is a linear function of the irradiance itself:

$$\frac{dF(x,\lambda)}{dx} = -b(x,\lambda)F(x,\lambda) \tag{A1}$$

where dx is the length of atmosphere traversed, $F(x,\lambda)$ is the spectral radiant flux density (irradiance) of the radiation at wavelength λ, with unit W/m^2, and $b(x,\lambda)$ is the spectral extinction coefficient of the atmosphere. The "optical depth" of the atmosphere, denoted as τ, is obtained by integrating the extinction coefficient from the top of the atmosphere (TOA) to the surface:

$$\tau_{0,\lambda} = \int_0^{z_{TOA}} b(z',\lambda)dz' \tag{A2}$$

where $b(z',\lambda)$ is the extinction coefficient at altitude z', and z_{TOA} the altitude of the TOA. Using this definition of optical depth, we reach the following relationship shown in equation (1), which links spectral irradiance at the surface to that at the TOA:

$$F_{0,\lambda} = F_{TOA,\lambda}\exp(-\tau_{0,\lambda}) \tag{A3}$$

where $F_{0,\lambda}$ and $F_{TOA,\lambda}$ are the spectral radiant flux density (irradiance) at the ground surface and at the TOA, respectively. The $\tau_{0,\lambda}$ is hence a measurement of the atmospheric opacity in general. Atmospheric physics measures the impact of aerosol on τ by separating the value into aerosol optical depth (AOD) and gaseous molecules. The AOD hence refers to the value of $\tau_{0,\lambda}$ attributable to the aerosol molecules in the atmosphere.

Traditionally, AOD was measured from *in situ* observatories involving spectral radiome-

ters (Holben et al., 2001), which was severely limited in its geographical coverage. The advent of spaceborne multispectral sensors, such as the MODIS sensor on board the Terra and Aqua satellites, significantly relaxed this problem by providing a daily measurement of AOD across the entire globe at a 1km spatial resolution. Validation of MODIS AOD measurements from ground *in situ* measurements have consistently confirmed its strong performance over different geographies (Wei et al., 2019).

MODIS algorithm of AOD measurement largely mimics that used in *in situ* observatories; the process involves combining retrievals from different spectral bands, and using this data to remove the share of non-aerosol molecules in the columnar $\tau_{0,\lambda}$ (Remer et al., 2008). To be more specific, MODIS takes an approach called "look-up table inversion," where observations from different spectrum bands (in land AOD, retrievals from the 0.47μm, 0.66μm, and 2.1μm) are combined. Using such measurements, the MODIS sensor backtracks the composition of atmospheric aerosol. In this backtracking, *a priori* knowledge into the optical characteristics of aerosols in each region and season is used to calibrate the retrievals (e.g., aerosols in California are mostly assumed to be "non-absorbing" in the fall and winter.).

Lastly, a brief discussion on the empirical implications of this theoretical conclusion is in order. The rather simple theoretical result, shown in equation (A3), may suggest that given a log-transformed irradiance measure as the outcome variable, the coefficient for τ must be 1. However, whether this theoretical notion must be true in our setting is a matter of empirical verification, for two reasons. First, our outcome of interest is rooftop solar generation, not surface irradiance. Although between-region difference in surface-level irradiance is a good proxy of solar panel productivity, the performance of an individual rooftop solar panel under adverse air quality is also a function of the material and technology of panel production. We are also interested in the lagged effects of air opacity. Such differences may lead to deviations of the true coefficient from 1.

Second, note that atmospheric physics theory presented here only presents the relationship between *spectral* optical depth and *spectal* irradiance. While such a relationship is arguably sufficient to justify a log-linear specification, it remains to be empirically tested whether this spectrum-specific relationship can be extended to the general relationship between AOD and solar generation.

A.2 PM2.5 and atmospheric optical density

We now summarize the physical theories of the relationship between AOD and PM2.5. The relationship was first empirically proposed by Liu et al. (2005) and was subsequently buttressed by such theoretical breakthroughs as Koelemeijer et al. (2006).

Equation (A2) suggests that the AOD can be obtained from the vertical integration of the

extinction coefficient $b(z, \lambda)$. Theoretical and empirical analyses of vertical profiles of $b(z, \lambda)$ have confirmed that the extinction coefficient in general shows a negatively exponential distribution. This suggests that:

$$\tau_{0,\lambda} = \int_0^{z_{TOA}} b(0, \lambda) \exp(-z/H) dz \approx H b(0, \lambda) \tag{A4}$$

where H is the e-fold scale height. There is thus an approximately linear relationship between the AOD and the extinction coefficient at the surface. In the following discussions, we focus on AOD and extinction coefficient at wavelength $0.55 \mu m$; we hence drop the λ notation for conciseness.

Under almost all conditions, ambient particulate matter concentration shows a bimodal distribution with regards to size. The two modes are placed around diameters of 2.5μm and 10μm, respectively called "fine" and "coarse" modes. PM2.5 is a measurement of the mass density ($\mu g/m^3$) of particulate matters in the former mode. Holding all other factors (e.g., humidity, temperature) constant, the relationship between $b(0)$ and PM2.5 can be represented as

$$b(0) = \alpha c_{\text{PM2.5}} \text{PM2.5} \tag{A5}$$

where $c_{\text{PM2.5}}$ refers to the share of coarse and fine PM molecules in a given mass of air and α is the "mass extinction efficiency" (MEE) that converts the units to match the extinction coefficient.[1]. Using (A4), we get

$$\tau_0 = H \alpha c_{\text{PM2.5}} \text{PM2.5} \tag{A6}$$

This equation forms the basis of the linear specification we use in the main text.

[1]It is also possible to further convert the expression (A5) by multiplying the "hygroscopic growth factor" $\left(\frac{1-\text{RH}}{1-\text{RH}_0}\right)^{-\gamma}$ to the right hand side of the equation, where RH= relative humidity. We choose to model the impact of relative humidity as a polynomial transformation instead.

B Reverse causality in OLS estimates

Appendix Table B1. Reverse causality in OLS estimates

	OLS			IV	
	2010-2016	2010-2012	2013-2016	Gas	Non-Gas
	(1)	(2)	(3)	(4)	(5)
AOD	-0.294^{***}	-0.369^{***}	-0.262^{**}	-1.445^{***}	-1.542^{***}
	(0.048)	(0.031)	(0.050)	(0.170)	(0.162)
AOD*Gas	-0.088^{**}	-0.063	-0.087^{**}		
	(0.024)	(0.028)	(0.027)		
F-statistic				174.479	1806.721
Dep. var. mean	8.719	8.758	8.70	8.71	8.72
N	1361487	464056	897431	113219	1248268
Adj. R^2	0.963	0.955	0.967	0.940	0.924

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Columns 1 to 3 use OLS estimates of the equation (B1), while columns 4 and 5 present 2SLS estimates of β for equation (2) after dividing the samples into ZIPs with gas plants and those without.

In this section, we present suggestive evidence of reverse causality in the OLS estimates of the solar generation-air opacity model (2). As mentioned in the main text, our primary source of concern is that solar generation is representative of the solar yield in that ZIP code *per se*, and that an increased share of solar in the fuel mix may in turn affect AOD conditions. Such air pollution effect will be particularly strong in ZIP codes with gas plants, which was the primary source of thermal electricity generation in California throughout our study period.[2]

To test this possibility, we first obtain data on the location of gas-powered plants in California (California Energy Commission, 2022). We then define the indicator variable $\text{Gas}_i = 1$ if ZIP code i houses a gas power plant. The OLS estimation is:

$$\log(y_{iswt}) = \beta \text{AOD}_{iswt} + \theta \text{AOD}_{iswt}\text{Gas}_i + X'_{iswt}\lambda + \alpha_{is} + \gamma_w + \varepsilon_{iswt} \tag{B1}$$

where the notations are identical to (2). A negative θ would suggest that in ZIP codes with gas plants, the OLS estimates a weaker relationship between AOD and solar conditions; this

[2]Other important sources of fuel include hydroelectric power, nuclear energy, wind, and geothermal.

will be the case when solar generation reduces gas-powered electricity generation in these ZIP codes and reduce atmospheric opacity. To maximize statistical power, we use GHI, not solar generation, as the dependent variable here. We also use same-day GHI, as we expect the grid-level impacts of solar conditions to be limited to that given day.

Appendix Table B1 presents the results of this estimation. The results show that indeed, anomalies in AOD form a stronger negative association with the solar yield conditions in ZIP codes with gas power plants. Further, this estimate seems to be driven by the samples in years 2013-2016, when the share of solar in the fuel mix increased, along with its capacity to change fossil fuel combustion. This suggests that the θ estimates indeed are driven by reverse causality, and serves as an indirect proof that OLS estimations of (2) are vulnerable to biases from reverse causality. The IV estimates in columns 4 and 5 do not show heterogeneity.

C Violation of strict exogeneity

In this section, we present the consequences of strict exogeneity violation, coupled with serial correlation in the variable of interest. This is of particular concern in the context of our AOD-PM2.5 analysis, as daily PM2.5 may have lagged or leaded effects on daily AOD, and day-to-day variations in PM2.5 may show serial correlation.

We first show that under serial correlation and leaded/lagged effects, misspecification of the true DGP will lead to estimates that differ from the true coefficient. This implies that if we do not observe significant changes in the coefficients as we move from the simple model (4) to the augmented model (6), then (1) the true DGP does not have leaded or lagged effects, or (2) there is no serial correlation in PM2.5 data. In C.2, we find strong evidence suggesting that the PM2.5 in our dataset does indeed show serial correlation. The comparison of coefficients thus provide an indirect way of testing leaded/lagged effects in PM2.5 impact on AOD.

C.1 Motivation

C.1.1 Model set-up

We first assess the implications of strict exogeneity violation in the panel fixed effects framework. Throughout this exercise, we will assume the following lagged-and-leaded Data Generation Processes (DGP):

$$y_{it} = \zeta_0 X_{i,t-1} + \zeta_1 X_{it} + \zeta_2 X_{i,t+1} + \alpha_i + u_{it} \tag{C1}$$

where α_i is the site-fixed effect, X_{it} the regressor of interest in site i time t, and e_{it}, u_{it} the error term. Note that β_1 and ζ_1 are our coefficients of interest for both DGPs. We assume that strict exogeneity of u_{it} stands under each DGP:

$$E(u_{it}|X_{i,t-1}, X_{it}, X_{i,t+1}, \alpha_i) = 0, t = 2, \ldots, T-1 \tag{C2}$$

Now, assume that the X_{it} themselves are serially correlated. We assume the AR(1) model throughout for simplicity:

$$X_{it} = \phi X_{i,t-1} + \delta_i + \varepsilon_{it} \tag{C3}$$

where ε_{it} is white noise with mean 0 and variance σ_i^2, and δ_i is the site-fixed intercept of the AR(1) process that assures $E(X_{it}) = \frac{\delta_i}{1-\phi}$ under $|\phi| < 1$. We assume weak stationarity.

Lemma. Under DGP (C1) and AR(1) X_{it} with white noise (C3), $E(X_{it}\varepsilon_{i,s'}) = 0, \forall t \in \{1, \ldots, T\}, s' > t$.

Proof. Note that $X_{it} - \frac{\delta_i}{1-\phi}$ can be expressed as a linear combination of $X_{is}, s \leq t$:

$$X_{it} - \frac{\delta_i}{1-\phi} = \sum_{a=0}^{\infty} \phi^a \varepsilon_{i,t-a} \tag{C4}$$

From the independence of ε_{it},

$$E\left(\left[X_{it} - \frac{\delta_i}{1-\phi}\right]\varepsilon_{i,s'}\right) = 0$$

□

C.1.2 Asymptotic bias under misspecification

We show that under lagged-and-leaded true DGP (C1), misspecification with no leads and lags results in biased estimates:

Theorem (Asymptotic bias under misspecification: True DGP (C1))**.** Let the true DGP (C1) involve lags and leads of X_{it} . Assume strict exogeneity (C2). Then a misspecified model with only X_{it} on the RHS

$$y_{it} = \beta_1 X_{it} + \tilde{\alpha}_i + \tilde{e}_{it} \tag{C5}$$

leads to asymptotic bias:

$$\hat{\beta}_{1,FE} - \zeta_1 \to_p \frac{\zeta_0}{\phi} + \zeta_2\phi - \zeta_0\frac{\sigma_i^2}{\phi} \text{ as } NT \to \infty$$

where σ_i^2 is the variance of the error term in the AR(1) process (C3).

Proof. First see that lagged and leaded effects (C1), coupled with AR(1) process (C4), yield:

$$y_{it} = \left(\frac{\zeta_0}{\phi} + \zeta_1 + \zeta_2\phi\right) X_{it} + \left(\zeta_2 - \frac{\zeta_0}{\phi}\right)\delta_i + \alpha_i + \zeta_2\varepsilon_{i,t+1} - \frac{\zeta_0}{\phi}\varepsilon_{it} + u_{it} \tag{C6}$$

Now, see that a "within" transformation using demeaning of both sides of (C6) would result in

$$\ddot{y}_{it} = \left(\frac{\zeta_0}{\phi} + \zeta_1 + \zeta_2\phi\right) \ddot{X}_{it} + \zeta_2\ddot{\varepsilon}_{i,t+1} - \frac{\zeta_0}{\phi}\ddot{\varepsilon}_{it} + \ddot{u}_{it} \tag{C7}$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_i, \ddot{X}_{it} = X_{it} - \bar{X}_i$. $\bar{X}_i$ and $\bar{y}_i$ are the time-averaged value of X_{it} and y_{it},

respectively. The FE estimate of β_1 is obtained from the following estimator:

$$\hat{\beta}_{1,FE} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}'_{it}\ddot{y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}'_{it}\ddot{X}_{it}} = \tilde{\beta}_1 + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\beta_2\ddot{X}'_{it}\ddot{\varepsilon}_{i,t+1} - \frac{\beta_0}{\phi}\ddot{X}'_{it}\ddot{\varepsilon}_{i,t+1} + \ddot{X}'_{it}\ddot{u}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}'_{it}\ddot{X}_{it}} \tag{C8}$$

where $\tilde{\beta}_1 = \frac{\zeta_0}{\phi} + \zeta_1 + \zeta_2\phi$. By Weak Law of Large Numbers, the error term in (C8) converges in probability to a summation of three different expectations:

$$\hat{\beta}_{1,FE} - \tilde{\beta}_1 \to_p \zeta_2 E(\ddot{X}_{it}\ddot{\varepsilon}_{i,t+1}) - \frac{\zeta_0}{\phi}E(\ddot{X}_{it}\ddot{\varepsilon}_{it}) + E(\ddot{X}_{it}\ddot{u}_{it}) \text{ as } NT \to \infty \tag{C9}$$

where we have simplified the matrix notation by noting $\ddot{X}_{it} \in R$. See that the last term of (C9) is obviously zero by strict exogeneity under correct DGP (C2) and Law of Iterated Expectations:

$$E(\ddot{X}_{it}\ddot{u}_{it}) = E(E(\ddot{X}_{it}\ddot{u}_{it}|\ddot{X}_{it})) = 0$$

We then see that the remaining two terms can be expressed in terms of (a) σ_i^2, the variance of the error term in AR(1) expression ε_{it}, and (b) ϕ, the first-order correlation between X_{it} and $X_{i,t-1}$. For the derivation of this relationship, see first that the LHS of (C4) is equivalent to $\ddot{X}_{it}$ from $\bar{X}_i = \frac{\delta_i}{1-\phi}$. Thus,

$$\ddot{X}_{it} = \varepsilon_{it} + \phi\varepsilon_{i,t-1} + \phi^2\varepsilon_{i,t-2} + \cdots \tag{C10}$$

So for the first term of (C9),

$$\begin{aligned} E(\ddot{X}_{it}\ddot{\varepsilon}_{i,t+1}) &= -\frac{1}{N}\left\{E(\varepsilon_{it}\varepsilon_{it}) + E(\phi\varepsilon_{i,t-1}\varepsilon_{i,t-1}) + \cdots\right\} \\ &= -\frac{1}{N}\frac{\sigma_i^2}{1-\phi} \end{aligned}$$

under $|\phi| < 1$, which is true under weak stationarity. This bias will asymptotically converge to zero as $N \to \infty$. However, for the second term of (C9),

$$\begin{aligned} E(\ddot{X}_{it}\ddot{\varepsilon}_{it}) &= E(\varepsilon_{it}\varepsilon_{it}) - \frac{1}{N}\left\{\phi E(\varepsilon_{i,t-1}\varepsilon_{i,t-1}) + \phi^2 E(\varepsilon_{i,t-2}\varepsilon_{i,t-2}) + \cdots\right\} \\ &= \sigma_i^2 - \frac{\phi\sigma_i^2}{N(1-\phi)} \end{aligned}$$

The expectation does not converge to zero even under $N \to \infty$, $T \to \infty$. Combining these

three results together, we see that under $N \to \infty$:

$$\hat{\beta}_{1,FE} - \tilde{\beta}_1 \to_p -\zeta_0 \frac{\sigma_i^2}{\phi} \text{ as } NT \to \infty \tag{C11}$$

□

The above theorem has three interesting implications. First, the results are counter-intuitive in that under misspecification, OLS estimates is not a simple "weighted mean" of the true parameters ζ_0, ζ_1, and ζ_2. That is, $\hat{\beta}_{1,FE}$ does not converge in probability to $\tilde{\beta}_1 = \frac{\zeta_0}{\phi} + \zeta_1 + \zeta_2\phi$. There still is asymptotic bias of $\hat{\beta}_{1,FE}$ from $\tilde{\beta}_1$. This bias is entirely driven by the β_0 and σ_i^2, while β_2 is unimportant asymptotically.

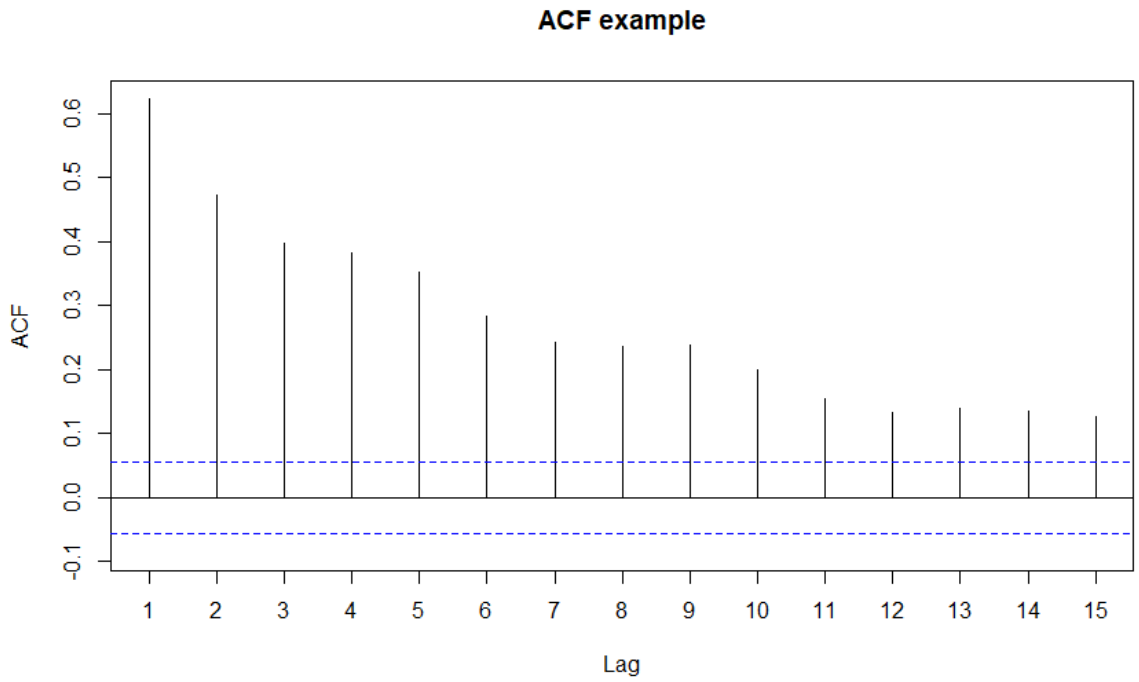
Second, the *sign* of this bias is decided by that of ϕ. With positive autocorrelation of X_{it}, $\beta_{1,FE}$ will be *smaller* than the true $\tilde{\beta}_1$. With negative autocorrelation, $\beta_{1,FE} > \tilde{\beta}_1$ asymptotically. In the context of our study, if today's PM2.5 concentration is positively correlated with that of the preceding day, then OLS estimates of the marginal effect will underestimate the true value.

Lastly, the *size* of this bias is driven by both ϕ and σ_i^2. Larger σ_i^2 and smaller $|\phi|$ lead to larger biases. This is intuitive, as larger σ_i^2 means that more noise is introduced *via* ε_{it}, and smaller $|\phi|$ means that it no longer makes sense to represent the effect of present-day PM2.5 as a linear combination of that of previous days.

C.2 Evidence of serial correlation

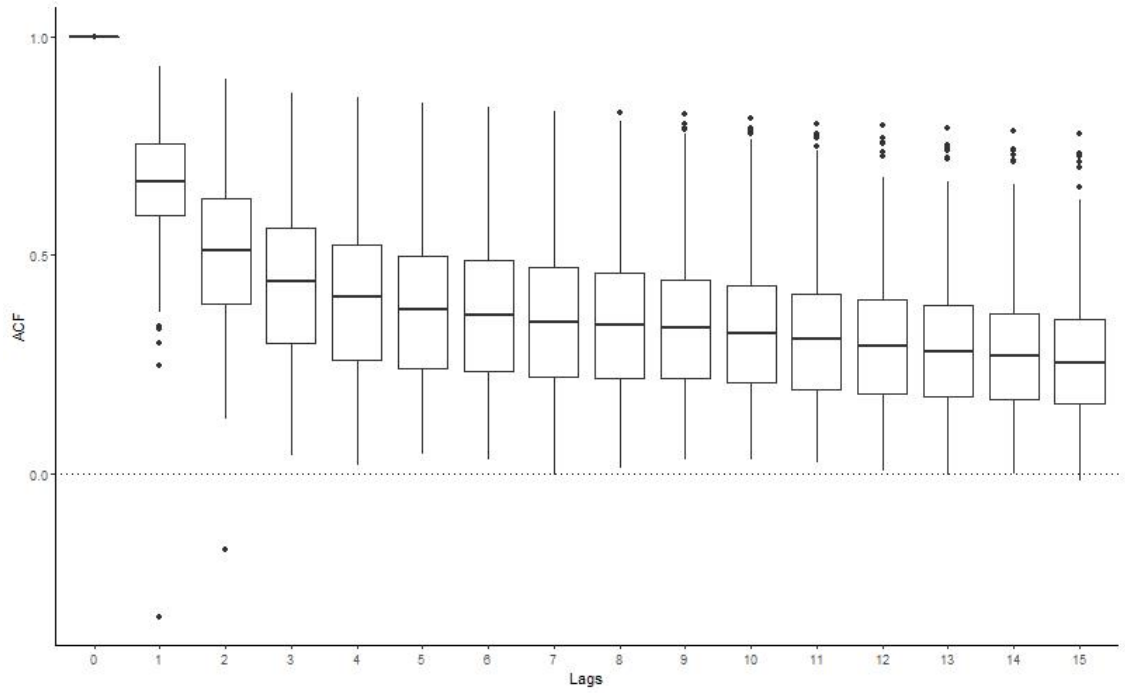
We assess the autocorrelation of PM2.5 using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. To this end, we calculate the ACF at each ZIP code, *à la* Appendix Figure C1, obtaining the ACF and PACF estimates for lags up to 15. These estimates are obtained for all 1,764 ZIP codes included in the data. We present the boxplot of such estimates at each lag in Appendix Figures C2 and C3.

The results show that indeed, the PM2.5 concentrations used in our data show significant serial correlation. For certain ZIP codes, the ACF is still at a significantly high level even after 15 lags, attesting to the strength of the serial correlation.



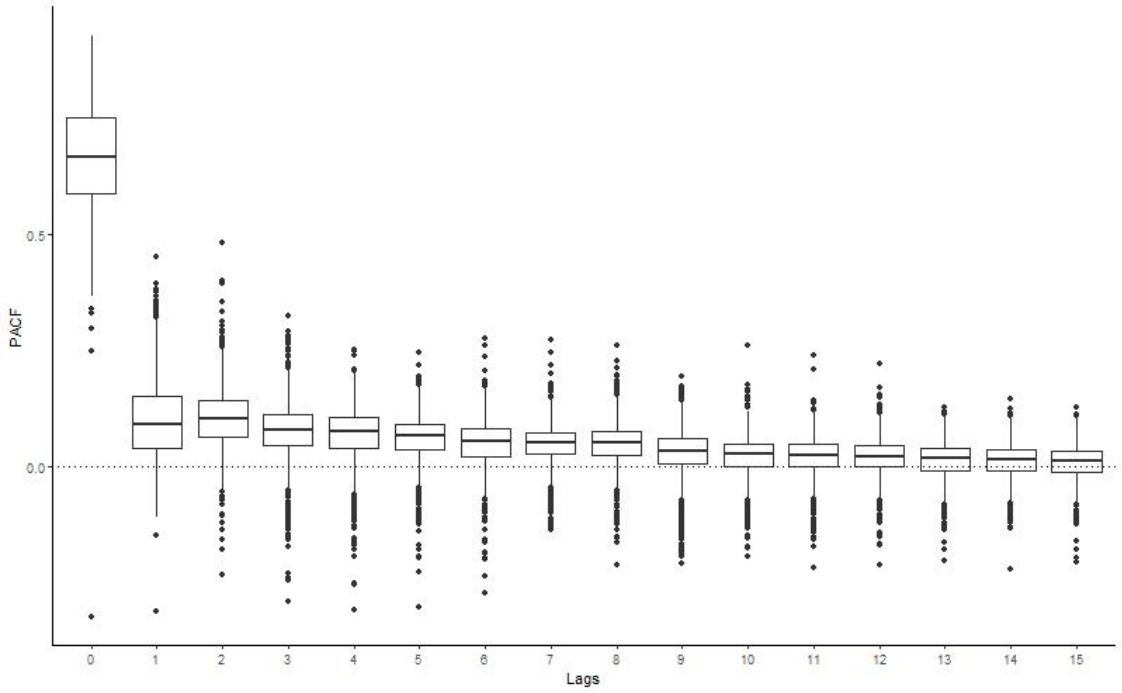
Appendix Figure C1. Example of ACF graph, for ZIP code 89439

Notes: Figure plots the Autocorrelation Function (ACF) for PM2.5 concentration in ZIP code 89439.



Appendix Figure C2. Boxplot of ACF estimates for 1,468 ZIP codes

Notes: Figure plots the ACF estimates for lags up to 15, calculated at each ZIP code.



Appendix Figure C3. Boxplot of PACF estimates for 1,468 ZIP codes

Notes: Figure plots the PACF estimates for lags up to 15, calculated at each ZIP code.

D Additional robustness checks

Appendix Table D1. Additional robustness checks for Table 2 results

	Main	Reduced-form		Fixed effects	
	(1)	(2)	(3)	(4)	(5)
AOD	-1.41^{***}	-1.46^{***}	-1.28^{***}	-1.27^{***}	-1.42^{***}
	(0.26)	(0.31)	(0.25)	(0.26)	(0.25)
F-statistic	56.20	56.63	54.68	54.93	56.07
Adj. R^2	0.85	0.85	0.86	0.77	0.85
Dep. var. mean	3.89	3.89	3.89	3.89	3.89
AOD mean	0.10	0.10	0.10	0.10	0.10
AOD SD	0.06	0.06	0.06	0.06	0.06
N	157,747	157,747	157,747	157,747	157,747

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Column 1 replicates the main result in Table 2, panel B. Column 2 includes fourth power of the covariates in the reduced-form equation. Column 3 removes cubic transformation of the covariates. Column 5 substitutes ZIP code-by-year fixed effect with ZIP code *and* year fixed effect, separated. Column 6 includes a monthly fixed effect.

We present additional robustness checks of the solar generation-air opacity model (2) with 2SLS estimation. Column 1 of Appendix Table D1 replicates the main result. Columns 2 and 3 introduce changes in the specification of the reduced-form equation, the former including a fourth power of the covariates and the latter excluding the third power of covariates. Both changes do not lead to substantially different estimates.

Columns 4 and 5 assess the sensitivity of our results to fixed effects. Column 4 substitutes ZIP code-by-year fixed effect by ZIP code *and* year fixed effects, separately. This is to address the concern that the β may be biased if an AOD shock in a given ZIP code leads to a change in behavior common to that ZIP code's year, which may be particularly important if we are assessing *grid-wise* outcome. (e.g., a high-pollution event changing energy consumption behaviors for that ZIP code throughout the year). This does not change the results. Lastly, column 5 introduces monthly fixed effects, as weekly fixed effect may not be sufficient to capture the changes in solar generation conditions within the year. This also does not change the results in an important manner.

Appendix Table D2 provides additional robustness checks for the second part of our

Appendix Table D2. Additional robustness checks for Table 4 results

	(1)	(2)	(3)	(4)	(5)
PM2.5	0.0021***	0.0022***	0.0021***	0.0021***	0.0021***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Sample subsets	All	All	All	All	All
Dropped monitors	None	>90%, < 10%	< 10%	>90%	None
Negative PM2.5	No	No	No	No	Yes
AOD mean	0.1114	0.1111	0.1118	0.1106	0.1113
PM2.5 mean	10.3439	10.3485	10.7722	9.9195	10.3114
PM2.5 SD	7.4980	7.3608	7.6975	7.1468	7.5113
N	151,119	113,844	132,669	132,736	151,561
Adj. R^2	0.3103	0.3112	0.3119	0.3098	0.3105

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Column 1 replicates the main result in Table 4, panel A. Columns 2 to 4 drop EPA monitors that were either in high- or low-pollution regions within their county. Column 5 includes negative PM2.5 measurements.

analysis, linking PM2.5 to AOD. All regressions are on "full" data set with all four seasons. Column 1 replicates the main result (column 5, panel A of Table 4). Column 2 drops EPA monitors that were placed in census tracts with top and below 10% percentile of PM2.5 level within each county. Column 3 drops those in below 10% percentile, and column 4 drops top 10% percentile monitors. Column 5 includes negative PM2.5 data. These all do not change the OLS results.

E Comparison with He et al. (2020) methods

Appendix Table E1. Comparison with He et al. methods

	(1)	(2)	(3)	(4)	(5)	(6)
AOD	−1.556***		−1.309	−89.838***	−45.972	
	(0.322)		(1.871)	(14.063)	(96.877)	
PM2.5		−0.016***				2.423
		(0.003)				(3.492)
Endog. var.	AOD	PM2.5	AOD	AOD	AOD	PM2.5
First-stage	TS	TS	He	TS	He	He
Log-transformed	Yes	Yes	Yes	No	No	No
F-statistic	89.108	361.000	11.397	89.108	11.397	6.902
Dep. var. mean	3.897	3.795	3.897	59.986	59.986	55.452
AOD/PM2.5 mean	0.105	10.041	0.105	0.105	0.105	10.041
AOD/PM2.5 SD	0.063	4.446	0.063	0.063	0.063	4.446
N	202807	283064	202807	202807	202807	283064
Adj. R^2	0.840	0.836	0.835	0.880	0.884	0.862

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Column 1 replicates the main result in Table 2, column 1 panel B. Column 2 changes the endogenous variable of interest to PM2.5 instead of AOD. Column 3 uses the first-stage regression suggested by He et al. (2020). Column 4 uses this study's first-stage regression, but uses a linear specification. Column 5 uses He et al. (2020) first-stage regression and linear specification. Column 6 uses He et al. (2020) first-stage regression, linear specification, and chooses PM2.5 as the endogenous variable of interest. "TS" = This study.

As previously mentioned, our empirical strategy differs from He et al. (2020) in three important ways: (1) Usage of AOD, not PM2.5, as the endogenous variable of interest, (2) use of wildfire interacted with wind conditions, not just wind direction, in the first-stage regression, and (3) use of log-linear specification instead of linear specification. Here, we replicate to the He et al. (2020) approach to the greatest extent possible and compare the results to ours.

Column 1 in Appendix Table E1 replicates our column 1, Table 2 estimate. Column 2 relaxes the first difference in that it uses PM2.5, not AOD, as the endogenous variable. The marginal effect of PM2.5 on solar generation implied by column 2 is −1.58%, from $\exp(-0.016) - 1$. The question is whether this estimate is similar in size to that from combining column 1 with Table 4 estimates. Table 4, column 5 panel B implies that the marginal effect of PM2.5 on solar generation is on average 0.0025 across the four seasons.

This translates to a marginal loss in solar generation of 0.39% from PM2.5 (from $\exp(0.0025 \times 1.559) - 1$).

The considerable difference in the marginal estimates is consistent with fluctuations of PM2.5 being positively correlated with other factors that also affect solar generation. This will inflate the role that PM2.5 *per se* plays in increasing atmospheric opacity (and hence reducing solar generation). Our Table 4 estimates allows us to identify the share of AOD variation that is attributable to PM2.5.

Column 3 of Appendix Table E1 explores the implications of the second difference between this study and He et al. (2020). We drop our original first-stage specification, but rather use only the cosine of wind direction[3]. The results show that the He et al. (2020) first-stage specification simply fails to identify the impact of AOD on solar generation.

One possible explanation for this difference is that in California, contrary to Arizona, wind direction could violate the exclusion restriction. That is, wind direction-driven pollution could be dominated by a stationary source of pollution. The predictability of such pollution could lead to residential sorting in neighboring population. Interacting wind direction with wildfire smoke, which originates from transient and plausibly unanticipatable sources, defends us against this possible mechanism of exclusion restriction violation.

Column 4 explores the consequences of using a linear, not log-linear, specification, while keeping intact all remaining aspects of our empirical strategy. The results suggest that the effect of one SD change in AOD on three-day solar generation is around 5.66kWh (89.838×0.063). This is around 9.4% of the average three-day solar generation (59.986kWh), which is remarkably close to the 9.4% estimate from the log-linear specification.

The linear specification has very different interpretations compared to our preferred log-linear model. The linear approach argues that one SD increase in AOD reduces solar generation by 5.66kWh, *regardless of the PV system capacity*. On average, the PV systems in our data set show a three-day solar generation of anywhere between 11.84kWh and 320.34kWh. 5.66kWh is 48% and 2% of these quantities. Such remarkable difference in the impact of AOD has little theoretical ground in atmospheric physics. A policy analysis using the linear approach will simply consist of counting the number of PV systems and multiplying it by 5.66kWh. On the contrary, the log-linear approach will involve summing up the total number of solar energy generated, and attributing a certain share of this amount to AOD.

Columns 5 and 6 now test different combinations of He et al. (2020) empirical strategy, with column 5 using linear specification and wind direction-only first-stage regression. Col-

[3]To be precise, we follow He et al. (2020) in using the difference between local wind and "prevailing winds," the latter being defined as the county. Note that there may be minute differences in the actual wind direction variable, as He et al. (2020) uses a different weather data set with hourly wind observations.

umn 6 uses linear specification, wind direction-only first-stage regression, and PM2.5 as the endogenous variable. Column 6 thus fully replicates the He et al. (2020) empirical strategy. Both columns fail to identify the impact of PM2.5.

F Additional Tables and Figures

Appendix Table F1. Definition of used variables

Variable	Definition	Source
Daily solar generation	Average solar energy generated for rooftop PV systems in each ZIP code. Confined to residential rooftop systems. Unit: kWh/ZIP code.	CSI
GHI	Global horizontal irradiance. Total sum of solar energy incident on a horizontal surface at a given location. Unit: W/m^2.	NSRDB
DNI	Direct normal irradiance. Sum of solar energy coming from the direction of the sun, measured at the surface. Unit: W/m^2.	NSRDB
DHI	Diffused horizontal irradiance. Sum of solar energy that has been diffused in the atmosphere and enters the surface from a non-normal angle. Unit: W/m^2.	NSRDB
PM2.5	Concentration of fine particulate matter with diameter smaller than 2.5μm in the atmosphere. Unit: $\mu g/m^3$.	Reid et al., 2021
AOD†	Aerosol Optical Depth. Natural log of the solar radiation fraction that is attributable to the aerosols in the atmosphere. Multiplied by 1,000 for legibility. Dimensionless.	MODIS
Windspeed†	Wind velocity in the ZIP code.	gridMET
Wind direction†	Cosine of wind direction in the ZIP code.	gridMET
NO_2	Surface-level Nitrogen Dioxide concentration Unit: mol/km^2.	OMI
Precipitation†	Snowfall and rainfall per day. Unit: mm/day.	gridMET
Diurnal temperature†	Daily temperature minimum & maximum. Unit: °C.	gridMET
Diurnal relative humidity†	Daily relative humidity & maximum. Unit: %.	gridMET
Socioeconomic variables	% people of color, % low income, % less than high school education.	EJScreen

† denotes data accessed through Google Earth Engine (Gorelick et al., 2017).

Appendix Table F2. Heterogeneous effect of air opacity

	% people of color		% low income		% less than high school	
	Low	High	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)
AOD	-1.668^{***}	-1.372^{***}	-1.604^{***}	-1.449^{**}	-1.535^{***}	-1.513^{***}
	(0.367)	(0.357)	(0.291)	(0.454)	(0.350)	(0.360)
F-statistic	45.497	46.006	62.259	32.327	55.905	36.049
Adj. R^2	0.860	0.821	0.873	0.808	0.843	0.838
Dep. var. mean	3.920	3.875	3.893	3.902	3.920	3.871
Treat. var. mean	0.101	0.109	0.104	0.106	0.103	0.107
Treat. var. sd.	0.063	0.062	0.062	0.064	0.063	0.063
N	99,108	102,971	106,598	95,481	106,108	95,971

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Table reports IV estimations of equation (2). ZIP codes are divided into two groups depending on their socioeconomic characteristics. All columns use three-day measures of the dependent variables to account for lagged effects. All results include ZIP code-year and week fixed effects. Three-day diurnal temperature, precipitation, relative humidity, NO_2 concentration, and their polynomials, are used as covariates. Standard errors two-way clustered by year-week and ZIP code.

Appendix Table F3. Equity implications of AOD impact

	(1)	(2)	(3)	(4)	(5)	(6)
% people of color	0.055***			0.035***		
	(0.002)			(0.002)		
% low-income household		0.025***			0.019***	
		(0.004)			(0.004)	
% less than high school education			0.106***			0.067***
			(0.009)			(0.007)
Average PM2.5				0.350***	0.505***	0.477***
				(0.020)	(0.018)	(0.018)
PM2.5 included	No	No	No	Yes	Yes	Yes
Adj. R^2	0.285	0.018	0.080	0.394	0.319	0.340
Num. obs.	1757	1757	1757	1752	1752	1752

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

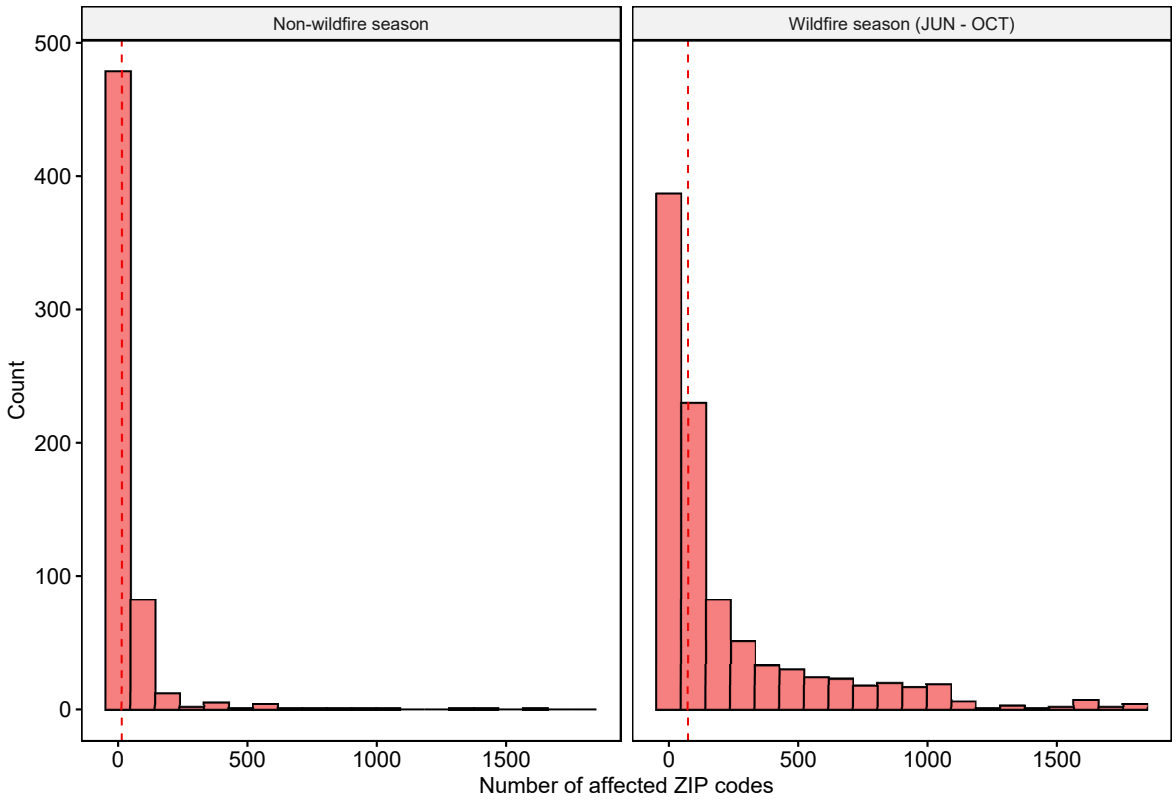
Notes: Table assesses the correlation between % of solar generation lost from AOD, averaged at the ZIP code level, versus socioeconomic characteristics. Columns 1 to 3 simply regress the average lost solar generation against socioeconomic characteristics, while columns 4 to 6 include average PM2.5 levels at each ZIP code to the right-hand side.

Appendix Table F4. Potential drivers of AOD impact heterogeneity

	(1)	(2)	(3)	(4)
Average PM2.5	0.494***	0.508***	0.481***	0.470***
	(0.018)	(0.019)	(0.022)	(0.023)
Population density	0.006***			
	(0.001)			
Average NO_2		−0.250		
		(0.325)		
Average wind speed			−0.222***	
			(0.078)	
Wind speed variability				−0.475***
				(0.146)
R^2	0.319	0.311	0.316	0.317
Adj. R^2	0.318	0.310	0.315	0.316
Num. obs.	1753	1761	1759	1759

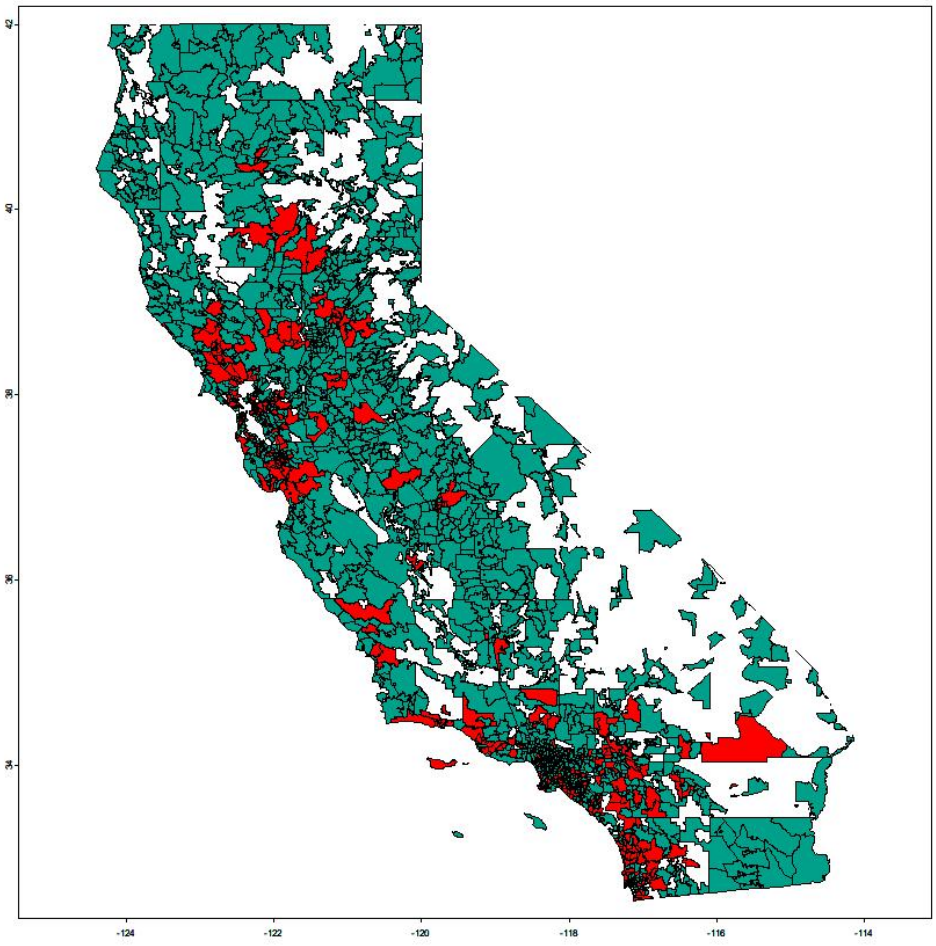
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Notes: Table explores the potential drivers of AOD-attributed solar generation loss. The units of population density, NO_2 density, and wind speed are 1,000 people per square mile, mol per km^2, and m/s. "Variability" is defined as the standard deviation of wind speed.



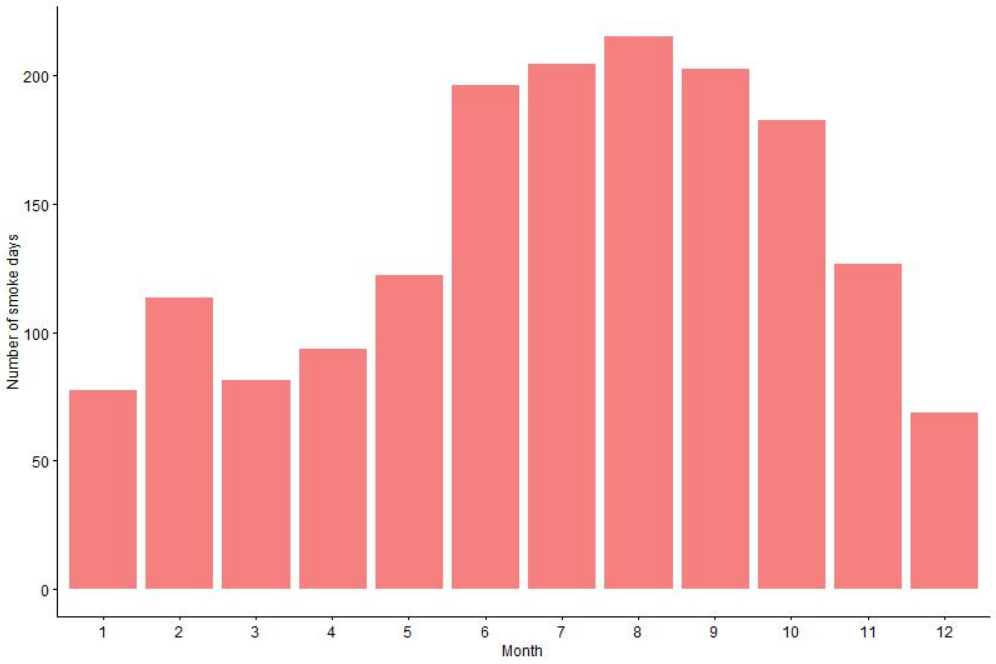
Appendix Figure F1. Number of ZIP codes affected per wildfire smoke event

Notes: Dashed lines show the median for each group (17 and 100 ZIP codes for non-wildfire season and wildfire season, respectively.).



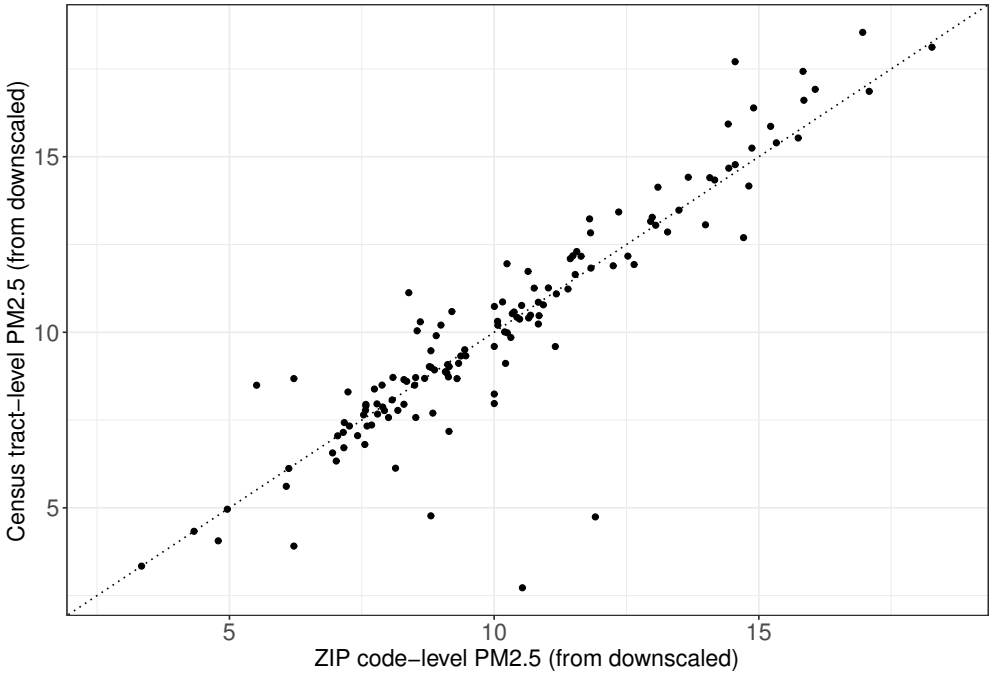
Appendix Figure F2. Expansion of geographical coverage from NSRDB

Notes: CSI-included ZIP codes shown in red, NSRDB-augmented ZIP codes shown in green.



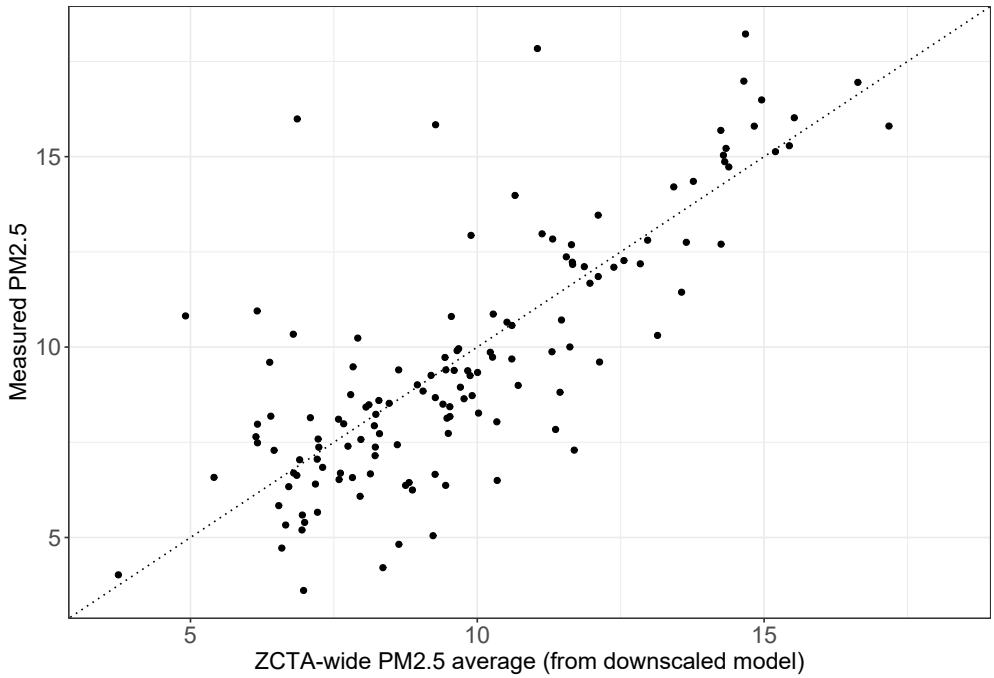
Appendix Figure F3. Monthly occurrence of smoke days

Notes: Figure shows the number of smoke days in each month observed over the analysis period (2010-2016).



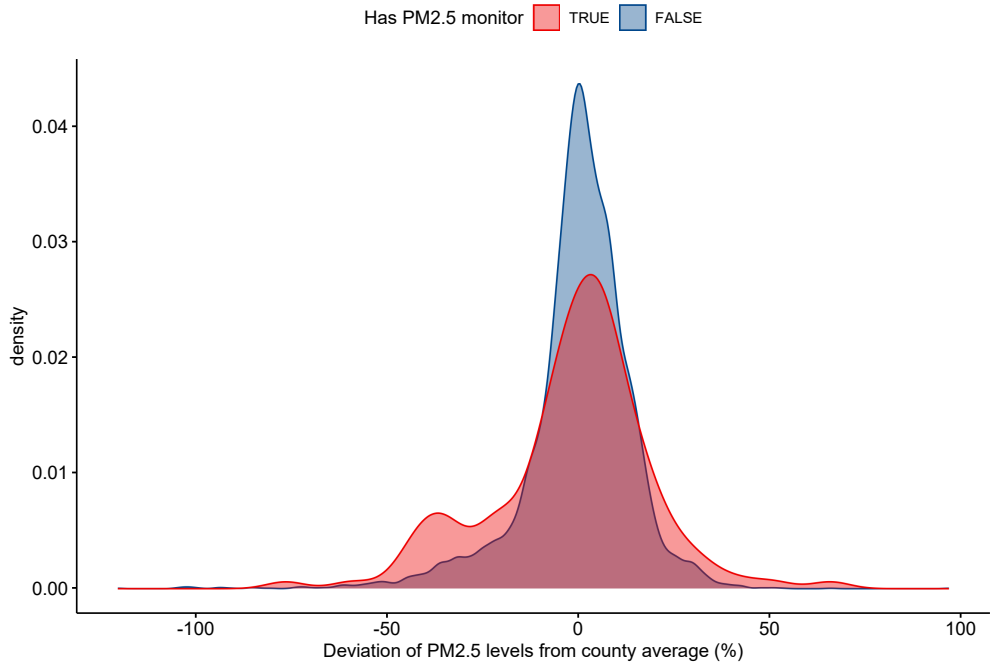
Appendix Figure F4. PM2.5 level of census tracts with EPA monitors

Notes: Figure compares PM2.5 levels in census tracts with EPA monitors with their ZIP code-wide average. Data from downscaled PM2.5.



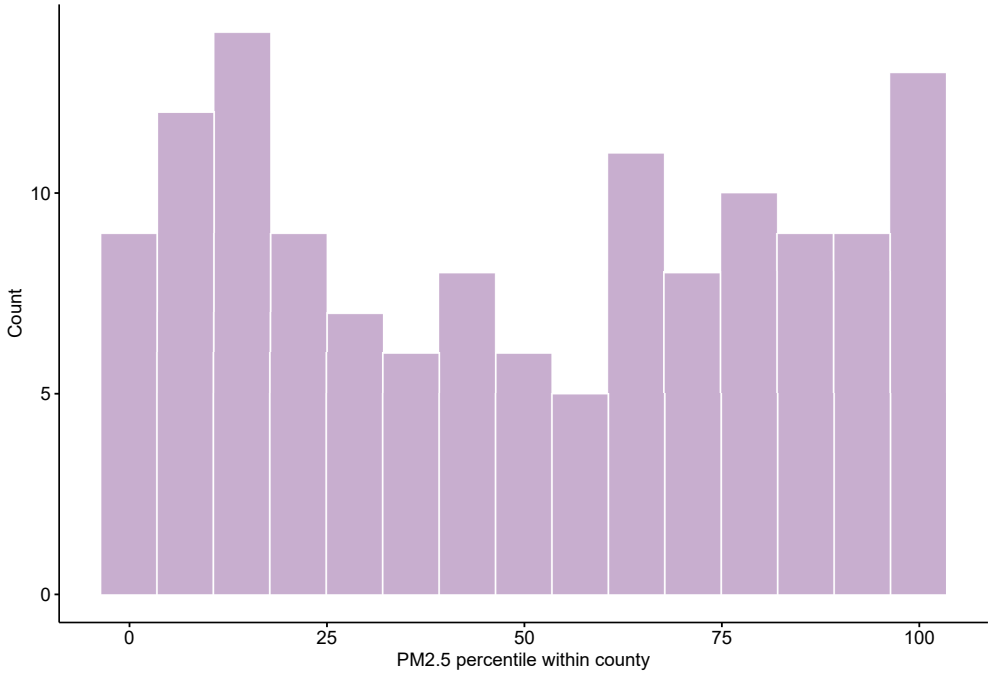
Appendix Figure F5. Measured PM2.5 levels versus ZIP code-wide average

Notes: Figure compares PM2.5 levels measured by EPA monitors with their ZIP code-wide average, the latter from downscaled PM2.5 data.



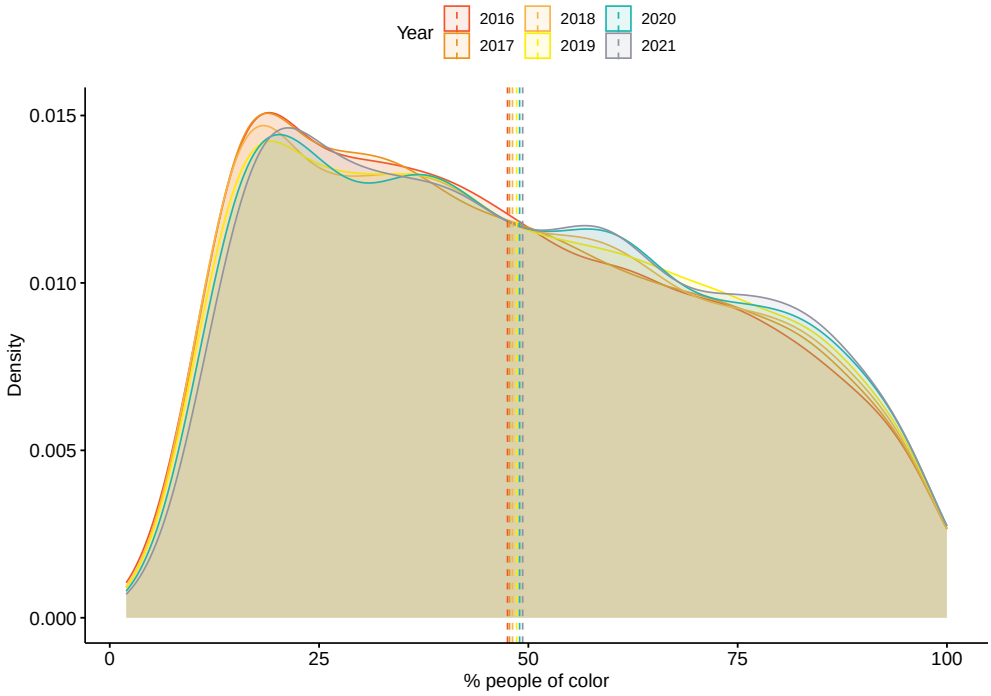
Appendix Figure F6. PM2.5 level of tracts with EPA monitors, versus county average

Notes: Figure compares, for tracts with and without EPA monitors, PM2.5 levels in census tracts to county-wide average. "Deviation" is defined as "(PM2.5 in census tract - PM2.5 in county)/PM2.5 in county." Data from downscaled PM2.5.



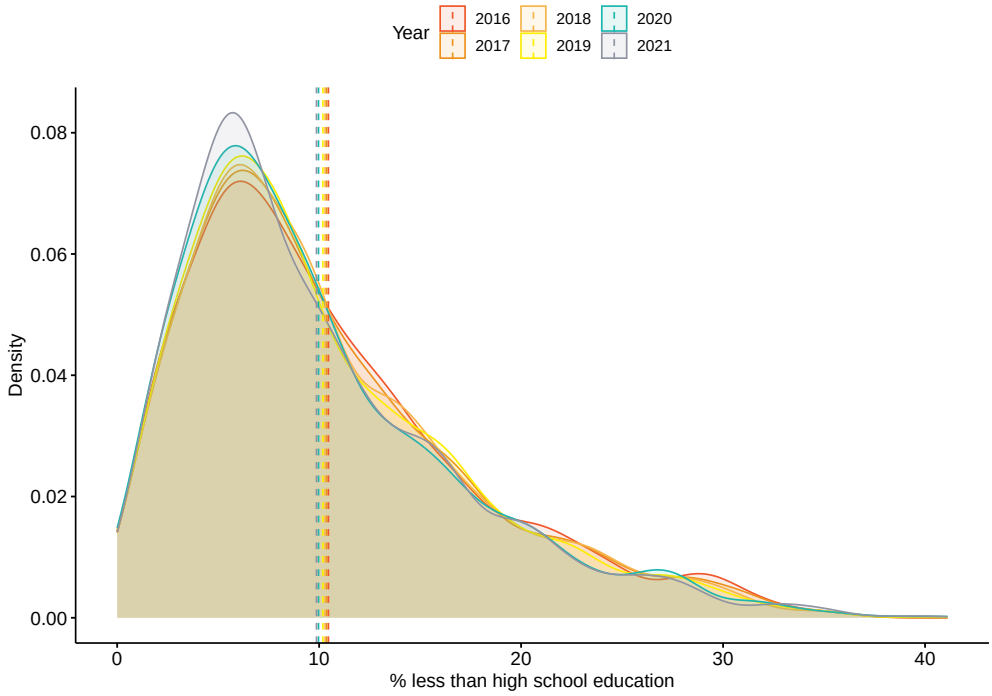
Appendix Figure F7. Within-county percentile of PM2.5 levels for tracts with EPA monitors

Notes: Figure plots, for census tracts with EPA monitors, the percentile of their PM2.5 levels within their county. Higher percentile means higher pollution in the monitor-installed tract compared to its county average. Data from downscaled PM2.5.



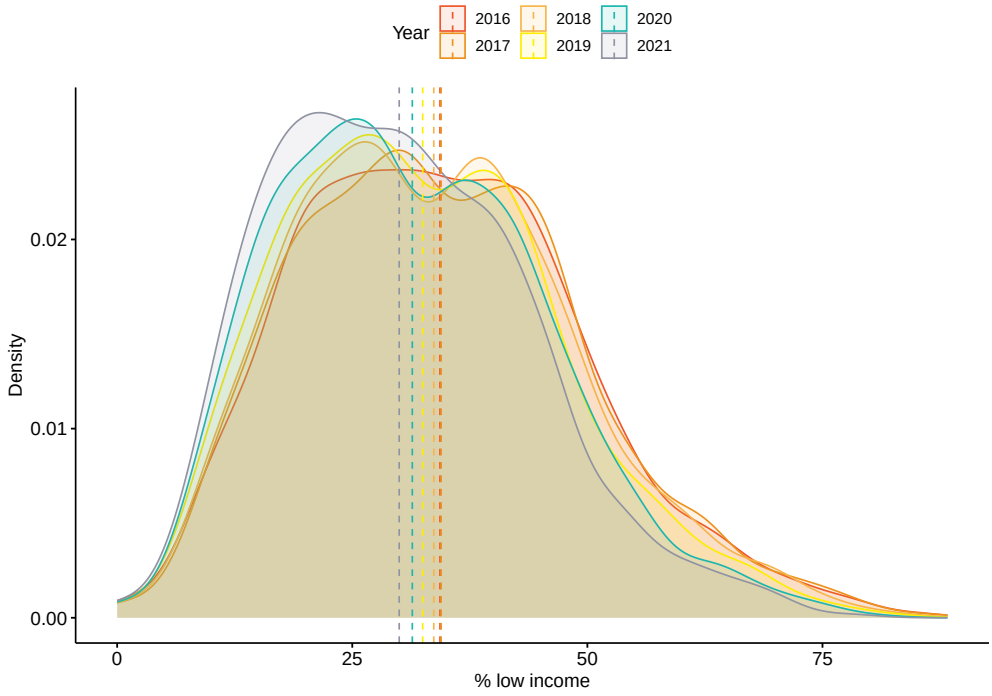
Appendix Figure F8. Distribution of "% people of color", from 2016 to 2021

Notes: Figure shows the distribution of "% people of color" for each ZIP code in the EJScreen data set from years 2016 to 2021. The dotted lines represent the average of the distributions in each year.



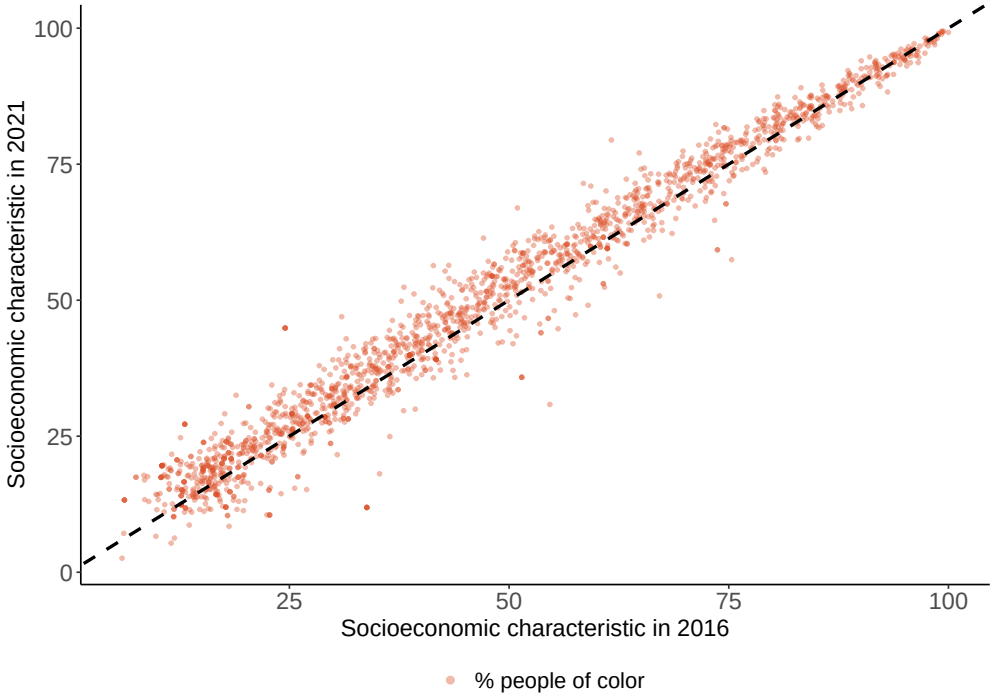
Appendix Figure F9. Distribution of "% less than high school educated", from 2016 to 2021

Notes: Figure shows the distribution of "% less than high school educated" for each ZIP code in the EJScreen data set from years 2016 to 2021. The dotted lines represent the average of the distributions in each year.



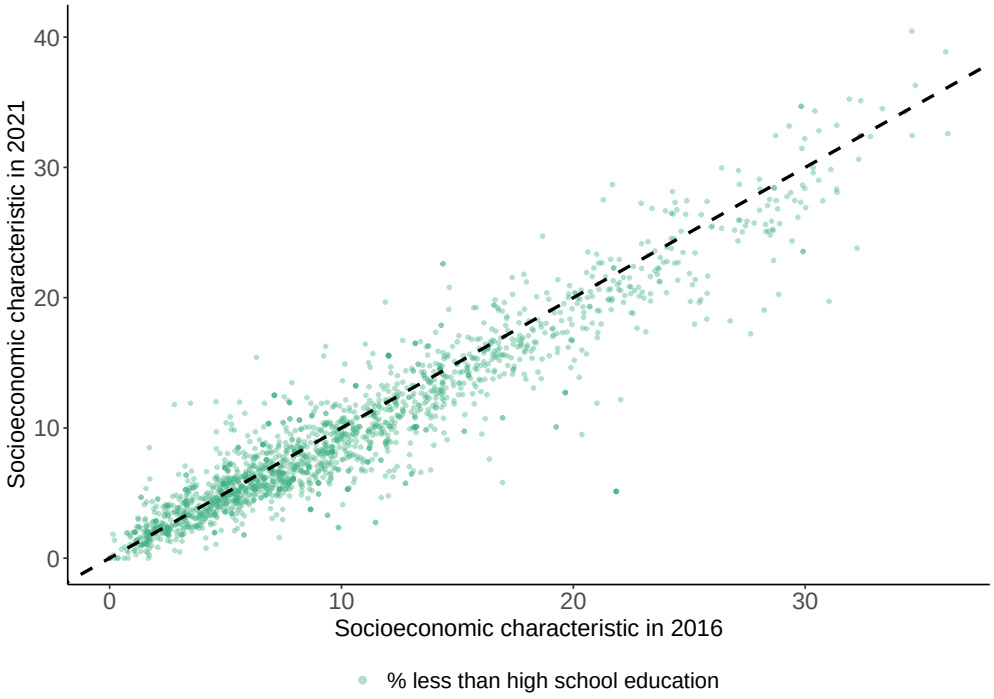
Appendix Figure F10. Distribution of "% low income", from 2016 to 2021

Notes: Figure shows the distribution of "% low income" for each ZIP code in the EJScreen data set from years 2016 to 2021. The dotted lines represent the average of the distributions in each year.



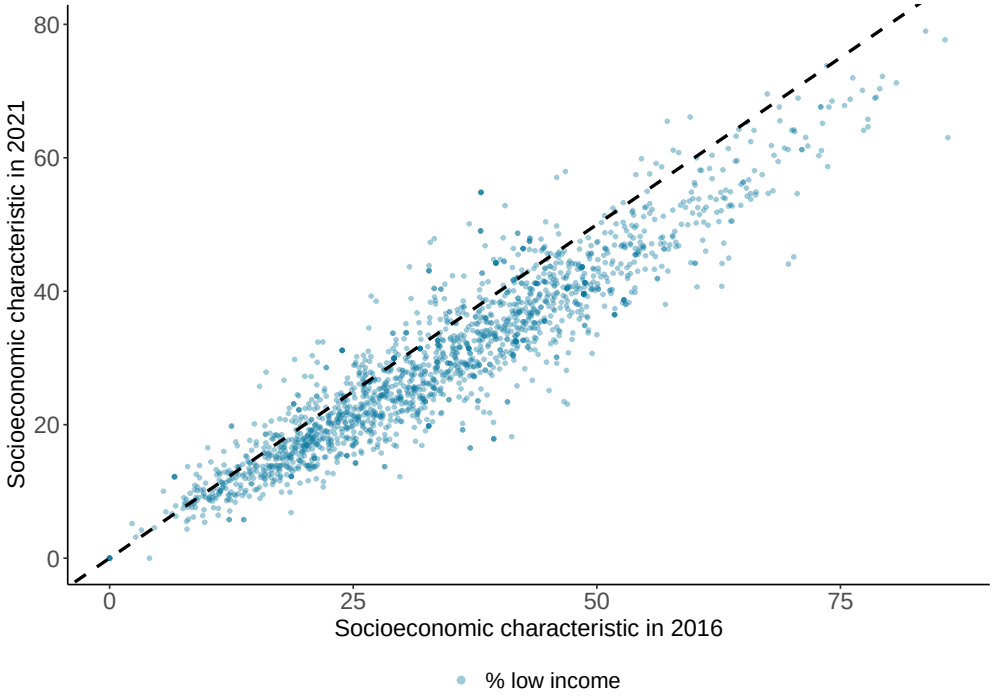
Appendix Figure F11. "% people of color", 2016 versus 2021

Notes: Figure compares, at each ZIP code, the "% people of color" variable in 2016 versus 2021. Dotted line represents $y = x$ line.



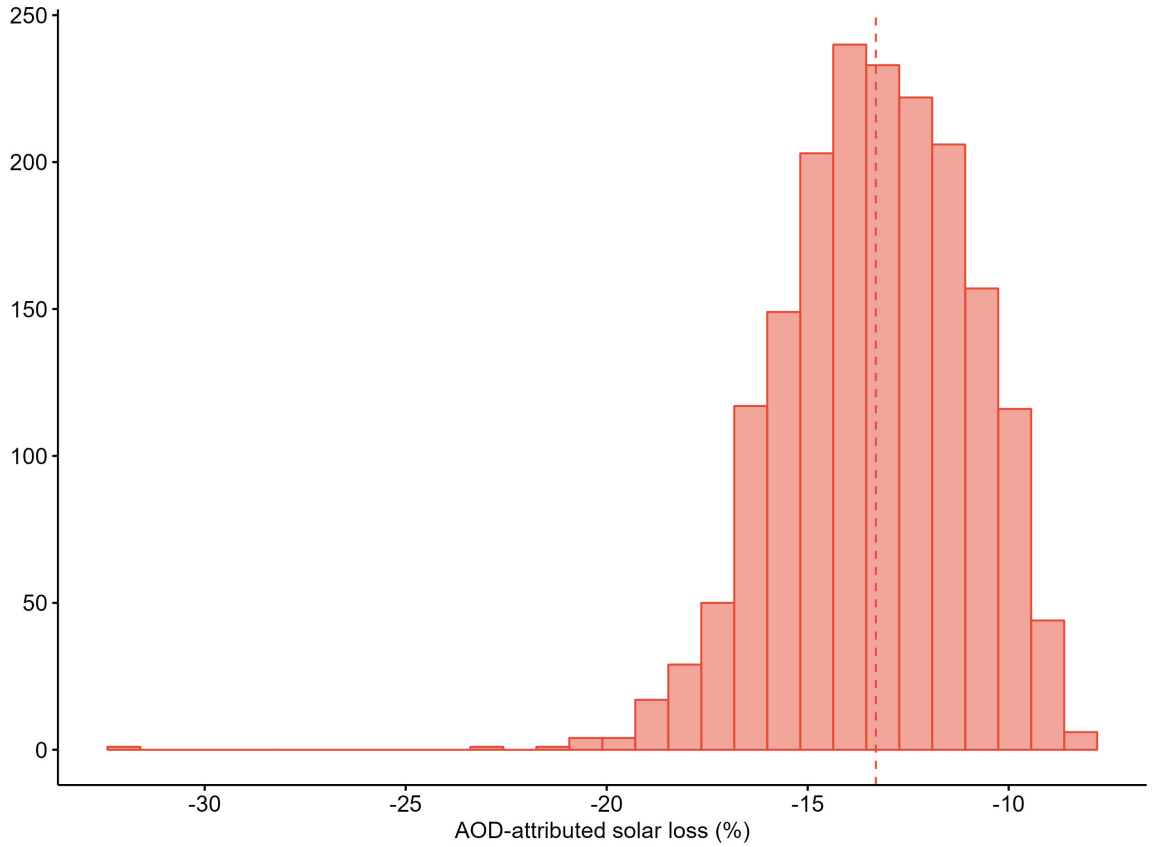
Appendix Figure F12. "% less than high school educated", 2016 versus 2021

Notes: Figure compares, at each ZIP code, the "% people of color" variable in 2016 versus 2021. Dotted line represents $y = x$ line.



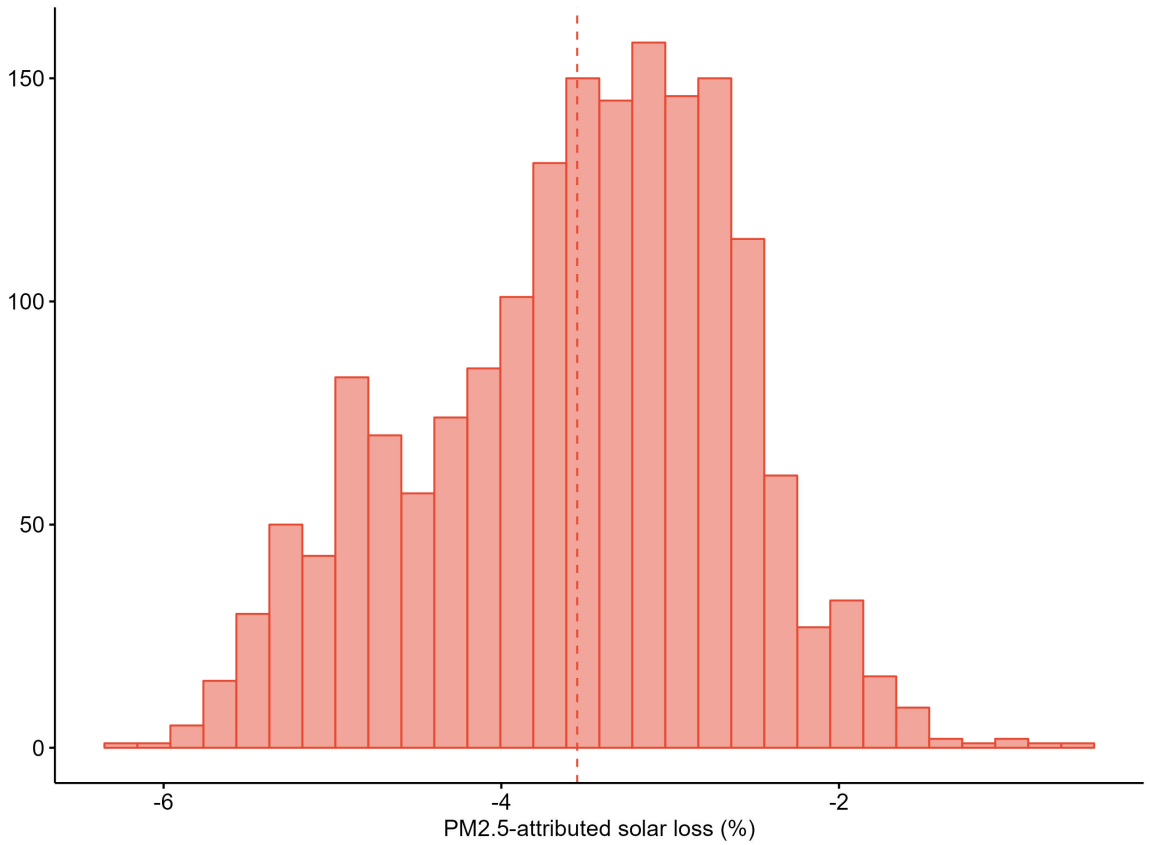
Appendix Figure F13. "% low income", 2016 versus 2021

Notes: Figure compares, at each ZIP code, the "% people of color" variable in 2016 versus 2021. Dotted line represents $y = x$ line.



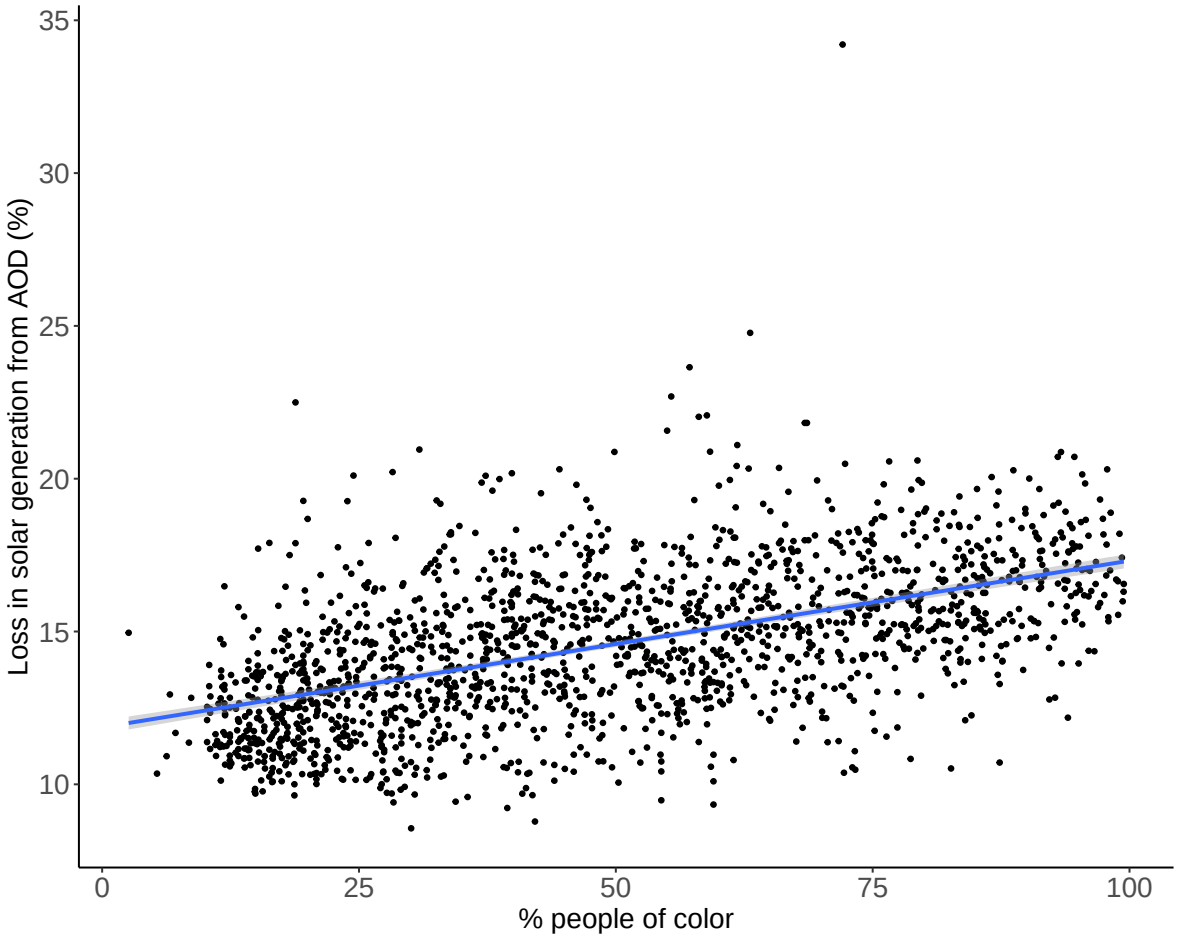
Appendix Figure F14. Histogram of AOD-attributed solar loss

Notes: Figure shows the ZIP code-average of loss of rooftop solar productivity from AOD levels, in each ZIP code. The red dotted line denotes the statewide average (−14.54%).



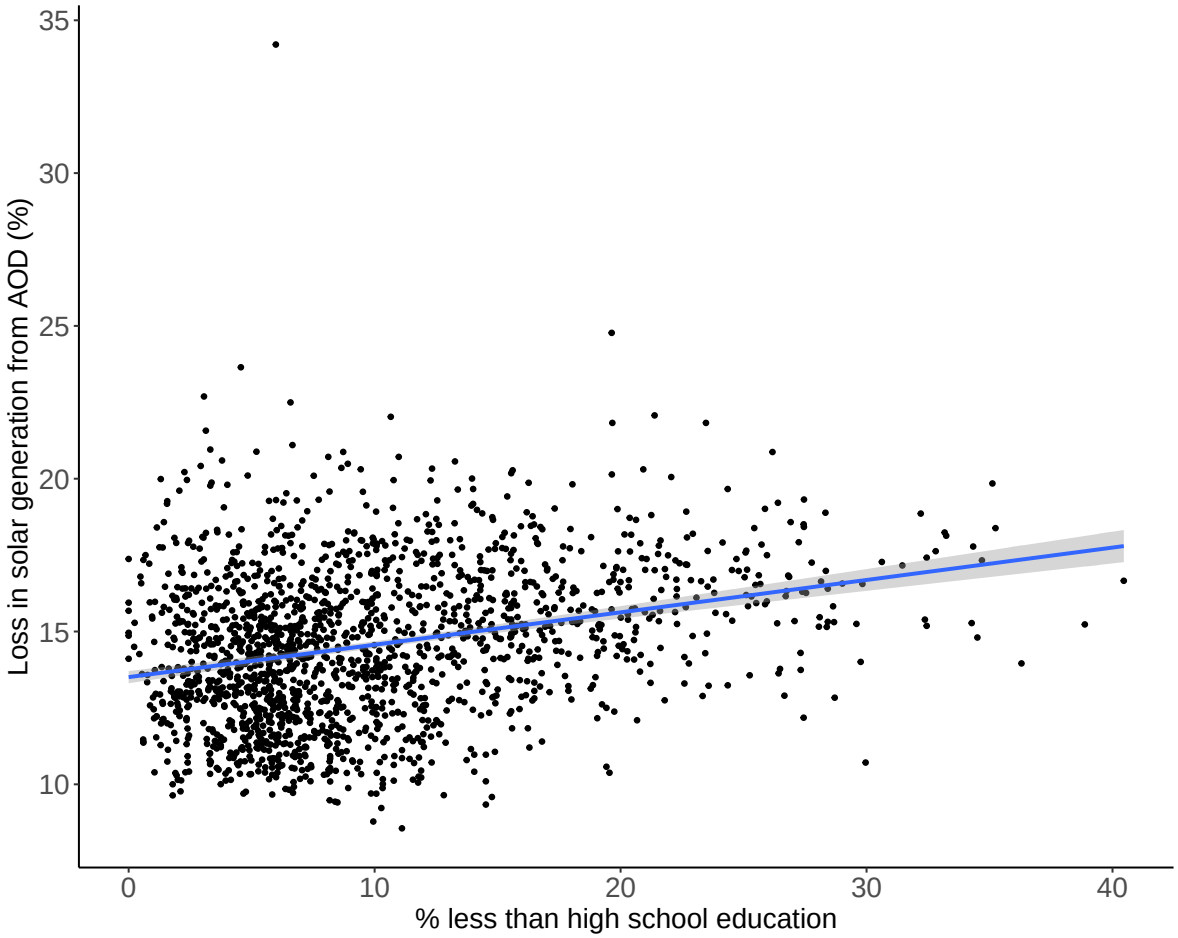
Appendix Figure F15. Histogram of PM2.5-attributed solar loss

Notes: Figure shows the ZIP code-average of loss of rooftop solar productivity from PM2.5 conditions, in each ZIP code. The red dotted line denotes the statewide average (−3.91%).



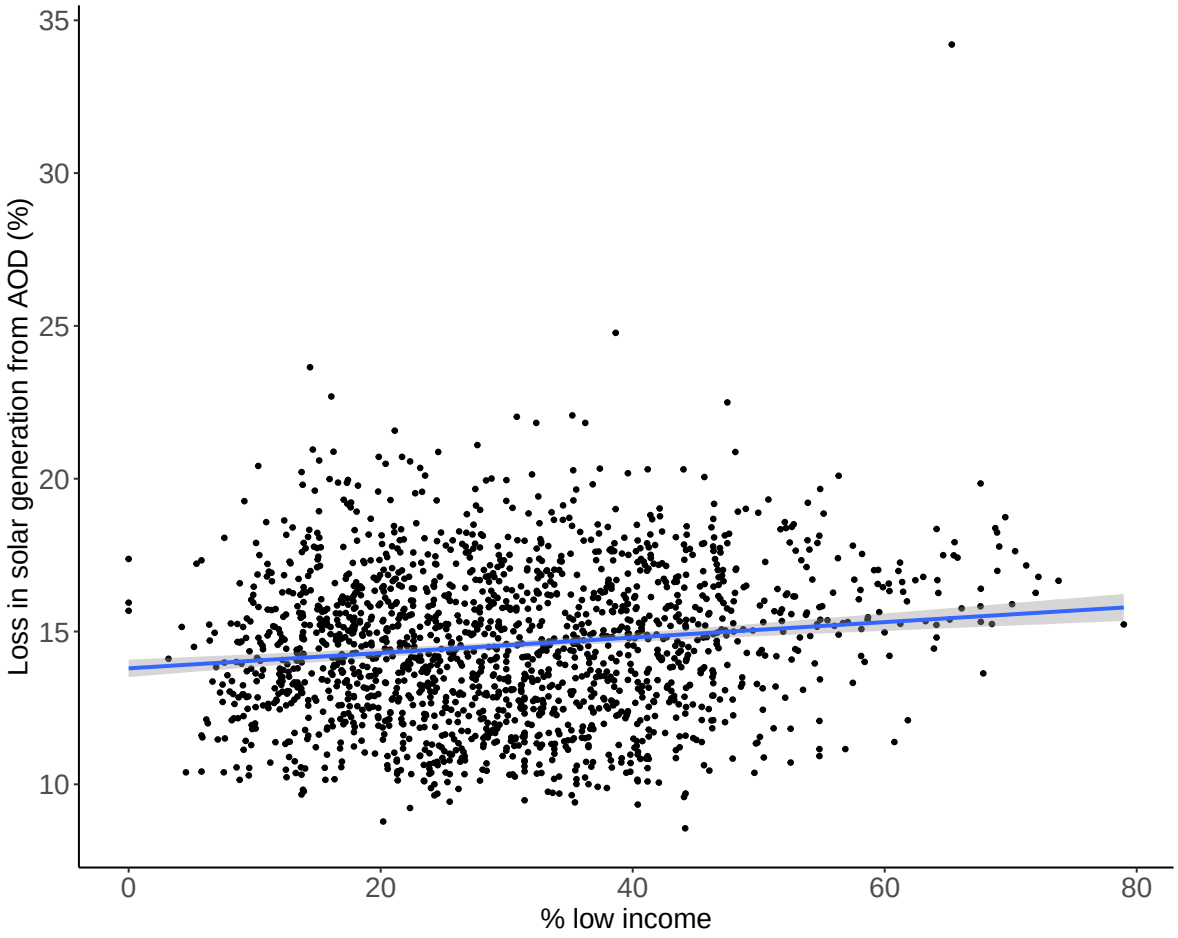
Appendix Figure F16. AOD-attributed solar loss versus % people of color

Notes: Figure places AOD-attributed solar loss on the y axis, and the % people of color on the x axis. Blue line and grey fill represent linear regression line and 95% confidence interval, respectively.



Appendix Figure F17. AOD-attributed solar loss versus % less than high school educated

Notes: Figure places AOD-attributed solar loss on the y axis, and the % less than high school educated on the x axis. Blue line and grey fill represent linear regression line and 95% confidence interval, respectively.



Appendix Figure F18. AOD-attributed solar loss versus % low income

Notes: Figure places AOD-attributed solar loss on the y axis, and the % low income household on the x axis. Blue line and grey fill represent linear regression line and 95% confidence interval, respectively.