report

Title of the report

Introduction

World Health Organization estimates 7 million death annually related to air pollution(WHO 2020). Air quality plays an vital role in public health. In order to improve air quality, this study aims to pinpoint the factors of air quality quantitatively. The factors under study are rainfall, population density, income per capita, added value of companies and adjacency to coast. To shed light on the relation between air quality and its factors, we adopts better subset selection algorithm. A better subset selection is favourable due to its better fit in terms of smaller residual sum of squares. The algorithm yields a better fit than a subset without optimization as the result of its monotonicity (Xiong 2014). Another strong motivation for a subset selection is the avoidance of unnecessary measurement or sampling. Smaller sets of variables may reduce measurement cost.
This study attempts to answer: Are societal impact stronger than geographic impact to air quality in 1972's California? (Ruoying: Here I group our variables into two groups. One is societal, the other is geographic. Open to discussion!!) We implemented better subset selection based on the algorithm of maximization in majorization. As the optimal subset is unknown, we ran the selection with all possible variable number. In this report, the implementation details in the Method section follows the introduction and preprocessing of data in the Data section. The research question is answered in the Result section.

Data

Previous study has proven or suggested the relation between the five chosen variables and air quality. Both natural and anthropogenic events attribute to air quality in the atmosphere. A difference between generating air pollutant and distributing air pollutant is noted. The generation of air pollution relates to production and consumption from society (Baklanov, Molina, and Gauss 2016). Accordingly, this study captures the relation by population density, income per capita and added values from companies. Furthermore, the distribution of air pollution mainly depends on the wind field (Leelőssy et al. 2014), which is quantified by the variable of adjacency to coast and rainfall in this study as they both reflect the wind field's condition.
From the econometrics dataset for 1972's air quality in California, this study examines five independent variables and one dependent variable (r-project 2020). The dependent variables under study are rainfall(inch), population density(per square mile), income per capita(dollar), added value of companies(dollar) and adjacency to coast(binary). The dataset has 30 set of observations for the 6 variables in this study. As the unit of each variable is heterogeneous, we scaled all independent variables before regression so as to ensure fair interpretation of the model coefficents. After scaling, each independent variable has the mean of 0 and the variance of 1.

	Dependent variable:				
	Air Quality				
	(1)	(2)	(3)	(4)	(5)
Coastal Area	-13.73^{***}	-13.83^{***}	-14.85^{***}	-14.98^{***}	-15.57^{***}
	(-2.54)	(-2.76)	(-2.98)	(-3.03)	(-3.19)
Value Added	-	9.26	3.63	3.33	4.090
		(0.86)	(0.34)	(0.31)	(0.39)
Median Income	-	-	6.3	7.2	6.9
			(0.58)	(0.67)	(0.65)
Population Density	-	-	-	-2.94	-3.04
				(-0.63)	(-0.66)
Rain	-	-	-	-	3.38
					0.72
Intercept	104.700^{***}	104.700^{***}	104.700^{***}	104.700^{***}	104.700^{***}
	21.34	(4.419)	(4.419)	(4.419)	(23.69)
Observations	30	30	30	30	30
R^2	0.383	0.383	0.383	0.383	0.383
Adjusted R^2	0.254	0.254	0.254	0.254	0.254
Note:			*p<0.1; **p<0.05; ***p<0.01		

Method

Result

Reference

Baklanov, Alexander, Luisa T Molina, and Michael Gauss. 2016. "Megacities, Air Quality and Climate." *Atmospheric Environment* 126: 235–49.

Leelőssy, Ádám, Ferenc Molnár, Ferenc Izsák, Ágnes Havasi, István Lagzi, and Róbert Mészáros. 2014. "Dispersion Modeling of Air Pollutants in the Atmosphere: A Review." *Central European Journal of Geosciences* 6 (3): 257–78.

r-project. 2020. "Ecdat: Data Sets for Econometrics." World Health Organization. https://cran.r-project.org/web/packages/Ecdat/index.html.

WHO. 2020. "Air Pollution." World Health Organization. https://www.who.int/health-topics/air-pollution#tab=tab_1.

Xiong, Shifeng. 2014. "Better Subset Regression." *Biometrika* 101 (1): 71–84.