ESSAYS ON INCOME INEQUALITY AND THE ENVIRONMENT

by

JOHN VOORHEIS

A DISSERTATION

Presented to the Department of Economics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2016

DISSERTATION APPROVAL PAGE

Student: John Voorheis

Title: Essays on Income Inequality and the Environment

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

Trudy Ann Cameron	Chair
Peter Lambert	Core Member
Caroline Weber	Core Member
Ronald Mitchell	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2016

DISSERTATION ABSTRACT

John Voorheis

Doctor of Philosophy

Department of Economics

June 2016

Title: Essays on Income Inequality and the Environment

This dissertation considers two of the most pressing concerns of the current time, income inequality and exposure to pollution, and provides evidence that these two concerns may in fact be causally linked. In order to do this, I assemble novel datasets on income inequality and pollution exposure, and propose an strategy for causally identifying the effect of the former on the latter.

In the first substantive chapter, I develop a new dataset on income inequality measured at the US state and metropolitan area level. I compare the trends in income inequality measured using different income definitions. In general, pre-tax, pre-transfer income inequality has increased in most states since 1980, but post-fiscal income inequality has seen slow or no growth since about 2000. I conduct inference on how income inequality has changed using a semi-parametric bootstrap method, and consider potential correlates with state-level income inequality. I find that de-unionization is perhaps the most important factor driving rising inequality.

In the second substantive chapter, I leverage satellite-derived remote sensing data on ground-level concentrations for two important pollutants (NO_x and PM2.5) to measure the distribution of pollution exposure. I propose a dashboard approach

to measuring environmental inequality and environmental justice, proposing and applying several candidate measures to the satellite datasets. I find that environmental inequality has largely decreased since 1998, as has average exposure. I consider potential correlations between neighborhood demographics and the distribution of exposure, but find inconclusive results.

In the third substantive chapter, I attempt to resolve this ambiguity by considering whether rising income inequality within metropolitan areas (the subject of the first chapter) might causally affect the distribution of exposure across people (the subject of the second). Using a simulated instrumental variables identification strategy designed to address potential endogeneity due to locational sorting, I find that income inequality decreases the average level of exposure, but increases environmental inequality. I argue this is consistent with the benefits of pollution reduction accruing to the most advantaged, and provide evidence that this may work through the political system: inequality increases the responsiveness of politicians to the environmental demands of the rich.

CURRICULUM VITAE

NAME OF AUTHOR: John Voorheis

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Eastern Michigan University, Ypsilanti, MI

DEGREES AWARDED:

Doctor of Philosophy, Economics, 2016, University of Oregon
Master of Science, Economics, 2012, University of Oregon
Master of Arts, Economics, 2011, Eastern Michigan University
Bachelor of Science, Economics, 2009, Eastern Michigan University

AREAS OF SPECIAL INTEREST:

Income Inequality
Environmental Economics
Political Economy
Public Finance

GRANTS, AWARDS AND HONORS:

Young Scholar Grant, Washington Center For Equitable Growth, 2015

Edward G. Daniel Scholarship, University of Oregon, 2015

PhD Research Paper Award, University of Oregon, 2014

Kleinsorge Summer Research Fellowship, University of Oregon, 2013

Everett D. Monte Scholarship, University of Oregon, 2013

Dale Underwood Award, University of Oregon, 2012

Graduate Teaching Fellowship, University of Oregon, 2011-2016

PhD Research Paper Award, University of Oregon, 2014

ACKNOWLEDGEMENTS

I am grateful for the advice and guidance I've received from from Trudy Ann Cameron, Peter Lambert and Caroline Weber. Many helpful comments from Joe Wyer made these papers much better. I have also benefitted from much useful feedback from participants at the University of Oregon Micro Group, the 2015 Winter School on Inequality and Social Welfare Theory, the 2015 AERE Annual Meeting, the 2015 Southern Economics Association Annual Meeting and the 2016 Society for Benefit Cost Analysis Annual Meeting. Any remaining errors are my own.

This research was made possible by a Major Research Instrumentation grant from the National Science Foundation, Office of Cyber Infrastructure, "MRI-R2: Acquisition of an Applied Computational Instrument for Scientific Synthesis (ACISS)," Grant #: OCI-0960354, and was directly supported by a grant from the Washington Center for Equitable Growth and by generous assistance from the Ray Mikesell Foundation at the University of Oregon.

To my father, James Voorheis.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The dual challenges of rising income inequality and environmental degradation are among the most pressing considerations facing policymakers in the contemporary United States. In broad terms, my research examines the relationship between the distribution of income and the distribution of exposure to air pollution. The three substantive chapters provide new insights concerning income inequality, environmental inequality, and examines the relationship between income ineuqality and the distribution of pollution exposure. Chapter II addresses the measurement of spatially disaggregated income inequality (within states and metropolitan areas). It offers a novel dataset on income inequality, and new insights into the causes of rising income inequality. Chapter III considers the broad question of how to measure environmental inequality in a normatively sensible way, and applies these measurement tools to data on pollution exposure from satellite remote sensing data. Chapter IV leverages the measurement work in each of the first substantive two chapters to explore the causal effect of income inequality on environmental quality, and explores potential political economy avenues by which these effects may occur.

Expanding on this brief summary, Chapter II considers the quantification of income inequality measured at the US state and metropolitan area level, using income micro-data from the Census Bureau. Measuring inequality using Census Bureau survey data can be challenging, chiefly due to the fact that incomes are top-censored. To address this challenge, I adapt a semi-parametric multiple imputation method proposed by Jenkins, et al. (2009) to the new task of measuring sub-national income inequality. Further, I extend this method to perform inference on changes in income inequality via semi-parametric bootstrapping. Using this new semi-parametric bootstrap method, I find that income inequality in most states and metropolitan areas has increased since the mid-1990's, a result which stands in contrast to some literature on inequality measured with survey data. I then consider potential determinants of state income inequality. Consistent with a long line of literature starting with Dinardo, Fortin and Lemieux (1996), among the most (if not the most) important explanation for the rise in inequality since the 1970s has been the decline in union density. In addition to these results, this paper produces a dataset of state and MSA income inequality over time. This dataset is additionally used in Chapter IV, as well as several projects by myself and other researchers that are currently in progress.

Chapter III considers how to measure environmental inequality in an analogous and normatively sensible fashion, and how environmental inequality has evolved within the United States. I use two data sources of data on the distribution of exposure to air pollution: satellite observations of NO_x from NASA's Aura satellite, and estimates of ground level PM2.5 derived from several extant satellites. I find that average exposure to these particular pollutants has decreased both nationally. Additionally, I find that the distribution of pollution exposure has become more equal since the late 1990's (using the PM2.5 data) and since 2005 (using the NO_x data). Using re-centered influence function regressions, which identify how variations in individual census tract demographics affect national inequality, I consider how census tract demographic characteristics are related to environmental inequality. There is evidence that many of the dimensions of disadvantage highlighted in the environmental justice literature (chiefly dimensions of race and class) are correlated with environmental inequality.

Chapter IV builds on the results from the first two substantive chapters by exploring the causal effect of income inequality on the distribution of pollution exposure. I leverage the dataset developed in Chapter II on MSA level income inequality as well as the dataset developed in Chapter III on the distribution of NO_x exposure, and use a simulated instrumental variables strategy to achieve causal identification. I find that metropolitan area income inequality decreases average pollution exposure. There is also evidence that income inequality increases pollution exposure inequality as well as the gap in exposure between advantaged and disadvantaged groups. Together, these results imply that the most advantaged members of society are disproportionately reaping the benefits of pollution exposure reduction. I propose a political economy explanation for these results, showing in particular that income inequality appears to increase the environmentalism of US politicians, a result consistent with increasing responsiveness to the environmental demands of the rich.

The final chapter of this dissertation concludes with a summary of this research agenda as it currently stands, highlighting the contributions to existing literatures on income inequality, environmental justice and political economy. This final chapter details potential future extensions of the line of research outlined in this dissertation. The final chapter also summarizes several research projects which are underway which run parallel to the work in this dissertation.

CHAPTER II

STATE AND METROPOLITAN AREA INCOME INEQUALITY IN THE UNITED STATES: TRENDS AND DETERMINANTS

Introduction

Rising income inequality has become one of the most pressing concerns of the modern era. An extensive literature has attempted to explain why the distribution of income in the United States has become more unequal. This literature has almost exclusively relied on national-level data, however, both as a matter of convenience (official, national income distribution statistics are often made available by national statistical agencies) and as a matter of analytical preference (for many studies, countries are a natural unit of observation). There has been less work done on the longer run trends in, and determinants of, income inequality measured at sub-national geographic scales.

There are several reasons, to study income inequality at smaller geographic resolution. Inequality can generally be thought of as being composed of between-subgroup and within-subgroup inequality. In the United States, inequality between states has actually been decreasing, implying that within-State income inequality has been the driver of overall income inequality. This in turn suggests that in order to understand national trends in income inequality it is necessary to first understand what has been happening at the sub-national level. Additionally, a small but growing literature has begun to examine whether rising income inequality affects other outcomes of interest. These effects may be the results of individuals' responses to rising inequality. Individuals are unlikely to be able to accurately perceive the national level of income inequality, but are more likely to be able to accurately perceive inequality in their own metropolitan area or state.[1]

This paper is innovative along two dimensions: first by offering a consistent time series of inequality measures at the state and metropolitan area levels—using Metropolitan Statistical Area (MSA) definitions—over a relatively long time period (as early as 1968 to the present), and second by applying recently developed tools for dealing with top-coded Census Bureau data to sub-national geographic scales. Specifically, I apply the Generalized Beta multiple imputation methodology proposed in Jenkins et al. (2011), previously used only with national-level data, to the State and MSA level. I believe this dataset to be more complete and better constructed than previous efforts.[2] I leverage the large set of information about income sources and household composition available in Current Population survey data to construct income inequality measures using five different definitions of income, which range from an Adjusted Gross income concept that coincides with IRS tax return data to a broad post-fiscal household income definition that considers the effect of taxes, transfers and in-kind social spending that most closely aligns with the disposable income available for consumption. In addition to allowing the analysis in this paper, these data has already proven to be useful in studies where researchers may require measures of inequality at the State or MSA level.

Methodologically, I utilize a number of strategies to analyze how income inequality has changed. I introduce a semi-parametric bootstrap technique for inference on inequality measures, which complements the Generalized Beta II multiple imputation process used to generate point estimates. I use this method throughout the analysis, first in pairwise comparison of inequality measures, and then in a semi-parametric extension of the Barrett

[1]Gimpelson and Treisman (2015) provide some suggestive evidence for this disparity.

[2]One notable exception being the Frank (2009) state-level dataset, which utilizes IRS Statistics of income data and Pareto interpolation from 1913 to 2005

et al. (2014) bootstrap test for Lorenz Dominance. I find that not only is there robust evidence of an increase in inequality, this increase in inequality is not a purely the result of increasing incomes at the top. Although the increase in income inequality is consistent with an increase in the share of the top 1%, there are also substantial changes within the bottom 99% of the distribution. Specifically, the top quartile of the distribution finds itself relatively better off over the past decade, while the bottom 50-75% is unambiguously relatively worse off. I examine several possible explanations for this increase in inequality in subsequent regression analysis, and find that evidence for associations between income inequality and both de-unionization and changes in top marginal tax rates, but less evidence for associations between inequality and changes in technology or human capital accumulation.

The paper proceeds as follows. Section 2 summarizes the related literature on local area income inequality as well as the recent literature on strategies for dealing with top-censored income microdata for the study of income distribution dynamics. Section 3 describes my strategy for constructing the state and MSA income inequality panel datasets, and discusses the dynamics of the resulting income inequality measures. Sections 4 and 5 perform inference concerning the changes in income inequality over the last decade, and use state-level data to analyze the determinants of these changes. Section 6 concludes with directions for future research, included suggested applications of the inequality datasets.

Related Literature

There is substantive literature documenting a substantial increase in income inequality in recent decades. It is almost too large to summarize here, though earlier thorough reviews exist (see Acemoglu (2002) or Atkinson et al. (2011)). The literature on income inequality in the United States can be categorized based on the underlying data source, the income-receiving unit, and the definition of income. Studies based on IRS tax return data (e.g. Piketty and Saez (2003)) use the tax unit as the unit of observation, and the income concept is market income (pre-tax, pre-transfer). Studies based on Census Bureau survey data (from the Current Population Survey, the decennial Census and the American Community Survey) have tended to use the household as the unit of observation, and pre-tax, post-transfer incomes as the income concept.[3] I will use Census Bureau survey data, so I highlight the most relevant literature using this data, as well as the relatively small literature dealing with income inequality within states or metropolitan areas.

Census Bureau Microdata

There are two main feature of Census Bureau data which must be addressed when using these data to analyze income inequality: (1) these data do not measure capital gains (either realized or unrealized) as part of income, and (2) incomes above a certain threshold, which is different for each income source, are "topcoded" for privacy reasons. The topcoding procedures have changed significantly over the history of the CPS, but the basic process has been to replace the true reported income with a top-code value which anonymizes the top income earners in the sample. Both topcoding and the exclusion of capital gains should, other things equal, lead to lower estimates of income inequality than a "true" inequality measure where there is no topcoding and the inequality measure includes capital gains.

[3] Recently there has been some interest in more expansive income definitions (see, e.g. Armour et al. (2014)). I use the standard pre-tax post-transfer income measure in this study, but using more expansive income definitions might be a useful direction for further work.

The exclusion of capital gains can be justified on definition-of-income grounds, as in Armour et al. (2014). Unrealized capital gains do not necessarily represent changes in wealth available for consumption, and hence a reasonable definition of income might exclude these gains. Topcoding presents a more pressing issue, however. One solution, which is still occasionally used, is to simply eschew the use of income inequality measures that are sensitive to the upper tail of the income distribution (e.g. the Gini coefficient of top 1% share) and to use instead measures such as the 90-10 ratio.[4] However, as Burkhauser et al. (2009) note, there is still substantial topcoding in the public-use micro data, even at the 90th percentile of the income distribution, so using the 90-10 ratio does not necessarily eliminate the top-coding problem.

Two approaches have been proposed for correcting for top-coded income data. The first involves collecting suitable cell means data for top incomes from the confidential CPS data, and imputing these cell means in place of topcoded income amounts. The second involves a multiple imputation approach wherein top-coded incomes are imputed as draws from a suitable distribution. The cell means approach was actually used by the Census Bureau in the public CPS data from 1996 to 2010[5], and Larrimore et al. (2008) extends this cell means series back to 1976. The Larrimore series is at least as good as the internal CPS data used by the Census Bureau for official analysis, but still probably understates inequality, since top incomes are still censored. The multiple imputation approach directly addresses the censoring by fitting a parametric model of the income distribution from the raw, topcoded data, and then simulates distributions semi-parametrically by drawing replacement values for topcoded incomes from the fitted model. Jenkins et al. (2011) offers a detailed description of the suggested process. This approach is the one I adopt

[4]The 90-10 ratio is the ratio of income at the 90th percentile to income at the 10th percentile.

[5]From 2011 onwards, the CPS has a new top-coding process wherein top incomes are randomly swapped within a cell, rather than replaced with the cell means.

in the analysis of local area income in this paper. More detail on my modifications of the multiple imputation approach can be found in section 2.3.

Local Area Income Inequality

The above discussion of the current debates and developments in the inequality literature has focused on national-level income inequality within the US. With notable exceptions, very little of the economics literature on inequality has discussed inequality at smaller geographic scales. Several early analyses of inequality at the state level do exist, for instance Bishop et al. (1991). Several other papers have examined inequality at smaller geographies as well. County-level analyses of trends in inequality have been conducted, including Moller et al. (2009) and Peters (2013). No thorough MSA-level analysis of income inequality has been done, at least to my knowledge.

Several papers have used state-level or local-level income inequality data as a way of analyzing how inequality might affect certain outcomes. Frank (2009) constructs a state-level panel of income inequality statistics to address a separate question (a time-series analysis of the inequality-growth relationship). Frank's data are derived from the public-use version of the IRS data used by Piketty and Saez (2003), and extends from 1916-2005. Several papers have used CPS or Census/ACS data for similarly separate ends, including Mellor and Milyo (2002), who use CPS data to calculate MSA-level inequality in service of analyzing the Wilkinson hypothesis of a link between inequality and health status. A recent working paper, Daly and Wilson (2013), utilize county-level inequality data to examine a similar relationship between inequality and health.

At least two recent papers have examined the determinants of inequality at the level of individual MSAs. Florida and Mellander (2013) utilize both inequality measured using the American Community Survey (ACS) well as ineuqality measured using wage data from the Current Population Survey. Glaeser et al. (2009) also use inequality measures calculated from ACS and Census data to examine the connection between home prices and inequality. However, compared to these papers I make several contributions. First, using the CPS, I construct an annual panel of inequality measures at the MSA (and state) levels.[6] Second, this study is unique among state or local area inequality studies in that it directly addresses the top-coded nature of the CPS microdata via a multiple imputation process along the lines of Jenkins et al. (2011).

Data

This paper seeks to add to this small but growing literature by examining income inequality at the sub-national level using CPS data in a way that addresses the potential pitfalls of the data (the right-censoring of incomes due to topcoding). To accomplish this, I adapt the Generalized Beta II multiple imputation approach of Jenkins et al. (2011) to a sub-national setting. Before carefully explaining the multiple imputation process, I present some institutional information about Census Bureau survey data. The focus here is on topcoding in the Current Population Survey, although there is also topcoding in the decennial Census and American Community surveys.

Census Bureau Topcoding

The March Supplement to the Current Population Survey is an annual survey of around 50,000 households which is designed to produce a nationally representative sample of the US population. There has been some debate concerning the representativeness of any given year's CPS sample for individual states or localities. The decennial Census and the ACS are designed to be representative for localities (e.g. the 1-year ACS public-

[6]In Appendix B, I construct a similar dataset using ACS data.

use files are intended to be representative for geographic areas with populations of at least 100,000). I proceed under the assumption that the CPS is reasonably close to representative, at least for sufficiently large MSAs and states.[7]

The CPS March Supplement includes a number of questions about income, which have become more detailed in subsequent iterations of the survey. This disaggregated individual income information is then aggregated up to form individual, family and household income amounts for each relevant responding unit in the survey. From 1968 to 1975, the CPS included eight questions about personal income components. In 1976, this was expanded to eleven items, and after 1988, the CPS included 24 separate questions about income sources.[8] The full list of income sources can be found in Table 1. Each of these income sources is subject to topcoding. In order to satisfy their mandate to protect individual privacy, the Census Bureau censors the raw reported amounts above a pre-defined threshold for each income source. There are two levels of topcoding—a "hard" topcode of the internal data, in which the raw survey results are replaced with a censored amount, and a second topcode for the public-use data, in which the internal data amount is replaced with a separate censored amount before the data are released for public-use. The internal topcodes are usually, but not always, different from the public-use topcodes.

Until 1996, the Census Bureau simply replaced each topcoded income amount with the topcoding threshold. From 1996 to 2010, however, the Census Bureau instituted a more informative topcoding regime. They first divided the topcoded individuals into cells (gender-by-race-by-employment status) and then calculated a mean of all the individuals in each cell above the topcoding threshold (but below the hard-coded censoring point).

[7]Census Bureau documentation suggests that the CPS sample is probably representative for the largest 100 MSAs.

[8]The more recent surveys merely divide broad categories of income (e.g. government transfers) into more disaggregated components. The various household income definitions I will use should be unaffected by the granularity.

TABLE 1. Income Sources in the CPS by Year

	1968-1976
Labor Sources:	Wages, Self Employment, Farm Income
Non-labor sources:	Social Security, Welfare, Government Programs, Interest, Dividends and Rents alimony, contributions, other
	1976-1987
Labor Sources:	Wages, Self Employment, Farm Income
Non-labor sources:	Social Security, Supplemental Security, Welfare, Interest, Dividends & Rental Income Veterans & Workers Comp., Retirement, Other
	1988-2012
Labor Sources:	Wages, Self Employment, Farm Income
Non-labor sources:	Social Security, Supplemental Security, Welfare, Interest, Dividends, Rental Income, Alimony, Child Support, Unemployment, Veterans Benefits, Workers Comp., Retirement, Survivor Benefits, Disability Benefits, Educational Assistance, Financial Assistance, Other

These cell means are then substituted for each respective topcoded income amount. Compared to the previous regime, the cell means provide a better picture of the right tail of the income distribution, although a substantial amount of information is still suppressed. The Census Bureau change in topcoding policies after 1996 can lead to discontinuities in any calculations of distributional statistics. However, Larrimore et al. (2008) provide a series of cell means that is consistent with the official Census Bureau series, and extends back to 1976.

Starting with the 2011 March CPS, the Census Bureau has implemented a new top-coding method, in which topcoded incomes are randomly swapped across individuals within relatively narrow income bins. At the national level, this means that there should be very little difference between the internal and public-use data when these are used simply for the purpose of income inequality analysis. The internal CPS data is topcoded at a higher threshold, so the bias due to topcoding is still present, albeit diminished substantially. Additionally, the Census Bureau has additionally provided researchers with "swap files" for the 1977–2010. I then use these "swap files" to perform the same rank-bin swap process on the March CPS public-use data from 1977 to 2010. I will then use this swap-file modified public-use CPS microdata from 1977 to 2010 and the CPS public-use data from 2011 onwards to calculate measures of income inequality and perform inference on trends.[9]

Generalized Beta II Imputation

Although the rank-proximity swap process reduces some of the problematic topcoding in the public-use CPS, it does not affect the internal "hard" topcodes, which are problematic for the study of income inequality.[10] A large literature has emphasized that the rise in income inequality since 1970 has been driven primarily by gains at the very top of the income distribution. Inequality measured using censored top-income values in the Census Bureau survey data is likely to miss changes at the very top of the income distribution. The use of cell mean or rank-proximity swap replacement partially address this censoring. One increasingly common method to address the remaining bias due to topcoding and under-reporting, which I utilize in this study, is to implement a multiple imputation approach. Jenkins et al. (2011) introduced this method for national level income inequality, and I extend this approach to the state (and MSA) level.

The basic methodology for the multiple imputation process is as follows. First, a parametric distribution is fitted to the observed size-adjusted household income data. Then, partially synthetic income distributions are formed by taking draws from the fitted

[9] A previous version of this paper used the Larrimore et al. (2008) cell means series and the GB2 imputation method described in what follows. The current version of this current version uses the same multiple imputation method, but with the swap-file-modified CPS microdata.

[10] Additionally, as noted by Diaz-Bazan (2015), there is reason to believe that incomes may be under-reported at the top of the income distribution in the March CPS.

distribution and imputing these draws to the topcoded or potentially under-reported individual incomes. Second, distributional statistics are calculated using the partially synthetic data. This process is repeated n times, and the n point estimates are combined according to the rules proposed by Reiter (2003). I use $n = 200$ for most of my analysis, although preliminary results in small-scale simulations seem to imply that $n = 100$ or even $n = 20$ is sufficiently large. In Jenkins et al. (2011), this process was conducted for each year of the CPS to produce national-level inequality statistics. However, since I am interested with sub-national-level inequality, there are slight modifications that must be made to the Jenkins et al. (2011) methodology. In the following, I describe the specifics of this modified process.

I merge the CPS swap-file with the public-use CPS microdata and then calculate household incomes from the resulting dataset.[11] I then flag all topcoded and potentially under-reported individual income source amounts in the combined CPS sample from 1968-2014.[12] As shown in Appendix A, using a cutoff at the 97.5th percentile allows me to reasonably approximate trends in top income shares and the Gini coefficient calculated using IRS tax return data. I adjust for household size by using an equivalence scale equal to the square root of the number of people in the household. Each individual in the household is then associated with this size-adjusted household income amount.

I assume that the distribution of size adjusted household income is well approximated by a Generalized Beta distribution of the second kind (GB2).[13] The four

[11]The CPS counts business losses as income, so there are some households which report negative total household incomes. I am utilizing some inequality measures (notably the Theil index) which are only defined for positive incomes. After aggregating to the household level, I truncate all non-positive household incomes to $0.01.

[12]All geographic identifiers are not available for all years. Some MSAs are identified starting in the 1968 March CPS, but states are not identified until 1977.

[13]See McDonald (1984), Majumder and Chakravarty (1990), Wilfling (1996) and McDonald and Ransom (2008)

parameter GB2 distribution has a probability density function

$$f(y) = \frac{ay^{ap-1}}{b^{ap}\beta(p,q)\left(1+\left(\frac{y}{b}\right)^{a}\right)^{p+q}}$$

where $\beta(\cdot)$ is the Beta function, and a, b, p and q are positive parameters. Several well known distributions used in the income inequality literature (including the Singh-Maddala and Pareto distributions) are special cases of the GB2 distribution. In the model fitting step of the imputation process, I estimate a GB2 distribution to fit the observed data via maximum likelihood.[14]

I am interested in inequality at sub-national levels, so it would be natural to modify the Jenkins et al. (2011) method to fit a distribution to local level incomes, and then proceed with the imputation. However, in practice, there are often too few observations at the local level. Hence I fit a single distribution for each year using observations from the entire US in the swap-file-modified CPS microdata.[15] I then perform the imputation for each MSA or state separately based on the estimated parameters from this fitted GB2 distribution. This method maximizes the number of top incomes employed to fit the GB2 distribution, and hence should yield a more-accurate approximation of the true upper tail of the distribution.

In the second step of the GB2 imputation process, I split the data for each year by MSA or state, and construct partially synthetic datasets for each geographic unit in each year. Non-topcoded households enter each partially synthetic dataset unchanged. For households with topcoded (or potentially under-reported) incomes, I replace the topcoded income with a draw from the fitted GB2 distribution. Specifically, each topcoded income

[14] I use the R package *GB2* package (`https://cran.r-project.org/web/packages/GB2/index.html`) to perform the maximum likelihood estimation.

[15] Following Jenkins et al. (2011), I use only the top 70% of incomes in the CPS sample in each year in the estimation. This is intended to improve the fit at the right tail of the distribution.

y_i is replaced by

$$y_i^* = F^{-1}\left(u_i\left(F\left(y_i\right),1\right)\right)$$

where $F\left(\cdot\right)$ and $F^{-1}\left(\cdot\right)$ are the CDF and inverse CDF associated with the fitted GB2 distribution, and $u_i\left(a,b\right)$ is a draw from a uniform distribution with lower bound a and upper bound b. I construct $n = 200$ such partially synthetic datasets for each MSA (or state) in each year. The final step in the GB2 imputation process is to estimate income inequality measures for each partially synthetic data set. I then follow the rules from Reiter (2003) for combining estimates from partially synthetic datasets. For each income inequality measure of interest, I produce a point estimate q^* equal to the simple average of the estimates q_i over the n partially synthetic datasets: $q^* = \frac{1}{n}\sum_{i=1}^{n} q_i$.[16]

I calculate several income inequality measures using this methodology.[17] The two most important are the Gini coefficient and the top 1% share. The Gini coefficient is calculated as:

$$G = \frac{1}{\mu_x 2n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\left|x_i - x_{i'}\right|$$

where x_i is household i's income. The top 1% income share is the income accruing to the top 1% of the income distribution as a share of total income. If incomes are arrayed in non-decreasing order $i = 1,...,n$, and k is the closest integer to $\frac{99n}{100}$, then the top 1% share is

$$Top1Share = \frac{\sum_{k}^{n} x_i}{\sum_{i}^{n} x_i}$$

[16]The variance of this point estimate is given by $V^* = \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(q_i - q^*\right)^2 + \sum_{i=1}^{n} v_i\right)$ where v_i is the variance (calculated using asymptotic variance formulas) of the point estimate q_i from the ith partially synthetic dataset.

[17]In addition to the measures listed, I calculate the 90-10 ratio, Theil, Atkinson and Schutz indices, various Generalized entropy indices (varying the α parameter from 0.5-2), the coefficient of variation, the 99-median and 95-median ratios, and the 80-20 ratio, as well as Kuznets ratios.

I also calculate a number of Lorenz ordinates for a variety of analyses. The pth ordinate of the Lorenz curve is

$$L(p) = \frac{1}{\mu} \int_0^{F^{-1}(p)} x f(x)\, dx$$

Note that the top 1% share can therefore also be expressed as a function of Lorenz ordinates:

$$Top1Share = 1 - L(0.99)$$

I calculate these measures for all 50 states (and the District of Columbia) Unfortunately not all MSAs have enough observations in the CPS to perform reliable income distribution analysis. Hence, I focus on MSAs which have at least fifty household observations in each year of the CPS. The observations from MSAs with less than fifty households are recoded as non-MSA households, so that they may still be used for the fitting of the GB2 distribution. Additionally, since the evolution of income inequality is of primary importance, households in MSAs which have large gaps in their time series are likewise recoded as non-MSA households. After these modifications, 177 MSAs remain in the dataset. The state-level and MSA-level inequality measures that I estimate here are available in an online data appendix.[18] I also perform similar exercises, generating datasets of income inequality measures for all 50 states and 277 MSAs using microdata on income from the American Community Survey, which are available in the same online data appendix. The specifics of the ACS-based analysis can be found in Appendix B.

From Market Income to Haig-Simons Income

The GB2 imputation method addresses potential concerns about topcoding and under-reporting of income at the top of the distribution. Using inequality measures

[18]Available at http://pages.uoregon.edu/jlv/state-and-msa-inequality.html

generated using this method, it is possible to examine how inequality has changed over the last several decades. Before doing so, however, it is necessary to define carefully the income concept to be used. The most conceptually attractive definition both i) treats income receiving units equally (by adjusting for household size using an equivalence scale), and ii) defines income as the change in net worth that can be used for consumption (where this is usually referred to as "Haig-Simons income"). Measuring Haig-Simons income requires information about changes in wealth due to capital gains, which are unavailable in the Census Bureau surveys used in this study. Nonetheless, it is possible to use all of the information available in the March CPS to construct an income definition that is as close as possible to the Haig-Simons definition.

In the spirit of Armour et al. (2014), I will construct a "crosswalk" from the most commonly used income concept in inequality studies using tax return data (e.g. Piketty and Saez (2003)) to a broader measure of income that I will call "post-fiscal" income. Table 2 summarizes each step along the crosswalk. The crosswalk consists of five income definition, arranged in order of their "closeness" to Haig-Simons income, where each step corresponds to an additional source of income or a change in income-receiving unit. The first step is tax-unit Adjusted Gross Income (AGI), which includes all "market income" received by a tax unit (the individuals included on a tax return). The second step, pre-transfer income, changes the income receiving unit to the household, and adjusts for household size by applying a square root equivalence scale.[19] The third step, post-transfer income, adds cash transfers from the government. The fourth step, post-tax income, is post-transfer income after state and federal income taxes (including tax credits). The final step, post-fiscal income, adds the cash value of in-kind aid to post-tax income. The contrast between post-fiscal and pre-transfer income incorporates not just how market

[19] I divide each household's income by the square root of the number of members in the household and assign this size-adjusted household income to each individual in the household.

income has evolved, but also how state interventions to reduce income inequality have evolved, in the form of changes to the tax and transfer system and, increasingly, the provision of in-kind benefits to households in the bottom half of the income distribution.

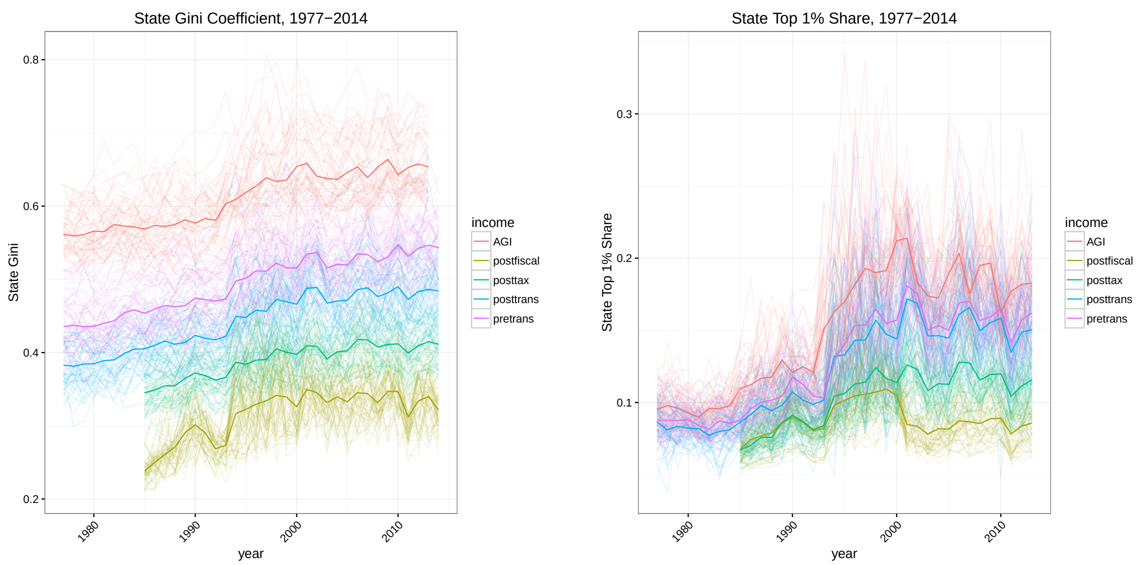
TABLE 2. Crosswalking from Market to Disposable Income

	Definition
AGI Income	Tax unit adjusted gross income
Pre-transfer Income	Pre-tax Household income (wages + investment)
Post-transfer Income	Pre-transfer Income plus cash transfers
Post-tax Income	Post-transfer income less taxes and tax credits
Post-fiscal income	Post-tax income plus the monetary value of noncash benefits

Cash transfers include: TANF/AFDC, Social Security, Disability (SSDI or SSI), unemployment benefits.
Non-cash benefits include SNAP (food stamps), the value of public housing and rental subsidies, the value of home heating subsidies, the value of government provided school lunch, the fungible value of Medicaid and Medicare benefits and the value of WIC benefits

To visualize how recent trends in inequality vary across states and across income definitions, I present several graphs. First, I visualize the trends for all states and all five income definitions simultaneously as a "spiderweb graph" in Figure 1 for the Gini coefficient and the top 1% share. The central tendency across all 50 states for each income definition is emphasizes to permit easy visualization of the average trend in inequality across states. As expected, there are large differences in the level of income inequality when using different income measures. Income inequality is highest for the AGI definition of income, and lowest for post-fiscal income. The levels of measured income inequality differ substantially depending on the income definition. Especially in the period after about 2000, trends in income inequality vary by income definition as well. Income inequality measured by the Gini coefficient using AGI income is on average increasing (albeit at a slower rate after about 2000 than the run-up in inequality before 2000). There is a noticeable decline in the post-fiscal Gini coefficient after 2000. This is also true of the post-fiscal top 1% share, with particularly sharp drop occurring right after 2000. [20]

FIGURE 1. State Income Inequality, 1977–2014: Crosswalking from Market Income to Post-fiscal Income

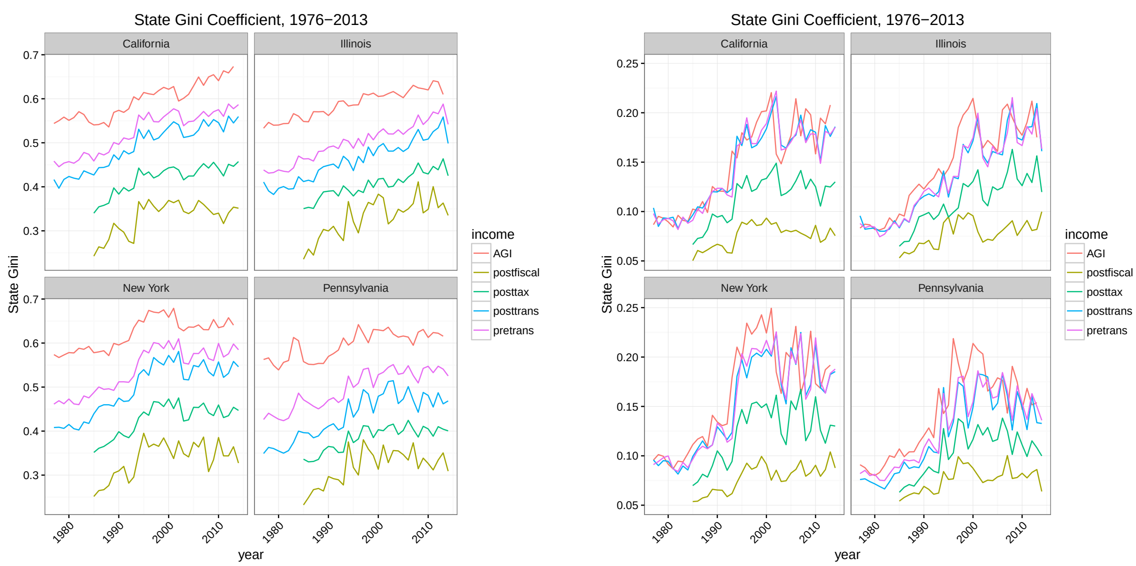


Since the "spiderweb" graphs by design abstract from trends in any individual state, it can be instructive to examine trends in a few states individually. Figure 2 illustrates the Gini and top 1% share income inequality crosswalks from AGI market income to post-fiscal income for the four largest states. For these large states, the pattern seen in the central tendency across all states is even more readily apparent: measured using market income, income inequality has increased almost monotonically since 1977, but measured using a broader, post-fiscal definition, inequality has actually decreased since 2000.

[20]This may be due in part to the rising value of rental subsidies and Medicare/Medicaid given that both health care costs and rents have been rising over this period.

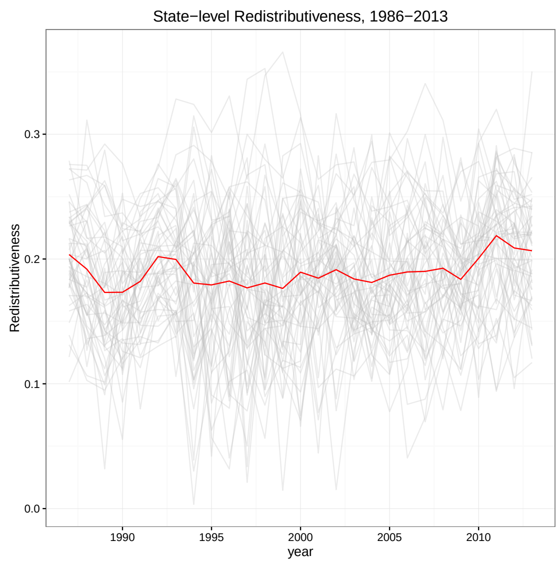
Additionally, it is clear that the primary drivers of the difference between market and post-fiscal income vary depending on the type of inequality measure used. If inequality is measured by the top 1% share, household size is relatively unimportant, but the opposite is true for inequality measured by the Gini.

FIGURE 2. State Income Inequality, 1977–2014: Crosswalking from Market Income to Post-fiscal Income, 4 largest States



The difference in the level of income inequality when using pre-transfer versus post-fiscal income is one way of quantifying the degree to which government intervention is redistributive (Kakwani (1977). Examination of this Kakwani-style measure of redistributiveness across states and over time can shed light on the degree to which different states have adjusted policy in response to rising income inequality. Figure 3 summarizes the state-level trends in redistributiveness measured as the difference between the Gini coefficient using pre-transfer income, and the Gini coefficient using post-fiscal income. It appears to be the case that, on average, redistributiveness has increased since 2000, especially in the period of time corresponding to the Great Recession of 2007-2010.

FIGURE 3. State-level Redistributiveness, 1977-2014



Inference on Changes in Income Inequality

Having generated a dataset measuring point estimates of state-level and MSA-level income inequality as accurately as is possible with public-use data, the logical next step is to analyze how inequality has been changing over time, and what might be driving these changes. Towards that end, I pursue two lines of inquiry. First, I can perform inference on scalar inequality measures and Lorenz ordinates. The scalar metrics and different Lorenz ordinates are sensitive to changes in different parts of the distribution, and hence comparing the results of these tests between these different inequality measures can provide qualitative information on how inequality has been changing. A second line of inquiry uses the bootstrap method of Barrett et al. (2014) to test directly the hypothesis of Lorenz dominance. I can then make direct welfare statements based on this inference.

I will use bootstrap methods to make inferences based on the data. This is a partial departure from much of the survey-based literature, which has relied on inference using asymptotic variance formulas. There is reason to believe, however, that this type of inference may lead to incorrect conclusions. Flaichaire and Davidson (2007) and Brzezinski (2013) note that bootstrap procedures have lower nominal type I error probabilities than inference using asymptotic variance formulas for the Theil coefficient and top income shares respectively, and Mills and Zandvakili (1997) show that the same is true for the Gini. Additionally, in simulation experiments, it appears that the hypothesis tests using asymptotic variance formulas for Lorenz curves have extremely low power for the sort of hypotheses I will be interested in testing, relative to bootstrap-based inference.

Mills and Zandvakili (1997) develop a simple bootstrap method for conducting inference on the Gini coefficient. To extend this method to topcoded and potentially under-reported data, I extend the semi-parametric bootstrap technique of Flaichaire and Davidson (2007). For clarity, I first describe the simple Mills and Zandvakili (1997) method, and then describe my extension of the semi-parametric bootstrap technique of Flaichaire and Davidson (2007) for conducting inference on changes in inequality in the presence of topcoded or potentially under-reported data. The main object of interest is the difference between an inequality metric θ calculated using samples from two populations: $\phi = \theta_2 - \theta_1$.[21] In the simple bootstrap method, one would bootstrap resample (sample with replacement) from the samples of the two populations in the data, calculate the inequality measures θ_1, θ_2 using the bootstrap samples, and calculate the difference ϕ_j^*. After performing this n times and collecting the bootstrap replicates $\Phi = \left\{\phi_j^*\right\}_{j=1}^n$, inference can then be done by constructing confidence intervals $\left(\Phi_{\alpha/2}, \Phi_{1-\alpha/2}\right)$, where

[21] I will be conducting inference where these two populations are a single state in two different years, but this method could also be used to test a hypothesis of no difference in inequality between two different states in a given year.

subscripts refer to percentiles of the distribution of bootrap replicates. Bootstrap p-values can also be calculated, e.g. for the null hypothesis of no change, as

$$p^* = \frac{1}{n}\sum_{j=1}^{n}\mathbb{1}\left(\left(\phi_j^* - \phi\right)^2 \geq \phi^2\right)$$

where ϕ is the full sample ("true") estimate of the difference in inequality.

To adapt this to our environment, I move to semi-parametric bootstrapping.[22] For each time period, I partition the data into topcoded and non-topcoded observations. The non-topcoded observations are bootstrap resampled as in the simple case. The topcoded observations are imputed as in the GB2 multiple imputation process used to generate point estimates. That is, for each bootstrap replication, each topcoded observation y_i is replaced by a draw from the tail of the fitted GB2 distribution. In practice, for a fitted distribution $\hat{F}(y)$, I take a uniform draw $u_i \in \left[\hat{F}(y_i), 1\right)$, and then replace y_i with $\hat{F}^{-1}(u_i)$. I then estimate θ_j from the full non-topcoded sample and the topcoded draws, and θ_j^* from the bootstrap sample of the non-topcoded observations and the topcoded draws. From these I obtain ϕ_j^*, ϕ_j as before. After repeating this process n times, I can then calculate a point estimate of the difference in inequality as $\hat{\phi} = \frac{1}{n}\sum_{j=1}^{n}\phi_j$ as in the multiple imputation case. The bootstrap confidence interval is the same as before, and the two sided p-value for a null of no change is

$$p^* = \frac{1}{n}\sum_{j=1}^{n}\mathbb{1}\left(\left(\phi_j^* - \hat{\phi}\right)^2 \geq \hat{\phi}^2\right)$$

[22] This semi-parametric method is similar to Flaichaire and Davidson (2007), differing in the parametric distribution used (the Generalized Beta II distribution)

or, for a one-sided p-value for the null hypothesis of no change against the alternate that the the change is negative,

$$p^* = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\left(\left(\phi_j^* - \hat{\phi}\right) \leq \hat{\phi}\right)$$

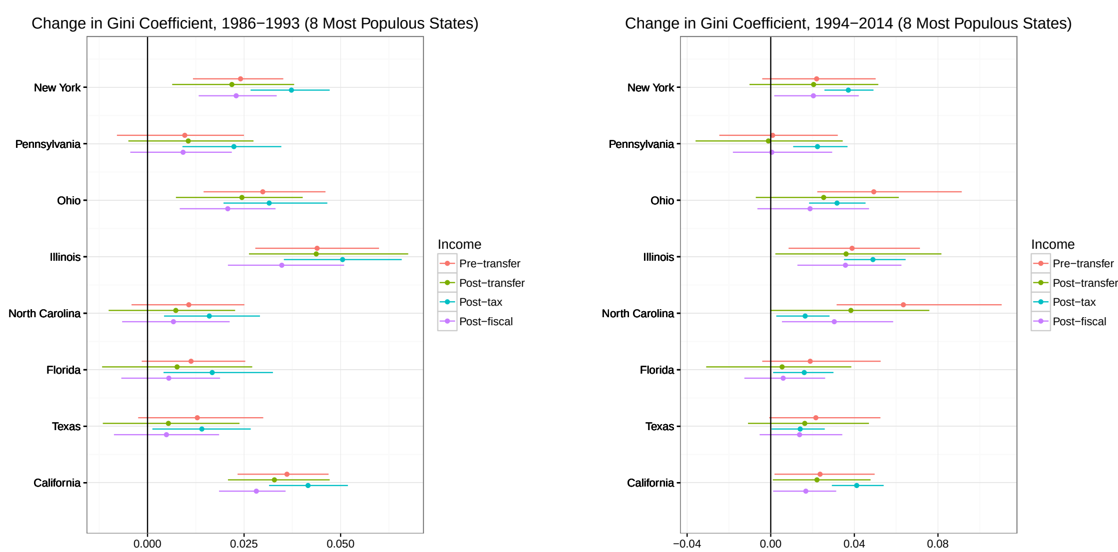
With this semi-parametric bootstrap method, I perform inference on the same scalar inequality metrics for which I have produced point estimates in the previous section. Performing inference in this case involves many more hypothesis tests than are feasible to display individually, given that the focus is on with State-level and MSA-level inequality. I summarize the results by the proportion of states or MSAs for which I can reject the null hypothesis of no change in inequality, and show individual results for only the most populous of states and MSAs.

To analyze trends in income inequality, I will perform tests of the null hypothesis that inequality did not change over a specific period of time. Most but not all previous studies utilizing Census microdata have shown little to no increase in inequality since the mid-1990's at the national level. Studies utilizing tax return data find that inequality has increased dramatically, although in general at a slower rate than the 1980's. Further, these studies tend to find that the increase in inequality is driven by an increasing share of income accruing to the richest 1%. In light of this, I will consider two questions: whether income inequality increased from 1986 through 1993, and whether income inequality has increased from 1994 onward.[23] Due to a change in the way the Current Population Survey was administered, estimates of income inequality from before 1994 are not easily comparable with estimates after 1994, and hence dividing the sample at this discontinuity has an intuitive appeal.

[23]For the pre-1993 test, I choose 1986 as the starting year, since this is earliest year in which the post-fiscal income inequality series can be constructed.

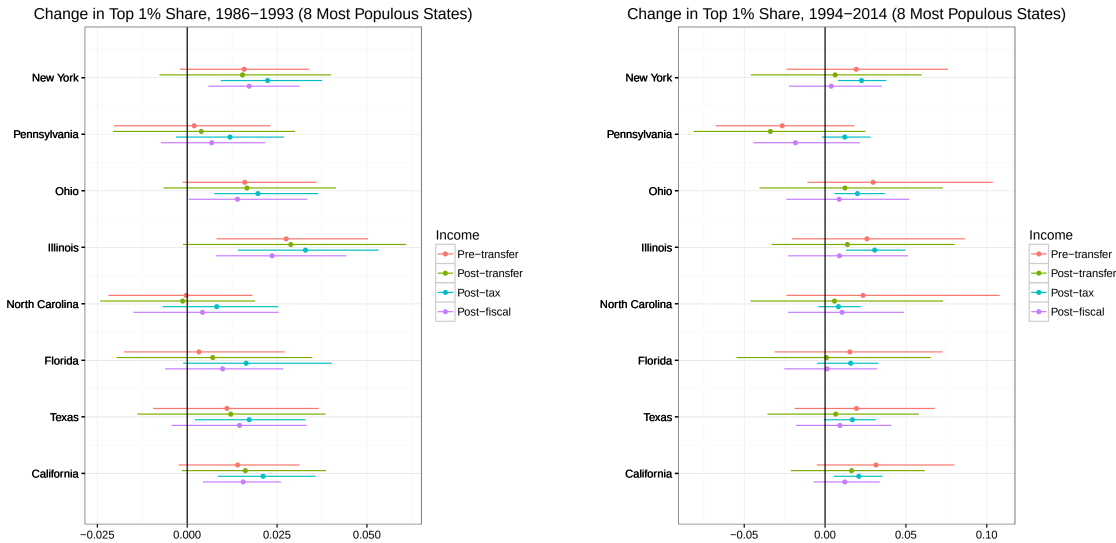
Figures 4 and 5 summarizes inference using CPS microdata for two scalar inequality measures—the Gini coefficient and the top 1%'s share of income, for the periods 1986–1993 and 1994–2014 at the State level. These figures show the 95% confidence interval around the change in the inequality measure in question for the eight most populous states. These largest states all have positive point estimates for the change in inequality by either measure in both time periods. For about half of all states, the change in the Gini coefficient is not statistically different from zero in either time period. On the other hand, the change in the top 1% share is statistically insignificant for almost all states from 1994 to 2014 and for most states from 1986 to 2014. There is some interesting heterogeneity across the income definitions—notably the confidence intervals around the differences in post-tax income inequality are substantially smaller than they are for the other income concepts.

FIGURE 4. Difference in Gini Coefficient, 1986–1993 and 1994–2014



Given the properties of these two income inequality measures, these results suggest that increases in income inequality between 1994 and 2014 might be driven by changes

FIGURE 5. Difference in top 1% Share, 1986–1993 and 1994–2014



within the bottom 99% of the income distribution. Although there is less evidence for a top incomes-driven change in income inequality, it is difficult to disentangle why there is a failure to reject the null hypothesis of no change for the top 1% share. This may be due to (a) inadequate coverage of top incomes in the CPS survey, or (b) because top income shares did not increase. In an effort to further characterize the changes in income distribution, I next consider inference concerning Lorenz curves, which allow fopr the identification of the regions of the income distribution driving these changes in inequality.

Recall that the ordinates of a Lorenz curve

$$L(p) = \frac{1}{\mu} \int_{0}^{y=F^{-1}(p)} x f(x)\, dx$$

can be interpreted as the cumulative share of income accruing to the bottom $(p)\,100\%$ of the income distribution. The change in specific Lorenz ordinates then describes the change in income shares of the bottom p percent of the income distribution. Note that

$$\frac{dL(p)}{dp} = \frac{y}{\mu}$$

where $y = F^{-1}(p)$. Thus for two populations with income distributions subscripted as 1 and 2 for convenience,

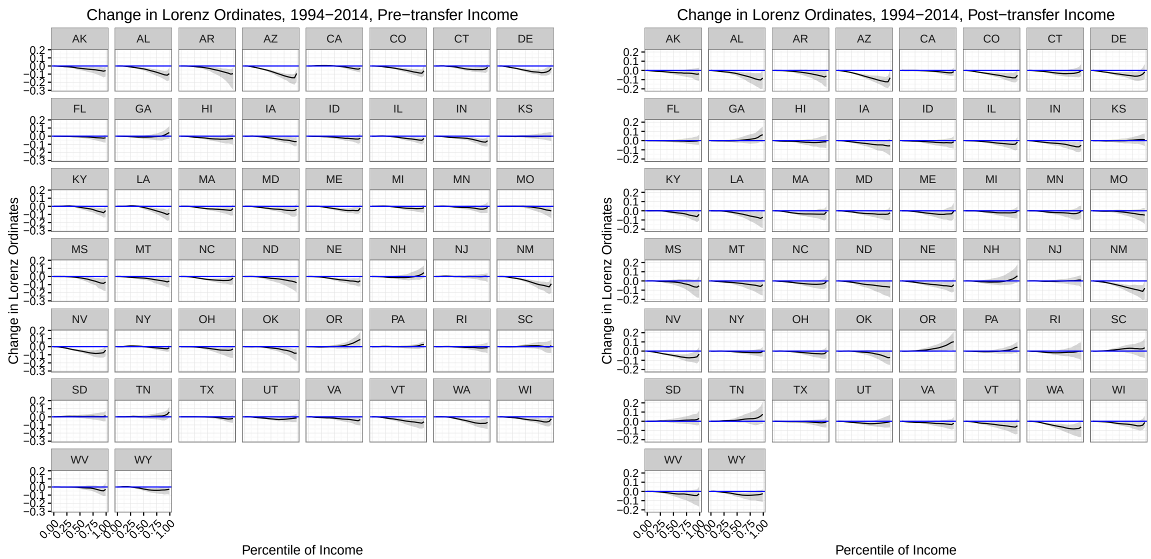
$$\frac{d(L_1(p) - L_2(p))}{dp} = \frac{y_1}{\mu_1} - \frac{y_2}{\mu_2}$$

Which is to say, the slope of the difference in Lorenz ordinates graphed in $(p, L_1(p) - L_2(p))$ space describes the change in the pth percentile's relative income. This can be thought of as as how much the pth percentile is relatively better off. A slope change can then be interpreted as delineating the portion of the income distribution which is relatively better off under distribution 1 and which is relatively better off under distribution 2.

Consider a case in which changes in income inequality are driven solely by increases in the top 1%'s incomes while the bottom 99% stays the same. In this case, $L_1(p) - L_2(p) < 0, \forall p \leq 0.99$ and $\frac{d(L_1(p)-L_2(p))}{dp} < 0, \forall p \leq 0.99$. On the other hand, if $L_1(p) - L_2(p) < 0, \forall p \leq 0.99$ but $\exists p \leq 0.99$ s.t. $\frac{d(L_1(p)-L_2(p))}{dp} > 0$ However if the former holds but the latter does not, then this is evidence for both rising top incomes and changes in inequality within the bottom 99%. If $L_1(p) - L_2(p)$ has only one local minimum, this further suggests that there is income polarization within the bottom 99%, with the top of the bottom 99% seeing relative improvements, while the bottom of the bottom 99% is relatively worse off.

Figures 6 and 7 inventory the evidence concerning changes in Lorenz curves at the state level for the four income concepts of interest over the period 1994-2014. Each figure visualizes the point estimate of the change in Lorenz ordinate, as well as 95%

FIGURE 6. Change in Lorenz Ordinates, 1994–2014, Pre-transfer and Post-transfer Income

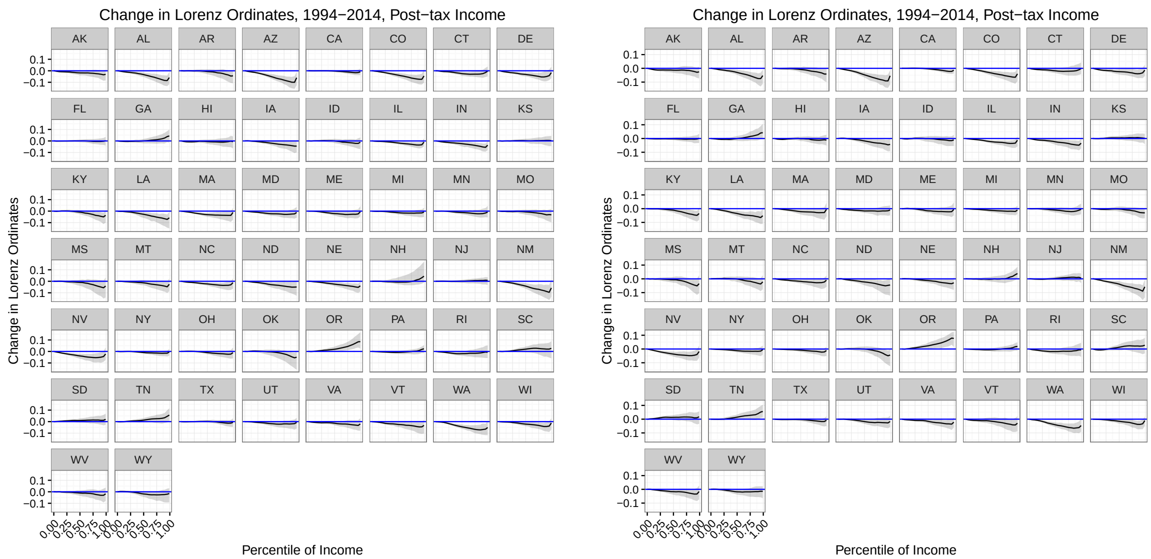


bootstrap confidence intervals (obtained via the semi-parametric bootstrap process described above). For almost all states, the Lorenz curve for the income distribution in 2014 is lower than the Lorenz curve for 1994, although the changes at any specific ordinate are not necessarily statistically different from zero.[24] Further, the change in the Lorenz curve ordinates is negative and decreasing with p until around the 80th percentile for many states, at which point the change in the Lorenz curve ordinates begins to increase (although it remains negative). In general, the Lorenz curves for the different income definitions exhibit largely similar changes over the period 1994-2014. These patterns in the change in Lorenz ordinates over the income distribution are consistent with both an increase in top incomes, and increasing income inequality within the bottom 99%.

I have established that there have been statistically significant increases in income inequality since the 1990's, especially within the bottom 99% of the income distribution. I

[24]Oregon is the notable exception to this general trend.

FIGURE 7. Change in Lorenz Ordinates, 1994–2014, Post-tax and Post-fiscal Income



next examine Lorenz dominance, which gives normative content to the changes in Lorenz curves shown previously. Distribution 1 is said to "Lorenz dominate" distribution 2 if

$$L_1(p) \geq L_2(p)\, \forall p \in (0,1)$$

and

$$\exists p \in (0,1) \text{ s.t. } L_1(p) > L_2(p)$$

I extend the bootstrap test for Lorenz dominance suggested by Barrett et al. (2014) to accommodate topcoding by using the semi-parametric approach used previously. I then perform this Lorenz dominance test for each state using the four household income concepts.

I first review the Barrett et al. (2014) method, and then describe my extension. Consider two distributions (numbered 1 and 2 for convenience) with associated Lorenz curves L_1, L_2. Define $\phi(p) = L_2(p) - L_1(p)$, the functional $I(\phi) = \int_0^1 \phi(p) \mathbb{1}(\phi(p) > 0)\, dp$ and $T_n = \frac{n_1 n_2}{n_1+n_2}$. To test the null hypothesis H_0^1

$$L_2(p) \leq L_1(p) \;\; \forall p \in [0,1] \tag{2.1}$$

i.e. that distribution 2 does not Lorenz dominate distribution 1, I perform m bootstrap replications, using the same semi-parametric scheme as above: I calculate $\hat{\phi}_i^*(p)$ from a sample composed of bootstrap resampled observations for non-topcoded observations, and GB2 draws for the topcoded observations, and $\hat{\phi}_i^M(p)$ from a sample composed of the full non-topcoded sample and draws from the GB2 distribution imputed for the topcoded observations. I then calculate $\hat{\phi}(p) = \sum_{i=1}^m \hat{\phi}_i^M(p)$. Thus, $\hat{\phi}_i^M(p)$ is just the GB2 multiple imputation estimate of $L_2(p) - L_1(p)$. Finally, I construct a one-sided bootstrap p-value from the bootstrap replications

$$\hat{p}_1 = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\left(T_n I\left(\hat{\phi}_i^*(p) - \hat{\phi}(p)\right) > T_n I\left(\hat{\phi}(p)\right)\right)$$

A test of the hypothesis in equation 2.2 can be conducted based on the rule "reject if $\hat{p} < \alpha$". This is equivalent to a test of weak Lorenz dominance. It is straightforward to test the opposite hypothesis H_0^2:

$$L_2(p) \geq L_1(p) \;\; \forall p \in [0,1] \tag{2.2}$$

by reversing the order of the two distributions and constructing the bootstrap p-value $\hat{p}_2$. A test of strong Lorenz dominance can then be conducted by examining both bootstrapped p-values. I conclude that distribution 1 *strongly* Lorenz dominates distribution 2 if both $\hat{p}_1 < \alpha$ and $\hat{p}_2 > \alpha$.

It is most straightforward to conduct a test for the Lorenz dominance of an early distribution over a later distribution (e.g. when comparing 2014 to 1994, the 1994 distribution plays the role of distribution 2, and the distribution in 2014 takes the place of distribution 1). Table 3 summarizes these results for the eight largest states and each of the four income concepts. The table summarizes the two p-values which can be used jointly to test for strong Lorenz dominance—if $p_1 < \alpha$ and $p_2 > \alpha$ the conclusion is that the income distribution in 1994 Lorenz-dominates the distribution in 2014. I conclude that this Lorenz dominance has occurred for many states, although the number of states where this occurs diminishes with the level of redistributiveness encapsulated in the income definition. Assuming $\alpha = 0.05$, Lorenz dominance holds for 29 states using pre-transfer income, and for 31 states using post-transfer income. However, Lorenz dominance holds for only 23 states using post-tax income, and for only 20 states using post-fiscal income. This is the strongest evidence yet that income inequality may not have risen as much if taxes and transfers (including in-kind programs) are taken into account. This result is in line with Armour et al. (2014).

Potential Explanations for Changing State-level Income Inequality

I have demonstrated that there have been significant changes in State and MSA-level income inequality concentrated in the bottom 99% of the income distribution. Although I do not find strong direct evidence of top-incomes-driven inequality changes, I cannot rule these out. At any rate, such changes are not incompatible with the effects I observe in the lower part of the distribution. To complete the examination of state-level income inequality, I examine potential explanations for the rising income inequality observed since

TABLE 3. Lorenz Dominance Results

State	Pre-Transfer p_1	Pre-Transfer p_2	Post-Transfer p_1	Post-Transfer p_2
California	0.006	1	0.002	0.130
Florida	0.136	0.090	0.048	1
Illinois	0.014	0.070	0.008	0.998
New York	0.098	0.076	0.016	1
North Carolina	0.002	0.052	0	1
Ohio	0.002	0.070	0.006	1
Pennsylvania	0.012	0.094	0.016	1
Texas	0.004	1	0.014	0.054
Lorenz Dominance:	29 States		31 States	

State	Post-tax p_1	Post-tax p_2	Post-Fiscal p_1	Post-Fiscal p_2
California	0.004	0.074	0.006	0.090
Florida	0.106	1	0.074	1
Illinois	0.002	0.998	0	0.996
New York	0.012	0.998	0	1
North Carolina	0.002	0.998	0.006	1
Ohio	0.012	1	0.016	1
Pennsylvania	0.002	1	0.002	1
Texas	0.020	0.056	0.036	0.048
Lorenz Dominance:	23 States		20 States	

the 1990's. Previous studies of inequality have suggested a number of factors that might account for the increase in inequality observed in recent decades.

The evidence presented so far suggests that rising income inequality has been driven both by rising top incomes and by changes within the bottom 99%. To examine each of these factors, I will perform a "horse-race" type analysis, using fixed effects panel regressions to compare the relative influence of five common explanations for rising income inequality. These five candidate explanations are 1) unionization rates, 2) minimum wages, 3) top marginal tax rates, 4) human capital attainment and 5) technological advancement. The first two factors can be measured directly.[25] To capture state variation in top tax rates, I include both the overall marginal capital gains tax rate and the overall marginal income tax rate in subsequent regressions. I capture human capital attainment by the fraction of a state's population with at least a bachelor's degree, and technological advancement by the number of patents granted per capita.

I first consider how these five potential factors are related to two scalar measures of inequality (the Gini Coefficient and top 1% share) and then move to the use of ordinates of the the Lorenz curve as dependent variables. As in Bishop et al. (1991), changes in the slope of the size of the effect of a variable on the Lorenz share with respect to the cumulative proportion p of the population can be interpreted as the marginal effect on the income share (non-cumulative) of the pth percentile of the population.[26]

Other things equal, I expect that unionization rates and the real value of minimum wages will affect primarily the bottom 99% of the distribution, while technological advancement and top marginal tax rates will affect primarily top incomes. Educational

[25] I measure the minimum wage as the binding statutory minimum wage in a state, deflated by the CPI-U price index.

[26] This is because the Lorenz curve is a sum of infinitesimal income shares, and therefore the effect on cumulative income shares can be expressed as a sum of the effects on infinitesimal income shares

attainment may affect both parts of the distribution. Note that there is no ex ante reason to expect that each of these factors might affect income inequality identically for each income concept. In particular, note that the broader income definitions (e.g. post-fiscal income) incorporate the effects of policies which may be designed in response to the various factors' effects on market income.

In the current setting, causal identification is difficult, especially given free labor mobility between states. Neither unionization rates nor human capital attainment have been subject to the types of exogenous discrete policy variation necessary for difference-in-difference or regression discontinuity methods, and obviously exogenous instruments are not readily available. To overcome potential simultaneity problems, the best available course of action is to control for unobserved heterogeneity across states and over time to the fullest extent possible via fixed effects, and to compare results across a number of models and specifications. If most of the estimated effects lie in a relatively narrow band this can be taken as suggestive evidence as to the true effect size.

The baseline horse-race model compares the influence of these 5 effects in a model including State and year fixed effects:

$$\begin{aligned} Ineq_{it} = \alpha_i + \alpha_t + \beta_1 Union_{it} + \beta_2 MinWage_{it} + \\ \beta_3 Tax_{it} + \beta_4 Educ_{it} + \beta_5 Patents_{it} + \gamma X_{it} + \epsilon_{it} \end{aligned} \tag{2.3}$$

To modify the above to allow for more flexibility in absorbing time-varying heterogeneity, I can allow for State-specific linear trends and/or quadratic trends, as in:

$$\begin{aligned} Ineq_{it} = \alpha_i + \alpha_t + \theta_{1,i} t + \beta_1 Union_{it} + \beta_2 MinWage_{it} + \\ \beta_3 Tax_{it} + \beta_4 Educ_{it} + \beta_5 Patents_{it} + \gamma X_{it} + \epsilon_{it} \end{aligned} \tag{2.4}$$

$$Ineq_{it} = \alpha_i + \alpha_t + \theta_{1,i}t + \theta_{2,i}t^2 + \beta_1 Union_{it} + \beta_2 MinWage_{it} + \\ \beta_3 Tax_{it} + \beta_4 Educ_{it} + \beta_5 Patents_{it} + \gamma X_{it} + \epsilon_{it} \tag{2.5}$$

In each model, X_{it} is a vector of other potential time-varying confounding factors which might be related to inequality (including population density, government spending per capita, demographic characteristics, changes in household size and composition, industry composition, real state personal income per capita, the state unemployment rate and age composition of the state).

Tables 4 and 5 report results from regressions estimated using the Gini coefficient as the dependent variable, for each of the four difference income concepts. In line with expectations, the unionization rate has a statistically significant and negative impact on the Gini coefficient for three of the four income concepts, suggesting that unionization reduces income inequality. The exception to this trend is post-fiscal income inequality, with which neither unionization nor any of the other determinants of interest has any statistically significant relationship. Top marginal tax rates also appear to have a substantial impact on changes in income inequality, although interestingly top marginal capital gains rates rather than income tax rates appear to drive this.[27] The effects of the other determinants have the expected signs for most income concepts. the level of the minimum wage has reduces inequality while human capital attainment and technological advancement increase inequality, although these estimated effects are not statistically different from zero. The estimated effect of unionization appears to have a larger effect on inequality using pre-transfer or post-transfer income concepts, which is consistent with unionization primarily affecting wage income. Top marginal tax rates affect both pre-tax and post-tax income inequality, which suggests that the effect of tax rates working

[27] This is interesting given that capital gains are not included in any of the income definitions. Changes in capital gains tax rates may indirectly affect dividend income, however, which is included in the income definitions.

through the elasticity of taxable income rather than the mechanical effect of redistributive progressive taxes.

TABLE 4. Determinants of State Income Inequality (Pre-transfer and Post-transfer)

	Dependent variable:					
	Pre-transfer Gini			Post-transfer Gini		
	(1)	(2)	(3)	(4)	(5)	(6)
Union Coverage	-0.231^{***}	-0.323^{***}	-0.327^{***}	-0.183	-0.320^{**}	-0.338^{**}
	(0.085)	(0.108)	(0.124)	(0.121)	(0.142)	(0.163)
Minimum Wage	-0.0001	-0.001	-0.002	0.0004	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Capital Gains Tax Rate	-0.006^{***}	-0.007^{**}	-0.008^{*}	-0.008^{***}	-0.008^{**}	-0.009
	(0.002)	(0.003)	(0.004)	(0.002)	(0.004)	(0.006)
% College Educated	0.077	0.130	0.0001	0.163	0.183	0.062
	(0.144)	(0.169)	(0.178)	(0.188)	(0.234)	(0.247)
Patents per cap.	0.013	0.024	0.015	0.021	0.028	0.013
	(0.011)	(0.015)	(0.023)	(0.014)	(0.020)	(0.028)
Linear Trends?	No	Yes	Yes	No	Yes	Yes
Quad. Trends?	No	No	Yes	No	No	Yes
Observations	1,000	1,000	1,000	1,000	1,000	1,000

Note: *p<0.1; **p<0.05; ***p<0.01
All models include State and Year fixed effects

Other variables included in regressions but omitted from this table: top marginal income tax rate, real personal income per capita, the mean and standard deviation of years of education, % employed in manufacturing, population density, R&D spending per capita, % black, % Latino, median age, divorce rate, unemployment rate, % over 55, % under 25, % non-citizens, % foreign born, government expenditures per capita

Tables 6 and 7 report analogous results using the top 1%'s share as a dependent variable. When examining how the five factors of interest are related to state-level income inequality when measured by the top 1%'s share of income, largely similar patterns of estimated effects obtain in terms of signs, although not necessarily significance. Of the

TABLE 5. Determinants of State Income Inequality (Post-tax and Post-fiscal)

	Dependent variable:					
	Post-tax Gini			Post-fiscal Gini		
	(1)	(2)	(3)	(4)	(5)	(6)
Union Coverage	−0.160**	−0.220**	−0.195*	−0.115	0.009	0.001
	(0.080)	(0.097)	(0.113)	(0.128)	(0.145)	(0.173)
Minimum Wage	0.001	0.00002	−0.001	−0.003	0.001	−0.0003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Capital Gains Tax Rate	−0.006***	−0.006**	−0.007*	−0.005	−0.003	−0.005
	(0.002)	(0.003)	(0.004)	(0.003)	(0.006)	(0.007)
% College Educated	0.084	0.082	−0.014	−0.147	−0.056	−0.142
	(0.130)	(0.157)	(0.168)	(0.198)	(0.271)	(0.300)
Patents per cap.	0.012	0.021	0.016	0.004	0.040*	0.043
	(0.010)	(0.014)	(0.020)	(0.016)	(0.023)	(0.037)
Linear Trends?	No	Yes	Yes	No	Yes	Yes
Quad. Trends?	No	No	Yes	No	No	Yes
Observations	1,000	1,000	1,000	1,000	1,000	1,000

Note: *p<0.1; **p<0.05; ***p<0.01
All models include State and Year fixed effects
For more details see Table 4

five potential factors outlined above, only the effect of tax rates has an individually statistically significant effect on the top 1%'s share (although not for all specifications).

TABLE 6. Determinants of State Top 1% Share (Pre-transfer and Post-transfer)

	Dependent variable:					
	Pre-transfer Top 1% Share			Post-transfer Top 1% Share		
	(1)	(2)	(3)	(4)	(5)	(6)
Union Coverage	−0.106	−0.155	−0.206	−0.109	−0.180	−0.248*
	(0.080)	(0.110)	(0.134)	(0.099)	(0.123)	(0.150)
Minimum Wage	−0.001	−0.001	−0.001	−0.0002	−0.0002	0.0003
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Capital Gains Tax Rate	−0.007***	−0.006	−0.007	−0.009***	−0.006	−0.008
	(0.002)	(0.005)	(0.006)	(0.002)	(0.005)	(0.006)
% College Educated	−0.034	0.085	−0.011	−0.013	0.052	−0.031
	(0.181)	(0.230)	(0.248)	(0.201)	(0.251)	(0.270)
Patents per cap.	0.010	0.025	0.020	0.012	0.026	0.019
	(0.013)	(0.017)	(0.030)	(0.014)	(0.018)	(0.032)
Linear Trends?	No	No	No	No	No	No
Quad. Trends?	No	No	No	No	No	No
Observations	1,000	1,000	1,000	1,000	1,000	1,000

Note: *p<0.1; **p<0.05; ***p<0.01
All models include State and Year fixed effects
For more details see Table 4

As noted, although the Generalized Beta II multiple imputation method used to estimate income inequality attempts to address an important weakness in the current population survey, the underlying microdata ultimately represent only a sample of top incomes. As a robustness check for the analysis, I estimate regressions using state-level estimates of the top 1% share from Frank (2009). Table 8 summarizes these results. Notably, the estimated effects of all five major factors are qualitatively similar to the

TABLE 7. Determinants of State Top 1% Share (Post-tax and Post-fiscal)

	Dependent variable:					
	Post-tax Top 1% Share			Post-fiscal Top 1% Share		
	(1)	(2)	(3)	(4)	(5)	(6)
Union Coverage	-0.067	-0.096	-0.119	-0.144^{**}	-0.141^{**}	-0.153^{*}
	(0.061)	(0.082)	(0.099)	(0.068)	(0.071)	(0.083)
Minimum Wage	0.0003	0.0003	-0.0003	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Capital Gains Tax Rate	-0.005^{***}	-0.004	-0.006	-0.002	-0.001	-0.002
	(0.002)	(0.004)	(0.004)	(0.001)	(0.001)	(0.002)
% College Educated	-0.041	0.035	-0.042	0.025	0.015	-0.012
	(0.140)	(0.175)	(0.191)	(0.125)	(0.158)	(0.167)
Patents per cap.	0.007	0.017	0.012	-0.001	0.007	-0.009
	(0.010)	(0.013)	(0.022)	(0.008)	(0.010)	(0.015)
Linear Trends?	No	Yes	Yes	No	Yes	Yes
Quad. Trends?	No	No	Yes	No	No	Yes
Observations	1,000	1,000	1,000	1,000	1,000	1,000

Note: *p<0.1; **p<0.05; ***p<0.01
All models include State and Year fixed effects
For more details see Table 4

effects reported in Table 6. In addition, perhaps because the Frank (2009) data contains the whole population of top incomes, estimated effects are actually more precise here. In particular, note that at least for some specifications, rates of unionization have a statistically significant effect on top income shares, as does, interestingly, the real minimum wage. Top marginal tax rates exhibit a similarly sized effects as in Table 6, although the estimates are less precise than those for the effects of unionization and minimum wages.

TABLE 8. Determinants of State Top 1 % Share, Frank (2009) Data

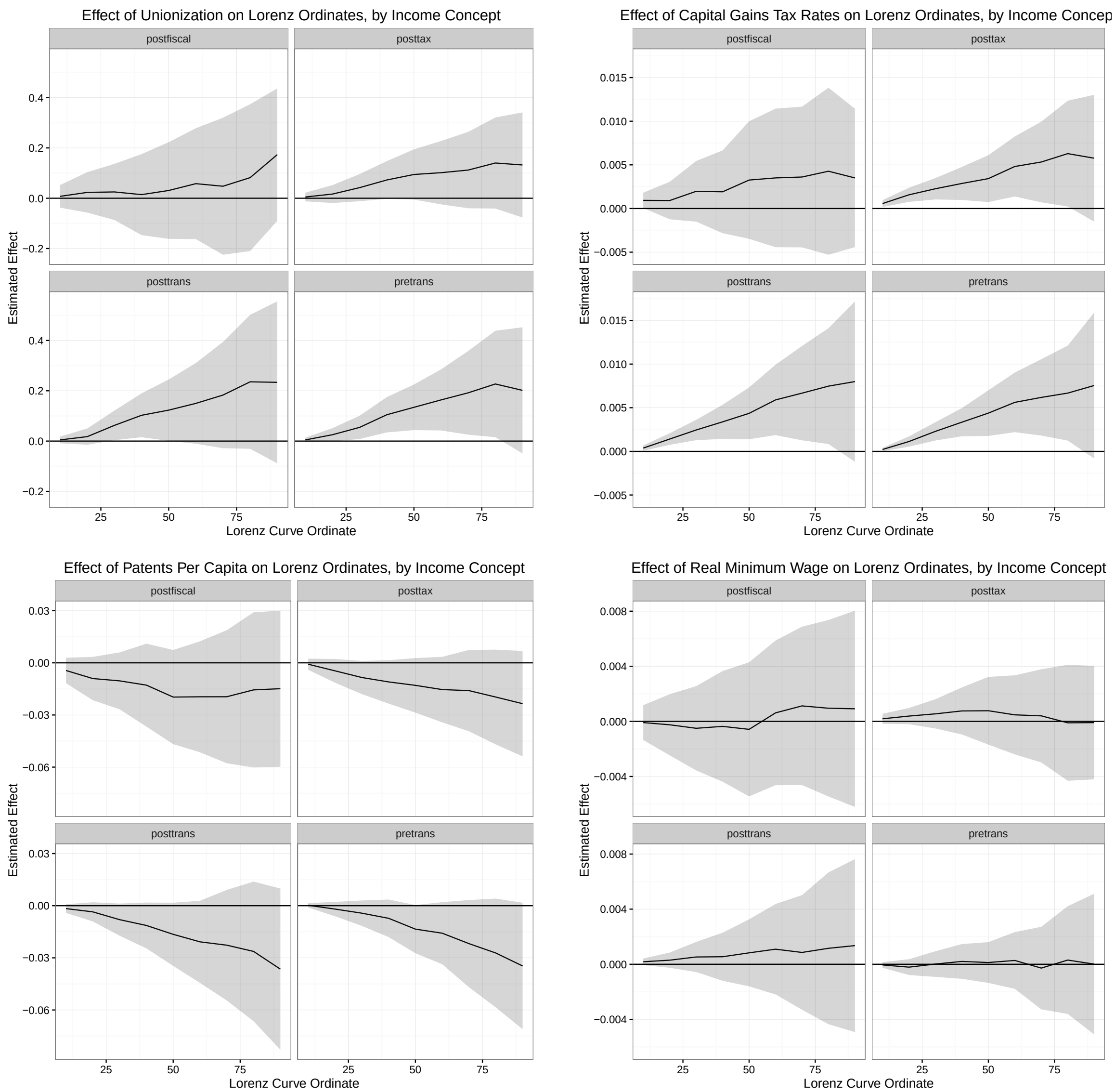
	Dependent variable:		
	(1)	(2)	(3)
Union Coverage	−0.075	−0.172***	−0.041
	(0.064)	(0.057)	(0.071)
Minimum Wage	−0.005	−0.009***	−0.012***
	(0.003)	(0.003)	(0.004)
Capital Gains Tax Rate	−0.002*	−0.001	−0.001
	(0.001)	(0.001)	(0.001)
% College Educated	0.085	−0.046	−0.031
	(0.133)	(0.102)	(0.113)
Patents per cap.	−0.005	0.005	0.020
	(0.014)	(0.014)	(0.014)
Linear Trends?	No	Yes	Yes
Quad. Trends?	No	No	Yes
Observations	1,000	1,000	1,000
Note:	*p<0.1; **p<0.05; ***p<0.01		
	All models include State and Year fixed effects		
	For more details see Table 4		

To further explore how these five factors might affect income inequality at different points along the income distribution, I estimate regressions using the Lorenz ordinates as dependent variables. Doing so allows me to examine of which factors might be closely related to within-bottom-99% inequality, as opposed to those which are primarily driving top incomes. Positive coefficients suggest a reduction in income inequality (a positive coefficient implies an increase in the cumulative income going to incomes below a given percentile) and negative coefficients imply an increase in income inequality. As in the previous Lorenz curve inference, examining how estimated effects change across the distribution has implications about the relevance of the factor in question for top income inequality versus bottom 99% income inequality. Suppose that $E(p;X)$ is the "effect curve" for variable X, describing the estimated effect of X on the Lorenz curve at percentile p. Then if $E(p;X) < 0, \forall p$ or $E(p;X) > 0 \forall p$ then variable X has an effect primarily on top incomes. On the other hand, if there is a slope change in the effect curve, this implies that variable X has an effect primarily on within-99% inequality.

Figure 8 shows the effect of unionization, capital gains tax rates and patents per capita on Lorenz ordinates for the four income concepts. For each concept save post-fiscal income, the effect of unionization has its greatest effect around the 80th percentile of the income distribution, suggesting an effect on income inequality within the bottom 99%. Top marginal tax rates, have increasing effects across the distribution (at least for pre-tax inequality), suggesting an effect primarily on top income shares. Similarly, technological advancement, proxied by patents granted per capita, has an increasing and negative effect on Lorenz ordinates. This suggests that technology increases top incomes, although these effects are not statistically significant. It is clear, however, that this is a borderline case: 95% confidence intervals contain only small areas greater than zero.

Of these five potential factors of interest, it can be argued that unionization is the most important because it has the most robust statistical impact on income inequality. However, this may be primarily affecting income inequality within the bottom 99% of the income distribution. Top marginal tax rates are also important, and are probably

FIGURE 8. Determinants of Lorenz Ordinates, Selected Covariates



responsible for changes in top incomes. Interestingly, variations in human capital acquisition across states appear to have little to no effect on state-level changes in income inequality.

Conclusion

This study has generated a new dataset consisting of inequality measures at the State and MSA level using four different income concepts. The analysis of income inequality measured using these different income concepts in this paper sheds new light on the trends in income inequality within sub-national entities. This analysis also deepens our understanding of which factors may be the driving forces behind rising income inequality. Using semi-parametric multiple imputation and bootstrap methods, I show that income inequality has increased substantially over the last two decades, and note that this change has coincided with both rising top incomes as well as increasing inequality within the bottom 99% of the distribution.

One important way forward in the study of state-level and MSA-level income inequality is to attempt to bridge the gap between state-level income inequality datasets such as Frank (2009) which use tax return data, and the survey based approach taken here (using CPS microdata). Given matched survey and administrative data, it may be possible to retain the benefits of both the survey based approach (more information about household structure and non-taxable income sources) as well as the tax return data (the full universe of top incomes), while potentially incorporating better administrative data on program participation for in-kind transfers.[28] Fully utilizing all available information

[28] Leveraging administrative records on transfer payments and participation in in-kind will require linking these administrative records to Census survey records. This will require access and support of official statistical agencies.

on incomes is important for deepening our understanding of the dynamics of income inequality.

In addition to the analysis presented in this paper on potential determinants of state-level inequality, there are a number of potential applications for this data. Several papers have already made use of this data, including Voorheis (2016), which examines the effect of income inequality on political polarization, Voorheis et al. (2015), which examines the effect of income inequality on carbon emissions, as well as Chapter IV of this dissertation, which uses the MSA-level income inequality dataset to examine the connection between income inequality and environmental justice.

Income inequality is an important and much-discussed topic of obvious relevance to both researchers and policymakers. This study has created a novel, and hopefully quite useful dataset on state-level and MSA-level inequality that corrects for the censoring in the underlying Census Bureau microdata. This dataset allows for inferences concerning the trend in inequality since the 1990's, about which there is some controversy in the literature. I find that for most states and MSAs, there has been a statistically significant increase in inequality (either in terms of Lorenz dominance or pairwise comparison of inequality statistics.) This increase in inequality is driven by both top incomes (i.e. an increase in the top 1% share) and by increasing disparities in income within the bottom 99%. This inequality increase seems to be driven by a decrease in workers' bargaining power (via union density). In contrast, human capital attainment and technological advancement do not appear to explain the observed changes in income inequality within the bottom 99%. Insofar as greater inequality is generally considered a poor outcome, one policy implication is that laws strengthening the institutional position of unions may be an effective tool for decreasing income inequality as a complement to the redistributive effects of the tax and transfer system.

CHAPTER III

TRENDS IN ENVIRONMENTAL INEQUALITY IN THE UNITED STATES: EVIDENCE FROM SATELLITE DATA

Introduction

Concerns about air quality and its negative health and ecological impacts are widespread, both in developed countries and, increasingly, in developing countries. A wide literature, spawned from the environmental justice movement, has documented differences across sub-groups in average exposure to pollutants and toxic chemicals. However, little is known what these differences imply about the overall inequality in the size distribution of pollution exposure. This paper adds to the stock of knowledge about environmental inequality by proposing a dashboard approach[1], combining recently developed theoretical tools for measuring environmental inequality with conventional environmental justice measures. I also examine potential explanations for the variation in this environmental inequality over time.

To examine trends in, and determinants of, environmental inequality, I utilize two datasets of satellite-derived observations of ground-level pollution exposure. These datasets provide fine-grained information about exposure to two important pollutants: nitrogen oxides (NO_x) and particulate matter smaller than 2.5 microns (PM2.5). These two pollutants are relevant to human health. In addition to the direct health impact of exposure to NO_x and PM2.5, they are both highly correlated with other pollutants (e.g. ozone) and as such can be taken as an index of overall air quality. These satellite-derived datasets have been studied widely in other disciplines (e.g. in the atmospheric science

[1] I borrow the "dashboard" terminology from the multi-dimensional poverty literature, e.g. Alkire and Foster (2011)

literature), however they have not yet found their way into the economics literature. These data provide several advantages over conventional tropospheric air quality data derived from ground-level air quality monitors or generated from air quality models using data on emissions from point sources.

Using these two sources of data, I describe trends in environmental inequality for the entire United States, as well as for states and major metropolitan areas within the United States. I consider two different ways of considering the degree to which the distribution of pollution exposure is "unequal." The first follows from Sheriff and Maguire (2014) and defines "environmental inequality" in terms of commonly used inequality measures adapted from the income distribution literature. The second follows from the environmental justice literature and quantifies "environmental justice" as the difference in exposure between advantaged and disadvantaged subgroups. These two concepts can be viewed as capturing "vertical equity" and "horizontal equity."[2]

I examine how traditional environmental justice concerns relate to environmental inequality (rather than simply average environmental quality) by examining how census-tract-level demographic characteristics correlate with different measures of environmental inequality. To accomplish this, I adopt a re-centered influence function (RIF) regression approach. RIF regression is a general estimation strategy for examining how individual characteristics affect a summary statistic of a distribution. RIF regressions have been used mostly in the study of income or wage distributions (e.g. Dube (2013)) and I extend the use of RIF regressions to the study of pollution exposure distributions.[3]

[2] To capture horizontal equity it would be best to compare exposure between groups conditional on income levels. Doing so is difficult with the group demographic data available at the Census tract level, however.

[3] In Appendix C, I further use an extension of Oaxaca-Blinder style decompositions to examine to what degree observable demographics explain differences in exposure between advantaged and disadvantaged census tracts across the exposure distribution, and to what degree demographic characteristics appear to have contributed to changes in exposure across the distribution of exposure over time.

I obtain several key results. First, I confirm that the average level of pollution exposure has decreased since 1998, a result consistent with other studies of trends in environmental quality. Second, I examine how the measures of environmental inequality and environmental justice comprising my “dashboard” have changed over time. I find that most measures in the dashboard are declining over time, with heterogeneity across the two pollutants of interest. In general, decreases in exposure inequality for NO_x have occurred through decreased exposure at the top of the distribution, while decreases in exposure inequality for PM2.5 have come via changes at the bottom of the distribution.

By considering more than just average pollution exposure in examining the correlation between demographic factors and pollution exposure (using RIF regressions), I can expand upon the usual types of empirical results in the environmental justice literature. I find that the African American proportion of the population of a census tract is positively related to the level of pollution in very polluted tracts (increasing inequality), but the relationship is negative at the median (reducing inequality), and insignificant for less polluted tracts. However, the African American proportion of a census tract increases national environmental inequality.

The rest of the paper proceeds as follows. First I briefly discuss the related literature on environmental justice and inequality. I then describe the data to be used in the analysis, and the process for assigning pollution exposure levels to each census tract. I define the various measures of environmental inequality in the “dashboard”, and investigate trends in exposure inequality, contrasting these to trends just average exposure. Finally, I present results from re-centered influence function regressions, and conclude with directions for future work.

Previous Literature

There is a large literature, scattered across disciplines, that has attempted to quantify the degree to which environmental harms might be felt disproportionately by disadvantaged communities. This literature has arisen at least in part as a response to activists from the Environmental Justice movement. In fact, this literature takes its name and often its terminology from these same political advocates, although in practice most environmental justice scholarship is concerned with documenting environmental disparities rather than making normative claims about alternative policies to address these disparities.

One early study by an advocacy group that stimulated the subsequent environmental justice literature is the so-called UCC study (Chavis and Lee (1987)), which noted a correlation between the locations of toxic waste sites and local ethnic minority populations. Subsequent analysis (e.g. Bryant and Mohai (1992)) of toxic waste sites has confirmed that the African-American population share in a neighborhood is sometimes an important correlate with the probability that a toxic waste site will be located in a neighborhood. Other dimensions of disadvantage are also important correlates with toxic waste siting. Despite this consensus, there remains some disagreement as to whether this correlation implies racist siting policies by firms or local authorities. Several studies (e.g. Been and Gupta (1997) and Wolverton (2009)) have persuasively shown that this correlation may merely capture the subsequent hedonic general equilibrium effects—toxic waste sites depress local property values, which leads to an inflow of poor, disadvantaged individuals and an outflow of richer, advantaged individuals. This phenomenon is often termed "coming to the nuisance." Among others, Mohai et al. (2009) and Brulle and Pellow (2006) ably summarize this vast literature.

The simple proximity of toxic waste sites to disadvantaged communities was the primary early concern for the environmental justice literature, but this proximity itself merely implies exposure without actually measuring it. Measuring the actual disproportionate exposure to airborne and waterborne toxics and pollutants (and the negative health impacts of these exposures) has also been a large concern. However, the measurement of ambient pollution exposure is much more complicated than measuring proximity to the point locations of fixed toxic waste sites. Studies of disparate pollution exposure have utilized air pollution models derived from Toxic Release Inventory (TRI) or National Air Toxic Assessment (NATA) data (e.g. Morello-Frosch and Jesdale (2006), Zwickl and Moser (2015)) or data from the US Environmental Protection Agency's (EPA) network of air quality monitoring stations. At least one paper, by Clark et al. (2014) has used satellite data on pollution exposure to document spatially disparate exposure levels. This, and almost all other papers in this literature have considered only single cross sections of data, and thus have not been able to consider how these disparate impacts might be changing over time.

The literature has achieved consensus that in any given cross section at a point in time, disadvantaged communities are exposed to disproportionately high levels of environmental hazards. However, there is not yet a consensus on how best to summarize or quantify the extent of these disparities for use either in policy analysis or for the comparison of trends in the distribution of environmental hazards over time. This confusion about measurement continues, despite the fact that, at least since Executive order 12898, issued in 1994, US government agencies have been required to take environmental justice concerns into account when enacting or changing regulations.

Several approaches to measuring environmental inequality (and environmental injustice) have been advanced in the literature. The simplest possible measure of environmental justice/injustice is of course just the difference in average exposure between advantaged and disadvantaged populations—Clark et al. (2014) use this approach. To address the possibility that there may be differential exposure across the distribution, Boyce et al. (2016) have proposed comparing quantiles of the race-specific exposure distributions. Another approach sporadically used by Zwickl et al. (2014), Zwickl and Moser (2015) and Boyce and Voirnovytskyy (2010) among others has been to import measurement tools from the income distribution literature to describe environmental inequality. This approach has been extended and formalized by Maguire and Sheriff (2011) and Sheriff and Maguire (2014), who adapt several commonly used income inequality measures to produce normatively sensible conclusions. The important difference between income and environmental hazards is that environmental hazards are “bads”. The normative conclusions of unmodified income inequality measures calculated using pollution exposure data may produce potentially unethical conclusions about policy—someone with high income is highly advantaged, while someone with high pollution exposure is highly *disadvantaged*.

In this paper, I advance the literature on environmental inequality and environmental justice along several dimensions. First, I extend the use of satellite data first used by Clark et al. (2014) to study environmental justice beyond a single cross-section. This allows me to describe not just disparities in exposure across groups at a point in time, but also how these disparities have been changing. Second, I propose a dashboard approach to environmental justice analysis by cataloging and categorizing different measures which quantify disparities in pollution exposure. I apply these measures both to document how environmental inequality has been changing over time, and also to reveal potential correlates of these changes in environmental inequality.

Data and Institutional Details

Any examination of the distribution of pollution exposure across space and across sociodemographic groups requires measurements of ground-level pollutant concentrations at a relatively fine spatial scale. Ideally, I would like to observe the actual exposure of each individual as they go about their day. Such a dataset is infeasible without a universal monitoring regime, a prospect that seems unlikely outside of a George Orwell novel. As a next best alternative, I will estimate pollution concentrations at the level of US census tracts.[4] I can then estimate measures of inequality and environmental justice measures using these concentrations, weighted by tract population. These measures of inequality can be interpreted as the inequality across individuals in exposure.

There are two limitations that come with using tract-level average pollution concentrations to estimate pollution exposure inequality or environmental justice measures. First, weighting by tract population essentially assigns the tract-average exposure to each individual residing in the census tract. This is equivalent to assuming no tract-level inequality in pollution exposure. Second, using ground-level concentrations as a measure of exposure ignores potential adaptation behavior on the part of individuals. Adaptation behavior may further be related to sociodemographic characteristics such as income, especially if adaptation requires purchasing expensive equipment. If likelihood to engage in adaptation is positively related to, e.g. income, then tract-level average concentrations will over-estimate the exposure of rich individuals, and hence underestimate the gap in exposure between the rich and the poor. Thus each of these limitations suggest that using tract-level average concentrations to calculate exposure inequality and environmental justice measures will produce a lower bound estimate.

[4] Due to the limitation in spatial coverage of the satellite data, I am only able to calculate tract-level concentrations for the contiguous United States.

This paper capitalizes on the use of two novel sources of data on pollution exposure derived from remote sensing satellite observations. The first dataset, described in detail by Lamsal et al. (2008), infers ground-level NO_x concentrations using a chemical air transport model and observations of the tropospheric vertical column densities of NO_x. The second dataset, described in detail by van Donkelaar et al. (2016), infers ground-level PM2.5 concentrations from aerosol optical depth observations from multiple satellite sources.[5]

Each of these datasets is available at relatively fine geographic resolution—the NO_x data are on a (0.1×0.1)-degree grid, while the PM2.5 data are on a (0.01×0.01)-degree grid (this corresponds to 10 km and 1 km at the equator, respectively). Both datasets provide observations for most of the globe, spanning approximately 70 degrees S through 70 degrees N. These datasets provide substantially more-comprehensive spatial coverage than measurements of ground level exposure from unevenly distributed fixed monitors. By way of comparison, note that there are about 500 NO_x monitors in the EPA's monitoring network, while there are 100,000 uniformly distributed fixed geographic grid points within the contiguous United States in the NO_x satellite data.

To infer person-level exposure to NO_x and PM2.5 respectively, it is necessary to link information on the detailed spatial variation in remotely sensed ground-level pollutant concentrations with information about where people are located. In the absence of any more-fine-grained information about the spatial distribution of households, I assume that each person in a census tract is exposed to, the average ground-level pollutant concentration at the population-weighted centroid of her Census tract. For each satellite data source in each year, I interpolate over the fixed grid to the centroid of each census tract using inverse distance weighting. I use all gridpoints within a 10km radius from the

[5] These datasets are provided for public use by the Atmopsheric Composition Analysis Group at Dalhousie University, and can be accessed at `http://fizz.phys.dal.ca/~atmos/martin/`

centroid of a census tract in calculating the IDW estimate of tract-level exposure.[6] I using this IDW interpolation, I obtain annual average concentrations for PM2.5 for each year from 1998-2014, and for NO_x from 2005-2011.

The spatial distribution of person-level exposure to PM2.5 and NO_x can be visualized using a choropleth map, as in Figure 9, which visualizes the data on exposure to PM2.5 and NO_x for 2005 for the contiguous United States. At this point in time, it is clear that pollution exposure is concentrated in urban areas, although elevated ambient exposure levels are also present in large parts of the rural eastern US. These higher exposures may be in part due to this region's relatively heavy reliance on coal-fired power plants for electricity generation. Coal-fired power plants are a major source of emissions of chemical precursors to NO_x and PM2.5.

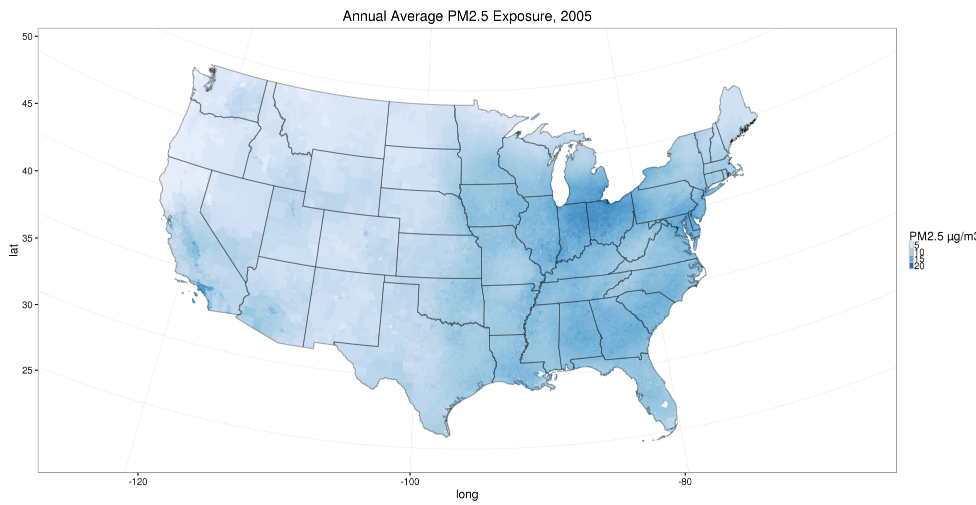
On average, pollution exposure has been declining over the sample periods for the two satellite-derived pollution exposure datasets. Figure 10 shows the population-weighted annual average exposure to PM2.5 and NO_x exposure for the contiguous US. The PM2.5 data span the period 1998–2014, while the NO_x span 2005–2011. NO_x and PM2.5 exposure have both also decreased markedly over these periods. Both datasets suggest that decreases in exposure to pollution were most pronounced in the period before 2008. After 2008, average exposure to both NO_x and PM2.5 continued to decline, but at a slower rate than the previous period.

Quantifying Environmental Inequality and Environmental Justice

Using two independent satellite-derived datasets, it is clear that average pollution exposure has decreased markedly since the 1990s. There are a number of ways, however, in which this reduction in average pollution exposure could have occurred. The uneven

[6]Because of the differing resolution of the gridded data, this translates to using the 4 nearest gridpoints for the NO_x data, and the 100 nearest gridpoints for the PM2.5 data.

FIGURE 9. Annual Average PM2.5 and NO_x Exposure, 2005



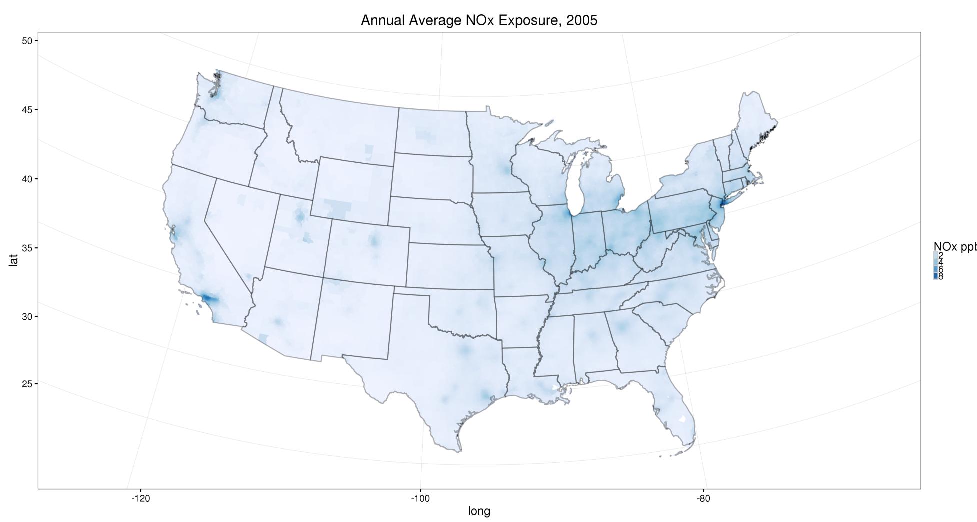
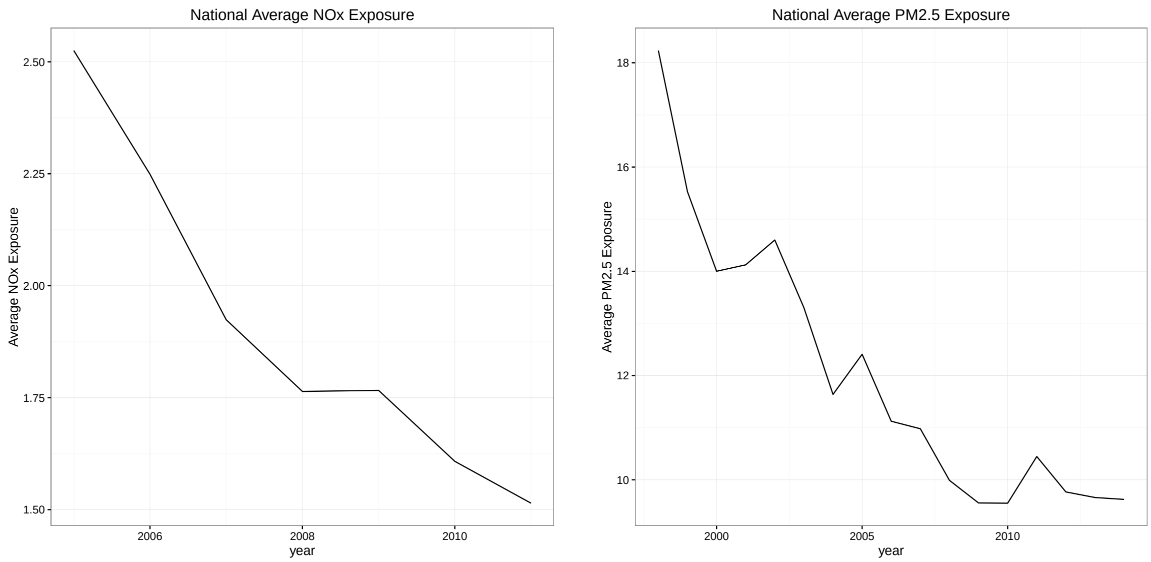


FIGURE 10. National Average NO_x Exposure (in ppb), 2005-2011



distribution of these pollution exposure reductions is of particular interest. Any effort to explore how the changes in the distribution of pollution exposure that have accompanied the average improvements in environmental quality requires some methods for quantifying pollution exposure inequality, and normative tools to rank distributions according to some social evaluation function.

Rather than seeking to identify a single measure which summarizes the distribution of pollution exposure across and within different groups, and its relation to other dimensions of disadvantage, I propose a "dashboard" consisting of several different measures which together can be used for distributional policy evaluation. Considering the whole dashboard provides a transparent view of how the distribution of pollution exposure is evolving that is resistant to cherry-picking individual measures to provide ex-post justification for preferred policy. This dashboard considers two ways of thinking about the distribution of pollution exposure. The first, which I term "environmental inequality" can be viewed as a vertical equity concept. Environmental inequality measures summarize the size distribution of pollution exposure while preserving anonymity of the exposed — the identities of the exposed individuals do not matter, only their exposure levels. The second type, which I term “environmental justice”, can be viewed as a horizontal equity concept, and summarizes the distribution of pollution exposure across demographic categories.

The measurement of environmental inequality can adapt some of the methods used in the literature on the measurement of income inequality. However, it is not immediately obvious that one can simply calculate income inequality measures using pollutant concentration and obtain ethically sensible measures of environmental inequality. Pollution exposure, unlike income, is a ”bad”, which means that “advantage” in pollution exposure runs in the opposite direction from “advantage” in income. Thus the bottom 10% of the pollution exposure distribution is actually the most *advantaged* segment of society along an environmental dimension, whereas the bottom 10% of the income distribution is the most *disadvantaged* segment of society along the income dimension.

One apparently straightforward way to transform the distribution of a “bad” into a good, suggested by Sheriff and Maguire (2014) is to “reverse the sign”—use the negative values of pollutant concentrations when calculating measures of environmental inequality. This has the appeal of imposing the intuitively “correct” (or ethically sensible) ordering. Individuals experiencing the largest absolute value of pollutant concentration are now at the bottom of the distribution, and those who experience the lowest absolute value of pollution exposure are at the top. The environmental inequality components of the “dashboard” consist of three variants of the Lorenz curve, and two scalar environmental inequality measures (the Atkinson index and the Kolm-Pollak index).

First, I consider the relative Lorenz curve

$$L\left(p\right)=\frac{1}{\mu}\int_{-\infty}^{F^{-1}(p)}xf\left(x\right)dx$$

Each ordinate $L(p)$ can be estimated with discrete data by the Kovacevic and Blinder (1997) estimator:

$$\hat{L}(p) = \frac{1}{\hat{N}\hat{\mu}} \sum w_i y_i I\left(y_i \leq F^{-1}(p)\right)$$

where the w_i are weights, $\hat{N} = \sum w_i$, and $\hat{\mu}$ is the weighted mean of the outcome. Pollution concentrations are expressed as negative quantities, so the Lorenz curve will lie above the line of perfect equality, contrary to the "usual" case of the income-based Lorenz curve. For any two distributions 1 and 2, Lorenz dominance of distribution 2 over distribution 1 requires

$$L_2(p) \leq L_1(p) \;\; \forall p \in [0,1]$$

For any two equal-mean distributions, relative Lorenz dominance has normative content: a relative Lorenz dominating distribution would be preferred by every concave social welfare function, as shown by Maguire and Sheriff (2011). Even for distributions with unequal means, the relative Lorenz curve can still provides useful information about the fairness of the distributions, although it no longer generates a unanimous ordering of distributions by all concave social welfare functions.

Second, I consider the Generalized Lorenz curve, which is just the relative Lorenz curve scaled by the mean of the outcome distribution (in this case, mean pollution exposure):

$$GL(p) = \mu L(p) = \int_{-\infty}^{F^{-1}(p)} x f(x)\, dx$$

Generalized Lorenz ordinates can be estimated by a modification of the Kovacevic and Binder (1997) estimator:

$$\hat{GL}(p) = \hat{\mu}\hat{L}(p) = \frac{1}{\hat{N}} \sum w_i y_i I\left(y_i \leq F^{-1}(p)\right)$$

The Generalized Lorenz dominance condition is the same for pollution the income generalized Lorenz dominance criteria from Shorrocks (1983):

$$GL_2\left(p\right) \geq GL_1\left(p\right) \forall p \in [0,1]$$

Generalized Lorenz dominance takes into account the outcome distribution as well as the average level of the outcome. The normative content of generlized Lorenz dominance is more general than the relative Lorenz case—any concave social welfare function will prefer the Generalized Lorenz dominant distribution, regardless of the means of either of the distributions.[7]

I also consider the absolute Lorenz curve proposed by Moyes (1987), which captures the absolute cumulative gap in exposure between the overall average exposure μ, and the average exposure of the bottom pth percent of the population. The absolute Lorenz curve can be expressed as

$$AL\left(p\right) = \int_{-\infty}^{F^{-1}(p)} \left(x-\mu\right) f\left(x\right) dx$$

However, note that the absolute Lorenz curve can be expressed in terms of either the relative Lorenz or generalized Lorenz curves:

$$AL\left(p\right) = \mu\left(L\left(p\right) - p\right)$$

or

$$AL\left(p\right) = GL\left(p\right) - \mu p$$

[7]The transformation of the pollution exposure data ensures that this result, proved by Shorrocks (1983) for income, will hold for pollution exposure

Hence it is possible to modify the Kovacevic and Binder (1997) estimator for the absolute Lorenz case:

$$\hat{AL}(p) = \frac{1}{\hat{N}} \sum w_i y_i I\left(y_i \leq F^{-1}(p)\right) - \hat{\mu} p$$

where $\hat{\mu}$ is the mean exposure, weighted by tract population. Absolute Lorenz dominance requires

$$AL_2(p) \geq AL(p), \forall p \in [0,1]$$

Absolute and relative Lorenz curves are measuring the the two “unequal inequalities” concepts of Kolm (1976). The absolute Lorenz curve captures an absolute environmental inequality concept summarized by its translation invariance property: if every individual’s pollution exposure increases by a constant amount, the Absolute Lorenz ordinates are unchanged. In contrast, the relative Lorenz curve captures a relative inequality concept captured by its scale invariance property: if every individual’s pollution exposure increased by a constant proportion, the relative Lorenz ordinates would be unchanged.

None of these variants of the Lorenz curve guarantee a complete ordering of distributions. If any pair of these Lorenz curves cross, then no normative conclusion is possible. In income distribution studies, scalar indices (e.g. the Gini coefficient) are often used to induce a complete ordering of distributions. However, it is not clear that many of the most common inequality measures used in the income distribution literature are directly applicable to a case where the distribution concerns a bad (or a transformation of a bad, in this case). Sheriff and Maguire (2014) show that a transformation of the Atkinson index induces a complete and sensible orderings of pollution exposure distributions. The transformed Atkinson index is

$$I_A = \left(\frac{1}{N} \sum \left(\frac{x_i}{\bar{x}}\right)^{1+\alpha}\right)^{\frac{1}{1+\alpha}} - 1$$

where α is an inequality aversion parameter. I modify this formula so that I can weight by tract populations w_i

$$\hat{I}_A = \left(\frac{1}{\hat{N}} \sum w_i \left(\frac{x_i}{\bar{x}} \right)^{1+\alpha} \right)^{\frac{1}{1+\alpha}} - 1$$

where $\hat{N} = \sum w_i$, and $\bar{x}$ is the weighted mean of the outcome distribution. The Atkinson index has a normative interpretation. It is derived from a specific welfare function, so it directly ranks income distributions (higher Atkinson indices are less preferred). Additionally, the transformed Atkinson index has a cardinal interpretation. The Atkinson index of environmental inequality can be interpreted as the percent increase in the average pollution exposure necessary to maintain a constant level of welfare, if that higher average pollution exposure level were to be distributed equally.

The Atkinson index, like the relative Lorenz curve, embeds a relative inequality concept that corresponds to scale invariance: it aggregates ratios rather than gaps. Sheriff and Maguire (2014) show that the Kolm-Pollak index is a suitable absolute inequality measure. The Kolm-Pollak index, like the Absolute Lorenz curve, embeds an absolute inequality concept corresponding to translation invariance. The Kolm-Pollak Index is defined as

$$K_I(x) = -\frac{1}{\kappa} \ln \frac{1}{N} \sum_{i=1}^{N} e^{-\kappa(x_i - \mu_x)}, \kappa < 0 \tag{3.1}$$

where κ can be interpreted as an environmental inequality aversion parameter for the associated social welfare function.

In addition to these five vertical equity (environmental inequality) measures, I include two measures of horizontal equity (environmental justice) in the "dashboard." These measures be capture by differences in exposure between demographic subgroups. I follow the environmental justice literature and focus on differences in exposure by race and ethnicity, specifically on the difference in exposure between non-Latino whites and

African-Americans. The first of these horizontal equity measures is a simple comparison in averages, calculating the difference in average exposure between subgroups. The second compares the distributions across subgroups by calculating the differences in exposure at percentiles of the race-specific exposure distribution (i.e. comparing the 10th percentile of the African-American exposure distribution to the 10th percentile of the white exposure distribution.)

Trends in Environmental Inequality and Environmental Justice

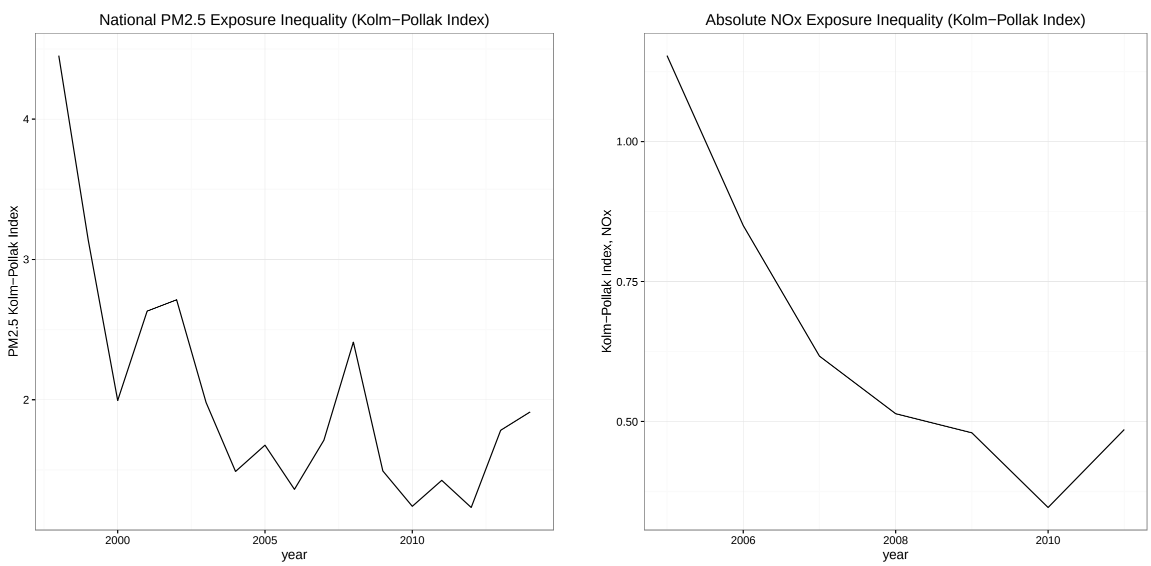
Having defined this dashboard of measures to capture environmental inequality and environmental justice, it is possible to examine how environmental inequality has changed over time, utilizing the new satellite data on PM2.5 and NO_x exposure that have not been widely studied by environmental economists. This use of ambient air pollution as an environmental disamenity in the study of environmental justice is in contrast to some earlier literature focusing on proximity to toxic sites. Concerns about discriminatory siting are less important when studying the distribution of pollution exposure.[8]

As with the calculation of average exposure levels, it is necessary to match ground-level pollutant concentrations with census tract population data to calculate the various environmental inequality and environmental justice measures. To fill in the gaps between the 2000 Census and the American Community Survey, I linearly interpolate between decennial censuses to estimate each census tracts' population and racial demographics from 2001-2004, and use weighted averages of overlapping ACS 5-year file estimates for the period 2005-2014. I will calculate each of the components of the dashboard for the contiguous US using the tract-level pollution exposure and population data

[8] Note however that there are steep gradients in exposure near freeways, as shown by Currie and Walker (2011), which combined with the routing of freeways through minority or poor neighborhoods could be seen as an analogue to discriminatory siting of toxic facilities.

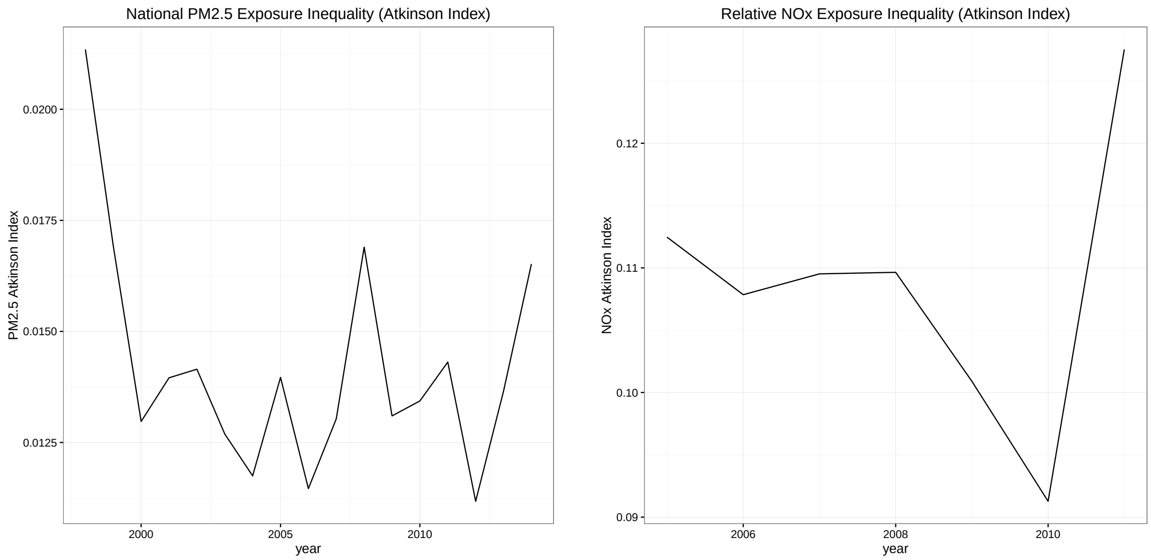
Trends in the scalar environmental inequality measures are summarized in Figure 11 for the Kolm-Pollak index and 12 for the Atkinson Index. Absolute environmental inequality has generally seen a downward trend for both types of pollutants over the relevant sample, although decreases in the Kolm-Pollak index appear to have slowed in the period after 2007 or so. Relative environmental inequality has a less clear overall trend, on the other hand. NO_x exposure is consistently more unequally distributed than PM2.5 according to the Atkinson index, but both pollutants exhibit seem to behave in a roughly stationary way over time.

FIGURE 11. Kolm-Pollak Index, PM2.5 and NO_x



Examining the three variants of the Lorenz curve can shed some light on which parts of the pollution exposure distribution may be driving the scalar inequality trends seen for the Kolm-Pollak and Atkinson indexes. Select relative Lorenz curves are shown for the period 2005-2011 for NO_x and for PM2.5 from 1998-2014 in Figure 13. Relative Lorenz curves provide unambiguous inference about environmental inequality, since the relative Lorenz curves for, e.g. 2005 and 2011 cross at least once. Note that, unlike the relative

FIGURE 12. Atkinson Index, PM2.5 and NO_x



Lorenz curves for NO_x, the largest changes in the PM2.5 relative Lorenz curve seem to be concentrated in the more advantaged (less exposed) part of the exposure distribution.

One the other hand, the point estimates of the the absolute and generalized Lorenz curves are more unambiguous. Figure 14 visualizes the absolute Lorenz curves calculated for NO_x from 2005-2011, and for PM2.5 from 1998-2014. Absolute Lorenz curves for NO_x exposure are higher in 2011 than 2005 throughout the exposure distribution, and absolute Lorenz curves for PM2.5 are similarly higher in 2014 versus 1998. Similar results hold for generalized Lorenz curves, summarized in Figure 15 These results are suggestive of absolute and generalized Lorenz dominance for both NO_x and PM2.5.

The trends in the horizontal equity (environmental justice) components of the dashboard show similar evidence of increasing equality over time. Figure 16 shows the trends in the gap between the average exposure of African-Americans and whites for PM2.5 and NO_x. For both pollutants, this gap has shrunk considerably over the duration of the sample, falling by a factor of almost two for PM2.5. Note that since the

FIGURE 13. Relative Lorenz Curves, NO_x (2005-2011) and PM2.5 (1998-2014)

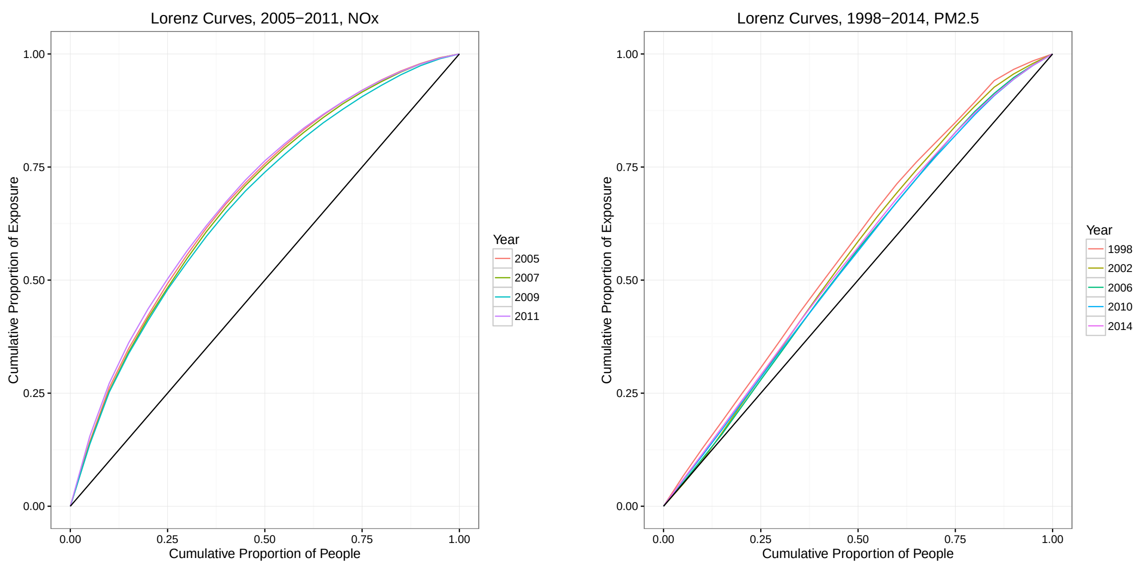


FIGURE 14. Absolute Lorenz Curves, NO_x (2005-2011) and PM2.5 (1998-2014)

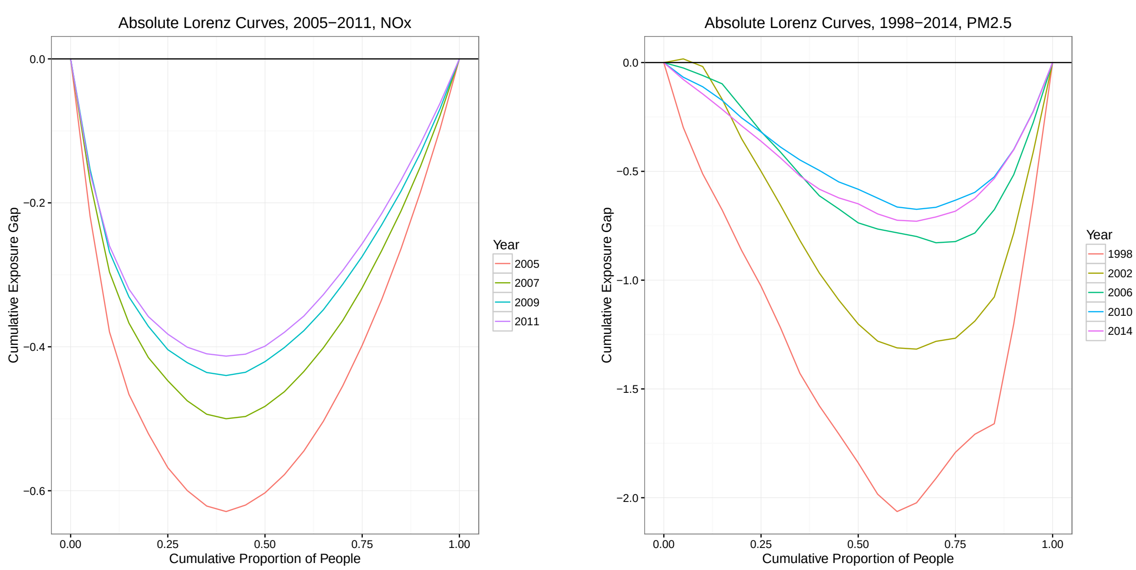
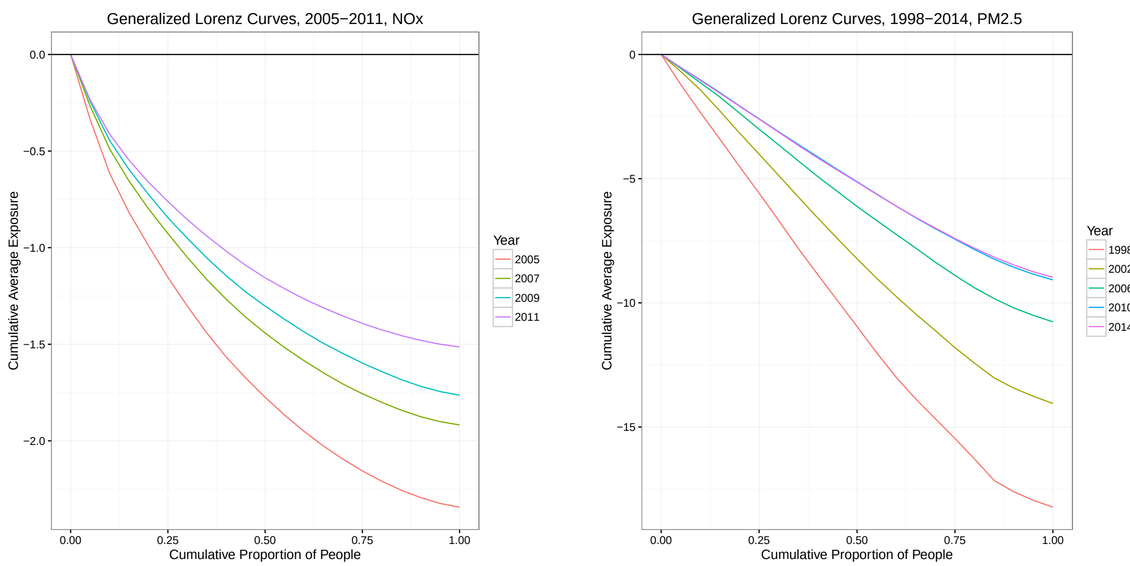


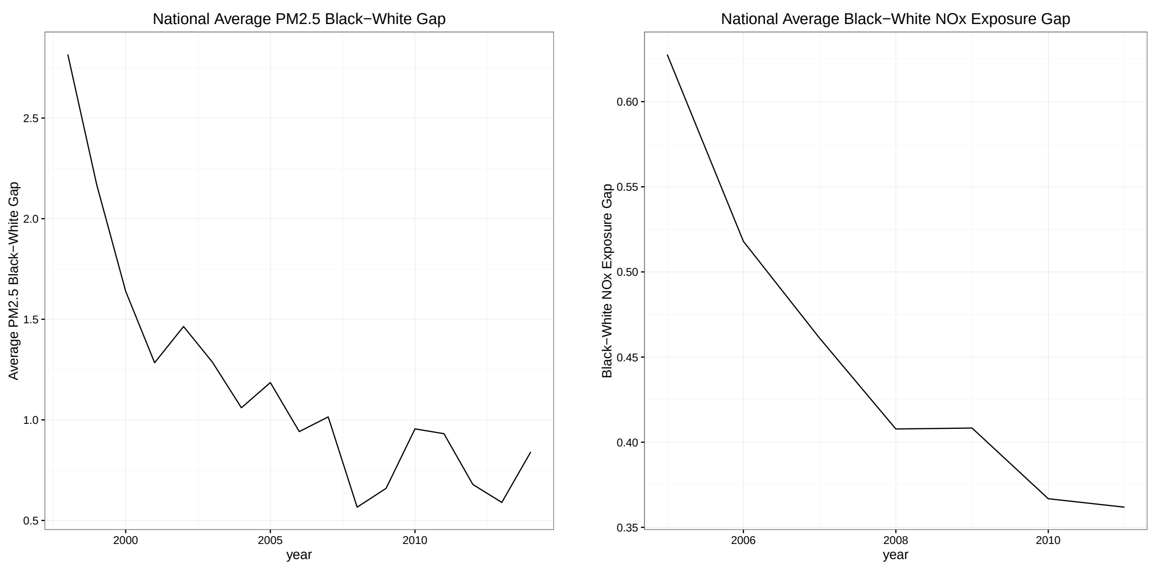
FIGURE 15. Generalized Lorenz Curves, NO_x, 2005-2011



two pollutants are measured in different units, the two black-white gaps are not directly comparable. Black-white ratios are unit-free and hence directly comparable, however, and are shown in Figure 17. Note also that the time periods of the two datasets differ, with no NO_x observations before 2005. In the overlapping period of time (2005-2011), the black-white gap for NO_x has a clear downward trend while the trend for PM2.5 is less clear.

There are, of course, a number of ways in which the gap in exposure between two subgroups may have decreased, as is the case for both NO_x and PM2.5 exposure from the beginning to the end of the relevant time periods. In the case of NO_x, the reduction in the gap in exposure between blacks and whites has occurred because average black exposure was declining faster than average white exposure. Average black exposure to NO_x declined from 2.89 ppb in 2005 to 1.71 in 2011 (a decline of 1.17 ppb) , while average white exposure declined from 2.27 ppb in 2005 to 1.35 in 2011 (a decline of only 0.92 ppb). A similar trend is evident in the decline in the black-white gap in PM2.5 exposure. Average black exposure declined from 20.73 $\mu g/m^3$ in 1998 to 9.75 in 2014 (a decline of 10.97 $\mu g/m^3$, while average white exposure to PM2.5 declined from 18.08 $\mu g/m^3$ in 1998

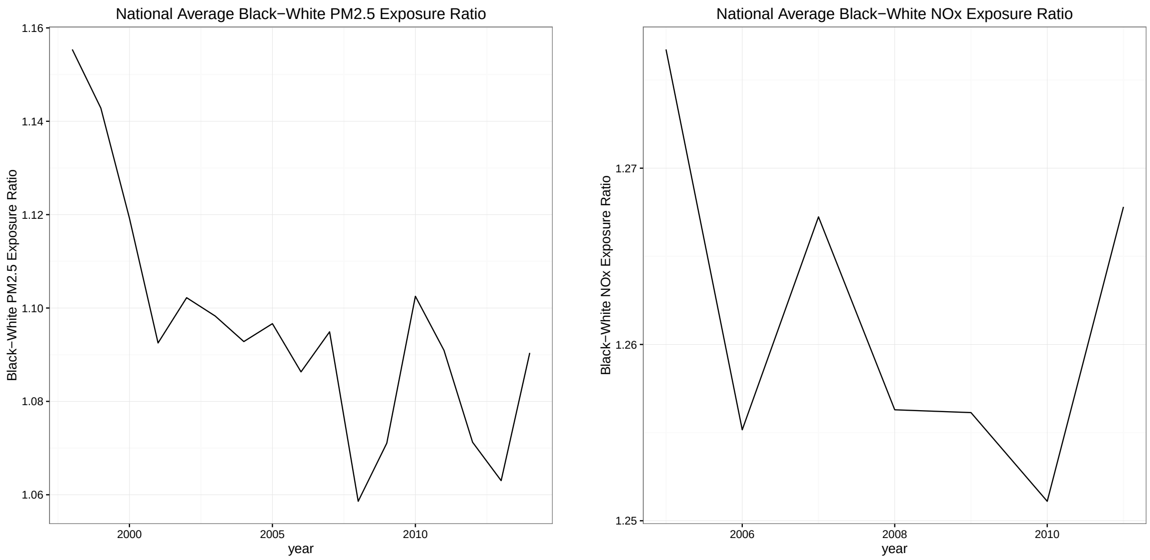
FIGURE 16. National Black-White Exposure Gap, PM2.5, 1998-2014 and NO_x, 2005-2011



to 8.83 μg/m^3 in 2012 (a decline of 9.25 μg/m^3). Thus the decline in the black white gap for both pollutants represents not just absolute but also relative improvements for the disadvantaged group.

One way to examine how this trend of reduction in the average difference in exposure across racial groups has played out across the exposure distribution is to examine how the gap in exposure between blacks and whites has evolved at specific percentiles of the exposure distribution. Figure 18 shows the black-white percentile gap curves for PM2.5 from 1998-2014 and NO_x from 2005-2011 respectively. For PM2.5, a notable feature of the percentile gap curves in each year is that the gap in exposure is actually *larger* at the less exposed part of the exposure distribution. This may reflect in part the differing rural/urban population distributions across the two groups. The percentile gap curve has flattened over the period 1998-2014, largely due to the decline in the gap at low levels of exposure. Environmental justice with respect to NO_x appears to have evolved in roughly the opposite manner, however. The black-white exposure gap is much larger at the highly

FIGURE 17. National Black-White Exposure Ratio, PM2.5, 1998-2014 and NO_x, 2005-2011

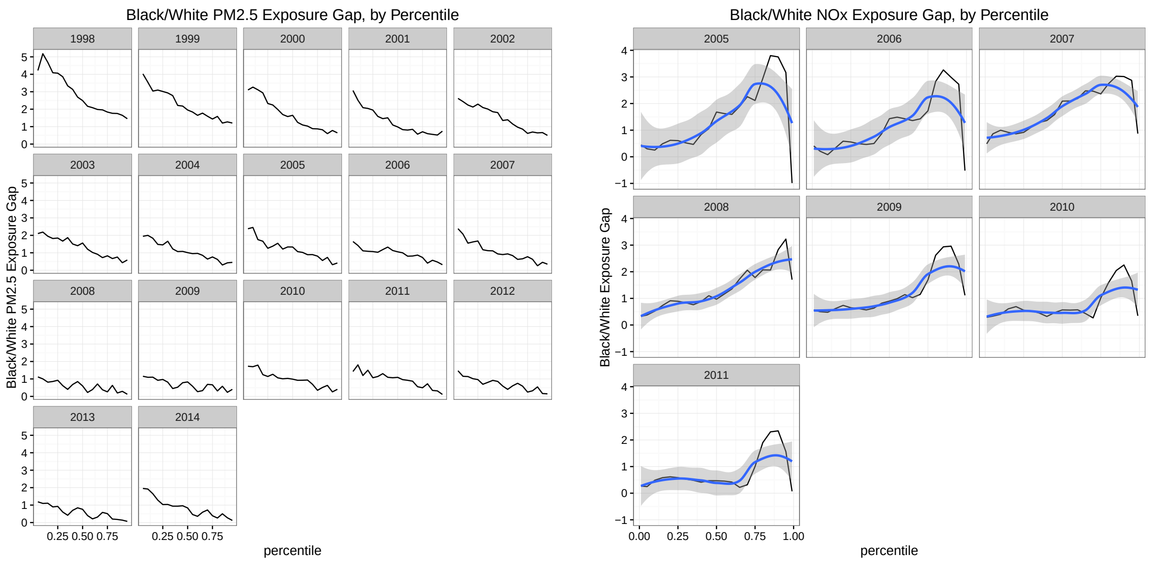


exposed end of the distribution, and the flattening of the percentile gap curve have largely occurred through upper percentile gap reductions.

The evidence on whether the distribution of pollution exposure has become more unequal is not consistent across the measures in the dashboard. One way of synthesizing the evidence presented thus far is to examine the common features of the measures which show relatively unambiguous evidence of decreasing inequality in contrast to those measures for which such evidence is ambiguous. The Kolm-Pollak index, the absolute Lorenz curve and the gap in average exposure between blacks and whites all point towards declining exposure inequality, while the Atkinson index, relative Lorenz curve and black-white exposure ratio all show somewhat ambiguous trends over time.[9] Each of the former group of measures aggregates, in one way or another, differences in absolute exposure, while the latter group aggregates relative exposure. Thus, the trends in inequality in these

[9]The generalized Lorenz curve also shows declining inequality, although this is driven by declines in overall average exposure.

FIGURE 18. National Black-White Exposure Gap, by Percentile (PM2.5 and NO_x)



two groups of measures in the dashboard are consistent with more or less proportional declines in pollution exposure across the pollution exposure distribution. This, in turn, implies greater absolute improvements (declines in exposure) for disadvantaged populations.

In general, the above patterns point toward the conclusion that average exposure to NO_x and PM2.5 has decreased, and so has the inequality in this exposure. These results are less clear for the relative measure of environmental inequality than for measures capturing absolute environmental inequality. The two pollutants have very different cross-sectional distributions, and the two pollutants' exposure distributions have evolved in different ways. Nonetheless the most robust finding is that not only has exposure to these two pollutants been decreasing over time, it is also likely becoming more equally distributed.

Explaining the Distribution of Pollution Exposure

The trends in the distribution of pollution exposure at the national level point to both a decrease in average pollution exposure and a decrease in pollution exposure inequality. There are a number of potential reasons for this trend. In the context of the environmental justice literature, it is perhaps most interesting to consider how these changes may be related to the underlying demographics of the relevant geographic units (here, census tracts). I will examine the relationship between tract-level demographics and the distribution of pollution exposure by utilizing re-centered influence function (RIF) regressions to describe the variation in a functional of the *national* pollution exposure distribution as a function of tract-level demographics.[10]

Re-centered Influence Function Regressions

The RIF regression method is a way to estimate the effect that variation in individual characteristics might affect a functional of the entire distribution. This method has been used most extensively in the income distribution literature, where the distributional functional is often the quantile function. In the quantile case, the RIF regression can be interpreted as an unconditional quantile regression Firpo et al. (2009). Essentially any functional of the distribution in question can be used in an RIF regression. Essama-Nssah and Lambert (2012) catalog the RIFs for most commonly used distributional functionals, including the Lorenz and Generalized Lorenz curve ordinates.[11] I use RIF regressions to attempt to explain how census tract demographics may be

[10] Additionally, RIF regressions can be used to perform a decomposition analysis of the change in exposure over time, and of the difference in exposure between diverse and non-diverse census tracts, as in Appendix C.

[11] Essama-Nssah and Lambert (2012) also provide an expression for the standard Atkinson index, but not for the transformed Atkinson index from Sheriff and Maguire (2014).

related to the observed trends in environmental inequality noted above. Recentered influence function regressions have become an increasingly popular way of estimating distributional effects, both in the conventional wage or income distribution setting (e.g. Zhu (2016), Essama-Nssah and Lambert (2016), Dube (2013)) as well as in the study of health inequality (e.g. Heckley et al. (2016), Gaskin et al. (2015)) but have not been used in the study of environmental inequality.

To review, the influence function is an analytic device commonly used in robust statistics. For a distribution of outcome variable y and a distributional functional ν, the influence function $IF(y, \nu)$ describes how each individual's observed y affects $\nu(F(y))$. Mathematically, the influence function is merely the directional derivative of the functional from the observed distribution towards a distribution with all probability weight at observed outcome y. The re-centered influence function merely adds the functional back into the influence function: $RIF(y, \nu) = IF(y, \nu) + \nu(F(y))$. Assuming that the conditional expectation of the RIF is a linear function of some observable demographic characteristics X, so that

$$E(RIF(y, \nu)) = \beta X$$

then the parameters β can be estimated via OLS.

Estimating these RIF regressions can be seen as an extension of a common technique used in the environmental justice literature. In this commonly used (e.g. by Morello-frosch et al. (2002) and Clark et al. (2014)) technique, some measure of environmental hazard (e.g. number of toxic sites or air pollutant concentrations) in a neighborhood or census tract is regressed on neighborhood characteristics. Note that the RIF for the mean is

$$RIF(y, \mu) = y$$

so that the usual linear regression $E(y|X) = \beta X$ is a special case of the RIF regression. Thus the RIF regression method I propose in fact nests the commonly used regressions in the environmental justice literature, but allowing for functionals other than the mean as outcomes.

I will consider four such functionals: the quantile function, and ordinates of each of the three Lorenz curve variants: the relative, generalized and absolute Lorenz curves. Following Essama-Nssah and Lambert (2012), the RIF for the pth quantile is

$$RIF(y, Q(p)) = \begin{cases} \hat{Q}(p) + \frac{p}{f(\hat{Q}(p))} & y > \hat{Q}(p) \\ \hat{Q}(p) - \frac{1-p}{f(\hat{Q}(p))} & y < \hat{Q}(p) \end{cases}$$

where $\hat{Q}$ is the empirical quantile point at p. The RIF for the pth Lorenz curve ordinate is

$$RIF(y, L(p)) = \begin{cases} \frac{y-(1-p)\hat{Q}(p)}{\mu_y} - \hat{L}(p)\frac{y}{\mu_y} & y < \hat{Q}(p) \\ \frac{p\hat{Q}(p)}{\mu_y} - \hat{L}(p)\frac{y}{\mu_y} & y \geq \hat{Q}(p) \end{cases}$$

where $\hat{L}(p)$ is the pth empirical Lorenz ordinate. Finally, the RIF for the pth Generalized Lorenz ordinate is

$$RIF(y, GL(p)) = \begin{cases} y - (1-p)\hat{Q}(p) & y < \hat{Q}(p) \\ p\hat{Q}(p) & y \geq \hat{Q}(p) \end{cases}$$

Finally, absolute inequality in pollution exposure, and not just the relative inequality captured by the relative Lorenz and Generalized Lorenz curves is of interest. A suitable way of capturing this is the absolute Lorenz curve due to Moyes (1987). The Absolute

Lorenz curve is defined as

$$AL(p) = GL(p) - p\mu$$

Because the recentered influence function is a linear operator, I can use the previous expressions (plus the fact that $RIF(\mu(y)) = y$) to define a recentered influence function for the absolute Lorenz curve:

$$RIF(y, AL(p)) = \begin{cases} y - (1-p)\hat{Q}(p) - py & y < \hat{Q}(p) \\ p\hat{Q}(p) - py & y \geq \hat{Q}(p) \end{cases}$$

Or, more compactly, $RIF(y, AL(p)) = RIF(y, GL(p)) - py$

Firpo et al. (2009) show that in the case of the quantile function, the estimated coefficients of an RIF regression can be interpreted as "unconditional partial effects", namely the effect of a small location shift in the distribution of X on the functional $\nu(F(y))$. This logic directly extends to the Lorenz curve cases. The unconditional partial effects estimated from an RIF regression allow for the examination of how demographic characteristics of census tracts are related to the distributional statistics of interest, but they do not directly speak to how demographics might be driving the observed trends in environmental inequality above.[12]

RIF Results

I first report the results of RIF regressions estimated using pooled data from 2005 through 2011. I supplement the data on tract-level NO_x and PM2.5 exposure with tract-level demographic data from the ACS 5-year summary files. The ACS provides

[12]However, the RIF regression method can be used to perform a decomposition analysis that can address this concern. This decomposition, performed in Appendix C allows for an examination of the extent to which demographics have shaped the change in environmental inequality over time.

data on racial and ethnic composition, median income, poverty status, age distribution, employment and educational status at the census tract level. However, these estimates are only released as 5-year average summary files (2005-2009, etc). To back out yearly estimates of the sociodemographic variables, I take a weighted average of all the five year file estimates that contain a given year.[13] So for instance, the estimate for 2005 is merely the 2005-2009 5-year file estimate, but the estimate for 2007 is an average of the estimates from the 2005-2009, 2006-2010, and 2007-2011 5 year files.[14] The environmental justice literature suggests that variables capturing various types of "disadvantage" will be of interest. I will highlight in particular the proportion of the census tract that is African-American, proportion Latino, and proportion under the poverty line, the proportion with only a high school degree and the proportion with less than a high school degree.

My empirical strategy is to estimate a fixed-effects version of the RIF regression outlined above:

$$RIF_{i,t} = \delta_i + \delta_t + \gamma_1 Black_{i,t} + \gamma_2 Latino + \gamma_3 Poverty + \gamma_4 HS + X_{i,t}\beta + e_{i,t}$$

where δ_i, δ_t are census tract and year fixed effects, and $X_{i,t}$ is a vector of other sociodemographic variables.[15] I repeat this for four different distributional functionals — quantiles, relative Lorenz ordinates, Absolute Lorenz ordinates and Generalized Lorenz ordinates. For each functional, I will estimate separate regressions for each $p \in \{0.05, 0.1, ..., 0.95\}$. Note that this linear and additively separable specification may

[13] The weights are linearly increasing in the difference between the middle year of the 5-year file and the year I want an estimate for, placing more weight on files that contain more years that are "closer" to the year in question.

[14] This approach is admittedly imperfect, and may induce measurement error. Other alternatives to matching 5-year ACS files to annual data are similarly imperfect, which highlights the limits of the data being used in this study.

[15] See Table 9 for a full list of covariates.

be subject to multicollinearity if the relevant degrees of disadvantage are correlated.[16] However, note that among the independent variables, there are no pairwise correlations above 0.85 in absolute value in the estimating sample.

Tables 9 and 10 summarize the results for RIF regressions estimated using quantiles as the distributional statistic(s) of interest for NO_x exposure and PM2.5 exposure respectively. There is some evidence for a correlation between the minority population of census tracts — blacker and more heavily Hispanic tracts are more likely to be exposed to higher amounts of NO_x and PM2.5. There is interesting heterogeneity; the proportion of a tract's population that is black has statistically significant and positive effects on exposure mostly in less-exposed tracts, while the opposite is true for Hispanic populations. Median tract income appears to be positively related to exposure at the more exposed end of the exposure distribution, although this may be a function of the fact that high income tracts are often found in central cities (with concomitant higher levels of exposure).

The unconditional quantile partial effect of a tract's population proportion black is actually negative at the top of the exposure distribution. This implies that tracts with higher concentrations of African Americans are actually less exposed, at least at the highly exposed end of the exposure distribution. The effect of educational attainment within tracts is highly varied across the distribution; tracts with higher high school drop out populations are more exposed at the bottom of the exposure distribution, while the high school-only population is related to higher exposure near the top of the distribution.

Tables 11 and 12 summarizes the results of RIF regressions estimated using the Lorenz curve ordinates as the distributional functionals of interest. Due to the multiplication of raw NO_x exposure by -1 in calculating the Lorenz curve ordinates, the Lorenz curve will lie strictly above the line of perfect equality. This means that positive coefficient estimates imply an increase in environmental inequality in conjunction

[16]One possible way to test for the appropriateness of the specification would be to more fully saturate the model with interaction terms.

TABLE 9. Quantile RIF Regression Results (NOx Exposure)

	Dependent variable:				
	NO_x Quantiles				
	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
	(1)	(2)	(3)	(4)	(5)
Black	0.124***	0.149***	−0.065	−0.620***	−2.141***
	(0.041)	(0.046)	(0.073)	(0.153)	(0.403)
Latino	−0.064	−0.052	0.245**	0.029	−1.353**
	(0.052)	(0.061)	(0.106)	(0.227)	(0.573)
Poverty	−0.121***	−0.077	−0.245***	0.161	0.956**
	(0.043)	(0.048)	(0.079)	(0.165)	(0.388)
UR	−0.447***	−0.290***	0.192**	1.066***	−1.299**
	(0.055)	(0.056)	(0.093)	(0.190)	(0.507)
HS Only	−0.028	0.002	0.178*	−0.419*	1.425***
	(0.052)	(0.062)	(0.104)	(0.221)	(0.495)
Less than HS	0.256***	0.351***	0.050	−1.176***	0.499
	(0.061)	(0.071)	(0.118)	(0.245)	(0.553)
Log Med. Income	−0.006	0.021	0.115***	−0.058	0.403***
	(0.015)	(0.018)	(0.030)	(0.061)	(0.141)
Observations	470,569	470,569	470,569	470,569	470,569

Note: *p<0.1; **p<0.05; ***p<0.01
All models include census tract and year fixed effects
Negative values can be interpreted as increasing inequality

Other variables included in regressions but omitted from this table: linguistic isolation, median age, median home value, % between 20-24, % between 25-44, % between 45-64, % between 5-19, % aged 65+, % asian/pacific islander, labor force participation rate, % native, % other race, % more than bachelor's, % unaffordable rent (30%+ of HH income), tract Gini, % some college, % commuting by bike/walk, % commuting by car, % commuting by transit, tract population, % unmarried parents, % veteran

TABLE 10. Quantile RIF Regression Results (PM2.5 Exposure)

	Dependent variable:				
	Absolute Lorenz Ordinate				
	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
	(1)	(2)	(3)	(4)	(5)
Black	0.403***	0.531***	−0.002	−0.663***	−0.286
	(0.128)	(0.131)	(0.112)	(0.176)	(0.445)
Latino	0.265	−0.125	−0.042	0.522**	1.823***
	(0.239)	(0.218)	(0.169)	(0.258)	(0.603)
Poverty	−0.203	0.392***	0.298**	−0.056	−1.379***
	(0.151)	(0.146)	(0.120)	(0.179)	(0.444)
UR	−0.064	0.054	−0.130	−1.614***	−1.530***
	(0.188)	(0.172)	(0.139)	(0.210)	(0.529)
HS Only	0.159	−0.004	−0.131	0.001	0.227
	(0.205)	(0.203)	(0.157)	(0.241)	(0.572)
Less than HS	0.263	0.297	0.288	0.430	0.014
	(0.237)	(0.226)	(0.178)	(0.274)	(0.662)
Log Med. Income	−0.159***	−0.139**	0.044	0.484***	1.099***
	(0.059)	(0.056)	(0.044)	(0.067)	(0.165)
Observations	615,630	615,630	615,630	615,630	615,630

Note: *p<0.1; **p<0.05; ***p<0.01
Negative values can be interpreted as increasing inequality
All models include census tract and year fixed effects
For more details see Table 9

with an increase in the regressor in question, whereas negative coefficient would imply greater equality in exposure. The most immediate result is that across the exposure distribution, racial composition of census tracts has a large and significant effect on inequality. Increases in the black proportion of a census tract are positively related to the cumulative share of pollution exposure up through the 25th percentile, but negatively related to cumulative exposure past the median. This implies that racial composition is inequality-enhancing for more advantaged (less exposed) tracts, i.e. a more racially diverse population is associated with greater disparities in exposure among less-exposed tracts.

The effect of median income is, as expected, negative throughout most of the pollution exposure distribution — implying that higher income tracts exhibit lower cumulative pollution shares. High poverty areas are strongly correlated with more unequal pollution exposure — the effect of the proportion of the population poverty implies an increase in environmental inequality across the whole pollution exposure distribution. There is some evidence for a correlation between educational attainment and environmental inequality — the effect of the proportion of the population with less than a high school degree is negative and significant through most of the pollution exposure distribution.

Tables 13 and 14 summarize the results of RIF regressions estimated using the Generalized Lorenz ordinates as the distributional functionals of interest. The same inference applies here as in the unconditional quantile partial effects case. Negative coefficients imply movements towards more environmental inequality, positive coefficients towards less inequality. The black and Hispanic proportions of population increase inequality throughout the pollution exposure distribution. The Hispanic effect reaches a maximum around the median of the distribution, while the black effect is largest for the most polluted tracts. Poverty seems to have similar effects as in the Lorenz curve

TABLE 11. Relative Lorenz RIF Regression Results (NOx Exposure)

	Dependent variable:				
	Relative Lorenz Ordinate				
	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
	(1)	(2)	(3)	(4)	(5)
Black	0.120***	0.019	−0.021**	−0.011**	−0.005***
	(0.017)	(0.016)	(0.010)	(0.005)	(0.002)
Latino	0.015	−0.039	0.016	0.007	0.0003
	(0.040)	(0.031)	(0.017)	(0.007)	(0.003)
Poverty	0.068***	0.078***	0.053***	0.019***	0.007***
	(0.021)	(0.018)	(0.011)	(0.005)	(0.002)
UR	−0.096***	−0.048**	0.058***	0.049***	0.022***
	(0.025)	(0.022)	(0.014)	(0.006)	(0.002)
HS Only	0.025	0.010	0.005	0.001	0.001
	(0.029)	(0.024)	(0.015)	(0.006)	(0.002)
Less than HS	−0.058*	−0.083***	−0.108***	−0.049***	−0.012***
	(0.034)	(0.028)	(0.017)	(0.007)	(0.003)
Log Med. Income	−0.015*	−0.021***	−0.006	−0.0004	0.001
	(0.009)	(0.007)	(0.004)	(0.002)	(0.001)
Observations	470,569	470,569	470,569	470,569	470,569

Note: *p<0.1; **p<0.05; ***p<0.01
Positive values can be interpreted as increasing inequality
All models include census tract and year fixed effects
For more details see Table 9

TABLE 12. Relative Lorenz RIF Regression Results (PM25 Exposure)

	Dependent variable:				
	Relative Lorenz Ordinate				
	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
	(1)	(2)	(3)	(4)	(5)
Black	-0.004	-0.016^{**}	-0.026^{***}	-0.018^{***}	-0.006^{***}
	(0.004)	(0.006)	(0.005)	(0.003)	(0.001)
Latino	0.011^{**}	0.017^{**}	0.012	0.003	-0.002
	(0.005)	(0.008)	(0.008)	(0.005)	(0.003)
Poverty	-0.017^{***}	-0.029^{***}	-0.015^{***}	-0.001	0.003
	(0.004)	(0.006)	(0.005)	(0.004)	(0.002)
UR	0.010^{**}	-0.013^{*}	-0.026^{***}	-0.010^{**}	-0.001
	(0.005)	(0.008)	(0.007)	(0.004)	(0.002)
HS Only	0.008	0.009	-0.001	-0.005	-0.005^{**}
	(0.005)	(0.008)	(0.007)	(0.005)	(0.002)
Less than HS	-0.008	-0.007	-0.004	-0.006	-0.001
	(0.006)	(0.009)	(0.008)	(0.005)	(0.003)
Log Med. Income	0.008^{***}	0.016^{***}	0.016^{***}	0.009^{***}	0.002^{***}
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Observations	615,630	615,630	615,630	615,630	615,630

Note: *p<0.1; **p<0.05; ***p<0.01
Positive values can be interpreted as increasing inequality
All models include census tract and year fixed effects
For more details see Table 9

RIF regressions, increasing exposure inequality near the middle-top of the pollution distribution. Likewise, median income seems to decrease inequality throughout the distribution, suggesting a significantly negative income-exposure relationship, which was also visible in the relative Lorenz RIF results.

TABLE 13. Generalized Lorenz RIF Regression Results (NOx Exposure)

	Dependent variable:				
	Generalized Lorenz Ordinate				
	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
	(1)	(2)	(3)	(4)	(5)
Black	−0.230***	−0.052	0.014	−0.010	−0.022
	(0.039)	(0.049)	(0.050)	(0.046)	(0.044)
Latino	−0.042	0.044	−0.071	−0.064	−0.054
	(0.093)	(0.097)	(0.094)	(0.089)	(0.086)
Poverty	−0.161***	−0.209***	−0.195***	−0.151***	−0.137***
	(0.049)	(0.056)	(0.057)	(0.053)	(0.051)
UR	0.210***	0.148**	−0.019	0.016	0.074
	(0.058)	(0.067)	(0.067)	(0.063)	(0.061)
HS Only	−0.061	−0.045	−0.051	−0.051	−0.054
	(0.065)	(0.074)	(0.074)	(0.070)	(0.067)
Less than HS	0.187**	0.302***	0.421***	0.359***	0.308***
	(0.079)	(0.088)	(0.087)	(0.082)	(0.079)
Log Med. Income	0.029	0.040*	0.013	0.002	0.0003
	(0.020)	(0.022)	(0.022)	(0.021)	(0.020)
Observations	470,569	470,569	470,569	470,569	470,569

Note: *p<0.1; **p<0.05; ***p<0.01
Negative values can be interpreted as increasing inequality
All models include census tract and ear fixed effects
For more details see Table 9

The effects of tract-level demographic characteristics on the final functional, the Absolute Lorenz curve, are summarized in Tables 15 and 16. Once again, racial composition is correlated with environmental inequality, although with heterogeneity

TABLE 14. Generalized Lorenz RIF Regression Results (PM25 Exposure)

	Dependent variable:				
	Generalized Lorenz Ordinate				
	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
	(1)	(2)	(3)	(4)	(5)
Black	0.046	0.163^{**}	0.273^{***}	0.192^{***}	0.075
	(0.044)	(0.075)	(0.082)	(0.073)	(0.065)
Latino	-0.178^{***}	-0.310^{***}	-0.374^{***}	-0.370^{***}	-0.373^{***}
	(0.061)	(0.107)	(0.129)	(0.124)	(0.114)
Poverty	0.204^{***}	0.356^{***}	0.257^{***}	0.153^{*}	0.127^{*}
	(0.047)	(0.078)	(0.088)	(0.083)	(0.077)
UR	-0.016	0.337^{***}	0.622^{***}	0.596^{***}	0.578^{***}
	(0.058)	(0.093)	(0.104)	(0.098)	(0.090)
HS Only	-0.095^{*}	-0.122	-0.046	-0.023	-0.028
	(0.057)	(0.099)	(0.117)	(0.110)	(0.100)
Less than HS	0.044	-0.006	-0.097	-0.127	-0.210^{*}
	(0.066)	(0.114)	(0.132)	(0.124)	(0.113)
Log Med. Income	-0.117^{***}	-0.248^{***}	-0.314^{***}	-0.293^{***}	-0.258^{***}
	(0.017)	(0.028)	(0.033)	(0.031)	(0.029)
Observations	615,630	615,630	615,630	615,630	615,630

Note: *p<0.1; **p<0.05; ***p<0.01
Negative values can be interpreted as increasing inequality
All models include census tract and year fixed effects
For more details see Table 9

across pollutants. Tract African-American populations are associated with higher absolute inequality at the more exposed end of the distribution for NO_x, but are associated with lower absolute inequality for PM2.5. Tract Hispanic population is associated with higher absolute inequality for both pollutants, but this association is only statistically significant for PM2.5. Interestingly, median tract income is actually associated with lower absolute inequality for NO_x, but with higher absolute exposure inequality across the distribution for PM2.5.

TABLE 15. Absolute Lorenz RIF Regression Results (NOx Exposure)

	Dependent variable:				
	Absolute Lorenz Ordinate				
	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
	(1)	(2)	(3)	(4)	(5)
Black	−0.227***	−0.044	0.030	0.014	0.007
	(0.036)	(0.040)	(0.029)	(0.015)	(0.006)
Latino	−0.037	0.058	−0.044	−0.023	−0.005
	(0.086)	(0.077)	(0.052)	(0.025)	(0.010)
Poverty	−0.148***	−0.177***	−0.131***	−0.056***	−0.023***
	(0.045)	(0.045)	(0.033)	(0.016)	(0.006)
UR	0.198***	0.119**	−0.077**	−0.072***	−0.031***
	(0.053)	(0.053)	(0.039)	(0.020)	(0.008)
HS Only	−0.056	−0.032	−0.024	−0.011	−0.005
	(0.060)	(0.059)	(0.043)	(0.021)	(0.008)
Less than HS	0.158**	0.228***	0.275***	0.139***	0.044***
	(0.073)	(0.070)	(0.050)	(0.025)	(0.010)
Log Med. Income	0.029	0.040**	0.012	0.001	−0.001
	(0.019)	(0.018)	(0.013)	(0.006)	(0.002)
Observations	470,569	470,569	470,569	470,569	470,569

Note: *p<0.1; **p<0.05; ***p<0.01
Negative values can be interpreted as increasing inequality
All models include census tract and year fixed effects
For more details see Table 9

TABLE 16. Absolute Lorenz RIF Regression Results (PM2.5 Exposure)

	Dependent variable:				
	Absolute Lorenz Ordinate				
	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
	(1)	(2)	(3)	(4)	(5)
Black	0.045	0.160^{**}	0.268^{***}	0.185^{***}	0.066^{***}
	(0.041)	(0.065)	(0.058)	(0.035)	(0.017)
Latino	-0.137^{**}	-0.206^{**}	-0.167^{*}	-0.060	0.0001
	(0.056)	(0.090)	(0.088)	(0.060)	(0.033)
Poverty	0.188^{***}	0.314^{***}	0.174^{***}	0.028	-0.023
	(0.044)	(0.067)	(0.061)	(0.040)	(0.021)
UR	-0.076	0.188^{**}	0.323^{***}	0.148^{***}	0.040
	(0.054)	(0.081)	(0.072)	(0.048)	(0.026)
HS Only	-0.086	-0.100	-0.001	0.043	0.052^{*}
	(0.053)	(0.085)	(0.082)	(0.054)	(0.027)
Less than HS	0.067	0.051	0.017	0.044	-0.004
	(0.062)	(0.097)	(0.092)	(0.060)	(0.031)
Log Med. Income	-0.093^{***}	-0.186^{***}	-0.190^{***}	-0.107^{***}	-0.035^{***}
	(0.015)	(0.024)	(0.023)	(0.015)	(0.008)
Observations	615,630	615,630	615,630	615,630	615,630

Note: *p<0.1; **p<0.05; ***p<0.01
Negative values can be interpreted as increasing inequality
All models include census tract and year fixed effects
For more details see Table 9

These results together shine additional light on the interconnectedness of the concepts of environmental justice and environmental inequality — it appears that racial differences and poverty are important correlates of environmental inequality. However, these results are at best only potentially interesting correlations, and cannot claim to have uncovered the causes of environmental inequality. In particular, I cannot speak to whether any correlation between the racial composition of census tracts and exposure inequality necessarily represents environmental racism, although the generally more robust correlations between income variables (median income and poverty) and exposure inequality suggests that the level and distribution of income may be important for explaining changes in the exposure distribution, a possibility I take up more formally in Chapter IV of this dissertation.

Conclusion

This paper proposes a "dashboard" approach to the considerations of distributional concerns in environmental policy analysis. Rather than choosing a single summary statistic (e.g. the black-white exposure gap, or the Atkinson index) as a social evaluation function when analyzing the (environmental) distributional effects of policy, looking at several indicators can be a more fruitful approach that more fully illustrates how policy is affecting the distribution of exposure to harmful pollutants.

I propose two ways of thinking about the degree to which the distribution of pollution exposure is "unequal": a horizontal equity concept, where the primary concern is the inequality in exposure across subgroups, and a vertical equity concept, where the primary concern is the degree of inequality across the whole population. I apply scalar measures of each equity concept (e.g. the Atkinson index for vertical equity, and the average black-white exposure gap for horizontal equity), in addition to measures that allow for the examination of the whole distribution (black-white gaps by percentile for horizontal equity, and Lorenz curves for vertical equity).

I apply this dashboard of measures of environmental inequality and environmental justice to novel data on pollution exposure derived from remote sensing observations of ground level NO_x and PM2.5 concentrations. By matching these remote sensing data with census data on the distribution of population, I am able to measure not just average exposure, but also all of the environmental inequality and environmental justice measures, on an annual basis over the period 1998-2014 (only 2005-2011 for NO_x exposure). Average exposure to both pollutants has decreased markedly over time, as have most measures of environmental inequality and environmental justice.

I use a re-centered influence function regression estimation strategy to isolate how individual tract-level demographic characteristics affect measures of environmental inequality as a whole. I find that many characteristics that correlate with disadvantage — such as poverty rates, education levels and racial minority populations — have statistically significant associations with environmental inequality, although notably the patterns of sign and significance often differ between pollutants, suggesting the interaction between the distribution of people and the production and the fate-and-transport of pollution may depend in part of the chemical properties of the pollutants in question.

This largely descriptive exercise has not, by design, involved the identification of causal factors that might be related to the observed trends in environmental quality, environmental inequality and environmental justice. The final substantive chapter of this dissertation takes up this challenge by synthesize the first substantive chapter's discussion of trends in income inequality with this chapter's discussion of environmental inequality, and considers whether income inequality within metropolitan areas might have a causal effect on changes in the distribution of pollution exposure.

CHAPTER IV

ENVIRONMENTAL JUSTICE VIEWED FROM OUTER SPACE: HOW DOES GROWING INCOME INEQUALITY AFFECT THE DISTRIBUTION OF POLLUTION EXPOSURE?

Introduction

Income inequality has increased substantially in the last several decades, both within the United States as a whole (Piketty and Saez (2003)) and within individual US states and metropolitan areas (Frank (2009), and Chapter II of this dissertation). This fact has spawned a contentious debate, both within and outside of academia. Much of this debate has concerned itself with the *causes* of this increase in income inequality. Substantially less time and effort has been expended on considering the potential *effects* of rising inequality. In this paper, I will examine how increases in income inequality might affect the distribution of environmental disamenities, specifically, exposure to ground-level nitrogen oxides (NO_x). Exposure to NO_x itself, and to the ozone and smog generated when ground-level NO_x interacts with sunlight and volatile organic compounds, is a major health hazard, contributing as many as 10,000 excess deaths per year in the United States (Caiazzo et al. (2013)). Concern about NO_x has become particularly salient in light of the recent Volkswagen emissions scandal, wherein software on several car models was designed to circumvent emissions testing regimes in the US and Europe, resulting in thousands of tons of excess NO_x emissions.

Using data on ground-level NO_x concentrations inferred from remote sensing observations by NASA's Aura satellite, I am able to measure pollution exposure at a fine geographic resolution. I combine these data with information about income distributions in US metropolitan statistical areas (MSAs) to examine the causal effect of changes in within-MSA income inequality on the distribution of NO_x exposure within MSAs. There is likely to be some joint endogeneity between the income inequality and pollution exposure, operating through migration between metropolitan areas. To identify a causal effect in the presence of this endogeneity, I employ an instrumental variables approach, using a version of the simulated instrument introduced by Boustan et al. (2013). I construct an instrument for income inequality by simulating counterfactual MSA-level income distributions which are independent of changes in the MSA-level distribution of pollution exposure, thereby eliminating the potential endogeneity bias due to locational sorting.

Using this empirical strategy, I find that increases in metropolitan area income inequality lead to decreases in the *average* level of NO_x exposure within MSAs. However, increases in income inequality are also associated with an increase in *environmental* inequality. I conclude that the decrease in average exposure caused by income inequality disproportionately benefits the most-advantaged, although the least-advantaged are still better off in absolute terms. I then consider to what extent the political system might serve as a potential mechanism for the effect of inequality on pollution exposure. Specifically, I examine how income inequality is related to the pro-environmental voting records of legislators, as measured by their League of Conservation Voters score. I find that increases in income inequality within a US senator's state lead to an increase in the LCV scores of Democratic senators, but have no discernible effect on the LCV scores of Republican senators.

The remainder of the paper proceeds as follows. I review the relevant literature on environmental justice and income inequality. I then offer a conceptual model for my analysis, and describe the data sources I use. Finally, I introduce my identification strategy, and present estimation results. I discuss potential mechanisms, and examine how the effect of income inequality on the distribution of environmental amenities may work through the political system. I conclude with some ideas for potential extensions and directions for future research.

Previous Literature

Environmental Justice

The earliest environmental justice literature was motivated by the identification of differences in the levels of pollution exposure across discrete socio-economic categories. There are many capable reviews of this literature, including a recent volume collected by Banzhaf (2012) and surveys including Mohai et al. (2009) and Brulle and Pellow (2006). The central claim in this literature is that minority and poorer households are disproportionately more likely to be exposed to pollution, compared to white and richer households. In practice, papers in this literature are often able to establish only that areas near toxic emitting firms and areas with higher measured levels of pollution tend to be less white, to be less educated and/or to have lower median income than less-polluted areas.

There are slightly fewer papers that deal explicitly with environmental inequality rather than environmental injustice. By environmental inequality I am referring to a measure capturing the entire distribution of the level of exposure to pollution within an area.[1] This contrasts with the concept of environmental injustice, which refers to differences in subgroup means. This distinction is not entirely standard across the broad environmental injustice literature (e.g. Downey (2007)). Maguire and Sheriff (2011) and Sheriff and Maguire (2014) provide a direct adaptation, from the income inequality literature, of tools to rank distributions of environmental amenities or

[1]For the purposes of this paper, Environmental Inequality will be captured by a scalar index, although it is possible to measure environmental inequality using, e.g. Lorenz curves without sacrificing normative content.

disamenities. They suggest that normatively based scalar indices should be used for studying the environmental justice effects of a given policy. In a similar paper, although in a different context, Harper et al. (2013), apply some of the tools from the literature on the measurement of income inequality to the measurement of health outcomes, again in the context of policy analysis.

Income Inequality and Environmental Quality

Scholars in environmental economics have long sought to model the simple relationship between incomes and pollutant emissions. This relationship is often embedded within considerations of the Environmental Kuznets Curve (EKC)—the idea being that pollution levels first rise as countries develop (i.e. as average incomes increase), and then, at some point, begin to decline, as demand for environmental amenities increases. Although the EKC model concerns only average incomes, and not the full distribution of income, the EKC model has spawned a few papers that examine how income inequality, as distinct from just average incomes, might affect environmental quality. Berthe and Elie (2015) summarize this small but growing literature, and attempt to synthesize its findings by suggesting a theoretical structure which can explain all of the potential pathways through which income inequality might affect environmental quality. This literature has not yet reached any consensus on whether there is a relationship between income inequality and environmental quality, let alone has it established the sign of this relationship or whether this relationship is causal.

Boyce (1994) is perhaps the first to consider this relationship, followed by Scruggs (1998), Ravallion et al. (2000) and others.[2] Papers in this literature share a common

[2] A non-exhaustive list includes Torras and Boyce (1998), Heerink et al. (2001), Magnani (2000) and Neumayer (2004). More-recent papers in this literature include Zwickl and Moser (2015) and Baek and Gweisah (2013)

framework. Consider a propensity-to-emit function (PEF) which describes the amount of emissions embedded in each household's consumption activity at any given level of household income, a function that I assume is non-decreasing over some range of income.[3] To consider how income inequality might affect environmental quality, it is necessary only to know the concavity of the PEF. If the function is concave, so that the marginal propensity to emit is decreasing in income, then Pigou-Dalton transfers (mean-preserving progressive transfers of income) may actually increase average emissions levels. Conversely, if the PEF is convex, so that the marginal propensity to emit is increasing in income, Pigou-Dalton income transfers will reduce average emissions levels. In other words, *increases* in income inequality may account for environmental degradation if the PEF is convex, while *decreases* in income inequality may account for environmental degradation if the PEF is concave.

There is, however, no obvious reason to expect, a priori, that the propensity to emit function need necessarily be either convex or concave. No consensus has emerged from the resulting empirical literature, although patterns have sometimes been observed. In cross-country regressions, generally suggest that rising income inequality increases environmental degradation (e.g. Drabo (2011)). Basing the analysis simply on within-country variation, however, has produced a wide array of estimates. There appear to be rather different relationships for developing or middle income countries (e.g. China, in the case of Golley and Meng (2012)) than for developed countries (e.g. Sweden, in the case of Brännlund and Ghalwash (2008)). In particular, the weight of the evidence suggests that the propensity to emit function may be convex for developing countries, but concave for developed nations. This implies that developing countries may be able to reduce emissions

[3]In a recent working paper, Levinson and O'Brien (2015) introduce the related concept of an environmental Engel curve

and enhance equity simultaneously, whereas developed countries face a trade-off between equity and environmental quality.

There are two difficulties faced by earlier papers in this literature. First, the observability of environmental quality is often incomplete, relying on irregularly spaced ground monitors. Second, there is potential endogeneity between environmental quality and the distribution of income. No paper in this literature has proposed a credible strategy to recover estimates of the causal effect of income inequality on the environment in the presence of this potential endogeneity. I address these issues by (1) utilizing satellite data that provide information about air quality in unmonitored areas, and (2) using an instrument for inequality that addresses endogeneity due to locational sorting.

Data and Empirical Strategy

My goal is to capture the impact of income inequality on some functional of the pollution exposure distribution. As with many questions of distributional measurement, the choice of the proper geographic scope is important. My analysis will be performed at the level of metropolitan statistical areas (MSAs) in the United States. Metropolitan areas are typically the geographic context in which the health effects of exposure to non-uniformly mixing pollutants like NO_x are most acute. The differences in these exposures within an individual metropolitan area contribute to the environmental inequality I measure. Metropolitan areas are also sufficiently large to allow for estimation of income inequality measures using publicly available Census Bureau data at an annual frequency. Thus MSAs are a natural choice as the geographic setting for this analysis.

To examine the relationship between income inequality and the distribution of pollution exposure, I utilize two novel data sources. I use data on NO_x concentration levels from the Aura satellite to calculate both metropolitan-area average exposure levels and environmental inequality, which I capture via two alternative measures (functionals of the pollution exposure distribution). I supplement these pollution exposure data with metropolitan area income inequality measures from Chapter II of this dissertaion, and socio-demographic variables calculated from the American Community Survey (ACS) 1-year files.

The Aura satellite was launched in 2004 with a mission to measure the atmospheric composition both for the troposphere (i.e. at ground-level) and for the stratosphere (i.e. the ozone layer). The Ozone Monitoring Instrument (OMI), carried by the Aura satellite, provides comprehensive global observations of the tropospheric vertical column density of NO_x on a fixed grid. Vertical column densities do not correspond with ground-level concentrations of NO_x on a one-for-one basis. In general, it is necessary to infer the ground-level concentrations via the use of ground-level observations from monitoring stations and a chemical air-transport model. I use the *SP_GC v*1.01 dataset provided by the Air Composition Analysis Group (ACAG) at Dalhousie University, which is estimated using the GEOS-CHEM chemical transport model.[4] More information about these data are available in Lamsal et al. (2008) and Lamsal et al. (2010).[5] Figure 19 summarizes the national distribution of NO_x concentrations in the form of choropleth maps.[6]

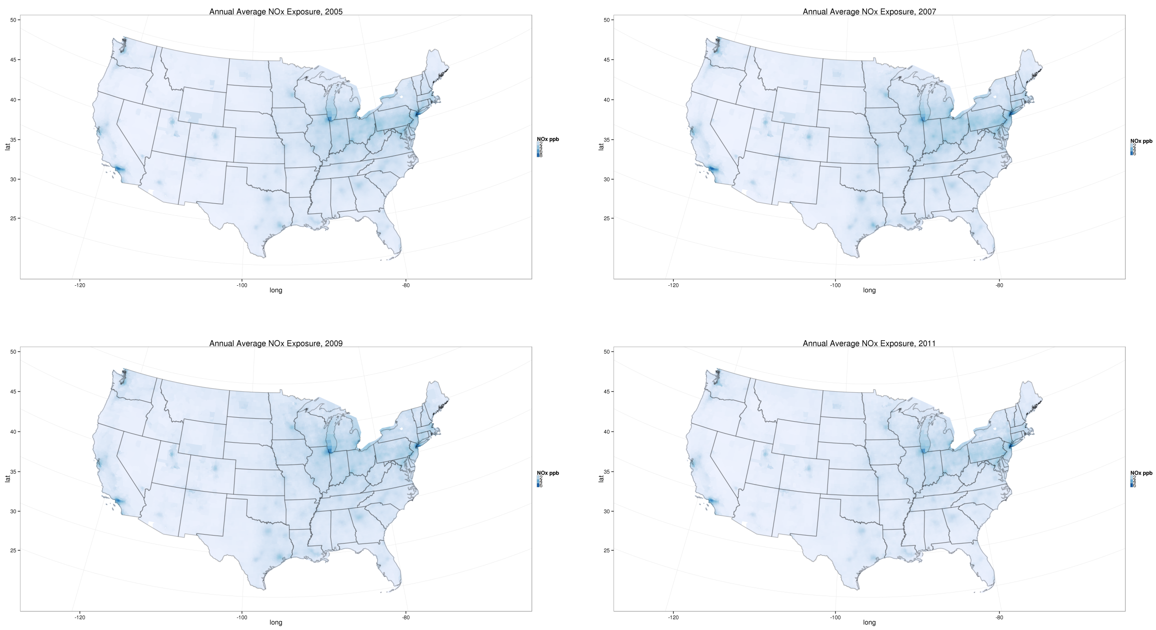
The ACAG data provide yearly average NO_x concentrations on a 0.1×0.1-degree grid for North America. To measure the person-level distribution of pollution exposure, I geographically interpolate the ACAG data to the census-tract level by

[4] A previous version of this paper used an alternate approach to inferring ground-level concentrations by comparing the distribution of ground-level observations from monitoring stations with the distribution of vertical column densities.

[5] These data are available from the ACAG website at: `http://fizz.phys.dal.ca/~atmos/martin/`

[6] More detailed choropleths, including metropolitan area maps, are available in an online appendix accessible at `http://pages.uoregon.edu/jlv/jmp_appendix.html`. An additional set of interactive maps visualizing the distribution of NO_x exposure in 2005 is available at `http://pages.uoregon.edu/jlv/interactive_NOX_maps.html`

FIGURE 19. Average Annual NO_x Exposure by Census Tract, 2005–2011



inverse distance weighting.[7] These annual estimates of NO_x concentrations at the census-tract level are then used to calculate metropolitan area average exposure and the necessary environmental inequality measures, which are functionals of the NO_x exposure distribution.

The final estimating sample includes observations for 265 MSAs, for the period from 2005 to 2011. Metropolitan area income inequality is measured by the well-known Gini coefficient, estimated from income microdata provided in the 1-year American Community Survey (ACS) files. I use the Gini coefficient estimates from chapter II of this disseration, which address topcoding by adopting a multiple imputation approach, simulating the censored right tail of the income distribution as following a Generalized Beta II

[7]Inverse distance weighting interpolates to unobserved locations x_i by calculating the weighted average of observed locations nearby (x_j). The weights are calculated as $w_j = \frac{1}{d(x_i,x_j)^2}$, where $d(.)$ is the Euclidean distance operator.

distribution. Income inequality estimates are calculated for all MSAs identified in the public-use ACS microdata files (265 in total). I supplement my two main variables with several demographic measures including median income, race, industry characteristics, poverty, and transportation patterns.

Measuring Environmental Inequality

Most previous research focuses on the average level of pollution exposure within a jurisdiction, generally calculated as the population-weighted mean of ground monitor readings. I extend this by examining not just the average (i.e. first moment of the distribution), but also functionals of the entire distribution of pollution exposure within each MSA.[8] I define two types of measures which summarize the pollution exposure distribution. The first type I describe as measures of "environmental inequality." These measures summarize the marginal distribution of pollution exposure, across the population, without considering individual attributes other than the level of exposure. The second type I term measures of "environmental justice." These summarize the joint distribution of pollution exposure and other demographic characteristics, likewise across the population.

Measures of environmental inequality thus summarize just the unconditional distribution of pollution exposure, without considering subgroup differences. Sheriff and Maguire (2014) describe how to adapt common measures of inequality to a case where the distribution of interest is a "bad" (as for NO_x exposure). If inequality measures are interpreted purely as a description of the spread of a distribution, then this distinction is largely irrelevant. However, normative interpretations of these inequality measures

[8] While I will not be using the higher moments (e.g. variance, skewness) of the distribution per se, the measures I use can be thought of as capturing the same sort of information about the shape and dispersion of exposure distributions.

are important for analyzing the tradeoff between equity and efficiency when making environmental policy decisions, and these measures are not symmetric across the good/bad distinction. I adopt the approach of Sheriff and Maguire (2014) and modify commonly used inequality measures so that it is possible to rank distributions in an ethically sensible fashion.

In contrast, measures of environmental justice are defined in terms of the differences in average pollution exposure between subgroups, where the subgroups correspond to conventional notions of advantage and disadvantage. I will consider differences in exposures across racial lines: specifically, the difference in pollution exposures between African-Americans and whites, as well as the difference in exposures between Latinos and whites. I will consider differences in exposures across income levels as well, defined as the difference in average pollution exposure between the top quintile of the income distribution and the bottom quintile of the income distribution.

I will capture environmental inequality by two types of scalar inequality measures, each of which induce a complete ordering of pollution exposure distributions.[9] I will divide these types of measures into two groups: *relative* inequality measures and *absolute* inequality measures. The substantive difference in these measures lies in their invariance properties. A *relative* inequality measure $I_R(x)$, for a vector x characterizing the empirical pollution exposure distribution, satisfies the property of scale invariance:

$$\forall x : I_R(x) = I_R(kx), k > 0 \tag{4.1}$$

[9]Environmental inequality can also be captured using a Lorenz curve (or variations thereof). However, Lorenz curves induce only a partial ordering of pollution exposure distributions, since the Lorenz dominance criteria requires that the curves do not cross.

In contrast, an absolute inequality measure, $I_R(x)$, satisfies the property of translation invariance:

$$\forall x : I_A(x) = I_A(x+k) \tag{4.2}$$

where k is any vector, in the domain of x, whose entries are all equal.

There has been some controversy over which of these two measures more accurately captures moral intuitions about inequality in general. Following Kolm (1976), the key distinction between the two types of measures lies in how the worst-off (most-exposed) individuals are treated. Consider the two inequality-preserving transformations above. A proportional increase in pollution exposure as in equation (4.1) will in fact result in much larger changes for the most exposed individual. A equiproportional increase in pollution of, e.g., 20% will by definition leave *relative* inequality unchanged, but will increase absolute inequality. Likewise, an equal-sized increase in pollution will decrease relative inequality, but leave *absolute* inequality unchanged.

Of the two measures, a strong case can be made that absolute inequality measures may be more appropriate when considering the distribution of pollution exposures. Absolute differences in exposure have negative health effects, while relative differences in exposure may not necessarily correspond to substantive health disparities.. Thus, the absolute measures will respect the spirit of the environmental justice movement's concept of equity (defined in terms of absolute differences in exposure), unlike the relative measures.

I use the Atkinson index, as modified by Sheriff and Maguire (2014) to quantify *relative* inequality:

$$A_I(x) = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i}{\mu_x}\right)^{1-\alpha}\right]^{\frac{1}{1-\alpha}} - 1,\ \alpha \leq 0 \tag{4.3}$$

Here, α is an environmental inequality aversion parameter in the associated social welfare function (SWF). As $\alpha \to 0$, the associated SWF becomes increasingly utilitarian, and as $\alpha \to -\infty$ the implied SWF becomes increasingly Rawlsian.

Likewise, I use the (modified) Kolm-Pollak index to quantify *absolute* inequality:

$$K_I(x) = -\frac{1}{\kappa} \ln \frac{1}{N} \sum_{i=1}^{N} e^{-\kappa(x_i - \mu_x)}, \kappa < 0 \tag{4.4}$$

where κ can be interpreted as an environmental inequality aversion parameter for the associated social welfare function. Symmetrically with the α parameter in the previous case, as $\kappa \to 0$ the SWF becomes increasingly utilitarian, and as $\kappa \to -\infty$, it becomes increasingly Rawlsian.

There are two reasons to believe that the estimates of these measures of environmental justice and environmental inequality should be regarded as lower bound estimates. First, by construction and due to data limitations, I assume that all census tract residents have the same level of exposure, which is equivalent to assuming that within-tract inequality is zero. Second, I use the actual ground level concentration of NO_x as a measure of exposure. However, individuals may be able to engage in averting behavior to avoid exposure to a given ambient concentration level. If the ability to engage in averting behavior is related to the degree of advantage then the estimates of environmental justice and environmental inequality based on concentration levels will overestimate exposure by advantaged groups and hence *underestimate* the gap in exposure between advantaged and disadvantaged groups.

IV Approach

I assume, as a baseline, that the relationship between a functional of an MSAs pollution exposure distribution ν and an MSA's income inequality, measured by the Gini coefficient, can be modeled as

$$\nu\left(Pollution\right)_{i,t} = \beta_1 Gini_{i,t} + \beta_2 X_{i,t} + u_{i,t} \tag{4.5}$$

$$u_{i,t} = \alpha_i + e_{i,t} \tag{4.6}$$

where α_i is the MSA-specific fixed component of the error term, and $e_{i,t}$ is white noise. To recover the effect of income inequality on the pollution exposure distribution, I can estimate a model in first-difference form:[10]

$$\Delta\nu\left(Pollution\right)_{i,t} = \beta_1 \Delta Gini_{i,t} + \beta_2 \Delta X_{i,t} + \Delta u_{i,t} \tag{4.7}$$

Alternatively, I could estimate equation (4.5) directly using a fixed-effects specification. This baseline model does not absorb time-varying unobserved heterogeneity. I address this heterogeneity by allowing for MSA-specific linear trends:

$$\nu\left(Pollution\right)_{i,t} = \beta_1 Gini_{i,t} + \beta_2 X_{i,t} + \alpha_i t + u_{i,t} \tag{4.8}$$

which can be estimated either in a first-difference or fixed-effects specification. Even after allowing for time-varying heterogeneity, however, none of these estimates of β_1 are sufficient on their own to establish a causal effect.

When attempting to investigate the relationship between inequality in the distribution of income and inequality in the distribution of pollution exposure, there

[10] This is the preferred OLS specification in Boustan et al. (2013).

are two significant challenges to identifying a causal effect. First, there may be long-run reverse causality, as implied by the literature on the intergenerational transmission of inequality (for example Currie (2011)). Increases in environmental inequality might lead to increases in future income inequality, although only on intergenerational time scales. Second, on shorter time scales, there may be endogenous locational sorting *between* metropolitan areas, where sorting differs systematically across the income distribution.

For example, if rich households disproportionately migrate out of metropolitan areas with high average levels of pollution exposure, or high inequality in pollution exposures, then there is the potential for reverse causality. At the other end of the income distribution, poorer households may disproportionately migrate into metropolitan areas with higher levels of pollution (or alternatively with less equitable distributions of pollution exposure) which may change the income distribution in these areas.[11] These two potential migration flows have conflicting effects on the income distribution—if rich households migrate out, this will decrease the Gini coefficient of income inequality, however if poor households migrate in, this will increase the Gini coefficient of income inequality.[12] If both types of flows occur, the effect on the Gini coefficient is ambiguous.

To address these concerns about endogeneity, I adopt a simulated instrumental variables strategy. Specifically, I construct an instrument for income inequality that definitionally rules out any between-MSA sorting by "freezing" the MSA-level income distribution in an initial year and simulating the evolution of counter-factual income distributions. These counter-factual income distributions are constructed by allowing

[11]These are related to the concept of "coming to the nuisance", wherein environmental disamenities depress local rents, which in turn attracts lower income households.

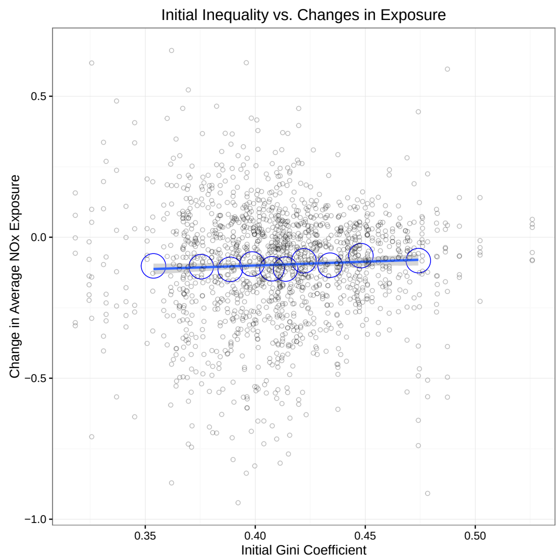
[12]To see this, consider the following toy example. Suppose there are 4 people in a city, with incomes of 10,25,50 and 150. The Gini coefficient in the city is 0.473. If the richest person leaves, the Gini coefficient drops to 0.314. If, on the other hand, a new poor person (with income 10) moves in, the Gini coefficient increases to 0.52

each decile of each MSAs income distribution to follow the growth trends observed for the corresponding deciles of the national income distribution. Using this instrument makes it possible to identify the causal effect of income inequality on the distribution of NO_x exposure by cutting off any potential variation due to between-MSA sorting.

This instrument is very similar to the instrument used in Boustan et al. (2013). This type of identification strategy has also been used in other settings, as in Enamorado et al. (2014), who examine the causal effect of income inequality on crime in Mexican municipalities. This class of simulated instruments are examples of "Bartik-style" instruments, which have been used in a wide variety of settings. As Baum-Snow and Ferreira (2015) note, the ability of this strategy to identify the causal effect of income inequality hinges on the assumption that the initial level of income inequality within metropolitan areas is independent of changes in the distribution of pollution exposure, except through its effect on the actual trends in MSA-level income inequality. One test of this identifying assumption would then be to examine the correlation between initial income inequality and subsequent changes in the distribution of NO_x exposure. Figure 20 visualizes the relationship between initial income inequality and changes in average exposure. The slope of the line of best fit through this scatterplot is not statistically different from zero, which can be taken as evidence that the identifying assumptions hold. Essentially identical results hold for all the other functionals of the NO_x exposure distribution used as outcome variables.

To construct the instrument, I use the same microdata on incomes from the ACS 1-year files that I used to calculate the measures of MSA-level income inequality (using the Generalized Beta II multiple imputation method described in Chapter II of this

FIGURE 20. Initial MSA Income is Unrelated to Subsequent Changes in NO_x Exposure



dissertation).[13] For each MSA identified in the ACS microdata, I calculate the mean income of each decile of the MSA income distribution in the first year of the sample (2005). I also calculate the means of each decile of the national distribution for each year, and calculate the growth rates for each of these decile means from 2006 to 2011. I leave each MSA's information out of calculations of national decile average income levels and trends when constructing that MSA's simulated counterfactual income distribution. This eliminates the possibility that large MSAs, e.g. New York, may be disproportionately driving national level income trends.[14]

I then use these national decile-mean growth rates to construct synthetic income distributions for each MSA in each year in the period 2006-2011. I assign each MSA decile

[13]The income concept used here is pre-tax, post-transfer household income, adjusted for household size by applying an equivalence scale equal to the square root of household size.

[14]Previous versions of this paper used the same national decile income levels and trends for each MSA. There is no qualitative difference in results when using the instruments, although the "leave-one-out" estimates are more precise.

in 2005 to its matching national decile in 2005. I calculate the instrument by assuming that every MSA decile grew at the same rate as the corresponding national decile mean for each year in 2006-2011. In other words, I "freeze" the individual MSA income distribution in 2005, and then simulate future counterfactual distributions based only on nationwide trends. Once I have these sets of simulated decile means, I calculate a Gini coefficient using the simulated decile means for each year between 2006 and 2011.

With these simulated Gini coefficients as my instrumental variable, I can estimate the relationship between income inequality and environmental inequality using two-stage least squares methods. The baseline model can be estimated in first-differences as

$$\text{First stage: } \Delta Gini_{i,t} = \alpha + \gamma_1 \Delta SynthGini_{i,t} + \delta \Delta X_{i,t} + v_{i,t} \tag{4.9}$$

$$\text{Second stage: } \Delta EnvIneq_{i,t} = \alpha + \beta \Delta \widehat{Gini}_{i,t} + \Gamma \Delta X_{i,t} + \epsilon_{i,t} \tag{4.10}$$

To account for potential time-varying heterogeneity, as mentioned above, I additionally estimate models that include MSA-specific trends α_i:

$$\text{First stage: } \Delta Gini_{i,t} = \alpha_i + \gamma_1 \Delta SynthGini_{i,t} + \delta \Delta X_{i,t} + v_{i,t} \tag{4.11}$$

$$\text{Second stage: } \Delta EnvIneq_{i,t} = \alpha_i + \beta \Delta \widehat{Gini}_{i,t} + \Gamma \Delta X_{i,t} + \epsilon_{i,t} \tag{4.12}$$

where in either case $X_{i,t}$ is a vector of exogenous sociodemographic covariates that enter into both the first and second stages. These controls include median income, race, industry characteristics (proportion of employment in manufacturing), poverty, and transportation patterns (such as the percent of the population that commutes by car, and the average commute time). Like fixed-effects models, these first difference models will absorb any permanent features (terrain, climate, etc) which will affect between-MSA variation in environmental inequality. The remaining variation in pollution exposure can be attributed to socio-demographic factors and economic activity.

Figure 21 and Table 17 summarize the first-stage results of the model. As Figure 21 shows, my simulated instrument for MSA-level income inequality is highly correlated with actual MSA-level income inequality. The first column of Table 17 reports first-stage results for a model without MSA-specific linear trends, while the second column reports first-stage results for a model with MSA-specific linear trends. The estimated coefficient associated with the instrument is positive and close to unity, and the F-statistics for the two first-stage specifications are 70.5 and 55.34 respectively. These F-statistics are well above the rule-of-thumb value of 10, which indicates that a weak first stage is not a problem. Together with the construction of the instrument, this can be taken as evidence that I obtain unbiased causal estimates of β. Additionally, Figure 22 summarizes the reduced form effect of the simulated instrument on the outcomes of interest (average NO_x exposure, the black-white gap, and the Kolm-Pollak and Atkinson indexes).

FIGURE 21. Actual Gini Coefficient as a function of Simulated Gini Instrument, for 265 MSAs, 2005–2011

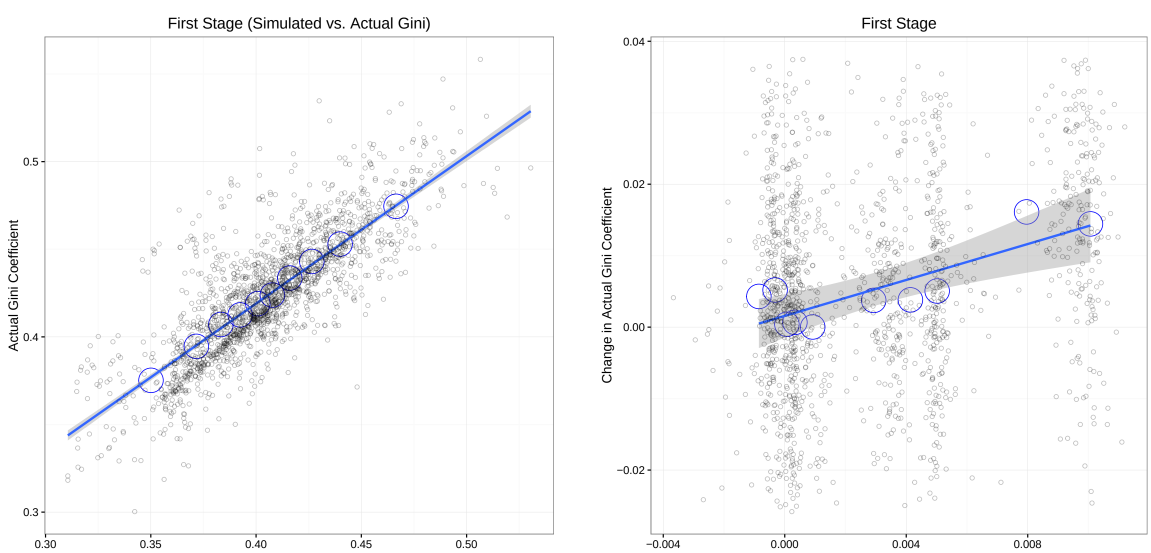
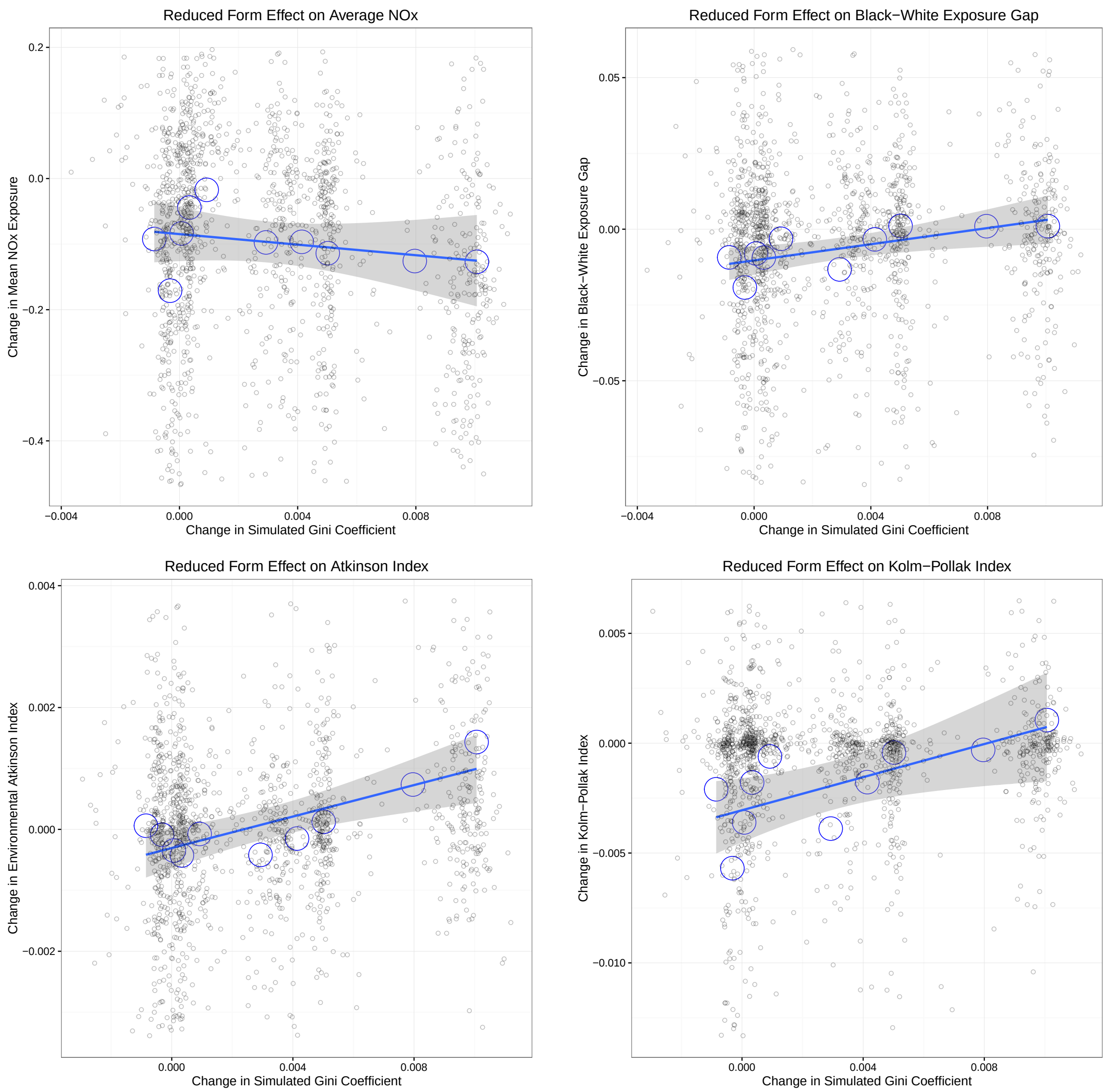


TABLE 17. First Stage, key coefficient only

	Dependent variable:	
	MSA-Level Gini	
	(1)	(2)
Simulated Gini	1.302***	1.600***
	(0.175)	(0.263)
MSA-specific Linear Trend?	No	Yes
F-stat	70.5	55.342
Observations	1,578	1,578
Note:		*p<0.1; **p<0.05; ***p<0.01
		Standard Errors allow for MSA clustering

Other control variables not shown: homeownership rates, % linguistic isolated, mean commute time, mean # of bedrooms in housing stock, mean household size, % latino, % black, % in school, % native, % asian % other race, % with a high school education , % with some college, % with a bachelor's degree, % with postgraduate education, % female, unemployment rate, % non-English speakers, Average Duncan Socioeconomic Index, % who commute by car, median age, median home value, % employed in 16 broad NAICS categories

FIGURE 22. Reduced Form Visualizations Showing the Effect of Simulated Income Inequality on Pollution Exposure



Results

I will consider three main sets of results. The previous literature has largely discussed the effect of income inequality on environmental degradation in terms of average exposure or emissions. Thus, to ensure comparability with previous results, I begin by examining the effect of MSA-level income inequality on *average exposure* to NO_x. Second, I consider whether increases in MSA-level income inequality affect measures of environmental *injustice* (which I define as the disparity in average exposure across subgroups). Third, I consider whether increases in income inequality affect measures of environmental *inequality*, which I define in terms of a functional of the distribution of pollution exposure. Examination of these last two relationships constitutes one of the main innovations in this research.

Table 18 reports the estimates of the key coefficient measuring the effect of income inequality on the average level of pollution exposure within a metropolitan area.[15] The top panel in the table shows results from an IV model using the simulated Gini coefficient described above as an instrument for actual income inequality. For comparison, the bottom panel shows results from a naive OLS regression of average NO_x exposure on the MSA-level Gini coefficient, ignoring potential endogeneity. The first column reports results for a model with no other covariates. The second column includes the full set of time-varying controls (see Table 17 for a full list). Finally, the third column reports results from a model with all time-varying controls and MSA-specific linear trends (equivalent to an MSA fixed effect in first differences). All models allow for MSA-level clustering in the standard errors.

[15]For this and other specifications, coefficient estimates for additional control variables are not presented. Full tables of all parameter estimates are available in supplemental material.

TABLE 18. Effect of Income Inequality on Average NO_x Exposure

		(1)	(2)	(3)
IV Results:	Gini	-3.772^{***}	-5.698^{***}	-6.405^{***}
		(1.346)	(1.681)	(2.111)
OLS Results:	Gini	-0.330	-0.503	-0.507
		(0.359)	(0.364)	(0.426)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

For specifications with and without MSA-specific linear trends the sign of the effect of income inequality on pollution exposure is negative, implying that rising income inequality *decreases* average exposure.[16] These estimated effects are statistically significantly different from zero, and quantitatively important. To put this in perspective, the average cumulative change in NO_x exposure across metropolitan areas from 2005-2011 is -0.582. According to the results in the third column of Table 18, a one Gini-point increase (i.e. a 0.01 increase in the Gini coefficient) in income inequality reduces average NO_x exposure by about 0.064 parts per billion, which is approximately 11% of the average cumulative change in NO_x concentration in our sample (from 2005 to 2011). A one-standard-deviation change in inequality (approximately 0.035) would decrease NO_x

[16]There is a relatively large difference in the estimated coefficients between a model with no control variables and a model with the full set of control variables. This difference appears to be driven by population density and total population.

concentrations by 0.22 ppb, which can account for approximately 38.5% of the average cumulative change in exposure over the study period.

These results are consistent with some, but not all, of the literature on the environmental effects of income inequality. Ravallion et al. (2000) and Scruggs (1998), among others, find that greater income inequality decreases the average level of emissions or pollutant concentrations. However, a recent study by Zwickl and Moser (2015), which utilizes variation within regions of the United States over time as I do here, finds that income inequality, when measured by the Gini coefficient, increases average pollution exposure. My results, when they differ from the previous literature, however, have two distinct advantages that tend to make them more arguably credible. First, unique in this literature, my results directly address potential time-varying endogeneity due to household locational sorting. Second, my data on pollution exposure, derived from satellite observations, uniformly covers areas which are not directly observed by EPA monitoring stations. Measuring pollution exposure from outer space thus allows for more-accurate measurement of the distribution of exposure.

Income inequality's effect on the pollution exposure distribution is unlikely to be fully summarized by changes in the first moment of the exposure distribution, however. To move beyond the mean, I first examine how increasing income inequality within metropolitan areas affects measures of "environmental injustice," which capture the differences in average exposure across sub-groups which are traditionally assumed to experience different degrees of advantage or disadvantage. Table 19 presents results from models using within-MSA differences in average exposure between African-Americans and whites as a dependent variable. The structure of Table 19 mimics Table 18, with increasing sets of controls from left to right (no controls in the first column, control variables in the second, and control variables and MSA-specific trends in the third).

Regardless of whether the models include MSA-specific trends, greater income inequality *increases* the difference in exposure between blacks and whites. This pattern is also broadly apparent when examining the effect of increasing income inequality on the Latino-white exposure differences in Table 20 and the rich-poor exposure gap in Table 21. The effects of increasing income inequality on black-white differences are larger in terms of effect size, and are statistically different from zero in almost all models, while the effect of inequality on the rich-poor exposure gap, and the Latino-white exposure gap are at best marginally significant.

TABLE 19. Effect of Income Inequality on Black-White Exposure Gap

		(1)	(2)	(3)
IV Results:	Gini	1.000***	1.091***	1.024**
		(0.320)	(0.382)	(0.455)
OLS Results:	Gini	0.096	0.064	0.067
		(0.069)	(0.075)	(0.089)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

As an alternative way of moving beyond the mean, I also consider whether rising income inequality affects the distribution of NO_x exposure as summarized by measures of environmental inequality. Table 22 reports the results of models using my preferred measure of absolute environmental inequality, the Kolm-Pollak index. I

TABLE 20. Effect of Income Inequality on Latino-White Exposure Gap

		(1)	(2)	(3)
IV Results:	Gini	0.595***	0.549**	0.684**
		(0.217)	(0.228)	(0.285)
OLS Results:	Gini	0.030	0.014	0.006
		(0.054)	(0.058)	(0.069)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

TABLE 21. Effect of Income Inequality on Poor-Rich Exposure Gap

		(1)	(2)	(3)
IV Results:	Gini	0.478***	0.462**	0.488**
		(0.183)	(0.194)	(0.239)
OLS Results:	Gini	0.048	0.027	0.033
		(0.038)	(0.041)	(0.052)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

select an environmental inequality aversion parameter $\kappa = 0.5$, which corresponds to relatively utilitarian social preferences. The Kolm-Pollak index is sensitive to the absolute differences in exposure between highly exposed and minimally exposed individuals, and is ethically similar to the environmental justice measures used previously in this paper. It is perhaps unsurprising that, consistent with the environmental justice results above, I find that greater income inequality increases absolute environmental inequality across all models.

TABLE 22. Effect of Income Inequality on Absolute Environmental Inequality

		(1)	(2)	(3)
IV Results:	Gini	0.283***	0.203**	0.150*
		(0.098)	(0.092)	(0.090)
OLS Results:	Gini	0.013	0.002	0.003
		(0.012)	(0.012)	(0.013)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

The choice of an absolute environmental inequality aversion parameter, κ, that corresponds to relatively utilitarian preferences is conservative in that it produces the smallest estimated effect size. To illustrate the influence of this parameter choice on the results, I allow for more Rawlsian preferences by allowing the absolute environmental inequality aversion parameter to vary from $\kappa = 0.5$ to $\kappa = 5$. The size of the effect of income inequality on the Kolm-Pollak index of absolute environmental inequality is monotonically increasing in κ over this range of κ.[17] These results are summarized in Figure 23. Intuitively, as social preferences become more Rawlsian (as $\kappa \rightarrow \infty$), these preferences are more sensitive to the the fate of the most exposed individual.

FIGURE 23. Illustration of how the estimated effect of Income Inequality on *Absolute* Environmental Inequality changes as the assumed value of κ, capturing Absolute Environmental Inequality aversion, increases

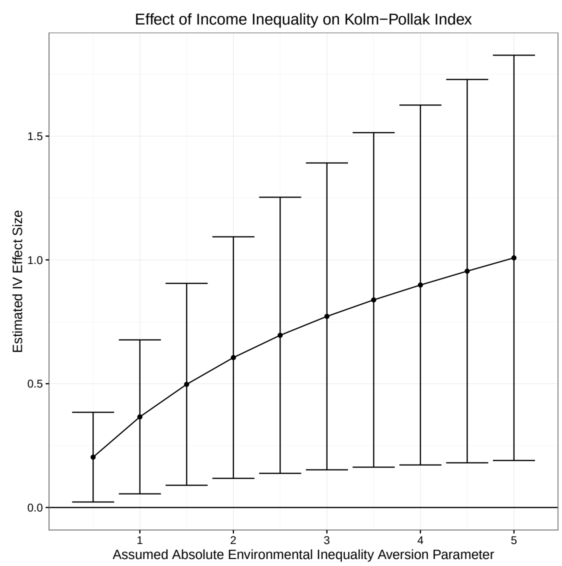


Table 23 reports results from models which use, alternatively, the Atkinson index of relative environmental inequality as a dependent variable. I find that increasing income inequality universally increases relative environmental inequality. This is unsurprising given the previous results on the effect of income inequality on the Kolm-Pollak index. Recall that absolute environmental inequality indexes are translation-invariant and relative inequality indexes are scale-invariant. Hence, if increasing income inequality reduces average exposure in a manner that increases the differences between highly and minimally exposed individuals, then the ratio between the highly and minimally exposed individuals

[17]This pattern continues for larger values of κ, although around $\kappa = 30$ the effect seems to level off.

will also increase (so the Atkinson index would be expected to increase). These results use a relative inequality aversion parameter of $\alpha = 0.5$, which again corresponds to relatively utilitarian social preferences. Paralleling the previous analysis, selecting larger values of α (more Rawlsian preferences) results in larger effect sizes, as shown in Figure 24.

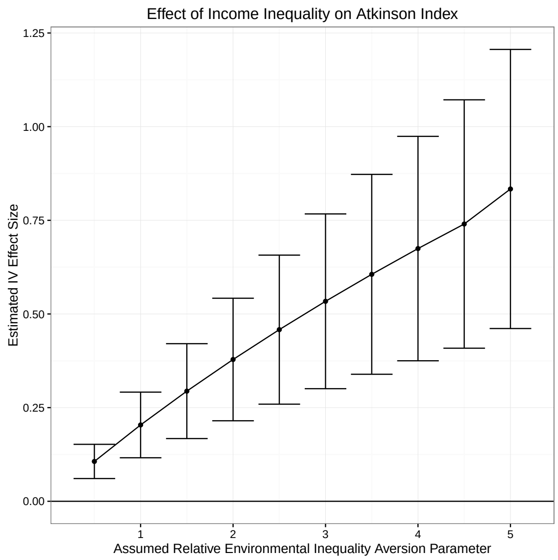
TABLE 23. Effect of Income Inequality on Relative Environmental Inequality

		(1)	(2)	(3)
IV Results:	Gini	0.101***	0.106***	0.100***
		(0.020)	(0.023)	(0.027)
OLS Results:	Gini	0.008**	0.007*	0.006
		(0.004)	(0.004)	(0.005)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

Income inequality has a complex effect on environmental quality: it decreases the average level of NO_x exposure within metropolitan areas, but it also causes this exposure to be more unequally distributed, measured either in terms of environmental justice (across race/class lines) or in terms of an environmental inequality index. This implies that the benefits of pollution reduction induced by greater income inequality accrue disproportionately to the most-advantaged. This does not necessarily mean, however, that income inequality makes the most-disadvantaged worse off in an absolute sense. Table 24 shows results of models using the within-MSA average exposure among African-Americans

FIGURE 24. Effect of Income Inequality on *Relative* Environmental Inequality, varying Relative Environmental Inequality Aversion



as a dependent variable. Increased income inequality decreases pollution exposure among African-Americans, across different specifications and for different sub-samples. Similarly, Table 25 summarizes results for models using average white exposure as a dependent variable. Rising income inequality decreases white exposure, and the estimated effect is larger than is the effect on average black exposure.

Figure 25 illustrates how a five Gini-point increase in income inequality (roughly the size of the observed change in the aggregate national Gini coefficient between 1990 and 2010) would affect a hypothetical distribution of NO_x exposure.[18] I take as given the estimated effect sizes in the second columns of Tables 24 and 25 as the race-specific marginal effects of income inequality on pollution exposure. I assume for simplicity that African-Americans and whites each experience a decrease in pollution exposure equal to the race-specific marginal effect multiplied by a five Gini-point change in inequality. The overall result is depicted by the distributional changes between the top and bottom panels

[18]The hypothetical pollution distributions are generated by simulating distributions from a Generalized Beta distribution fitted to the aggregate national data in 2005 for whites and blacks separately.

TABLE 24. Effect of Income Inequality on Average Black Exposure

		(1)	(2)	(3)
IV Results:	Gini	-3.311^{**}	-4.951^{***}	-5.680^{***}
		(1.436)	(1.759)	(2.190)
OLS Results:	Gini	-0.243	-0.427	-0.423
		(0.384)	(0.387)	(0.452)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

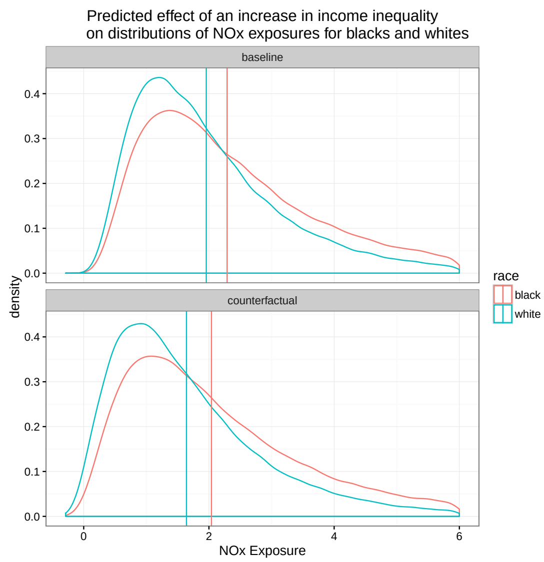
TABLE 25. Effect of Income Inequality on Average White Exposure

		(1)	(2)	(3)
IV Results:	Gini	-4.311^{***}	-6.041^{***}	-6.704^{***}
		(1.328)	(1.671)	(2.109)
OLS Results:	Gini	-0.339	-0.491	-0.490
		(0.355)	(0.360)	(0.422)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

of Figure 25. An increase in income inequality decreases the average exposure of both groups in absolute terms, but increases the difference in average exposures across the two groups.

FIGURE 25. Effect of an increase in income inequality on NO_x exposure.



This illustrates the predicted effect of a 5 Gini-point increase in inequality (approximately the size of the increase from 1990 to 2010 nationally) on the distribution of NO_x exposure among blacks and whites. I observe (1) average exposure decreases for both groups and (2) the gap between the two groups increases in the counterfactual higher income inequality scenario. Baseline scenario is the smoothed national distribution of NO_x exposure in 2005, counterfactual assumes that the baseline distributions are "shifted" by the marginal effect (column 2 in tables 24 and 25) for a 5 Gini-point change in inequality.

Mechanisms

Across a variety of model specifications, and using a number of different measurement tools for summarizing the NO_x exposure distribution in each of 265 different MSAs, I have shown that increasing income inequality tends to reduce the average level of pollution exposure, but also increases exposure inequality and the differences in exposure between advantaged and disadvantaged subgroups. My findings are inferred from reduced-form models, so they cannot directly address the underlying mechanisms by which these results might occur (i.e the pathways through which changes in the income distribution might affect the pollution exposure distribution). I can, nonetheless, appeal to some other studies to suggest potential mechanisms for the effects I observe. There are several potential pathways from income inequality to environmental quality, including (1) differential demands for clean air across the income distribution, (2) residential or industrial sorting *within* MSAs, and (3) the political process working through legislator ideology or legislator voting behavior.

As noted, the Environmental Kuznets curve literature provides some insight as to how increasing income inequality might affect the level of pollution exposure. One can think of this in two ways. First, suppose that the NO_x emission intensity of a household's consumption bundle can be summarized by a "propensity to emit" function. Over the range of incomes in the United States, this function is assumed to be non-decreasing in income. If the PEF is concave in income, so that the marginal propensity to emit is decreasing with income, then an immediate conclusion is that regressive income transfers should decrease emissions. The marginal decrease in emissions by a poor household at one end of the regressive transfer is larger in absolute terms than the marginal increase in emissions by the rich household at the other end of the transfer. Although there is not much direct evidence on the concavity of the PEF, Levinson and O'Brien (2015) use US expenditure data to show that the closely related Environmental Engel curve—the pollution embedded in consumption along a household's income expansion path—is concave in household income.

A second potential pathway from increased income inequality to decreased average pollution exposure (and also increased environmental inequality) might occur through residential sorting within MSAs.[19] If increases in income inequality are associated with differential rates of migration towards relatively more polluted areas by disadvantaged groups, perhaps due to property market dynamics, this might increase my measures of environmental inequality, despite decreasing pollution exposure on average. Matlack and Vigdor (2008) show, for example, that increases in income inequality within metropolitan areas lead to higher rents and home prices at the bottom of the home price distribution. This effect may induce poorer and more-disadvantaged households to "come to the nuisance" in search of cheaper housing (see also Banzhaf (2012)).

A final mechanism for the relationship between income inequality and environmental quality might work through the political system. This mechanism, however, is specific to the institutional arrangement of US politics, and may not generalize to other countries. Within the US political system, the party of the left (the Democrats) is generally considered to also be the party of the environment. In other research, Voorheis et al. (2015) investigate the effect of within-state income inequality on the average ideological position of political parties within state legislatures. Perhaps the most striking result in Voorheis et al. (2015) is that increasing income inequality moves state Democratic parties further to the left.

Another small but growing literature in political science has suggested that income inequality may be connected to what is termed "unequal democracy". As an exemplar, Bartels (2009) shows that US Senators are more responsive to the political opinions of rich constituents than poor constituents. This suggests a potential mechanism for the relationship between income inequality and pollution exposure observed in the present study. This political mechanism can be summarized as a two-stage process. First, if

[19]my identification strategy addresses potential reverse causality from pollution exposure to income inequality due to between-MSA sorting. Within-MSA residential sorting poses no such endogeneity problems, since the Gini coefficient satisfies the property of anonymity. Within-MSA locational sorting will not affect the Gini coefficient for that MSA.

income inequality increases the responsiveness of legislators to rich constituents, and demand for clean air is increasing in income, then this may induce legislators to support more environmentally friendly legislation. Second, increased likelihood of enacting or enforcing environmental regulation, would be expected to decrease average pollution exposure.

Inequality and LCV Scores

I can directly examine the first stage of the proposed political mechanism in the last section by modeling how increasing income inequality affects the environmental voting record of elected legislators. However, this makes it necessary to change the geographic scale of analysis to account for the specifics of the US political system. Most environmental regulation occurs at the Federal level, making members of the US House and Senate the obvious legislators to analyze. For reasons of data availability, I will further restrict my attention to just the US Senate.[20]

To measure the environmental voting record of US senators, I will use the League of Conservation Voters' *National Environmental Scorecard*.[21] The LCV has published the scorecard annually since 1970. The scorecard gives each federal legislator (senator or member of the House of Representatives) a score between 0 and 100, based on their votes on a list of bills selected by the LCV as being important for the environment and natural resources. A score of 100 represents a voting record in a given year where a legislator voted in agreement with the LCV on all bills, while a score of 0 represents a voting record where a legislator always voted against the LCV's policy position. The LCV

[20] Income inequality data for US House districts is available only in the decennial Censuses or in the ACS 5-year files, while Chapter II of this dissertation produces State-level inequality measures annually for the period 1977-2014.

[21] The nominal, unadjusted LCV scorecard data can be accessed via the LCV website, at: `http://scorecard.lcv.org/`

gives a separate score in each year a senator serves, so there can be non-trivial within-senator variation that can be exploited.[22] Since there are different sets of bills that are voted on in each year, I follow the method proposed by Groseclose et al. (1999) for scaling nominal LCV scores to ensure comparability over time. These scores are correlated with the ideology of legislators. The most commonly used data on the ideology of senators are the DW-NOMINATE [23] scores, which place legislators on a scale from -1 (most left-wing) to 1 (most right-wing). Figure 26 summarizes the correlations between LCV scores and DW-NOMINATE scores.

As expected, there is party polarization along the environmentalism dimension. To see this further, Figure 27 plots the histogram of LCV scores by party for each Senator serving between 1977-2014. Although the Republican party is substantially more unified on environmental issues than is the Democratic party, as can been seen by the pronounced peak in Republican LCV scores near zero, there is still substantial heterogeneity within both parties.

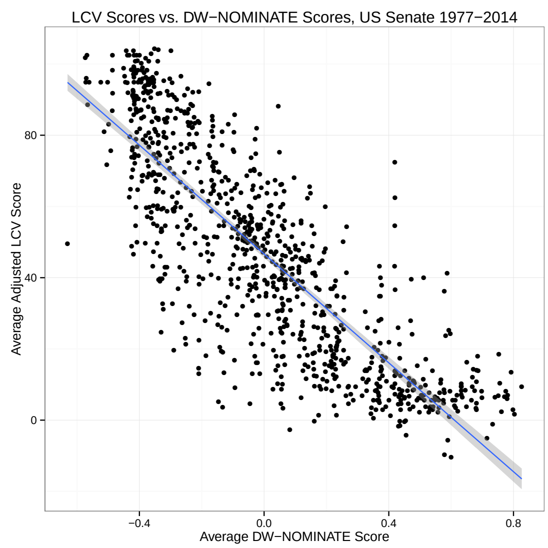
I will analyze the effect of state-level income inequality on these adjusted LCV scores for all senators who served during the period 1977-2014. State-level income inequality data is available over the same time period from Chapter II of this dissertation. Unlike the MSA-level data used above, these state-level inequality measures are calculated using the Current Population Survey, and address Census Bureau censoring and potential under-reporting by modeling the right tail of the income distribution as following a Generalized Beta II distribution. I supplement these two variables of interest (State-level income inequality and LCV scores) with state-level demographic and economic information from

[22]This within-legislator variation is in contrast to ideal-point estimates of legislator ideology which assume that legislators do not change positions over time.

[23]Dynamically-Weighted Nominal Three-step Estimation

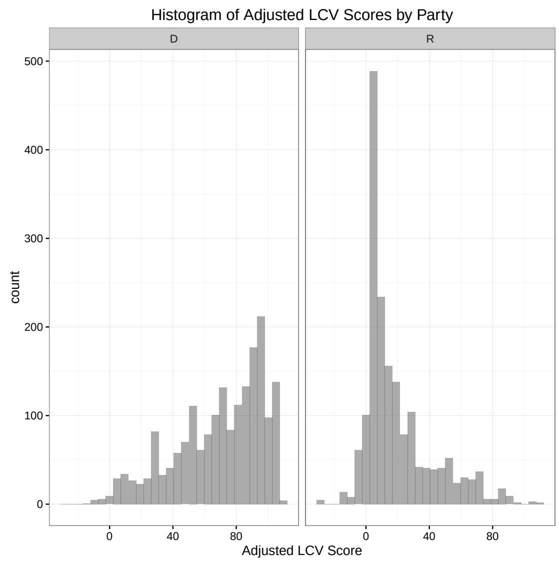
FIGURE 26. Correlation Between League of Conservation Voter Scores and Ideology, US senators

(a)



Nominal League of Conservation Voter (LCV) scores represent the the percentage of votes cast on environmental legislation by a senator that agree with the LCV position. The adjusted LCV Scores shown here rescale each year's Nominal scores to make them comparable over time, and hence can be less than 0 or greater than 100. DW-NOMINATE scores represent the "ideal point" of a senator on a latent ideology scale ranging from -1 (most liberal) to 1 (most conservative)

FIGURE 27. Histogram of Adjusted LCV Scores by Party for US Senators, 1977-2014



the Current Population Survey and the Bureau of Economic Analysis' national income accounts.

I assume that each senator is potentially responsive to changes in income inequality within the state that he or she represents. As with the previous analysis of the effect of MSA-level inequality on environmental quality, and similar to Voorheis et al. (2015), there is potential endogeneity between income inequality and the outcome of interest (in this case, the environmental voting record of senators). In addition to the possibility of locational sorting by a senator's constituents, the examination of the effect of inequality is potentially complicated by an additional source of endogeneity bias stemming from the distributional consequences of environmental policy. To address these potential sources of endogeneity bias, I implement an instrumental variables identification strategy similar to that used in the earlier sections of this paper.

As before, I require an instrument for state-level income inequality that is uncorrelated with changes in a senator's voting record, but is correlated with the actual state-level income inequality experienced by that senator. I propose a version of the simulated inequality instrument used in the previous analysis. In the state-level setting, I construct the instrument by freezing each state's income distribution in 1976 (the first year in which the state-level income inequality data are available), and simulating future counterfactual state income distributions based on nationwide trends in decile income growth. This is instrument is constructed identically to the instrument used in Voorheis et al. (2015), with the sole difference being an earlier starting year.

I estimate the effect of income inequality on a senator's LCV score using a two-stage least squares model in first differences:

$$\text{First Stage:} \quad \Delta Gini_{i,t} = \delta_i + \gamma \Delta SynthGini_{i,t} + \theta \Delta X_{i,t} + \nu_{i,t} \tag{4.13}$$

$$\text{Second Stage:} \quad \Delta Score_{i,t} = \alpha_i + \beta \Delta \widehat{Gini}_{i,t} + \Gamma \Delta X_{i,t} + \epsilon_{i,t} \tag{4.14}$$

where i now indexes senators, t indexes time in years, α_i and δ_i are senator-specific linear trends, and $X_{i,t}$ is a vector of time varying controls, including the proportion of the population that is black or Hispanic, the median age, median household income, poverty rates and educational attainment in the state each senator represents.

Table 26 summarizes the basic results from this estimation. The first two columns report the full-sample effect of income inequality on senators' LCV scores with and without senator-specific linear trends, respectively. The final columns report the effects (again with and then without senator-specific linear trends) for just the sub-sample of senators serving after 2004. This latter subsample corresponds to the time period in the previous MSA-level results describing the effect of income inequality on the distribution of NO_x exposure. Across all specifications and all subsets of the data considered, however, I find that rising income inequality seems to have a positive effect on a senator's environmental voting record. Using results with the senator-specific trends, a one-Gini-point increase in income inequality would increase a senator's adjusted LCV score by 1.6 points (i.e. this greater degree of income inequality would increase the proportion of times a senator agrees with the LCV by 1.6 percentage points).

TABLE 26. Effect of State Inequality on Senators' LCV Scores

	Adjusted LCV Score			
	1977–2014		2005–2014	
	(1)	(2)	(3)	(4)
Gini	163.324***	160.154**	281.250*	278.416
	(60.356)	(64.949)	(158.769)	(193.750)
Control Variables?	Yes	Yes	Yes	Yes
Senator-specific Trend?	No	Yes	No	Yes
N	3,356	3,356	742	742

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
See Table 1 for other variables not shown here

Politics within the US Senate has become increasingly one-dimensional over the period of the estimating sample—over 80% of roll votes can be correctly classified based only on a senator's DW-NOMINATE score, which captures conservative-liberal differences (see McCarty et al. (1997)). As politics has become more polarized, it is likely that the effect of income inequality on senators' environmental voting records might also be polarized along party lines. Table 27 summarizes the effect of income inequality on the environmental voting records of senators, stratified by political party.[24] The first two columns report results for the subsample of Democratic senators (with and without

[24]There were 5 senators without formal party affiliation who were elected during this period. I exclude these senators from the party-stratification analysis.

senator-specific trends), while the final two columns report results for Republican senators. The effect of income inequality on the environmental voting record of Democratic senators is statistically significant, positive, and larger than the full sample results. On the other hand, the effect on the environmental voting record of Republican senators is negative, substantially smaller in absolute value, and very imprecisely estimated.

TABLE 27. Effect of State Inequality on Senators' LCV Scores, By Party

	Adjusted LCV Score			
	Democrats, 1977–2014		Republicans, 1977–2014	
	(1)	(2)	(3)	(4)
Gini	499.237***	495.279***	−105.548*	−102.157
	(138.356)	(149.892)	(56.372)	(62.321)
Control Variables?	Yes	Yes	Yes	Yes
Senator-specific Trend?	No	Yes	No	Yes
N	1,708	1,708	1,597	1,597

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
See Table 1 for other variables not shown here

These results imply that increasing income inequality improves the environmental voting record of Democratic senators, but has no discernible effect on Republican senators. If demand for environmental quality is increasing in income, this is consistent with the concept of "unequal democracy", wherein the positions of elected officials are more closely aligned with the preferences of the rich than the poor or middle class. In this case, an increase in income inequality, and hence an increase in the local bargaining power of the affluent, moves Democratic senators towards the typical position of the affluent in terms of more environmentally friendly legislation. The differential effect of inequality on

Democratic and Republican senators is likely the result of differing underlying levels of environmentalism, or a different income-environmentalism gradient among party elites.

This suggests that the political system may be a plausible pathway through which increases in income inequality affect the distribution of environmental amenities. Note that the previous result that effects of income inequality on the distribution of NO_x exposure were stronger for the subsample when Democrats controlled federal institutions is consistent with the Democratic-party-specific political mechanism implied in these state-level results. To reiterate, however, these results are context-specific, and may not describe how changes in income inequality affect environmental policy at other levels of government (e.g. the restrictiveness of state regulations or local land use ordinances).

Conclusion

Viewing environmental justice from outer space—using satellite remote-sensing data to infer the distribution of ground-level exposure to harmful pollutants like NO_x—provides a viable way forward for the measurement and analysis of environmental justice and inequality. Leveraging these remotely sensed data to study environmental justice and environmental inequality is one important contribution of this study. The near-universal coverage and fine geographic resolution of the satellite data used in this study are far superior to the incomplete coverage (and the endogenous placement) of ground-monitor data so often used in previous studies.

I have shown that there is consistent and robust evidence that rising income inequality decreases the level of average NO_x exposure within metropolitan areas. Increasing income inequality also appears to increase the differences in exposure between advantaged and disadvantaged groups, and increases measures of environmental inequality. This pattern implies that the benefits of pollution reduction tend to disproportionately accrue to the most advantaged. However, income inequality still decreases the absolute level of exposure among the most disadvantaged—either in terms of class, race or degree of exposure. These results hold for a variety of specifications, and for various normative assumptions underlying the empirical measurement of environmental inequality (such as the assumed degree of environmental inequality aversion in the social welfare function).

I offer some evidence that the effect of increasing income inequality on the distribution of environmental amenities may work through the political system. Using a state-level simulated instrument strategy similar to that employed elsewhere in this paper for MSAs, I use state-level data to identify the causal effect of income inequality on the environmental voting record of US senators. I find that income inequality increases the LCV scores of senators, and that this effect appears to be concentrated among Democratic senators.

NO_x exposure is a major public health concern in the United States and Europe, contributing to thousands of premature deaths. The fact documented here—that rising income inequality appears to mitigate this exposure to a measurable extent—suggests that there may be unanticipated local effects of inequality that have not previously been recognized. These reductions in exposure are not equally shared, because greater income inequality increases the difference in exposure between advantaged and disadvantaged groups. These results add nuance to the discussion of the effects of increasing inequality in the United States. One particularly important implication of this research is that the effect of inequality on the distribution of pollution exposure may lead to a further propagation of inequality into the future, due to the widening gap in exposure between children born to rich and poor parents.

There is considerable further work to be done on this topic. In particular, I have shown an effect only for one pollutant. Using satellite data and/or monitoring data to examine the relationship for other important pollutants, including particulate matter and ozone, may be instructive. It will also be relevant to examine the relationship between inequality and pollution exposure for other countries and regions, including the EU, India and China. Further investigating the possible mechanisms behind the observed reduced-form relationship between income inequality and the environment is another important direction for research.

CHAPTER V

CONCLUSION

In the three substantive chapters of this dissertation, I have shown how rising income inequality may be an important driver of improvements in environmental quality and, more sinisterly, inequality in the distribution of environmental disamenities. These results are highly relevant in the current policy environment, in which rising income inequality and the environmental challenges of climate change and the health impacts of air pollution are salient topics. These chapters add to several important literatures across sub-fields within economics, and introduce new data, modelling tools and extend identification and inference strategies in ways that will benefit future researchers.

Chapter II adds to the stock of knowledge about one of the most pressing concerns of the current time: the growth in income inequality. The new datasets of income inequality measures at the state and metropolitan area level represent an extension of the literature on using survey data to measure income inequality, specifically by extending the method of using the Generalized Beta II distribution to address topcoding and under-reporting first proposed by Burkhauser et al. (2011). Using this data, I also contribute to understanding the determinants of rising inequality, identifying declining unionization and reductions in top marginal tax rates as the most important among potential explanations. I find less compelling evidence for explanations relying on changes in human capital or skill-biased technological change. This bolsters claims by, among others, DiNardo et al. (1996) and Piketty et al. (2014) that the recent rise in inequality may be the result of policy, such as changes in tax rates and legislation designed to reduce the bargaining power of unions.

Chapter III provides two main contributions to the literature on environmental inequality and environmental justice. First, I leverage satellite-derived remote sensing data to measure the distribution of exposure to two important pollutants. Although the use of satellite data to measure pollution exposure is common in other fields, it has not yet become common in the environmental economics literature (this chapter, and Chapter IV represent the first use of these data in economics to my knowledge). Due to the fact that satellite data provide substantially better spatial coverage than data derived from ground monitoring stations, and provide a better picture of actual exposure than conventional air quality models, since the satellite data captures ground-level concentrations deriving from all sources, whereas most air quality models only capture exposure to pollution from stationary emissions sources. In addition to bringing this potentially quite valuable data source into the economics literature, I also contribute to the literature on measuring environmental inequality. I propose a dashboard approach to measuring environmental inequality, combining the vertical equity approach of Sheriff and Maguire (2014) with the horizontal equity approach of the environmental justice literature (e.g. Mohai et al. (2009)). Using the satellite data to calculate the various environmental inequality measures that compose this "dashboard" for the entire United States annually over the period 1998-2014.

Chapter IV combines the datasets introduced in the Chapters II and III to examine whether rising income inequality within metropolitan areas might be related to changes in the distribution of pollution exposure. Identifying the causal effect of inequality on an outcome such as pollution requires care, as locational sorting may bias naive OLS estimates which do not account for reverse causality. I extend an approach to causal inference on the effects of income inequality first proposed in Boustan et al. (2013), in which I simulate an instrument for MSA-level income inequality using national level trends in income growth at deciles of the income distribution. This approach is an important tool in the small but growing literature on the causal effects of rising income inequality on other outcomes of interest. Using the new datasets produced in the previous chapters with the causal identification approach presents a way forward for the literature on the environmental effects of income inequality, which, since Boyce (1994) has been plagued by insufficient or incomplete data and a lack of clean identification. I also examine a potential political economy explanation for the environmental effects of inequality. This adds to the political economy literature on the connection between income inequality and the political process (e.g. McCarty et al. (1997), Bartels (2009)), and connects this political economy literature to the literature of environmental inequality and the environmental effects of income inequality.

The results presented in this dissertation represent the core of a broader research agenda examining interrelations among income inequality, the environment and the political process. One important project running parallel to this is dissertation is Voorheis et al. (2015), an examination of the effect of income inequality on political polarization in US State legislatures. Using the data from Chapter II and a similar identification strategy to that used in chapter IV, we find that income inequality increases political polarization by inducing the replacement of moderate Democratic legislators with more conservative Republican legislators. We provide suggestive evidence that campaign contributions may serve as a potential mechanism for this effect, by allowing the effect of income inequality on polarization to vary systematically in state-years in which pre-Citizens United caps on independent campaign expenditures were in effect. I further examine the effect of income inequality on the environment in Voorheis (2016), which presents evidence that income inequality might lead to a reduction in carbon emissions, a result consistent with Chapter IV.

APPENDIX A

MEASURING STATE INCOME INEQUALITY BY COMBINING CPS AND IRS DATA

Introduction

We have income information from two sources: public use microdata from the Current Population Survey, and data on number of tax returns and adjusted gross income for income bands. The first substantive chapter of this dissertation uses the former to estimates measures of state-level income inequality, while Frank (2009) and uses the latter. Each data source has drawbacks. The CPS data may incorrectly capture top incomes for two reasons: (1) top earners may systematically under-report (or differentially non-respond), and (2) the Census Bureau censors incomes above a certain threshold ("topcoding"). Both of these will bias estimates of income inequality downwards. The IRS data, on the other hand, contains no information on the income of non-filers. Estimates of inequality from IRS data will likely overstate the level of inequality, especially when inequality is captured by top income shares.

One obvious way to improve on estimates of inequality using each of the above data sources individually is to combine information on top incomes from tax return data with information on the rest of the income distribution using survey data. Atkinson (2007) and Alvaredo (2011) suggest an approach for the Gini coefficient based on the approximation that for top income share S, the Gini coefficient is approximately $G^* (1 - S) + S$, where G^* is the Gini coefficient estimated for the non-top incomes. Diaz-Bazan (2015) shows that is possible to estimate the population Gini coefficient by combining estimates of the Gini coefficient computed using conditional distributions estimated from survey and tax data.

We propose an alternative, but similar, approach to this problem which improves on the previous methods for combining survey and tax data in two dimensions. This approach relies on simulating top incomes rather than relying on an approximation based on asymptotic behavior. This means that in practice this method can be used to estimate any income inequality measure, not just the Gini coefficient. Additionally, as in Flaichaire and Davidson (2007), it is possible to perform inference via semi-parametric bootstrapping and calculate point estimates of inequality measures simultaneously.

I proceed as follows: I describe the data in detail and the steps necessary to make income information in the CPS comparable to income information in the IRS data. I describe the proposed method for combining information from the two series, as well as an approach to selecting the optimal cutoff percentile. I then present estimates of inequality using this method for US states, and compare these estimates to estimates calculated using an approach using only information from CPS data.

Methodology

In order to combine the information from the two data sources, I need to first accomplish three tasks. First, I need to ensure that the income concepts in the two data sources are identical to ensure comparability. Second, I must identify the point at which top incomes may be censored or underreported in the survey data. Finally, I need to translate information on top incomes from the tax return data back to the survey data in order to estimate income inequality.

Microdata from the March Current Population survey provides information on income from a variety of sources for each member of a household. These sources include earnings (wages, small business income, farm income), unearned income (rent, interest, dividends) as well as taxable and non-taxable transfer income (unemployment insurance, social security and AFDC/TANF). The summary data available from the IRS reports "Adjusted Gross Income" (broad income), which is the sum of all taxable gross income less a set of pre-defined allowances.

In order to incorporate information from both sources, it is necessary to transform the CPS income data to conform with the IRS AGI definition. I do this in two steps. I first form "tax units" from the CPS households as follows. I first identify all dependents in each year's sample (all children under 18 and all students currently in school under the age of 25). I then assign these children to the head of household (or to their parent if their parent is not the head). I then define all non-dependents in a household as either married, single of head of household filers depending on their marital status and whether they have children. Finally, I define all dependents with total personal income above a threshold ($3000) as dependent taxpayers. For each of the tax units I form, I simulate AGI using NBER's TAXSIM, using the income information in the CPS. Since there is no information on essentially all "above-the-line" deductions in the CPS, this may overestimate AGI.

Next, I determine the point past which I believe CPS income data is either censored (topcoded) or under-reported; this will determine the cutpoint at which I will use IRS information. Heuristically, I search for a point at which the CPS distribution and the IRS distribution "match up". For each percentile $p \in \{0.9, 0.901, 0.902, ..., 0.999\}$, I calculate the threshold income at this percentile in the CPS AGI distribution, and estimate the percentile p_P at which the CPS threshold income falls in the IRS AGI distribution using Pareto interpolation. I choose as the cutpoint past which I will use IRS information only by minimizing $|p - p_P|$.

Finally, I combine the information from the survey and tax return data to estimate measures of inequality by simulating incomes above the optimally chosen percentile p. For each state and year, I estimate the Pareto parameter α using IRS data, again via Pareto interpolation. I then simulate n partially synthetic income distributions. For each replication, I combine all the CPS tax units below the cutoff p, and replace all tax units above p with random draws from a Pareto distribution with shape parameter α and location parameter equal to the threshold income at p. I calculate income inequality measures $\nu(F_i)$ for each partially synthetic distribution. The resulting estimate of inequality, considering information from both the survey and tax return data is $\frac{1}{n}\sum_{i=1}^{n}\nu(F_i)$

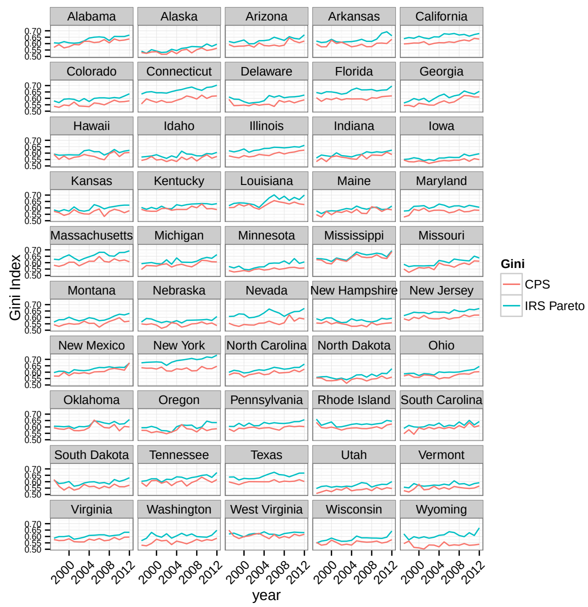
Data and Results

I will compare estimates of two measures of income inequality (the Gini coefficient and the top 1%'s share of income) using the IRS Pareto imputation against two baselines. I first compare inequality measures estimated via the IRS Pareto imputation with estimates calculated using just the CPS data with no corrections for topcoding, and second with the estimates from Frank (2009), which use just the IRS data. The readily available IRS SOI summary data used in the Pareto imputation spans 1997-2012, a relatively short period. Data for years before 1997 was published in the SOI bulletin, although it is not made freely available in machine readable form by the IRS.

Figure 28 shows the estimates of the Gini coefficient estimated with the IRS Pareto simulation method (in blue) and with just CPS microdata (with no correction for topcoding, in red). For all states, the Gini coefficient estimated via the Pareto simulation process is higher than the Gini estimated using just CPS data. There is a wide degree of heterogeneity in the degree to which the Pareto simulation estimates diverge from the CPS-only estimates, however. Unsurprisingly, states which are relatively richer on average (e.g. Connecticut, New York) exhibit the widest differential between the two estimates, while states which are poorer on average (e.g. Mississippi) exhibit only small differences. For many (though not all) states, the trends in the Gini coefficient are largely similar for the two measures.

Figure 29 shows the estimates of inequality measured by the top 1%'s share of income, again comparing estimates using the Pareto simulation to estimates using only CPS data. As with the Gini coefficient, using the Pareto simulation method substantially

FIGURE 28. State Gini Coefficients, 1997-2012: IRS Simulation vs. CPS Baseline

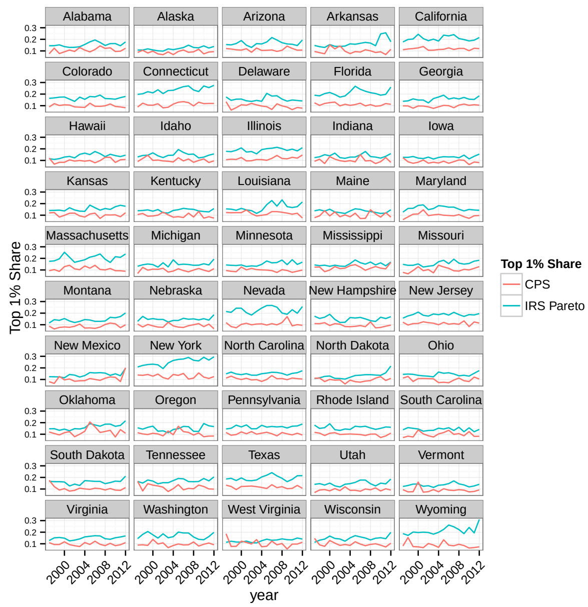


increases estimates of the top 1% share relative to the CPS-only baseline. Unlike the Gini coefficient case, however, the Pareto simulation estimates deviate from the CPS baseline not just in the level of inequality, but also in trends over time for many states. This is especially true in the last few years of the sample, where most states exhibit increasing top 1% shares from 2009-2012 using the Pareto simulation method, while top 1% shares are flat or falling in the CPS baseline.

Figures 30 and 31 compare IRS simulation method estimates of the Gini coefficient and top 1% share, respectively with estimates from Frank (2009), which are estimated using only IRS data.[1] The Pareto simulation estimates of the top 1% share are much closer to the Frank estimates than they are to the CPS baseline. The largest deviations occur between 2004-2008 for most states, where the Frank estimates of the top 1%

[1]This dataset has been updated through 2013, and is available at `http://www.shsu.edu/eco_mwf/inequality.html`

FIGURE 29. State Top 1% Shares, 1997-2012: IRS Simulation vs. CPS Baseline



share are much higher than the IRS simulation results. The Frank estimates of the Gini coefficient deviate much more considerably from the Pareto simulation results, however, both in level and trends. Overall, as expected, the IRS simulation method seems to produce estimates that are a compromise between the IRS and CPS estimates. It is notable that the IRS simulation method can roughly match the level and trends from the updated State inequality dataset from Frank (2009).

The IRS simulation method is only feasible for circumstances where the income definitions in survey data can be aligned to the IRS definition, and where summary data on tax returns and aggregate AGI by income group is available. In cases where the desired income concept is not tax unit AGI, such as equivalized household income, or a broad income concept that includes in-kind transfers, it is not clear how to integrate information on the tail of the AGI distribution into an estimate of inequality in another distribution. Nonetheless, concerns about under-reporting and censoring in the survey data are still

FIGURE 30. State Gini Coefficient, 1997-2012: IRS Simulation vs. Frank (2009)

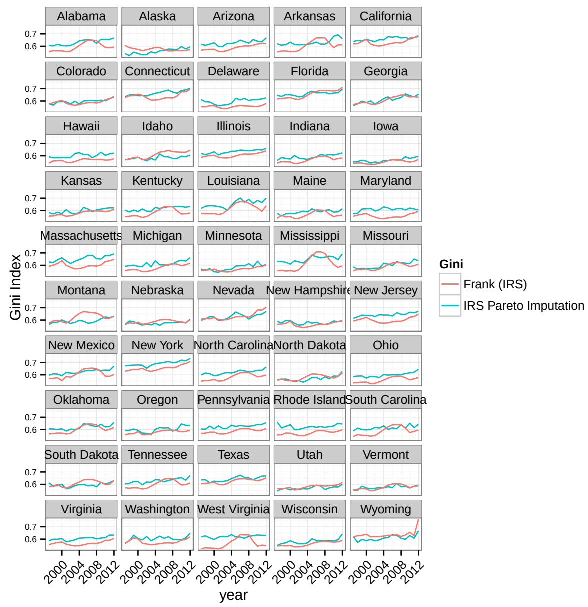
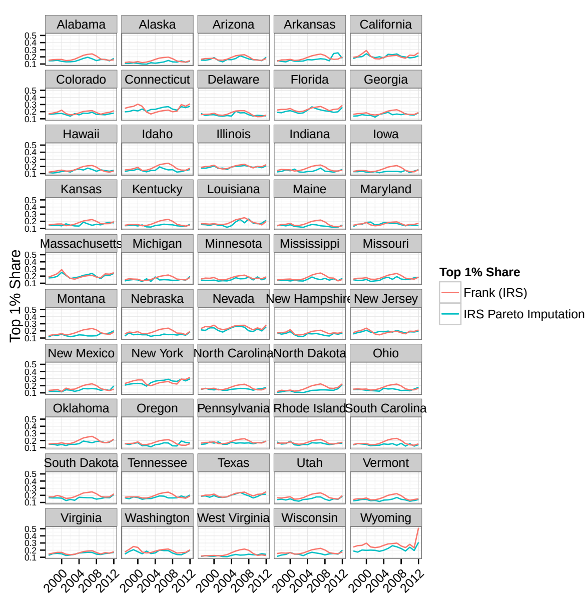


FIGURE 31. State Top 1% Shares, 1997-2012: IRS Simulation vs. Frank (2009)

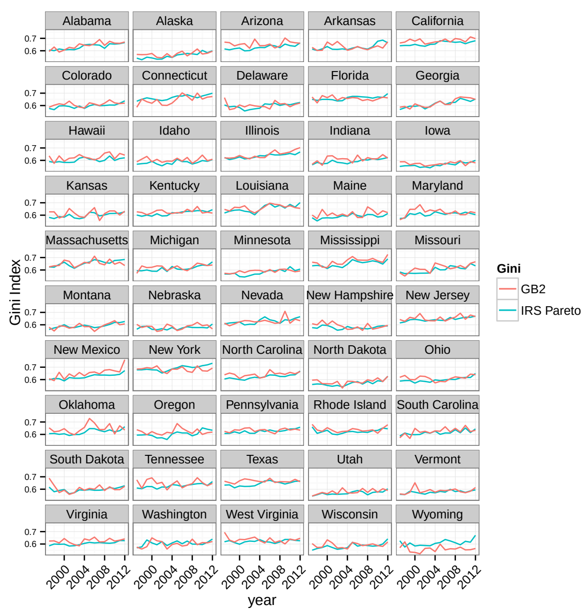


salient in these cases. One way to address these concerns is to use a semi-parametric method similar to the Pareto simulation method described above using only survey data. I will use a specific version of this survey data-only approach to estimate state inequality and compare it to the Pareto simulation results produced above to judge the quality of the survey-data only approach.

One specific way to address potentially downward biased top incomes in the survey data is to use a multiple imputation approach. In this approach, a Generalized Beta II distribution is fitted to the national income distribution from the CPS microdata in each year. A cutoff is selected, past which incomes are believed to be censored or underreported. To maintain comparability with the Pareto simulation estimates above, I will use the cutoffs identified above using the "lining up" method. The two data sources line up, on average, at about the 97.5th percentile. This, and the fact that it is only possible to use the "lining up method" for AGI income, supports the use of the 97.5th percentile as a cutoff for non-AGI income definitions (as in Chapter II of this dissertation). Then, as in the Pareto simulation method above, N partially synthetic datasets are formed; each partially synthetic dataset is formed by concatenating all incomes below the cutoff with n_1 draws from the fitted GB2 distribution, where n_1 is the number of incomes above the cutoff. Inequality measures $\nu(F_i)$ are calculated from each partially synthetic dataset, and the resulting estimate of inequality is again $\hat{\nu} = \frac{1}{N}\sum_{i=1}^{N}\nu(F_i)$.

Figure 32 compares the Pareto simulation estimates of the Gini coefficient with estimates produced using the GB2 approach using only CPS data. Remarkably, the GB2 simulation method produces estimates that very closely matches the level and trends in inequality estimated using the Pareto simulation approach, but it does so using only information available in the CPS. The one notable exception appears to be Wyoming, for which the GB2 method spectacularly fails. This is likely due to the combination of low population and extreme inequality in the right tail of the income distribution that is unique to Wyoming.[2] Figure 33 compares Pareto simulation estimates of the top 1% share to the GB2 simulation estimates. Here the GB2 estimates are not quite as impressively matched. For many states, the GB2 estimates are relatively close to the Pareto simulation estimates, although not nearly as close as the Gini estimates are. Nonetheless, the GB2 estimates are a substantially improvement on the baseline CPS estimates with no correction for topcoding or underreporting.

FIGURE 32. State Gini Coefficient, 1997-2012: IRS Simulation vs. GB2 Simulation

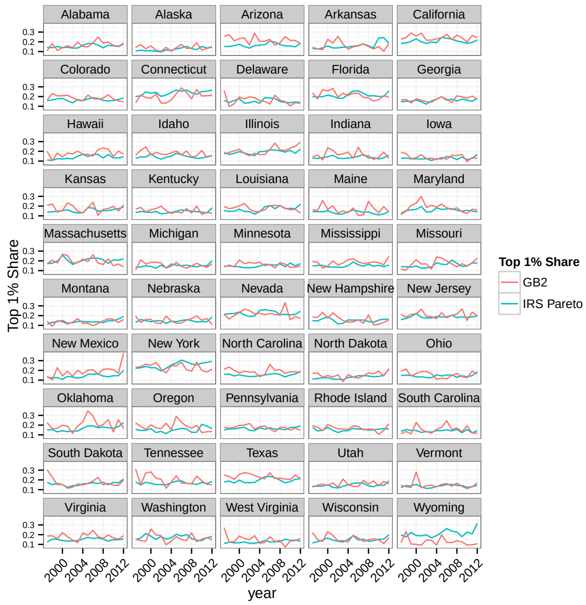


Pareto Interpolation

In order to estimate the Pareto parameter α using IRS summary data, I use Pareto interpolation. I describe the process using a single state (Alabama) in a single year (2012)

[2]In 2012, the average income of the top 1% in Wyoming was $4,844,205, the highest in the nation, despite the fact that Wyoming has the smallest population of any state.

FIGURE 33. State Top 1% Shares, 1997-2012: IRS Simulation vs. GB2 Simulation



by way of example. This process and notation is adapted from Sommeiller and Price (2014). Table 28 shows the information available in Historical Table 2 in the Statistics of Income Bulletin for Alabama.

Pareto Interpolation allows us to calculate the threshold incomes at a given percentile p, and to estimate the Pareto shape parameter α past this threshold. To do the interpolation, I need to proceed in steps. First, define s_i as the lower bound of each income bin i. Next, define N^* and Y^* as the cumulative sums of Returns and AGI, from highest bin to lowest:

$$N_i^* = \sum_{j=9}^{i} N_j$$

$$Y_i^* = \sum_{j=9}^{i} Y_j$$

TABLE 28. Alabama Total AGI and Number of Returns, by Size of AGI, 2012

Income Bins	Number of Returns	AGI
1 under 10,000	$340,890$	$-116,087,000$
10,000 under 25,000	$575,210$	$9,742,551,000$
25,000 under 50,000	$491,320$	$17,622,052,000$
50,000 under 75,000	$254,210$	$15,629,937,000$
75,000 under 100,000	$159,820$	$13,821,966,000$
100,000 under 200,000	$182,240$	$24,071,292,000$
200,000 under 500,000	$37,950$	$10,701,664,000$
500,000 under 1,000,000	$6,280$	$4,271,645,000$
1,000,000 or more	$2,970$	$8,525,047,000$

Define $y_i = \frac{Y_i^*}{N_i^*}$. Then I can define $b_i = \frac{y_i}{s_i}$, and hence the Pareto shape parameter is $\alpha_i = \frac{b_i}{b_i - 1}$. Let $p_i = \frac{N_i^*}{N_i}$ and $k_i = s_i \left(p_i^{\frac{1}{\alpha_i}} \right)$. I can finally derive an expression for the threshold income value at percentile p

$$TI\left(p\right) = \frac{k_i}{\left(1-p\right)^{\frac{1}{\alpha_i}}}$$

Or, equivalently, if I know a threshold income, the equivalent percentile is

$$p = 1 - \left(\frac{TI\left(p\right)}{k_i} \right)^{-\alpha_i}$$

This process yields estimates of $TI\left(p\right), p, \alpha_i$ using information from each income bin. The final step in the interpolation process is to select which income bin's estimates are to be used. To do this, I select the income bin for which $|p - p_i|$ is minimized (in other words, the bin whose cumulative percentage of incomes are higher is closest to percentile p).

I illustrate how this Pareto interpolation method can be used to improve estimates of inequality. Let's suppose that I know that incomes above the 99th percentile are understated in the CPS data. For Alabama in 2012, the 99th percentile of AGI in the CPS data is \$215527.10. To improve these estimates, I would like to replace incomes above this threshold with draws from the Pareto distribution implied by the IRS data. To do this, I first calculate the percentile at which the CPS cutoff income falls in the IRS AGI distribution p_i^P, and then select the income bin for which $\left|p_i^P - p_i\right|$ is minimized. The Pareto parameter α_i for the selected bin can then be used to simulate top incomes.

Table 29 summarizes the intermediate calculations of $y_i, b_i, \alpha_i, p_i, k_i$, and Table 30 summarizes the calculation of cutoff percentiles p_i^P and the choice of income bin, and the Pareto parameter selected.

TABLE 29. Intermediate Calculations in the Pareto Interpolation Process

Income Bin	y_i	b_i	α_i	p_i	k_i
1 under 10,000	50,841.380	5,084,137.000	1.000	1	0.010
10,000 under 25,000	61,044.540	6.104	1.196	0.834	8,589.864
25,000 under 50,000	83,401.870	3.336	1.428	0.553	16,518.010
50,000 under 75,000	119,697.200	2.394	1.717	0.314	25,459.060
75,000 under 100,000	157,713.600	2.103	1.907	0.190	31,373.360
100,000 under 200,000	207,329.400	2.073	1.932	0.112	32,177.260
200,000 under 500,000	497,846.500	2.489	1.671	0.023	20,944.210
500,000 under 1,000,000	1,383,426.000	2.767	1.566	0.005	15,885.180
1,000,000 or more	2,870,386.000	2.870	1.535	0.001	14,123.570

TABLE 30. Selecting the Pareto Parameter

Income Bin	p_i^P	$\lvert p_i^P - p_i \rvert$	α_i
1 under 10,000	1.000	1.000	1.000
10,000 under 25,000	0.979	0.813	1.196
25,000 under 50,000	0.974	0.528	1.428
50,000 under 75,000	0.974	0.288	1.717
75,000 under 100,000	0.975	0.164	1.907
100,000 under 200,000	0.975	0.086	1.932
200,000 under 500,000	0.980	**0.003**	**1.671**
500,000 under 1,000,000	0.983	0.012	1.566
1,000,000 or more	0.985	0.014	1.535

Conclusion

Previous estimates of income inequality within US states have relied on data from either the Current Population Survey or IRS tax returns. How ver, each data source has conceptual shortcomings. I have described an approach to incorporate information from both sources in a way that addresses the shortcomings while preserving the advantages. This method produces estimates of income inequality which are generally much larger than the naive results estimated from CPS data, while generally somewhat smaller than estimates using just IRS data.

Additionally, I show that I can closely match the level and trends in the Gini coefficient estimated using information from both the IRS and the CPS using a semi-parametric method that uses only CPS data. This suggests that CPS-based estimates of income inequality using income definitions other than tax unit Adjusted Gross Income should be regarded as with significantly less skepticism than that which they are often greeted.

APPENDIX B

STATE AND MSA INCOME INEQUALITY IN THE AMERICAN COMMUNITY SURVEY

As noted earlier, the issues presented by topcoding are most severe in the Current Population Survey. However, topcoding is not limited to the CPS. The decennial Census and the American Community Survey[1] both topcode incomes, although at a much higher threshold than in the CPS. The Census/ACS also has much less disaggregated income information, collecting data on only eight income sources. Each income source has a state-specific topcode in the public use data equal to the 99.5th percentile of that income source's distribution in a given state. Additionally, there is a hard topcode of $999,999 in the internal Census data. For the 1990 and 2000 decennial Censuses, and for the ACS from 2005-2011, the public use data imputes topcoded incomes with state-specific cell means calculated in a similar manner to the CPS cell-means, although the cells are state-based rather than demographic-based. Applying the GB2 multiple imputation methodology to Census and ACS data is then relatively straightforward. It should be noted, however, that there are slight differences in sample design between the CPS and Census/ACS that must be addressed to make the two series comparable.[2] I calculate the same scalar inequality metrics and Lorenz ordinates for the Census and ACS microdata as those produced from the CPS microdata.

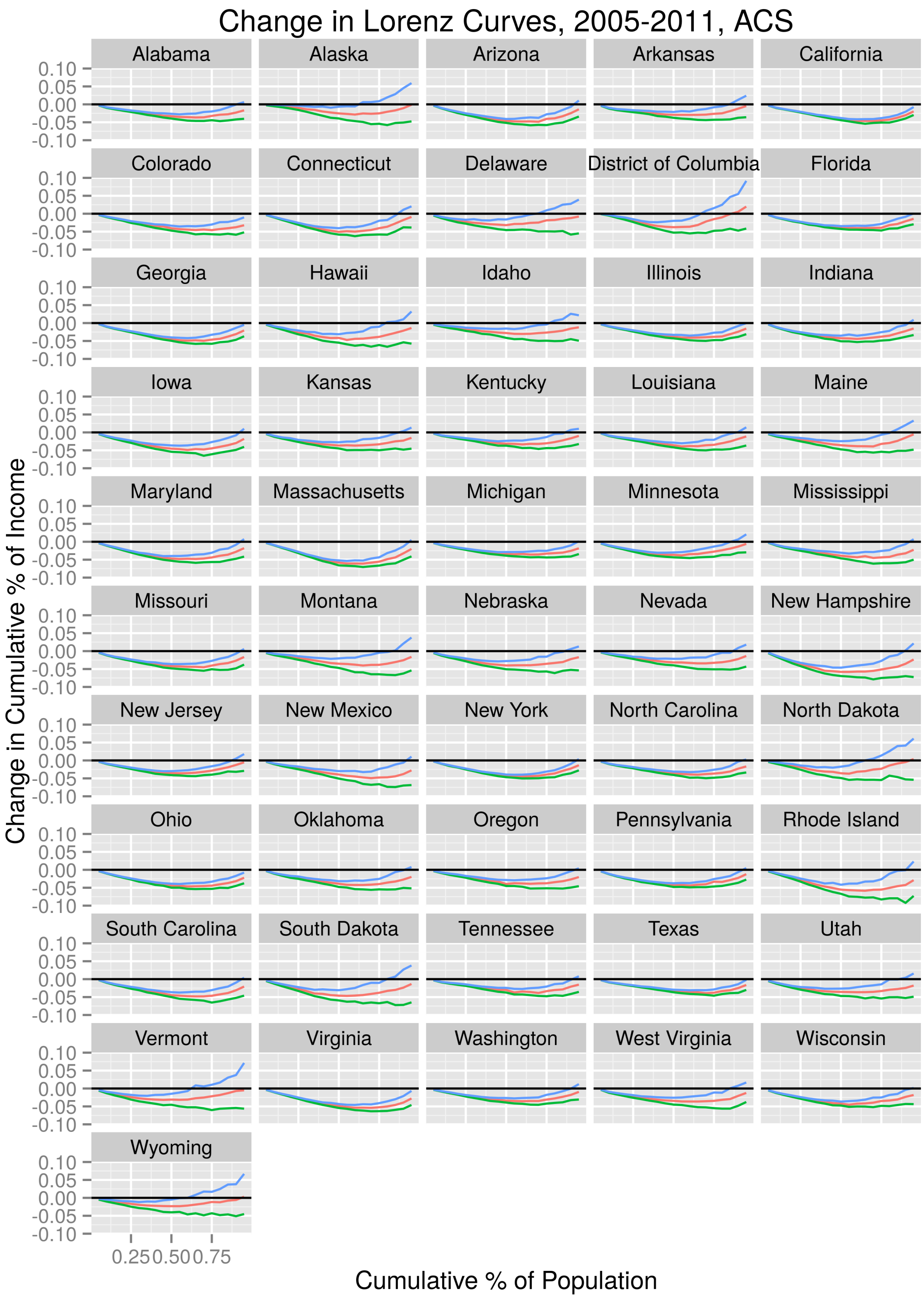
Figure 34 visualizes the point estimates and bootstrap confidence intervals for the change in Lorenz curve ordinates from 2005–2011 using ACS. I see that for almost all

[1]The ACS can be thought of as a successor to the decennial Census, since it replaces the "long form" survey after 2000. The ACS is smaller, providing 1% samples of the US each year, compared to the 5% sample available in the public use long form Census data.

[2]The chief difference here is that the ACS/Census and CPS have different coverage of individuals in group quarters. Following I dropped observations from the ACS and Census coded as living in institutions, old age homes or prisons.

states, the size of the change in Lorenz ordinates reaches a local minimum around the middle of the income distribution (generally between the 50th and 75th percentiles). This is very strong evidence against a pure "top incomes" story. It appears that not just the top 1%, but much of the top half of the distribution, is relatively better off. This also implies that the changes in inequality since 2005 are characterized not just by income gains among the top 1%, but by the fact that the bottom half of the distribution is relatively worse off than it was a decade ago. This important fact would be missed by analyses that focus solely on the share of income going to the top 1% of the distribution.

FIGURE 34. Lorenz Curve Results, ACS (2005-2011)



APPENDIX C

DECOMPOSITION ANALYSIS OF CHANGES IN ENVIRONMENTAL INEQUALITY

RIF regressions can be used to decompose the changes between two different distributions. Suppose that I want to analyze the difference in some functional ν between two distributions A and B. Firpo et al. (2011) show that the overall change in the functional $\nu_B - \nu_A$ can be decomposed as

$$\nu_B - \nu_A = \sum_{k=1}^{K} \overline{X}_{B,k} \left(\hat{\beta}_{B,k} - \hat{\beta}_{A,k}\right) + \sum_{k=1}^{K} \hat{\beta}_{A,k} \left(\overline{X}_{B,k} - \overline{X}_{A,k}\right)$$

where $\overline{X}_{i,k}$ is the mean of the kth demographic covariate in distribution $i \in \{A, B\}$, and the $\hat{\beta}_{i,k}$ are the parameter estimates from an RIF regression using distribution i data. Borrowing terminology from Oaxaca-Blinder wage decompositions, the first term in the decomposition is the "structure effect," where some part of the change is due to a change in parameters, and the second is the "composition effect," where some part is due to a change in variables.

In the aggregate, these two effects can be viewed as the unexplained and explained variation in the change in the distributional functional of interest. Additionally, each element of the aggregate composition effect can be examined individually (this is often termed a "detailed decomposition"). The structure effect captures all of the variation in environmental inequality that is not due to changes in observable demographic characteristics. The estimated structure effect in a decomposition of the change in environmental inequality in 2005 vs. 2011 can then be thought of as an upper bound on the environmental policy that applies to all census tracts.

I decompose the difference in each of the distributional functionals of interest between 2005 and 2011, providing a detailed decomposition of the composition effect, which describes how changes in each sociodemographic variable contribute to the overall change in the functional.[1] The aggregate structural effect then captures all variation that is not explained by sociodemographic and economic factors. One unexplained factor that I cannot separately identify is the aggregate influence of environmental policies (chiefly, the continuing regulations associated with the Clean Air Act). The structure effect can be interpreted as a rough upper bound of the aggregate effect of environmental policy on environmental inequality.

Tables 31 and 32 summarize the results of a Oaxaca-Blinder style decomposition using the quantile points as distributional functionals for NO_x exposure and PM2.5 exposure respectively. The aggregate composition effect is positive and statistically significant at conventional levels for the entire pollution exposure distribution, while the aggregate structure effect is negative throughout the distribution (and larger at each quantile). Aggregate sociodemographic changes appear to be responsible for a net increase in exposure at any given quantile, but this effect is swamped by the effect of policy (the structure effect).

Tables 33 and 34 decompose the changes in Lorenz ordinates for the two pollutants.[2] It is possible to glean information about the contribution of changes in demographics from the detailed decomposition terms. I can see that race and ethnicity correlate positively with measured inequality according to the Lorenz curve, with proportionally larger effects on the cumulative share of pollution exposure for the most polluted tracts. The black share of a tract's population contributes positive changes in Lorenz curves across the exposure distribution for both PM2.5 and NO_x. The opposite is true for the Hispanic share of the population, suggesting that the proportion African-American contributes

---

[1] The detailed decomposition of the structure effect, as noted by Firpo et al. (2011) has no clear interpretation in most cases.

[2] Tables 37, 38, 39 and 40 repeat the exercise for generalized and absolute Lorenz curves.

to rising environmental inequality, while the proportion Hispanic contributes to falling environmental inequality.

As noted in chapters III and IV of this dissertation, the environmental justice literature highlights the degree to which black and Latino populations are exposed to excess pollution exposure relative to non-Hispanic whites. I can perform decomposition exercises to further illuminate this disparity, and its connection to environmental inequality. Following Fowlie et al. (2012), I divide census tracts into "diverse" tracts (where the proportion of African-Americans and Latinos exceeds 30%) and non-diverse tracts. I have observations of annual average PM2.5 and NO_x exposure (and demographic characteristics) for all tracts for each year from 2005-2011. This means I can use a panel data extension of the RIF-based decomposition above.

Following Firpo et al. (2011), I allow each census tract to have a tract-specific fixed effect θ_i, and assume that the return to the fixed effect is 1 for the non-diverse tracts (now distribution A in my notation), and σ for the diverse tracts (now distribution B). The change in any functional ν can be decomposed as

$$\nu_B - \nu_A = \sum_{k=1}^{K} \overline{X}_{B,k} \left(\hat{\beta}_{B,k} - \hat{\beta}_{A,k}\right) + \overline{\theta}_B \left(\sigma - 1\right) + \sum_{k=1}^{K} \hat{\beta}_{A,k} \left(\overline{X}_{B,k} - \overline{X}_{A,k}\right) + \left(\overline{\theta}_B - \overline{\theta}_A\right)$$

Where there are two new terms: $\overline{\theta}_B \left(\sigma - 1\right)$ is the structure effect of the fixed effect, and $\left(\overline{\theta}_B - \overline{\theta}_A\right)$ is the composition effect of the fixed effects.

Tables 35 and 36 summarize decompositions of the difference in quantile points between diverse and non-diverse census tracts for NO_x and PM2.5 respectively. I find that the first evidence of education effects — for heavily polluted tracts, the composition effect of the less than high school educated proportion is negative and statistically significant. Poverty and the unemployment rate both contribute positively to the diverse vs. non-diverse gap for highly exposed tracts, but negatively for lightly-exposed tracts.

TABLE 31. NOx Quantile Detailed Decomposition, 2005 vs. 2011

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	0.575***	0.854***	1.958***	2.578***	4.908***
	(0.030)	(0.038)	(0.055)	(0.071)	(0.170)
aggregate structure	-0.672***	-0.992***	-1.085***	-2.059***	-3.503***
	(0.046)	(0.048)	(0.066)	(0.102)	(0.191)
black	0.000***	0.001***	0.000	-0.005***	-0.008***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
highschool	-0.008***	-0.017***	-0.036***	-0.046***	-0.034***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.005)
incometoneeds50to99	-0.003***	-0.005***	-0.009***	-0.021***	-0.042***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)
latino	-0.021***	-0.023***	-0.041***	-0.017***	0.002
	(0.001)	(0.001)	(0.003)	(0.003)	(0.009)
lessthanhs	-0.001***	0.000*	0.001	-0.003***	0.007***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)
medianinc	0.011***	0.021***	0.039***	0.032***	-0.037***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.005)
unemployment	0.006***	0.027***	0.062***	0.071***	0.087***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.008)

Estimates show the contribution to the change in quantile points from 2005-2011.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

TABLE 32. PM2.5 Quantile Detailed Decomposition, 2005 vs. 2011

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	3.043***	1.664***	1.779***	2.420***	4.940***
	(0.280)	(0.140)	(0.169)	(0.113)	(0.235)
aggregate structure	-2.910***	-2.863***	-3.983***	-3.463***	-2.650***
	(0.391)	(0.165)	(0.161)	(0.152)	(0.167)
black	-0.010***	-0.010***	-0.017***	-0.004***	0.004***
	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)
highschool	-0.057***	-0.039***	-0.062***	-0.029***	0.009
	(0.008)	(0.004)	(0.006)	(0.004)	(0.008)
incometoneeds50to99	-0.021***	-0.021***	-0.036***	-0.024***	-0.020***
	(0.004)	(0.002)	(0.004)	(0.002)	(0.004)
latino	-0.129***	-0.179***	-0.131***	-0.022***	0.082***
	(0.013)	(0.008)	(0.009)	(0.005)	(0.012)
lessthanhs	-0.032***	-0.026***	-0.013***	-0.006***	-0.017***
	(0.005)	(0.004)	(0.002)	(0.001)	(0.003)
medianinc	0.085***	0.040***	0.036***	-0.007***	-0.065***
	(0.007)	(0.003)	(0.004)	(0.002)	(0.006)
unemployment	0.060***	0.097***	0.189***	0.131***	0.209***
	(0.009)	(0.005)	(0.009)	(0.006)	(0.011)

Estimates show the contribution to the change in quantile points from 2005-2011.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

TABLE 33. NOx Relative Lorenz Curve Detailed Decomposition, 2005 vs. 2011

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	0.001	-0.006*	0.038***	0.107***	0.197***
	(0.002)	(0.003)	(0.007)	(0.009)	(0.008)
aggregate structure	-0.006***	-0.023***	-0.031***	-0.023**	-0.036***
	(0.001)	(0.003)	(0.006)	(0.010)	(0.008)
black	0.000	0.000***	0.000***	0.000*	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
highschool	0.000	0.000	-0.002***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
incometoneeds50to99	0.000***	0.000***	0.000***	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
latino	0.000***	-0.001***	-0.005***	-0.005***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
lessthanhs	0.000***	0.000***	0.000*	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
medianinc	0.000***	0.001***	0.005***	0.009***	0.008***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
unemployment	-0.001***	0.000***	0.003***	0.008***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Estimates show the contribution to the change in RLC points from 2005-2011.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

TABLE 34. PM2.5 Relative Lorenz Curve Detailed Decomposition, 2005 vs. 2011

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	0.019***	0.038***	0.014***	-0.010***	0.003
	(0.002)	(0.004)	(0.005)	(0.003)	(0.002)
aggregate structure	0.004**	-0.008**	-0.020***	-0.024***	-0.014***
	(0.002)	(0.004)	(0.004)	(0.003)	(0.002)
black	0.000**	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
highschool	0.000*	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
incometoneeds50to99	0.000	0.000**	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
latino	-0.001***	-0.004***	-0.005***	-0.004***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
lessthanhs	0.000***	-0.001***	-0.001***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
medianinc	0.001***	0.002***	0.002***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
unemployment	0.000**	0.000	0.000*	0.000**	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

EEstimates show the contribution to the change in RLC points from 2005-2011.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

TABLE 35. NOx Quantile Detailed Decomposition, Diverse vs. Non-diverse tracts

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	-0.117***	-0.174***	0.000	0.528***	1.193***
	(0.024)	(0.025)	(0.033)	(0.064)	(0.136)
aggregate structure	0.167	-0.096	0.597	-2.655***	-6.035***
	(0.265)	(0.279)	(0.427)	(0.949)	(1.599)
black	-0.192***	-0.201***	-0.025	0.454***	0.928***
	(0.019)	(0.019)	(0.026)	(0.052)	(0.097)
highschool	-0.001***	0.000	0.000	-0.001	0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)
incometoneeds50$_9$9	0.010**	0.005	0.007	0.046***	-0.037*
	(0.004)	(0.004)	(0.007)	(0.009)	(0.020)
latino	0.056***	0.009	-0.008	0.033	0.336***
	(0.018)	(0.019)	(0.025)	(0.050)	(0.103)
lessthanhs	-0.060***	-0.085***	-0.077***	0.028	0.210***
	(0.010)	(0.010)	(0.014)	(0.024)	(0.049)
medianinc	-0.007*	0.012***	0.013*	-0.112***	-0.276***
	(0.004)	(0.004)	(0.007)	(0.013)	(0.032)
unemployment	0.047***	0.026***	0.014***	-0.035***	-0.028**
	(0.003)	(0.004)	(0.005)	(0.008)	(0.014)

Estimates show the contribution to the change in quantile points between Diverse vs. Non-diverse tracts.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

TABLE 36. PM2.5 Quantile Detailed Decomposition, Diverse vs. Non-diverse tracts

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	-0.486***	-0.106	0.190***	1.491***	2.042***
	(0.088)	(0.117)	(0.064)	(0.170)	(0.225)
aggregate structure	2.869**	4.826***	-1.206	-5.141***	-3.520
	(1.284)	(1.007)	(0.840)	(1.769)	(2.648)
black	-0.168***	-0.201**	-0.135***	0.572***	1.291***
	(0.052)	(0.089)	(0.050)	(0.126)	(0.165)
highschool	-0.005**	-0.002	-0.002	0.004	0.000
	(0.002)	(0.003)	(0.001)	(0.004)	(0.005)
incometoneeds50$_9$9	-0.030*	-0.031*	0.003	0.042	0.151***
	(0.018)	(0.019)	(0.011)	(0.027)	(0.031)
latino	-0.274***	0.257***	0.315***	0.647***	0.233
	(0.086)	(0.084)	(0.042)	(0.115)	(0.153)
lessthanhs	-0.042	-0.320***	-0.254***	-0.416***	-0.074
	(0.042)	(0.048)	(0.024)	(0.065)	(0.082)
medianinc	-0.026	-0.093***	-0.094***	-0.150***	0.148***
	(0.021)	(0.022)	(0.014)	(0.032)	(0.043)
unemployment	0.008	0.127***	0.092***	0.184***	0.053**
	(0.014)	(0.014)	(0.008)	(0.023)	(0.026)

Estimates show the contribution to the change in quantile points between Diverse vs. Non-diverse tracts.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

TABLE 37. NOx Generalized Lorenz Curve Detailed Decomposition, 2005 vs. 2011

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	0.045***	0.140***	0.490***	1.071***	1.644***
	(0.002)	(0.006)	(0.014)	(0.025)	(0.033)
aggregate structure	-0.050***	-0.178***	-0.438***	-0.813***	-1.201***
	(0.003)	(0.008)	(0.020)	(0.033)	(0.042)
black	0.000**	0.000***	0.000***	0.000***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
highschool	-0.001***	-0.002***	-0.009***	-0.019***	-0.026***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
incometoneeds50to99	0.000***	-0.001***	-0.003***	-0.006***	-0.010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
latino	-0.001***	-0.004***	-0.013***	-0.020***	-0.022***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)
lessthanhs	0.000***	0.000***	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
medianinc	0.001***	0.003***	0.011***	0.021***	0.023***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
unemployment	0.000	0.002***	0.014***	0.031***	0.044***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)

Estimates show the contribution to the change in GLC points from 2005-2011.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

TABLE 38. NOx Absolute Lorenz Curve Detailed Decomposition, 2005 vs. 2011

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	-0.143***	-0.330***	-0.450***	-0.339***	-0.048***
	(0.003)	(0.008)	(0.014)	(0.015)	(0.011)
aggregate structure	0.105***	0.210***	0.338***	0.350***	0.195***
	(0.005)	(0.012)	(0.022)	(0.026)	(0.024)
black	0.000***	0.000***	0.001***	0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
highschool	0.002***	0.004***	0.005***	0.001*	-0.002***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
incometoneeds50to99	0.001***	0.003***	0.005***	0.005***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
latino	0.002***	0.003***	0.000	0.000	0.003**
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
lessthanhs	0.000***	0.000***	0.000**	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
medianinc	-0.001***	-0.001***	0.004***	0.010***	0.010***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
unemployment	-0.004***	-0.007***	-0.005***	0.003***	0.010***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)

Estimates show the contribution to the change in ALC points from 2005-2011.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

TABLE 39. PM2.5 Generalized Lorenz Curve Detailed Decomposition, 2005 vs. 2011

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	0.340***	0.857***	1.237***	1.736***	2.363***
	(0.031)	(0.062)	(0.093)	(0.101)	(0.115)
aggregate structure	-0.117***	-0.621***	-1.506***	-2.411***	-2.867***
	(0.022)	(0.059)	(0.094)	(0.108)	(0.109)
black	-0.001***	-0.003***	-0.006***	-0.009***	-0.009***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
highschool	-0.003***	-0.012***	-0.025***	-0.033***	-0.037***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)
incometoneeds50to99	-0.001***	-0.005***	-0.013***	-0.019***	-0.023***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)
latino	-0.014***	-0.052***	-0.091***	-0.107***	-0.104***
	(0.001)	(0.003)	(0.005)	(0.005)	(0.006)
lessthanhs	-0.002***	-0.008***	-0.013***	-0.014***	-0.016***
	(0.000)	(0.001)	(0.002)	(0.002)	(0.002)
medianinc	0.008***	0.020***	0.029***	0.032***	0.028***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
unemployment	0.005***	0.022***	0.059***	0.097***	0.127***
	(0.001)	(0.002)	(0.003)	(0.004)	(0.005)

Estimates show the contribution to the change in GLC points from 2005-2011.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

TABLE 40. PM2.5 Absolute Lorenz Curve Detailed Decomposition, 2005 vs. 2011

	p=0.1	p=0.25	p=0.5	p=0.75	p=0.9
aggregate composition	0.068***	0.178***	-0.123***	-0.304***	-0.085***
	(0.024)	(0.042)	(0.042)	(0.029)	(0.016)
aggregate structure	0.196***	0.162***	0.060	-0.062**	-0.048**
	(0.019)	(0.038)	(0.045)	(0.027)	(0.019)
black	0.000***	0.000***	-0.002***	-0.002***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
highschool	0.000	-0.004***	-0.008***	-0.008***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
incometoneeds50to99	0.001***	0.001	0.000	-0.001**	-0.001***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
latino	-0.004***	-0.029***	-0.046***	-0.039***	-0.023***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
lessthanhs	-0.001***	-0.004***	-0.004***	-0.001***	-0.001***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
medianinc	0.006***	0.015***	0.018***	0.016***	0.009***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
unemployment	-0.009***	-0.013***	-0.011***	-0.008***	0.002**
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)

Estimates show the contribution to the change in ALC points from 2005-2011.
Bootstrapped standard errors shown in parentheses.
Other sociodemographic variables omitted.

APPENDIX D

ADDITIONAL TABLES AND FIGURES

Chapter II Miscellaneous Tables and Figures

FIGURE 35. State-level Gini Coefficient, Pre-transfer Income

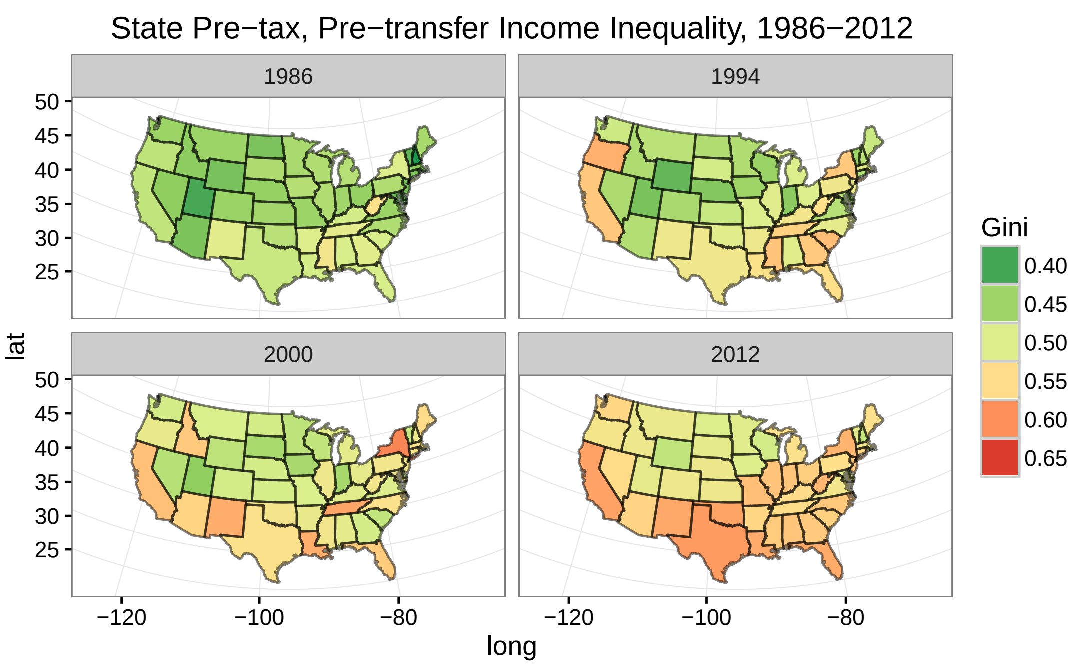


FIGURE 36. State-level Gini Coefficient, Post-transfer Income

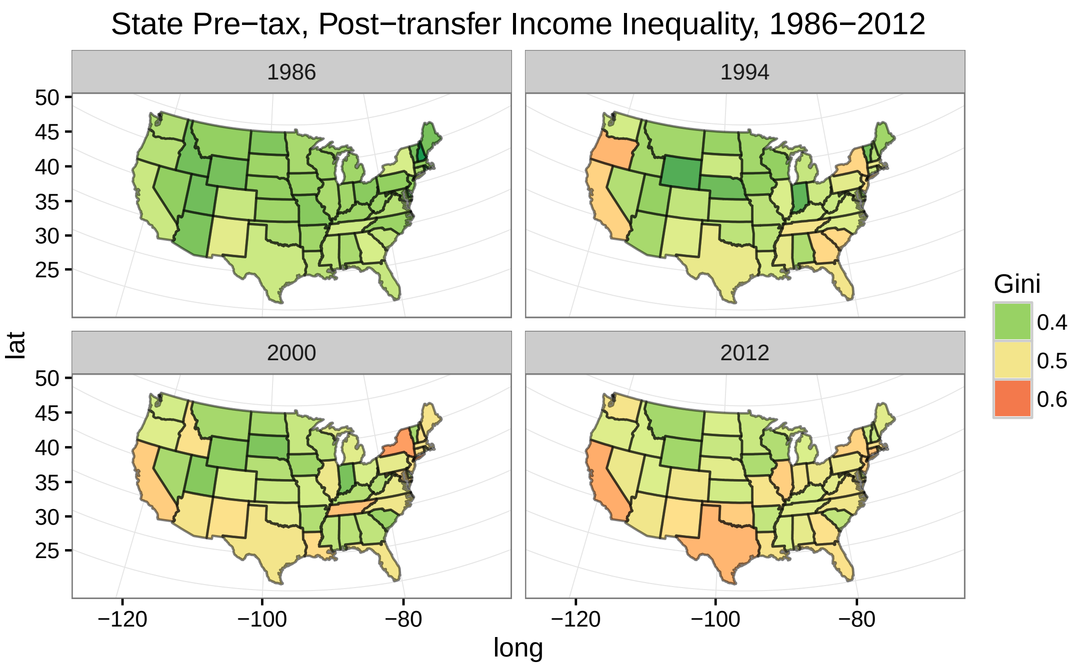


FIGURE 37. State-level Gini Coefficient, Post-tax Income

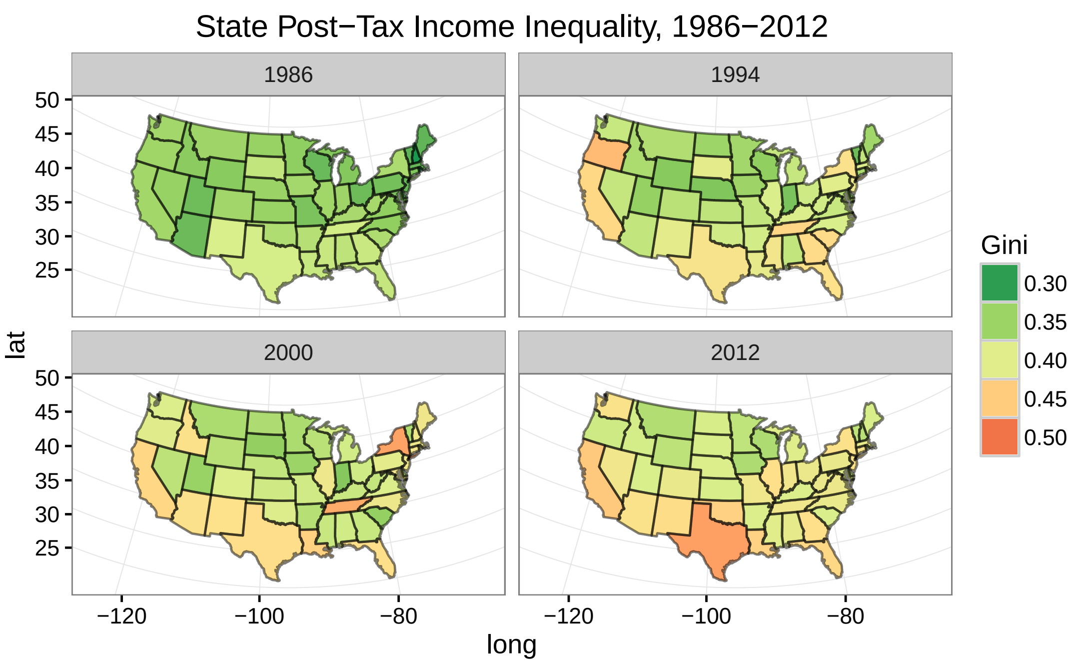


FIGURE 38. State-level Gini Coefficient, Post-fiscal Income

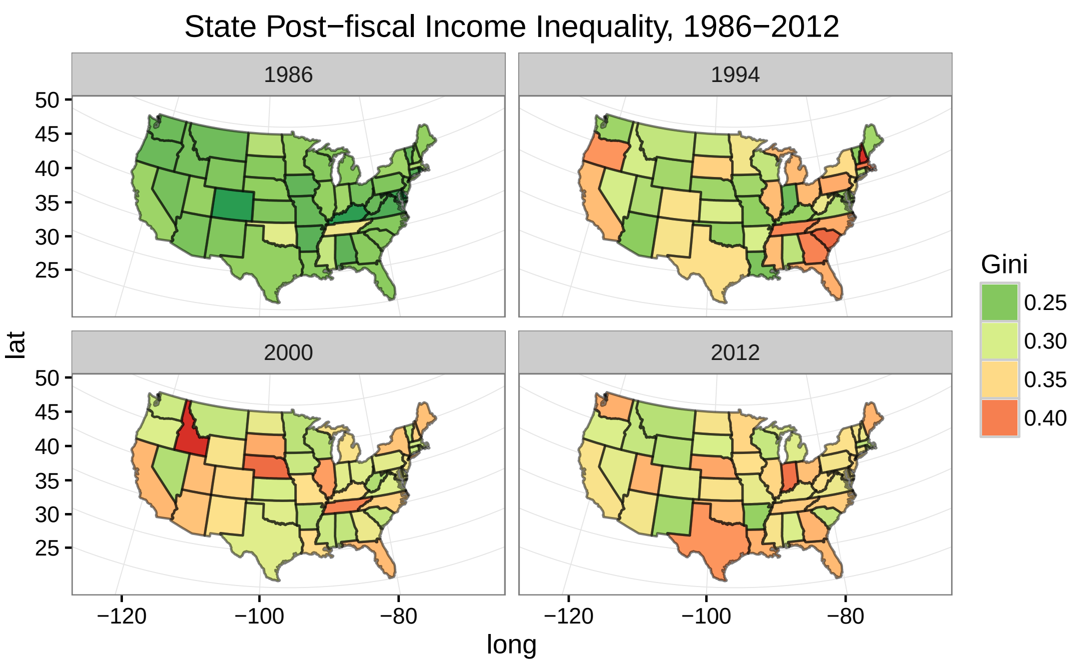


TABLE 41. Determinants of State Income Inequality (Pre-transfer and Post-transfer Gini), All Covariates

	Dependent variable:					
	(1)	(2)	(3)	(4)	(5)	(6)
union_cov	-0.231^{***}	-0.323^{***}	-0.327^{***}	-0.183	-0.320^{**}	-0.338^{**}
	(0.085)	(0.108)	(0.124)	(0.121)	(0.142)	(0.163)
state_minwage	-0.0001	-0.001	-0.002	0.0004	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Total_rate_capgains	-0.006^{***}	-0.007^{**}	-0.008^{*}	-0.008^{***}	-0.008^{**}	-0.009
	(0.002)	(0.003)	(0.004)	(0.002)	(0.004)	(0.006)
Total_rate_wages	0.003	0.004	0.004	0.006^{*}	0.005	0.004
	(0.003)	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)
UR	0.004^{***}	0.003^{**}	0.002	0.004^{**}	0.004^{*}	0.002
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Real_PersIncPC	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000^{*}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
college_prop	0.077	0.130	0.0001	0.163	0.183	0.062
	(0.144)	(0.169)	(0.178)	(0.188)	(0.234)	(0.247)
std_educ	0.009	0.012	0.011	0.017	0.022	0.016
	(0.017)	(0.023)	(0.021)	(0.022)	(0.030)	(0.025)
mean_educ	-0.021	-0.032	-0.013	-0.005	-0.007	0.002
	(0.022)	(0.026)	(0.030)	(0.028)	(0.036)	(0.039)
manufacturing	-0.035	0.044	-0.073	0.022	0.090	-0.063
	(0.070)	(0.111)	(0.132)	(0.090)	(0.144)	(0.170)
popdens	-0.0001	-0.0002	-0.001	-0.0001	0.00004	-0.001
	(0.0001)	(0.0005)	(0.001)	(0.0001)	(0.0005)	(0.001)
RD_percap	-0.042^{**}	-0.125^{***}	-0.048	-0.034	-0.159^{***}	-0.055
	(0.018)	(0.039)	(0.062)	(0.023)	(0.047)	(0.070)
total_patents_PC	0.013	0.024	0.015	0.021	0.028	0.013
	(0.011)	(0.015)	(0.023)	(0.014)	(0.020)	(0.028)
GovtPC	-7.169	-15.692^{**}	-9.810	-4.184	-13.262	-9.147
	(4.772)	(7.443)	(8.851)	(6.066)	(9.290)	(10.275)
black	0.108^{**}	0.121^{**}	0.097	0.099	0.105	0.115
	(0.054)	(0.061)	(0.073)	(0.069)	(0.083)	(0.095)

latino	0.087	0.085	0.099	0.198***	0.149	0.160
	(0.064)	(0.091)	(0.099)	(0.076)	(0.119)	(0.129)
age	0.001	0.0001	−0.001	0.002	0.001	−0.001
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
married	0.158	0.098	−0.036	0.328***	0.240*	0.082
	(0.102)	(0.115)	(0.121)	(0.126)	(0.144)	(0.147)
divorced	0.181	0.057	−0.115	0.159	0.023	−0.156
	(0.166)	(0.172)	(0.152)	(0.218)	(0.228)	(0.197)
over55	0.038	0.098	0.040	−0.100	0.024	−0.017
	(0.085)	(0.099)	(0.116)	(0.116)	(0.130)	(0.149)
under25	0.120	−0.017	−0.177	0.162	0.019	−0.171
	(0.135)	(0.147)	(0.127)	(0.175)	(0.184)	(0.143)
noncitizen	0.260*	−0.003	−0.064	0.255	0.069	−0.037
	(0.140)	(0.205)	(0.243)	(0.155)	(0.258)	(0.285)
nativeborn	0.143	−0.011	−0.006	0.226	0.088	0.076
	(0.127)	(0.155)	(0.178)	(0.155)	(0.193)	(0.221)
Linear Trends?	No	Yes	Yes	No	Yes	Yes
Quad. Trends?	No	No	Yes	No	No	Yes
Observations	1,000	1,000	1,000	1,000	1,000	1,000
Note:	*p<0.1; **p<0.05; ***p<0.01 All models include State and Year fixed effects					

TABLE 42. Determinants of State Income Inequality (Post-tax and Post-fiscal Gini), All Covariates

	Dependent variable:					
	(1)	(2)	(3)	(4)	(5)	(6)
union_cov	−0.160**	−0.220**	−0.195*	−0.115	0.009	0.001
	(0.080)	(0.097)	(0.113)	(0.128)	(0.145)	(0.173)
state_minwage	0.001	0.00002	−0.001	−0.003	0.001	−0.0003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Total_rate_capgains	−0.006***	−0.006**	−0.007*	−0.005	−0.003	−0.005
	(0.002)	(0.003)	(0.004)	(0.003)	(0.006)	(0.007)
Total_rate_wages	0.004*	0.004	0.004	0.004	0.004	0.005
	(0.002)	(0.004)	(0.005)	(0.004)	(0.007)	(0.009)
UR	0.003***	0.003**	0.003	0.001	0.001	−0.0004

	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Real_PersIncPC	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
college_prop	0.084	0.082	−0.014	−0.147	−0.056	−0.142
	(0.130)	(0.157)	(0.168)	(0.198)	(0.271)	(0.300)
std_educ	0.006	0.011	0.009	−0.018	−0.027	−0.027
	(0.015)	(0.020)	(0.018)	(0.022)	(0.027)	(0.031)
mean_educ	−0.015	−0.016	−0.003	0.031	0.015	0.026
	(0.020)	(0.025)	(0.028)	(0.029)	(0.038)	(0.044)
manufacturing	0.006	0.044	−0.054	0.192*	0.106	−0.006
	(0.062)	(0.099)	(0.117)	(0.098)	(0.164)	(0.182)
popdens	−0.0001	0.00001	−0.0004	−0.0003**	0.0001	0.00002
	(0.0001)	(0.0003)	(0.0004)	(0.0001)	(0.001)	(0.001)
RD_percap	−0.025*	−0.116***	−0.031	−0.054*	0.052	−0.051
	(0.015)	(0.032)	(0.054)	(0.033)	(0.075)	(0.115)
total_patents_PC	0.012	0.021	0.016	0.004	0.040*	0.043
	(0.010)	(0.014)	(0.020)	(0.016)	(0.023)	(0.037)
GovtPC	−3.729	−11.331*	−6.775	4.831	−1.080	0.248
	(4.034)	(6.566)	(7.622)	(8.164)	(9.778)	(13.843)
black	0.110**	0.118**	0.098	0.124	0.135	0.085
	(0.049)	(0.057)	(0.067)	(0.139)	(0.149)	(0.173)
latino	0.102**	0.111	0.115	0.066	0.298*	0.326*
	(0.052)	(0.085)	(0.091)	(0.111)	(0.160)	(0.172)
age	0.001	0.0003	−0.001	0.003	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
married	0.184**	0.129	0.021	−0.063	−0.053	−0.185
	(0.083)	(0.097)	(0.101)	(0.201)	(0.213)	(0.247)
divorced	0.162	0.067	−0.081	−0.009	−0.075	−0.217
	(0.148)	(0.156)	(0.137)	(0.199)	(0.215)	(0.229)
over55	−0.059	0.020	−0.007	−0.218*	−0.176	−0.270
	(0.079)	(0.088)	(0.107)	(0.121)	(0.148)	(0.178)
under25	0.110	0.001	−0.137	−0.112	−0.181	−0.330
	(0.121)	(0.130)	(0.109)	(0.217)	(0.242)	(0.281)
noncitizen	0.191*	0.001	−0.027	0.115	−0.815**	−0.787*

	(0.103)	(0.179)	(0.210)	(0.182)	(0.354)	(0.414)
nativeborn	0.124	0.028	0.053	0.032	−0.247	−0.187
	(0.109)	(0.134)	(0.149)	(0.146)	(0.203)	(0.247)
Linear Trends?	No	Yes	Yes	No	Yes	Yes
Quad. Trends?	No	No	Yes	No	No	Yes
Observations	1,000	1,000	1,000	1,000	1,000	1,000

FIGURE 39. National Black-White PM2.5 Exposure Ratio (by Percentile), 1998-2014

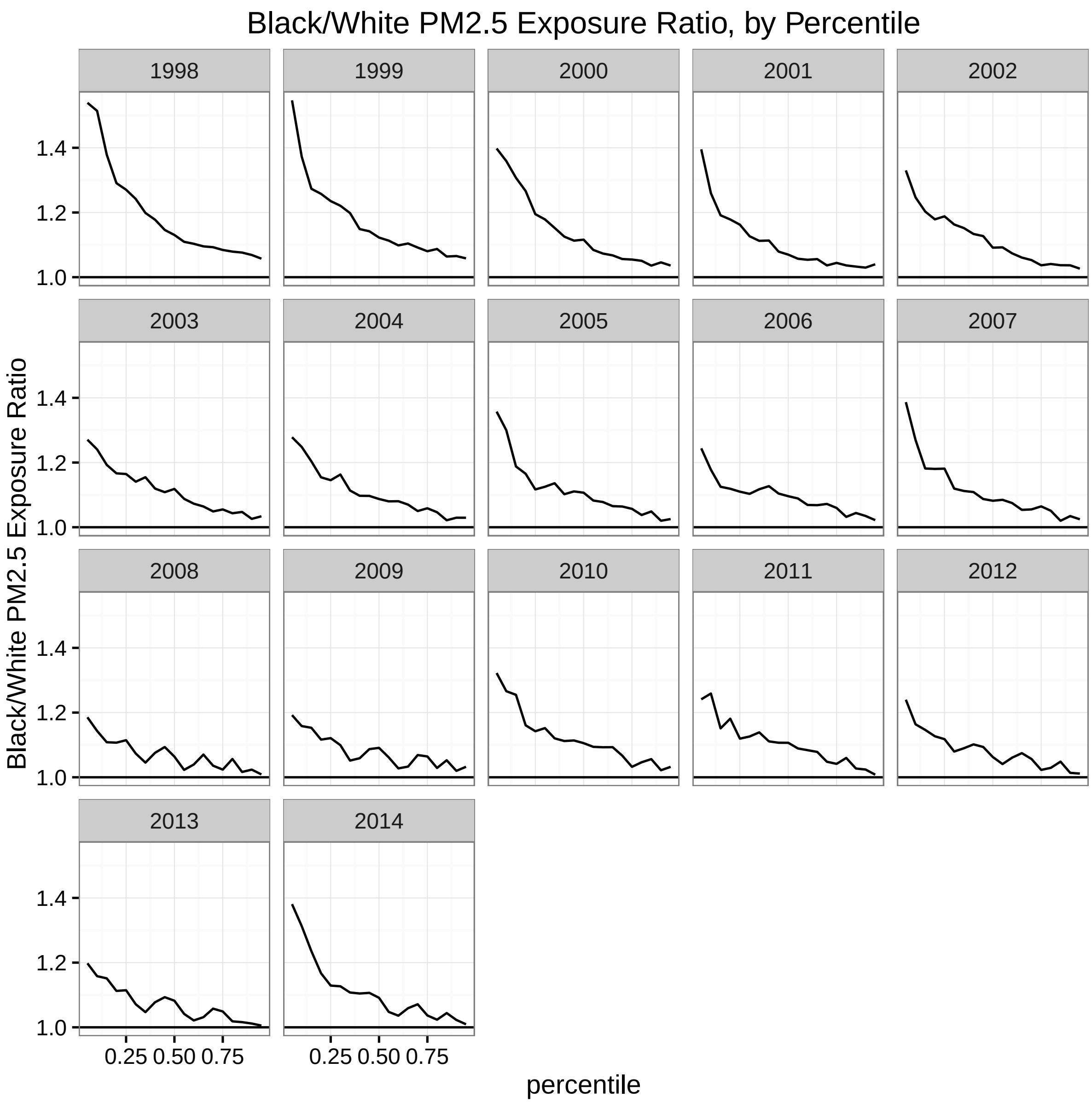


FIGURE 40. National Black-White NO_x Exposure Ratio (by Percentile), 2005-2011

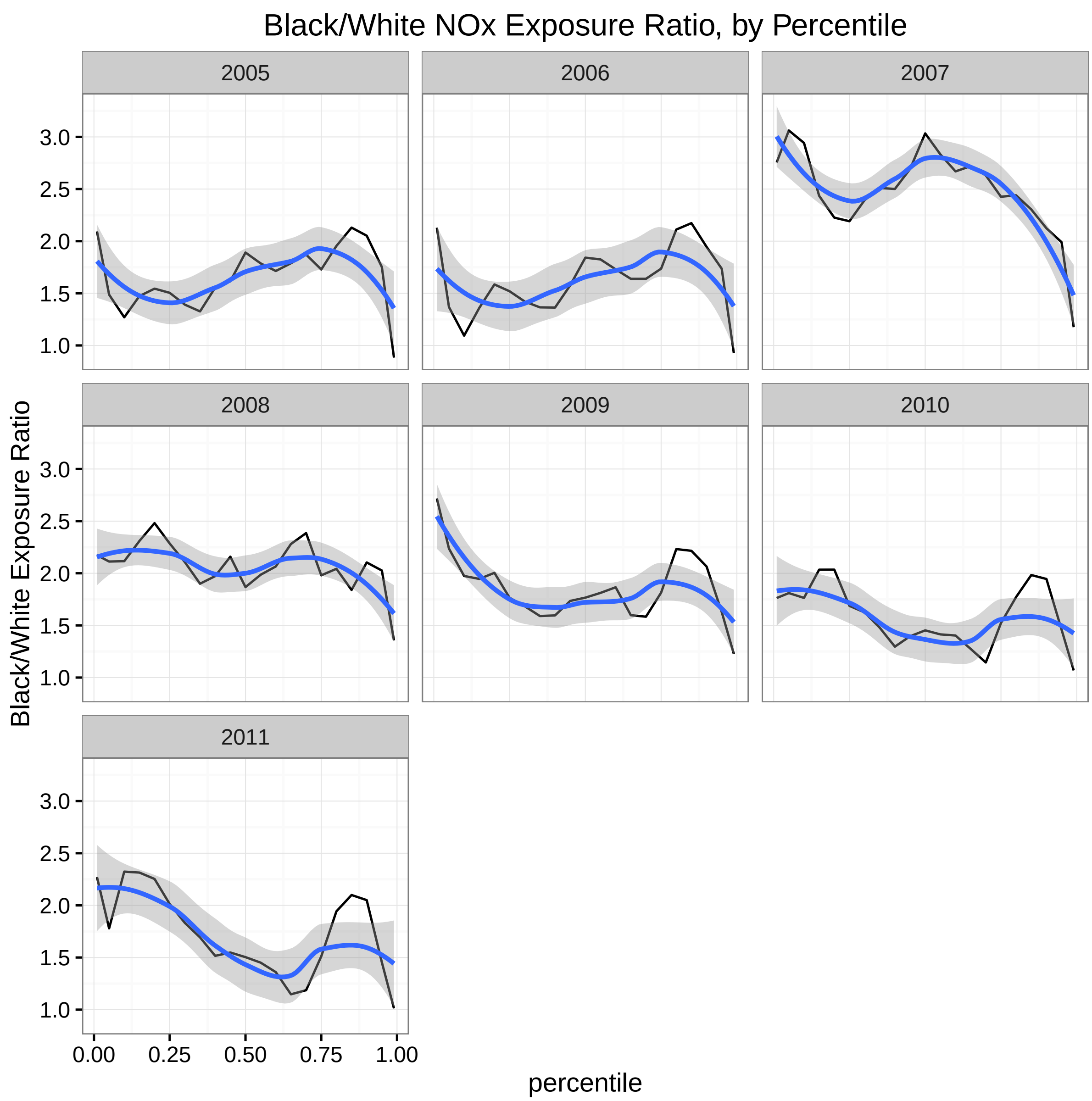


TABLE 43. Effect of Income Inequality on Average Latino Exposure

		(1)	(2)	(3)
IV Results:	Gini	-3.717^{***}	-5.493^{***}	-6.020^{***}
		(1.388)	(1.702)	(2.114)
OLS Results:	Gini	-0.309	-0.477	-0.484
		(0.373)	(0.377)	(0.445)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

TABLE 44. Effect of Income Inequality on Average Poor Exposure

		(1)	(2)	(3)
IV Results:	Gini	-3.696^{***}	-5.460^{***}	-6.117^{***}
		(1.374)	(1.701)	(2.131)
OLS Results:	Gini	-0.287	-0.465	-0.461
		(0.368)	(0.373)	(0.437)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

TABLE 45. Effect of Income Inequality on Average Rich Exposure

		(1)	(2)	(3)
IV Results:	Gini	-4.174^{***}	-5.922^{***}	-6.605^{***}
		(1.341)	(1.683)	(2.121)
OLS Results:	Gini	-0.335	-0.491	-0.494
		(0.359)	(0.366)	(0.428)
Observations		1,578	1,578	1,578
First Stage F		86.22	70.5	55.34
Control Variables?		No	Yes	Yes
MSA-specific Trend?		No	No	Yes

Notes: See Table 17 for further details

REFERENCES CITED

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature*, 3(1).

Alkire, S. and Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95.

Alvaredo, F. (2011). A note on the relationship between top income shares and the gini coefficient. *Economics Letters*, 110(3):274–277.

Armour, P., Burkhauser, R., and Larrimore, J. (2014). Levels and trends in united states income and its distribution: A crosswalk from market income towards a comprehensive haig-simons income measure. *Southern Economic Journal*, 81(2).

Atkinson, A. (2007). Measuring top incomes: Methodological issues. In Atkinson, A. and Piketty, T., editors, *Top Incomes over the Twentieth Century: A Contrast BetweenContinental European and English-Speaking Countries*. Oxford University Press.

Atkinson, A., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, 49(1953):3–71.

Baek, J. and Gweisah, G. (2013). Does income inequality harm the environment?: Empirical evidence from the United States. *Energy Policy*, 62:1434–1437.

Banzhaf, S., editor (2012). *The Political Economy of Environmental Justice*. Stanford University Press.

Barrett, G., Donald, S., and Bhattacharya, D. (2014). Consistent nonparametric tests for lorenz dominance. *Journal of Business & Economic Statistics*, 32(1).

Bartels, L. (2009). Economic inequality and political representation. In Jacobs, L. and King, D., editors, *The Unsustainable American State*. Oxford University Press, Oxford.

Baum-Snow, N. and Ferreira, F. (2015). Causal inference in urban economics. In Duranton, G., Henderson, V., and Strange, W., editors, *Handbook of Regional and Urban Economics, vol. 5A*. Elseview North-Holland.

Been, V. and Gupta, F. (1997). Coming to the nuisance or going to the barrios - a longitudinal analysis of environmental justice claims. *Ecology Law Quarterly*, 24.

Berthe, A. and Elie, L. (2015). Mechanisms explaining the impact of economic inequality on environmental deterioration. *Ecological Economics*, 116.

Bishop, J., Formby, J., and Smith, W. J. (1991). Lorenz dominance and welfare: Changes in the us distribution of income, 1967-1987. *The Review of Economics and Statistics*, 73(1).

Boustan, L., Ferreira, F., Winkler, H., and Zolt, E. (2013). The Effect of Rising Income Inequality on Taxation and Public Expenditures: Evidence from US Municipalities and School Districts, 19702000. *Review of Economics and Statistics*, 95(October):1291–1302.

Boyce, J. and Voirnovytskyy, M. (2010). Economic inequality and environmental quality: evidence of pollution shifting in Russia. *Working Paper*.

Boyce, J., Zwickl, K., and Ash, M. (2016). Measuring environmental inequality. *Ecological Economics*, 124.

Boyce, J. K. (1994). Inequality as a Cause of Environmental Degradation. *Ecological Economics*, 11(3).

Brännlund, R. and Ghalwash, T. (2008). The income-pollution relationship and the role of income distribution: An analysis of Swedish household data. *Resource and Energy Economics*, 30:369–387.

Brulle, R. J. and Pellow, D. N. (2006). Environmental justice: human health and environmental inequalities. *Annual review of public health*, 27(102):103–124.

Bryant, B. and Mohai, P. (1992). Environmental racism: reviewing the evidence. In Bryant, B. and Mohai, P., editors, *Race and the Incidence of Environmental Hazards: A Time for Discourse*, page 16376. Westview.

Brzezinski, M. (2013). Asymptotic and bootstrap inference for top income shares. *Economic Letters*, 120(1).

Burkhauser, R., Feng, S., and Jenkins, S. (2009). Using the p90/10 index to measure us inequality trends with current population survey data: A view from inside the census bureau vaults. *Review of Income and Wealth*, 55(1).

Burkhauser, R., Feng, S., Jenkins, S., and Larrimore, J. (2011). Estimating trends in us income inequality using the current population survey: The importance of controlling for censoring. *Journal of Economic Inequality*, 9(1):373–415.

Caiazzo, F., Ashok, A., Waitz, I. A., Yim, S. H., and Barrett, S. R. (2013). Air pollution and early deaths in the united states. part i: Quantifying the impact of major sectors in 2005. *Atmospheric Environment*, 79.

Chavis, B. and Lee, C. (1987). Toxic wastes and race in the united states. Technical report, United Church Christ.

Clark, L. P., Millet, D. B., and Marshall, J. D. (2014). National Patterns in Environmental Injustice and Inequality: Outdoor NO2 Air Pollution in the United States. *PloS one*, 9(4).

Currie, J. (2011). Inequality at Birth: Some Causes and Consequences. *American Economics Review*, 101(3).

Currie, J. and Walker, R. (2011). Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal: Applied Economics*, 3(1):65–90.

Daly, M. and Wilson, D. (2013). Inequality and mortality: New evidence from u.s. county panel data. *Working Paper*.

Diaz-Bazan, T. (2015). Measuring inequality from top to bottom. *Working Paper*.

DiNardo, J., Fortin, N., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5).

Downey, L. (2007). US Metropolitan Area Variation in Environmental Inequality Outcomes. *Urban Studies*, 44(5-6):953–977.

Drabo, A. (2011). Impact of income inequality on health: Does environment quality matter? *Environment and Planning A*, 43:146–165.

Dube, A. (2013). Minimum wages and the distribution of family incomes. *Working Paper*.

Enamorado, T., Lopez-Calva, L.-F., Rodriquez-Castelan, C., and Winkler, H. (2014). Income Inequality and Violent Crime: Evidence from Mexico's Drug War. *Working Paper*.

Essama-Nssah, B. and Lambert, P. J. (2012). Influence functions for policy impact analysis. In Bishop, J. A. and Salas, R., editors, *Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber*. Emerald Group.

Essama-Nssah, B. and Lambert, P. J. (2016). Counterfactual decomposition of pro-poorness using influence functions. *Journal of Human Development and Capabilities*, 17(1):74–92.

Firpo, S., Fortin, N., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Firpo, S., Fortin, N., and Lemieux, T. (2011). Decomposition Methods In Economics. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*.

Flaichaire, E. and Davidson, R. (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics*, 141.

Florida, R. and Mellander, C. (2013). The Geography of Inequality. *Working Paper*.

Fowlie, M., Holland, S., and Mansur, E. (2012). What do emissions markets deliver and to whom? evidence from southern california's nox trading program. *American Economic Review*, 102.

Frank, M. W. (2009). Inequality and Growth in the United States: Evidence From a New State-Level Panel of Income Inequality Measures. *Economic Inquiry*, 47(1):55–68.

Gaskin, D. J., Zare, H., Haider, A. H., and LaVeist, T. A. (2015). The quality of surgical and pneumonia care in minority-serving and racially integrated hospitals. *Health Services Research*.

Gimpelson, V. and Treisman, D. (2015). Misperceiving inequality. *Working Paper*.

Glaeser, E. L., Resseger, M., and Tobio, K. (2009). Inequality in Cities. *Journal of Regional Science*, 49(4):617–646.

Golley, J. and Meng, X. (2012). Income inequality and carbon dioxide emissions: The case of Chinese urban households. *Energy Economics*, 34(6):1864–1872.

Groseclose, T., Levitt, S., and Snyder, J. (1999). Comparing interest group scores across time and chambers: Adjusted ada scores for the u.s. congress. *American Political Science Review*, 93.

Harper, S., Ruder, E., Roman, H. a., Geggel, A., Nweke, O., Payne-Sturges, D., and Levy, J. I. (2013). Using inequality measures to incorporate environmental justice into regulatory analyses. *International journal of environmental research and public health*, 10(9):4039–59.

Heckley, G., Gerdtham, U.-G., and Kjellsson, G. (2016). A general method for decomposing the causes of socioeconomic inequality in health. *Journal of Health Economics*, 48.

Heerink, N., Mulatu, a., and Bulte, E. (2001). Income inequality and the environment: Aggregation bias in environmental Kuznets curves. *Ecological Economics*, 38(3):359–367.

Jenkins, S., Burkhauser, R., Feng, S., and Larrimore, J. (2011). Measuring inequality using censored data: a multiple-imputation approach to estimation andinference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(1):63–81.

Kakwani, N. C. (1977). Measurement of tax progressivity: An international comparison. *The Economic Journal*, 87(345):71–80.

Kolm, S.-C. (1976). Unequal inequalities. I. *Journal of Economic Theory*, 12(3):416–442.

Kovacevic, M. and Binder, D. (1997). Variance estimation for measures of income inequality and polarization - the estimating equations approach. *Journal of Official Statistics*, 13(1).

Lamsal, L., Martin, R., van Donkelaar, A., Boersma, E. C. R., Dirksen, R., Luo, C., and Wang, Y. (2010). Indirect validation of tropospheric nitrogen dioxide retrieved from the omi satellite instrument: Insight into the seasonal variation of nitrogen oxides at northern mid-latitudes. *Journal of Geophysical Research*, 115.

Lamsal, L., Martin, R., van Donkelaar, A., Steinbacher, M., Celarier, E., Bucsela, E., Dunlea, E., and Pinto, J. (2008). Ground-level nitrogen dioxide concentrations inferred from the satellite-borne ozone monitoring instrument (omi). *Journal of Geophysical Research*, 113.

Larrimore, J., Burkhauser, R., Feng, S., and Zayatz, L. (2008). Consistent cell means for topcoded incomes in the public use march cps (1976-2007). *Journal of Economic and Social Measurement*, 33:89–128.

Levinson, A. and O'Brien, J. (2015). Environmental Engels Curves. *Working Paper*.

Magnani, E. (2000). The environmental Kuznets Curve, environmental protection policy and income distribution. *Ecological Economics*, 32(3):431–443.

Maguire, K. and Sheriff, G. (2011). Comparing distributions of environmental outcomes for regulatory environmental justice analysis. *International journal of environmental research and public health*, 8(5):1707–26.

Majumder, A. and Chakravarty, S. R. (1990). Distribution of personal income: Development of a new model and its application to u.s. income data. *Journal of Applied Econometrics*, 5(2):189–196.

Matlack, J. L. and Vigdor, J. L. (2008). Do rising tides lift all prices? income inequality and housing affordability. *Journal of Housing Economics*, 17(3):212 – 224.

McCarty, N., Poole, K., and Rosenthal, H. (1997). *Income Redistribution and the Realignment of American Politics*. AEI Press, Washington, DC.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 52(3):647–663.

McDonald, J. B. and Ransom, M. (2008). The generalized beta distribution as a model for the distribution of income: Estimation of related measures of inequality. In Chotikapanich, D., editor, *Modeling Income Distributions and Lorenz Curves*, pages 147–166. Springer New York, New York, NY.

Mellor, J. and Milyo, J. (2002). Income inequality and health status in the united states: Evidence from the current population survey. *The Journal of Human Resources*, 37(3):510–539.

Mills, J. and Zandvakili, S. (1997). Statistical inference via bootstrapping for measures of inequality. *Journal of Applied Econometrics*, pages 133–150.

Mohai, P., Pellow, D., and Roberts, J. T. (2009). Environmental Justice. *Annual Review of Environment and Resources*, 34:405–430.

Moller, S., Alderson, A., and Nielsen, F. (2009). Changing Patterns of Income Inequality in US Counties, 197020001. *American Journal of Sociology*, 114(4):1037–1101.

Morello-Frosch, R. and Jesdale, B. M. (2006). Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in U.S. metropolitan areas. *Environmental health perspectives*, 114(3):386–393.

Morello-frosch, R., Jr, M. P., Porras, C., and Sadd, J. (2002). Environmental Justice and Regional Inequality in Southern California : Implications for Future Research. 110(April):149–154.

Moyes, P. (1987). A new concept of lorenz domination. *Economics Letters*, 23:203–207.

Neumayer, E. (2004). The environment, left-wing political orientation and ecological economics. *Ecological Economics*, 51:167–175.

Peters, D. J. (2013). American income inequality across economic and geographic space, 1970-2010. *Social Science Research*.

Piketty, T. and Saez, E. (2003). Income inequality in the United States, 1913-1998. *Quaterly Journal of Economics*, 118(1).

Piketty, T., Saez, E., and Stantcheva, S. (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, 6(1):230–71.

Ravallion, M., Heil, M., and Jalan, J. (2000). Carbon emissions and income inequality. *Oxford Economic Papers*, 52:651–669.

Reiter, J. (2003). Inference for partially synthetic, public use microdata sets. *Survey Methodology*, 29.

Scruggs, L. A. (1998). Political and economic inequality and the environment. *Ecological Economics*, 26:259–275.

Sheriff, G. and Maguire, K. (2014). Ranking Distributions of Environmental Outcomes. *Working Paper*.

Shorrocks, A. (1983). Ranking income distributions. *Economica*, 50.

Sommeiller, E. and Price, M. (2014). The increasingly unequal states of america income inequality by state, 1917 to 2011.

Torras, M. and Boyce, J. K. (1998). Income, inequality, and pollution: a reassessment of the environmental Kuznets Curve. *Ecological Economics*, 25(2):147–160.

van Donkelaar, A., Martin, R., Brauer, M., Hsu, N. C., Kahn, R., Levy, R., Lyapustin, A., Sayer, A., and Winker, D. (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental Science And Technology*.

Voorheis, J. (2016). Income inequality and carbon emissions: Evidence from state-level data. *Working Paper*.

Voorheis, J., McCarty, N., and Shor, B. (2015). Unequal incomes, ideology and gridlock: How rising inequality increases political polarization. *Working Paper*.

Wilfling, B. (1996). Lorenz ordering of generalized beta-ii income distributions. *Journal of Econometrics*, 71(12):381 – 388.

Wolverton, A. (2009). Effects of Socio-Economic and Input-Related Factors on Polluting Plants' Location Decisions. *The B.E. Journal of Economic Analysis & Policy*, 9(1):1–32.

Zhu, R. (2016). Wage differentials between urban residents and rural migrants in urban china during 20022007: A distributional analysis. *China Economic Review*, 37:2 – 14. Special Issue on Human Capital, Labor Markets, and Migration.

Zwickl, K., Ash, M., and Boyce, J. (2014). Regional variation in environmental inequality: Industrial air toxics exposure in US cities. *Working Paper*.

Zwickl, K. and Moser, M. (2015). Informal environmental regulation of industrial air pollution: Does neighborhood inequality matter? *Working Paper*.