Every Breath You Take — Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970*

Adam Isen
Office of Tax Analysis
Department of the Treasury
adam.isen@treasury.gov

Maya Rossin-Slater
University of California, Santa Barbara
mrossin@econ.ucsb.edu

W. Reed Walker
University of California, Berkeley
rwalker@berkeley.edu

June 2013

Abstract

This paper examines the long-term impacts of *in-utero* and early childhood exposure to ambient air pollution on adult labor market outcomes. We take advantage of a new administrative data set that is uniquely suited for addressing this question because it combines information on individuals' quarterly earnings together with their counties and dates of birth. We use the sharp changes in ambient air pollution concentrations driven by the implementation of the 1970 Clean Air Act Amendments as a source of identifying variation, and we compare cohorts born in counties that experienced large changes in total suspended particulate (TSP) exposure to cohorts born in counties that had minimal or no changes. We find a significant relationship between TSP exposure in the year of birth and adult labor market outcomes. A 10 unit decrease in TSP in the year of birth is associated with a 1 percent increase in annual earnings for workers aged 29-31. Most, but not all, of this effect is driven by an increase in labor force participation. In present value, the gains from being born into a county affected by the 1970 Clean Air Act amount to about $4,300 in lifetime income for the 1.5 million individuals born into these counties each year.

JEL: H40, H51, I12, I14, J17, J18, J31, Q51, Q53, Q58

*This paper has been previously circulated under the title "Does Improved Air Quality at Birth Lead to Better Long-Term Outcomes? Evidence from the Clean Air Act of 1970". We would like to thank Doug Almond, Michael Anderson, David Card, Janet Currie, Lucas Davis, Olivier Deschenes, Will Dow, Ilyana Kuziemko, and Matt Neidell as well as seminar participants at Columbia, Georgia State, UC-Berkeley, the Robert Wood Johnson Foundation, and the AEA 2013 meetings for valuable comments and suggestions. All results have been reviewed to ensure that no confidential information is disclosed. This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following NSF Grants SES-9978093, SES-0339191 and ITR-0427889; NIA Grant AG018854; and grants from the Alfred P. Sloan Foundation. Isen acknowledges additional support from the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090015 of the U.S. Department of Education. Walker acknowledges additional support from the Robert Wood Johnson Foundation and the University of California Center for Energy and Environmental Economics. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Census Bureau or the Robert Wood Johnson Foundation. We thank David Silver for helpful research assistance.

1 Introduction

The desire to protect human health and welfare motivates much of modern environmental regulation.[1] While there is growing evidence in both epidemiology and economics pointing to the contemporaneous influences of ambient air pollution on population health and other measures of welfare (see Graff-Zivin and Neidell, Forthcoming for a recent review), there is a relative dearth of empirical evidence on the long-run and cumulative impacts of environmental toxins. We view this gap in the literature as particularly important because research has suggested a critical link between population health and wealth throughout the life cycle (Currie, 2009).Therefore, contemporaneous measures of the dose-response relationship between environmental conditions and health outcomes may substantially underestimate the total welfare impact of environmental toxins.

This paper provides some of the first quasi-experimental evidence linking short-run environmental exposure in very early childhood to life-long measures of well-being. We explore this topic by focusing on a policy experiment in the early 1970's that generated large changes in ambient pollution levels in hundreds of counties in the United States. We then examine whether cohorts that were born just before and just after these large changes in air pollution exhibit persistent differences in outcomes measured 30 years after birth. Our focus on pollution exposure in early periods of life is motivated by the emerging empirical evidence on the fetal origins of life-long well-being (Barker, 1990; Almond and Currie, 2011) combined with the mounting evidence on the particularly severe impacts of pollution on infant and fetal health.[2]

We combine this policy experiment with newly available administrative data from the U.S. Census Bureau's Longitudinal Employer Household Dynamics (LEHD) file that allows us to observe adult outcomes linked to location and exact date of birth for 5.7 million individuals born around the time of the policy experiment. We focus on measures of labor market performance at age 30 that broadly encompass (i) changes to cognitive and non-cognitive skill formation that may have been "imprinted" in early childhood, (ii) any persistent health effects attributable to early-life air pollution exposure, and (iii) any reinforcing or compensatory parental investments.[3] As such, our outcomes represent quantifiable summary measures that may be particularly relevant for cost-benefit calculations in environmental policy design.

The policy experiment in the paper stems from the introduction of the 1970 Clean Air Act Amendments (CAAA), which imposed county-level restrictions on the maximum-allowable concentrations of total suspended particulates (TSP). As a result, the set of counties that exceeded these new restrictions (nonattainment counties) were forced to reduce their TSP concentrations, while counties that had air pollution levels below the regulatory ceiling (attainment counties) were not legally required to change their TSP concentrations. This legislation induced substantial variation in county-level pollution changes over this time period that has been previously used to study the effects of air pollution on infant mortality (Chay and Greenstone, 2003a), adult mortality (Chay, Dobkin, and Greenstone, 2003), and fetal mortality (Sanders and Stoecker, 2011a). We use this variation to estimate whether cohorts exposed to lower levels of ambient air pollution *in-utero* and in early childhood exhibit improved labor market outcomes measured 30 years later. Our baseline empirical specification compares cohorts of individuals born just before and after the the mandated improvements in air quality in nonattainment counties,

[1]See, for example, the Environmental Protection Agency's (EPA) mission statement at: http://www2.epa.gov/aboutepa/our-mission-and-what-we-do (accessed on June 13, 2013).

[2]The "fetal origins hypothesis", originally put forth by British epidemiologist David J. Barker, argues that poor nutrition *in-utero* "programs" the fetus to have metabolic characteristics that can lead to future disease in adulthood. For recent evidence on the link between pollution and infant/fetal health, see, for example: Chay and Greenstone, 2003a; Chay and Greenstone, 2003b; Currie and Neidell, 2005; Currie, Neidell, and Schmieder, 2009; Currie and Walker, 2011; Sanders and Stoecker, 2011a.

[3]See Becker and Tomes (1976) for economic theory regarding parental responses to initial endowments, Almond and Mazumder (2012) for recent empirical evidence on parental responses with regards to the fetal origins hypothesis, and Gelber and Isen (2013) for some related empirical evidence on complementarity in schooling and parental investment.

using cohorts born in attainment counties as a counterfactual control group. While nonattainment status is not randomly assigned, we show that observable characteristics of nonattainment and attainment counties in the years prior to regulation are similar in both levels and, more importantly, trends.

We find a strong association between ambient air pollution in a cohort's year of birth and labor market outcomes measured 30 years later. We first show that nonattainment status is associated with an over ten percent reduction in ambient TSP levels in the years after the regulation went into effect. We then show that this regulation-induced reduction in air pollution is associated with an increase in the labor force participation rate for affected cohorts. This estimated impact translates to an increase in mean annual earnings of about one percent. Assuming a constant earnings effect over the lifecycle, our results suggest that the cumulative lifetime income gain is approximately $4,300 in present value terms (using a 5% annual discount rate). This calculation implies that the present discounted total wage bill attributable to the improvements in early life air quality amounts to about $6.5 billion for each affected cohort. We view these estimates as lower bounds on the true value due to various sources of potential bias that would tend to attenuate our baseline estimates. Nevertheless, our estimates suggest that the long-run impacts of environmental exposure may be as large or larger than the short-run impacts on infant mortality examined in previous research (e.g., Chay and Greenstone, 2003a).

This paper provides three primary contributions: First, prior literature estimating the health effects of the CAAA typically focuses on contemporaneous changes in infant health. While infant mortality is an important outcome to study, it reflects some of the most severe consequences of adverse environmental conditions. There may be other consequences for individuals that survived, and, as human capital is an engine for long-run economic growth (Romer, 1986; Schultz, 1961), in aggregate these effects may be larger and far more long-lasting than those associated with infant mortality gains (Graff-Zivin and Neidell, Forthcoming). Although there is some evidence of a contemporaneous relationship between pollution and economic outcomes[4], there is very little work that examines how the short-run benefits of environmental policy may persist in the long run.[5]

Second, we provide additional quasi-experimental empirical support for the theory of fetal origins and early-life determinants of long-run outcomes. Much of the existing empirical literature in this area focuses on rare natural disasters, disease outbreaks, or famines, which are difficult to forecast or protect against (Almond, 2006; Almond, Edlund, and Palme, 2009; Almond, Edlund, Li, and Zhang, 2010).[6] In contrast, we examine the long-run returns to environmental regulation, an intervention over which policy makers have direct control. The dose-response relationship between ambient air pollution and long-run labor market performance is an important policy parameter for which we have very few estimates. Moreover, since less advantaged individuals live in more

[4]For example, Hanna and Oliva (2011) examine labor supply, while Graff-Zivin and Neidell (2013) study labor productivity. Studies also show that contemporaneous pollution exposure can affect human capital accumulation by increasing school absenteeism (Ransom and Pope, 1992; Gilliland, Berhane, Rappaport, Thomas, Avol, Gauderman, London, Margolis, McConnell, Islam, et al., 2001; Currie, Hanushek, Kahn, Neidell, and Rivkin, 2009), and impairing cognitive performance on high-stakes tests (Lavy, Ebenstein, and Roth, 2012). There is also a possibility that pollution can affect adult income if parents have to forego work to take care of asthmatic children (Currie, Hanushek, Kahn, Neidell, and Rivkin, 2009).

[5]Within the United States, we are only aware of two papers that study these questions, although they focus on non labor market outcomes. Sanders (2012) analyzes the relationship between early-life air pollution and high school test scores in Texas, while Reyes (2007) examines the effects of early-life lead exposure on young adult crime. However, an important limitation of both studies is the lack of information on place of birth. As a result, Sanders (2012) effectively assigns birth location based on county of high school attendance, while Reyes (2007) assigns exposure based on state of crime occurrence around age 20. These analyses may therefore be affected by bias from endogenous mobility responses and measurement error. Nonetheless, these studies suggest that there might be an earnings effect of pollution. Outside the U.S., a few recent studies have also estimated the impacts of early-life lead exposure on adult outcomes in Sweden (Nilsson, 2009) and Chile (Rau, Reyes, and Urzua, 2013). However, these studies are limited in their ability to directly observe labor market outcomes (specifically Rau, Reyes, and Urzua, 2013), while also focusing on a very different and far more toxic pollutant (lead) in a context outside of the United States.

[6]A recent exception includes work by Hoynes, Schanzenbach, and Almond (2012), who study the long-term consequences of the roll-out of the Food Stamps program.

polluted areas, our results may have important implications for characterizing disparities in earnings as a function of the early childhood environment.

Third, this paper introduces a new resource for studying long-term outcomes in the United States: the Longitudinal Employer-Household Dynamic Files from the U.S. Census Bureau. Previous work focusing on long-run implications of early-life interventions in the United States is typically challenged by the fact that very few publicly available datasets contain detailed information on birth location linked to long-run outcomes (the few that do are of limited use in this context because of small sample sizes).[7] In contrast, our administrative earnings data contain the near-universe of the employed workforce, with precise information on both location and date of birth.

The rest of the paper proceeds as follows: Section 2 presents a basic conceptual framework to help guide the empirical analysis. Section 3 provides a brief overview of the CAAA and related literature. Section 4 provides a description of the data used in the analysis, with a more complete discussion found in Appendix B.1. Section 5 outlines the various econometric models used, and Section 6 discusses the results of those models. Sections 7 and 8 discusses the implications of our findings and concludes, respectively.

2 Conceptual Framework

How might early-life exposure to ambient air pollution affect adult outcomes? An existing literature in biology, epidemiology, and (more recently) economics provides motivation for treating exposure to air pollution as an input into an individual's health stock. While precise biological mechanisms underlying the relationship between environmental toxins and human health remain unclear, there is consensus among researchers that air pollution can affect respiratory function, lung development, and further internal problems as particulates can be transferred from lungs into the bloodstream. Fetal health can be impacted as a result of health impairments to the mother as well as direct effects due to particulates transferred to the fetus through the bloodstream.[8] As a shock to *in-utero* and early childhood health, exposure to environmental toxins can further influence an individual's cognitive and physiological development, thereby affecting returns to human capital investments. Additionally, there may also be a direct effect of pollution on human capital accumulation as health can affect an individual's occupational choice, his decision to enter or exit the labor force, or the benefit-cost calculation associated with his educational attainment decision.

We present a simple framework designed to formalize the relationship between early-life inputs into health and long-run human capital accumulation while also motivating our empirical approach.[9] Let an individual's health stock be a function of inputs during two time periods: $h = h(I_1, I_2)$, where I_t are inputs in each period t, and $t = 1, 2$. In our case, we can think of $t = 1$ as representing early childhood, while $t = 2$ as representing the rest of life up to the point of observation.

An individual's earnings are a function of his health stock h and his education level e, where education also depends upon the health stock. Formally, $y = y(e, h) = y(e(h(I_1, I_2)), h(I_1, I_2))$, where y represents earnings

[7]The restricted version of the Panel Study of Income Dynamics (PSID) is the best currently available dataset that gives information on location and date of birth linked to long-run earnings (Johnson and Schoeni, 2007; Hoynes, Schanzenbach, and Almond, 2012; Johnson, 2011).

[8]For more information on the link between air pollution and health, see World Health Organization (2005), available at: http://www.euro.who.int/__data/assets/pdf_file/0010/74728/E86575.pdf.

[9]Our framework is closely related to the model described in Bleakley (2010). This framework abstracts away from modeling parental investments in response to health shocks (Becker and Tomes, 1976), or the dynamic complementarities between shocks and investments across different time periods (Cunha and Heckman, 2007). We instead focus on the reduced-form relationship between early-life inputs into health and adult earnings because this is what we can measure in our data. The framework could also model pollution as a more general direct input into earnings without hypothesizing that the mechanism occurs solely through the health stock.

and e represents years of schooling. We are interested in the impact of health inputs in period 1 (I_1) on earnings. The CAAA improved ambient levels of pollution for certain counties, and the cohorts born just before the CAAA had relatively "low-quality" inputs in period 1 (i.e. high pollution levels) versus relatively "high-quality" inputs in period 2 (i.e. lower pollution levels). In contrast, the cohorts born just after the CAAA in affected counties had "high-quality" inputs in both periods. Thus, the thought experiment in the paper is to use the variation in CAAA implementation to isolate the impact of changes to I_1 investments on adult earnings:

$$\frac{\partial y}{\partial I_1} = \left[\frac{\partial y}{\partial e} \times \frac{\partial e}{\partial h} \times \frac{\partial h}{\partial I_1} + \frac{\partial y}{\partial h} \times \frac{\partial h}{\partial I_1}\right] \tag{1}$$

If we assume that earnings are observed after an individual has completed his desired schooling, the partial derivatives in equation (1) should be evaluated at the optimal schooling level, e^*. The envelope theorem tells us $\frac{\partial e}{\partial h}|_{e^*}$ must equal zero such that Equation (1) simplifies to:[10]

$$\frac{\partial y}{\partial I_1} = \frac{\partial y}{\partial h} \times \frac{\partial h}{\partial I_1} \tag{2}$$

Hence, the long-run impact of changes in period 1 investments consist not only of the direct health impacts of shocks to period 1 investments $\frac{\partial h}{\partial I_1}$ (i.e. that which is typically measured in dose-response, environmental health estimates) but also the interactive effect of how these early life health impacts influence long run earnings or human capital accumulation $\frac{\partial y}{\partial h}$.

The goal of the rest of the paper is to deliver estimates of $\frac{\partial y}{\partial I_1}$, where the change in period 1 inputs stems from changes in the levels of ambient air pollution experienced by cohorts surrounding the 1970 CAAA. The precise details of both the research design and econometric strategy are described more fully below.

3 The Clean Air Act

The Clean Air Act regulates air pollution in the United States and is the largest environmental program in the country. The Clean Air Act requires the Environmental Protection Agency (EPA) to develop and enforce regulations to protect the general public from exposure to airborne contaminants that are known to be hazardous to human health. The Act was passed in 1963 and significantly amended in 1970, 1977, and 1990. The enactment of the Clean Air Act Amendments of 1970, by authorizing federal regulations to limit emissions, resulted in a major shift in the federal government's role in air pollution control. In doing so, the EPA established national ambient air quality standards (NAAQS), which specify the minimum level of air quality acceptable for six criteria air pollutants.[11]

In a series of pathbreaking papers, Henderson (1996) first showed how nonattainment designations led to large changes in ambient air concentrations; Chay and Greenstone (2003a, 2005) then showed how these regulatory induced changes may be exploited as a source of quasi-experimental variation to better understand the underlying relationships between ambient air pollution, infant health, and willingness to pay for air quality more generally. Chay and Greenstone (2003a) showed how the 1970 Clean Air Act Amendments led to large reductions in ambient air pollution in newly regulated counties, and they then showed how this reduction was causally related to the significant decrease in the infant mortality rate in these affected set of counties. We ask whether these

[10]The basic logic of the envelope theorem suggests that while the change in period t investments may induce changes in behavior, these behavioral responses cannot have a first order effect on income — if they did, the individuals would not have been optimizing.

[11]These pollutants consist of sulfur dioxide (SO_2), particulates (TSP, PM2.5, and PM10), nitrogen dioxide (NO_2), carbon monoxide (CO), ozone, and lead.

same changes in air pollution in the 1970's led to any long run consequences for the cohorts of individuals born in these counties.

While the Chay and Greenstone (2003a, 2005) papers serve as the underlying basis for our research design, they also presage potential sources of confounders with regards to earnings determination in later life. For example, the reduction in the infant mortality rate in the years after the CAAA suggests that the wage distribution, conditional on survival, may be influenced by the "marginal" births that were saved due to the air quality improvements. Moreover, Chay and Greenstone (2005) show how these regulation-induced improvements in air quality led to increases in housing prices in communities most affected by the changes. These findings at least raise the possibility that if households have heterogeneous preferences for air quality that are correlated with the underlying health endowment of their children, then households may differentially sort after the improvements in air quality so that the cohorts born before and after the changes are of different underlying quality. Our research design and subsequent analysis intends to address these and many other sources of confounding variation. The exact details are specified in subsequent sections.

4 Data

The primary source of data for this project comes from the Longitudinal Employer Household Dynamics File (LEHD). We link this dataset to various other proprietary and public use datasets described below. Additional details can be found in Appendix B.1.

Longitudinal Employer Household Dynamics File (LEHD)

The Census Bureau's LEHD file provides administrative, quarterly earnings records for over 90% of the United States workforce.[12] The earnings records correspond to the report of an individual's UI-covered earnings by an employer in a given quarter. While the LEHD earnings records are fairly comprehensive, notable exceptions include the self-employed, agricultural workers, and some state, local, and federal employees.

The LEHD provides longitudinal employment and earnings histories for workers along with some basic demographic characteristics such as sex and race.[13] Crucially for our analysis, the LEHD also provides information on both the place and exact date of birth. The place of birth variable in the LEHD is a string variable detailing in most cases the city and state of birth. We developed a matching algorithm to connect this string variable to the Census Bureau database of places, counties, and minor civil divisions as well as the United States Geological Survey's Geographic Names Information System (GNIS) file. This provides us with a crosswalk between the LEHD place of birth string variable and County FIPS codes. A full description of the matching algorithm is detailed in Appendix B.2. Over 95 percent of the individuals in the LEHD file were matched to their county of birth. Lastly, we use the Bureau of Economic Analysis "county-equivalent" as our baseline definition of a county, both to maintain a consistent definition of counties throughout our sample frame as well as to match the BEA's Regional Economic Information System (REIS) data described below.

While the LEHD provides extraordinary levels of detail for a large fraction of the United States workforce, several limitations bear mention. Firstly, the LEHD is assembled by combining various state's administrative earnings records. This means that states have varying degrees of temporal coverage in the main dataset, with most states entering the sample by the late 1990s. The second challenge is that it is not possible to distinguish between unemployment and non-participation. For example, we would observe a missing earnings record for an

[12] See Abowd, Stephens, Vilhuber, Andersson, McKinney, Roemer, and Woodcock (2008) and McKinney and Vilhuber (2008) for a comprehensive discussion of both the construction and contents of the LEHD files.

[13] The race variable is divided into 6 mutually exclusive categories: White, Black, Other, Asian, Hispanic, and American Indian.

individual both if he/she were to move to a state not covered in the LEHD and if he/she became unemployed or self-employed in a given year. Since our treatment variable may also covary with this form of sample attrition, we are careful to construct a sample which tries to address such concerns.

Specifically, we limit our sample to the 24 states which continuously contain earnings records between 1998-2007.[14] Furthermore, we limit ourselves to examining earnings records of individuals who were *born* in one of those 24 states.[15] Workers are able to move from their state of birth to other states, but they will only be in our sample if they ever work in one of these 24 states between 1998-2007.

Our research design exploits the sharp changes in ambient TSP exposure experienced as a result of the 1970 CAAA. Since the data on labor market outcomes begins in 1998, we are limited with respect to the ages for which we may observe cohorts born in the 1970s. The correlation between annual earnings and lifetime income rises rapidly as individuals enter the labor market and begins to stabilize only in the late twenties (Chetty, Friedman, and Rockoff, 2011). Therefore, our focus is on labor market outcomes of affected cohorts well into their 20s or early 30s. Our main outcomes of interest are the average annual number of quarters employed and the average annual earnings of an individual between the ages of 29-31. We study labor market outcomes averaged over a set of years rather than outcomes in a particular year in order to (i) minimize the residual variance in the observed unemployment/earnings distribution, and (ii) ameliorate concerns that any effects we see are driven by a contemporaneous economic shock in one particular earnings year. Our data allow for pre-post comparisions of cohorts of individuals born in the treated counties before the sharp changes in TSP took place, relative to those born after. Additionally, in several robustness checks, we also study outcomes measured at all ages between 28 and 32.

For each individual in our sample, we calculate the years at which they turn 29 to 31, and we search for their earnings record in the employment history file. We take the combined earnings for a worker in a given year, adding over both employers and states (in the event of multiple job spells within a year). We also calculate the number of quarters for which that worker has positive earnings in a given year. In the event that a worker works in more than one state in a given year, we assign the worker to be working in the state for which he/she has the maximum earnings in that year. If the earnings record is missing for a particular age category (i.e. because the worker is unemployed or has attritted from the data), we estimate specifications where we either keep this earnings record as missing or we replace it with a zero. For the workers for which we do not observe an earnings records in "the sample period", we impute the current state of work using the state for which the worker has the most quarterly earnings observations either in future or previous years. In the event that a worker has the same earnings observations in more than one state, we randomly assign the worker to one of these states.[16]

We express all monetary variables in 2008 dollars, adjusting for inflation using the Consumer Price Index. For each cohort, we cap earnings at age 28 equivalent \$100,000 allowing for 2% annual growth in earnings in order to limit the influence of outliers.[17] Our baseline sample consists of earnings records from 5.7 million individuals born between the 1969 and 1974. We use these earnings records to construct a balanced panel of county×year cohort data, for 148 counties in our 24 states. The mean earnings between the ages of 29-31 is \$23,563 for those

[14]We exclude the year 2008 from our analysis because some states only have partial quarterly coverage in 2008. The states in our sample include: CA, CO, FL, GA, ID, IL, IN, LA, MD, ME, MT, NC, NJ, NM, OR, RI, TX, WA, WI, WV, TN, SC, NV, and VA. Total non-farm employment in these 24 states accounted for 61% total United States non-farm workforce in 2000.

[15]Formally, we take the full sample of individuals who ever worked in one these 24 states in the years for which the state is covered within the LEHD by pooling over the individuals in the LEHD Individual Characteristics Files for all 24 states.

[16]For the purposes of our analysis, it does not particularly matter where an individual ends up. We are primarily focused on place of birth. However, in some econometric specifications, we control for underlying earnings heterogeneity in a particular work-state×year.

[17]Here, we follow Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan (2011) although the results are robust to windsorizing at other points in the distribution or using raw earnings. Specifically, we cap earnings at \$100,000 for 28 year olds, \$102,000 for 29 year olds, \$104,040 for 30 year olds, \$106,121 for 31 year olds, and \$108,243 for 32 year olds.

individuals born in 1969 (in 2008 dollars).

Additional County×Year Data

We match the LEHD earnings records to the Regional Economic Information System data from the Bureau of Economic Analysis at the "county-equivalent" by birth-year level. This provides us with a variety of time-varying information at the county-level for the year of birth.[18] We make use of the data on population counts, employment, per capita income, and information regarding transfer receipts (i.e. total unemployment spending and total transfer receipts in a county×year).

We match all of our data to data from the universe of individual-level natality and mortality files from the National Center from Health Statistics (NCHS). These data provide a rich source of time-varying information on maternal, paternal, and child characteristics for each birth county and birth year. Moreover, these data allow us to examine how infant health responds to adverse environmental conditions for our particular subsample of states and compare our results to those found in Chay and Greenstone (2003a).

Our measure of pollution comes from the EPA's air pollution monitoring network, which provides annual readings for the universe of air pollution monitors scattered throughout the United States. Following Chay and Greenstone (2003a, 2005), we use this data to construct two measures of county-level air pollution in each year. The first consists of a weighted average of annual air pollution over all monitors, with weights proportional to the number of monitor observations within a given year. The second measure of air pollution consists of the second highest TSP reading in a county (taken over all monitors in a county). These two pollution measures form the basis for the NAAQS standards, central to the Clean Air Act and county nonattainment designations. Specifically, a county is designated as nonattainment if one of the following criteria are met in a given year: (i) the annual geometric mean concentration exceeds 75 $\mu g/m^3$, or (ii) the second-highest daily concentration exceeds 260 $\mu g/m^3$. We use these definitions to classify counties into nonattainment based on their 1970 ambient air quality readings.[19] The empirical analysis restricts the sample to only include monitors which had more than 15 readings in a given year.

Lastly, we bring in data on temperature and precipitation in the county and year of birth from Schlenker and Roberts (2009) to control for potential concomitant variation between ambient pollution levels and weather. Further details about the data may be found in Appendix B.1.

5 Econometric Specification

Our baseline empirical specification examines the effect of ambient TSP exposure in the year of birth on labor market outcomes 29 to 31 years later. Earnings and labor force participation are complex functions of age, experience, education, and various other labor market conditions, both current and past. In addition, costs of living differences across locations imply further earnings heterogeneity based on location. In order to understand the relationship between early-life environmental conditions and long-run outcomes, we need an econometric model that explicitly addresses these different sources of variation in labor market outcomes. For each of the 5.7 million individuals in our sample, we observe the place of birth, date of birth, age 29-31 earnings, and the state of work. We collapse the data to the county×year cohort level, where cohorts are defined by their county and

[18] In subsequent sections, we also control for earnings heterogeneity in the state of work, for which we use state-level REIS data (population weighted). Recall, that we only observe the state in which a worker is currently employed.

[19] As highlighted in Chay and Greenstone (2003a, 2005), the EPA does not maintain historical records of county-level nonattainment status dating back prior to 1978. Thus, we are forced to reconstruct which counties were likely to be in nonattainment based on historical pollution levels.

year of birth. We collapse the data in order to limit the computational burden of working with the full set of individuals, while also estimating the models at the level of variation (i.e. a county×year). We estimate various forms of the following cohort-based econometric model:

$$y^a_{ct} = \beta_0 + \beta_1 TSP_{ct} + X'\Phi + \gamma_c + \eta_{st} + \mu_{ct} \tag{3}$$

where outcome y^a_{ct} is either annual earnings or quarters employed for workers at age a who were born in county c in year t. We also control for a number of time-varying socioeconomic and demographic characteristics $X'\Phi$ in the county and year of birth that may also influence earnings determination and labor force participation at ages 29-31. The exact controls vary across specifications and will be described in more detail both below and in Appendix B.1. All specifications of Equation (3) include county fixed effects γ_c to control for time-invariant, unobserved determinants of labor market outcomes for workers born in a particular county as well as birth-state×year fixed effects η_{st} to control for time-varying determinants of long run outcomes that are common to all individuals born in a particular state×year.

There are many micro-level determinants of labor market outcomes that are effectively ignored when collapsing the data to the birth-county×year level. To control for some of this observed heterogeneity, we estimate Equation (3) via a two-step estimator:[20]

$$\text{Step 1: } y^a_{ic\tilde{s}t} = \Gamma'_{ic\tilde{s}t}\pi + \theta_{ct} + \epsilon_{ic\tilde{s}t} \tag{4a}$$

$$\text{Step 2: } \widehat{\theta_{ct}} = \phi_0 + \phi_1 TSP_{ct} + X'\rho + \gamma_c + \eta_{st} + \nu_{ct} \tag{4b}$$

First, we use the micro-data to estimate employment and earnings regressions using equation (4a), controlling for micro covariates such as race, sex, month of birth using indicator variables. In terms of notation, any variable with a tilde, represents a contemporaneous relationship (i.e. in the year the worker was observed working at age a), while those covariates without a tilde represent information in the year of birth. Equation (4a) also controls for some contemporaneous state-level outcomes to explain some of the labor market outcome heterogeneity in a given work-state×year. All of the control variables, both in the county of birth and state of work, are denoted by the vector $\Gamma_{ic\tilde{s}t}$.[21] In this same regression, we include a full set of birth-county×year indicators θ_{ct}. The coefficient estimates, $\widehat{\theta_{ct}}$, represent covariate-adjusted average labor market outcomes for a birth-county×year after controlling for race, sex, month of birth, and work-state×year covariates. We then use the coefficient estimates, $\widehat{\theta_{ct}}$, from equation (4a) as the dependent variable in equation (4b). Equation (4b) is the exact same regression model as equation (3) with the exception of using the group-level, composition-adjusted averages, rather than the unadjusted group means.[22] Equation (4b) is estimated using weighted least squares, weighting by the group-level cell size.

Exposure to TSP in the year of birth is likely correlated with many observable and unobservable determinants of long-run labor market potential. Including county-cohort fixed effects γ_c will absorb any time-invariant determinants of long run human capital unique to a specific cohort-county, and including cohort-state×year fixed effects will control for transitory determinants of long-run outcomes common to all cohorts within a given state×year. However, there may exist local *and* transitory determinants of long-run outcomes that may also

[20] Similar two-step estimators may be found in Angrist and Lavy (2009), Baker and Fortin (2001), and Donald and Lang (2007).

[21] The exact controls include indicators for race, sex, and month of birth and the following work-state level controls (all in per-capita terms): Supplemental Nutrition Assistance Program (SNAP) spending, Medical benefit spending, public medical care benefits, family assistance spending, personal transfer receipts, income maintenance, and unemployment insurance benefits. The state-level data was constructed by taking population weighted averages over all counties in a state×year from the REIS data.

[22] Composition adjusted earnings measures, such as those used here, can be found in work by Albouy (2009a,b); Notowidigdo (2011); Shapiro (2006).

covary with ambient air pollution. For example, local economic conditions are strong predictors of ambient TSP (Chay and Greenstone, 2003b) but also have been shown to correlate with both infant health and fertility decisions (Dehejia and Lleras-Muney, 2004; Lindo, 2011; Schaller, 2012), as well as long-run outcomes such as mortality (Van den Berg, Doblhammer-Reiter, and Christensen, 2011; Van Den Berg, Lindeboom, and Portrait, 2006). Any unobserved transitory, local shocks that covary with both TSP and long-run outcomes will lead to bias in the OLS estimate of η_1. In addition, exposure to TSP is measured with error for most individuals. If this measurement error is classical, it will tend to attenuate estimates of η_1, especially in models with additional covariates (Griliches, 1986).

A common solution to both the measurement error problem and the possible endogeneity of pollution exposure is to find or construct an instrumental variable that is correlated with changes in TSP but is unrelated to long-run determinants of labor market outcomes, except through the instrument's effect on TSP. In the case of either endogeneity of TSP_{ct} and/or classical measurement error, the assumptions required for a consistent estimate of β_1 in equation (4b) are stronger than before. In addition to the traditional OLS assumptions, the instrumental variable must be correlated with the observed measure of pollution but uncorrelated with ν_{ct} in equation (4b).[23]

We exploit the introduction of the 1970 CAAA and county-level nonattainment status as an instrumental variable for ambient TSP.[24] Several papers have shown the strength of the first stage relationship between county-level nonattainment status and ambient levels of air pollution (see e.g. Auffhammer, Bento, and Lowe (2009); Chay and Greenstone (2003a, 2005); Grainger (2012); Henderson (1996); Sanders and Stoecker (2011a)). In subsequent sections, we present similar evidence that nonattainment designation led to significant and persistent declines in ambient TSP concentrations in the years after nonattainment went into place. We also present evidence that nonattainment designation may satisfy the excludability condition required for a consistent estimate of β_1. While the identifying assumption is inherently untestable, data from other time periods permit several indirect tests. In particular, we use data from years prior to the change in regulations to examine trends in covariates and outcomes prior to the change in policy, finding little evidence of statistically significant differences.

There are certain reasons to believe that nonattainment status may not perfectly satisfy the exclusion restriction as CAAA enactment plausibly has effects on counties beyond pollution reduction. For example, prior work has shown that while nonattainment designation reduces ambient concentrations of particulates, it does so at the cost of some economic competitiveness (Greenstone, 2002; Greenstone, List, and Syverson, 2012; Walker, 2011, 2012). Therefore, nonattainment may also contribute to declining economic conditions which may impact long-run earnings capacity of children born into those counties.[25] Any impact of nonattainment on the local economy that subsequently affects long-run earnings capacity would likely downwardly bias our results. Nevertheless, caution is warranted, and for all outcomes, we present both the "reduced-form" effect of nonattainment and the 2SLS estimate. By scaling the reduced-form coefficient by the effect of CAAA on pollution abatement, the 2SLS estimate allows readers to interpret the effect of the policy on earnings in terms of a pollution dose-response relationship. We explore these issues more fully in the subsequent analysis.

[23] In the case of heterogeneous effects of pollution on health, we require stronger assumptions than those listed here, namely a monotonicity condition in that the instrument affects pollution in a singular direction (Imbens and Angrist, 1994).

[24] Since the CAAA regulations apply non-linearly in the level of pollution, it is in principle possible to estimate models using a regression discontinuity (RD) design. However, our limited number of sample counties, combined with the lack of sufficient density around the threshold, preclude a formal RD analysis.

[25] The actual impact of nonattainment status on the broader local economy is fairly small. Work in Walker (2012) suggests that the implied impact of nonattainment on county employment is less than 0.7% of the total workforce.

6 Results

Fixed Effects Relationships Between TSP Exposure and Long Run Outcomes

Appendix Table A1 presents fixed effects OLS estimates which characterize the relationship between ambient TSP exposure in the year of birth and labor market outcomes of individuals aged 29-31. These results come from estimating equation (4b), weighted by the number of workers in a county×year cell. We conduct inference using cluster-robust standard errors, clustering by county commuting zone to account for any spatial dependence in nonattainment designations within the same metropolitan area.[26] Panels A and B present results where the dependent variable consists of the average annual quarters of employment $\in [0, 4]$ and mean annual earnings in a county×year, respectively. Columns (1)-(4) vary the control variables that are used in the baseline regression model, whereby the set of controls increases in stringency as one moves from left to right across the table. There is suggestive evidence that TSP exposure in the year of birth is correlated with long-run earnings capacity. However, the results suggest that increases in ambient TSP *improve* long-run outcomes of cohorts born into high TSP counties. As mentioned before, OLS estimates relating TSP to long-run earnings may be confounded by other unobserved factors that covary with pollution such as stronger economic activity and/or per capita income. Moreover, OLS estimates may suffer from various forms of bias associated with measurement error in the independent variable. Thus, we next turn towards specifications that attempt to address both of these concerns.

The CAAA as the Basis for a Research Design

As suggested above, instrumental variables provide a solution to both issues of classical measurement error and the potential endogeneity of TSP exposure. A valid instrumental variable must be strongly correlated with changes in ambient TSP exposure while also remaining excludable in the "structural" regression equation of interest. Appendix Table A2 presents evidence of the first requirement — county-level nonattainment status is a strong predictor of a decline in county ambient TSP concentration. Formally, we augment equation (3) by placing TSP_{ct} on the left hand side, and examining the effect of the regulation as mediated through county nonattainment status:

$$TSP_{ct} = \alpha_0 + \alpha_1(Non_{1970,c} \times 1[\tau > 1971]) + X'\rho + \gamma_c + \eta_{st} + \nu_{ct} \tag{5}$$

The parameter of interest is α_1 which provides a difference-in-differences estimate of the impact of nonattainment designation on county TSP levels in the years after CAAA regulations went into place.

The first term inside the parentheses of equation (5), $Non_{1970,c}$ is a time-invariant county indicator equal to 1 if a county exceeds the NAAQS TSP standards based on its 1970 pollution readings. The second term, $1[\tau > 1971]$, is equal to 1 in the years after the CAAA went into effect. This regression model mimics both the controls and sample used in the baseline OLS regressions from before and the subsequent IV and "reduced-form" regressions below. The regressions are weighted by the number of workers in each county×year cell and the standard errors are clustered by commuting zone.

The results in Appendix Table A2 correspond to estimates of α_1 in equation (5). Consistent with the previous literature, we find a strong relationship between CAAA implementation and ambient concentrations of TSP in

[26]The USDA Economic Research Service used county-level commuting data from the 1990 Census data to create 741 clusters of counties that are characterized by strong commuting ties within CZs and weak commuting ties across CZs (Tolbert and Sizer, 1996). Subsequent researchers have used similar levels of Census geography for economic research on local labor markets (e.g. Autor and Dorn (Forthcoming); Walker (2012)).

nonattainment counties. This relationship is robust across the specifications and suggests that CAAA reduced TSP concentrations by 8-12 μg/m^3. Relative to a mean value of 95.9 μg/m^3, this amounts to about a 10 percent reduction in air pollution for the set of counties in our sample. The regression model from equation (5) implicitly assumes that the CAAA improved air quality instantly, and these improvements lasted forever. In order to understand the dynamics of the process more explicitly, we augment equation (5) into an event-study style regression.[27] Figure 1 plots estimates of the event-time coefficients for newly regulated counties in the years before and after CAAA went into place. This exact regression specification mimics the controls in column (5) of Table A2, although the sample frame is widened to span the years of 1969-1977. Consistent with the results in Table A2, we see a persistent decline in ambient TSP in the years after nonattainment went into place. A useful feature of the event-study specification is the ability to examine trends between treatment and control counties in the years prior to the change in nonattainment status. We see that in the years immediately preceding CAAA initiation, trends between eventual nonattainment and attainment counties evolve similarly.[28] The figure provides suggestive evidence that changes in attainment counties serve as a useful counterfactual for what would have happened to nonattainment counties in the absence of the regulation, a key condition for identification in a difference-in-difference estimator.

Table 1 provides additional suggestive evidence for the validity of our research design. Columns (1) and (2) of Table 1 present a range of observable characteristic means for both attainment and nonattainment counties in 1969, whereas Columns (3) and (4) present the same statistics in log differences between 1969 and 1971. Columns (5) and (6) present p-values from tests of the null hypotheses that the levels and pre-trends in characteristics of attainment and nonattainment counties are not statistically different. While Column (5) makes clear that nonattainment counties are observably different than attainment counties, Column (6) suggests that trends in observable characteristics between attainment and nonattainment counties are similar in the years prior to the 1970 CAAA. Across most specifications, we cannot reject the null hypothesis that the difference in trends is the same. This lends further credibility towards using cohort trends in attainment counties as counterfactuals for cohorts in nonattainment counties. There is one covariate which exhibits significant differences — total transfers per capita. Columns (3) and (4) suggest that nonattainment counties exhibit about 2% more growth in pre-period total per capita transfers. As a result, we attempt to control flexibly for total transfers per capita in all regressions by interacting the pre-determined levels of transfers per capita (i.e. county transfers in 1969) with the 1st-4th order polynomial trends.

The Effect of CAAA on Long Run Outcomes

We next examine how CAAA implementation affected labor market outcomes of individuals born into nonattainment counties 29-31 years later. We modify equation (4b) by replacing the independent variable of interest, TSP_{ct}, with county nonattainment status interacted with an indicator for a post-1971 cohort birth year, $(Non_{1970,c} \times 1[\tau > 1971])$. Panel A of Table 2 presents results using the mean quarters of employment in a

[27]Specifically, we include leads and lags in event-time indicators (i.e. before and after CAAA implementation), and we interact these event-time indicators with the county nonattainment indicator. The exact regression model is the following:

$$TSP_{ct} = \lambda_0 + \sum_{k=1969}^{1977} \lambda_k (Non_{1970,c} \times 1[\tau = k]) + X'\rho + \gamma_c + \eta_{st} + \nu_{ct} \tag{6}$$

In the presence of state×year fixed effects, the baseline event-time indicators are not identified. The coefficients of interest are the λ_k's which provide an estimate as to the time-path of ambient TSP in a newly regulated county before and after nonattainment goes into place. In the presence of county fixed effects, not all of the λ_k's are identified. We make the normalization $\lambda_{1971} = 0$.

[28]The first year for which we have data on air pollution is 1969, which limits our ability to examine pre-trends prior to 1969.

county×year cell ($\in [0, 4]$) as a dependent variable for a variety of econometric specifications.[29] Column (1) shows that cohorts born into nonattainment counties in the years after CAAA went into effect work on average 0.020 quarters more, relative to the counterfactual. Relative to a mean number of employed quarters of 2.74, this amounts to about a 0.7% increase in quarters employed. Column (2) adds in flexible controls for climate and weather in the year of birth to absorb some of the potentially confounding relationships between temperature, precipitation, and ambient air pollution.[30] The results increase in magnitude slightly with statistical precision also increasing. Columns (3) and (4) add in further controls for time-varying birth, mother, and father characteristics in a county×year.[31] The magnitudes remain similar across these additional specifications.

Panel B presents results using mean annual earnings in a county×year as the dependent variable. The results across all columns are similar in magnitude, although some of the specifications are not significant at conventional levels. Relative to baseline earnings in nonattainment counties of $23,623, estimates suggest that CAAA implementation increased the earnings of cohorts born into "cleaner" nonattainment counties by about 1 percent.[32] The estimates in Panel B are slightly larger in magnitude to those in Panel A; an increase in labor force participation by 0.020 quarters is equivalent to about $150 in annual earnings,[33] suggesting that some of the estimated impact in earnings is driven by either intensive margin effects, labor market participation impacts that vary at levels below the quarterly level, and/or extensive margin effects for individuals away from the mean of the earnings distribution.

Appendix Table A3 presents additional estimates, exploring the sensitivity of the baseline earnings estimates. Panel A uses mean log earnings in a county×year as the dependent variable, whereas Panel B explores the sensitivity of the baseline estimates to excluding workers who were non-employed in all 4 quarters of the year. Since the log of zero is undefined, the underlying samples in Panel A and B are identical, and the differences in regression estimates reflect differences associated with functional form of the underlying regression model. Panel A suggests that nonattainment is associated with an approximately 1% increase in annual earnings. The results in Panel B also suggest improvements in earnings, although the effects are smaller in magnitude and are not statistically different from zero. The differences between the log and level specifications in Appendix Table A3 may be due to the approximate log-linear distribution of earnings and/or may suggest that real earnings differences across cohorts/counties may be an important source of unobserved heterogeneity and that log earnings measures are better at addressing such concerns. It may also be that a log-linear model better captures the underlying relationship of interest. Nevertheless, Appendix Table A3 suggests that the earnings effects we see are not entirely driven by workers with zero earnings, providing some reassurance that our baseline estimates are not simply an artifact of sample attrition and/or sample construction.

Figure 2 replicates the event-study research design from equation (6), replacing the dependent variable with

[29] The number of quarters employed is the number of quarters in the calendar year when the individual was earning positive compensation.

[30] Weather controls include a linear, quadratic, and cubic terms in annual county precipitation. We also include flexible temperature controls, calculated as the number of "degree days" in a given county year above 0, 5, 8, 10, 12, 15, 20, 25, 29, 30, 31, 32, 33, 34 degrees Celsius (i.e. 14 separate terms).

[31] Controls in the "Natality Basic" column include a continuous measure of both mother and father education, mother's age, and indicators for marital status of mother, month of the first prenatal care, and an indicator for no prenatal care. Controls for the "Natality Unrestricted" columns include the "Natality Basic controls" in addition to: indicators for years of education of both the mother and father (<12, =12, 13-15, 16+), father's age, indicators for mother's age (10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40+), delivery outside of hospital indicator, physician present at delivery indicator, previous live birth indicator (1, 2+), previous fetal death indicator (1, 2+), last pregnancy was live birth indicator, last pregnancy was fetal death indicator, indicators for 1-11, 12-17, 18 or more months since last live birth, indicators for 1-11, 12-17, 18 or more months since termination of last pregnancy.

[32] The average effect across all columns in Panel B of Table 2 is $259, where the average is taken across columns, weighting by the inverse of the standard error.

[33] This statistic is calculated by noting that 1 quarter is equal to 91.25 days out of the year. The average daily earnings is $64 ($23,623/365). Hence, $64×91.25×0.020=$116.8.

mean annual quarters employed between the ages of 29-31. As before, the event-time coefficients plot the time path of changes for the nonattainment group relative to the attainment group, before and after the regulations go into place. The event-study regression mimics Column (4) in Panel A of Table 2 with respect to control variables; the only difference is that the sample frame is extended to include years 1969-1977. The figure shows a general shift both in trends and levels in the outcome variable as the regulations go into place, with the magnitudes mimicking the magnitudes found in the corresponding table. The temporal dynamics of the changes also match those found in Figure 1, with little evidence of pre-trend differences combined with shifts in the post-period.

Instrumental Variable Estimates of the Effect of TSP on Long Run Outcomes

Panels A and B of Table 3 presents instrumental variable estimates of the effect of TSP exposure in the year of birth on non-employment and mean earnings between the ages of 29-31. The results suggest that a ten-unit increase in ambient TSP exposure in the year of birth reduces average annual earnings at age 30 by around 1%. The magnitude of the IV estimates in Table 3 are of equal and opposite sign to that of the reduced form estimates in Table 2. This is to be expected since the estimated first stage (i.e. Table A2) showed a decrease in ambient TSP of around 10 μg/m^3, and the TSP variable in Panel B is scaled by a factor of 10, representing a 10 μg/m^3 increase in ambient TSP. As before, the point estimates increase in statistical precision as we reduce the residual variance in long-run earnings determination by including additional control variables.

Table 4 presents estimates using labor market outcomes at different ages. Each column corresponds to a different regression using a different age-earnings sample. Column (6) replicates column (4) in Table 2 as a basis for comparison. In each earnings year, the results are qualitatively consistent with the baseline results from before; nonattainment in the year of birth increases long-run labor market outcomes. While there is some heterogeneity across age categories, the confidence intervals overlap across all age groups. These results may be due to the fact that earnings are highly correlated across ages, but they also provide some degree of assuredness that (i) positive results are found at more than one (ultimately somewhat arbitrary) age category, and (ii) our results are not confounded by some contemporaneous change in earnings determinants in later years. As evidence of the latter point, consider that Columns (1)-(5) are estimated using the same individuals, but the earnings are collected at different years (i.e. cohorts born in 1970 show up between 1998 (Column 1) and 2002 (Column 5)). Our preferred earnings measure, Column (6) serves as a type of "summary-index" test over the various age categories while also reducing the residual variance in annual earnings. Appendix Table A4 presents the corresponding IV estimates for the same set of outcomes, and the results are consistent with both the results in Tables 3 and 4; elevated levels of TSP in the year of birth are associated with a reduction in long-run earnings capacity.

Heterogeneity in Relationship Between Earnings and TSP Exposure

Thus far we have concentrated on the average effects of nonattainment designation and/or TSP exposure on long-run outcomes. However, this focus obscures a tremendous amount of heterogeneity both across individuals and within different parts of the earnings distribution. Table 5 explores heterogeneity in effects of CAAA and TSP exposure by race and sex. Columns (1)-(4) present estimates separately for White, African American, Asian, and Hispanic individuals. Race is collected in the LEHD, and each race category is mutually exclusive. The estimates for white individuals are centered around our baseline estimates. However, estimates for both Asian and African American populations are lower and are not well estimated. The responsiveness of labor force participation for Hispanics is larger than the baseline estimate. The heterogeneity in dose-response relationships may be driven by racial differences in the health endowment and/or the fact that Hispanic populations live in more polluted areas on average, and the underlying dose-response relationship is non-linear. However, the confidence intervals in all regressions overlap and thus preclude us from making strong conclusions pertaining to heterogeneity. Columns (5) and (6) of Table 5 suggest that males are less responsive to adverse environmental conditions in the year of birth. Appendix Table A5 presents heterogeneity results using annual mean earnings for a county×year, and the results are qualitatively similar.

Next, we explore heterogeneity across different parts of the earnings distribution. We estimate a series of regression models that explore how CAAA implementation and TSP exposure affect the fractions of cohorts in various percentiles of the earnings distribution.[34] We begin by calculating the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentiles of the within-county earnings distribution for the 1969 cohort. Then for each subsequent cohort, we classify individuals into bins based on their place in the "pre-treatment" 1969 earnings distribution. Collapsing the data to the county×year yields the fraction of individuals in a given cohort that are in the binned percentile of the 1969 earnings distribution. Appendix Table A6 presents results for the various quantiles, and the results are graphically summarized in Figure 3. The estimates suggest that most of the mean earnings effect is being driven by the bottom tail of the distribution. In particular, CAAA implementation is associated with a relative decrease in the fraction of individuals at the bottom tail of the distribution and a relative increase in the fractions in middle parts of the distribution. These results are consistent with changes in the extensive/work margin explaining much of the observed earnings impacts.

Non-Random Population Sorting Before and After the 1970 CAAA

An important concern for our study is that improvements in air quality might change the composition of the population in nonattainment counties if preferences for clean air are correlated with other differences in socioeconomic status. For example, Chay and Greenstone (2005) find that nonattainment designation, and the associated improvements in air quality, were associated with increases in housing values nearly 10 years after the improvements occurred. While these increases in housing values may reflect movements along indifference curves for the same individuals, they may also arise from changes in the underlying population characteristics as individuals with higher socioeconomic status migrate to areas with cleaner air (Banzhaf and Walsh, 2008). Table 6 investigates whether nonattainment designation is associated with a compositional shift in the underlying county population. Each column represents a different dependent variable. Columns (1)-(3) use data from the NCHS Vital Statistics records to estimate whether the maternal education or the fraction of white or black children differentially change in the years after nonattainment designation. Column (4) uses data from the BEA to estimate whether nonattainment status is correlated with differential changes in per capita income in newly regulated counties. Lastly, Column (5) uses the LEHD earnings records to form a predictive earnings index based on sex and race of workers.[35] The results in Table 6 suggest little evidence for differential sorting along observables that might bias our estimates. The point estimates are not only statistically insignificant, but also small in magnitude, and the sign of the bias tends to point in the opposite direction of our baseline estimates.

Although observable population characteristics do not seem to be changing differentially with respect to CAAA implementation, unobserved changes in the population health endowment might remain. However, we might expect the scope of migration to be more limited in the years immediately after the 1970 CAAA went into place. Appendix Tables A7 and A8 present estimates where we restrict our estimation sample to the period between 1970-1972. Appendix Table A7 presents the results pertaining to the effect of nonattainment designation on labor market outcomes, and the results are similar to our baseline estimates and remain statistically signif-

[34] Sample sizes preclude us from estimating quantile treatment effects directly using our micro data.

[35] Specifically, we use the micro data to estimate earnings regressions controlling for sex and race indicators. We then use the predicted values from this regression as a summary index measure of sorting in Column (5).

icant.[36] Appendix Table A8 presents results from the IV models for the same 1970-1972 window. The results are somewhat larger in magnitude, but they are less well estimated. The first stage F-statistics, pertaining to the strength of the instrument at predicted variation in air pollution in this shortened window, are all below 10. Thus, part of the estimated imprecision of the 2SLS estimates may be driven by a weak instrument.

We also explore whether fertility or the total number of workers observed in the data are correlated with our regulatory variation in Appendix Table A9. Columns (1) and (2) present results using log(# births) and log(# workers) in a county×year as the dependent variable (unweighted). Columns (3) and (4) explore whether the regulatory variation is correlated with the sex ratio at birth or the ratio of male to female workers in the data.[37] We find some evidence of more workers in the LEHD born in nonattainment counties in the years after the 1970 CAAA, although the estimated effect is weakly significant.[38]

Lastly, state migration patterns may be correlated with non-attainment status. This issue is especially relevant for our analysis as we only observe individuals who ever appear in one of the 24 sample states. Consequently, our results may be biased if the CAAA affect the likelihood that an individual moves out of the set of sample states between the first time he appears in our data and the time of observation for labor market outcomes (ages 29-31). While we cannot test this conjecture explicitly, we are able to study the relationship between CAAA implementation and out-of-birth-state-mobility within our 24 sample states. Appendix Table A10 presents results where the dependent variable is the fraction of individuals in a cohort who are working at ages 29-31 in a state other than the state in which they born. Across all specifications in columns (1)-(4), we find no statistically significant relationship between mobility and CAAA initiation. These results provide reassurance that our main findings are unlikely to be severely confounded by differential mobility out of the 24 sample states.

7 Interpretation and Discussion

Our focus thus far has been to ask whether early-life environmental conditions are predictive of long-run outcomes. Evidence suggests the answer is "Yes", but there are some important caveats. The thought experiment in this paper has been to compare individuals born in counties that experienced large regulatory-induced changes in ambient TSP exposure, before and after the regulations went into place. As suggested by Figure 1, the declines in TSP were persistent in the years after CAAA implementation. This means that individuals born before CAAA went into effect (e.g. in years 1969-1971) likely experienced some of the benefits of reduced TSP exposure at ages 1-3, but were not afforded the luxury of exposure to cleaner air over their entire gestation/early childhood as did those born 1973-1974. If the former group did experience earnings benefits associated with the improved air quality, then our relative comparisons would bias the estimated results towards zero (relative to a measure of the cumulative effect of childhood exposure to pollution).

A few other features of our analysis suggest that our results may suffer from additional attenuation. For example, while mobility in very early childhood is fairly limited, any amount of cross-county mobility in the early years of a child's life would likely attenuate our treatment assignment. There may also be measurement error associated with matching our "place of birth" string variable to county FIPS codes, which should generate bias towards zero.

There are also sources of bias that may cause us to underestimate the effect (i.e. downwardly bias our results

[36] For the 1970-1972 specifications, we replace our 1st-4th order baseline polynomial trends with linear trends in pre-determined (1969) employment, population, total transfers per capita, and unemployment transfers per capita. The full set of quartic interactions cannot be fit with 3 years of data.

[37] Population sex ratios may be impacted by CAAA if there are effects on fetal deaths (see Sanders and Stoecker, 2011b for evidence on this topic).

[38] Note, the observed increase is consistent with the main finding of increased labor force participation.

as opposed to biasing our results towards zero). For example, nonattainment designation influences production decisions of firms (Greenstone, 2002; Walker, 2011), and these may have consequences for the broader economic community (Walker, 2012). This being said, the estimates from Walker (2011, 2012) suggest that the effect of nonattainment designation on the overall labor market to be small. Additionally, a regression of contemporaneous county level per capita income on CAAA enactment, as presented in Table 6, failed to detect an effect, which also suggests at most a minor effect on the economy. An additional source of bias could arise from changes in infant mortality rates resulting from CAAA (Chay and Greenstone, 2003b). While the effect on the overall population or cohort size is modest, an increase in the number of individuals in a cohort may change the equilibrium wage rate of that particular cohort (i.e. a shift outwards in the labor supply curve) (Macunovich, 1998; Welch, 1979).

We replicate Chay and Greenstone (2003b) for our subset of counties and states, and present results for both the 1969-1974 window as well as the 1970-1972 window in Appendix Tables A11 and A12, respectively. The estimates from Appendix Table A11 suggest that CAAA implementation is associated with a decrease in mean birthweight and a reduction in the infant mortality rate.[39] However, the latter effect is not statistically significant at conventional levels. These two results are not inconsistent with one another, as a reduction in the infant mortality rate typically "saves" marginal births who are more likely to be born low birthweight. Appendix Table A12 shows that regression models focusing on a narrower window before/after the 1970 CAAA estimate a significant decrease in the infant mortality rate. The results in Appendix Table A12 suggest the 1970 CAAA are associated with a 13.8% decrease in the infant mortality rate in newly regulated counties.[40] The average nonattainment county in our sample had 102 infant deaths in 1972, which suggests that roughly 5 additional lives were saved per county. Relative to a mean cohort size of around 2990, this suggests the mortality induced selection effect should be relatively small (i.e. an increase in the size of the labor force of 0.16 percent) and therefore very unlikely to be driving our results.[41]

In summary, we find suggestive evidence that TSP exposure is correlated with increases in the infant mortality rate and less evidence that air quality affects infant health as proxied by birthweight.[42] So, while it seems that cohorts born around the 1970 CAAA were differentially affected by the regulations, we see little evidence of observable impacts on morbidity. This is consistent with a range of recent work suggesting that birth weight and infant morbidity may be an unreliable metric for predicting long-term health (Painter, Roseboom, Bleker, et al., 2005; Royer, 2009; Kelly, 2011). For example, Almond, Edlund, and Palme (2009) find that radiation exposure exhibits latent effects that shape human capital development later in life, but they find little evidence of health effects as measured by birth outcomes and child hospitalizations.

The Long Run Benefits of the 1970 Clean Air Act Amendments

Our baseline regression estimates suggest that nonattainment designation (or a 10 μg/m^3 reduction in TSP more generally) leads to a 1 percent increase in earnings at age 29-31, or $260 annually. To interpret the magnitude

[39] The internal infant mortality rate is defined as the number of infant deaths due to internal causes divided by the number of live births in a given year.

[40] Chay and Greenstone (2003a) find about a 3-6 percent decrease in the infant mortality rate for the full set of nonattainment counties.

[41] The results are robust to construction bounds following **?**.

[42] The lack of evidence surrounding air pollution and birthweight is consistent with the findings in Chay and Greenstone (2003a,b), but inconsistent with evidence from more recent time periods (Currie, Davis, Greenstone, and Walker, 2012; Currie, Neidell, and Schmieder, 2009; Currie and Walker, 2011). One proposed explanation for the discrepancy in findings concerns both different medical technologies in more recent time periods as well as non-linearities in the dose-response relationship between survival, birthweight, and pollution. An alternative explanation may be that pollution exposure leads to two countervailing forces for infants: (i) improvement in infant health which might raise birthweight of those born (ii) a shift in the survival threshold means that more unhealthy infants survive which may lower average birthweight through compositional forces.

of the effect of air pollution in the year of birth on earnings at ages 29-31, we calculate the lifetime earnings impact implied by our estimates. We assume that the percentage gain in earnings remains constant at 1% over the life-cycle and that earnings are discounted at a 3% real rate (i.e., a 5% discount rate with 2% wage growth) back to age zero.[43] Under these assumptions, the mean present value of lifetime earnings at age zero in the U.S. population is approximately $434,000. We calculate this number using the mean wage earnings from the March 2008 Current Population Survey to obtain an earnings profile over the lifecycle.[44] Thus, the financial value of being born into a nonattainment county in the years after CAAA went into effect is 1% of $434,000 or $4,340 per person.

In 1972, there were approximately 1.5 million births in newly designated nonattainment counties, implying that the total increase in lifetime earnings for this cohort amounted to about $6.5 billion (2008 dollars) annually. To the extent that this reduction in TSPs was permanent (recall Figure 1), these benefits would have accrued in each of the 40 years since then. In this case and assuming a 5% discount rate, the present discounted value of these earnings gains is around $118 billion (2008 dollars).[45] Moreover, these calculations ignore any of the non-wage amenity benefits associated with cleaner air as well as the benefits of any further reductions in TSPs caused by the Clean Air Act in later years. Of course, a full cost-benefit analysis also requires precise information about the costs of these regulations.[46]

8 Conclusion

In this paper, we provide the first empirical quasi-experimental examination of the relationship between individuals' *in-utero* and early childhood exposure to environmental toxins and their labor market outcomes measured nearly 30 years later. We exploit variation induced by the introduction of the CAAA in 1970, which imposed restrictions on the maximum-allowable concentrations of TSP emissions in nonattainment counties. Our analysis compares cohorts in nonattainment counties born just before and after the legislation-mandated reductions in air pollution relative to the same difference among cohorts in attainment counties.

We find that an individual's exposure to lower ambient air pollution levels in the year of birth positively impacts earnings 30 years later. Specifically, we show that the approximate ten percent reduction in TSP that resulted from CAAA implementation is associated with a one percent increase in age-30 earnings among affected cohorts in our sample states. Assuming a constant effect over the lifecycle, we calculate an approximate $4300 average cumulative lifetime income gain in present value terms, implying that early-life air quality contributes a total of $6.5 billion in lifetime earnings for each affected cohort.

Disadvantaged populations disproportionately live in more polluted areas in the United States. This fact has complicated research examining the effects of pollution on health, as disadvantaged individuals, who also have worse health outcomes, tend to be exposed to higher levels of air pollution because of their residential location. However, these persistent disparities also raise interesting questions: why does the pollution-health relationship remain and how does this persistence influence broader social and economic inequality in this country? Our analysis suggests that both economic and environmental inequality may be reinforcing. If this circular relationship is robust, then policies designed to improve air quality may also play a role in thwarting this

[43] This exercise using similar assumptions is conducted in Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan (2011) and Chetty, Friedman, and Rockoff (2011) to evaluate the long-term benefit of having a high value-added teacher in the classroom for an additional year.

[44] Based on this exercise, the cumulative (non-discounted) average lifetime earnings is around $1.5 million (2008$). Poterba, Venti, and Wise (2010) estimate that average lifetime earnings in the United States are $1.6 million (2010$).

[45] This assumes that otherwise, pollution levels as well as the marginal effect of pollution would have remained the same for the last forty years (e.g. technology and supply-demand factors also remained the same).

[46] See Greenstone, List, and Syverson (2012) and Walker (2012) for recent evidence on the costs associated with the Clean Air Act.

cycle, thus serving not only as environmental health policies but also as effective social and economic policies for reducing inequality.

Our paper additionally contributes to a large and growing literature in economics that documents a lasting relationship between early-life conditions and adult well-being by studying a policy-driven shock to the pre- and post-natal environment of modern U.S. cohorts. Previous work on this topic has been challenged by data limitations, with very few large-scale datasets that combine detailed information on location and date of birth together with adult outcomes. We introduce the LEHD data set as a new resource for studying these issues in the United States.

While our study provides new estimates of the impacts of early-life air quality on long-run economic outcomes, it also raises important questions that might motivate future research. Future work should explore in more detail the mechanisms by which these estimated changes in human capital accumulation occur, ideally studying the impacts on health and cognitive ability throughout childhood and adolescence. It would also be interesting to explore the differences in impacts of early-life air quality across family circumstances to help inform our understanding of human capital production and the design of targeted and cost-effective policies. With the growing availability of linkages across various administrative datasets in the United States, this seems like a fruitful area for future research.

References

Abowd, J. M., B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock (2008): "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators," *LEHD Working Paper*.

Albouy, D. (2009a): "The unequal geographic burden of federal taxation," *Journal of Political Economy*, 117(4), 635–667.

——— (2009b): "What are cities worth? Land rents, local productivity, and the capitalization of amenity values," Discussion paper, National Bureau of Economic Research.

Almond, D. (2006): "Is the 1918 Influenza pandemic over? Long-term effects of in utero Influenza exposure in the post-1940 US population," *Journal of Political Economy*, 114(4), 672–712.

Almond, D., and J. Currie (2011): "Human Capital Development before Age Five," in *Handbook of Labor Economics*, ed. by O. Ashenfleter, and D. Card, vol. 4, pp. 1315–1486. Elsevier.

Almond, D., L. Edlund, H. Li, and J. Zhang (2010): *Long-Term Effects of Early-Life Development: Evidence from the 1959 to 1961 China Famine*pp. 321–345. University of Chicago Press.

Almond, D., L. Edlund, and M. Palme (2009): "Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden," *The Quarterly Journal of Economics*, 124(4), 1729–1772.

Almond, D., and B. Mazumder (2012): "Fetal origins and Parental Responses," Working Paper 2012-14, Federal Reserve Board of Chicago.

Angrist, J., and V. Lavy (2009): "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 99(4), 1384–1414.

Auffhammer, M., A. Bento, and S. Lowe (2009): "Measuring the effects of the Clean Air Act Amendments on ambient PM10 concentrations: The critical importance of a spatially disaggregated analysis," *Journal of Environmental Economics and Management*, 58(1), 15–26.

Autor, D., and D. Dorn (Forthcoming): "The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market," *American Economic Review*.

Baker, M., and N. Fortin (2001): "Occupational Gender Composition and Wages in Canada, 1987–1988," *Canadian Journal of Economics*, 34(2), 345–376.

Banzhaf, S. H., and R. P. Walsh (2008): "Do people vote with their feet? An empirical test of Tiebout's mechanism," *The American Economic Review*, 98(3), 843–863.

Barker, D. J. (1990): "The fetal and infant origins of adult disease.," *BMJ: British Medical Journal*, 301(6761), 1111.

Becker, G. S., and N. Tomes (1976): "Child Endowments and the Quantity and Quality of Children," *The Journal of Political Economy*, 84(4), S143–S162.

Bleakley, H. (2010): "Health, human capital, and development," *Annual Review of Economics*, 2(1), 283–310.

Chay, K., C. Dobkin, and M. Greenstone (2003): "The Clean Air Act of 1970 and adult mortality," *Journal of Risk and Uncertainty*, 27(3), 279–300.

Chay, K., and M. Greenstone (2003a): "Air quality, infant mortality, and the Clean Air Act of 1970," Working Paper 10053, National Bureau of Economic Research.

——— (2003b): "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," *The Quarterly Journal of Economics*, 118(3), 1121–1167.

Chay, K., and M. Greenstone (2005): "Does air quality matter? Evidence from the housing market," *Journal of Political Economy*, 113(2).

Chetty, R., J. Friedman, and J. Rockoff (2011): "The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood," Discussion paper, National Bureau of Economic Research.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011): "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR," *Quaterly Journal of Economics*, 126(4), 749–804.

Cunha, F., and J. Heckman (2007): "The Technology of Skill Formation," *The American Economic Review*, 97(2), 31–47.

Currie, J. (2009): "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development," *Journal of Economic Literature*, pp. 87–122.

Currie, J., L. W. Davis, M. Greenstone, and R. Walker (2012): "Do Housing Prices Reflect Environmental Health Risks? Evidence from More than 1600 Toxic Plant Openings and Closings," *Working Paper*.

Currie, J., E. A. Hanushek, E. M. Kahn, M. Neidell, and S. G. Rivkin (2009): "Does Pollution Increase School Absences?," *The Review of Economics and Statistics*, 91(4), 682–694.

Currie, J., and M. Neidell (2005): "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?," *Quaterly Journal of Economics*, (3), 1003–1030.

Currie, J., M. Neidell, and J. Schmieder (2009): "Air pollution and infant health: Lessons from New Jersey," *Journal of health economics*, 28(3), 688–703.

Currie, J., and R. Walker (2011): "Traffic Congestion and Infant Health: Evidence from E-ZPass," *American Economic Journal: Applied Economics*, 3(1), 65–90.

Dehejia, R., and A. Lleras-Muney (2004): "Booms, busts, and babies' health," *The Quarterly Journal of Economics*, 119(3), 1091–1130.

Donald, S., and K. Lang (2007): "Inference with Difference-in-Differences and Other Panel Data," *Review of Economics and Statistics*, 89(2), 221–233.

Gelber, A., and A. Isen (2013): "Children's Schooling and Parents' Behavior: Evidence from the Head Start Impact Study," *Journal of Public Economics*.

Gilliland, F. D., K. Berhane, E. B. Rappaport, D. C. Thomas, E. Avol, W. J. Gauderman, S. J. London, H. G. Margolis, R. McConnell, K. T. Islam, et al. (2001): "The effects of ambient air pollution on school absenteeism due to respiratory illnesses," *Epidemiology*, 12(1), 43.

Graff-Zivin, J., and M. Neidell (2013): "The Impact of Pollution on Worker Productivity," *American Economic Review*.

——— (Forthcoming): "Environment, Health and Human Capital," *Journal of Economic Literature*.

Grainger, C. (2012): "The Distributional Effects of Pollution Regulations: Do Renters Fully Pay for Cleaner Air?," *Journal of Public Economics*.

Greenstone, M. (2002): "The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 clean air act amendments and the census of manufactures," *Journal of Political Economy*, 110(6).

Greenstone, M., J. List, and C. Syverson (2012): "The Effects of Environmental Regulation on the Competitiveness of US Manufacturing," Discussion paper, National Bureau of Economic Research.

Griliches, Z. (1986): "Economic Data Issues," *Handbook of Econometrics*, 3, 1465–1514.

Hanna, R., and P. Oliva (2011): "The Effect of Pollution on Labor Supply: Evidence from a Natural Experiment in Mexico City," Working Paper 17302, National Bureau of Economic Research.

Henderson, V. (1996): "Effects of air quality regulation," *The American Economic Review*, 86(4), 789–813.

Hoynes, H., D. Schanzenbach, and D. Almond (2012): "Long Run Impacts of Childhood Access to the Safety Net," Discussion paper, National Bureau of Economic Research.

Imbens, G., and J. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Johnson, R. (2011): "Long-Run Impacts of School Desegregation and School Quality on Adult Attainments," Working Paper 16664, National Bureau of Economic Research.

Johnson, R. C., and R. F. Schoeni (2007): "The influence of early-life events on human capital, health status, and labor market outcomes over the life course," .

Kelly, E. (2011): "The Scourge of Asian Flu In utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children," *Journal of Human Resources*, 46(4), 669–694.

Lavy, V., A. Ebenstein, and S. Roth (2012): "The Impact of Air Pollution on Cognitive Performance and Human Capital Formation," .

Lindo, J. (2011): "Parental job loss and infant health," *Journal of Health Economics*, 30(5), 869–879.

Macunovich, D. (1998): "Relative cohort size and inequality in the United States," *The American Economic Review*, 88(2), 259–264.

McKinney, K. L., and L. Vilhuber (2008): "LEHD Infrastructure files in the Census RDC - Overview," *LEHD Working Paper*, Revision : 219.

Nilsson, P. (2009): "The Long run effects of Early Childhood Lead Exposure: Evidence from the Phase-out of Leaded Gasoline," *Working Paper*.

Notowidigdo, M. (2011): “The incidence of local labor demand shocks,” Discussion paper, National Bureau of Economic Research.

Painter, R., T. Roseboom, O. Bleker, et al. (2005): “Prenatal exposure to the Dutch famine and disease in later life: an overview.,” *Reproductive toxicology*, 20(3), 345.

Poterba, J., S. Venti, and D. Wise (2010): “The rise of 401 (k) plans, lifetime earnings, and wealth at retirement,” *National Bureau of Economic Research*.

Ransom, M. R., and C. A. Pope (1992): “Elementary school absences and PM¡ sub¿ 10¡/sub¿ pollution in Utah Valley,” *Environmental research*, 58(1), 204–219.

Rau, T., L. Reyes, and S. Urzua (2013): “The Long-term Effects of Early Lead Exposure: Evidence from a case of Environmental Negligence,” University of Maryland, working paper.

Reyes, J. W. (2007): “Environmental Policy as Social Policy? The Impact of Childhood Lead Exposure on Crime,” *The BE Journal of Economic Analysis & Policy*, 7(1).

Romer, P. M. (1986): “Increasing returns and long-run growth,” *The Journal of Political Economy*, pp. 1002–1037.

Royer, H. (2009): “Separated at girth: US twin estimates of the effects of birth weight,” *American Economic Journal: Applied Economics*, 1(1), 49–85.

Sanders, N. (2012): “What Doesn’t Kill You Makes You Weaker Prenatal Pollution Exposure and Educational Outcomes,” *Journal of Human Resources*, 47(3), 826–850.

Sanders, N., and C. Stoecker (2011a): “Where have all the young men gone? using gender ratios to measure fetal death rates,” Discussion paper, National Bureau of Economic Research.

Sanders, N. J., and C. F. Stoecker (2011b): “Where Have All the Young Men Gone? Using Gender Ratios to Measure Fetal Death Rates,” Working Paper 17434, National Bureau of Economic Research.

Schaller, J. (2012): “Booms, Busts, and Fertility: Testing the Becker Model Using Gender-Specific Labor Demand,” University of Arizona, unpublished manuscript.

Schlenker, W., and M. Roberts (2009): “Nonlinear temperature effects indicate severe damages to US crop yields under climate change,” *Proceedings of the National Academy of Sciences*, 106(37), 15594–15598.

Schultz, T. W. (1961): “Investment in human capital,” *The American Economic Review*, pp. 1–17.

Shapiro, J. (2006): “Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital,” *The Review of Economics and Statistics*, 88(2), 324–335.

Tolbert, C. M., and M. Sizer (1996): “U.S. Commuting Zones and Labor Market Areas. A 1990 Update.,” *Economic Research Service Staff Paper*, (No. 9614).

Van den Berg, G., G. Doblhammer-Reiter, and K. Christensen (2011): “Being born under adverse economic conditions leads to a higher cardiovascular mortality rate later in life: Evidence based on individuals born at different stages of the business cycle,” *Demography*, 48(2), 507–530.

Van Den Berg, G., M. Lindeboom, and F. Portrait (2006): “Economic conditions early in life and individual mortality,” *The American Economic Review*, pp. 290–302.

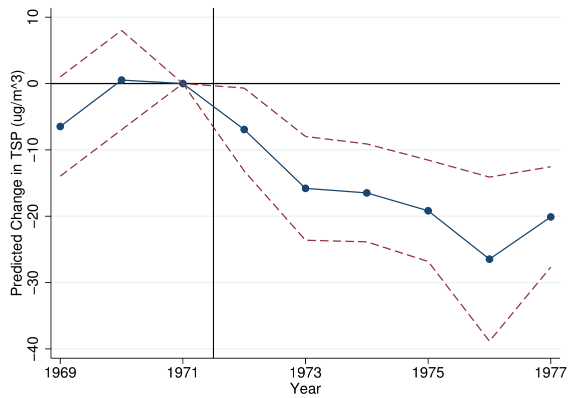
Walker, R. (2011): “Environmental Regulation and Labor Reallocation,” *American Economic Review: Papers and Proceedings*, 101(2).

Walker, R. (2012): “The Transitional Costs of Sectoral Reallocation: Evidence From the Clean Air Act and the Workforce,” .

Welch, F. (1979): "Effects of cohort size on earnings: The baby boom babies' financial bust," *The Journal of Political Economy*, pp. 65–97.

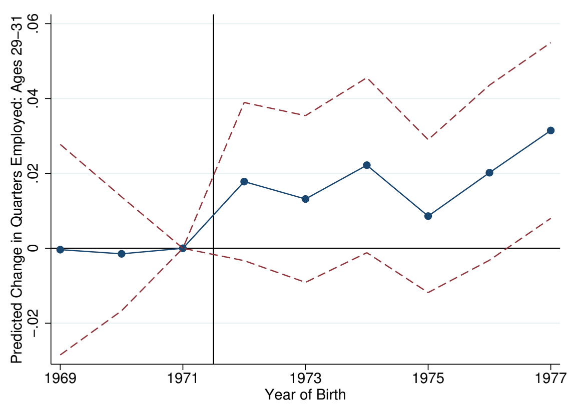
World Health Organization (2005): "Effects of Air Pollution on Children's Health and Development," Report, Available at: http://www.euro.who.int/__data/assets/pdf_file/0010/74728/E86575.pdf, (accessed on March 25, 2013).

Figure 1: Predicted Change in Ambient TSP Pollution As a Function of County Non-Attainment Status



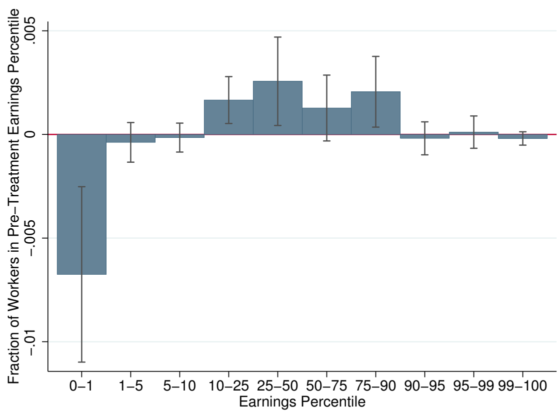
NOTE: This is an event study plot created by regressing the annual mean TSP reading in a given county on a full set of event time indicators interacted with an indicator for "nonattainment", county fixed effects, state by year fixed effects, local economic controls, weather controls, and unrestricted natality controls, weighting by the cohort-level cell size. Reported are the coefficients for event-time, which plot the time path of county TSP in "nonattainment" relative to "attainment" counties before and after the Clean Air Act. The dashed lines represent 95% confidence intervals, where standard errors are computed using cluster-robust standard errors. The event-time coefficients are normalized to zero in 1971. See text for details. Source: EPA.

Figure 2: Predicted Change in Employment at Age 29-31 As a Function of County Non-Attainment Status



NOTE: This is an event study plot created by regressing the mean annual quarters employed between ages 29-31 on a full set of event time indicators interacted with an indicator for "nonattainment", county fixed effects, state by year fixed effects, local economic controls, weather controls, and unrestricted natality controls, weighting by the cohort-level cell size. Reported are the coefficients for event-time, which plot the time path of labor force participation in "nonattainment" relative to "attainment" counties for cohorts born before and after the Clean Air Act. The dashed lines represent 95% confidence intervals, where standard errors are computed using cluster-robust standard errors. The event-time coefficients are normalized to zero in 1971. See text for details. Source: LEHD, EPA.

Figure 3: Predicted Change in the Earnings Distribution As a Function of County Non-Attainment Status



NOTE: This graph plots the coefficient estimates from 10 separate regressions, where the dependent variable in each regression is the proportion of individuals in a cohort whose earnings falls in within the 1969 earnings quantile indicated along the x-axis. The plotted coefficients correspond to the estimate of the effect of county-level "nonattainment" status on the location of individuals within the pre-treatment earnings distribution between the ages of 29-31, where Nonattainment$\times 1(\tau > 1971)$ is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See Appendix Table A6 for point estimates. Source: LEHD, EPA.

Table 1: Summary Means in 1969 and Pre-Trend Differences Between 1969-1971

	(1) Nonattain	(2) Attainment	(3) Nonattain (1969-1971)	(4) Attainment (1969-1971)	(5) (2)-(1) p-value	(6) (4)-(3) p-value
Cohort Averages: Ages 29-31 (Source: LEHD)						
Earnings (2008$)	23623	23294	-0.024	0.027	0.095	0.834
Earnings: 4-Quarter (2008$)	37109	36654	0.022	0.037	0.107	0.185
Quarters Employed	2.74	2.741	-0.038	-0.017	0.650	0.482
Fraction Working in Non-Birth State	0.31	0.30	-0.007	-0.030	0.879	0.129
County Environment: Age 0 (Source: EPA, Schlenker and Roberts (2009))						
Total Suspended Particulate (μg/m^3)	95.89	58.59	-0.058	0.044	0.000	0.662
Average Annual Precipitation	988.1	1164	-0.034	-0.041	0.243	0.811
Average Temperature	12.51	12.78	0.020	0.007	0.047	0.485
County Socio-Economic: Age 0 (Source: BEA)						
Income Per Capita (2008$)	17181	16613	0.037	0.035	0.178	0.233
Total Employment	194709	67169	0.025	0.034	0.041	0.265
Transfers UI Per Capita	4.374	4.33	0.243	0.224	0.131	0.304
Total Transfers Per Capita	7.35	7.36	0.031	0.029	0.564	0.002
Cohort Demographics: Age 0 (Source: NCHS)						
% White	0.85	0.86	-0.011	-0.002	0.779	0.903
% African American	0.14	0.11	0.055	-0.002	0.612	0.160
Mother's Education	11.82	11.73	0.001	-0.008	0.588	0.249
Father's Education	12.31	12.2	0.012	-0.016	0.464	0.127
Mother's Age	24.3	24.11	-0.007	-0.009	0.148	0.839
Father's Age	27.45	27.35	-0.011	-0.011	0.321	0.915
Totals						
Workers Born in 1969 working in 1998-2000	749,000	167,000				
Counties	97	51				

NOTE: This is a table of summary means from our baseline estimation sample. The summary statistics are calculated using data from 1969 for a single cross-section of counties. An observation is a county, and the means have been weighted by county population. All dollar amounts are in 2008 dollars. Column (1) presents summary statistics for those counties that exceeded the EPA's NAAQS TSP standard based on monitor readings in 1970. Column (2) presents summary statistics for the subset of counties that were in compliance with the NAAQS standards in 1970. Columns (3) and (4) present log differences between 1969 and 1971 for nonattainment and attainment counties, respectively. Column (5) presents the p-values from a test of the null hypothesis that the levels in columns (1) and (2) are the same in 1969. Column (6) presents the p-values from a test of the null hypothesis that the log differences between 1969 and 1971 are the same for both nonattainment and attainment counties. The p-values in Columns (5) and (6) are generated from a regression of either levels or log differences on an indicator for nonattainment, using robust standard errors for inference, and weighting by the county population.

Table 2: The Effect of County Nonattainment Designation on Labor Market Outcomes at Ages 29-31

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Nonattainment$\times 1(\tau > 1971)$	0.020**	0.023***	0.019**	0.022***
	(0.010)	(0.008)	(0.008)	(0.008)
Sample Size	888	888	888	888
	Panel B: Annual Earnings			
Nonattainment$\times 1(\tau > 1971)$	224.59	278.74**	250.57**	276.66**
	(148.36)	(126.23)	(118.63)	(115.45)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in Panel A is the annual mean quarters employed $\in [0, 4]$ from ages 29-31, and the dependent variable in Panel B is mean earnings from ages 29-31. The data have been aggregated to county-cohort cells. The regression sample remains the same in all columns and panels. Both Panel A and B report estimates of the effect of county-level "nonattainment" status on labor market outcomes, where Nonattainment$\times 1(\tau > 1971)$ is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 3: The Effect of Total Suspended Particulate Pollution in the Year of Birth on Labor Market Outcomes at Ages 29-31

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Mean TSPs (/10)	-0.017*	-0.024**	-0.021*	-0.028*
	(0.010)	(0.012)	(0.011)	(0.016)
	Panel B: Annual Earnings			
Mean TSPs (/10)	-186.21	-286.83*	-266.88*	-351.74*
	(139.30)	(158.61)	(145.62)	(190.39)
First Stage F	14.7	15.0	14.8	14.2
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in Panel A is the annual mean quarters employed $\in [0, 4]$ from ages 29-31, and the dependent variable in Panel B is mean earnings from ages 29-31. The data have been aggregated to county-cohort cells. The regression sample remains the same in all columns and panels. Both Panel A and B report 2SLS estimates pertaining to the effect of ambient TSP levels in the year of birth on labor market outcomes between the ages of 29-31, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the "first-stage" strength of the instrument. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 4: The Effect of Nonattainment Designation in Year of Birth on Labor Market Outcomes at Various Ages

	(1) Age 28	(2) Age 29	(3) Age 30	(4) Age 31	(5) Age 32	(6) Age 29-31
	Panel A: Quarters Employed					
Nonattainment$\times 1(\tau > 1971)$	0.023***	0.022***	0.022**	0.021***	0.015**	0.022***
	(0.008)	(0.008)	(0.009)	(0.008)	(0.006)	(0.008)
	Panel B: Annual Earnings					
Nonattainment$\times 1(\tau > 1971)$	97.86	209.35**	286.41**	332.94**	219.31**	276.66**
	(111.47)	(105.70)	(132.00)	(130.48)	(104.09)	(115.45)
Counties	148	148	148	148	148	148
First Year	1969	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974	1974
Sample Size	739	888	888	888	888	888
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Unrestricted	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 12 separate regressions, 6 per panel. The dependent variable in Panel A is the annual mean quarters employed $\in [0, 4]$, and the dependent variable in Panel B is mean earnings. The outcome variables in both panels change with respect to follow up ages (indicated in the column headings), but the underlying "cohort" or sample remains the same. The data have been aggregated to county-cohort cells. The regression sample remains the same in all columns and panels. Both Panel A and B report estimates of the effect of county-level "nonattainment" status on labor market outcomes, where Nonattainment$\times 1(\tau > 1971)$ is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 5: Investigating Heterogeneity: The Effect of Air Pollution Exposure in Year of Birth on Labor Market Outcomes at Ages 29-31 by Race and Gender

	(1) White	(2) African American	(3) Asian	(4) Hispanic	(5) Male	(6) Female
	Panel A: Quarters Employed					
Non×1(τ > 1971)	0.020***	0.003	0.010	0.024	0.017**	0.020**
	(0.008)	(0.012)	(0.033)	(0.016)	(0.007)	(0.009)
	Panel B: Quarters Employed					
Mean TSPs (/10)	-0.017*	-0.003	-0.009	-0.039	-0.019*	-0.023*
	(0.010)	(0.014)	(0.029)	(0.025)	(0.010)	(0.013)
First Stage F	12.1	13.6	31	9.86	16.4	17.5
Counties	147	126	124	135	148	148
First Year	1969	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974	1974
Sample Size	882	756	744	810	888	888
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Unrestricted	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 12 separate regressions, 6 per panel. The dependent variable in all regressions is the annual mean quarters employed $\in [0, 4]$ from various races and sex, indicated in the column headings. Columns (1)-(4) are mutually exclusive, and each column reflects a different race as indicated in the underlying earnings records. Columns (5) and (6) report results separately by sex. The data have been aggregated to county-cohort cells. Panel A reports estimates of the effect of county-level "nonattainment" status on employment status, where Nonattainment×1(τ > 1971) is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. Panel B reports 2SLS estimates pertaining to the effect of ambient TSP levels in the year of birth on employment status between the ages of 29-31, using Nonattainment×1(τ > 1971) as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the "first-stage" strength of the instrument. All regressions control for cohort (i.e. birth-county) fixed effects, state×year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 6: The Effect of County Nonattainment on Sorting and Population Characteristics

	(1) Maternal Education	(2) Fraction White	(3) Fraction Black	(4) Income Per Capita	(5) Predicted Earnings Ages 29-31
Nonattainment$\times 1(\tau > 1971)$	-0.155	0.007	-0.004	-37.23	-24.58
	(0.099)	(0.005)	(0.003)	(106.20)	(29.82)
Variable Mean	12	0.839	0.140	18506	23587
Counties	148	148	148	148	148
First Year	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974
Sample Size	888	888	888	888	888
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic					
YOB Natality Unrestricted					

NOTE: This table reports regression coefficients from 5 separate regressions. The dependent variable in columns (1)-(4) pertain to the population average characteristics in a county$\times$birth-year. Column (5) is a predicted earnings measure, where earnings at ages 29-31 are regressed on a set of indicators for race and sex, and the predicted values from this regression are used as a summary index measure of population characteristics (averaged to the birth-county$\times$year level). The regression sample remains the same in all columns and panels. The table reports estimates of the effect of county-level "nonattainment" status on labor market outcomes, where Nonattainment$\times 1(\tau > 1971)$ is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. All regressions control for county (or cohort) fixed effects and state$\times$year fixed effects. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, NCHS, EPA.

A Appendix Tables

Table A1: OLS Estimates: The Effect of Air Pollution Exposure in Year of Birth on Labor Market Outcomes at Age 29-31

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Mean TSPs (/10)	0.001	0.001*	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
	Panel B: Annual Earnings			
Mean TSPs (/10)	11.76*	12.67*	12.99**	11.88**
	(7.08)	(6.58)	(5.29)	(5.89)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in Panel A is the annual mean quarters employed $\in [0, 4]$ from ages 29-31, and the dependent variable in Panel B is mean earnings from ages 29-31. The data have been aggregated to county-cohort cells. The regression sample remains the same in all columns and panels. Both Panel A and B report OLS estimates pertaining to the effect of ambient TSP levels in the year of birth on labor market outcomes between the ages of 29-31. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A2: The Effect of County Nonattainment Designation on Ambient TSP Exposure

	(1)	(2)	(3)	(4)
Nonattainment×1(τ > 1971)	-12.06***	-9.71***	-9.38***	-7.86***
	(3.31)	(2.78)	(2.56)	(2.84)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Weather Controls		Yes	Yes	Yes
Natality Basic			Yes	Yes
Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 4 separate regressions. The dependent variable the annual mean TSP reading in a county×year. The regression sample remains the same in all columns. The table reports OLS estimates pertaining to the effect of "nonattainment" on ambient TSP levels in the years after the CAAA. Nonattainment×1(τ > 1971) is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. All regressions control for county fixed effects, state×year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: EPA.

Table A3: The Effect of County Nonattainment Designation on Labor Market Outcomes at Ages 29-31

	(1)	(2)	(3)	(4)
	Panel A: Log Annual Earnings			
Nonattainment$\times 1(\tau > 1971)$	0.006	0.010**	0.010**	0.011**
	(0.005)	(0.005)	(0.004)	(0.004)
Sample Size	888	888	888	888
	Panel B: Non-Zero Earnings			
Nonattainment$\times 1(\tau > 1971)$	2.465	68.805	61.501	31.623
	(89.243)	(91.387)	(82.138)	(91.918)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in Panel A is the mean log annual earnings from ages 29-31, and the dependent variable in Panel B consist of mean annual, non-zero earnings from ages 29-31. The data have been aggregated to county-cohort cells. The regression sample remains the same in all columns and panels. Both Panel A and B report estimates of the effect of county-level "nonattainment" status on labor market outcomes, where Nonattainment$\times 1(\tau > 1971)$ is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A4: The Effect of Total Suspended Particulate Pollution in the Year of Birth on Labor Market Outcomes at Various Ages

	(1) Age 28	(2) Age 29	(3) Age 30	(4) Age 31	(5) Age 32	(6) Age 29-31
	Panel A: Quarters Employed					
Mean TSPs (/10)	-0.025*	-0.028*	-0.028*	-0.026*	-0.019	-0.028*
	(0.013)	(0.014)	(0.016)	(0.015)	(0.012)	(0.016)
	Panel B: Annual Earnings					
Mean TSPs (/10)	-105.52	-260.61*	-356.50*	-414.17*	-272.77*	-351.74*
	(121.43)	(149.69)	(204.88)	(221.57)	(160.41)	(190.39)
First Stage F	15	14.2	14.2	14.2	14.2	14.2
Counties	148	148	148	148	148	148
First Year	1969	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974	1974
Sample Size	740	888	888	888	888	888
Year of Birth Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year of Birth Natality Basic	Yes	Yes	Yes	Yes	Yes	Yes
Year of Birth Natality Unrestricted	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 12 separate regressions, 6 per panel. The dependent variable in Panel A is the annual mean quarters employed $\in [0, 4]$, and the dependent variable in Panel B is mean earnings. The outcome variables in both panels change with respect to follow up ages (indicated in the column headings), but the underlying "cohort" or sample remains the same. The data have been aggregated to county-cohort cells. The regression sample remains the same in all columns and panels. Both Panel A and B report 2SLS estimates pertaining to the effect of ambient TSP levels in the year of birth on log earnings between the ages of 29-31, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the "first-stage" strength of the instrument. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A5: Investigating Heterogeneity: The Effect of Air Pollution Exposure in Year of Birth on Labor Market Outcomes at Ages 29-31 by Race and Gender

	(1) White	(2) African American	(3) Asian	(4) Hispanic	(5) Male	(6) Female
	Panel A: Annual Earnings					
Non×1($\tau > 1971$)	230.45*	22.73	246.37	103.79	153.11	281.18**
	(118.46)	(173.94)	(571.51)	(226.38)	(109.12)	(117.77)
	Panel B: Annual Earnings					
Mean TSPs (/10)	-199.75	-25.54	-214.03	-165.01	-173.18	-317.30**
	(128.12)	(195.78)	(498.43)	(336.72)	(129.66)	(161.17)
First Stage F	12.1	13.6	31	9.86	16.4	17.5
Counties	147	126	124	135	148	148
First Year	1969	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974	1974
Sample Size	882	756	744	810	888	888
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Unrestricted	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 12 separate regressions, 6 per panel. The dependent variable in all regressions is the mean annual earnings from ages 29-31 for various races and sex, indicated in the column headings. Columns (1)-(4) are mutually exclusive, and each column reflects a different race as indicated in the underlying earnings records. Columns (5) and (6) report results separately by sex. The data have been aggregated to county-cohort cells. Panel A reports estimates of the effect of county-level "nonattainment" status on employment status, where Nonattainment×1($\tau > 1971$) is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. Panel B reports 2SLS estimates pertaining to the effect of ambient TSP levels in the year of birth on employment status between the ages of 29-31, using Nonattainment×1($\tau > 1971$) as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the "first-stage" strength of the instrument. All regressions control for cohort (i.e. birth-county) fixed effects, state×year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A6: Effects of Nonattainment and Air Pollution on the Distribution of Earnings at Ages 29-31

	(1) $0 \leq p \leq 1$	(2) $1 < p \leq 5$	(3) $5 < p \leq 10$	(4) $10 < p \leq 25$	(5) $25 < p \leq 50$
Nonattainment$\times 1(\tau > 1971)$	-0.007***	-0.000	-0.000	0.002***	0.003**
	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)
	(6) $50 \leq p \leq 75$	(7) $75 < p \leq 90$	(8) $90 < p \leq 95$	(9) $95 < p \leq 99$	(10) $99 < p \leq 100$
Nonattainment$\times 1(\tau > 1971)$	0.001	0.002**	-0.000	0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Counties	148	148	148	148	148
First Year	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974
Sample Size	888	888	888	888	888
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic	Yes	Yes	Yes	Yes	Yes
YOB Natality Unrestricted	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 10 separate regressions, 5 per panel. The dependent variable in each column is the proportion of individuals in a cohort whose earnings falls in within the 1970 earnings quantile indicated in the column heading. Panel A reports estimates of the effect of county-level "nonattainment" status on the fraction of individuals in the indicated part of the earnings distribution, where Nonattainment$\times 1(\tau > 1971)$ is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. Panel B reports 2SLS estimates pertaining to the effect of ambient TSP levels in the year of birth on the location within the pre-treatment earnings distribution between the ages of 29-31, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the "first-stage" strength of the instrument. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. Source: LEHD, EPA.

Table A7: The Effect of County Nonattainment Designation on Labor Market Outcomes at Ages 29-31

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Nonattainment$\times 1(\tau > 1971)$	0.020***	0.019**	0.015*	0.018**
	(0.007)	(0.008)	(0.008)	(0.008)
	Panel B: Annual Earnings			
Nonattainment$\times 1(\tau > 1971)$	168.77*	210.91**	187.85*	171.90*
	(97.85)	(95.42)	(95.85)	(102.41)
Counties	148	148	148	148
First Year	1970	1970	1970	1970
Last Year	1972	1972	1972	1972
Sample Size	444	444	444	444
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in Panel A is the annual mean quarters employed $\in [0, 4]$ from ages 29-31, and the dependent variable in Panel B is mean earnings from ages 29-31. The data have been aggregated to county-cohort cells. The regression sample remains the same in all columns and panels. Both Panel A and B report estimates of the effect of county-level "nonattainment" status on labor market outcomes, where Nonattainment$\times 1(\tau > 1971)$ is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with linear trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A8: The Effect of Total Suspended Particulate Pollution in the Year of Birth on Labor Market Outcomes at Ages 29-31

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Mean TSPs (/10)	-0.019**	-0.029	-0.027	-0.033
	(0.009)	(0.022)	(0.024)	(0.035)
	Panel B: Annual Earnings			
Mean TSPs (/10)	-165.99*	-327.65	-332.94	-326.46
	(100.50)	(230.84)	(265.17)	(334.88)
First Stage F	7.94	9.78	9.52	6.47
Counties	148	148	148	148
First Year	1970	1970	1970	1970
Last Year	1972	1972	1972	1972
Sample Size	444	444	444	444
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in Panel A is the annual mean quarters employed $\in [0, 4]$ from ages 29-31, and the dependent variable in Panel B is mean earnings from ages 29-31. The data have been aggregated to county-cohort cells. The regression sample remains the same in all columns and panels. Both Panel A and B report 2SLS estimates pertaining to the effect of ambient TSP levels in the year of birth on labor market outcomes between the ages of 29-31, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the "first-stage" strength of the instrument. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with linear trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A9: The Effect of County Nonattainment Designation on Fertility and Workforce Composition

	(1) Log Number of Workers	(2) Log Number of Births	(3) Ratio M/F Births	(4) Ratio M/F Workers
	Panel A: Nonattainment and the Workforce/Fertility			
Nonattainment×1(τ > 1971)	0.029*	-0.006	0.002	0.001
	(0.015)	(0.016)	(0.002)	(0.001)
	Panel B: TSPs and the Workforce/Fertility			
Mean TSPs (/10)	-0.017	0.003	-0.002	-0.001
	(0.012)	(0.009)	(0.003)	(0.002)
Variable Mean	7.96	7.89	.513	.506
First Stage F	4.64	4.64	14.2	14.2
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	174
Sample Size	888	888	888	888
Weather Controls	Yes	Yes	Yes	Yes
Natality Basic	Yes	Yes	Yes	Yes
Natality Unrestricted	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 10 separate regressions, 5 per panel. The dependent variable in each column is corresponds to a different outcome in a county×year. The regression sample remains the same in all columns and panels. Panel A reports estimates of the effect of county-level "nonattainment" status on the number and fraction of workers and births in a county×year, where Nonattainment×1($\tau > 1971$) is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. Panel B reports 2SLS estimates pertaining to the effect of ambient TSP levels in the year of birth on the number and fraction of workers and births, using Nonattainment×1($\tau > 1971$) as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the "first-stage" strength of the instrument. All regressions control for cohort (i.e. birth-county) fixed effects, state×year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions in columns (3)-(4) are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: NCHS, EPA.

Table A10: The Effect of County Nonattainment Designation on Cross-State Mobility Between Birth and Ages 29-31

	(1)	(2)	(3)	(4)
Nonattainment$\times 1(\tau > 1971)$	0.002	0.001	0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 4 separate regressions. The dependent variable is the fraction of individuals who are working at ages 29-31 in a state other than the state in which they were born. The data have been aggregated to county-cohort cells. The regression sample remains the same in all columns and panels. The table reports estimates of the effect of county-level "nonattainment" status on cross-state mobility, where Nonattainment$\times 1(\tau > 1971)$ is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A11: The Effect of County Nonattainment Designation on Infant Mortality and Morbidity

	(1) Birthweight	(2) Low Birthweight	(3) Internal IMR Rate	(4) External IMR Rate
	Panel A: Nonattainment and Infant Health			
Nonattainment×1($\tau > 1971$)	-6.838**	0.000	-0.001	0.000
	(3.009)	(0.001)	(0.001)	(0.000)
	Panel B: TSPs and Infant Health			
Mean TSPs (/10)	8.694*	-0.001	0.001	-0.000
	(4.588)	(0.002)	(0.001)	(0.000)
Variable Mean	3273	0.078	0.034	0.001
First Stage F	14.2	14.2	14.2	14.2
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Weather Controls	Yes	Yes	Yes	Yes
Natality Basic	Yes	Yes	Yes	Yes
Natality Unrestricted	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in each column is corresponds to a different infant health outcome in a county×year. The regression sample remains the same in all columns and panels. Panel A reports estimates of the effect of county-level "nonattainment" status on the infant mortality and morbidity in a county×year, where Nonattainment×1($\tau > 1971$) is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. Panel B reports 2SLS estimates pertaining to the effect of ambient TSP levels in the year of birth on the infant mortality and morbidity, using Nonattainment×1($\tau > 1971$) as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the "first-stage" strength of the instrument. All regressions control for cohort (i.e. birth-county) fixed effects, state×year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: NCHS, EPA.

Table A12: The Effect of County Nonattainment Designation on Infant Mortality and Morbidity

	(1) Birthweight	(2) Low Birthweight	(3) Internal IMR Rate	(4) External IMR Rate
	Panel A: Nonattainment and Infant Health			
Nonattainment$\times 1(\tau > 1971)$	-4.837	-0.001	-0.005***	0.001*
	(4.549)	(0.002)	(0.001)	(0.000)
	Panel B: TSPs and Infant Health			
Mean TSPs (/10)	9.185	0.002	0.009	-0.001
	(10.951)	(0.004)	(0.007)	(0.001)
Variable Mean	3272	0.078	0.036	0.001
First Stage F	6.47	6.47	6.47	6.47
Counties	148	148	148	148
First Year	1970	1970	1970	1970
Last Year	1972	1972	1972	1972
Sample Size	444	444	444	444
Weather Controls	Yes	Yes	Yes	Yes
Natality Basic	Yes	Yes	Yes	Yes
Natality Unrestricted	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in each column is corresponds to a different infant health outcome in a county$\times$year. The regression sample remains the same in all columns and panels. Panel A reports estimates of the effect of county-level "nonattainment" status on the infant mortality and morbidity in a county$\times$year, where Nonattainment$\times 1(\tau > 1971)$ is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. Panel B reports 2SLS estimates pertaining to the effect of ambient TSP levels in the year of birth on the infant mortality and morbidity, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the "first-stage" strength of the instrument. All regressions control for cohort (i.e. birth-county) fixed effects, state$\times$year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with linear trends. Additional controls are listed in the tables. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: NCHS, EPA.

B Data Appendix

B.1 Data Appendix

Longitudinal Employer Household Dynamics Files

Table B1: Cohort Age Coverage in the LEHD

Age	Birth Years	Earnings Years
28	1970-1979	1998-2007
29	1969-1978	1998-2007
30	1968-1977	1998-2007
31	1967-1976	1998-2007
32	1966-1975	1998-2007

B.2 Matching Algorithm

In order to match Clean Air Act nonattainment status and county pollution levels at birth to individuals' earnings records, we take advantage of the LEHD variables that detail each worker's place of birth. Derived from Social Security Administration records, the two variables pobcity (place of birth) and pobst (state, or foreign born, country of birth) should in theory identify the county of birth. However, in practice, a major challenge with the data is that they are string variables (filled out by hand when applying for a Social Security card) that can be difficult to match to County FIP codes; while the place of birth variable is supposed to list the city of birth (or county, if born in a non-incorporated community), there are almost 2,280,002 unique combinations of Pobcity and Pobst among the non-foreign born in the data whereas there are fewer than 22,800 actual Census places and counties in the United States, implying over two million errors in the string variables. The several orders of magnitude more places of birth in the LEHD stem from (using the example of Los Angeles) misspellings (e.g. Las Angeles), alternative abbreviations (e.g. L.A.), extraneous information (e.g. Los Angeles, Los Angeles County), neighborhoods within the city (e.g. Hollywood), incorrect states (e.g. CO), some combination thereof, and other errors. In order to match these string variables to county FIP codes in both an accurate and comprehensive manner, we follow a 5-step process. In each step, we draw on the Census Bureau's database of Census places, counties, and minor civil divisions as well as the United States Geological Survey's Geographic Names Information System (GNIS) file, which includes information on non-Census geographic localities.

Step 1: We take the LEHD string variables as is and merge them to the string variables from the Census and GNIS databases. The process successfully matches up 68,698 unique combinations of pobcity and pobst from the LEHD to their county fips codes, which constitutes 3.01% of the total unique combinations of the two variables, and when weighted by the number of workers in the LEHD that report each combination, 56.12% of the sample. Broken down by file type, 16,956 (0.75% unweighted and 48.86% weighted) are matched to Census places, 2,063 to counties (0.09% unweighted and 1.41% weighted), 2,585 to minor civil divisions (0.12% unweighted, 0.61% weighted), and 47,094 to GNIS localities (3.01% unweighted and 5.24% weighted).

Step 2: We split up the pobcity string variable into separate strings delimited by spaces (to split up the string variable into separate words) and then merge the first word from each string to the string variables from the Census and GNIS databases. We carry out this step because examination of several thousand string variables by hand indicated that the most prominent error in the data occurs when the city is listed followed by one or more letters that represent the county in which the city is located. In this step, we make one additional restriction; the first word from the pobcity variable cannot be matched when that word also matches the first word of a two or more word geographic locality (i.e. if both Los and Los Angeles exist, we do not match up a string with a first word of Los). The process successfully matches up 528,126 unique observations, representing 23.16% of all unique combinations unweighted and 29.48% of weighted combinations. Broken down by file type, 329,650 are matched to Census places (14.43%% unweighted and 25.38% weighted), 29,837 to counties (1.34% unweighted and 0.87% weighted), 17,432 to minor civil divisions (0.76% weighted and 0.48% unweighted), and 151,207 to GNIS localities (6.63% unweighted and 2.75% weighted).

Step 3: We split up the pobcity string variable into separate strings delimited by spaces and then merge twenty different combinations of the multiple strings variables to the string variables from the Census and GNIS databases (e.g. the first two words of the string, the first three words of the string, etc). A full list of these combinations is available upon request. This process successfully merges 15,846 places of birth, which constitutes 0.68% of all listed places of birth (0.92% weighted)

Step 4: We split the pobcity string variable up into separate strings delimited by spaces and replace the first word in the string if they appear to abbreviate a direction (e.g. N, No, Nor, S, etc) or the word Saint (St, Snt) with the full word (e.g. North, Saint), and then merge the new string variable to the string variables from the Census and GNIS databases. The process successfully merges 7,003 unique combinations, representing 0.32% of all unique combinations unweighted and 0.28% of weighted combinations.

Step 5: We take the remaining combinations (1,660,729, 72.82% unweighted and 13.20% weighted) and select high frequency combinations (defined as those for which at least 100 individuals in the data report that combination or that shares the first word in pobcity with a combination that at least 100 individuals report). We then take these combinations (98,705, 4.33% unweighted, and 10.77% weighted) and attempt to match them by hand. This process is primarily undertaken to deal with high frequency spelling mistakes. We are able to hand match 82,928 of these combinations, representing 3.64% of all unique combinations unweighted and 10.32% of weighted combinations.

Duplicates: There are several sources of duplicate localities within a state that share the same name but are in different counties. Whenever a Census place or county shares a name with a minor civil division or non-Census locality, we assume that the individual was born in either the Census place or county (since the proper way to fill out the form is to list either the city or county of birth). If there are multiple cities within a state in different counties or multiple minor civil divisions or non-Census localities within the same state in different counties, we treat these observations as unmatched (80,256 combinations, 3.52% unweighted and 1.31% weighted). On the other hand, if a city and county share the same name but the city is in a different county, we assume the correct place of birth is the city (as the large majority of individuals list their city), unless the string indicates that it is a county (e.g. the last word in the string is Co, Cou, Count, County, etc). However, the results are robust to making the opposite assumption or treating these observations as unmatched. The following table shows the distribution of matches by step. In all, we match 622,345 unique combinations of pobst and pobcity, which constitutes 27.30% of all unique LEHD combinations and 95.81% of LEHD individuals' places of birth.

Table B2: Match Statistics

	Number of unique places of birth	Percent of total	Percent of total (weighted by no. of persons)	Cumulative Weighted Match Rate
Step 1	68,698	3.01	56.12	56.12
Step 2	528,126	23.16	29.48	85.60
Step 3	15,846	0.68	0.92	86.52
Step 4	7,003	0.32	0.28	86.80
Step 5	82,928	3.64	10.32	97.12
Duplicates	80,256	-3.52	-1.31	95.81
Total match	622,345	27.30	95.81	
Total unmatched	1,657,657	72.70	4.19	

Note: Source: LEHD, USGS