A Particular Look at Urban Particulate: Analyzing Observed Variations in PM2.5 Levels

Dan Quasney
GitHub: DQOfficial, NYUID: DQ333

Abstract

Air pollution has received particular attention over the last few months due to global coverage of the dangerous conditions in China and wildfires that have devastated the Western United States and parts of Canada. This type of pollution, called PM2.5, is caused by particulate matter rather than gases such as carbon dioxide and is particularly dangerous because it can damage lung tissue when inhaled in large amounts. This paper attempts to identify the features of an urban environment that may drive high levels of this pollutant.

Introduction

Smog in China. Wildfires in California. Haze in New York City. All of these phenomena have garnered considerable attention over the last few years due to the fact that humans are almost always responsible for their creation, and the debilitating impacts each can have on the environment where it occurs. Though each occurs in vastly different geographic areas, they do have something in common: they all generate huge amounts of particulate matter which mixes into the air people breathe. This particulate matter, specifically particles between 0.1 and 2.5 μm called PM2.5, are especially dangerous because their small size allows them to bypass air filters and become even more deeply embedded in pulmonary tissue, perhaps even entering the bloodstream.[1] These particles are generated when exhaust emissions from combustion engines, ambient dust, and industrial exhaust react with atmospheric water and sunlight, and their composition is both seasonal and distinct to different regions in the US. For instance, contaminant levels are at their highest in the eastern United States between July and August, whereas levels peak in the West during the winter months.[2] The health risks that high levels of PM2.5 pose to the public is the primary reason for selecting this topic for the foundations project. This paper attempts to identify the features that contribute to a city's PM2.5 levels and how the relationships between those features influence observable variations in those levels.

I am missing a section that describes the data. e.g. later you go on to talk about cars/ busses vs bikes and i have no idea how you got that data, and if it is a believably unbiassed measure of a CBSA transportation mode.

Methodology and Analysis

To identify an adequate municipality size, this study gives careful consideration not only to population, regional location, and proximity to a major metropolitan area but also to the implications of analyzing one municipal level over another. Statewide data tend to be too broad and fail to capture the impacts of smaller stimuli that impart changes on an urban system. This study uses Core-based Statistical Areas (CBSAs) as the primary municipal level to be the focus of the analysis. It also assumes that PM2.5 particulate observed in a given CBSA was also generated in that same CBSA since these areas are tied to surrounding areas socioeconomically by means of transportation and commuting patterns.

Additionally, energy sources on a state level are assumed to be an adequate proxy for energy production plants within a vicinity of CBSAs. Variables within the commuting habits section together represent a valid quality measurement of transportation infrastructure for each metropolitan area. This measurement is crucial for explaining PM2.5 levels since traffic is a major source for urban air pollution. Next, industry characteristics included for each urban agglomeration account for additional pollution sources, specifically manufacturing, transportation/warehousing and utilities due to the high level of combustion required for each. Finally, demographic factors assist in observing differences in CBSAs in

[1] "Understanding Particle Pollution," Environmental Protection Agency, accessed 11/14/2015 http://www3.epa.gov/airtrends/aqtrnd04/pmreport03/pmunderstand_2405.pdf.

[2] Joseph Pinto, Allen Lefohn, Douglas Shadwick, "Spatial Variability of PM2.5 in Urban Areas in the United States," *Journal of the Air & Waste Management Association (2004),* 440-449.

terms of their respective size and median income, as this study hypothesizes that denser cities emit less pollution.

Attempting to measure PM2.5 pollution across such a large area is difficult largely because of the long lifecycle of the particles[3] and limited number of measuring stations present in each CBSA. Particulate rates can be heavily impacted by variations in local geography, meaning that mountainous regions or regions with very few measuring stations may only accurately represent hyper-localized particulate levels. Consequently, the analysis incorporated only urban areas with a population size of more than 100,000 residents that are also monitored by more than 3 measurement stations, which reduced the sample size from 352 CBSAs to 98.

Results

Visualizing the data and investigating the normality of the distribution of the dependent variable provided useful insight into the shape of the data before running a regression. The yearly pollution mean appeared to be normally while also having a positive relationship with the CBSA population density (see Figure 1-1 and 1-2 for distributions and K-S test results). In order to select the relevant variables to be included in the model, the correlation matrix served as a critical tool for avoiding multicollinearity (See Table 1-2). As expected, some of the variables are highly correlated, especially the variables within the Industry, Energy Consumption, and Transportation sectors.

? normally distributed

The final regression yielded an R-squared value of 0.538, meaning that approximately 54% of the variance can be explained using the chosen variables. An increase in CBSA density increases pollution levels, while higher usage of petroleum and natural gas decreases pollution, which may stem from the correlation of total fossil and coal. While establishing causality is difficult, denser urban areas are also characterized by increased industrial activity, both of which are likely to cause higher consumption of energy and a greater dependency on fuel intensive modes of transportation. The impact of transportation on observed PM2.5 levels is represented solely by the share of commuters using bikes and not by car or public transportation as one might expect (See Figure 1-3). Biking therefore appears to be the differentiator between two cities of same density despite being the lowest percentage of commuter method used on average, as car usage or public transit is not quantitatively different across similarly dense metropolitan areas.

Prevailing indicators of variations in observed levels of PM2.5 are transportation as a function of metropolitan bikeability and energy usage by percentage of petroleum and natural gas used. The increase in the bikeability of an area has been shown to reduce levels of PM2.5 when controlling for other variables. Energy usage was driven primarily by the presence of petroleum and natural gas, as those negatively correlate with coal usage - a considerably worse polluter than the previous two. With greater density in urban areas comes an increase in energy usage and, subsequently, higher PM2.5 levels.

Future Work:

This analysis represents only the first step in better understanding the sources of PM2.5 and the features of cities that can either increase or decrease its creation. To better understand this question, further analysis could include a cluster analysis that grouped the 98 CBSAs into similar categories, then look at the similarities between the observations within each.

[3] Jeffrey R. Brook, Richard L. Poirot, Tom F. Dann, Patrick K.H. Lee, Carrie D. Lillyman, Thera Ip, "Assessing Sources of PM2.5 in Cities Influenced by Regional Transport," *Journey of Toxicology and Environmental Health (2007)*, 191-9.

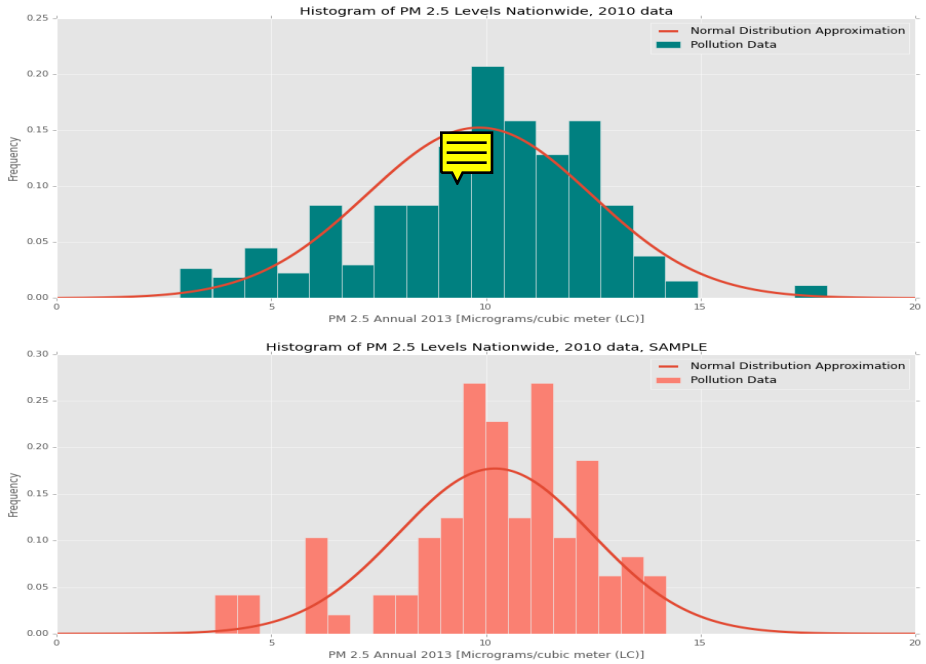

Figure 1-1: Distribution of annual PM2.5 distribution across sampled regions

OLS Regression Results

Dep. Variable:	np.log(pollution_mean)	R-squared:	0.538
Model:	OLS	Adj. R-squared:	0.517
Method:	Least Squares	F-statistic:	25.31
Date:	Sat, 14 Nov 2015	Prob (F-statistic):	6.39e-14
Time:	17:54:50	Log-Likelihood:	26.346
No. Observations:	92	AIC:	-42.69
Df Residuals:	87	BIC:	-30.08
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf.	Int.]
Intercept	1.2286	0.170	7.233	0.000	0.891	1.566
np.log(cbsa_density)	0.0894	0.018	4.897	0.000	0.053	0.126
np.log(petroleum + np.exp(-100))	-0.2961	0.117	-2.526	0.013	-0.529	-0.063
np.log(natural_gas + np.exp(-100))	-0.0676	0.030	-2.242	0.027	-0.127	-0.008
np.log(bike + np.exp(-100))	-0.1504	0.025	-6.065	0.000	-0.200	-0.101

Omnibus:	5.906	Durbin-Watson:	1.701
Prob(Omnibus):	0.052	Jarque-Bera (JB):	9.426
Skew:	0.017	Prob(JB):	0.00898
Kurtosis:	4.568	Cond. No.	62.1

Regression Table 1-2: Regression with only significant explanatory variables

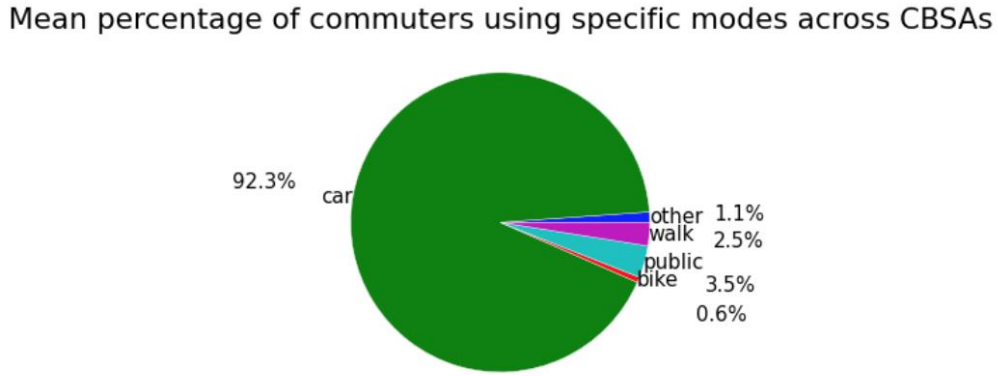

Figure 1-3: Observed commuter modal split