Neighborhood Dynamics and State Policy Variation: A spatio-temporal investigation of racial disparities in air pollution exposure

Austin Carter and Victoria Sass

March 13, 2017

Introduction

Research on environmental inequality grew out of several studies conducted during the 1980s which found evidence of the disproportionate presence of environmental hazards in neighborhoods composed largely of people of color and those from the lower rungs of the socioeconomic ladder (GAO 1983; CRJ 1987). While these studies highlighted an important area of research, they were limited in their ability to generalized to a broader population. Spurred by these findings, however, researchers began investigating this topic more thoroughly during the 1990s to better understand both the causes and mechanisms through which this inequitable distribution of environmental ills came about. Due to a variety of inconsistencies in geographic scope and scale, the definition of environmental hazards, and data quality more generally, the findings from this literature were somewhat mixed. However, with the improvement of both data quality and the methods used to explore the question of environmental inequality, the evidence overwhelmingly supports the argument that people of color and those of lower socioeconomic status are more likely to be exposed to environmental hazards (Brown 1995, Szasz and Meuser 1997, Institute of Medicine (US) Committee on Environmental Justice 1999, Evans and Kantrowitz 2002, Bullard, Mohai et al. 2008).

One of the main questions explored by this research in recent years has been to ascertain the mechanisms through which these inequalities came about and tend to persist. The two major processes put forth are what Mohai and Saha refer to as "disparate siting" and "post-siting demographic change" (Mohai and Saha 2015). In the former, the siting of a polluting facility or hazardous waste sites happens more often in neighborhoods where people of color and the poor tend to live whereas in the latter the demographic makeup of the area around a facility becomes more heavily concentrated by people of color and the poor over time. Understanding how these mechanisms operate and which process explains the lionshare of the variation in neighborhood disparities is extremely important for urban and environmental policymakers as it has the ability to inform appropriate ways by which equality can be achieved. However, it has been noted by scholars in this field that the structural nature of racism tends to make any clear-cut distinctions among these mechanisms purely abstract (Bonilla-Silva 2007, Mohai et al 2009). Additionally, in the decades since the emergence of this literature, there have been fundamental shifts in standards for emissions as well as the structure of the U.S. economy. The former of these has resulted in a secular decline in air pollution emissions monitored by the Environmental Protection Agency (EPA 2017). Simultaneously, the shift of the American economy from one structured around manufacturing and industrial output to one that is mostly service-based has changed the literal landscape of American neighborhoods (Wilson 1987). Lastly, the level of residential segregation has been declining in most American cities since its height in the 1980s (Cutler et al 1999). Despite these changes, it has yet to be determined what influence they have had on the share to which certain groups of people are exposed to air pollution. Furthermore, despite stricter federal emissions standards put forth in amendments to the Clean Air Act in 1990, states are tasked with enacting their own methods and rules for implementation so long as the federal levels are not exceeded. This has led to significant variability in state-level policy regarding air emissions. The way this state-level variation has played out at the neighborhood level has not yet been explored in the environmental inequality literature. This study seeks to fill this gap in the literature and contribute to a better understanding of how broad environmental policy agendas such as the Clean Air Act may be interacting with other macro structural processes at the local level.

Data

The data for this research comes from three sources: the Neighborhood Change Database (NCDB), the EPA's Toxic Release Inventory (TRI), and the EPA's Enforcement and Compliance History Online (ECHO) Database.

Neighborhood Change Database

Created by Geolytics in association with the Urban Institute, this database allows users to compare U.S. census tracts over time by employing a methodology that normalizes all tracts to 2010 geographical boundaries. For this analysis we utilize data on census tracts from 1990, 2000, and 2010 including variables on racial/ethnic composition (proportions of non-hispanic white, non-hispanic black, non-hispanic other race, and hispanic), proportion employed in manufacturing, population density, median family income, proportion of high school graduates, proportion of college (or greater) graduates, and poverty rate.

Toxic Release Inventory

The EPA requires all large-scale, industrial facilities to report the number (in pounds) of toxic emissions they release each year. This dataset therefore contains the location (latitude and longitude) and estimated pounds of emissions for each facility that handles more than 10,000 pounds of a toxic chemicals each year. We were able to utilize the air emissions portion of these data to create a relative average emissions for each tract in the continental United States. This was done using GIS software to create a raster heat map from the vector map of facilities. Using the Heatmap plugin in QGIS and setting the radius for each kernel density function to 1.5 miles, (consistent with prior literature) a raster map of the continental United States was populated with 400m by 400m pixels. The values of these pixels reflected their relative distance to polluting facilities with higher values for pixels that were closer to facilities. This method was particularly appealing because it allowed for a pixel that was near multiple facilities to be more heavily weighted and for those pixels that were near extremely high emitters to reflect that quantitative difference.

Using this heatmap raster the SpatialStats Plugin was used to average all pixels within a census tract (using 2010 boundaries) to arrive at an average emissions level for each tract. This same process was carried out for the years 1990, 2000, and 2010 and merged with the Neighborhood Change Database. It should be noted that the "average tract emissions" variable does not have an interpretable unit of analysis due to the way it is created. It can, however, be understood as a relative measure of toxic emissions, which will allow for an analysis comparing relative emissions by tract.

Enforcement and Compliance History Online

Lastly, to get a sense of the state-level variation with respect to the enforcement and compliance of the Clean Air Act standards we utilized the ECHO database which contains information on evaluations, violations, and enforcement actions. To maintain consistency, we subset the data to include only TRI-reporting facilities and matched it with the ones contained within our TRI dataset for the years 1990, 2000, and 2010. Due to the volume of information contained within this database we decided to run a Principal Component Analysis (PCA) on eight of the variables we felt captured important aspects of enforcement and compliance: number of facilities, evaluations, high priority violations, formal actions, informal actions, total stack tests, passed stack tests, and failed stack tests. From the resulting PCA we selected the first three loadings, which captured 85% of the variation contained within the eight aforementioned variables. We then merged these three state-level PCA variables to our full dataset for each of the years 1990, 2000, and 2010.

Methods

The two main investigations of interest driving this analysis are 1) the extent to which the racial composition of a neighborhood has been predictive of air pollution levels, and how this has changed since the passage of the amendments to the Clean Air Act in 1990 and 2) how state-level variation in enforcement and compliance of these national EPA standards helps to explain such tract-level variation in the relationship between race and emissions.

Cross-sectional analysis

To assess the first of these questions we carried out three cross-sectional multilevel models, one for each year in our dataset. In each model we included random effects for the state and fixed effects for our predictor variables. Our relative average air emissions variable was the outcome of interest. In order to assess which control variables should be included, we used Bayesian Model Averaging for generalized linear models with a gaussian distribution. Any variable that were found to be significant in any of the iterations were included as controls in each of the models. For each year of cross-sectional data we tested the same set of permutations of race and state level enforcement and compliance variables. Our full set of models for cross-sectional analysis were:

1. A baseline linear model including only controls
2. A multilevel model including controls and state random intercepts
3. Race variables
4. Enforcement and compliance variables[1]
5. Race variables and enforcement and compliance variables[1]
6. All race variables interacted with enforcement and compliance variables[1]
7. Black interacted with enforcement and compliance variables[1]
8. White interacted with enforcement and compliance variables[1]
9. Hispanic interacted with enforcement and compliance variables[1]
10. Other interacted with enforcement and compliance variables[1]

[1] Including controls and state random intercepts

For each time period, we chose our best model using the Bayesian information criterion (BIC), which compares the likelihood of each model while penalizing for the number of parameters included.

Change over time analysis

For the second question we restructured our data and ran two change models, 1990 to 2000 and 2000 to 2010. This allowed us to incorporate the time dependency between our observations and therefore attempt to capture the multiple structural changes (i.e. segregation, manufacturing, and air pollution decline) occurring over the course of our data's time frame. These models were also hierarchical with state random effects. Using the same methodology for choosing control variables, we built out and compared the following set of models:

1. A baseline linear model including only controls
2. A multilevel model including controls and state random intercepts
3. Average race variables
4. Change in race variables[2]
5. Average and change in race variables[2]
6. Enforcement and compliance variables[2]
7. All race variables and enforcement and compliance variables[2]
8. All average race variables interacted with enforcement and compliance variables[2]
9. Average black interacted with enforcement and compliance variables[2]
10. Average white interacted with enforcement and compliance variables[2]
11. Average hispanic interacted with enforcement and compliance variables[2]
12. Average other interacted with enforcement and compliance variables[2]
13. All change in race variables interacted with enforcement and compliance variables[2]
14. Change in black interacted with enforcement and compliance variables[2]
15. Change in white interacted with enforcement and compliance variables[2]
16. Change in hispanic interacted with enforcement and compliance variables[2]
17. Change in other interacted with enforcement and compliance variables[2]

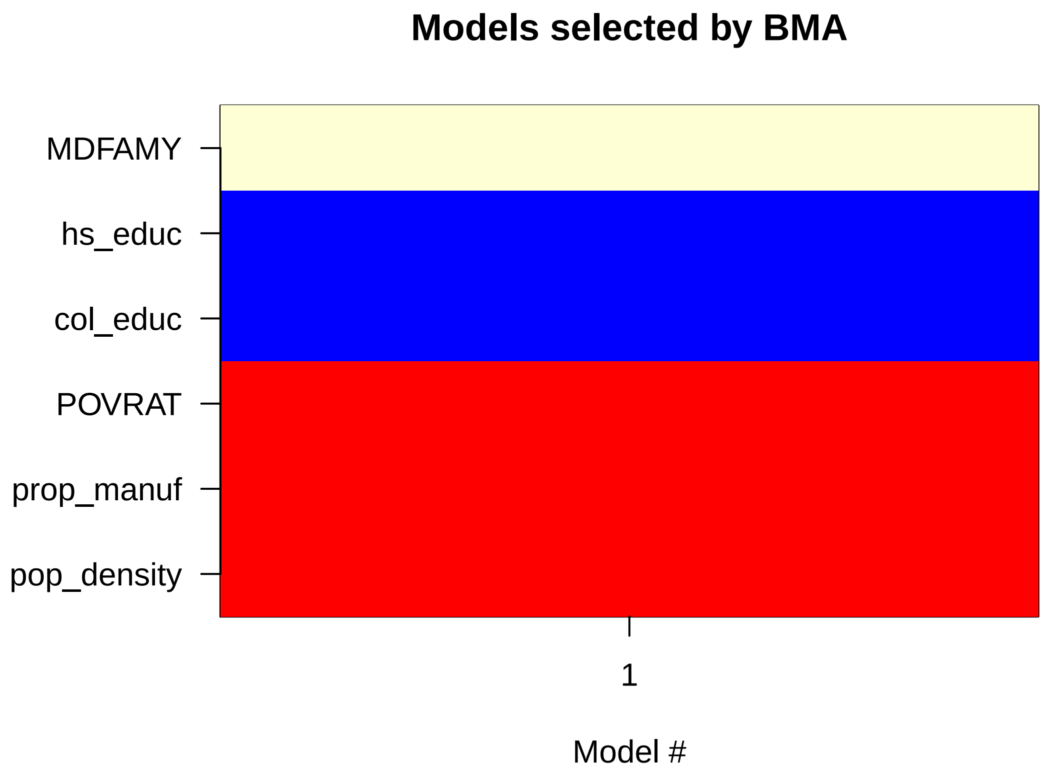
To assess whether or not there was sufficient state level variation to include state random intercepts, we plotted the random intercepts from each of our best models.

Results

1990 cross-sectional analysis

Education, poverty, manufacturing and population density were found to be significant controls. The best model was the multilevel model including interactions between the proportion of the neighborhood that was Hispanic and the enforcement and compliance variables. In this model, the enforcement and compliance variables were not significant, but the interaction of the first and second PCA variable with the proportion were significant and positive.

[2] Including controls and state random intercepts

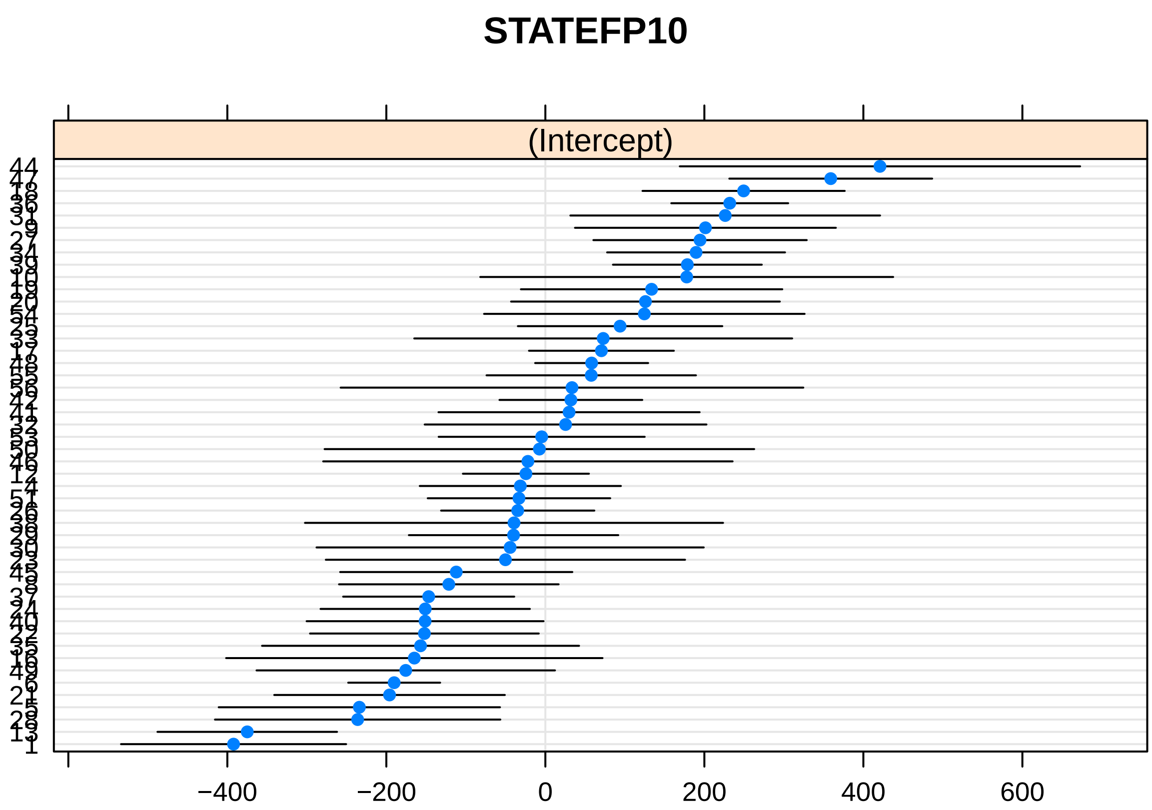



```
##     model     bic        diff
##  1:    M0 1345551 3625.89072
##  2:    M1 1345218 3293.08083
##  3:    M2 1345249 3324.14507
##  4:    M3 1342078  152.81503
##  5:    M4 1342109  183.75728
##  6:    M5 1341974   48.78215
##  7:    M6 1342107  182.24826
##  8:    M7 1342034  109.03017
##  9:    M8 1341925    0.00000
## 10:    M9 1342097  171.89189
```

```
##
## \begin{table}
## \begin{center}
## \begin{tabular}{l c }
## \hline
##  & Model 1 \\
## \hline
## (Intercept)                  & $395.22^{***}$   \\
##                              & $(114.86)$       \\
```

```
## Log\_PC1                    & $54.41$          \\
##                             & $(30.41)$        \\
## SHRHSP                      & $3253.19^{***}$  \\
##                             & $(297.28)$       \\
## Log\_PC2                    & $-29.67$         \\
##                             & $(48.16)$        \\
## Log\_PC3                    & $-38.88$         \\
##                             & $(49.41)$        \\
## hs\_educ                    & $-1676.60^{***}$ \\
##                             & $(185.24)$       \\
## col\_educ                   & $-793.60^{***}$  \\
##                             & $(116.47)$       \\
## POVRAT                      & $1437.59^{***}$  \\
##                             & $(105.84)$       \\
## prop\_manuf                 & $2009.69^{***}$  \\
##                             & $(131.69)$       \\
## pop\_density                & $-3013.97$       \\
##                             & $(3048.25)$      \\
## Log\_PC1:SHRHSP             & $1332.34^{***}$  \\
##                             & $(108.78)$       \\
## SHRHSP:Log\_PC2             & $1032.17^{***}$  \\
##                             & $(208.03)$       \\
## SHRHSP:Log\_PC3             & $-290.87$        \\
##                             & $(153.78)$       \\
## \hline
## AIC                         & 1341787.30       \\
## BIC                         & 1341925.07       \\
## Log Likelihood              & -670878.65       \\
## Num. obs.                   & 72049            \\
## Num. groups: STATEFP10      & 48               \\
## Var: STATEFP10 (Intercept) & 37129.43          \\
## Var: Residual               & 7157866.89       \\
## \hline
## \multicolumn{2}{l}{\scriptsize{$^{***}p<0.001$, $^{**}p<0.01$, $^*p<0.05$}}
## \end{tabular}
## \caption{Statistical models}
## \label{table:coefficients}
## \end{center}
## \end{table}
##  FALSE

## $STATEFP10
```



2000 cross-sectional analysis

Education, poverty, and manufacturing were found to be significant controls. The best model was the multilevel baseline model that only included controls. None of the models that included our variables of interest sufficiently improved the likelihood of the model given the increase in parameters.

2010 cross-sectional analysis

Family income, education, poverty, and manufacturing were found to be significant controls. The best model was the multilevel model that had only the racial variables. None of the models that included our variables of interest sufficiently improved the likelihood of the model given the increase in parameters.

1990-2000 change analysis

Emissions in 1990, poverty, high school education, and manufacturing were found to be significant controls. The best model was the multilevel model including interactions between proportion White and the enforcement and compliance variables. Every parameter coefficient was statistically significant, and the enforcement and compliance variables had the expected positive correlation with declines in emissions. Both the proportion white and its interaction with enforcement and compliance had negative coefficients, implying that the declines in emissions for a given level of enforcement and compliance were smaller for tracts with a higher proportion of white residents.

2000-2010 change analysis

Emissions in 2000 and education were found to be significant controls. The best model was the multilevel model including interactions between the change in proportion White and the enforcement and compliance variables. The every parameter coefficient was statistically significant, and the enforcement and compliance variables had the expected positive correlation with declines in emissions. Change in proportion white, as well as two of the interactions between change in proportion white and the enforcement and compliance variables, had negative coefficients. The third interaction had a smaller positive coefficient.

Random intercepts

The plots of the random intercepts from each of our best models exhibit high levels of within- and between-state variation. In each of the models, a significant proportion of random intercepts do not contain zero, and there is heterogeneity in the variance of intercepts.

Discussion & Conclusion

Our investigation highlights both the historical racial disparities in exposure to emissions and the critical contributions of environmental regulations and enforcement to ending this unequal exposure. Even in 2010, 20 years after the passing of the Clean Air Act, communities with a greater share of black or hispanic residents are exposed to more pollution as was the case in 1990. At the same time, the average proportion white in a tract and increases in the proportion white were found to be associated with smaller declines in emissions for that tract. Our results imply that white communities experience less improvement in air quality for a given level of enforcement and compliance. While some may consider this inequality problematic, we believe that policies should be equitable, which in the case of air pollution means that larger declines in air pollution need to occur in communities of color because of the unequal distribution of emissions prior to the institution of the act.

The inclusion of random intercepts in our models was justified by their contribution to an improved BIC. Further, the high level of variance between states and the high within-state variance in a number of states implies that our multilevel modelling impacted state-level inference. Future investigation of the dynamics of emissions control and race should similarly utilize random effects.

There were a number of limitations to our study, including the large time gaps between observations and the aggregate nature of the data. The addition of stricter air pollution standards to the Clean Air Act in 1990 made that year a natural starting point for this analysis, especially given that it coincided with required reporting of large-scale polluting facilities and the accompanying evaluation and compliance data. However, census variables on our controls and main predictors of interest were only available in 1990, 2000, and 2010 (unfortunately the American Community Survey which provides estimates for years between decennial censuses was not fully implemented until 2005). Therefore, our data was limited to assessing large-scale structural trends on micro-neighborhood changes with data that was only able to capture three points in time. Ideally, future exploration of this area of research will have more temporal variation, allowing for a better assessment of the mechanisms characterizing the macro- and micro- changes.

Relatedly, while this study sought to investigate neighborhood change, much more inference could be made if these data were individual-level rather than tract-level aggregations. Given individual-level covariates, as well as the original point-location sources of polluting facilities, a much more fine-grained analysis could be conducted at this level, allowing for more causal conclusions to potentially be drawn from the results.

References

Bonilla-Silva E 1997 Rethinking racism: toward a structural interpretation Am. Sociological Rev. 62 465-80.

Brown, P. (1995). "RACE, CLASS, AND ENVIRONMENTAL-HEALTH - A REVIEW AND SYSTEMATIZATION OF THE LITERATURE." Environmental Research 69(1): 15-30.

Bullard, Robert D., Paul Mohai, Robin Saha, and Beverly Wright. 2008. "Toxic Wastes and Race at Twenty: Why Race Still Matters after All of These Years Environmental Justice: Making It a Reality." Environmental Law 38:371-412.

Commission for Racial Justice (1987). Toxic wastes and race in the United States: A national report on the racial and socioeconomic characteristics of communities with hazardous waste sites. New York, United Church of Christ.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." Journal of Political Economy 107(3):455-506.

EPA, US. "National Air Quality: Status and Trends of Key Air Pollutants." (https://www.epa.gov/air-trends).

Evans, G. W. and E. Kantrowitz (2002). "Socioeconomic status and health: The potential role of environmental risk exposure." Annual Review of Public Health 23: 303-331.

GAO, (U. S. Government Accountability Office). 1983. "Siting of Hazardous Waste Landfills and Their Correlation With Racial and Economic Status of Surrounding Communities." (RCED-83-168). (http://www.gao.gov/products/RCED-83-168).

Institute of Medicine (US) Committee on Environmental Justice (1999). Toward environmental justice: Research, education, and health policy needs, National Academies Press (US).

Mohai P, Pellow D and Roberts J T 2009a Environmental justice Annu. Rev. Environ. Resour. 34 405-30.

Mohai, Paul and Robin Saha. 2015. "Which Came First, People or Pollution? A Review of Theory and Evidence from Longitudinal Environmental Justice Studies." Environmental Research Letters 10(12):125011.

NCDB. "GeoLytics - US Census Demographic and GIS Data Products." (http://www.geolytics.com/USCensus,Neighborhood-Change-Database-1970-2000,Products.asp).

Szasz, A. and M. Meuser (1997). "Environmental Inequalities: Literature Review and Proposals for New Directions in Research and Theory." Current Sociology 45(3): 99-120.

TRI. "Toxics Release Inventory (TRI) Program." (https://www.epa.gov/toxics-release-inventory-tri-program).

Table 1: Change Models Variables

Name	Description
	Change Models
emissions	Average tract-level emissions
SHRNHB	Proportion non-hispanic Black living in tract
SHRHSP	Proportion Hispanic/Latinx living in tract
prop__other__race__all	Proportion non-hispanic Other race living in tract
MDFAMY	Median family income of tract
hs__educ	Proportion of high school graduates in tract
col__educ	Proportion of college (or more) graduates in tract
POVRAT	Poverty rate in tract
prop__manuf	Proportion of civilian workforce 16+ years working in manufacturing in tract
pop__density	Population density of tract
Log__PC1	Logged version of first principal component for state-level evaluation/ compliance variation
Log__PC2	Logged version of second principal component for state-level evaluation/ compliance variation
Log__PC3	Logged version of third principal component for state-level evaluation/ compliance variation
STUSAB	State Abbreviations

Wilson, William J. 1987. The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy. University of Chicago Press.

Variable dictionary

Table 2: Cross-Sectional Models Variables

Name	**Description**
	Cross-sectional Models
emissions_decrease	The amount average tract-level emissions decreased from time1 to time2
emissions_1990	Average tract-level emissions in 1990
emissions_2000	Average tract-level emissions in 2000
avg_white	Average proportion non-hispanic White at time1 and time2 in tract
white_change	Difference between proportion white from time1 to time2 in tract
avg_hs_educ	Average proportion of high school graduates at time1 and time2 in tract
avg_col_educ	Average proportion of college (or more) graduates at time1 and time2 in tract
avg_povrat	Average poverty rate at time1 and time2 in tract
avg_manuf	Average proportion of civilian workforce 16+ years working in manufacturing at time1 and time2 in tract
avg_log_PC1	Average logged version of first principal component for state-level evaluation/compliance variation at time1 and time2
avg_log_PC2	Average logged version of second principal component for state-level evaluation/compliance variation at time1 and time2
avg_log_PC3	Average logged version of third principal component for state-level evaluation/compliance variation at time1 and time2
STUSAB	State Abbreviations