UNIVERSITY OF CALIFORNIA
Santa Barbara

Three Papers in Environmental and Natural Resource Economics

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy
in
Economics
by

Christopher Colin Goodwin

Committee in Charge:
Professor Olivier Deschenes, Chair
Professor Douglas Steigerwald
Professor Robert Deacon

December 2015

.

The Dissertation of
Christopher Colin Goodwin is approved:

Professor Robert Deacon

Professor Douglas Steigerwald

Professor Oliver Deschenes, Committee Chair

December 2015

. .

VITA OF CHRISTOPHER COLIN GOODWIN

December 2015

EDUCATION

Bachelor of Forestry Science, First Class Honours, University of Canterbury, 2000
Master of Science in Forestry, Virginia Polytechnic & State University, 2002
Master of Arts in Economics, University of California, Santa Barbara, 2008
Doctor of Philosophy in Economics, University of California, Santa Barbara, 2015 (expected)

Three Papers in Environmental and Natural Resource Economics

Christopher Colin Goodwin

Abstract

The first chapter concerns the eruption of Mount St. Helens in Washington State on May 18th, 1980 which resulted in a massive and unanticipated particulate air pollution shock. I use the incidence of the ash cloud fallout across Washington as a natural experiment to estimate the effect of a particulate shock on birth outcomes and infant mortality. I find that while there is no statistical effect on infant mortality, there were statistically fewer low birth weight babies born. The measured effect indicates about 230 fewer babies were born as a result of exposure while in the womb. These results suggest that about 1 in 10 pregnancies were terminated from the 20th percentile group of the weight distribution, increasing to almost 1 in 5 in the lowest 5th percentile. The effect is found to be the strongest in the early stages of pregnancy. Using these findings I estimate that the cost of a single particulate shock of typical magnitude on a metropolitan area of median size is $3.7 million. This quasi-experimental treatment is unique since the ash particulate is bioreactively inert and uncorrelated with other pollution. It clearly identifies the pernicious nature of all particulates, not just particulate categorized by source. In addition, the discrete timing of the event identifies the most vulnerable window for suspended particulate shocks on expectant mothers.

In the second chapter, my colleague Daniel Moncayo and I investigate the informational effect that social comparisons play in household electricity use. We employ a randomized field experiment to present individuals with a social norm that encourages them to conserve electricity by comparing their consumption to that of an "energy effi-

cient neighbor," consisting of the average of the 10th percentile of participants' electricity consumption. Utilizing smart-meter data, we find households that received our informational treatment reduced their electricity consumption by an average of 7%, even though the treatment was private and there was no financial incentive to conserve. We also discover that conservation gains are largest during the peak morning and evening hours.

The third chapter examines the effect of country-level political stability and investment security on forestland use. Using cross-section data I find that these measures are associated with benign outcomes for overall rates of forest area change and roundwood production. These associations are robust when instrumented for endogeneity, and reveal stronger impacts than OLS estimates would imply. Two-stage least squares results indicate that a one standard deviation increase in political stability, as currently measured by the World Bank, increases forest area by 12% over 10 years, and increases roundwood production by a factor of 10 to 19. Targeting political stability and investment security may be one of the most effective tools in mitigating carbon emissions through forest expansion and increasing forest productivity.

Contents

Chapter 1

Air Quality and Birth Outcomes: The Mount St. Helens Eruption - A Natural Experiment.

Summary

The eruption of Mount St. Helens in Washington State on May 18th, 1980 resulted in a massive and unanticipated particulate air pollution shock. I use the incidence of the ash cloud fallout across Washington as a natural experiment to estimate the effect of a particulate shock on birth outcomes and infant mortality. I find that while there is no statistical effect on infant mortality, there were statistically fewer low birth weight babies born. The measured effect indicates about 230 fewer babies were born as a result of exposure while in the womb. These results suggest that about 1 in 10 pregnancies were terminated from the 20th percentile group of the weight distribution, increasing to almost 1 in 5 in the lowest 5th percentile. The effect is found to be the strongest in the early stages of pregnancy. Using these findings I estimate that the cost of a single particulate shock of typical magnitude on a metropolitan area of median size is $3.7 million. This quasi-experimental treatment is unique since the ash particulate is bioreactively inert and uncorrelated with other pollution. It clearly identifies the pernicious nature of all particulates, not just particulate categorized by source. In addition, the discrete timing of the event identifies the most vulnerable window for suspended particulate shocks on expectant mothers.

1.1 Introduction

Since 1970, the U.S. Environmental Protection Agency (EPA) has regulated particulate air pollution as a criteria pollutant[1] for the public health and wellbeing. Particulate pollution[2] has been linked in epidemiological studies to increased risks of cardiopulmonary disease and mortality in humans.[3] More importantly, short term particulate shocks are thought to directly affect mortality (U.S. EPA, 2011). However, there are no large scale clinical trials of human exposure (in vivo) to particulates[4] and the suggestion of causality is actively questioned. Natural experiments, where the incidence of exposure is distributed independent of population characteristics, offer a way for researchers to overcome many of the issues regarding self selection and simultaneity that characterize epidemiological research.

The ashfall from the May 18th, 1980, eruption of Mount St. Helens constitutes a quasi-experimental treatment in particulate exposure. Using the universe of birth and death records from the State of Washington I find two important things. First, there are statistically fewer babies born in the lower tail of the birth weight distribution within the nine month timing window after the Mount St. Helens eruption. Second, there is no statistical evidence of an increase in the infant death rate for infants (children less than one-year old) exposed either as infants or whilst in utero. This quasi-experimental event affected about 9000 pregnant women and represents a close ideal to the research design where air pollution is randomly assigned across mothers and infants, as discussed by

[1]The EPA is required to set National Ambient Air Quality Standards for six pollutants under the Clean Air Act (42 U.S.C. 7401 *et seq.*); these six pollutants are known as criteria pollutants.

[2]Various particulate pollution categorizations have been promulgated since 1970. Prior to 1987 all particulates up to around 45 micrometers in diameter were measured as Total Suspended Particulates (TSP). Between 1987 and 1997 particulates were measured and categorized as PM10 - particulates less than 10 microns. After 1997, two sets of standards have been in force - one set for fine particulates called PM2.5 (particulates less than 2.5 microns) and one set for coarse particulates (PM10).

[3]See Pope (2000) for a review of the more than 150 epidemiological studies.

[4]U.S. EPA (2009) tabulates 58 small scale and limited controlled human exposure studies. There are no more than 50 participants in any one study, with most involving 10-15 subjects.

Chay and Greenstone (2003).

There are number of significant reasons why the Mount St. Helens eruption is a compelling natural experiment in particulate exposure. The eruption was a discrete exogenous shock that sent millions of tons of particulate matter[5] across parts of Washington, Idaho, and Montana. More than 90% of this particulate matter was in the respirable range ($< 10\mu m$ diameter) (Baxter et al., 1981) and recorded on air quality monitors as Total Suspended Particulate (TSP). The city of Yakima, located in eastern Washington, recorded[6] fine particulates ($< 2.5\mu m$ diameter) in excess of 33,000 $\mu g/m^3$ - almost 10,000 times the now current 24-hour standard.[7] Elevated levels of TSP were recorded at Yakima for 7 days following the ashfall; this transience is clearly seen in Figure 1.1 against ambient background readings through time. Furthermore, elevated levels of particulates were recorded inside homes, schools, commercial establishments, and automobiles in affected areas (Center for Disease Control, 1980c). The diffusion of this particulate matter through the atmosphere, and the subsequent deposition, created a pattern of particulate exposure to humans that was independent of demographic or socioeconomic profiles.[8] Prior to the eruption, the scale of the Mount St. Helens ashfall was neither expected nor predicted (Warrick et al., 1981). During the event, official emergency warnings ambiguously reported the extent of and the recommended response to the ash fallout.[9] After the event there were no coordinated public health messages on the effects of exposure

[5] Green et al. (1982) estimates the eruption ejected about 4km^3 of material into the atmosphere.

[6] Low flow rate, less than 24 hour sample.

[7] To meet current attainment standards the 98th percentile of 24-hour PM2.5 readings must not exceed $35\mu g/m^3$ averaged over three years.

[8] Actual individual exposure could be correlated with income or other social-economic characteristics, this possibility is discussed in Section 6.

[9] The first teletype message issued by the Washington State Department of Emergency Services was at 10:15am, 30 minutes after the ash cloud had passed over the city of Yakima, and about 1.5 hours after the eruption itself. This teletype simply stated, "precautionary measures are encouraged". The next emergency message was not issued until after 1:30pm - the ash had already passed into Montana. This teletype instructed residents to stay indoors.

to the ash fallout in part because the effects were not known.[10] In fact, the exposure to the particulate was thought to be relatively benign,[11] though there was little scientific evidence to support that assumption at the time.

Using the incidence of the ashfall as a quasi-experimental treatment for a particulate shock is relevant for three reasons. First, being uncorrelated with other types of air pollution it is unique in identifying the effects of a particulate shock.[12] Second, the discrete timing sheds light onto the prenatal effects of pollution. Third, it clearly identifies the importance of particle toxicity for a pollutant that is only determined by particle size. The remainder of the paper is organized as follows. Section 2 briefly describes background information on the connection between air pollution and health effects generally, as well as what is known about the specific health impacts of the Mount St. Helens eruption. Sections 3 and 4 present the data and methods respectively and describes the covariates in comparison between the affected and non-affected areas. Section 5 presents the main results and outlines robustness test outcomes. Section 6 discusses the validity of the natural experiment and details conclusions.

[10] The Mount St. Helens eruption spurred much of the subsequent research into the effects of volcanic ash on human health Horwell and Baxter (2006).

[11] The first hand account of Russel Ephraim, M.D., a Centers for Disease Control emergency team leader, reported a few days after the eruption that children treated the ash as a play thing such as sand or fresh snow by making wetted ash-balls to throw, or riding bicycles and tricycles through ash piles and down ashen streets to make ash plumes. Adults also seemed ambivalent to any potential risk since masks were given out but most people did not wear any kind of protection, even when actively cleaning or clearing ash (Center for Disease Control, 1980e).

[12] Criteria air pollutant levels recorded by the EPA during this period remained at normal ambient levels.

1.2 Background

1.2.1 Health effects of air pollution

The EPA regulates six pollutants under the National Ambient Air Quality Standards framework. The primary impact of these pollutants occurs through inhalation.[13] Each has a different patho-physiological pathway[14] but all generally lead to increased morbidity and mortality. In the case of particulates, this morbidity and mortality risk is a result of lung irritation, a generalized stress response, erratic heart rate, and dis-function of the cells lining the lungs. The causal effects have been inferred based on a now, very large number of studies that fall into one of several types.[15] Among these, epidemiological investigations are the most numerous and generally show a positive correlation between exposure levels and medical visits, shorter life expectancy, and mortality. However epidemiological studies suffer from identification issues that range from the confounding effects of weather, co-pollutant associations, seasonality, lifetime exposure, self selection, and concerns regarding harvesting[16] and/or life-shortening (Pope, 2000).

Considering the limitations of epidemiological studies, many researchers pursue empirical studies with clearer identification strategies. These strategies generally try to exploit variation induced by an argued exogenous event to overcome self-selection concerns. For example, Pope (1989) uses temporal variation in Utah TSP levels induced by a brief steel mill closure that resulted from a labor dispute. He finds respiratory related hospital admissions were significantly lower when the mill was closed. While Chay, Dobkin, and

[13]The EPA also regulates these six pollutants under secondary criteria for public welfare effects such as decreased visibility and damage to crops, livestock, vegetation, and buildings.

[14]Pathways and physiological effects of each criteria pollutant are briefly outlined in Table A1.1 in the Appendix.

[15]There are three broad types of study that constitute the body of evidence for causality used by the EPA. These are epidemiological studies, in vitro trials, and in vivo experiments.

[16]Harvesting anthropomorphically refers to Death metaphorically reaping those who are already sick or near the point of dying who might otherwise have died of some other factor in short order.

Greenstone (2003) use county level variation in TSP induced by the 1970 Clean Air Act Amendments (CAAA). They find that the regulation induced pollution reduction is not statistically associated with a reduction in adult mortality.

Natural experiments are not, in and of themselves, sufficient to overcome all the endogeneity issues listed above. For this reason researchers often focus on the infant health impacts of natural experiments to mitigate harvesting and life-time exposure concerns. During the temporary Utah steel mill closure mentioned above, Parker, Mendola, and Woodruff (2008) find fewer pre-term babies were born. Using the 1970 CAAA discontinuity, Sanders and Stoecker (2011) find that male baby births[17] drop by 3.1% as a result of a one standard deviation increase in pollution levels; using the same discontinuity, Isen, Rossin-Slater, and Walker (2014) find that a 10 point drop in TSP exposure during infancy results in a 1% increase in annual wage. Chay and Greenstone (2003) utilize county level variation in TSP levels from a short term recession in the early 1980's to leverage identification. They find a 1% reduction in annual TSP results in a 0.35% decline in infant mortality. Further, Sanders (2011) examines education outcomes for students born during that same time frame and finds a one standard deviation decrease in TSP level in the birth year is associated with an almost 2% increase in high school standardized test scores.

These studies focus their attention on infants who have arguably much lower lifetime exposure to pollutants and perhaps lower concerns about harvesting than those that estimate particulate effects on adults. However, endogeneity concerns still remain with regard to self-selection, and interactions with co-pollutants, weather, and seasonality. An

[17]Based on the hypothesized Trivers-Willard effect Sanders and Stoecker estimate between 21,000 and 134,000 fetal deaths were prevented by the CAAA. The Trivers-Willard effect is the hypothesis that mammalian females in poor maternal condition tend to produce more females than males, while those in good maternal health tend to produce more males (Trivers and Willard, 1973).

identification strategy that relies on induced variation over longer time frames and economic factors factors can result in differential impacts and effects by population groups on migration, unemployment, loss of insurance or access to healthcare, economic stress, etc. Since it is not clear how to control for these confounding factors, natural experiments or policy induced changes in pollutant levels are likely to still suffer from endogeneity. With regard to the ashfall that affected eastern Washington, there is no sense that residents chose to live in that part of the state as a trade-off between the likelihood of an eruption induced particulate shock and their income, or any other factor. As discussed below, the ashfall was a discrete random event that contained no other co-pollutants and was effectively exogenous.

Quasi-experimental settings offer clear advantages over epidemiological approaches. Researchers have also used this type of identification strategy to investigate the effects of air pollutants other than particulates. Notably, Moretti and Neidell (2011) use the incidence of ship traffic at the Ports of Los Angles as an instrument to measure the effect of ozone on area hospitalizations. Almond, Edlund, and Palme (2009) and Black, Butikofer, Devereux, and Salvanes (2013) use the, arguably exogenous, incidence of radioactive fallout over Scandinavian countries in the 1980's and 1960's respectively, to estimate the effect of prenatal radiation exposure on birth and educational outcomes. Currie, Neidell, and Schmieder (2009) investigate the effects of carbon monoxide on infant health by using variation in pollution exposure around birth between siblings. While Currie and Walker (2011) use difference-in-differences estimation to measure the effect of auto emissions on infant health before and after the installation of automated toll road equipment. Jayachandran (2009) exploits the incidence and timing of smoke across Indonesia from massive wildfires to estimate the effect on prenatal mortality, finding that exposure during the later end of pregnancy is the most damaging. Recently, Tanaka (2010) and Luechinger (2011) use air pollution policy changes in China and Germany respectively, to estimate the effect of airborne sulfur oxides on infant health. In all of these natural experiment or discontinuity settings however, co-pollutants still remain a relevant issue (among others).

Despite strong identification concerns about epidemiological studies of air pollution they inform the periodic EPA NAAQS review process.[18] With regard to particulate matter shocks, numerous, though not all, public health researchers find associations between spikes in particulate mater and human health impacts.[19] As a body of evidence, the EPA has determined causality between fine particulate shocks and cardiovascular effects, along with human mortality (U.S. EPA, 2009). However, current research is inconclusive on the relative toxicity of the different components and sources of fine particulate matter (U.S. EPA, 2011). Furthermore, the EPA, in categorically reviewing the epidemiological literature on the effects of particulates on infant mortality, concludes that there is no consistent window of exposure (U.S. EPA, 2009). Finally, there are no known studies on spikes in particulate matter levels and infant health in a quasi-experimental setting.

There are a few studies that examine the effects of average particulate exposure on fetal death and miscarriage.[20] Sanders and Stoecker (2011) is mentioned above. Mohallem et al. (2005) show that mice exposed to air pollution from Sao Paulo, Brazil (pollution that includes particulates) have fewer live births and more implantation failures.[21] While Pereira et al. (1998) find no association between Sao Paulo air pollution and human fetal deaths. Kim et al. (2007) find that stillbirths were positively associated with particulate levels in the third trimester of pregnancy for women in Seoul, South Korea. However,

[18] Section 109(d)(1) of 42 U.S.C. 7409 requires the EPA to review criteria pollutants and NAAQS every five years.

[19] See U.S. EPA (2009) for an extensive literature review.

[20] Fetal death (stillbirth) is the death of the fetus after 20 weeks of gestational age. Prior to 20 weeks, in-uterine death is categorized as miscarriage.

[21] Mohallem et al. (2005) did not find an effect on sex ratio vis-a-vis the Trivers-Willard effect.

Hwang, Lee, and Jaakkola (2011) find that first trimester exposure to particulate pollution is associated with an increase in stillbirths in Taiwan. Even so there remains a gap between our knowledge of transient particulate shocks and their effect on pregnancy outcomes.

The utilization of the Mount St. Helens eruption ashfall as a natural experiment in particulate exposure is clearly a powerful identification strategy that overcomes many significant problems in previous studies. It also fills a void in our understanding of the effect of a single particulate shock on infant/fetal health. The ashfall contains only particulate matter, there are no other co-pollutants as recorded by air quality monitors and investigated by ashfall composition assays. The ash is relatively inert – it establishes a baseline effect for all particulate matter regardless of source.[22] The ashfall is a discrete event, clearly discernible against the background of particulates and allows the determination of when particulate matter exposure is most pernicious to infant/fetal health. Furthermore, the ashfall affected a large population of pregnant women and infants at the same time.

1.2.2 Health effects of Mount St. Helens ash

The Centers for Disease Control confirmed 31 Mount St. Helens eruption-associated deaths as of August 22, 1980 (Center for Disease Control, 1980d). Despite TSP levels in excess 30,000 micro-grams per cubic meter[23] observed in some of the most heavily affected areas outside the blast and tree destruction area (Center for Disease Control,

[22] Constituent sources of particulate matter vary greatly by time and location. Generally most particulates originate from fossil fuel combustion and fugitive dust (U.S. EPA, 2009).

[23] The EPA 24hr standard for TSP at that time was $260\mu g/m^3$, with thresholds for the designations alert, warning, emergency, and significant harm at levels of 375, 625, 875, and $1000\mu g/m^3$ respectively (Center for Disease Control, 1980e)

1980e), there were no deaths associated with exposure to ash particulates.[24] In the ash-fall area there was, however, a significant spike in the number of hospital emergency room visits following the eruption with patients complaining of sore throats, shortness of breath, eye irritation, and emphysema like symptoms. But there was no increase in the number of patients diagnosed with chronic or acute bronchitis (Center for Disease Control, 1980a). Not long after the eruption the CDC concluded that there was no initial evidence of severe respiratory illness (Center for Disease Control, 1980b).

Many studies investigate or document the effects of the Mount St. Helens ash on humans, but none investigate the health effect on infants and birth outcomes. A number of these focus on the acute effects of ash exposure through hospital and emergency room records (Baxter et al., 1981; Nania and Bruya, 1982; Merchant et al., 1982; Baxter et al., 1983). Through surveys administered to two small rural towns (one in the affected area and one not) Shore, Tatum, and Vollmer (1986) report that the eruption caused some higher levels of mental anxiety and a possible elevation in risk for mental health issues. Adams and Adams (1984) find some evidence to conclude there was an increase in alcoholism and domestic abuse in a different rural town. Lines of investigation also examine the clinical effects of the ash on population subgroups, specifically children and loggers. These find short-term reduced lung function in some cases and no effects in others (Bernstein et al., 1981; Johnson, Loftsgaarden, and Gideon, 1982; Buist et al., 1983, 1986). Fruanfelder et al. (1983) examine the ocular effects of ash on loggers up to 18 months after the eruption, finding that the particulates were "well tolerated". Generally, these studies are small in scope and limited in the number of observations available for inference. Only one study related to Mt St. Helens ash has used a panel data set; Norton and Gunter (1999) investigate the relationship between respiratory disease and particulate matter in Idaho.

[24] Postmortem examination for 23 of the Mount St. Helen's deaths showed the cause to be asphyxiation from the inhalation of ash (Center for Disease Control, 1980a). All deaths were all immediately proximal to the volcanic blast.

Sampled particulate matter contained about 60% Mount St Helens ash. The authors do not find any correlation between particulate matter levels and respiratory disease in either the general population or in specific groups, such as farmers, typically exposed to higher levels of dust. Using the set of births in Washington, my paper is the first to investigate whether the ashfall had an effect on infants. It is also the first to reveal the evidence of fetal mortality as a result of exposure in the womb.

Other lines of research study the effect of Mount St Helens ash through in-vitro, and in some cases in-vivo clinical trials[25] (Fruchter et al., 1980; Beck, Brian, and Bohannon, 1981; Green et al., 1981, 1982; Craighead et al., 1983; Vallyathan et al., 1983a,b; McLemore et al., 1984). They show that exposure carries some risk of a fibrogenic effect and and a potential for pneumoconiosis for heavily exposed individuals. In general, however, the ash is decribed as non-toxic and similar in clinical effect to aluminum oxide dust which is considered relatively inert compared to other constituents typically found in particulate air pollution. That the ash, which constitutes the massive particulate shock, is considered bioreactively inert is a clear advantage in my quasi-experimental design. From this fact I am able to identify a lower bound on the effect of particulate shocks in general, other particulate constituents from combustion being more bioreactive. As found by the clinical studies of the ash above, my design also has the advantage of being unencumbered by other types of air pollution. Therefore, my research not only uncovers effects of the ashfall that have gone unnoticed, it also is also the first to identify effects of a single particulate shock.

[25]In-vitro refers to laboratory testing on isolated living cells, in-vivo refers to testing on whole living organisms, i.e animals.

1.3 Data

Data for this study comes from publicly available birth and mortality records collated by year by the Centers for Disease Control and Prevention.[26] The birth and death records for the State of Washington were extracted from these files for January 1973 through December 1988.[27] The extracted birth data constitutes the micro-level dataset of individual births used in the birth weight analyses. To create a panel of county-month infant mortality rates I aggregate birth and mortality data into month and county of residence cells, after deaths due to external causes are filtered out.

To determine treated and untreated observations I use the ashfall isomass contour lines from Sarna-Wojcicki et al. (1981).[28]. The extent of the ashfall (above trace amounts) across Washington state is presented in Figure 1.2, and shows the 8 counties designated as treated and 15 non-treated counties. My research design confines the investigation into the effect of the ashfall TSP shock to a binary treatment and control setting. While it is recognized that there is a heterogeneous distribution of TSP exposure, due to limitations in the publicly available data, it is not possible to identify exposure at a finer level than the county. Furthermore county level exposure is not directly determinable. Even though the work of Sarna-Wojcicki et al. (1981) describes the thickness of the ash deposited across Washington, is it not entirely clear how this relates to particulate levels in any given county. The historical EPA TSP monitor data is only available at the average weekly level and there were only two monitors located in the affected area at the time of the eruption. In addition the thickness of the ashfall is not a sufficient condition

[26] Data available at <http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm>

[27] The universe of birth records for Washington State are only available from 1978, before 1978 only a 50% sample is available – each of these records is doubled in the dataset used. After 1988 the date of birth in the natality files, and the date of death in the mortality files is not available.

[28] Counties on the margin of the ashfall and those counties that received some ashfall but where the population largely resides outside the affected area were not included in the primary analysis. These marginal counties are included later as a robustness check.

to estimate the exposure levels. The winnowing effect time and distance from the eruption, as well as idiosyncratic atmospheric mixing patterns preclude the use proximity or similar surrogates for county level exposure.

Given the prevailing winds on the day of the eruption were westerly, the eight treated counties lie in eastern Washington, a region that is primarily agricultural and dry. The fifteen non-treated counties are also generally rural and, for the most part, west of the Cascade ranges along the Pacific coast. Pre-eruption descriptive statistic comparisons between treated and non-treated counties are presented in Table 1.1. Between 1978 and 1979 there was no statistical mean difference between the percentage of males born, percentage white mothers, and percentage of babies born with no prenatal care. Though the other cofactor means are very similar, there are some statistical differences. About 2 per 100 fewer babies born received early prenatal care in the treated counties on average. Gestation length is significantly shorter in the treated counties by 1.5 days, mother's are younger by 3 months on average, and about 1 per 100 more births are to teenage mothers. About 3 per 100 more births on average are to immigrant mothers, while 2 per 100 more births are to non-married women, but 2 per 100 more births are delivered in hospitals in the treated counties as a whole.

Although I utilize a binary treatment for the ashfall TSP shock, it is relevant to compare the TSP levels for each of the two groups of counties before the eruption. Historical air quality data, available from the US EPA Air Quality System,[29] reveals a statistically significant mean difference in the levels of TSP. Between 1978-1979 the TSP levels appear to be parallel and trending slightly upward. Treated counties are, on average, $20\mu g/m^3$ higher than the non-treated group as shown in Figure 1.3. Using the point estimate results from Chay & Greenstone (2003), this would imply an estimated 1 per 1000 dif-

[29]EPA AQS data accessed from <http://www.epa.gov/ttnairs1/airsaqs/archived%20data/index.htm>

ference in the infant mortality rate, and a 6 gram difference in birth weight, between the two groups of counties.

Through the 1970's infant mortality was typically higher in the treated counties. On average, the annual difference was around 2.6 more infant deaths per 1000 births in the treated counties. Figure 1.4 plots the annual infant mortality rates during the 1970's between the two groups of counties, and generally shows a downward parallel trend. Comparing the number of infant deaths for May, June, and July, 1980 with the same period the previous year, there were 13 more infant deaths in the treated counties, compared to 17 more infant deaths for the non-treated counties. As shown in Table 1.2, taking into account the differential number of live births, the treated group infant mortality rate increased by 3.7 while the non-treated group rate only rose by 2.8, with a seeming impact of an extra 0.85 deaths per 1000 births as a result of the ash fall.

While basic comparisons of the data show that infant mortality apparently increased slightly over the same 3-month period from the previous year, mean infant birth weight was also higher. Table 1.3 presents the mean infant birth weight for the treated and non-treated counties for May, June, and July, 1980 and the same three months one year earlier. Average infant birth weight increased by 31 grams for babies in the treated group over those three months relative to the non-treated infants.[30] This outcome seems counter intuitive, but, as discussed in the next section, this comparison of birth weights does not take into account an upward bias by the failure to account for any increase in fetal deaths and miscarriages of at-risk fetuses.

[30] The mean difference is significant at the 10% level using a one-way ANOVA.

1.4 Methods

In order to address the issues of harvesting and lifetime exposure I focus my analysis on infant outcomes. Though infants are not a perfect group that has neither been exposed to pollution prior to the ashfall nor a likelihood of imminent death due to prior conditions, they are at least considered to be a group of minimal concern relative to older populations. I hypothesize that air pollution, and the ashfall particulate shock specifically, may affect infants prior to birth, and infant health after birth. Given the paucity of available infant health data I limit my analysis of these hypotheses to measuring the effect of the ashfall on infant birth weight and infant mortality.

1.4.1 Infant birth weight

As a measure of pre-natal infant health, birth weight is a common proxy variable in many studies, including those that investigate the effect of pollution on infant health.[31] Using a baseline difference in differences approach to estimate the average treatment affect of exposure to the ashfall TSP shock whilst in utero, the following log-linear model is estimated:

$$\begin{aligned} ln(birth\_wgt)_i \quad = \quad & \alpha_0 + \alpha_1 I(ash\_county)_i + \alpha_2 I(in\_utero)_i \\ & + \alpha_3 [I(ashcounty)_i \times I(in\_utero)_i] + \epsilon_i \end{aligned} \tag{1}$$

where i indexes the individual. $I(ash\_county)_i$ is an indicator variable that takes on the value 1 if the county where the individual is born is designated as 'treated' (the county received more than a trace amount of ashfall, see Figure:1.2) and a value of 0 otherwise, while $I(in\_utero)_i$ is also an binary variable that takes on the value 1 if the baby was in utero on the date of the eruption (May 18th, 1980) and 0 otherwise. The interaction

[31]For example see Chay and Greenstone (2003); Almond, Chay, and Lee (2005); Almond, Edlund, and Palme (2009); Currie, Neidell, and Schmieder (2009).

between the two indicator variables is the regressor of interest with α_3 the estimator for the average treatment effect across all babies exposed in utero. Given the hypothesized deleterious effect of TSP exposure on fetal health α_3 is assumed to be negative. In the case of complete random assignment of the treatment, model (1) would be sufficient for modeling an unbiased and consistent average treatment effect. However, since the treatment was not randomly assigned but imposed through a quasi-experimental natural event I add successive controls to Model (1). Models (2), (3), and (4), additively include relevant birth cofactors, month and year fixed effects, and county fixed effects respectively. Birth cofactors include gestation length and mothers age along with dummy variables for early prenatal care, no prenatal care, birth at a hospital, whether the baby is male, one of a multiple birth, born prematurely, and/or white, and whether the mother is married, an immigrant, and/or less than 18 years old.[32]

The ashfall TSP shock may affect infant birth weight depending on when the infant was exposed during the fetal development cycle. In order to investigate this I partition out the in utero indicator variable in Model (1) by trimester[33] in the following way:

$$\begin{aligned} ln(birth\_wgt)_i &= \beta_0 + \beta_1 I(ash\_county)_i + \sum_{j=2}^{4} \beta_j I(in\_utero\ trimester_{j-1})_i \\ &\quad + \sum_{k=5}^{7} \beta_k I(in\_utero\ trimester_{(k-4)})_i \times I(ashcounty)_i + \epsilon_i \end{aligned} \tag{5}$$

where the subscripts on *trimester* indicate the 1st, 2nd, or 3rd, trimester of pregnancy. There are now three indicator variables that describe whether or not the baby was in utero during a given trimester at the time of the eruption. The three difference in differences interactions are now the regressors of interest. Model (5) is further augmented by

[32] Gestation length and male births may themselves be outcome variables. The effect of the ashfall on these two variables is mentioned in the Results section.

[33] While embryonic and fetal development is a continuous process it is generally divided into three time periods representing embryogenesis, the beginning of the fetal period, and the beginning of the perinatal period.

including the successive controls previously mentioned and constitute regression models (6), (7), and (8).

A critical limitation of the models described above is the assumption that, conditional on fetal health, the likelihood of miscarriage and fetal death is unaffected by exposure to the TSP shock. This is to say that the error term (with respect to fetal health) is uncorrelated with exposure while in the womb and that there is no omitted variable bias. However, birth weight is the primary proxy for the health of the baby during fetal development. Larger fetuses are generally more healthy, have heavier birth weights, and are less likely to be terminated than their smaller counterparts. Failure to account for the possibility that less healthy fetuses exposed to the TSP shock may be at a higher risk of termination would result in an upward bias in the average treatment effect as measured by birth weight. It would be the more healthy babies, and thus the heavier babies, that are more likely to be born and have their birth weights recorded. Since there is no available data on fetal health other than the proxy 'birth weight' it is not possible to model birth weight conditional on fetal health. Given this limitation is not directly surmountable I model the effect of the TSP shock on the distribution of birth weight outcomes. I estimate the following linear probability model to characterize the distribution of birth weight:

$$
\begin{aligned}
I(birth\_wgt_i < a) \quad = \quad & \alpha_1 I(ash\_county)_i + \alpha_2 I(in\_utero)_i \\
& + \alpha_3 I(ashcounty)_i \times I(in\_utero)_i \\
& + x_i\gamma + T_i + C_i + \epsilon_i
\end{aligned}
\tag{9}
$$

where a, in grams, is the set $\{2000, 2500, 3000, 3500, 4000, 4500\}$ which roughly corresponds to the 2.5th, 5th, 20th, 50th, 85th, and 97.5th percentiles respectively.[34] Also, x, T, and C designate birth cofactor variables, month and year fixed effects, and county of birth fixed effects respectively.

Model (9) estimates the likelihood a baby is born below a particular partition of the birth weight distribution - the regressor of interest is again the difference in differences intersection of being born in a county that was affected by the ashfall and being in utero at any point during the nine months of fetal development at the time of the eruption. As above, I also separate out the in utero indicator based on trimester of fetal development. The linear probability model now splits out the effect of the TSP shock on the distribution of birth weight across each of the three trimesters as specified by:

$$\begin{aligned} I(birth\_wgt_i < a) &= \beta_1 I(ash\_county)_i + \sum_{j=2}^{4} \beta_j I(in\_utero\ trimester_{j-1})_i \\ &\quad + \sum_{k=5}^{7} \beta_k I(in\_utero\ trimester_{k-4})_i \times I(ashcounty)_i \\ &\quad + x_i\gamma + T_i + C_i + \epsilon_i \end{aligned} \tag{10}$$

where the set a is defined as above, and other variables as previously defined.

1.4.2 Infant mortality

In order to model the effect of the ashfall TSP shock on infant mortality I begin by defining the infant mortality rate in the standard fashion. Specifically, infant mortality rate is the number of infant deaths[35] in a given time period divided by the number of

[34] According to pediatric lexicon, a birth weight of 2500 grams or less is categorized as "low birth weight".

[35] Infants are categorized as children less than one year old. Intentional or accidental deaths are not included in this statistic.

births in that same period. Using a difference in differences model I estimate:

$$
\begin{aligned}
inf\_mort_{it} &= \alpha_0 + \alpha_1 I(ash\_county)_i + \alpha_2 I(ash\_timing)_t \\
&\quad + \alpha_3 I(ashcounty)_i \times I(ash\_timing)_t + \epsilon_{it}
\end{aligned} \tag{11}
$$

where i indexes county and t indexes month. $I(ash\_timing)_t$ is an indicator variable that takes the value of 1 for every county-month observation up to one year after the May eruption. A one year window of effect is chosen arbitrarily so as to encompass observations that not only include the effect on babies already born at the time of the eruption and who may succumb to the deleterious effects of the TSP shock over time, but also those babies that were treated in utero who were yet to be born up to 9 months later.[36] The coefficient on the interaction between the ash-county and the ash-timing indicators capture the average treatment effect of exposure to the ashfall TSP shock. Given the quasi-experimental setup I additively expand Model (11) to include controls for relevant grouped birth statistics - Model (12), month and year fixed effects - Model (13), and county fixed effects - Model (14). The grouped birth statistics include the percentage of births that are white, male, born in a hospital, a multiple birth, born premature, born to married parents, born to an immigrant mother, born to a teenage mother, had no prenatal care, and had early prenatal care.

Infant mortality risk can be characterized by a steep decay function as the child lives longer. Unconditionally, babies less than one month old are at a much higher risk of death than those close to one year old. Considering that the death of infants who are older than one month may be idiosyncratic, I adjust the numerator of the infant mortality rate to include only those infants who die that are less than one month old. I thus re-estimate the four models above as Models (15) through (18) with the adjusted

[36] Adjustment of the timing window results are discussed later in the paper.

mortality rate regressand to reduce the noise in the data.

A significant drawback in using the conventional measure of infant mortality above is, at worst, an implicit assumption that the effect of the TSP shock is homogeneous for infants of any age, and at best, it attenuates the hypothesized effect. Some infants in the numerator of the infant mortality statistic were exposed while in the womb and others exposed after they were born - the effect could be different depending on when and how the infant was exposed. To address this issue I construct the cohort infant mortality rate - a statistic that divides the number of babies born in a given month that subsequently die by the total number of babies born in that same month. To be clear, the infant deaths counted in each cohort rate observation share a common month of birth, where as in the standard mortality statistic they share a common month of death. The birth cohort mortality rate allows me to parse out the effect of the TSP shock on infants already born and on those that were in utero. I arbitrarily split out the babies born before the eruption into two groups – those born up to 3 months before the eruption and those born between 3 and 6 months before. The mortality cohort rate for babies that were in utero during the ashfall are separated out into trimester groups. I then estimate the difference in differences models (19) through (22) as such, which include successively more controls as above.

1.5 Results

1.5.1 Infant birth weight

Using micro data of births from Jan 1st, 1973 through Feb 18th, 1981, Table 1.4 presents the estimation results of Models (1) through (4). Recall these models estimate the basic difference in differences estimates for the effect of the ashfall TSP shock on mean infant birth weight in logs. In Model (1) the interaction between treated counties and those infants in the womb at the time of the ashfall is highly significant, suggesting that infant birth weight was on average about 9 percent greater than those babies not treated. Though this result is counter intuitive, the effect size diminishes once covariates are added in Models (2) though (4) and it is no longer significant.[37]

Per models (5) through (8), the results when the in utero timing window is split out into trimesters are presented in Table 1.5. In model (5) the basic difference in differences estimate between treated and non-treated infants at the 95% level seems to indicate infants treated during their 1st or 3rd trimesters are on average heavier by around 10 percent.[38] However, once controls are included in the regression, the mean birth weight difference in the 1st and 3rd trimester effect falls to about 8 percent and are significant only at the 10% level. The point estimates for the 2nd trimester effect in these models is not significant. These counterintuitive and slightly ambiguous results are cleared up in the results for Model (9) below.

The distributional effects of the TSP shock, Model (9), are reported in Table 1.6. These results show that babies that were in utero at the time of the eruption in the group of

[37]There is no statistically significant effect of the ashfall in these models when gestational length and percent male babies are used as regressands.

[38]The respective point estimates for Models (1) though (4) are a weighted average of the results partitioned by trimester in Models (5) though (8).

counties that experienced the ashfall are statistically under represented in the lower tail of the distribution. This accounts for the counter intuitive mean result found in Model (8). The effect is particularly strong and precisely estimated in the group that is less than 3000 grams at birth, a weight corresponding to about the first quintile. There is an estimated 2.45 percentage point decrease in the likelihood that a given baby is born in the 1st quintile as a result of exposure to the TSP shock while in utero.

The results of Model (9) also show the TSP effect on the distribution of birth weights did not apparently affect the median and upper ranges where there is an increasingly precise estimated effect around zero. One explanation for this result is that there was not a distributional right shift as is perhaps ambiguously implied by the results of Models (1) and (5)-(8), but that babies who would normally "fill out" the lower range of the birth weight distribution are missing, thus making it appear that the mean birth weight baby was born heavier, when in fact this is not the case. This post-ashfall pinching of the distribution as seen in Figure 1.5.[39] Given that during this period there were 9,055 babies born in the treated counties - 1,613 of which were born into the less than 3000 gram segment, the point estimate for babies less than 3000 grams would imply there are 227, or about 1 in 8 babies are missing from the bottom quintile.[40] Within the bounds of a 95% confidence interval, the accuracy of this estimate varies in effect from between 161 to 295 babies. Using the point estimate for babies less than 2000 grams the effect is larger, about 1 in 6 babies appear to be missing from this weight group.[41]

[39]The kernel density graph for the treated counties could also be seen as a post-eruption right shift in the distribution, but careful inspection reveals the density distortion between 2500 and 3000 grams, where there is no such distortion on the right side of the density. Furthermore, a symmetric right shift is not supported by the results of Models (9) and (10).

[40]The number of missing births is calculated by inflating the total number of births in the treated counties during this period by the reciprocal of 1 plus the point estimate. The proportion of missing births in the weight segment is simply the estimated number of missing babies divided by the total number of babies in the weight group (actual number plus the estimated number missing) expressed as a ratio.

[41]Though the point estimate is smaller for this weight group there are also fewer babies born into this group. The net effect, on balance, is larger.

The results of the estimation for Model (10), where the in utero indicator is split out into trimesters for the linear probability model, are presented in Table 1.7. These results further refine the distributional findings above. Babies exposed to the TSP shock while in utero during their 1st trimester were significantly less likely to be born in the 20th percentile. The magnitude of this effect diminishes with the age of the fetus and strongly suggests that the effect of the TSP shock is more deleterious on younger fetuses. This pattern of effect is mirrored in the point estimates for the low birth weight category (babies less than 2500 grams). As above, the point estimates at the median and in the higher percentiles of the distribution show no change in likelihood that babies were born with heavier birth weights. From these results I find no evidence to suggest a mean shift in the distribution of birth weight, but a pattern consistent with the hypothesis that fewer babies were born in the lower tails of the birth weight distribution.

My finding that the TSP shock is most pronounced during the first trimester and for those fetuses that would have filled out the lower distribution of birth weight is largely in line with teratogenic principles (Wilson and Fraser, 1977). Focusing on the estimated effect of the TSP shock on babies in utero during their first trimester; it is 3.21% less likely that a baby is born in the 20th percentile as a result of exposure. This result implies there were 94 fewer babies born in the lowest birth weight quintile and represents an attrition of 1 in 6.25 babies. The strength of the TSP shock increases for the 5th percentile; the point estimate implies that almost 1 in 4 babies who were exposed in the first trimester are missing.

In the spirit of Granger (1969) and Autor (2003) I expand the births micro-dataset to include births for the treated and non-treated counties up to Dec 31, 1998,[42] and employ

[42]This is the last date for which individual birth date data is publicly available.

a series of leads and lags on the difference in differences variables for Models (9) and (10). In this manner I estimate a series of falsification tests on the data, forward and backward through time, as if the ashfall happened in different 9-month periods, in the case of Model (9), and 3-month intervals, in the case of Model (10). From Model (9), the estimated effect of the TSP shock from Table 1.6 is dramatically seen at t=0 relative to the time periods before and after, on the less-than-3000 gram birth weight group in Figure 1.6. These results show that there appears to be no discernible affect of the ashfall on birth weight before the eruption, as well as after the 9-month window following the eruption. In other words there seems to be no effect on mothers who gave birth right before the eruption nor on those who conceived shortly afterwards either. The leads and lags also reveal the stability of the median (3500 grams) and higher percentiles (4000 and 4500 grams) to the TSP shock when the timing of effect is measured in the 9 month in utero window. Figure 1.7 displays the estimated effect of the TSP shock on the birth weight distribution using a series of 3-month leads and lags. This mimics Model (10) for the expanded micro-dataset. Similar to the results found for Model (9), the leads and lags reveal the stability of the median and higher percentiles. Also as found, there is a dramatic decrease in the likelihood that a baby is born in the 3000 gram or less range during the actual window of exposure. The figure also shows the relative strength of the effect for babies exposed in their first trimester, an effect that is lower for babies in their 2nd and 3rd trimester of fetal development.

The addition of county-specific time trends attenuates the point estimates of Models (9) and (10) but the basic result remains.[43] This attenuation is the largest in model (9) where the point estimate in the $<$ 3000 gram segment is reduced by almost one-half, but it remains strongly significant. Though the 95% confidence interval of the baseline Model

[43]Regression results with the addition of county-specific time trends for Models (9) and (10) are reported in the Appendix in Table A1.2 and Table A1.3 respectively.

(9) $<$ 3000 gram estimate covers the point estimate when county-specific time trends are employed, this result could indicate dynamic effects which are not easily captured in my difference-in-differences estimator. Potential dynamic effects are discussed in the next section. Adding county-specific time trends does not substantially alter the results above when I expand the dataset to employ leads and lags. Neither does replacing year and month indicators with year×month dummies. Similarly, re-estimating the models using the excluded counties that received a trace amount of ashfall or were marginally contiguous to the ashfall area (see Figure 1.2) does not appreciably alter the basic results.

Another consideration is lumpiness in the birth weight data. Weight at birth was typically measured in integer ounces and, as such, results in a distribution of mass points over the range of birth weight (in grams). This kind of mass point distribution can lead to estimation troubles when discrete regressors are used (see Almond et al. (2010); Barreca et al. (2010); Barreca, Lindo, and Waddell (2011); Barreca et al. (2011)). For this reason I check the robustness of the Model (9) results by two methods: I employ a quasi-Monte-Carlo technique detailed below, I also run the linear probability model using 1-gram increments of the difference-in-differences estimator. As expected, using the latter method results in some stair stepping estimates that reflect the nature of the mass points across integer ounce measurements. Despite this sensitivity in the indicator regressor, there are no instances where the point estimate makes significant jumps. The estimators are relatively robust.[44] For the Monte-Carlo technique, I assume that baby weight, measured in integer ounces, results in a pure measurement error of up to 0.5 ounces, and add noise (up to $\pm$ 14 grams) to each birth weight observation. I then re-estimate Model (9) and iterate – recording the point estimates for each iteration. The resultant bell-shaped point estimate distribution ranges between -0.0264 and -0.0208 for the $<$ 3000 gram weight group. Recall, the point estimate obtained from Model (9) was -0.0245, this

[44] The results of the estimation process is shown in the Appendix in Figure A1.1.

result appears to be robust in the context of the mass point distribution of birth weight.[45]

In addition to checking the robustness of the point estimates I check the robustness of the standard errors. Specifically with regard to the cluster-robust variance estimator, which I employ at the county level across models with 23 groups, researchers are concerned that few groups, correlations between groups, and asymmetric groups can lead to understated standard errors and thus, overstated statistical significance (Cameron and Miller, 2010). Carter, Schnepel, and Steigerwald (2013) indicate that test size rapidly increases as the number of groups descends below 20, and that significant reduction in the effective number of groups can occur with inter-correlation between groups. Cameron, Gelbach, and Miller (2008) find, in the presence of a small number of clusters, bootstrapping can be effective with as few as 6-8 groups. I employ the wild cluster bootstrap-t with H_0 imposed as outlined and recommended by Cameron, Gelbach, and Miller (2008) and find some inflation compared to the t-distribution for the results in Model (9). However, the statistical inference based on the cluster-robust standard error remains unchanged and the difference-in-differences estimates remain significant at the levels indicated in Table 1.6.

1.5.2 Infant mortality

Using monthly panel data by county between Jan 1972 and Feb 1982, I estimate the effect of the TSP shock on the infant mortality rate. The regression estimates for Models (11) through (18) are found in Table 1.8. Recall the regressand in Models (11) through (14) is the standard formulation of infant mortality (the numerator is the number of children less than one year old who die within the given month), while in Models (15) through (18) I adjust the numerator to be only those infants less than one month old

[45]The cluster robust standard errors for all iterations are significant at the 5% level. The distribution of point estimates is shown in the Appendix in Figure A1.2.

who die during the month. The interaction term is not of the expected sign and is not statistically significant in any model specification. As anticipated, the standard error is smaller on the interaction term for Models (15) through (18). However, the negative sign remains and the point estimates are not significant. Though a positive effect of the ash fall on infant mortality cannot be ruled out,[46] there appears to be no statistical basis to conclude that the TSP shock had the expected impact on infant mortality. In order to check this result, I run a Poisson regression on the deaths count data by county-month. Consistent with the results above, the Poisson regression point estimates for Models (14) and (18) are also not of the expected sign and not significant.[47]

Table 1.9 shows the regression results of reformulating the data from the standard infant mortality rate to birth-month cohort mortality rates. This allows me to estimate the effect of the TSP shock on infant mortality depending on which infants were in utero at the time of the eruption and which had already been born. In Model (22) there is a weakly significant effect on the mortality rate for infants who had been born 0 to 3 months before the eruption. However, the sign of the effect is negative and not easily interpretable. This sign is consistent when using a poisson regession model on the count of birth month cohort deaths and is not dependent on an influential observation or county, but it is sensitive to the assumed error variance-covariance structure.[48] The signs on the estimated coefficients for those babies exposed whilst in utero are also negative, but none are significant. Only the coefficient on the effect of the TSP shock for infants 3 to 6 months old at the time of the eruption is of the expected sign. However, this estimate is also not significant.

[46] One-sided tests on the interaction point estimate include zero.

[47] Poisson regression results are reported in the Appendix in Table A1.4.

[48] The standard error is larger using heteroskedastic robust estimation compared to the cluster robust estimation I use in the paper. This result is discussed briefly in the Appendix.

1.6 Discussion

1.6.1 Threats to validity

It is recognized that this study, like many natural experiments, is not without shortcomings. The treats to the validity in this case include dynamic effects, such as unobserved compensatory actions, and correlated or causal changes in activity and circumstance, and the whether or not the ash fall exposure constituted a TSP shock.

In general, it is likely that women who were expecting the birth of their child, or mothers and/or care-givers of infants, might have acted to compensate for a perceived health impact to themselves or their child as a result of the ash fall. Infants and pregnant women may have remained indoors for extended periods, visited their doctor more frequently, or moved out of the affected area. Furthermore, these hypothesized actions could be correlated with unobserved characteristics of the mother (such as income). Attenuation would result in the first case, the second could threaten identification. Such compensatory actions, have the potential to reduce the validity of the quasi-experimental treatment on birth weight and infant mortality. It has been discussed, though, that high levels of particulates were measured in all common indoor settings,[49] and that there was generally a high degree of ambivalence to ash exposure. Furthermore, interstate highways in the ashfall affected areas were closed resulting in no opportunity to leave the affected area for a number of days.

Any impact of the ash fall would also be a function of the associated effects as a result of a disruption in routine and/or circumstance. Higher levels of stress, concern, and uncertainty as a result of the eruption, increased time indoors and associated sedentary

[49] See Chen and Zhao (2011) for an overview and review of studies measuring the correlation between indoor and outdoor particulates.

activities, loss of real income, ingestion of ash on food or in water, etc, are potential concerns that affect the validity of the treatment. A mother's exposure to stress inducing events while pregnant is known to affect birth outcomes and infant mortality to some degree (Mansour and Rees, 2012; Currie and Rossin-Slater, 2013). It is possible that the results I find are a result of associated effects and not directly from the TSP shock itself. With regard to job loss, there appears to be no dip in State employment following the eruption, and the Monthly Labor Review published by the Bureau of Labor Statistics makes no mention in 1980 of any reference to the eruption.[50] However, employment is only one factor for a singular event that still remains in the consciousness of those affected by it.

In the case of ingesting water contaminated with ash – most of eastern Washington is arid or semi-arid with water pumped from relatively deep wells – there was little concern about contamination Warrick et al. (1981). Furthermore, the Food and Drug Administration performed laboratory testing of ash, it's solubility in water and acids, and found no hazardous trace elements. Milk samples were also analyzed with no evidence of health hazards (Center for Disease Control, 1980f). Never-the-less, the perception of risk of ingesting something harmful as a result of the event may have produced a placebo effect that cannot be ruled out.

The effect of the air quality shock from the ash fall may be more complex than the simple difference-indifferences estimator that I have used to model it. For this reason I estimate county-specific time trends and models with leads and lags. However, as discussed by Wolfers (2006), further dissection of the estimated effect may exacerbate or ameliorate a potential bias by mis-specifying the actual dynamic effects. This is certainly one reason

[50] Among similarly weighty employment matters in 1980, the MLR reports on events entitled *"Kansas City bakery workers get new contract"* and *"GPO female bookbinders awarded back pay"*.

for limiting the primary analysis to nine months after the eruption.[51] Another is that the exact physiological pathway of particulate pathology is also not well understood.

It may also be the case that what I have described as a TSP shock was in fact a massive dust shock. This is to say that the results I find are due more to the effects of larger dust size ash particles than the much smaller suspended particulates. In addition, the occurrence of dust particles with smaller particulate matter could mitigate the efficacy of the TSP shock, since, as shown by Horwell and Baxter (2006), smaller particles often adhere to coarser particles; this in turn reduces the ability for fine particulates to reach deep into the lung. Never the less, the scanning electron microscope images by Horwell and Baxter reveal particles in Mount St Helens ash uniformly less than 10 microns; the ashfall was indeed a TSP shock – if not a PM10 shock.

1.6.2 Costs of particulate matter shocks

While the particulate pollution shock of Mount St Helens is certainly an outlier event relative to those regulated by the EPA, it clearly delineates the effect of a single particulate event on at-risk fetuses. I use my results to compute the cost of a single 150 $\mu g/m^3$ PM10 particulate shock event,[52] as well as the benefit from the 1987 reduction of the 24-hour PM10 primary standard from 260 $\mu g/m^3$ to 150 $\mu g/m^3$.

Assuming a simple linear effect without a threshold, and using the result for the <3000 gram birth weight group from Model (9), I calculate the effect of a single, hypothetical,

[51] The exogenous loss of a fetus, either from a discrete event or from a systemic change, may result in a number of dynamic effects which may be difficult to estimate. For example, an expectant mother who miscarries may or may not seek to become pregnant again. In aggregate this could either lead to a later surge in births, or a dip as potential mothers differentially forgo having additional children that they might have conceived were it not for the exogenous fetal death that they suffered.

[52] 150 $\mu g/m^3$ is the current 24-hour standard set by the EPA, and has been since 1987.

150 $\mu g/m^3$ shock as a 0.000205 percentage point decrease in the number babies born.[53] Assuming this shock affects a metropolitan US area of about median size (population 250,000), and assuming the birth rate of this hypothetical city is the current national average (13 births per 1,000 population per year)[54], and that $\frac{3}{4}$ of the babies to be born within the next year are in utero[55] at the time of the particulate shock, I calculate that about one half (0.5) of a birth on average would be terminated as a result of a single 150 $\mu g/m^3$ PM10 shock.[56] Using the EPA's value of a statistical life,[57] I estimate the cost (only in terms of statistical life) of a single particulate shock at \$3.7 million (2006 dollars).[58]

To calculate the benefit from the 1987 reduction in the 24-hour PM10 standard I count the number of single-day PM10 events in excess of 150 $\mu g/m^3$ across all States for 1986 and 1988, then naively ascribe the difference to the 1987 policy change. In 1986 there were 187 events exceeding the 1987 24-hour standard , while in 1988 there were 153 such events between same counties[59] as measured by air quality monitors.[60] Using the county population estimates from the U.S. Department of Commerce[61] and the national average birth rate for 1988[62] I estimate around 80 births would have been terminated

[53]Estimate derived by dividing 0.0245 percent points by the average of the two known particulate readings, one in Yakima and the other in Spokane, on the day of the eruption, i.e. 18k $\mu g/m^3$, then multiplying by 150.

[54]NCHS CDC FastStats for 2010 http://www.cdc.gov/nchs/fastats/births.htm

[55]Gestational length is approximately 9 months – thus is it reasonable to assume at a given point in time, on average 3/4 of the babies to be born within the next 12 months are in utero.

[56]Calculation: $0.4996 = 250\text{k} \times 13 \times 0.000205 \times 0.75$

[57]EPA Mortality Valuation Guidelines estimate the value of a statistical life at \$7.4 million (2006 dolars), http://yosemite.epa.gov/ee/epa/eed.nsf/webpages/MortalityRiskValuation.html

[58]This value may understate the full cost of a single particulate shock since it does not include impacts to others (including the mother), both aesthetic and medical.

[59]There are 319 counties with single-day PM10 readings common between both years. The single-day sampling rate is higher in 1988 for those same counties by 2842 observations. Thus the difference between the number of PM10 events is attenuated and may be much higher.

[60]PM10 daily monitor data available from http://www.epa.gov/ttnairs1/airsaqs/archived %20data/index.htm

[61]http://www.census.gov/popest/data/historical/1980s/county.html

[62]Vital Statistics of the United States, 1988. Volume I, Natality.

had it not been for the reduction in 150 $\mu g/m^3$ PM10 events. This reduction in the estimated number of terminations corresponds to a benefit (only in terms of statistical life) of $592 million (2006 dollars). This estimate could be considered an underestimate given most particulates originate from the combustion of fossil fuels and are considered more deleterious than the relatively inert ash particulate from which the calculated effect is based. Furthermore, this estimate does not include the value of other health and aesthetic benefits.

The assumptions I make in deriving these estimates are certainly strong. However, for reasons outlined previously, I consider these estimates as lower bounds. To my understanding these estimates represent the only values placed on a single particulate matter shock. Given the EPA has continued to strengthen the 24hr PM standards since the inception of the Clean Air Act, these estimates are long overdue. Looking forward it would be advantageous to estimate the marginal effect of a single particulate shock in an quasi-experimental setting. It could be possible to utilize the Mount St. Helens ash fall in this endeavor by correlating the depth of the ashfall and the distance from the eruption with particulate levels. Determining this relationship would require a significant level of effort since there are only two particulate matter readings on the day of the eruption and there are significant differences between the ash mass to depth ratio across the affected area. Furthermore, this line of investigation would also require locational information of births at a finer scale than the county level. Despite this, the Mount St. Helens eruption might represent one of the best settings in which to estimate the marginal effect of a particulate shock.

1.7 Conclusion

I use the incidence of ashfall from the 1980 eruption of Mount St. Helens as a quasi-experimental treatment for exposure to a single particulate shock on infants. Ninety percent of this ashfall is in the coarse to fine size range ($> 10\mu g/m^3$) and categorically PM10 or lower. I find no evidence to suggest that exposure to this shock increased infant mortality, but do find significant evidence to indicate babies were prematurely terminated while in the womb. It could be the case that these premature deaths harvested the fetuses likely to have died as infants after birth, but by carefully utilizing the discrete timing of the event, I find no evidence to suggest that the infant mortality rate dropped among babies born after the eruption either. What is striking are the statistical results suggesting that babies that would have filled out the lower range of the birth weight distribution are particularly vulnerable to a particulate shock. As a result of exposure in utero I estimate there were about 227 fewer babies born in the 20th percentile of the birth weight distribution. Also striking is the evidence to suggest that these at-risk fetuses seem to be much more susceptible to the terminal effect of a particulate shock in the early stages of development. Of those at-risk fetuses exposed during the first trimester of development I estimate around 1 in 6 suffered early termination. Extrapolating these results down to the current 24-hr standard for coarse particulates I calculate that a 150 $\mu g/m^3$ shock on a median sized U.S. metropolitan area is likely to result, on average, in the early termination of one half of a birth.

This quasi-experimental setting and the results highlight effects of suspended particulates (and air pollutants in general) that until recently have received little consideration - that is their deleterious effect on babies whist in utero. This and other recent studies clearly indicate that air pollution affects human health at the very origins of individual life. Thus prior estimates of the effects of air pollution, and the policies that have been implemented on their basis, may significantly undervalue the costs.

Regarding the link between infant mortality and a given particulate shock I find no evidence to suggest there is any relationship between the two in this setting. Given that particulate pollution is the only criteria pollutant categorized solely on the basis of size, not chemistry, it could be the case that species difference in particulate origin is relevant for discussion and research. The particulate matter from the Mount St. Helens ashfall is relatively inert in composition; if, by convention or association we believe there is a causal connection between particulate shocks and infant death then an avenue of future research might look to investigate the relative toxicity of different types of particulate pollution.

References

Adams, Paul R. and Gerald R. Adams. 1984. "Mount St. Helens's ashfall: Evidence for a disaster stress reaction." *American Psychologist* 39 (3):252–260.

Almond, Douglas, Kenneth Y. Chay, and David S. Lee. 2005. "The costs of low birth weight." *The Quarterly Journal of Economics* 120 (3):1031–1083.

Almond, Douglas, Lena Edlund, and Marten 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, Douglas, Jr. Joseph J. Doyle, Amanda E. Kowalski, and Heidi Williams. 2010. "Estimating marginal returns to medical care: Evidence from at-risk newborns." *The Quarterly Journal of Economics* 125 (2):591–634.

Autor, David H. 2003. "Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing." *Journal of Labor Economics* 21 (1):1–42.

Barreca, Alan, Melanie Guldi, Jason M. Lindo, and Glen R. Waddell. 2010. "Running and jumping variables in RD designs: Evidence based on race, socioeconomic status, and birth weights." IZA Discussion Papers 5106, Institute for the Study of Labor (IZA).

Barreca, Alan I., Melanie Guldi, Jason M. Lindo, and Glen R. Waddell. 2011. "Saving babies? Revisiting the effect of very low birth weight classification." *The Quarterly Journal of Economics* 126 (4):2117–2123.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell. 2011. "Heaping-induced bias in regression-discontinuity designs." NBER Working Papers 17408, National Bureau of Economic Research, Inc.

Baxter, Peter J., Roy Ing, Henry Falk, Jean French, Gary F. Stein, Robert S. Bernstein, James A. Merchant, and Jack Allard. 1981. "Mount St. Helens eruptions, May 18 to June 12, 1980: An overview of the acute health impact." *Journal of the American Medical Association* 246 (22):2585–2589.

Baxter, Peter J., Roy Ing, Henry Falk, and Brian Pilkaytis. 1983. "Mount St. Helens eruptions: The acute respiratory effects of volcanic ash in a northern American community." *Archives of Environmental Health* 38:138–143.

Beck, Barbara D., Joseph D. Brian, and Dian E. Bohannon. 1981. "The pulmonary toxicity of an ash sample from the Mt. St. Helens volcano." *Experimental Lung Research* 2:289–301.

Bernstein, R. S, M. A McCawley, M. D. Attfield, F. H. Green, S. A. Olenchock, and D. D. Dollberg. 1981. "Evaluation of the potential pulmonary hazards of breathing - zone exposures to Mount St. Helens volcanic ash in loggers: The results of short-term toxicological, environmental and epidemiological studies." Tech. Rep. GHE-80-112, Department Health and Human Services, The National Institute for Occupational Safety and Health, Atlanta, GA.

Black, Sandra E., Aline Butikofer, Paul J. Devereux, and Kjell G. Salvanes. 2013. "This is only a test? Long-run impacts of prenatal exposure to radioactive fallout." NBER Working Papers 18987, National Bureau of Economic Research, Inc.

Buist, A. S., L.R. Johnson, W. M. Vollmer, G. J. Sexton, and P. H. Kanarek. 1983. "Acute effects of volcanic ash from Mount St. Helens on lung function in children." *The American Review of Respiratory Disease* 127 (6):714–719.

Buist, A. S., W. M. Vollmer, L. R. Johnson, R. S. Bernstein, and L. E. McCamant. 1986. "Four-year prospective study of the respiratory effects of volcanic ash from Mt. St. Helens." *American Review of Respiratory Disease* 4:526–534.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-based improvements for inference with clustered errors." *The Review of Economics and Statistics* 90 (3):414–427.

Cameron, A. Colin and Douglas L. Miller. 2010. "Robust inference with clustered data." Working Papers 107, University of California, Davis, Department of Economics.

Carter, Andrew V., Kevin T. Schnepel, and Douglas G. Steigerwald. 2013. "Asymptotic behavior of a t test robust to cluster heterogeneity." Mimeo September 6, 2013, University of California, Santa Barbara, Department of Economics.

Center for Disease Control. 1980a. "Morbidity and mortality weekly report." 29 (21).

———. 1980b. "Morbidity and mortality weekly report." 29 (22).

———. 1980c. "Mount St. Helens volcano health report." (12).

———. 1980d. "Mount St. Helens volcano health report." (19).

———. 1980e. "Mount St. Helens volcano health report." (3).

———. 1980f. "Mount St. Helens volcano health report." (10).

———. 1980g. "Mount St. Helens volcano health report." (9).

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

Chay, Kenneth Y. and Michael Greenstone. 2003. "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.

Chen, Chun and Bin Zhao. 2011. "Review of relationship between indoor and outdoor particles: I/O ratio, infiltration factor and penetration factor." *Atmospheric Environment* 45 (2):275–288.

Craighead, John E., Kenneth B. Adler, Gregory B. Butler, Robert J. Emerson, Brooke T. Mossman, and Craig D. Woodworth. 1983. "Health effects of Mount St. Helens volcanic dust." *Laboratory Investigation* 48 (1):5–12.

Currie, Janet, Matthew Neidell, and Johannes F. Schmieder. 2009. "Air pollution and infant health: Lessons from New Jersey." *Journal of Health Economics* 28 (3):688–703.

Currie, Janet and Maya Rossin-Slater. 2013. "Weathering the storm: Hurricanes and birth outcomes." *Journal of Health Economics* 32 (3):487–503.

Currie, Janet and Reed Walker. 2011. "Traffic congestion and infant health: Evidence from E-ZPass." *American Economic Journal: Applied Economics* .

Fruanfelder, Frederick T., Robert E. Kalina, A. Sonia Buist, Robert S. Bernstein, and David S. Johnson. 1983. "Ocular effects following the volcanic eruptions of Mount St. Helens." *Archives of Ophthalmology* 101 (3):376–378.

Fruchter, Jonathan S., David E. Robertson, John C. Evans, Khris B. Olsen, Elweood A. Lepel, Jagdish C. Laul, Keith H. Abal, Ronald W. Sanders, Peter O. Jackson, Ned S. Wogman, Richard W. Perkins, Harold H. Van Tuyl, Raymond H. Beauchamp, John W. Shade, J. Leland Daniel, Robert L. Erikson, George A. Sehmel, Richard N. Lee, Alfred V. Robinson, Owen R. Moss, James K. Briant, and William C. Cannon. 1980. "Mount St. Helens ash from the 18 May 1980 eruption: Chemical, physical, mineralogical, and biological properties." *Science* 206 (4461):1116–1125.

Granger, Clive W. J. 1969. "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica* 37 (3):424–38.

Green, Francis H., Linda Bowman, Vincent Castranova, Donald D. Dollberg, Julie A. Elliot, Jeffrey S. Fedan, Nicholas Hahon, Deloras J. Judy, Pervis C. Major, Michael S. Mentnech, Phillip R. Miles, Judith Mull, Stephen Olenchock, Tong-Man Ong, William M. Pailes, Harvey Resnick, Lloyd E. Stettler, James G. Tucker, Val Vallyathan, and When-Zong Whong. 1982. "Health implications of the Mount St. Helens eruption: Laboratory investigations." *The Annals of Occupational Hygiene* 26:921–933.

Green, Francis H., Val Vallythan, Michael S. Mentnech, James H. Tucker, James A. Merchant, Peter J. Kiessling, John A. Antonius, and Phillip Parshley. 1981. "Is volcanic ash a pneumoconiosis risk?" *Nature* 293:216–217.

Horwell, Claire J. and Peter J. Baxter. 2006. "The respiratory health hazards of volcanic ash: A review for volcanic risk mitigation." *Bulletin of Volcanology* 69 (1):1–24.

Hwang, Bing-Fang, Yungling L. Lee, and Jouni J. Jaakkola. 2011. "Air pollution and stillbirth: A population-based case-control study in Taiwan." *Environmental Health Perspectives* 119 (9):1345–1349.

Isen, Adam, Maya Rossin-Slater, and W. Reed Walker. 2014. "Every breath you take - every dollar you'll make: The long-term consequences of the Clean Air Act of 1970." NBER Working Papers 19858, National Bureau of Economic Research, Inc.

Jayachandran, Seema. 2009. "Air quality and early-life mortality: Evidence from Indonesia's wildfires." *Journal of Human Resources* 44 (4):916–954.

Johnson, Kit A., Don O. Loftsgaarden, and Rudy A. Gideon. 1982. "The effects of Mount St. Helens ash on the pulmonary function of 120 elementary school children." *American Review of Respiratory Disease* 126:1066–1069.

Kim, Ok-Jin, Eun-Hee Ha, Byung-Mi Kim, Ju-Hee Seo, Hye-Sook Park, Woo-Jae Jung, Bo-Eun, Lee Young-Ju Suh, Young-Ju Kim, Jong-Tae Lee, Ho Kim, and Yun-Chul HongMD. 2007. "PM10 and pregnancy outcomes: A hospital-based cohort study of pregnant women in Seoul." *Journal of Occupational and Environmental Medicine* 49 (12):1394–1402.

Luechinger, Simon. 2011. "Air pollution and infant mortality: A natural experiment from power plant desulfurization." Mimeo, London School of Economics.

Mansour, Hani and Daniel I. Rees. 2012. "Armed conflict and birth weight: Evidence from the al-Aqsa Intifada." *Journal of Development Economics* 99 (1):190–199.

McLemore, Theodore L., J. E. Mauldin, Milton V. Marshall, Robert Teague, G. Ford, David L. Busbee, Nelda P. Wray, A. Clark Griffin, and S. Donald Greenberg. 1984. "Biological effects of Mount St. Helens volcanic ash on cultured human alveolar macrophages." *International Journal of Toxicology* 3 (1):25–35.

Merchant, James, Peter Baxter, Robert Bernstein, Mike McCawley, Henry Falk, Gary Stein, Roy Ing, and Michael Attfield. 1982. "Health implications of the Mount St. Helens eruption: Epidemiological considerations." *The Annals of Occupational Hygiene* 26 (8):911–919.

Mohallem, Soraya Vecci, Debora Ja de Araujo Lobo, Celia Regina Pesquero, Joao Vicente Assuncao, Paulo Afonso de Andre, Paulo Hilario Nascimento Saldiva, and Marisa Dolhnikoff. 2005. "Decreased fertility in mice exposed to environmental air pollution in the city of Sao Paulo." *Environmental Research* 98 (2):196–202.

Moretti, Enrico and Matthew Neidell. 2011. "Pollution, health, and avoidance behavior: Evidence from the Ports of Los Angeles." *Journal of Human Resources* 46 (1):154–175.

Nania, Jim and Timothy Bruya. 1982. “In the wake of Mount St. Helens.” *Annals of Emergency Medicine* 11 (4):184–191.

Norton, Michael R. and Mickey E. Gunter. 1999. “Relationships between respiratory diseases and quartz-rich dust in Idaho, U.S.A.” *American Mineralogist* 84:1009–1019.

Parker, Jennifer D., Pauline Mendola, and Tracey J. Woodruff. 2008. “Preterm birth after the Utah Valley steel mill closure: A natural experiment.” *Epidemiology* 19 (6):820–823.

Pereira, Luiz A, Dana Loomis, Gleice M. Conceicao, Alfesio L. Braga, Rosangela M. Arcas, Humberto S. Kishi, Jtilio M. Singer, Gybrgy M. Bohm, , and Paulo H. Saldiva. 1998. “Association between air pollution and intrauterine mortality in Sao Paulo, Brazil.” *Environmental Health Perspectives* 106 (6):325–329.

Pope, C. Arden. 1989. “Respiratory disease associated with community air pollution and a steel mill, Utah Valley.” *American Journal of Public Health* 79 (5):623–628.

———. 2000. “Review: Epidemiological basis for particulate air pollution health standards.” *Aerosol Science and Technology* 32 (1):4–14.

Sanders, Nicholas J. 2011. “What doesn’t kill you makes you weaker: Prenatal pollution exposure and educational outcomes.” Discussion Papers 10-019, Stanford Instititute for Economic Policy Research.

Sanders, Nicholas J. and Charles F. Stoecker. 2011. “Where have all the young men gone? Using gender ratios to measure fetal death rates.” NBER Working Papers 17434, National Bureau of Economic Research, Inc.

Sarna-Wojcicki, Andrei M., Susan Shipley, Richard B. Waitt Jr, Daniel Dzurisin, and Spencer H. Wood. 1981. “Aerial distribution, thickness, mass, volume, and grain size of air-fall ash from the six major eruptions of 1980.” Geological Survey Professional Paper 1250, pp 577-600.

Shore, James H., Ellie L. Tatum, and William M. Vollmer. 1986. “Evaluation of mental effects of disaster - Mount St. Helens Eruption.” *American Journal of Public Health* 76:76–83.

Tanaka, Shinsuke. 2010. “Air pollution, infant mortality, and the environmental regulations in China.” Mimeo, Department of Economics, Boston University.

Trivers, Robert L. and Dan E. Willard. 1973. “Natural selection of parental ability to vary the sex ratio of offspring.” *Science* 179 (4068):90–92.

U.S. EPA. 2004. “Air quality criteria for particulate matter (Final Report).” Tech. Rep. EPA 600/P-99/002, U.S. Environmental Protection Agency, Washington, DC.

———. 2006a. “Air quality criteria for lead (Final Report).” Tech. Rep. EPA/600/R-

05/144aF-bF, U.S. Environmental Protection Agency, Washington, DC.

———. 2006b. "Air quality criteria for ozone and related photochemical oxidants (Final Report)." Tech. Rep. EPA/600/R-05/004aF-cF, U.S. Environmental Protection Agency, Washington, DC.

———. 2008a. "Integrated science assessment for oxides of nitrogen - health criteria (Final Report)." Tech. Rep. EPA/600/R-08/071, U.S. Environmental Protection Agency, Washington, DC.

———. 2008b. "Integrated science assessment for sulfur oxides - health criteria (Final Report)." Tech. Rep. EPA/600/R-08/047F, U.S. Environmental Protection Agency, Washington, DC.

———. 2009. "Integrated science assessment for particulate matter (Final Report)." Tech. Rep. EPA/600/R-08/139F, U.S. Environmental Protection Agency, Washington, DC.

———. 2010. "Integrated science assessment for carbon monoxide (Final Report)." Tech. Rep. EPA/600/R-09/019F, U.S. Environmental Protection Agency, Washington, DC.

———. 2011. "Policy assessment for the review of the particulate matter National Ambient Air Quality Standards (Final Report)." Tech. Rep. EPA 452/R-11-003, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Vallyathan, Val, M. Sharon Mentnech, Lloyd E. Stettler, Donald D. Dollberg, and Francis H. Green. 1983a. "Mount St. Helens' volcanic ash: Hemolytic activity." *Environmental Research* 30 (2):349–360.

Vallyathan, Val, M. Sharon Mentnech, James H. Tucker, and Francis H. Green. 1983b. "Pulmonary response to Mount St. Helens' volcanic ash." *Environmental Research* 30 (2):361–371.

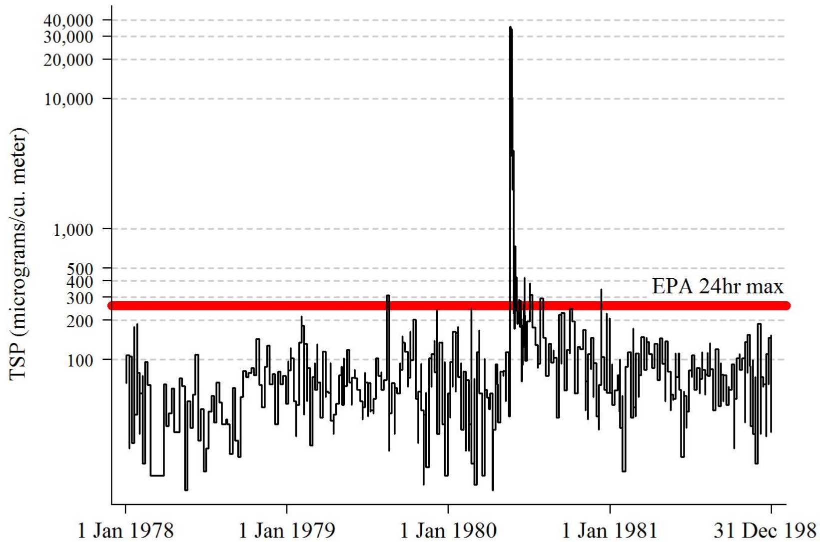
Warrick, Richard A., Jeremy Anderson, Thomas Downing, James Lyons, John Ressler, Mary Warrick, and Theodore Warrick. 1981. *Four communities under ash after Mount St. Helens*. Program on Technology, Environment, and Man Monograph 34. Institiute of Behavioral Science, University of Colorado.

Wilson, James G. and F. Clarke Fraser, editors. 1977. *Handbook of Teratology: General Principles and Etiology*, vol. 1. Plenum Publishing Company Limited.

Wolfers, Justin. 2006. "Did unilateral divorce laws raise divorce rates? A reconciliation and new results." *American Economic Review* 96 (5):1802–1820.

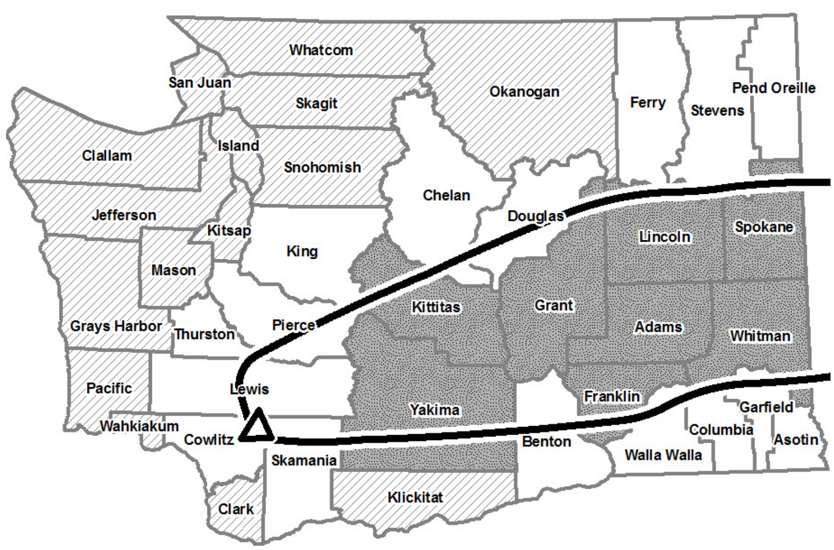
Figures and Tables

Figure 1.1: Periodic TSP monitor readings for Yakima county, Washington, in log scale with a reference line showing the, then current, EPA 24hr maximum standard.



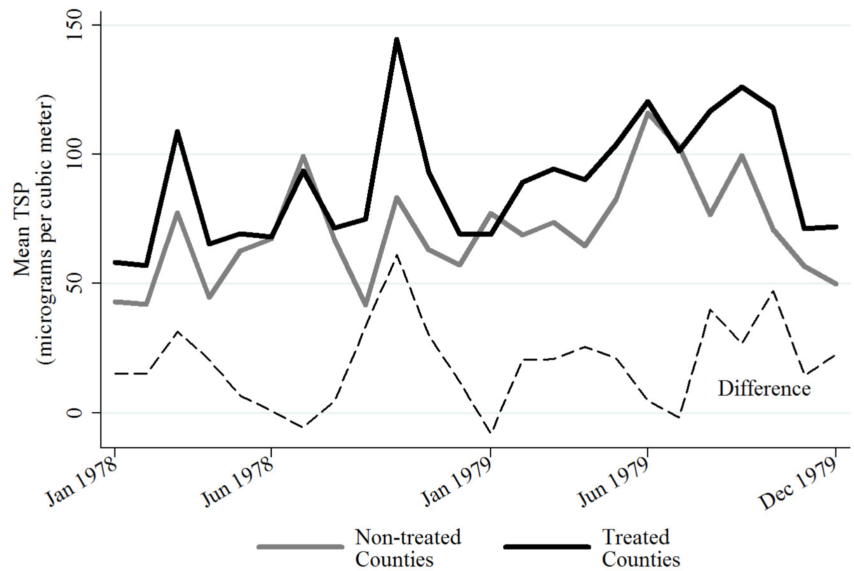
Sources: US EPA Air Quality System data and (Center for Disease Control, 1980g)

Figure 1.2: The extent of the May 18th Mt St Helens eruption ashfall over Washington.



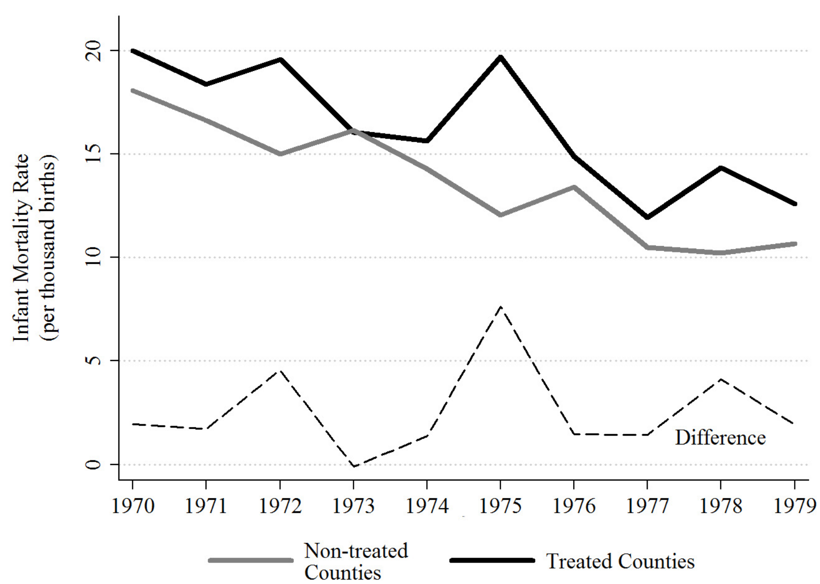
Note: Treated counties are shaded. Non-treated counties are hash marked. Marginal counties not used in the analysis are left unmarked.
Source: (Sarna-Wojcicki et al., 1981).

Figure 1.3: Mean TSP levels by month for treated and non-treated counties



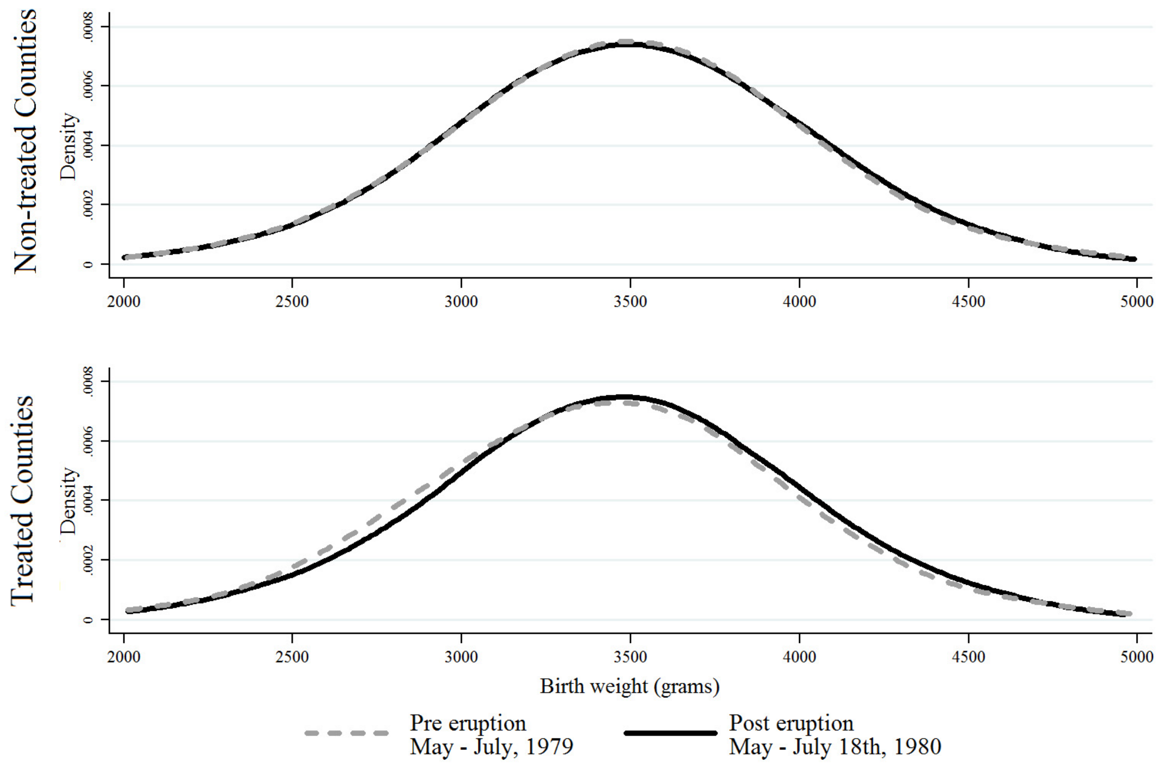
Source: Author's tabulation of EPA AQS data

Figure 1.4: Infant mortality rates by year for treated and non-treated counties



Source: Author's tabulation of NCHS Multiple Cause Mortality data and birth record data

Figure 1.5: Birth weight kernel densities for treated and non-treated counties



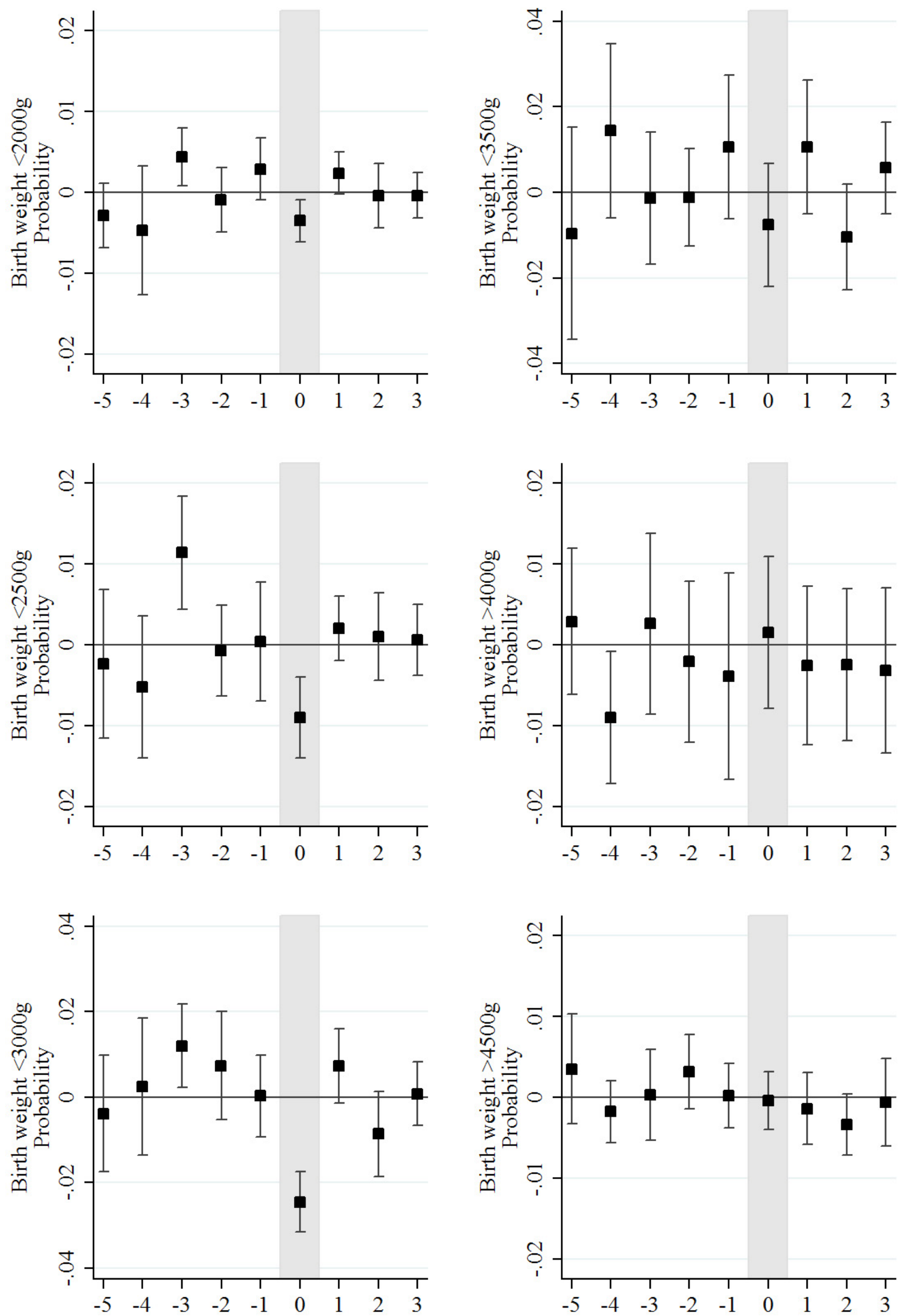
Source: Author's tabulation of NCHS birth record data

Table 1.1: Cofactor summary statistics for treated and non-treated counties, 1978-79

			Non-treated	Treated
	Min	Max	Mean	Mean
Early prenatal care	0	1	0.7936	0.7738
Gestation length (weeks)	17	52	39.93	39.72
Hospital birth	0	1	0.9722	0.9885
Male	0	1	0.5146	0.5130
Married	0	1	0.8882	0.8704
Mother's age (years)	13	49	25.047	24.767
Mother immigrant	0	1	0.0524	0.0837
Multiple birth	0	1	0.0174	0.0203
No prenatal care	0	1	0.0051	0.0061
Premature birth	0	1	0.0689	0.0776
Teenage mother	0	1	0.0415	0.0521
White	0	1	0.9300	0.9303
			n=27,815	n=18,014

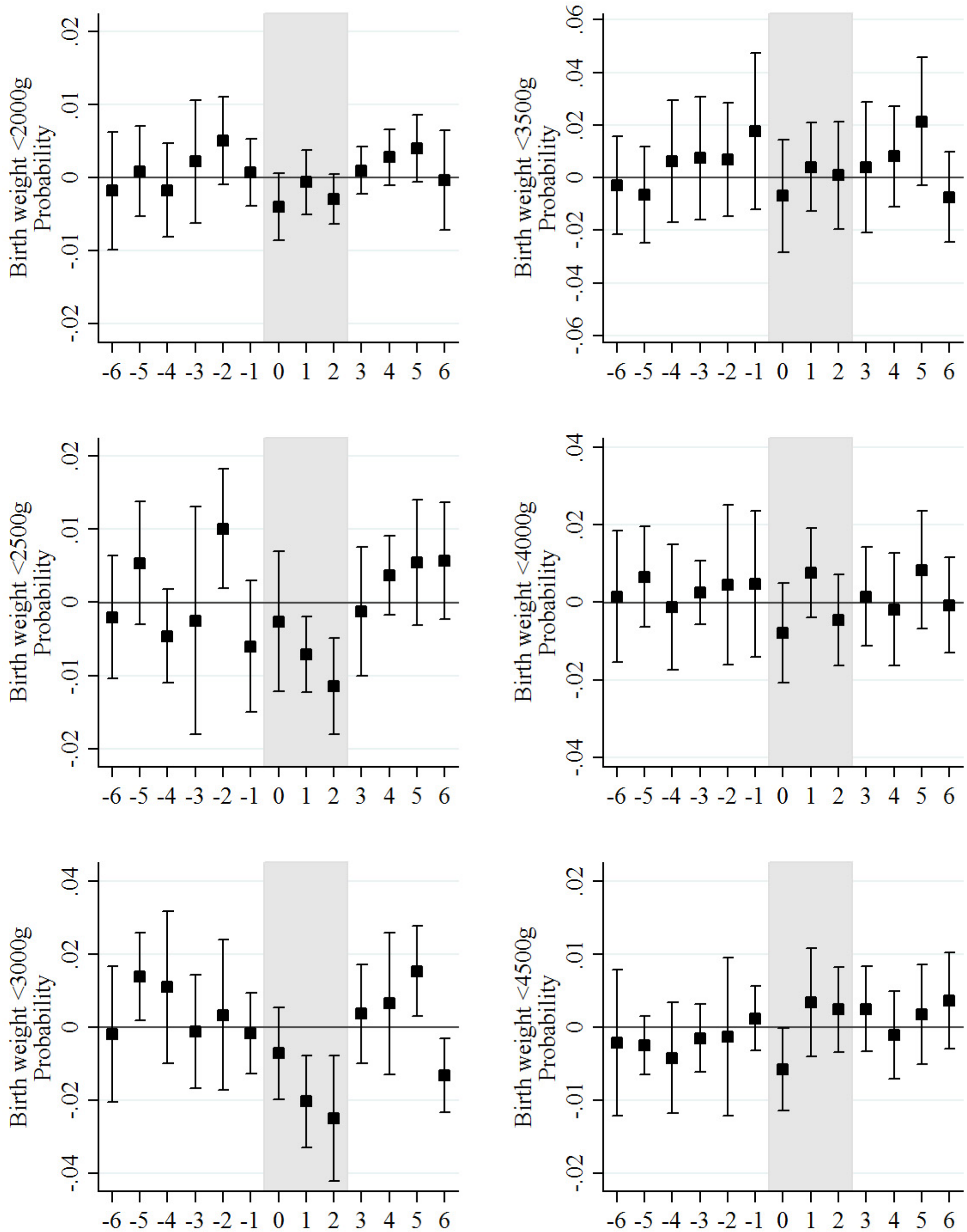
Source: NCHS Birth Data Files.

Figure 1.6: Effect of in utero TSP shock exposure on birth weight distribution by weight category with 9-month leads and lags



Note: Grey bands indicate eruption timing at t=0. Line caps indicate 95% confidence range (by cluster robust standard errors). The results are based on data ranging from 1/1/1973 through 12/31/1988.

Figure 1.7: Effect of in utero TSP shock exposure on birth weight distribution by weight category with 3-month leads and lags



Note: Grey bands indicate eruption timing effect window on 3rd, 2nd, and 1st trimester at t={0,1,2} respectively. Line caps indicate 95% confidence range (using cluster robust standard errors). The results are based on data ranging from 1/1/1973 through 12/31/1988.

Table 1.2: Comparison of infant mortality rates (deaths/births)

	Non-treated	Treated	Difference deaths per 1000 births
May-June-July 1979	51/4715	31/3101	-0.820
May-June-July 1980	68/4977	44/3213	0.032
	2.846	3.698	0.851

Source: NCHS Birth Data Files & Mortality Multiple Cause Files.

Table 1.3: Comparison of mean infant birth weights (grams)

	Non-treated	Treated	Difference
May-June-July 1979	3461	3399	-62
May-June-July 1980	3452	3421	-31
	-9	22	31

Source: NCHS Birth Data Files.

Table 1.4: Difference in differences estimates for TSP shock exposure on infant birth weight (natural log birth weight in grams)

	(1)	(2)	(3)	(4)
Ash county	-0.0189***	-0.0099***	-0.0099***	-
	(0.0021)	(0.0021)	(0.0021)	
In utero	0.0042**	0.0047***	-0.0021	-0.0021
	(0.0016)	(0.0014)	(0.0039)	(0.0038)
Ash × In utero	0.0092***	0.0041	0.004	0.004
	(0.0021)	(0.0031)	(0.0032)	(0.0031)
Group controls	no	yes	yes	yes
Month & Year FE	no	no	yes	yes
County FE	no	no	no	yes
R-squared	0.002	0.287	0.289	0.289
N	215,386	170,647	170,647	170,647

Note: cluster robust standard errors in parentheses (county).
* p<0.1, ** p<0.05, *** p<0.01.

Table 1.5: Difference in differences estimates for TSP shock exposure on infant birth weight during trimester (natural log birth weight in grams)

	(5)	(6)	(7)	(8)
Ash County	-0.0189***	-0.0099***	-0.0099***	-
	(0.0021)	(0.0021)	(0.0021)	
In utero 1st tri	0.002	0.0015	-0.0125***	-0.0125***
	(0.0033)	(0.0018)	(0.0039)	(0.0039)
In utero 2nd tri	0.0100***	0.0090***	0.0026	0.0027
	(0.0024)	(0.0022)	(0.0047)	(0.0046)
In utero 3rd tri	0.0005	0.0037	-0.0012	-0.0011
	(0.0033)	(0.0023)	(0.0045)	(0.0044)
Ash × In utero 1st	0.0126***	0.0092*	0.0089*	0.0087*
	(0.0042)	(0.0047)	(0.0047)	(0.0046)
Ash × In utero 2nd	0.0048	-0.0045	-0.0046	-0.0048
	(0.0029)	(0.0032)	(0.0033)	(0.0033)
Ash × In utero 3rd	0.0104**	0.0068	0.0067	0.0070*
	(0.0041)	(0.0041)	(0.0041)	(0.0040)
Group Controls	no	yes	yes	yes
Month & Year FE	no	no	yes	yes
County FE	no	no	no	yes
R-squared	0.002	0.287	0.289	0.289
N	215,386	170,647	170,647	170,647

Note: cluster robust standard errors in parentheses (county).
* p<0.1, ** p<0.05, *** p<0.01.

Table 1.6: Linear probability model estimates for TSP shock exposure on infant birth weight

	<2000g	<2500g	<3000g	<3500g	<4000g	<4500g
Ash county	-	-	-	-	-	-
In utero	0.003	0.0061**	0.006	0.0007	0.0041	-0.0006
	(0.0016)	(0.0028)	(0.0046)	(0.0078)	(0.0060)	(0.0023)
Ash × In utero	-0.0035**	-0.0090***	-0.0245***	-0.0076	-0.0015	0.0004
	(0.0013)	(0.0025)	(0.0035)	(0.0072)	(0.0047)	(0.0018)
Group controls	yes	yes	yes	yes	yes	yes
Month & year FE	yes	yes	yes	yes	yes	yes
County FE	yes	yes	yes	yes	yes	yes
Depd var mean	1.9%	5.5%	19.5%	55.3%	87.0%	97.5%
N	213,085	213,085	213,085	213,085	213,085	213,085

Note: cluster robust standard errors in parentheses (county).
* p<0.1, ** p<0.05, *** p<0.01.

Table 1.7: Linear probability model estimates for TSP shock exposure during trimester on infant birth weight

	<2000g	<2500g	<3000g	<3500g	<4000g	<4500g
Ash county	-	-	-	-	-	-
In utero 1st tri	0.0029	0.0093*	0.0252**	0.0161	0.0155*	-0.0037
	(0.0027)	(0.0051)	(0.0099)	(0.0099)	(0.0078)	(0.0028)
In utero 2nd tri	-0.0003	0.0033	0.0019	-0.0022	0.001	-0.0013
	(0.0019)	(0.0033)	(0.0070)	(0.0108)	(0.0068)	(0.0027)
In utero 3rd tri	0.0048**	0.0065**	0.0003	-0.0027	0.0018	0.0019
	(0.0019)	(0.0029)	(0.0049)	(0.0095)	(0.0072)	(0.0026)
Ash × In utero 1st	-0.0038**	-0.0135***	-0.0321***	-0.0054	-0.0038	0.0029
	(0.0017)	(0.0037)	(0.0078)	(0.0098)	(0.0064)	(0.0026)
Ash × In utero 2nd	-0.0016	-0.0093***	-0.0276***	-0.0027	0.0077	0.0039
	(0.0022)	(0.0028)	(0.0077)	(0.0094)	(0.0067)	(0.0034)
Ash × In utero 3rd	-0.0050*	-0.0047	-0.0146**	-0.0137	-0.0078	-0.0053*
	(0.0025)	(0.0051)	(0.0059)	(0.0110)	(0.0067)	(0.0031)
Group controls	yes	yes	yes	yes	yes	yes
Month & year FE	yes	yes	yes	yes	yes	yes
County FE	yes	yes	yes	yes	yes	yes
Depd var mean	1.9%	5.5%	19.5%	55.3%	87.0%	97.5%
R-squared	0.038	0.068	0.056	0.040	0.022	0.008
N	213,085	213,085	213,085	213,085	213,085	213,085

Note: cluster robust standard errors in parentheses (county).
* p<0.1, ** p<0.05, *** p<0.01.

Table 1.8: Difference in differences estimates for TSP shock exposure on infant mortality (per 1000 births)

	Infant Mortality Rate				Age <1mo. Mortality Rate			
	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Ash county	1.856*	1.423**	1.3938**	-	1.203*	0.954	0.840	-
	(0.999)	(0.638)	(0.604)		(0.645)	(0.602)	(0.539)	
Eruption window (12mo.)	1.256	0.813	2.220*	2.369*	-0.533	-0.638	1.391	1.446
	(1.156)	(1.067)	(1.223)	(1.206)	(0.798)	(0.768)	(1.047)	(1.053)
Ash × window	-1.802	-1.063	-1.113	-1.302	-1.536	-0.802	-0.904	-0.904
	(2.046)	(2.123)	(2.153)	(2.147)	(0.989)	(1.130)	(1.122)	(1.142)
Group Controls	no	yes	yes	yes	no	yes	yes	yes
Month & Year FE	no	no	yes	yes	no	no	yes	yes
County FE	no	no	no	yes	no	no	no	yes
R-squared	0.01	0.03	0.06	0.07	0.01	0.04	0.07	0.08
N	2,311	2,304	2,304	2,304	2,311	2,304	2,304	2,304

Note: regression weighted by #births, cluster robust std errors (county).
* p<0.1, ** p<0.05, *** p<0.01.

Table 1.9: Difference in differences estimates for TSP shock exposure on birth month cohort infant mortality (per 1000 births)

	(19)	(20)	(21)	(22)
Ash county	1.994*	1.411**	1.378**	-
	(1.027)	(0.6697)	(0.6357)	
In utero 1st tri	-1.313	-2.200	-6.659	-6.580
	(2.129)	(2.124)	(5.096)	(4.889)
In utero 2nd tri	-0.1227	-0.1978	-6.513**	-6.499**
	(1.136)	(1.200)	(2.410)	(2.396)
In utero 3rd tri	5.572**	4.792**	-0.8926	-0.8595
	(2.023)	(1.892)	(2.933)	(2.917)
Ex utero 0-3mo.	1.883	1.178	-5.232**	-5.229**
	(1.122)	(0.891)	(2.257)	(2.238)
Ex utero 3-6mo.	0.2132	-0.0785	-4.125*	-4.143*
	(1.205)	(1.301)	(2.046)	(2.064)
Ash × In utero 1st	-2.974	-2.274	-2.323	-2.601
	(3.383)	(3.545)	(3.516)	(3.523)
Ash × In utero 2nd	-0.4914	0.0027	-0.1141	-0.4087
	(2.118)	(2.185)	(2.169)	(2.174)
Ash × In utero 3rd	-4.526	-3.864	-3.943	-4.128
	(2.820)	(2.948)	(2.927)	(2.936)
Ash × Ex utero 0-3mo.	-3.899**	-2.749	-2.936*	-3.353*
	(1.849)	(1.601)	(1.591)	(1.642)
Ash × Ex utero 3-6mo.	2.551	2.506	2.476	2.375
	(1.783)	(1.790)	(1.781)	(1.797)
Group controls	no	yes	yes	yes
Month & Year FE	no	no	yes	yes
County FE	no	no	no	yes
R-squared	0.02	0.05	0.08	0.09
N	2,242	2,235	2,235	2,235

Note: regression weighted by #births, cluster robust std errors (county).
* p<0.1, ** p<0.05, *** p<0.01.

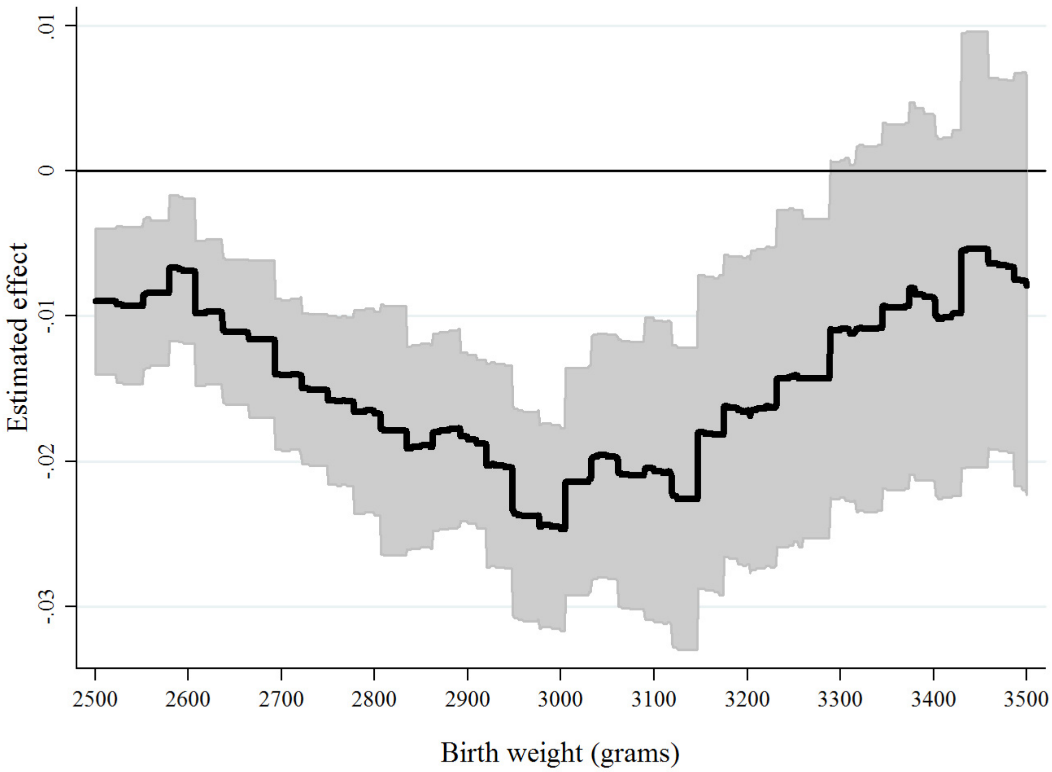
Appendix

Omitted Variables

Heteroskedastic robust standard errors calculated for the Poisson regression model were found to be significantly smaller than the cluster-robust standard errors presented in Table A1.4 and used elsewhere in my paper. This may be caused by negative correlation between the residuals and the explanatory variables – negative correlations exist between the residuals and immigrant status, gestation length (used in the regressions), as well as father's and mother's education (not used in the regressions due to the extremely large number of missing observations). While the exact cause of the difference in standard errors is not known, I speculate that it may be the result of omitted variables, such as household income, welfare assistance, and smoking, that are not available. The omission of these variables cannot be directly overcome and is a limitation. However, the residuals are not correlated with the indicators for those counties affected by the ashfall, the timing indicator for the 9 months after the eruption, or the product of the two (the difference-in-differences indicator).

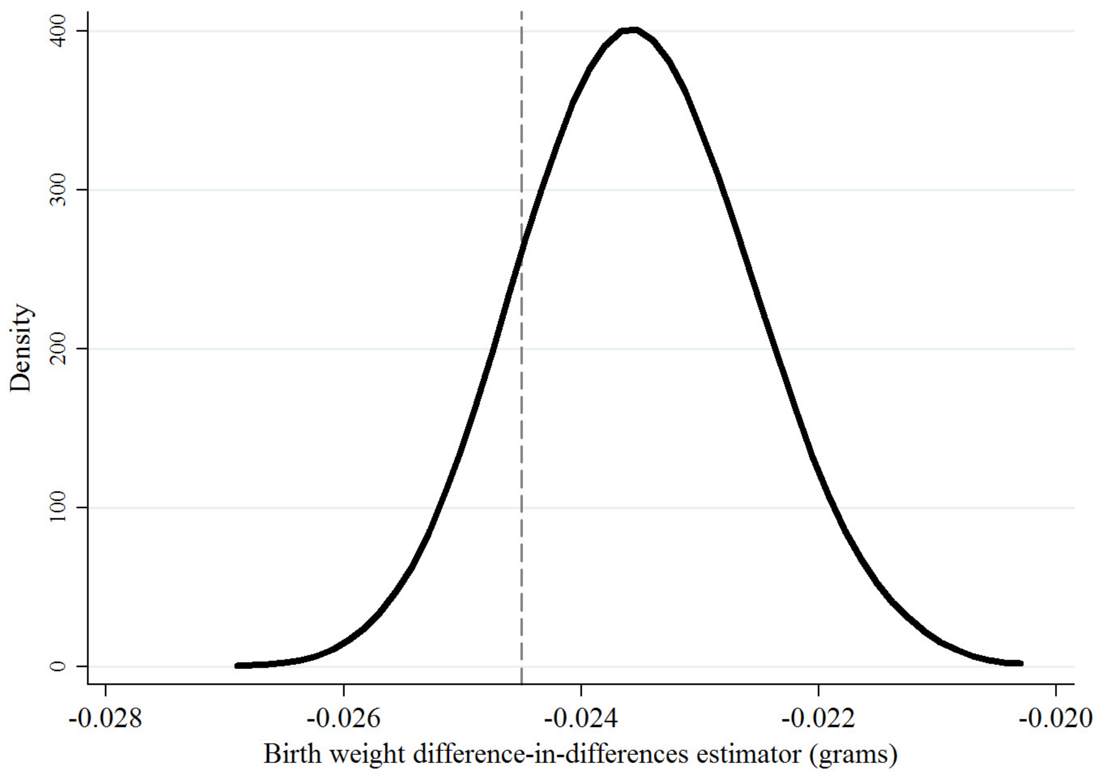
Figures and Tables

Figure A1.1: Linear probability model estimates for TSP shock exposure on infant birth weight



Note: The grey area covers the 95% confidence region (using cluster robust standard errors).

Figure A1.2: Quasi-Monte-Carlo difference-in-differences estimator distribution for the $<$ 3000g LPM [Model (9)]



Note: Kernel density plot of 2000 iterations; the dash reference line is located at the Model (9) regression point estimate.

Table A1.1: Summary of the criteria pollutants regulated by the EPA and their effects

Pollutant	Respiratory Pathway	Primary Effects	Source
Carbon monoxide	Readily binds to hemoglobin	Metabolic degradation, oxygen impairment	U.S. EPA (2010)
Lead	Absorption into the blood stream	Anemia, cellular membrane dysfunction, nerve dysfunction, neurochemical alteration	U.S. EPA (2006a)
Nitrogen oxide	Not fully determined (ozone precursor)	Internal epithelial cell damage, increased susceptibility to viral infection	U.S. EPA (2008a)
Ozone	Interaction with compounds in the epithelial lining fluid	Neural reflexes, air-way inflammation, lung disease	U.S. EPA (2006b)
Particulate matter	Lung irritant	Stress response, cardiac arrhythmia, endothelial dis-function	U.S. EPA (2004)
Sulfur oxide	Dissolves in the blood stream into constituent ions	Bronchial restriction and inflammation, cancerous cell mutation	U.S. EPA (2008b)

Table A1.2: Linear probability model estimates for TSP shock exposure on infant birth weight with county-specific time trends

	<2000g	<2500g	<3000g	<3500g	<4000g	<4500g
Ash county	-	-	-	-	-	-
In utero	0.0018	0.005	0.0014	-0.003	0.0046	-0.0008
	(0.002)	(0.003)	(0.004)	(0.007)	(0.006)	(0.003)
Ash × In utero	-0.0018	-0.0062*	-0.0126***	0.0022	-0.0029	0.0007
	(0.001)	(0.003)	(0.004)	(0.007)	(0.006)	(0.002)
Group controls	yes	yes	yes	yes	yes	yes
Month & year FE	yes	yes	yes	yes	yes	yes
County FE	yes	yes	yes	yes	yes	yes
Depd var mean	1.9%	5.5%	19.5%	55.3%	87.0%	97.5%
N	213,085	213,085	213,085	213,085	213,085	213,085

Note: cluster robust standard errors in parentheses (county).
* p<0.1, ** p<0.05, *** p<0.01.

Table A1.3: Linear probability model estimates for TSP shock exposure during trimester on infant birth weight with county-specific time trends

	<2000g	<2500g	<3000g	<3500g	<4000g	<4500g
Ash county	-	-	-	-	-	-
In utero 1st tri	0.0022	0.0082	0.0204*	0.0125	0.0164*	-0.0039
	(0.0029)	(0.0053)	(0.0106)	(0.0103)	(0.0080)	(0.0032)
In utero 2nd tri	-0.001	0.0022	-0.0025	-0.0059	0.0014	-0.0015
	(0.0022)	(0.0038)	(0.0062)	(0.0102)	(0.0064)	(0.0030)
In utero 3rd tri	0.0042**	0.0055*	-0.0042	-0.0063	0.0021	0.0017
	(0.0020)	(0.0029)	(0.0043)	(0.0089)	(0.0072)	(0.0027)
Ash × In utero 1st	-0.002	-0.0107***	-0.0195**	0.0049	-0.0055	0.0034
	(0.0013)	(0.0037)	(0.0080)	(0.0085)	(0.0073)	(0.0028)
Ash × In utero 2nd	0.0002	-0.0065	-0.0159*	0.0071	0.0063	0.0044
	(0.0024)	(0.0042)	(0.0091)	(0.0101)	(0.0083)	(0.0037)
Ash × In utero 3rd	-0.0033	-0.0021	-0.0033	-0.0042	-0.009	-0.0049
	(0.0025)	(0.0053)	(0.0055)	(0.0104)	(0.0078)	(0.0033)
Group controls	yes	yes	yes	yes	yes	yes
Month & year FE	yes	yes	yes	yes	yes	yes
County FE	yes	yes	yes	yes	yes	yes
Depd var mean	1.9%	5.5%	19.5%	55.3%	87.0%	97.5%
R-squared	0.038	0.068	0.056	0.040	0.022	0.008
N	213,085	213,085	213,085	213,085	213,085	213,085

Note: cluster robust standard errors in parentheses (county).
* p<0.1, ** p<0.05, *** p<0.01.

Table A1.4: Poisson regression difference in differences estimates for TSP shock exposure on infant mortality (per 1000 births)

	Infant Mortality Rate	Age <1mo. Mortality Rate
Ash county	-	-
Eruption window (12mo.)	0.2117**	0.1922
	(0.1064)	(0.1399)
Ash county × window	-0.1255	-0.1312
	(0.1940)	(0.1651)
Group Controls	yes	yes
Month & Year FE	yes	yes
County FE	yes	yes
N	2,304	2,304

Note: regression weighted by #births, cluster robust std errors (county).
* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Chapter 2

Free Energy and Social Norms: Electricity Conservation Through Peer Comparisons.

Summary

With my colleague Daniel Moncayo I investigate the informational effect that social comparisons play in household electricity use. We employ a randomized field experiment to present individuals with a social norm that encourages them to conserve electricity by comparing their consumption to that of an "energy efficient neighbor," consisting of the average of the 10th percentile of participants' electricity consumption. Utilizing smart-meter data, we find households that received our informational treatment reduced their electricity consumption by an average of 7%, even though the treatment was private and there was no financial incentive to conserve. We also discover that conservation gains are largest during the peak morning and evening hours.

2.1 Introduction

In 2009, United States households used approximately 10.2×10^{15} Btu's of electricity, worth an estimated value of $230 billion (US EIA, 2009). This level of consumption contributed to over 2 billion metric tons of carbon dioxide emitted into the atmosphere,[1] a non-trivial fraction of the estimated 32 billion metric tons of global carbon dioxide emissions that same year. Due to global concerns about the effect of these emissions on the environment, many supply-side interventions have been advanced as the leading solutions to reduce emissions. However, the effective implementation of these policies has been mixed (UNEP, 2012).

An alternative avenue of research suggests that social norms[2] can be successful in mitigating hard to manage externality problems, such as air pollution (McGinnis and Ostrom, 2008; Ostrom, 2014). Yet the influence of social norms is not well understood and is not always universal.[3] This is the case in the field of energy conservation, where results are not entirely definitive. Many confounding effects prevent researchers from reaching a consensus as to the direction and magnitude of the effect of social norms on households' electricity demand. Even so, their effectiveness to elicit energy conservation responses is generally encouraging (Allcott and Mullainathan, 2010).

We use a randomized control trial to estimate the informational effect of a particular type of social norm, a social comparison,[4] as a potential driver of household electricity conservation. The isolation of this effect is possible due to the near ideal setting where the field experiment takes place: participants live in near identical housing, have no pecuniary motives for conservation, are consumption naive, and can be mutually identified

[1] Calculated using the US EPA's eGRID data, http://cfpub.epa.gov/egridweb/ .

[2] Social norms are inter-relational constructs that define the boundaries and mores of acceptable behavior within a group.

[3] In some studies social norms are found to have little or no effect, see Viscusi, Huber, and Bell (2011) in the case of recycling, Beshears et al. (2011) for retirement savings, and National Science and Technology Council (2015) for pharmacist over prescription of controlled substances.

[4] A social comparison is a type of social norm message that, in this context, reveals one's level of consumption in comparison to others (be it a group or individual).

as a peer group due to their relative social homogeneity. We also analyze the strategic placement of the social comparison within the electricity consumption distribution of the participants. Using an agent-based model, we posit that a social comparison placed to the left of mean should elicit larger behavioral responses from treated participants. Simultaneously, this placement should also ameliorate a potential counter-productive norm reversion response.[5]

We test these hypotheses using a social comparison we called the "energy efficient neighbor,"[6] which is constructed using the average of the 10th percentile of electricity consumption. This comparison was presented weekly to treated households through graphs that compare their hourly consumption patterns to that of the "energy efficient neighbor" for 10 weeks. As a result we find that treated households reduced their electricity consumption by around 7% – as compared to an average reduction of 2% found in similar studies that use the mean as the social comparison. Conditioning on prior household energy consumption reveals that on average all treated participants differentially reduced their energy consumption. Even so, qualitative evidence, found by analyzing participants' survey responses, indicates that some individuals' behavior is not strictly consistent with the norm-reversion hypothesis. While treated participants' attitudes generally moved *toward* energy conservation, there are some portion of high-energy use participants that turned *away* from conserving electricity. This dissociative effect suggests there is an optimal placement for the social comparison, one that takes into account that some individuals are dissuaded from energy conservation when the social comparison is too far away.

As part of our study we find evidence that clarifies when households choose to conserve energy. The use of high frequency data allows us to estimate the timing and progress of

[5]Norm reversion occurs when individuals on either side of the distribution converge to the revealed social norm.

[6]This normative appellation was given for descriptive purposes only and should not convey to the reader any notion of economic efficiency per se.

changes in electricity consumption throughout the 10-week treatment period by hourly intervals. Our estimates reveal that treated households choose to conserve the most electricity during the waking morning and evening hours when consumption is typically the highest. We also find evidence to suggest that households learn and improve upon their conservation efforts over time, as indicated by an increase in treatment effectiveness as our intervention progresses.

The findings in this paper contribute to the literature on social norms by establishing a well-defined estimate for the informational effect of a social comparison. Our research shows that in a near ideal environment devoid of many competing factors, the social comparison approach can yield significant reductions in energy use. More importantly, these reductions can be achieved in a way that is easily scalable and replicable using electronic communications. On the theoretical side, our social comparison model clarifies the potential mechanism though which social comparisons affect individual behavior, and provides testable recommendations of ways to improve upon our experimental design. In all, the results presented in this paper and the potential of further improvements, strengthen the case that social comparisons may be used as a policy tool alongside pecuniary and supply-side interventions to address carbon emission levels.

2.2 Social norms and comparisons

2.2.1 Social norms research

There has been strong interest in social norm messaging in the context of energy policy to reduce household energy consumption. During the energy shortages of the 1970's there was an initial surge of this type of research,[7] and beginning in the 2000's there has been a resurgence with the popularization of climate change as a global concern.[8]

The overarching message of these studies is that social norms influence individual and household energy use, and that energy conservation can be encouraged by providing consumption feedback messaging. Conglomerating these studies Delmas, Fischlein, and Asensio (2013) perform a meta-analysis (156 individual studies) and find the average treatment effect (ATE) of some type of social message and informational consumption feedback on energy use is -7.4%, with a range that extends from -55% to +18.5%. Those characterized by Delmas, Fischlein, and Asensio as lower quality (primarily the earlier studies) have an ATE of almost -10% in energy use reduction, while those they designate of high-quality have an ATE around -2%.

The number of research studies conducted on social norms and energy use has not led to a clear consensus on the magnitude or direction of the effect that social norms have on electricity demand. This is not a surprise given the diversity of experimental designs. For instance, Allcott (2011) presents 600,000 households with a modified electricity bill that includes a comparison of their own historical consumption level to other households

[7] See Abrahamse et al. (2005) for a review of this research. For notable studies see Pallak and Cummings (1976); Seaver and Patterson (1976); Hayes and Cone (1977); Seligman and Darley (1977); Becker (1978); Bittle, Valesano, and Thaler (1979); Winett, Neale, and Grier (1979); Midden et al. (1983).

[8] There are a few studies scattered during the late 1980's and 1990's such as Sexton, Johnson, and Konakayama (1987); van Houwelingen and van Raaij (1989); Wilhite and Ling (1995); these do not cite specific policy motivations.

(a descriptive norm[9]) with smiley faces (an injunctive norm[10]), and energy reduction suggestions. This intervention reduced treated household energy consumption by 2% on average. On a much smaller scale Delmas and Lessem (2014) provided 66 university dorm room residents with public and private electricity use information. They use green and red dot stickers as the public injunctive norm, and energy-use levels as the private descriptive norm. The authors find that while private information has no statistical effect on energy consumption, the public treatment reduces dorm electricity use by 20%. In a similar public display study, Petersen et al. (2007) find that utilizing publicly visible energy-use indicators (injunctive norm green to red glowing 'orbs') in a college dorm competition reduced consumption by 32%. Conversely with no social message, Carrico and Riemer (2011) find that office employees collectively reduced energy consumption by 4% and 7% respectively when peer educators and monthly emails were used to disseminate energy use feedback with energy reduction goals.

The diversity of social messaging and experimental settings used in each study makes it difficult to compare their effectiveness. Differences between the types of social norms and messaging tend to obscure the mechanisms at play in behavioral responses. For example, consider a treatment that includes both descriptive and injunctive norm messages delivered to households that pay for electricity – it is natural to question the degree to which households might reduce consumption because they are influenced by the two types of social norms embedded in the informational treatment, or the desire to save money relative to their previous consumption levels or that of their peers, or because of reminders and suggestions about how to save electricity. These difficulties are exacerbated when social comparisons are utilized since households may or may not consider the information about those to whom they are compared as valid.

[9]Descriptive social norms are formed on the basis of perceiving the actions or outcomes of others. Individuals act to conform with the perceived norm.

[10]Injunctive social norms generally construe to individuals what is, or is not, acceptable behavior through social cues. Individuals act to receive or avoid (respectively) those social cues.

In this paper we disentangle many of these confounding factors by designing a field experiment that allows for a clean identification of the informational effect of a social comparison on electricity demand. Preceding the presentation of our experimental design and results we clarify the factors, both social and pecuniary, that determine individual household demand for electricity. Modeling preferences and social norms allows us to illustrate how our field experiment isolates a social norm as the underlying motive for the change in participant behavior. This also informs and motivates the selection of our social comparison. Finally, it outlines a basis from which to compare our results to the literature and suggests direction for future research.

2.2.2 Preferences over consumption and social norms

Our theoretical framework to study the effect of social norms on household energy consumption builds on the work of Levitt and List (2007) and Ferraro and Price (2013). It contains two main parts: utility from consumption and utility from social comparisons. The first component models utility from the consumption of electricity services and a composite private good. The second captures social factors associated with the consumption of electricity. These two types of utility allow us to study the mechanisms at play when we subject individuals to informational feedback interventions and their possible reactions to different kinds of social comparisons.

Individuals, indexed by $j \in 1, 2, ..., M_i$, are residents of households referenced by the subscript $i \in 1, 2, ..., N$. The set of N households represents a group of mutual peers. This relationship could be geographical, like a community, a college, a neighborhood, or an apartment complex. It could be determined by demographic characteristics such as age, gender, race, education, housing type, or wealth. Finally it could be based on ideological, historical, or cultural commonalities. The key feature of this peer group is that its members are influenced by social norms imposed on each other based on their common interests or characteristics, and their behavior is governed by implicit or explicit expectations about what is acceptable within the peer group.

For an individual j living in household i, preferences are represented by the utility function:

$$W_{i,j} = U\left(e_{ij}, x_{ij}\right) - D\left(\gamma_{ij}, \Lambda\right). \tag{2.1}$$

The first term, $U(\cdot)$, represents the utility derived from a consumption bundle of two goods – the quantity of personal services, $e_{i,j}$, conveyed by consuming electricity (such as the light from the use of a light bulb), and the consumption of a composite private good $x_{i,j}$. The good e_{ij} is produced by combining housing services, h_i, and electricity, $\epsilon_{i,j}$, using the household production function $g(h_i, \epsilon_{i,j})$. Housing services captures the characteristics of the residence i that are complementary to electricity consumption, such as the availability of natural light or the number of light bulb sockets available.[11] The expression $\epsilon_{i,j}$ is the quantity of electricity (kWhr) demanded by individual j in household i, and is aggregated to the household level from each individual's choices within the household through the function $E(\epsilon_{i,j})\ \forall j \in i$. Each individual determines their demand for electricity and the aforementioned goods based on their preferences while limited by their budget constraint:

$$y_{i,j} = q_h h_i / M_i + q_\epsilon \epsilon_{i,j} + q_x x_{i,j} \tag{2.2}$$

for any exogenous income $y_{i,j}$ and prices $\{q_h, q_\epsilon, q_x\}$.[12]

The second term in equation (2.1), $D(\cdot)$, represents the non-pecuniary aspects associated with the choice of $\epsilon_{i,j}$ as it relates to descriptive social norms. This type of social norm influences a household's electricity demand based on how its electricity consumption differs to that of other households' within the peer group. These norms are often

[11] In this paper we will not discuss the process by which households are formed. We assume households are formed ex-ante.

[12] The budget constraint implies that each individual within household i pays an equal proportion of the housing services cost.

presented as social comparisons where a normative standard and the participants' own level of consumption is presented side-by-side. We define γ_{ij} as the strength to which social comparisons affect individual ij's behavioral response and Λ as the social comparison function.

We model γ_{ij} and Λ using the theory of social comparisons developed by Festinger (1954). In his work Festinger argues that when there is no objective way to evaluate performance, individuals will opt for comparing themselves to the performance or opinions of others in the following ways:

(a) Individuals will choose a peer comparison group that closely resembles their characteristics relevant for such a comparison.

(b) Having chosen a peer group, individuals have the tendency to change their own position so as to move closer to the group's, and if possible, influence the group's position to move closer to their own.

(c) Individuals reevaluate their position in the peer group: an individual's tendency to compare them-self to a peer group decreases as the difference between their own opinions and actions and those of the peer group increases.

These claims inform the basis for our mathematical formalization of a social comparison function. This function depends on three main factors. First, it depends on a comparison group K, which may be a subset of the peer group N, or the entire group. Second, on information about the comparison group's consumption, as represented by the social comparison metric $\Gamma(\epsilon_K)$. This function summarizes any information given to individuals about the electricity consumption of their comparison group as part of an experimental or policy intervention.[13] As shown below and following previous studies, a possible specification for $\Gamma(\epsilon_K)$ is the arithmetic mean of the comparison group's

[13] The mean of the entire distribution is used in many instances, such as Ferraro and Price (2013). In the context of charitable giving, Shang and Croson (2006, 2009) use the upper percentiles (90th, 95th, 99th) of the distribution.

electricity consumption. Third, it depends on household i's electricity consumption, ϵ_i. Combining these elements we present the social comparison function Λ as:

$$\Lambda\big(\epsilon_i, \Gamma(\epsilon_K)\big) = \left(\frac{1}{K}\sum_{k=1}^{K}\epsilon_k - \epsilon_i\right)^2. \tag{2.3}$$

The functional form for Λ is consistent with claim (b), which states that individuals will want to move closer to the comparison group, i.e take steps to change energy usage towards the social comparison (norm reversion). A household that consumes an amount above the revealed comparison, $\Gamma(\epsilon_K)$, would receive dis-utility from their comparatively higher level of consumption, which in turn would incentivize them to reduce electricity consumption. Similarly, households that consume less than the revealed comparison would alter their behavior to consume more electricity.[14] The exponent term serves to make these incentives proportional to the difference in consumption. This means that the further away the household is from the social comparison, the greater their incentive to conform to the norm.[15]

The parameter $\gamma_{i,j}^K$ is a function of the comparison group K, and captures the effectiveness of the social comparison function in influencing an individual's behavior. This effect can be determined by various factors, such as the idea expressed in claim (a): the strength of the social comparison depends on the characteristics of the comparison group K. For example, Shang, Reed, and Croson (2008) find that participants in their experiment react more strongly to social comparisons when the comparison group is comprised of individuals of the same gender. This dependence on the compatibility between the treated individual and the comparison group's characteristics leads us to expect variation in the effectiveness of a social comparison across participants. In addition, individuals

[14] We make the assumption that the peer group in our experimental setting is large enough that the other part of Festinger's claim (b), where individuals try to influence the group to move closer to their position, is effectively nil.

[15] The exponent in this case may be any even natural number ie: $\mu \in 2\mathbb{N}$. Without loss of generality we assume $\mu = 2$.

may respond differently to different social comparison metrics as expressed in claim (c).[16] Together, $\gamma_{i,j}^{K}$ and the social comparison function $\Lambda(\cdot)$ from equation (2.3) describe the total effect of a social comparison on individual utility:

$$D(\gamma_{ij}, \Lambda) = \gamma_{ij}^{K} \left(\frac{1}{K} \sum_{k}^{K} \epsilon_k - \epsilon_i \right)^2 . \tag{2.4}$$

The individual's optimization problem

By outlining preferences, we can characterize the effect of social norms on individual electricity consumption. Assuming that individuals act rationally, and maximize their utility given a specific social environment $\{\gamma_{ij}, \Lambda(K, \Gamma)\}$. Individual j living in household i chooses a vector $\{\epsilon_{ij}, x_{ij}\}$ that solves the following problem:

$$\begin{aligned}
\max_{\epsilon_{ij}, x_{ij}} W_{ij} = &\quad U\big(e_{ij}, x_{ij}\big) - \gamma_{ij}^{K} \left(\tfrac{1}{K} \textstyle\sum_{k}^{K} \epsilon_k - \epsilon_i \right)^2 \\
s.t. & \\
y_{ij} = &\quad q_h h_i / M_i + q_\epsilon \epsilon_{ij} + q_x x_{ij} \\
e_{ij} = &\quad g(h_i, \epsilon_{ij}) \\
\epsilon_i = &\quad E(\epsilon_{ij})
\end{aligned}$$

The solution for electricity demand, $\epsilon_{ij}^{*}(y_{ij}, h_i, q_h, q_\epsilon, q_x, \Lambda(K, \Gamma))$ depends on a myriad of factors that can prevent researchers from making definitive statements about the informational effect of social comparisons in different experimental settings. For example, in a setting where households pay for electricity, the revealed information not only changes the shadow price of electricity, but also feeds through to the level of private good consumption. Thus researchers would need to adjust for income and substitution effects in order to estimate the pure informational effect of the social comparison.[17] Moreover,

[16] For example, Shang and Croson (2006) find that participants whose consumption is below the norm increase their consumption less than the decrease in consumption of participants above the norm. This suggests that individuals are affected not only by whether they are below or above the norm, but also their relative position to it.

[17] In other words, in this situation it may not be possible to hold other factors constant within $\epsilon_{ij}^{*}(\cdot)$ when adjusting an individual's access to descriptive norms.

there may be a negative correlation between electricity consumption and the strength of the response to social comparison norms (γ_{ij}) as evidenced by prior purchase of energy efficient appliances. Such a correlation would confound, or at least attenuate, the estimation of the informational effect of social comparisons. Additionally, the purchase of such appliances would signal differences in housing characteristics, which are difficult to analyze and control for when data on appliances is not available.

With respect to comparisons between experimental interventions, individuals and households are part of different social environments that influence their decisions. An estimate of the effect of social norms on electricity consumption is subject to differences in the peer groups to which the participants' belong to. Hence, comparisons between experiments hinges on knowledge about the characteristics of their respective peer groups and the social norms they enforce. As previously discussed, even though two experiments use the same social comparison, there is no reason to expect that the effect will be the same if the comparison group is perceived differently by those receiving the informational treatment. For this reason, the choice of comparison group has a direct impact on the magnitude and direction of its effect on electricity demand.

The behavioral response is also affected by which comparison metric the researcher chooses to disclose. For instance, if the distribution of electricity consumption is symmetric, and $\gamma_{i,j} = \gamma$, the model demonstrates that using the mean consumption level of the entire peer group as the social comparison would encourage those participants above the norm to decrease their consumption just as much as those below the norm are encouraged to increase theirs.[18] Under these assumptions, the simplest form of norm reversion leads to a zero estimate of a social comparison intervention. Contrast this result to the norm reversion prediction under the same assumptions if the revealed social comparison

[18] If we relax this assumption and match $\gamma_{i,j}$ to the results in Shang and Croson (2006), this functional form for $D(\cdot)$ could explain in part the small, yet significant, estimates of the effect of social norms on energy conservation where the mean level of energy consumption was chosen as the social comparison, as reviewed in Delmas, Fischlein, and Asensio (2013).

is above or below the mean. In the former case, the overall effect would be positive since a majority of participants increase their consumption. If the latter is presented, then the estimated net effect would be negative since the majority of participants decrease their consumption. Even with a simple assumption about $\gamma_{i,j}$, otherwise identical field experiments may elicit widely different results, depending on which social comparison is used and the distribution of electricity consumption.

Considering these arguments, the ideal research design to investigate the informational effect of social norms would involve a set of participants who do not have any pecuniary motives for electricity conservation. The social and demographic characteristics of these individuals should be such that they are credibly perceived as a peer group, ex ante. Furthermore, their social environment should be similar, so that comparisons within this group are regarded as appropriate and relevant. Likewise, residential characteristics and furnishings should be very similar, to the end that a reduction in electricity consumption across households is directly comparable. Finally, the chosen social comparison should be strategically placed to minimize the impact of possible norm reversion effect.

2.3 Experimental design

Experimental setting

The University of California Santa Barbara (UCSB) owns, operates, and manages a number of residence halls and apartment complexes. Our field experiment involves one particular off-campus UCSB apartment complex, which contains 200 apartments that provide housing for around 800 single undergraduate students on an annual (9-month) lease system. Of the apartments in the complex 10 are occupied by live-in staff and are not included in the experiment and the subsequent analysis. From the remaining

190 student occupied apartments, 95 were randomly selected to receive the experimental treatment described in the next section, while the other 95 formed the status quo group and received no treatment.[19]

Each apartment is built on an identical floor plan of around 900 square feet with a kitchen, two bedrooms, shower, toilet, and family room. Light fixtures, appliances, electrical outlets, and furnishings are identical across units; the apartments have no air conditioners and use gas for central heating, hot water, and the stove range. In terms of electricity use, each unit is metered for utilities, but residents only pay an individual fixed monthly rent. The university covers all utilities and residents have no access to any information on their usage. Therefore, all residents are simultaneously consumption naive and have no pecuniary incentives to conserve electricity. Furthermore, the consumption of electricity across apartments only differs by the behavior of its occupants and the plug load they use. Consequently, we expect that any conservation efforts are made on the intensive margin and are motivated by non-pecuniary benefits.

Participants' attitudes and demographics

In consequence of signing a rental agreement with UCSB, students living in the apartment complex are required to provide reasonably detailed demographic information about themselves. This information is retained by UCSB and is used to provide programming and services to the residents, and gives us a rich control dataset to work with. A selection of resident demographic characteristics and the types of apartments within the complex are shown in Table 2.1. In short, we know each resident's age, gender, major, college GPA, units taken last quarter, college standing, declared racial identity, vehicle make/model/year, home town and country, and the type of apartment they rent (up-

[19] Ex-ante power analysis based on pre-treatment electricity consumption data of these apartments indicated power is around 0.95 with $\alpha = 0.05$ and a 5% treatment effect.

stairs, downstairs, or ADA[20]). We also know whether roommates are self selected or assigned by the housing compex staff. These data are aggregated to the apartment level in our statistical analysis, and although treatment status is randomly assigned, there are a few mean differences between the control and treated groups. Namely, there are statistically fewer international students and undeclared majors in the treated group, as shown in Table 2.1.

Two months prior to the beginning of the treatment period, we conducted a survey of the apartment complex residents. This was administered as part of a University sponsored survey of UCSB undergraduates that is conducted biennially to assess the environmental and energy conservation attitudes of the student body. We use this survey to place participants in the context of the rest of UCSB's student body. A selection of comparative statistics from this survey is presented in Table 2.2. These self-reported measures indicate that the students living in the apartment complex are reasonably representative of the general undergraduate student body in their attitudes and behaviors. Summarizing the responses we deduce that most respondents recycle regularly, around half turn off their electronics and computers regularly when not in use, and about one quarter are willing to pay higher tuition (at least marginally) to help UCSB address its environmental impact. We also conducted an exit survey at the end of the 10-week treatment period, which we use to understand changes in participants' attitudes related to energy conservation.

Electricity consumption data

Electricity consumption data was provided by Southern California Edison (SCE), which is UCSB's electricity service supplier. One year prior to the commencement of our field experiment, SCE installed residential electricity smart meters[21] in the apartment com-

[20]ADA refers to Americans with Disabilities Act compliant apartments. NB: All ADA compliant apartments in the complex are located downstairs.

[21]Smart meters are electronic electricity meters that provide hourly, or better, interval monitoring, across a wireless network. Over 38 million residential units in the U.S. are equipped with smart-meters (US EIA, 2012).

plex. These meters provide SCE with billing-grade hourly consumption data on every apartment within the complex. As such we have the universe of hourly electricity consumption for all apartments over a 15 month period comprising of two residential cohorts. On average, weekly apartment electricity consumption is slightly lower at the weekend and higher earlier in the week. The diurnal electricity use follows predictable patterns with consumption increasing around sunrise, then falling away slightly during the day only to rise again in the evening before falling precipitously in the very early morning. The highest consumption levels occur around 9pm. These patterns of consumption are seen in Figure 2.1, which presents the distribution of apartment electricity use by hour and day of the week.

Treatment description

At the beginning of every week for ten weeks, between April and June 2013,[22] we emailed the occupants of every treated apartment[23] an electricity consumption report. The status quo group received no weekly email nor any information of any kind.[24] Each report included the amount of electricity in kilowatt hours consumed by that apartment over the previous week, the commensurate amount of carbon dioxide (CO_2) released as a result of electricity generation,[25] their peak 1-hour reading, a set of graphs depicting hourly rates of electricity use over the previous week, and hypertext links to unsubscribe and to the study Facebook page.

The apartment electricity usage graphs mentioned above served as our social com-

[22]The dates of treatment occurred over the Spring 2013 quarter of instruction for UCSB.

[23]We used the primary university email address of record for each resident (100% coverage). This email is used for all university correspondence.

[24]We endeavored to preserve the status quo to maintain consistency between the treatment period and prior quarters. Dummy or generic information emails were not sent; for the effect of such in general the reader can refer to Midden et al. (1983), Carrico and Riemer (2011), Ferraro and Price (2013).

[25]We used a conversion factor of 0.724 between kilowatt hours and pounds of CO_2. This is the conversion factor used by UCSB Sustainability, a University organization, for evaluating electricity reduction policies and programs and is based on local utility averages. As a point of reference, the U.S. average CO_2 emissions for 2009 was 1.22 pounds per kilowatt hour (US EPA, 2012).

parison, which showed residents their household's electricity consumption in comparison to what was called the "energy efficient apartment." This hypothetical apartment is the social comparison metric ($\Gamma(\epsilon_K)$), and was constructed from the average of the apartments within the 10th percentile of electricity use for the previous week.[26] One graph presented the recipient household's hourly consumption along with the hourly consumption of their peer "energy efficient apartment" and a second graph depicted the same information only that a deviation over the efficient apartment level for a given hour was filled with the color red, while a deviation under the efficient level was filled green. This color coding emphasized the information conveyed by the comparison with the intent to clarify the differences between above or below the efficient apartment for any given hour. An example of the electricity usage graphs and labeling of our social norm messages is shown in Figure 2.2.

The selection of the 10th percentile average as our social comparison norm was made to ameliorate a hypothesized norm reversion effect and to elicit a larger energy conservation effort from households. This choice can be justified by our model, which would predict that where $\gamma_{i,j} = \gamma$ and where the energy consumption distribution is symmetric, the reduction in electricity by those above the social comparison would far outweigh the increase by those below. Even if this particular assumption is not entirely accurate, based on the evidence presented in Shang and Croson (2006) we still expect a larger reduction in electricity consumption with our social comparison than if we used the mean of the electricity consumption distribution as our norm. Similar to the population mean, our comparison has the advantage of being simple and easily constructed for any population where the electricity consumption distribution is known.

Finally, we recognize that many unsolicited emails are never opened. For this reason we included the total amount of electricity used during the week in the subject line of

[26] The 10th percentile was determined using all the apartments in the complex with the exception of any apartments that seemed unoccupied for the relevant week.

each email, so that even if the email was not viewed in its entirety the most relevant information would still be noticed. However, this would not preclude email recipients from filtering the emails such that they are never seen. We do not know the extent to which this occurred. However, as a way to quantify the efficacy of our delivery method we tracked the number of emails that were directly opened, as shown by week in Figure 2.3, and the clicks to the field experiment Facebook page.

2.4 Experimental treatment effects

The randomized treatment-control setting of our field experiment lends naturally to the employment of a simple comparison of means to establish the statistical significance between groups. The hypothetical causal effect of the assigned binary treatment over the period of the experiment can defined as $\mathbb{E}[y_{1i}] - \mathbb{E}[y_{0i}] = \delta$, where i indexes the apartment and y is the natural log of electricity consumption. Assuming counter factual validity between groups and using Method of Moments, we define our estimator of δ as $\hat{\delta} = \bar{y}_1 - \bar{y}_0$, where $\hat{\delta}$ is the average treatment effect estimator.

A one-way ANOVA comparison of means reveals a statistical difference between groups at the 5% level over the 10-week treatment period, where the mean of the treated group is less than the control group. However, there are some characteristic differences between groups in spite of the assignment randomization (as shown in Table 2.1). A two-way factorial ANOVA testing the unconditional mean differences between the treatment groups, the pre-treatment period and treatment period, and the interaction of the two, also reveals a statistical difference between the treatment and control group during the treatment period at the 5% level.

Recognizing these mean differences we proceed by estimating the difference in electricity consumption between treatment and control groups using a difference-in-differences estimand. Using demographic variables, apartment characteristics, and fixed effects for apartment (by cohort), time, and calendar events,[27] we estimate the treatment effect while controlling for other variables. Even though we have already established the statistical significance of the treatment by our ANOVA tests, the use of a conditional estimator is beneficial since it allows us flexibility in parsing out various treatment effects and refines our standard errors. Our subsequent regression models follow the general form:

$$\begin{aligned} ln(kWh_{it}) &= \beta_0 + \beta_1(treated)_i + \beta_2(treat\_period)_t + \delta(treated_i \times treat\_period_t) \\ &\quad + \alpha_i + \tau_t + u_{it} \end{aligned}$$

where *treated* is a dummy variable for whether or not an apartment was treated, *treat_period* is an indicator for the Spring 2013 quarter of instruction when the experiment occurred, δ is our parameter of interest, α references the set of apartment level variables, and τ is the set of time relevant indicators.

2.4.1 Average and Heterogeneous Treatment Effects

Our regression results indicate apartments that received weekly emails which detailed their previous week's electricity consumption in comparison to an "energy efficient apartment" reduced their consumption by about 7% in comparison to apartments that did not.[28] As presented in Table 2.3, we find that the results are robust and significant across models with successive covariates, including apartment-cohort fixed effects.[29] As a simple robustness check we apply the difference-in-differences estimator to the quarter of instruction prior to the treatment period; there is no statistical effect.[30]

Using the point estimates from column (4) of Table 2.3 we back out the levels impact

[27] These calendar events include University and public holidays, examination periods, and special non-holiday dates (such as Halloween).

[28] Following Delmas and Lessem (2014) we sum the hourly consumption observations to the daily level. The results are robust when totaled to the weekly level or kept at the hourly level.

[29] Although we use apartment-cohort level cluster-robust standard errors, our statistical inference matches results obtained when using null-imposed wild cluster bootstrap-t inference per Cameron, Gelbach, and Miller (2008).

[30] Note that the data spans 15 months. The non-significance of the Spring 2013 quarter dummy in Table 2.3 also indicates there is nothing particularly different about that quarter compared to the Spring 2012 quarter when there was a different residential cohort and no intervention.

of the treatment by regressing kilowatt hours on the untransformed predicted values.[31] We estimate that had the status quo apartments been treated, they would have reduced their power consumption on average by 0.0118 kW. Over a 24-hour period, this reduction is equivalent to turning off a 60 watt light bulb for 4.7 hours when it would otherwise have remained on.[32]

We further analyze the difference in electricity consumption between treated and control apartments across time by partitioning out the indicator in three ways. First, we investigate whether this difference is constant through the treatment period across each week. Here, as shown in Figure 2.4, we find that over the 10 week period, substantial and significant reductions in electricity consumption did not occur until after the fifth week of treatment. This shift in effect corresponds to the week following mid-term exams, though we do not have a ready explanation for this correlation. There also appears to be a downward trend in the difference in electricity consumption between treated and control groups over the treatment time frame. This observation is consistent with qualitative evidence from our resident attitude surveys. The results of the ex-post survey suggest that treated residents increased their dialog within their apartment with respect to energy conservation. It is plausible that the effectiveness of a consistent social comparison treatment increases over time.

Second, we partition the treatment indicator by day of the week to see if residents differentially reduce their consumption on different days. These estimates are shown in Figure 2.5 where we see the largest decreases in consumption occur at the weekend and on Wednesday and Thursday. There appears to be no statistical effect early in the week. The latter is a curious result since the weekly emails were sent out on Mondays, and we expected that our communications might have a larger effect around the time they were viewed (Over 80% of the email-opens were on a Monday or a Tuesday). This result

[31]Failure to account for Jensen's Inequality, i.e. in our case $\mathbb{E}[ln(y)] < ln(\mathbb{E}[y])$, would result in a biased estimate of the treatment effect in levels. See Wooldridge (1994).

[32]4.72 hours = (24h $\times$ 0.0118kW)/0.06kW

suggests that the information conveyed is retained over the week and adjustments in electricity consumption are made when, we can only assume, it is convenient to do so.

Thirdly, we partition the difference in electricity consumption between treated and control apartments by hour of the day.[33] These results reveal, as shown in Figure 2.6, that the informational treatment significantly reduced electricity consumption in the peak morning and evening hours, and less so in the very early morning hours and in the afternoon. This result suggests residents may be more conscientious about reducing their consumption at times when they are actively using electricity in their daily routine.

Finally, by conditioning the difference in electricity consumption between treated and control groups on the quartile of consumption from the previous academic quarter, we find that the effect strengthens until the third quartile then drops, as shown in Figure 2.7. However, we cannot discern statistically significant energy conservation differences between quartile groups.

2.5 Discussion

2.5.1 Average Treatment Effect

Using billing-grade smart meter data we were able to present the residents of the treated apartments with detailed information about their previous week's electricity consumption and a comparison to their low-electricity-use peers. Recall, that for all intents and purposes, these are occupants of apartments that can only reduce energy consumption on the intensive margin and have no financial incentive to use less electricity. Even so, we measure a 7% reduction in electricity usage on average over the 10 week treatment period. This average treatment effect is equivalent to raising the price of electricity by

[33] These regression models utilize hourly observations and not total daily consumption used elsewhere.

27%-50% in the short run, and 12.8% in the long run.[34] This effect is larger than the 2% reduction in electricity consumption found in recent studies, such as Allcott (2011), Ayres, Raseman, and Shih (2009), and Costa and Kahn (2013).

The increased effectiveness of our informational treatment cannot be attributed to a single factor due to the various differences among the experimental designs of similar studies. From one perspective the increased effectiveness of our informational treatment is surprising given residents of the apartment complex have no monetary incentive to conserve electricity. On the other hand, households might find it easier to reduce their energy consumption with the use of higher frequency feedback. This explanation seems to be consistent with Ueno et al. (2006), who found a 9% reduction in electricity use for a small set of new homes where very detailed on-demand energy information feedback terminals were installed.

In addition to high frequency feedback, our social comparison contained detailed information about the differences in electricity consumption between treated apartments and the comparison. Residents were able to see the disparity between their energy consumption and that of their "energy efficient neighbor" by hourly intervals. They were able to compare their consumption patterns and use that information to strategically focus their conservation efforts. We observe this in the hourly disaggregation of the treatment effect, which showed that treated households reduced their electricity consumption during the peak morning and evening hours. We suggest that households sought to reduce their energy consumption when the observed disparities between their consumption and the social comparison were the greatest.[35]

The strength of our result might also be attributed to the similarity in housing characteristics. Our experimental units consisted of homogeneous apartments with no large

[34]We calculated this figure from the -0.10 to -0.18 Reiss and White (2008) estimate of the short-run elasticity of residential electricity demand in California with respect to a large, unanticipated price change, and the -0.39 long run elasticity estimate from (Reiss and White, 2005).

[35]The counter-factual levels and variance are both relatively larger during the peak use times than the off-peak times compared to those within the 10th percentile.

electrical appliances such as air-conditioners, electric ranges, electric washers and dryers. Households in other recent studies include a heterogeneous assortment of self-selected large appliances. Discretionary use of smaller appliances and plug load might constitute a much smaller fraction of the total electricity use in those households and could account for why our estimated effect is larger.

Another explanation for the strength of our estimate could be due to the greater applicability and credibility of the social comparison. Households in our study are relatively similar in all respects and the residents are all aware of their mutual similarities. The presentation of an "energy efficient" neighbor social comparison was proximal to each treated household's peers. In this way the validity and credibility of the comparison to the treated individuals is established. Even so, as discussed by Festinger (1954) we recognize that responses to social norms depend on the perception of the information to the individual. Apathy can result as suggested by Festinger's claim (c) where differences are perceived to be significant or the comparison group unrepresentative. In fact, we find evidence in our exit survey that corroborates this point: 9 out of 10 respondents who said that their attitude had moved *away* from energy conservation were from the treated group; most of which were in the 4th quartile of electricity consumption (i.e. those who used the most electricity for the 3 months prior to the experiment) and none from the 1st quartile. Selected exit survey results can be found in the Appendix.

Finally, our preferred explanation for the relative strength of our experimental treatment is the placement of our social comparison. Based on our model, we supposed two things with respect to the social comparison in the context of our experimental setting. First, the selection of a social comparison metric to the left of the mean would exact higher conservation efforts. Second, the strength of the social comparison response would increase the further the households consumption was from the comparison. The estimates of the treatment effect by quartile of prior electricity consumption, shown in Figure 2.7, appear to support our suppositions with this caveat. Those furthest away from our social comparison may have experienced some discouragement and thus their efforts were less that we might have predicted. This can be inferred from the fourth quartile estimate and from our exit survey results. The fourth quartile estimate appears to be attenuated compared to the third quartile, and those whose attitude moved *away* from energy conservation were correlated with those who used the most electricity.

2.5.2 Social comparisons as a policy tool

It has been suggested that non-pecuniary methods like social norms and social comparisons may be used as policy tools to encourage energy conservation. Though their efficacy is currently questionable, our results show that in a near ideal environment the social comparison approach can yield significant reductions in energy demand. We also show it can be done in a way that is easily scalable and replicable using electronic communications. Even so, there are further gains to be made in understanding social comparisons. Based on our model, we hypothesize that we can use participants' observable characteristics, particularly their historical levels of electricity consumption, to increase the overall effectiveness of our approach.

Our model theorizes the importance of a descriptive social norm response function, $D(\gamma_{i,j}, \Lambda)$, in determining the effectiveness of a social comparison intervention. The results we obtain from our field experiment provide useful information about some of the features of $D(\cdot)$, and allow us to map out some of its characteristics. Our initial hypothesis about the relative effectiveness of a social comparison based on its placement seems to have been confirmed. However, the results suggest that the strength of the norm reversion effect is subject to limitations. Some treated participants in our study, particularly in the 4th quartile of pre-treatment electricity consumption, turned away from energy conservation. This reduction in the effectiveness of the social comparison toward one end of the distribution opens the possibility of the existence of an optimal placement of the social comparison.

In order to find such an optimal placement of a social comparison (if one exists), the descriptive social norm response function would need to be sufficiently characterized.[36] As part of that characterization it seems possible that there can occur some dissociation from the revealed social comparison. Based on our results, this may be an increasing function of the distance between one's consumption and the social comparison.[37] Whether this dissociation occurs at discrete distance or as continuous function of the distance from our social comparison we cannot tell. If we assume that disassociation takes effect at a discrete distance from the social comparison, the distance between the revealed social comparison and the point of disassociation can be thought of as the effective range of the social comparison, which we denote as $r(\Gamma)$.[38] We characterize the point of disassociation as the point ϵ_Γ. In our case, ϵ_Γ would be located within the 3rd quartile, and $r(\Gamma)$ would span the distance between our chosen location for Γ (the average of the 10th percentile) and ϵ_Γ over the range of electricity consumption.

The location of a more effective social comparison might take into account the effective range $r(\Gamma)$ and be placed accordingly to increase the overall reduction of electricity use. Recall that the revealed social comparison is arbitrary and can be adjusted to suit research or policy objectives. Another alternative experimental or policy design might divide the treated population into groups and make use of more than one social comparison. The social comparison participants receive would be strategically placed to maximize, or at least increase, their behavioral response. Figure 2.8 shows one possible design with two divisions of a hypothetical population, where the first group receives social comparison norm Γ_1 and the second group receives norm Γ_2.[39]

[36] It is possible that there exists a universal descriptive social norm response function, but it is more likely there are families of response functions, each belonging to particular peer groups or subgroups of a population.

[37] This observation is consistent with Festinger's claim (c) mentioned previously.

[38] Recall that Γ is the social comparison metric.

[39] We have implicitly assumed $r(\Gamma)$ is constant across the distribution for any social comparison.

We postulate that these two approaches, either the repositioning a single social comparison, or the use of more than one comparison for different subgroups, would elicit a larger behavioral response than our original experimental specification for Γ. Similarly, we posit that using other observable characteristics to customize social norm messages could prove useful in settings where the participants are more heterogeneous than ours. Tailoring messages based on household characteristics other than just prior consumption levels could allow for more insightful and potentially effective social comparisons. In all, by conditioning the social comparison(s) on observable variables, we suggest that social scientists and policy makers interested in energy conservation could increase the effectiveness of revealed social comparison interventions. These issues and more are left for future research.

2.6 Conclusion

The field experiment we conducted provides convincing evidence that private social comparisons can encourage electricity conservation even when the marginal cost of consumption is effectively nil. This demonstrates the potential efficacy of demand-side alternatives as complements to supply interventions such as pollution taxes or constructed markets. Our work suggests that households alter their behavior depending on the type of social comparison they are exposed to. Our estimates are much larger than those of other studies that use the mean as the descriptive norm. Reasons discussed for this difference include our use of an increased frequency of feedback, temporally disaggregated information, the use of a challenging social comparison, and the relative population homogeneity.

Our study estimates the timing and progress of changes in electricity demand by showing that individuals choose to conserve when it is most convenient to do so, and that they become more effective in their conservation efforts over the course of the treatment period. On the other hand, we present evidence that the type of social comparison can have other effects on individuals – some participants who historically were high electricity consumers seem to have been dissuaded from conserving energy. We hypothesize that this is a function of the significant distance between their consumption levels and those of our social comparison.

We argue that characterization of the descriptive social norm response function, as we have modeled, opens the possibility of an optimal placement for a social comparison(s) to induce larger responses in similarly structured interventions. We hypothesize that future work in mapping out this response function will further improve the results of social norm interventions, making them more attractive and predictable as policy tools to mitigate pollution or to encourage other pro-social behaviors.

References

Abrahamse, Wokje, Linda Steg, Charles Vlek, and Talib Rothengatter. 2005. "A review of intervention studies aimed at household energy conservation." *Journal of Environmental Psychology* 25 (3):273–291.

Allcott, Hunt. 2011. "Social norms and energy conservation." *Journal of Public Economics* 95:1082–1095.

Allcott, Hunt and Sendhil Mullainathan. 2010. "Behavioral science and energy policy." *Science* 327 (5970):1204–1205.

Ayres, Ian, Sophie Raseman, and Alice Shih. 2009. "Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage." NBER Working Papers 15386, National Bureau of Economic Research, Inc.

Becker, Lawrence J. 1978. "Joint effect of feedback and goal setting on performance: A field study of residential energy conservation." *Journal of Applied Psychology* 63 (4):428–433.

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine L. Milkman. 2011. "The effect of providing peer information on retirement savings decisions." NBER Working Papers 17345, National Bureau of Economic Research, Inc.

Bittle, Ronald G., Robert M. Valesano, and Greg M. Thaler. 1979. "The effects of daily feedback on residential electricity usage as a function of usage level and type of feedback information." *Journal of Environmental Systems* 9 (3):275–287.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-based improvements for inference with clustered errors." *The Review of Economics and Statistics* 90 (3):414–427.

Carrico, Amanda R. and Manuel Riemer. 2011. "Motivating energy conservation in the workplace: An evaluation of the use of group-level feedback and peer education." *Journal of Environmental Psychology* 31 (1):1–13.

Costa, Dora L. and Matthew E. Kahn. 2013. "Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment." *Journal of the European Economic Association* 11 (2):680–702.

Delmas, Magali, Miriam Fischlein, and Omar Asensio. 2013. "Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975-2011." *Energy Policy* 61:729739.

Delmas, Magali A. and Neil Lessem. 2014. "Saving power to conserve your reputation?

The effectiveness of private versus public information." *Journal of Environmental Economics and Management* 67 (3):353–370.

Ferraro, Paul J. and Michael K. Price. 2013. "Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment." *The Review of Economics and Statistics* 95 (1):64–73.

Festinger, Leon. 1954. "A theory of social comparison processes." *Human relations* 7 (2):117–140.

Hayes, Steven C. and John D. Cone. 1977. "Reducing residential electric use: Payments, information, and feedback." *Journal of Applied Behavior Analysis* 10 (3):425–435.

Levitt, Steven D. and John A. List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic Perspectives* 21 (2):153–174.

McGinnis, Michael and Elinor Ostrom. 2008. "Will lessons from small-scale social dilemmas scale up?" In *New Issues and Paradigms in Research on Social Dilemmas*. Springer, 189–211.

Midden, Cees J. H., Joanne E. Meter, Mieneke H. Weenig, and Henk J. A. Zieverink. 1983. "Using feedback, reinforcement and information to reduce energy consumption in households: A field experiment." *Journal of Economic Psychology* 3:65–86.

National Science and Technology Council. 2015. "Social and Behavioral Sciences Team Annual Report." Tech. rep., Executive Office of the President, Washington DC.

Ostrom, Elinor. 2014. "A polycentric approach for coping with climate change." *Annals of Economics and Finance* 15 (1):97–134.

Pallak, Michael S. and William Cummings. 1976. "Commitment and voluntary energy conservation." *Personality and Social Psychology Bulletin* 2 (1):3–13.

Petersen, John E., Vladislav Shunturov, Kathryn Janda, Gavin Platt, and Kate Weinberger. 2007. "Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives." *International Journal of Sustainability in Higher Education* 8 (1):16–33.

Reiss, Peter C. and Matthew W. White. 2005. "Household electricity demand, revisited." *The Review of Economic Studies* 72 (3):853–883.

———. 2008. "What changes energy consumption? Prices and public pressures." *The RAND Journal of Economics* 39 (3):636–663.

Seaver, W. Burleigh and Arthur H. Patterson. 1976. "Decreasing fuel-oil consumption through feedback and social commendation." *Journal of Applied Bahavior Analysis* 9 (2):147–152.

Seligman, Clive and John M. Darley. 1977. "Feedback as a means of decreasing residential energy consumption." *Journal of Applied Psychology* 62 (4):363–368.

Sexton, Richard J., Nancy Brown Johnson, and Akira Konakayama. 1987. "Consumer response to continuous-display electricity-use monitors in a time-of-use pricing experiment." *Journal of Consumer Research* 14 (1):55–62.

Shang, Jen and Rachel Croson. 2006. "The impact of social comparisons on nonprofit fund raising." *Research in Experimental Economics* 11:143–156.

———. 2009. "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods." *Economic Journal* 119 (540):1422–1439.

Shang, Jen, Americus Reed, and Rachel Croson. 2008. "Identity congruency effects on donations." *Journal of Marketing Research* 45 (3):351–361.

Ueno, Tsuyoshi, Fuminori Sano, Osamu Saeki, and Kiichiro Tsuji. 2006. "Effectiveness of an energy-consumption information system on energy savings in residential houses based on monitored data." *Applied Energy* 83 (2):166–183.

UNEP. 2012. *The Emissions Gap Report 2012*. United Nations Environment Programme (UNEP), Nairobi.

US EIA. 2009. "Residential Energy Consumption Survey for 2009." Form EIA-457 A and C-G, United States Energy Information Administration, Office of Energy Consumption and Efficiency Statistics.

———. 2012. "Annual Electric Power Industry Report." Form EIA-861, United States Energy Information Administration.

US EPA. 2012. "The Emissions and Generation Resource Integrated Database for 2012 (eGRID2012)." Tech. Rep. Version 1.0, United States Environmental Protection Agency.

van Houwelingen, Jeannet H. and W. Fred van Raaij. 1989. "The effect of goal-setting and daily electronic feedback on in-home energy use." *Journal of Consumer Research* 16 (1):98–105.

Viscusi, W. Kip, Joel Huber, and Jason Bell. 2011. "Promoting recycling: Private values, social norms, and economic incentives." *American Economic Review* 101 (3):65–70.

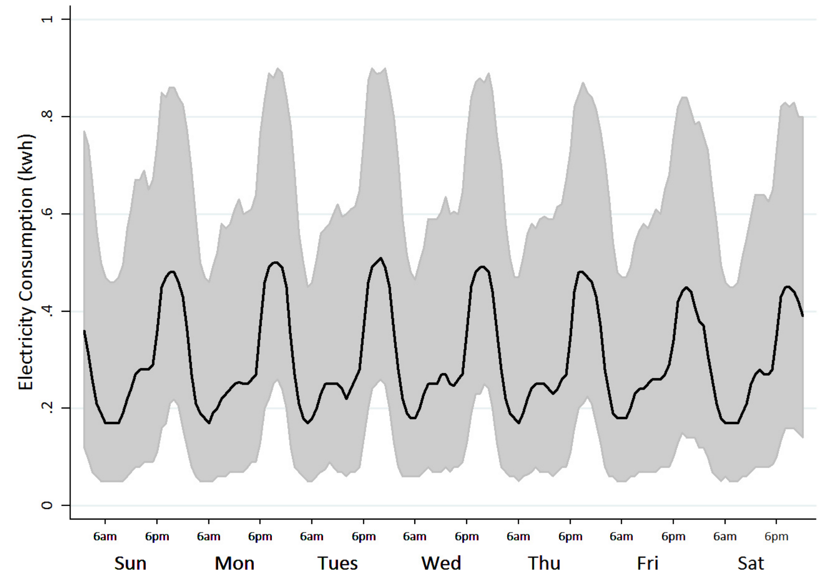
Wilhite, Harold and Rich Ling. 1995. "Measured energy savings from a more informative energy bill." *Energy and Buildings* 22 (2):145–155.

Winett, Richard A., Michael S. Neale, and H. Cannon Grier. 1979. "Effects of self-monitoring and feedback on residential electricity consumption." *Journal of Applied Behavior Analysis* 12 (2):173–184.

Wooldridge, Jeffrey M. 1994. "A simple specification test for the predictive ability of transformation models." *The Review of Economics and Statistics* 76 (1):59–65.

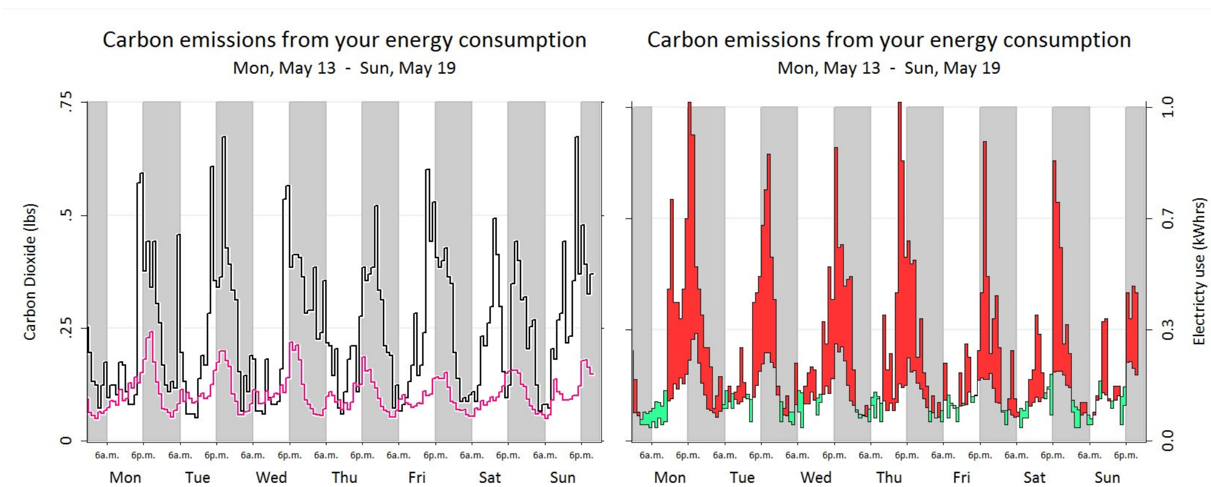
Figures and Tables

Figure 2.1: Hourly distribution of electricity use by day of the week



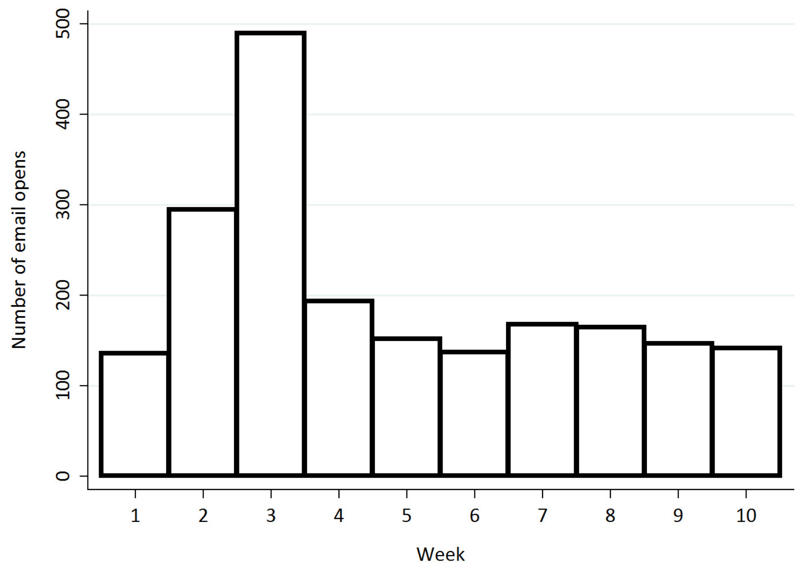
Note: Median hourly consumption in black, shaded area between 5th and 95th percentiles
Source: Authors' compilation of SCE smart meter data between Jan 7 and Mar 22, 2013

Figure 2.2: An example of graphs emailed to residents of an apartment



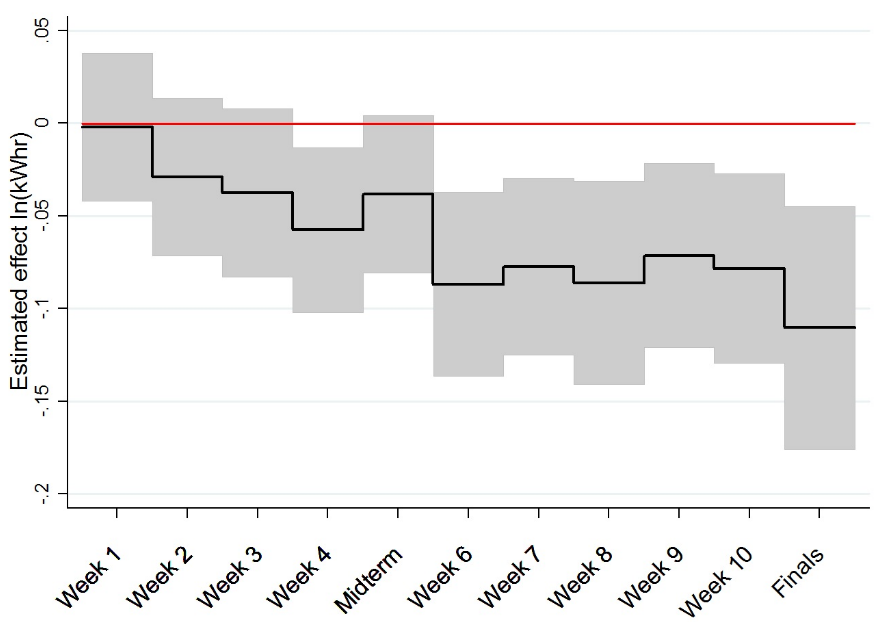
Source: Southern California Edison smart meter readings from May 2013

Figure 2.3: Number of email opens by week



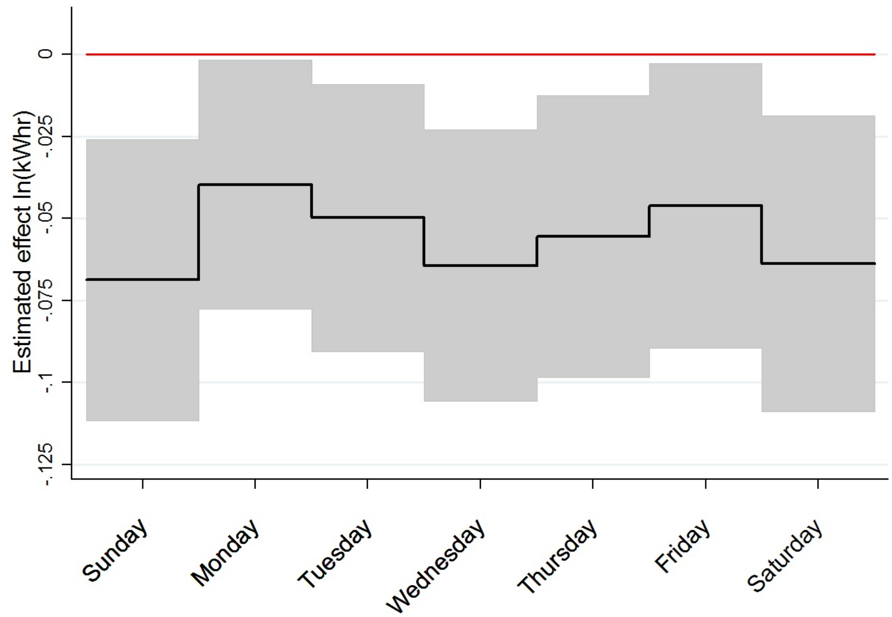
Source: Authors' compilation of email data over the treatment period

Figure 2.4: Treatment effect by week



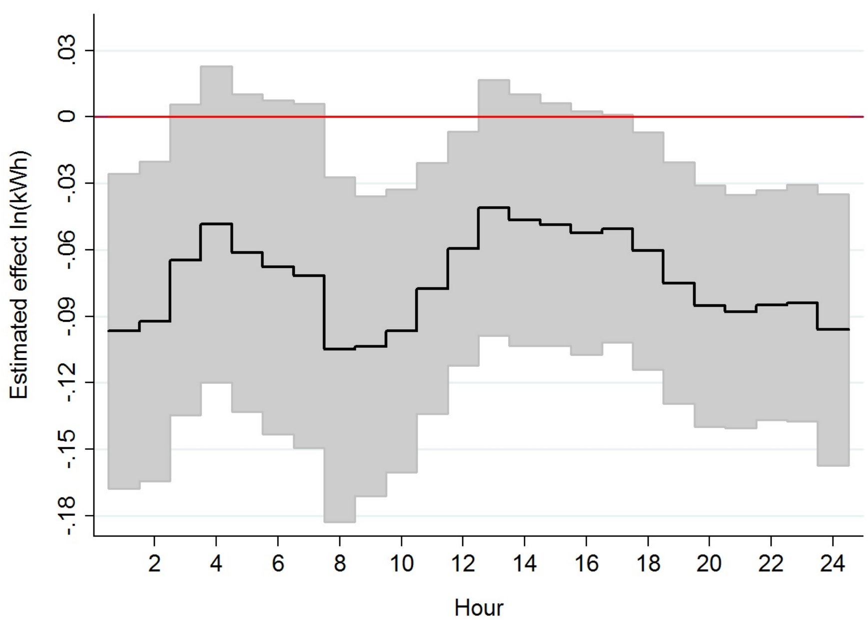
Note: Point estimates in black, 95% confidence intervals using cluster robust standard errors (apartment cohort) shown as the shaded area.

Figure 2.5: Treatment effect by day of the week



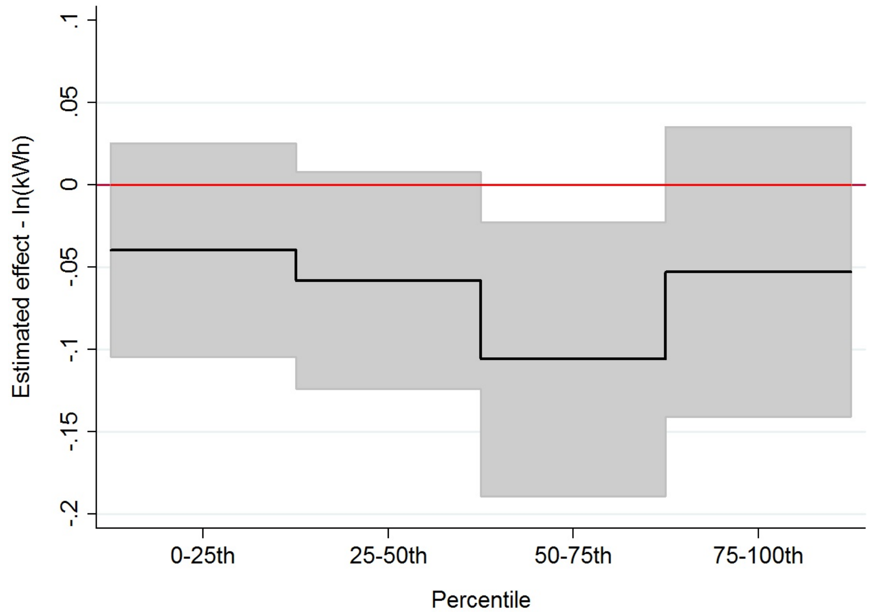
Note: Point estimates in black, 95% confidence intervals using cluster robust standard errors (apartment cohort) shown as the shaded area.

Figure 2.6: Treatment effect by hour of the day



Note: Point estimates in black, 95% confidence intervals using cluster robust standard errors (apartment cohort) shown as the shaded area.

Figure 2.7: Treatment effect by consumption quartile



Note: Point estimates in black, 95% confidence intervals using cluster robust standard errors (apartment) shown as the shaded area.

Figure 2.8: Two division social comparison design

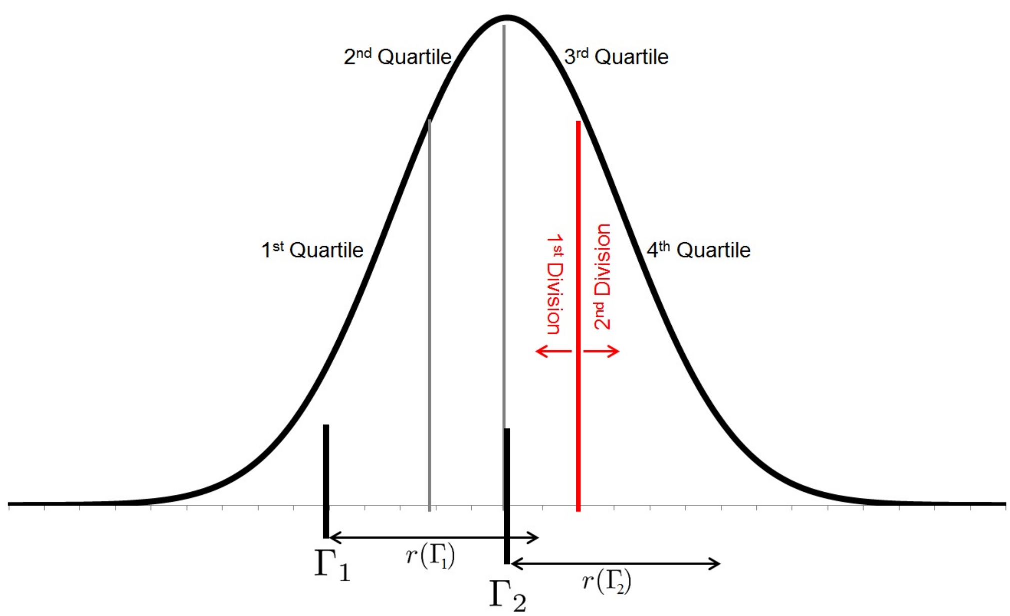


Table 2.1: Selected descriptive variable comparisons

			Control	Treated
	Min	Max	Mean	Mean
Male apartment	0	1	0.51	0.43
Average age (years)	20	25.8	21.3	21.3
Average units taken (prev. quarter)	11.25	18	13.9	14.1
Average GPA (4.0 scale)	2.1	4	3.2	3.1
Self assigned (%)	0	100	75.4	78.5
Senior (%)	0	100	39.0	42.5
International student (%) **	0	75	10.5	5.5
Transfer student (%)	0	100	54.1	56.8
Began UCSB as Freshman (%)	0	100	35.4	37.7
Major - undeclared (%) **	0	75	10.5	5.7
Major - Communications (%)	0	100	25.2	28.4
Major - Environmental Studies (%)	0	50	2.6	2.7
Race - Black (%)	0	50	3.8	3.1
Race - Hispanic (%)	0	100	30.9	30.2
Race - Asian (%)	0	100	25.0	28.1
Vehicle - New (%)	0	100	16.9	18.6
Vehicle - American make (%)	0	100	8.4	6.5
Vehicle - SUV model (%)	0	50	5.9	8.4
Upstairs apartment	0	1	0.49	0.41
Cottage apartment	0	1	0.11	0.08
ADA compliant apartment	0	1	0.08	0.06
N			95	95

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.2: Environmental attitudes survey results

	Apt. Complex	UCSB
Junior or Senior (%)	98.9	56.0
Male (%)	37.2	33.3
Important to address own impact on the environment (%)	95.1	92.4
Voted green in last election (%)	38.0	45.2
Member of an environmental group (%)	15.6	17.4
Regularly use recycle bins (%)	76.2	78.1
Regularly turn off electronics when not in use (%)	55.4	50.4
Willing to pay higher tuition to protect the environment (%)	26.0	27.7
Response rate	0.22	0.19
N	90	340

*** p<0.01, ** p<0.05, * p<0.1

Table 2.3: Average treatment effect estimates

Log(kilowatt hours)	(1)	(2)	(3)	(4)
Treated	0.0566	0.0566	0.0449	-
	(0.0416)	(0.0416)	(0.0316)	-
Spring 2013	0.0009	-0.0021	-0.0050	-0.0057
	(0.0119)	(0.0112)	(0.0120)	(0.0119)
Treated × Spring 2013	-0.0702***	-0.0702***	-0.0644***	-0.0630***
	(0.0209)	(0.0209)	(0.0211)	(0.0189)
Month & event FE	yes	yes	yes	yes
Day & hour FE	no	yes	yes	yes
Demographics & apt type	no	no	yes	yes
Apartment cohort FE	no	no	no	yes
Observations	56,807	56,807	56,807	56,807
R-squared	0.09	0.10	0.36	0.69

Cluster robust standard errors in parentheses (apartment cohort)
*** p<0.01, ** p<0.05, * p<0.1

Appendix

Energy conservation attitude changes

At the end of the experimental treatment period we conducted an exit survey that went to all residents of the apartment complex. The purpose of this survey was to qualitatively determine if there were any changes in energy conservation attitudes, and to shed light on any actions used to conserve energy. Selected results are presented in Table A2.1. We found significant differences between the responses of those in treatment and control groups in some areas but not in others.

Compared to the control group, a larger proportion of survey respondents in the treatment group reported talking *more* regularly with their roommates about energy conservation. The reciprocal response was also significant; we find that almost half of respondents in the control group *rarely* talked with their roommates about energy conservation, while only 19% of the respondents in the treated population made this claim.

About 40% of survey respondents who were in the treated group reported that their attitude has moved *toward* more energy conservation; but only about 20% of status-quo group respondents reported that shift. Fifty-two percent of the treated respondents reported that their energy conservation attitude stayed the same, with almost 80% of control group reporting no change. In addition, 8% of treated respondents said that their attitude had moved *away* from energy conservation. Almost all of these were in the upper percentiles of the electricity consumption distribution.

Tables

Table A2.1: Exit survey results by treatment status

	Mean	
	Control	Treated
Talked regularly about energy conservation? ***	0.05	0.19
Talked rarely about energy conservation? ***	0.42	0.19
Attitude moved toward energy conservation? ***	0.22	0.39
Attitude stayed the same? ***	0.77	0.52
Attitude moved away from energy conservation? **	0.01	0.08
4th quartile (high consumption users) ***	0.00	0.17
Took no effort to conserve? *	0.42	0.32
Regularly in agreement with roommates' energy use?	0.35	0.31
Used computer less?	0.18	0.14
Turned off computer regularly?	0.43	0.39
Turned lights off regularly?	0.85	0.85
Took shorter showers regularly?	0.22	0.25
Likely to conserve over next 3 months?	0.84	0.76

*** p<0.01, ** p<0.05, * p<0.1, two sided t-test

Chapter 3

Colonial Origins and Forestland Outcomes: The Impact of Institutions on Forests and Timber.

Summary

This chapter examines the effect of country-level political stability and investment security on forestland use. Using cross-section data I find that these measures are associated with benign outcomes for overall rates of forest area change and roundwood production. These associations are robust when instrumented for endogeneity, and reveal stronger impacts than OLS estimates would imply. Two-stage least squares results indicate that a one standard deviation increase in political stability, as currently measured by the World Bank, increases forest area by 12% over 10 years, and increases roundwood production by a factor of 10 to 19. Targeting political stability and investment security may be one of the most effective tools in mitigating carbon emissions through forest expansion and increasing forest productivity.

3.1 Introduction

What are the reasons for significant variation in the levels of forest cover and forest use among countries? Environmental Kuznets curves suggest a U-shaped relationship where economic development drives environmental quality, including forest cover.[1] Can the level of a country's development alone explain why some countries experience growing levels of forest cover and others do not? The implied cause and effect of the Kuznets reasoning belies the counter point that significant environmental degradation stifles economic development. Along another line of reasoning, the underlying quality of property rights and political institutions are considered to be both drivers of economic development and environmental quality. Yet it may be the case that richer nations can 'afford' the kinds of institutions and laws that promote development (Acemoglu, Johnson, and Robinson, 2001) and environmental quality.

The endogenous interplay between property rights, enforcement, and natural resources is clearly demonstrated in theoretical modeling by Copeland and Taylor (2009).[2] Yet this simultaneity has been largely ignored in the empirical literature on forestland use.[3] Failure to account for endogeneity between forestland use and political institutions would at best lead to a positive bias in empirical estimates if indeed richer countries can afford – and generally choose – higher levels of environmental quality. In the worst case,

[1] Kauppi et al. (2006) observe that no country with per capita GDP over $4600 (2003 U.S. dollars) experienced a declining level of forest stock in 2005. Naidoo (2004) finds a positive correlation between the level of *historical* forest clearance and current economic growth. Deacon (1994) reports an association between *contemporaneous* deforestation and lower economic growth.

[2] Copeland and Taylor present a model where "the success or failure in resource management" is endogenously determined by enforcement power, harvesting capacity, and resource rents.

[3] Of the many empirical deforestation studies reviewed, only Alston, Libecap, and Mueller (2000), and Araujoa et al. (2009) explicitly pursue an identification strategy that recognizes endogeneity between deforestation and political stability. Both these papers were limited to deforestation in Brazil. Arcand, Guillaumont, and Jeanneney (2008) instrument for exchange rates but not investment security in their empirical estimation of deforestation.

endogeneity would likely lead to ambiguous results.[4] Disentangling this relationship is required in order identify the causal impact of institutional robustness on forestland use outcomes.

In this paper I exploit the same source of arguably exogenous variation used by Acemoglu, Johnson, and Robinson (2001), to estimate the degree to which political stability, as measured by The World Bank (2015), and investment security, as indexed by Bohn and Deacon (2000), affect forest cover change, and industrial roundwood[5] production. As revealed in the results, these causal relationships are are often quite different than those which simple OLS correlations would imply. For example, while the OLS results indicate that a one standard deviation change in investment security is correlated with a 4 to 5% change in forest area over ten years, the IV estimates reveal the ceteris paribis effect is double this. For roundwood production the difference between the two is over four times larger. Compared to political stability, investment security stands out as the more reliable measure. It tends to hold up better among a battery of robustness tests.

Though many papers investigate and reveal much about forest cover change, particularly with respect to deforestation, this paper fills a gap in the empirical literature by applying a tried identification strategy to overcome the endogeneity between forestland use and institutions. The results also support the idea that afforestation and timber production are not strictly competing objectives, but strengthening institutions influences benign forest cover change and increases roundwood production. To that end this paper is laid out as follows – Section 2 presents a background of previous work in natural resource exploitation in general and forestland use in particular with respect to institutions and property rights. Section 3 describes the regression models and data I utilize,

[4]This may be seen to a degree in the environmental Kuznets curve literature with respect to the empirical relationship between GDP and deforestation where results are often ambiguous. See Mather and Needle (1999) for a brief discussion. Environmental Kuznets curve models, in general, are not statistically robust (Stern, 2014).

[5]All roundwood felled or otherwise removed or harvested, except fuelwood. Includes pulpwood. sawlogs, and veneer logs. FAO item code 1865.

while Section 4 presents the results of the empirical analysis. Sections 5 and 6 contain a discussion of the results and summary conclusions respectively.

3.2 Background

3.2.1 Resource outcomes, institutions, and property rights

Coase (1960) defines the basis for almost all discussion on environmental outcomes and property rights – the fact that property rights are not well defined, perfect enforcement does not exist, and/or negotiation transaction costs are non-negligible leads us to the conclusion that individuals and governments do not operate in a first-best world. To this end a great deal of research investigates the effect imperfect property rights and political institutions have on environmental outcomes. For example, modeling by Congleton (1992) indicates that authoritarian regimes face relatively higher marginal pollution abatement costs and will generally adopt less stringent environmental standards. While Fredriksson and Svensson (2003) theorize that although corruption unambiguously leads to less stringent environmental law, the effect is attenuated by the level of political instability – the incentive to bribe officials is reduced when the likelihood of a future payoff falls. Ungoverned common pool renewable resources subject to population or demand pressures can also lead to significant environmental degradation (Brander and Taylor, 1998; Taylor, 2011). In order to avoid these 'tragedies of the commons' Dietz, Ostrom, and Stern (2003) assert certain conditions, which include multi-layered governance (quality of institutions) and rule compliance (property rights enforcement), that should prevail.

Regarding natural resource extraction, similar outcomes are theorized and observed with respect to property rights and institutional quality. Unregulated fisheries tend to result in over-harvesting[6] and resource depletion (Gordon, 1953; Schaefer, 1957).[7] Costello, Gaines, and Lynham (2008) empirically demonstrate that well-governed catch shares can prevent fishery collapse. Oil exploration and production is greater when ownership is secure (Bohn and Deacon, 2000), while production volatility is lower among politically open governments (Metcalf and Wolfram, 2014). The status-quo of the political environment, property rights, and contractual law can either hinder or help the maximization of economic rents from both oil fields and fisheries (Johnson and Libecap, 1982; Libecap, 1984; Libecap and Wiggins, 1985).

3.2.2 Forestland uses, institutions, and property rights

Like the broad class of natural resource exploitation issues mentioned above, the role of property rights is advanced as one of the leading factors in deforestation. In the context of a Faustman model Mendelsohn (1994) demonstrates, even in the presence of small eviction rates (whether lawful or not), that destructive activities dominate over sustainable forestland use. Also using a Faustman approach, Amacher, Koskela, and Ollikainen (2008) show that migration pressures, expropriation risk, and insecure property rights lead to illegal logging, agricultural expansion, and deforestation. Empirically, Southgate, Sierra, and Brown (1991) find a positive correlation between land tenure insecurity and deforestation in Ecuador. Forest clearing by indigenous land holders in Ecuador was further found to be associated with rancher encroachment, but not other types of pressures by Godoy et al. (1998). Alston, Libecap, and Mueller (2000) not only discover that land reform and weak property rights in the Brazilian Amazon lead to forest clearing, forest clearing also reduces the potential for violent conflict between landowners and squatters (see also de Oliveira (2008); Araujoa et al. (2009)). In the case of Sierra Leone, conflict

[6] Over-harvesting with respect to long term economic rent.

[7] Open access fishery exploitation modeling of this sort has lead to a large body of bio-economic fishery modeling; see Conrad (1989); Deacon (1989) for examples.

itself tends to reduce deforestation (Burgess, Miguel, and Stanton, 2015). Somanathan (1991) observes that the system of land tenure in the Central Himalaya is a fundamental cause of deforestation. Conversely, Barbier and Burgess (2001) find no significant association between a property rights index and agricultural land expansion in Asia, Africa, and Latin America; while Culas (2007) finds some evidence for an effect only in Latin America.

Government stability and institutional quality is also thought to influence forestland use.[8] Deacon (1999), and Bohn and Deacon (2000) report broad significant associations between their investment security index (which is based off government indicators and political stability measures) and afforestation rates. Similarly, Didia (1997), and Mather and Needle (1999) document a negative correlation between the level of democracy and deforestation. Burgess et al. (2012) report increased deforestation in relation to increased fragmentation of jurisdictional areas in Indonesia. Using the investment security index Arcand, Guillaumont, and Jeanneney (2008) find only a week correleation between investment security and deforestation in some models. However, they do find evidence to suggest deforestation may *increase* with 'better' institutions in developing countries. Ferreira (2004) finds trade liberalization affects deforestation only conditional on its interaction with measures of governance quality. Barbier and Burgess (2001) report an overall significant positive association between political stability and agricultural land expansion; though regionally the significance only extends to Latin America, but not Africa or Asia. On the other hand, Bhattarai and Hammig (2001) present correlations between institutional quality and deforestation for each of those three regions. There is a degree of empirical conjecture regarding the roles institutions play in forest area change. By using an instrumental variables identification strategy I am able to address what seem to be endogeneity issues that encumber a number of previous studies.

[8] Some authors regard government and institutional indicators as proxies for property rights (Deacon, 1994; Knack and Keefer, 1995).

Outside of deforestation the literature on political stability and forestland use is sparse. Regarding forest burning Wright et al. (2007) observe that countries beset by corruption are less effective at maintaining tropical forest preserves from forest fire. Iliadisa, Papastavroub, and Lefakis (2002) find that political instability in Greece is associated with burned area and numbers of forest fires. Pokharel et al. (2007) very briefly mention a lower incidence of fire in Nepal for forests under community management. On the question of forest products, Ferreira and Vincent (2010) find that improvements in governance indicators in countries with weak regimes are associated with increases in forest product output. While others document the role of natural resource extraction, which includes timber, in a political economy and conflict context (Lambin et al., 2001; Le Billon, 2001; Smith et al., 2003; Johnston, 2004; Snyder, 2006; Caselli and Tesei, 2011). In order to expand the literature a little further beyond deforestation this paper also investigates the effect of robust institutions on industrial roundwood production. This approach also has the benefit of determining whether afforestation and timber production are competing outcomes or not across countries.

The significant consequences of deforestation have spurred a large body of empirical research – some of which I have reviewed above. Were it not for the fact that identification strategies seem to be lacking from this literature generally, one might question the need for more research on the subject. However, as Reduced Emissions from Deforestation and forest Degradation (REDD) becomes a primary vehicle and program in the developing world to mitigate the effects of global climate change, estimation of the causal importance of political institutions is necessary.[9] Since prior literature is largely limited to deforestation, clearly more is needed to not only evaluate causal forest area change determinants but also the drivers of timber production. This paper seeks to fill these gaps.

[9]The programmatic potential of REDD to centralize forestland use planning at the national level may or may not be uniformly beneficial (Phelps, Webb, and Agrawal, 2010; Thompson, Baruah, and Carr, 2011).

3.3 Political stability and forestland use

3.3.1 Regression models

I begin by following an outline of the afforestation regression model used by Deacon (1999), and Bohn and Deacon (2000) where change in forest cover is a linear relationship of the regressors of interest[10] – measures of political stability and investment security, and the environmental determinants of forest stock, along with demand variables for forest services. I estimate regressions of the following basic form:

$$(1) \qquad y_i = \beta_0 + \beta_1 stability_i + \beta_2 security_i + \mathbf{X}_i\gamma + \epsilon_i,$$

where y_i is the forestland use outcome for country i (either the change in forest area, or the log of roundwood production), $\mathbf{X}$ is a vector of the other explanatory variables, and ϵ is a random error term. In order that the comparative impact of stability and security on the forestland use variable may be assessed, I transform the two measures into standardized (z-score) vectors. I assume the causal effect of *stability* and *security*, i.e. β_1 and β_2, are greater than 0. All other things being equal, where there is an idiosyncratic difference in either political stability or investment security between one country and another, the country with the higher level will on average have a higher level of afforestation (or at least an attenuated level of deforestation) and roundwood production.

The reasoning for the assumptions about β_1 and β_2 is clear – in general, the provision and effective enforcement of property rights decreases the discount rate associated with postponing the realization of forest value.[11] Thus instability and insecurity engender the types of land use decisions that tend to increase the opportunity cost of waiting to

[10]Observationally, the linearity assumption is justified in Figure 3.1 as discussed in the Data section.
[11]This line of reasoning is modeled theoretically by Deacon (1999).

liquidate the forest asset in favor of a short term cash crop. With respect to roundwood production I follow the argument made by Bohn and Deacon (2000), that when property rights are insecure, timber will be cut at an earlier age and replanting/regeneration efforts will be lower, reducing the overall timber supply from the country.[12]

Within the vector $\mathbf{X}_i$ I include environmental attributes assumed to affect the proliferation of forests. These include the initial or starting level forested area, the relative abundance of freshwater resources, and the ratio of coastline to land area. As a physical process, a larger starting extent of forest area and a larger ratio of coastline to land area will tend to reduce the potential for forest expansion. With respect to roundwood harvest, a larger forest will increase the opportunities for larger harvests, whereas a higher ratio of coastline will tend to reduce those opportunities. A relative abundance of freshwater resources will tend to increase potential forest growth, positively affecting the likelihood for forest expansion. While this same effect might positively affect the potential for roundwood harvest, an over abundance could also hinder the ease of harvest, thus the square of the water resources variable is also included.

The vector $\mathbf{X_i}$ also includes variables assumed to affect the demand for forest land and products – namely population and road density. Though the effect of population density on deforestation is actively questioned (Carr, Suter, and Barbieri, 2005) I include it for completeness in comparison to other papers. Roads are also thought to affect the demand for forest land and timber products. In the former case it is argued that roads into forested areas reduce the cost of forest exploitation (Nelson and Hellerstein, 1997), in the latter case road density may proxy for economic development, which generally increases demand for timber products (Chou and Buongiorno, 1984; Turner and Buongiorno, 2004)). Since

[12] I implicitly argue that, on average, the long run effect on timber supply will dominate over the short run liquidation effect in my empirical analysis. Given that the security and stability measures do not vary much between decades this assumption may seem reasonable. A departure from this argument is the hypothesized influence of corruption on increased timber supply from illegal logging (Seneca Creek Associates, LLC and Wood Resources International, LLC, 2004), where the short run effect may dominate in some countries.

the influence of GDP is often empirically ambiguous and also not included in the forest modeling of Deacon (1999), and Bohn and Deacon (2000), I do not include it in the body of this analysis. A brief discussion of results that include GDP as a regressor is found in Appendix B.

I recognize that each country's level of political stability and investment security is not a random draw. If indeed countries with higher levels of stability and security tend to demand higher levels of environmental quality, and if simultaneously environmental quality positively affects national stability and security,[13] then we might expect a positive bias where an OLS estimator is employed. However, some have argued that an abundance of natural resources may engender instability and insecurity (Le Billon, 2001; Snyder, 2006; Lujala, 2010),[14] in which case we might expect a negative bias in an OLS estimator. If both countervailing effects are valid then the estimator may not be identified and estimates will be ambiguous.

Given endogeneity issues exist, I utilize an instrumental variables approach to identify quasi-experimental impacts. The effectiveness of this approach hinges on two factors – the exogeneity and validity of the instrument. I argue that settler mortality is a reasonable instrument that satisfies both criteria. The validity condition is carefully established by Acemoglu, Johnson, and Robinson (2001), which I summarize as follows: the current quality of a country's institutional governance is a function of earlier institutions set up during the colonial expansion by European countries. The quality of these early institutions is function of the experience of early settlements, and the difficulty of their establishment is measured by the level of settler mortality. Heterogeneous difficulty in establishing settlements endures through to present levels of political stability and investment security, and this establishes the validity of the instrument. However, the IV system is under-identified with only only instrument, and as such the IV results below

[13] Miguel, Satyanath, and Sergenti (2004) find lack of rainfall is a precursor to civil conflict in Africa. Hsiang, Meng, and Crane (2011) discover a correlation between a warm climate phase and civil conflict.

[14] See Ross (2004) for an evaluation and discussion of this literature.

include one of each of the two regressors of interest. With some caution I present IV estimates where both stability and security are included. These regressions use settler mortality and early institutional variables as instruments.

To establish the exclusion restriction, I argue that the level of settler mortality in previous centuries is functionally ergodic to current forest use. It is known that early settlers utilized forests in particular ways that have some influence on the composition of forests today.[15] By in large settlers and colonialists cleared forests for agricultural land and harvested wood products for timber and fuel; their methods and practices were common. While the successful exploitation of forest resources for settler establishment would have been influenced by the numbers of settlers who did not die, subsequent forestland use today, in terms of deforestation and logging production, has no other influence than the residual effect through the quality of the institutions that remain. Land that may have been entirely cleared of forest cover by settlers has, since that time, had more than a hundred years to recover (if left fallow or replanted) – a sufficient time span for the establishment of mature stands in many cases,[16] and/or to be converted to agricultural, residential, or industrial uses.

3.3.2 Data and descriptive statistics

The data used in this analysis are arranged into two country-level cross-sections, one for the 1990 decade and one for the 2000 decade. These data are presented as descriptive statistics in Table 3.1 for the whole world as well as the settler subsample. Country-level physical attributes are constant between the two decades and are only listed once in the table. The settler subsample is not representative of the whole world. On average those

[15] See Brender (1974) as an example.

[16] Note that mature secondary forests are not akin to old-growth or primary forests. Of the estimated 5.6 billion hectares of primary forest prior to the 1700's, approximately 1.4 billion hectares remain. The balance of the current 4 billion hectares of forestland in the world is secondary forest (i.e. forest that has been cut-over, cleared, or burned by humans at least once). See FAO (2012, 2015).

countries have a higher rate of deforestation and lower political stability score. There are also many fewer island states in the settler subsample, which is reflected in the much lower coastline and roads to area ratio.

The dependent variables, namely change in forest area and log average annual roundwood production are constructed from data in the U.N. Food and Agriculture database FAOSTAT and the FAO Forestry database CountrySTAT. Change in forest area is the difference between the 1990 and 2000, and the 2000 and 2010 forest area for each country divided by the initial forest area (either 1990 or 2000, respectively). The log average roundwood production cross-section for each decade is the natural log of the average annual industrial roundwood production numbers for each country the the years 1990 through 1999, and 2000 through 2009. These and all other data used in this paper are described in Appendix A.

Explanatory variables for land area and population density, as well as political stability are taken from the World Bank World Development Indicators, and Worldwide Governance Indicators project respectively. The political stability measure is one of six measures tracked as part of the World Governance Indicators project at the World Bank. It "measures the perceptions of the likelihood or political instability and/or politically motivated violence".[17] The values for each year are averaged for each decade for this analysis. Roads and coastline data is obtained from CIA (2000). The investment security index generated by Bohn and Deacon (2000), and settler mortality[18] data utilized by Acemoglu, Johnson, and Robinson (2001), are from their respective personal websites. The investment security index data is annual and is averaged for each of the two decades for use in the regression models.

The country-level relationship between forest area change for the 1990 and 2000 decades, and standardized measures for political stability and investment security, are

[17] As quoted from `http://info.worldbank.org/governance/wgi/pv.pdf`

[18] See Appendix B of Acemoglu, Johnson, and Robinson (2000) for a complete description of the settler mortality data and sources.

shown in Figure 3.1. Observationally, all four plots demonstrate a positive correlation, indicating that higher levels of political stability and investment security are associated with forest area expansion. This relationship seems to be similar between decades and appears to fit a linear approximation, which is at least a cursory justification for the linear model proposed previously. However, there is some clustering around 0 percent change, which could indicate at least one of two things. Conditional on a static political and investment environment, countries may approach a steady state of forest cover. Or, the process by which forest cover data is determined may be imprecise and biased toward no change in certain countries. Both of these concepts could indicate processes that may not be in keeping with a strict linear interpretation and are revisited later in the paper. It should not be taken as given that the visual association between these two variables is strictly one-directional. As has been discussed, stablity and security may lead to benign forest area outcomes, at the same time the presence of conflict and instability can also result in a reduction of malign forest outcomes (Burgess, Miguel, and Stanton, 2015).

The relationship of average roundwood production with political stability and investment security, as shown in Figure 3.2, appears less distinct than that for forest area change but generally reveals a similar positive correlation. Higher levels of stability and security are associated with greater roundwood output and the relationship can also be interpreted by a linear approximation. There seems to be no particular difference in the correlation between the two decades, indicating a degree of constancy in the system between decades. However, as with forest area change mentioned above, it would be a challenge to justify that trend lines through these points represents a functional relationship between roundwood production and the two institutional robustness measures. For this I rely on the IV approach that an instrument is exogenous and only influences the regressor of interest.

I present the 1st stage relationship between political stability and investment security with settler mortality in Figure 3.3. There appear to be reasonable downward linear trends with each of these measures - higher rates of settler mortality in the previous two centuries are negatively correlated with stability and security for the previous two decade cross sections. Notably the political stability variable seems to diverge, indicating a particular heteroskedastic relationship where the variability of political stability increases as settle mortality increases. Given political stability and investment security tend not to have changed significantly between the two decades the graphs appear very similar. In each of the four graphs there are two data points that stand out from the downward trend - both The Gambia and Mali have among the highest rates of settler mortality yet their stability and security both hover around zero.[19] As potential outliers the impact of these two countries is assessed in robustness checks that follow.

3.4 Model results

3.4.1 Forest area change

Ordinary least squares

Ordinary least squares estimates for the effect of investment security and political stability on forest area change for the whole world and the settler subsample are presented in Table 3.2. The top panel of Table 3.2 contains the results for the whole world, the bottom panel results are the for the settler subsample. Columns (1) through (3) contain the cross section results when the change in forest area between the year 1990 and 2000 is used as the dependent variable, while (4) through (6) contain the cross section results for the change in forest area between the year 2000 and 2010. Columns (1) and (4) present results when the z-score of investment security is used as the regressor of interest. Similarly, the regression results in columns (2) and (5) use only the political stability regressor

[19]The particular reasons for these outliers is not known. Although USAID (November, 2011) characterized Mali "as a stable democracy in the midst of the troubled West African region." `http://www.usaid.gov/locations/sub-saharan_africa/countries/mali/`

of interest. Regressions in columns (3) and (6) use both measures and list the p-value for the Wald test where the null is for both coefficients equal to 0. All models use the environmental and demand factor variables discussed previously; the p-values for the Wald tests where the respective variables of these two groupings are jointly equal 0 are listed in each column.

The investment security point estimate in columns (1) and (4) for the world and settler groupings varies between 0.027 and 0.053 with the standard error varying between 0.013 and 0.020. At face value a one standard deviation difference in investment security is correlated with 2.7 to 5.3% difference in forest area over 10 years. This level of investment security would be the difference between countries like Botswana, Costa Rica, or Bolivia, and Ireland, Finland, or Canada. The point estimate results for political stability in columns (2) and (5) range from 0.028 to 0.037 between the world and settler groupings, with standard errors varying between 0.010 and 0.016. A one standard deviation difference in political stability is correlated with a 2.8 to 3.7% difference in forest area over 10 years. The correlations between higher levels of investment security and political stability are each significant at the 5% level or better for each of these four sets of results. When both regressors are included in the model, as shown in columns (3) and (6), the political stability measure is never significant and generally imprecisely estimated close to 0. Investment security appears to be strongly correlated with benign outcomes for forest area, much more so than political stability. The one departure from this generalization is in the column (6) settler subsample result, here neither measure is significant and both point estimates are about the same with similar standard errors. In this regression the two measures are jointly not significantly different from 0 either.

Environmental and demand factors seem to be poorly correlated with changes in forest area. As expected the extent of initial forest area is generally negatively correlated with expanding forest coverage, but it is not often statistically significant. Point estimates for the effect of the water to land ratio and the square of the water to land area ratio are generally of the expected sign but never significant. The coast to land area ratio is always negatively correlated with higher levels of forest expansion and in some cases the point estimate is significantly different from 0. There are only three of the twelve regression results where environmental factors are jointly significant and then only at the 10% level.

Demand factor variables are never jointly significant. Individually, population density and road extent point estimates often vary in sign between regression models and are also never significantly different from 0. It seems there is little association for population and roads with changes in forest area.

Instrumental variables

The IV results for the change in forest area between 1990 and 2000, and between 2000 and 2010 are presented in Table 3.3. The instrument used is log settler mortality and the first stage results are presented in the lower panel of the table. The two regressors of interest are included separately for each of the two decades of forest cover change. These results compose columns (1) and (4), and (2) and (5), for investment security and political stability respectively. As mentioned previously, I also use additional instruments, comprised of early institutional variables, in order to include both regressors of interest in the IV models. This I do with some caution since the separation from current institutional performance may not be as distant as with settler mortality.[20] Early institutions may have a more direct role in current forest policy than settler mortality. As seen in the first stage results in columns (3) and (6), the early institution measures are jointly significant from zero for both the security and stability regressors of interest – while, in all but one of the sets of results in columns (3) and (6), settler mortality is not significant in the first stage. As might have been expected, the early institution variables overshadow the settler mortality data in the first stage.

[20]Glaeser et al. (2004) investigate whether early institution variables or early human capital are precursors of economic growth. They conclude the latter exerts a stronger influence.

Utilizing the IV approach reveals that the quasi-experimental effect of the security and stability measures on forest cover change to be much stronger than the OLS results would imply. Where each measure is utilized separately, as in columns (1), (2), (4), and (5), the point estimate is around 2 to 4 times larger, the larger difference being for the political stability estimates. The IV estimates for investment security imply that a one standard deviation change in this measure results in a 9 to 11% change in forest area over ten years. A one standard deviation change in political stability would result in a 12 to 15% change in forest area. While the political stability measure now seems a little larger in its effect on forest area change, with some trepidation I note that when both variables are included in the IV regression, investment security is statistically significant while political stability is not. The point estimate for security is also twice as large in value compared to when it is the only regressor, as in columns (1) and (4). Stability is not only not significant in columns (3) and (6), its sign is also negative. This indicates that one should interpret the regression results with caution where both measures are included.

The IV results appear to be reasonably robust to outliers. I mentioned earlier that while Mali and The Gambia have among the highest levels of settler mortality their security and stability measures are around 0, much higher than might be expected given the downward trend observed in Figure 3.3. Excluding those two countries from the IV regressions does not appreciably alter the point estimates or standard errors. On the other end of the scale, dropping the four neo-Europe colonial countries of Australia, Canada, New Zealand, and the United States does not meaningfully alter the results in Table 3.3 for investment security. The point estimates and standard errors for political stability are, however, larger, and the result in column (5) is no longer significant. Similarly, if the countries that report zero forest cover change are excluded from the IV regressions, the security index results do not change, but the political stability result in column (5) is no longer significant. Generally, the investment security index seems to be robust, but there is some question as to whether political stability is consistently reliable.

There is a body of literature that draws inference on economic outcomes from a country's legal origin (Glaeser and Shleifer, 2002; Beck, Demirguc-Kunt, and Levine, 2003). Adding dummies for countries that were British and French colonies, and for whether the country has a French legal origin does not change the IV results. However, adding continent dummies to the IV regressions for Africa and Asia alters the results somewhat. While the point estimates remain largely unchanged, political stability in column (2) and investment security in column (4) are no longer significant even though those two continent dummies are neither individually or jointly significant. For those two regressions the continent dummies appear to affect the first-stage, while they do not for the regressions in columns (1) and (5). Others suggest a link between economic performance and religious composition (La Porta et al., 1999; Barro and McCleary, 2003; Noland, 2005). However, adding variables for the percentages of Catholic and Muslim adherents does not change the forest area change IV results.[21]

I further test the robustness of the IV models by including additional controls for natural resource abundance and climate. Adding temperature and humidity controls affects the significance of models that utilize political stability but not investment security. Similar to the continent dummies, these controls are all jointly not significantly different from zero, but they affect the first stage result for political stability. When the temperature and humidity controls are included, the point estimates in the investment security models are halved when compared to the baseline results in columns (1) and (4). When controls for natural resource abundance (percent of the world's reserves of gold, iron, oil, silver, and zinc) are added, the significance of political stability in column (5) is altered, the other point estimate values and significance remains unchanged.

Glaeser et al. (2004) critiques Acemoglu, Johnson, and Robinson (2001) by arguing

[21]It is interesting to note that while Muslim density has no significant effect in any model, Catholic density is negatively and significantly correlated with forest area change in the models which utilize investment security.

that human capital, not institutions are primary drivers of economic growth. I check the robustness of the settler mortality instrument by replacing it with the log years of schooling in 1960 from Barro and Lee (2001).[22] The point estimates for security and stability in columns (4) and (5) are remarkably close when schooling is used, in columns (1) and (2) the estimates are about twice as large. The statistical inference is unchanged, each regressor remains significant.[23]

3.4.2 Roundwood production

Ordinary least squares

Ordinary least squares estimates for the cross-section regressions of average annual roundwood production between 1990-1999 and 2000-2009 on investment security and political stability are presented in Table 3.4. Columns (1) through (3) contain the results for the first decade cross section, while columns (4) through (6) are for the second decade. In columns (1) and (4) investment security is the only regressor of interest. In columns (2) and (5) only political stability is included. Both measures are used in the column (3) and (6) regressions. The table is partitioned horizontally between results for the whole world (above) and the settler subsample (below). Between the two partitions there is little difference in the point estimates for investment security in columns (1) and (4), though the standard error is a little higher in the settler subsample. This is not the case for political stability, where the the point estimates for the settler subsample in columns (2) and (5) is half the value estimated for the whole world. The political stability estimates in these columns are also not significant for the settler subsample, though they are of

[22]There are only 46 country observations common between the settler and schooling data – too few to allow for a direct comparison of the point estimates as the instruments are very weak with this number of observations. The described results using schooling as the instrument includes countries not available in the setter subsample.

[23]The robustness of the IV results using log years of schooling in 1960 as an instrument should be considered in the light of the critique of the settler mortality data by Albouy (2012) who parses out the settler data according to particular apparent inconsistencies, and refuted by Acemoglu, Johnson, and Robinson (2012).

the expected sign. At face value, a one standard deviation difference in investment security is associated with about a 45 to 50 percent difference in average annual roundwood production.

When security and stability are both included in the OLS regression for the whole world, political stability stands out as having the stronger correlation with roundwood output than investment security. In both cases in columns (3) and (6) the point estimates for stability are larger and are statistically significant, while security is smaller and not significant. Comparing the results for stability when it is entered by itself and when combined with security, the point estimates for the whole world regressions are relatively similar – this is not the case for investment security. However, these observations are reversed in the lower panel for the settler subsample results. In this case the investment security point estimates in columns (1) and (4) are strikingly similar to the point estimates in columns (3) and (6), while it is the point estimates for political stability that are much smaller when the two measures are included in the regression. The standard errors are also somewhat larger in the settler subsample results, and there-in political stability is not significant in the four regressions where it is included. In fact neither the security nor the stability measures are significant when they are included together in the settler subsample results. Additionally, they are not jointly significant either. Their combined correlation with roundwood production is ambiguous, at least for the set of countries that make up the settler subsample.

The environmental and demand cofactors are all jointly significant in the 12 different sets of results. The extent of initial forest area has the strongest correlation with average roundwood production - a larger starting endowment of forest area, unsurprisingly, is highly correlated with higher rates of average annual harvest. Other environmental cofactors are generally of the expected sign though not significant, with the exception of the coast to land area ratio where the negative point estimate is significant in several of the regressions using only the settler subsample. From which we conclude that for those countries more coastline is somewhat inhibitive to logging.

Among the demand cofactors, the log of population density is always significant - denser population is positively correlated with larger roundwood harvests. The ratio of road length to land area is also significant, but only for the whole world results. Though the point estimates are always positive, the ratio of roads is not significant when considering the settler subsample.

Instrumental variables

The instrumental variables regression results for roundwood production are presented in Table 3.5. As with previous tables, the results are separated into six numbered columns - one for each regression specification. The first three columns utilize the log average annual roundwood production for 1990 through 1999 as the dependent variable, and the last three columns are for the 2000-2009 cross-section results. The table is split between two panels where the top presents the two-stage least squares results, and the bottom contains the first stage results. Upon examination of the first stage results there is some indication that the settler mortality instrument may be a little weak regarding its influence on political stability.

The difference between the OLS and IV point estimates is quite pronounced. Considering the difference when each regressor of interest is entered separately for each decade, the point estimates for security and stability are between 4 and 6 times larger than the OLS results would imply. This suggests that the arguably causal effect of these factors is masked by endogenous factors mentioned previously. Other factors being equal, a one standard deviation increase in investment security would increase roundwood production by a factor of around 5 to 8. The point estimates for political stability suggest a larger effect, a one standard deviation change causing change in production by a factor of 10 to 19. An increase of this level would be comparable to a shift from the political stability and roundwood production observed in a country like Mali, Mozambique, or Vietnam to that of a country like New Zealand, Norway, or Sweden.

When both security and stability are included in the IV regressions neither variable is significant, as seen in columns (3) and (6). The point estimates are smaller and the standard errors are a little larger, political stability is particularly affected in this way. In this case a one standard deviation change in investment security would imply only a 2 to 4 factor change in output. Since both measures often covary it is instructive to note the the sum of the security and stability point estimates in columns (3) and (6) is about the same in magnitude as the point estimates when each measure is included singly in columns (1), (2), (4), and (5). Also, the two measures are jointly significant at the better than 1% level for both cross-sections. There may not be enough variation between the two measures to parse out the partial effects of each with respect to roundwood production. However, these joint results must be viewed with some caution since they rely on employing early institution variables as instruments and these may not be as exogenous as settler mortality. Embedded within these early institutions may be laws and practices that continue to be enforced or strongly influence current logging practices.

As in the OLS set of results the environmental and demand cofactors remain, for the most part, jointly significant. The joint significance of the environmental variables seems to be driven largely by the initial area of forest cover as it was in the OLS regressions. A larger initial endowment of forest area is correlated with larger average annual roundwood output. Similarly, the coast line to land area ratio point estimate is always negative and often, but not always, statistically significant - for a given land area, a more undulating coastline is associated with lower roundwood harvest. Regarding the demand cofactors, the roads to land area ratio does not play a measurable role, the sign on the point estimate is not consistent and it is never significant. However, population density is almost always significant and always has the expected positive sign.

I test the robustness of the roundwood IV results by removing specific observations and by adding in additional controls for continent dummies, colonial origins, religion, climate, and natural resource abundance. I also swap out the settler mortality instrument and use log years of schooling in 1960 in its place. When the neo-European countries Australia, Canada, New Zealand, and the United States are dropped the results do not change except for the effect of political stability in the 2000 decade cross section. The standard error in column (5) for political stability doubles when those 4 countries are excluded. Dropping the two possible outliers in the settler mortality data, Mali and The Gambia, does not affect the statistical inference. However, the point estimates are attenuated by 30% with a similar reduction in the standard errors.

Adding in continent indicators for Africa and Asia complicates the models. When those dummies are included, investment security and political stability are no longer significant. However, the continent dummies are not significant individually or jointly in the second stage. Their inclusion in the first stage seems to affect the inference of the regressors of interest since removing them from the first stage and including them only in the second stage restores the results as presented in Table 3.5. It could be that the investment security and political stability measures are picking up a continental effect correlated with roundwood production.

Likewise the results do not seem so robust when climate controls are included. Although humidity variables do not affect the baseline inference, different combinations of temperature controls do and do not affect the statistical significance of the security and stability variables. The climate variables have no individual or joint significance in the first or second stage, yet they affect the regressions such that the standard errors are larger when compared to the baseline results. Given the idiosyncrasy of these results they are difficult to interpret. There may be some unknown avenue through which the security and stability measures are picking up an effect on roundwood production related to temperature.

Including Catholic and Muslim population density in the IV models does not appreciably alter the results presented in Table 3.5. And although including dummies for

British and French colonies, and French legal origin attenuates the point estimates by 20-30%, the standard errors also drop - the inference remains unchanged. When adding in the set of natural resource abundance factors the significance for political stability on column (5) is impacted by a larger standard error. The other results are not affected. It is interesting to note that percent gold reserves is often significant in the second stage of these regressions. Furthermore, the the joint combination of the gold, iron, oil, silver, and zinc abundance are significantly different from zero in each model. Whether the percentage of the worlds natural reserves in a given country affects its roundwood production directly or through some other channel is not known, and could be an interesting line of future research.

Utilizing log years of schooling in 1960 in place of the settler mortality instrument results in little change for the point estimates for investment security and political stability in columns (1) and (2). In columns (4) and (5) the point estimates are about half than when settler mortality is used. However, stability and security remain statistically significant at the 5% level or better in each regression.

3.5 Discussion

There are many reasons to question whether variation in political security and investment security are causal agents of forest cover change and roundwood production. I take the approach that these measures are indicative of the social and institutional environment that influence the provision and enforcement of property rights, and as such are only proxies. Whether the two factors I use as regressors in this paper are functionally causal remains to be determined. At face value security and stability appear to benignly influence forest cover change and timber production as evidenced by the strength of the instrumental variables results. Comparing the IV to the OLS results, we observe that for all sets of dependent variables the IV point estimates are larger and statistically significant, where oft-times the OLS results using the same subsample are not. The fact there are substantial differences is not surprising. This outcome generally falls in line with the previous discussion above and in other papers that indicate correlations may be inaccurate assessments of the potentially causal influences of institutional robustness on forestland use.

The fact that differences between the OLS relationship and the causal influence are expected, and that they are realized in a particular set of IV estimates is not sufficient to pronounce the IV estimates as authoritative. The effectiveness of the instrument or instruments to uncover causal evidence depends on whether or not they are actually exogenous from the primary system, and valid in influencing the endogenous regressor. I argue that the settler mortality instrument used by Acemoglu, Johnson, and Robinson (2001) in identifying the influence institutions play in the differences of GDP between countries can also be used to identify the influence institutions play in forestland use. This assertion generally holds up to a number of robustness checks, such as the removal of potential outliers, the inclusion of colonial origin and basis of law, religious density of difference types, natural resource abundance, and climate. But there is not complete unanimity in this regard - the results for political stability vary, and the results when climate variables are included are not particularly robust.

There is a substantial amount of variability in the point estimates for political stability among the regression results. Whereas political stability is significant in each set of baseline IV results where it is entered by itself, it is never significant when investment security is included. Furthermore, the point estimates for stability in the IV models is substantially lower than when it is entered singly, and in two cases the sign is negative. We see similar variation in the results when robustness measures are applied. In the 1990s decade cross-section of forest area change, the significance of political stability is affected by including continent dummies and climate controls. In the 2000's decade cross-section it is affected by dropping the four neo-Europe countries, dropping countries with 0 forest cover change, adding climate controls, and including natural resource abundance variables. Similar patterns exist in the two roundwood production cross-sections when using political stability.

The reasons why the political stability measure does not hold up well in the regressions are the subject of speculation at this stage. It may be the case that political stability is correlated with other influences and is simply a proxy when those other influences are not included in the regression. Perhaps the political stability data I use may not be representative or internally consistent to a sufficient degree. The political stability variable is constructed by the World Bank as one of six measures that constitute their World Governance Indicators dataset. It is the distillation of almost 20 factors to measure the "perceptions of the likelihood of political instability"; it may be the subjective nature of this variable that is the cause of its lack of robustness. It could be that 'political stability', as a concept, is fundamentally nebulous and means different things in different contexts. A single measure of such may not have much identifiable impact.

When climate controls are included in the IV regressions the results are substantially different from the baseline IV results. The point estimates are halved and only the effect of investment security on forest area change remains significant. None of the regressors of interest in the roundwood production IV models are significant. However, since the climate variables are not jointly significant in any model in either the first or second stage, the reasons for this lack of robustness is not clear. Parsing out the constituent climate measures in iterative IV models reveals that different combinations of the temperature variables affect the outcome, while humidity does not. This outcome is the same if settler mortality is replaced with years of schooling in 1960 as the instrument. There may be some unknown mechanism through which temperature interacts with forestland use. An alternative way of assessing this is to utilize a different measure of institutional robustness. When a measure of corruption is used in place of investment security or political stability in IV models that include temperature, the point estimate has an expected negative sign and is significant at the 5% level or better.

3.6 Conclusion

The purpose of this paper is to estimate what might be causal effects of institutional robustness, through the measures of investment security and political stability, on forest cover change and roundwood production. Using settler mortality as an arguably exogenous source of variation in the current heterogeneity of political institutions currently observed among countries, it appears that institutions play a particularly strong role in whether forest cover expands or recedes. Although countries with stronger security and stability tend to have higher rates of afforestation, the hypothesized endogeneity occludes the estimated quasi-experimental effect which is 2 to 4 times larger than OLS estimates would suggest. The IV estimates imply that a one standard deviation change in investment security results in a 9 to 11% change in forest area over ten years. While a one standard deviation change in political stability would result in a 12 to 15% change in forest area over the same time period. A one standard deviation change would be akin to shift in investment security at the current levels of Bolivia, Botswana, or Costa Rica, to that of Ireland, Finland, or Canada.[24] [25]

Using the same approach, I find institutional strength also influences average roundwood production at the national level. The IV results indicate casual effects larger than those which simple correlations imply. A one standard deviation change in investment security would increase roundwood production by a factor of 5 to 8, ceteris paribus; while the simple correlation between the two indicates that this difference is only associated with a 40 to 50% difference in production. For political stability the effect seems to be greater – a one standard deviation increase would yield an expansion in production of 10 to 19 times. Based on this result political stability would appear to exert a larger influence on roundwood production than investment security, though joint regression results are not able to confirm this.

While these results seem to indicate a causal effect of institutional quality on forestland use they are not uniformly robust. As such further investigation is warranted. The investment security measure tends to hold up better to the removal of potential outliers, and to the inclusion of additional control variables which others have observed may be correlated with development and institutions. The IV models are generally robust to replacing settler mortality with the log of years of school in 1960, as used by Glaeser et al. (2004).

The two measures I use should be considered proxies for the social and political climate of institutions, and as such do not specify how a given country may increase its forest area or roundwood production by strengthening those factors. This fact is a short coming in this analysis – it is not prescriptive of policies that might influence benign

[24]Since the investment security index constructed by Bohn and Deacon (2000) is lucid, a governmental change from a monarchy to a non-parliamentary democracy, military dictatorship, or strong executive regime would be sufficient in most cases to cause a one standard deviation change. A change from one of the previous three regimes to a parliamentary democracy would also be sufficient.

[25]An investment stability induced 12% change in forest area for Bolivia, Botswana, and Costa Rica would increase forest area by around 8.5 million hectares, potentially sequestering a little over 1 gigatonne of carbon in above-ground vegetation. Estimates calculated using data from Dixon et al. (1994).

forestland use outcomes. It may be the case that there are no exogenous institutional treatments that a given country might apply to realize increases in forest extent and timber production. However, the results are not purely academic either. The effectiveness of international programs to reduce deforestation and forest degradation might be increased when consideration of security and stability is applied. Certainly, if nothing else this research highlights that, despite the large body of literature on deforestation, there are still avenues of future research which can illuminate causal effects of processes that policies might target and influence.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2000. "The colonial origins of comparative development: An empirical investigation." Working Paper 7771, National Bureau of Economic Research.

———. 2001. "The colonial origins of comparative development: An empirical investigation." *American Economic Review* 91 (5):1369–1401.

———. 2012. "The colonial origins of comparative development: An empirical investigation: Reply." *American Economic Review* 102 (6):3077–3110.

Albouy, David Y. 2012. "The colonial origins of comparative development: An empirical investigation: Comment." *American Economic Review* 102 (6):3059–3076.

Alston, Lee J., Gary D. Libecap, and Bernardo Mueller. 2000. "Land reform policies, the sources of violent conflict, and implications for deforestation in the Brazilian Amazon." *Journal of Environmental Economics and Management* 39 (2):162–188.

Amacher, Gregory S., Erkki Koskela, and Markku Ollikainen. 2008. "Deforestation and land use under insecure property rights." *Environment and Economic Development* 14 (3):281–303.

Araujoa, Claudio, Catherine Araujo Bonjeana, Jean-Louis Combesa, Pascale Combes Motela, and Eustaquio J. Reis. 2009. "Property rights and deforestation in the Brazilian Amazon." *Ecological Economics* 68 (8-9):2461–2468.

Arcand, Jean-Louis, Patrick Guillaumont, and Sylviane Guillaumont Jeanneney. 2008. "Deforestation and the real exchange rate." *Journal of Development Economics* 86 (2):242–262.

Barbier, Edward B. and Joanne C. Burgess. 2001. "The economics of tropical deforestation." *Journal of Economic Surveys* 15 (3):413–433.

Barro, Robert J. and Jong-Wha Lee. 2001. "International data on educational attainment: Updates and implications." *Oxford Economic Papers* 53 (3):541–563.

Barro, Robert J. and Rachel McCleary. 2003. "Religion and economic growth." Working Paper 9682, National Bureau of Economic Research.

Beck, Thorsten, Asli Demirguc-Kunt, and Ross Levine. 2003. "Law and finance: Why does legal origin matter?" *Journal of comparative economics* 31 (4):653–675.

Bhattarai, Madhusudan and Michael Hammig. 2001. "Institutions and the environmental Kuznets curve for deforestation: A crosscountry analysis for Latin America, Africa and Asia." *World development* 29 (6):995–1010.

Bohn, Henning and Robert Deacon. 2000. "Ownership risk, investment, and the use of natural resources." *American Economic Review* 90 (3):526–549.

Brander, James and M. Scott Taylor. 1998. "The simple economics of Easter Island: A Ricardo-Malthus model of renewable resource use." *American Economic Review* 88 (1):119–38.

Brender, Ernst V. 1974. "Impact of past land use on the lower piedmont forest." *Journal of Forestry* 72 (1):34–36.

Burgess, Robin, Matthew Hansen, Benjamin A. Olken, Peter Potapov, and Stefanie Sieber. 2012. "The political economy of deforestation in the tropics." *Quarterly Journal of Economics* 127 (4):1707–1754.

Burgess, Robin, Edward Miguel, and Charlotte Stanton. 2015. "War and deforestation in Sierra Leone." *Environmental Research Letters* 10 (9).

Carr, David L., Laurel Suter, and Alisson Barbieri. 2005. "Population dynamics and tropical deforestation: State of the debate and conceptual challenges." *Population and Environment* 27 (1):89–113.

Caselli, Francesco and Andrea Tesei. 2011. "Resource windfalls, political regimes, and political stability." Working Paper 17601, National Bureau of Economic Research.

Chou, Jieh Jen and Joseph Buongiorno. 1984. "Demand functions for United States forest product exports to the European Economic Community." *Wood and Fiber Science* 16 (2):158–168.

CIA. 2000. "The World Factbook, 2000." Report, Central Intelligence Agency, Virginia, USA.

Coase, Ronald Harry. 1960. "The problem of social cost." *Journal of Law and Economics* 3:1–44.

Congleton, Roger D. 1992. "Political institutions and pollution control." *Review of Economics and Statistics* 74 (3):412–421.

Conrad, Jon M. 1989. "Bioeconomics and the bowhead whale." *Journal of Political Economy* 97 (4):974–987.

Copeland, Brian R. and M. Scott Taylor. 2009. "Trade, tragedy, and the commons." *American Economic Review* 99 (3):725–749.

Costello, Christopher, Steven D. Gaines, and John Lynham. 2008. "Can catch shares prevent fisheries collapse?" *Science* 321 (5896):1678–1681.

Culas, Richard J. 2007. "Deforestation and the environmental Kuznets curve: An institutional perspective." *Ecological Economics* 61 (2):429–437.

de Oliveira, Jose Antonio Puppim. 2008. "Property rights, land conflicts and deforestation in the Eastern Amazon." *Forest Policy and Economics* 10:303–315.

Deacon, Robert T. 1989. "An empirical model of fishery dynamics." *Journal of Environmental Economics and Management* 16 (2):167–183.

———. 1994. "Deforestation and the rule of law in a cross-section of countries." *Land Economics* 70 (4):414–430.

———. 1999. "Deforestation and ownership: Evidence from historical accounts and contemporary data." *Land Economics* 75 (3):341–359.

Didia, Dal. 1997. "Democracy, political instability and tropical deforestation." *Global Environmental Change* 7 (1):63–76.

Dietz, Thomas, Elinor Ostrom, and Paul C. Stern. 2003. "The struggle to govern the commons." *Science* 302 (5652):1907–1912.

Dixon, Robert K., Sandra Brown, R. A. Houghton, A. M. Solomon, M. C. Trexier, and Jan Wisniewski. 1994. "Carbon pools and flux of global forest ecosystems." *Science* 263 (5144):185–190.

FAO. 2010. "Global forest resources assessment 2010." Report, Food and Agriculture Organization of the United Nations, Rome, Italy.

———. 2012. "State of the worlds forests 2012." Report, Food and Agriculture Organization of the United Nations, Rome, Italy.

———. 2015. "Global forest resources assessment 2015." Report, Food and Agriculture Organization of the United Nations, Rome, Italy.

Ferreira, Susana. 2004. "Deforestation, property rights, and international trade." *Land Economics* 80 (2):174–193.

Ferreira, Susana and Jeffrey Vincent. 2010. "Governance and timber harvests." *Environmental & Resource Economics* 47 (2):241–260.

Fredriksson, Per and Jakob Svensson. 2003. "Political instability, corruption and policy formation: The case of environmental policy." *Journal of Public Economics* 87 (7):1383–1405.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer. 2004. "Do institutions cause growth?" *Journal of Economic Growth* 9 (3):271–303.

Glaeser, Edward L. and Andrei Shleifer. 2002. "Legal origins." *Quarterly Journal of Economics* :1193–1229.

Godoy, Ricardo, Marc Jacobson, Vianca Aliaga Joel De Castro, Julio Romero, and Allison

Davis. 1998. “The Role of Tenure Security and Private Time Preference in Neotropical Deforestation.” *Land Economics* 74 (2):162–170.

Gordon, H. Scott. 1953. “An economic approach to the optimum utilization of fishery resources.” *Journal of the Fisheries Board of Canada* 10 (7):442–457.

Hsiang, Solomon, Kyle Meng, and Mark Crane. 2011. “Civil conflicts are associated with the global climate.” *Nature* 476:438–441.

Iliadisa, Lazaros S., Anastasios K. Papastavroub, and Panagiotis D. Lefakis. 2002. “A heuristic expert system for forest fire guidance in Greece.” *Journal of Environmental Management* 65 (3):327–336.

Jaggers, Keith and Ted Gurr. 1996. “Polity III: Regime change and political authority, 1800-1994. 2nd ICPSR version.” Data collection, Inter-university Consortium for Political and Social Research.

Johnson, Ronald N. and Gary D. Libecap. 1982. “Contracting problems and regulation: The case of the fishery.” *American Economic Review* 72 (5):1005–1022.

Johnston, Patrick. 2004. “Timber booms, state busts: The political economy of Liberian timber.” *Review of African Political Economy* 31 (101):441–456.

Kauppi, Pekka E., Jesse H. Ausubel, Jingyun Fang, Alexander S. Mather, Roger A. Sedjo, and Paul E. Waggoner. 2006. “Returning forests analyzed with the forest identity.” *Proceedings of the National Academy of Sciences* 103 (46):17574–17579.

Knack, Stephen and Philip Keefer. 1995. “Institutions and economic performance: Cross-country tests using alternative institutional indicators.” *Economics and Politics* 7 (3):207–227.

La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny. 1999. “The quality of government.” *Journal of Law, Economics, and Organization* 15 (1):222–279.

Lambin, Eric F., Bi L. Turner, Helmut J. Geist, Samuel B. Agbola, Arild Angelsen, John W. Bruce, Oliver T. Coomes, Rodolfo Dirzo, Gunther Fischer, Carl Folke, P.S. Georgej, Katherine Homewoodk, Jacques Imbernonl, Rik Leemansm, Xiubin Lin, Emilio F. Morano, Michael Mortimorep, P.S. Ramakrishnanq, John F. Richardsr, Helle Skaness, Will Steffent, Glenn D. Stoneu, Uno Svedinv, Tom A. Veldkampw, Coleen Vogelx, and Jianchu Xuy. 2001. “The causes of land-use and land-cover change: Moving beyond the myths.” *Global Environmental Change* 11 (4):261–269.

Le Billon, Philippe. 2001. “The political ecology of war: Natural resources and armed conflicts.” *Political Geography* 20 (5):561–584.

Libecap, Gary D. 1984. “The political allocation of mineral rights: A re-evaluation of Teapot Dome.” *Journal of Economic History* 44 (2):381–391.

Libecap, Gary D. and Steven N. Wiggins. 1985. "The influence of private contractual failure on regulation: The case of oil field unitization." *Journal of Political Economy* 93 (4):690–714.

Lujala, Paivi. 2010. "The spoils of nature: Armed civil conflict and rebel access to natural resources." *Journal of Peace Research* 47 (1):15–28.

Mather, Alexander S. and Coby L. Needle. 1999. "Development, democracy and forest trends." *Global Environmental Change* 9 (2):105–118.

Mendelsohn, Robert. 1994. "Property rights and tropical deforestation." *Oxford Economic Papers* 46:750–756.

Metcalf, Gilbert E. and Catherine Wolfram. 2014. "Cursed resources? Political conditions and oil market outcomes." *Energy Journal* .

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy* 112 (4):725–753.

Naidoo, Robin. 2004. "Economic growth and liquidation of natural capital: The case of forest clearance." *Land Economics* 80 (2):194–208.

Nelson, Gerald C. and Daniel Hellerstein. 1997. "Do roads cause deforestation? Using satellite images in econometric analysis of land use." *American Journal of Agricultural Economics* :80–88.

Noland, Marcus. 2005. "Religion and economic performance." *World development* 33 (8):1215–1232.

Parker, Philip M. 1997. *National cultures of the world: A statistical reference.* Greenwood Publishing Group.

Phelps, Jacob, Edward L. Webb, and Arun Agrawal. 2010. "Does REDD+ threaten to recentralize forest governance." *Science* 328 (5976):312–313.

Pokharel, Bharat, Peter Branney, Mike Nurse, and Yam Malla. 2007. "Community forestry: Conserving forests, sustaining livelihoods and strengthening democracy." *Journal of Forest and Livelihood* 6 (2):8–19.

Ross, Michael L. 2004. "What do we know about natural resources and civil war?" *Journal of Peace Research* 41 (3):337–356.

Schaefer, Milner B. 1957. "Some considerations of population dynamics and economics in relation to the management of the commercial marine fisheries." *Journal of the Fisheries Board of Canada* 14 (5):669–681.

Seneca Creek Associates, LLC and Wood Resources International, LLC. 2004. "Illegal logging and global wood markets: The competative impacts on the U.S. wood products industry." Report, prepared for the American Forest & Paper Association, Washington DC, USA.

Smith, Joyotee, Krystof Obidzinski, Subarudi Subarudi, and Iman Suramenggala. 2003. "Illegal logging, collusive corruption and fragmented governments in Kalimantan, Indonesia." *International Forestry Review* 5 (3):293–302.

Snyder, Richard. 2006. "Does lootable wealth breed disorder? A political economy of extraction framework." *Comparative Political Studies* 39 (8):943–968.

Somanathan, Eswaran. 1991. "Deforestation, property rights and incentives in Central Himalaya." *Economic and Political Weekly* 26 (4):PE37–PE46.

Southgate, Douglas, Rodrigo Sierra, and Lawrence Brown. 1991. "The causes of tropical deforestation in Ecuador: A statistical analysis." *World Development* 19 (9):1145–1151.

Stern, David I. 2014. "The environmental Kuznets curve: A primer." Working Paper 1404, Centre for Climate Economics & Policy, Crawford School of Public Policy, The Australian National University.

Taylor, M. Scott. 2011. "Buffalo hunt: International trade and the virtual extinction of the North American bison." *American Economic Review* 101 (7):3162–3195.

The World Bank. 2015. "World Development Indicators." Data collection, The World Bank, Washington DC, USA.

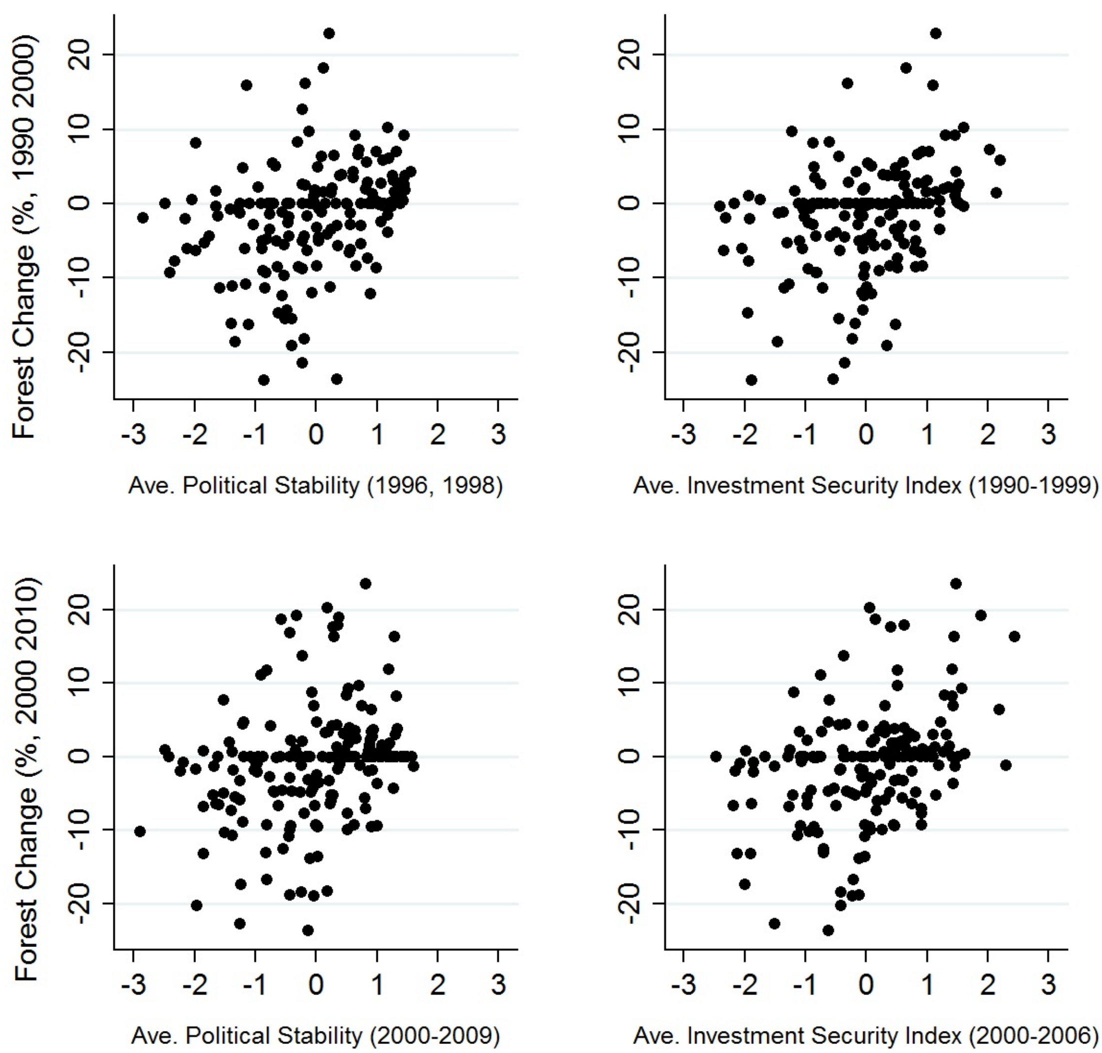
Thompson, Mary C., Manali Baruah, and Edward R. Carr. 2011. "Seeing REDD+ as a project of environmental governance." *Environmental Science and Policy* 14 (2):100–110.

Turner, James A. and Joseph Buongiorno. 2004. "Estimating price and income elasticities of demand for imports of forest products from panel data." *Scandinavian Journal of Forest Research* 19 (4):358–373.

Wright, S. Joseph, G. Arturo Sanchez-Azofeifa, Carlos Portillo-Quintero, and Diane Davies. 2007. "Poverty and corruption compromise tropical forest preserves." *Ecological Applications* 17 (5):1259–1266.

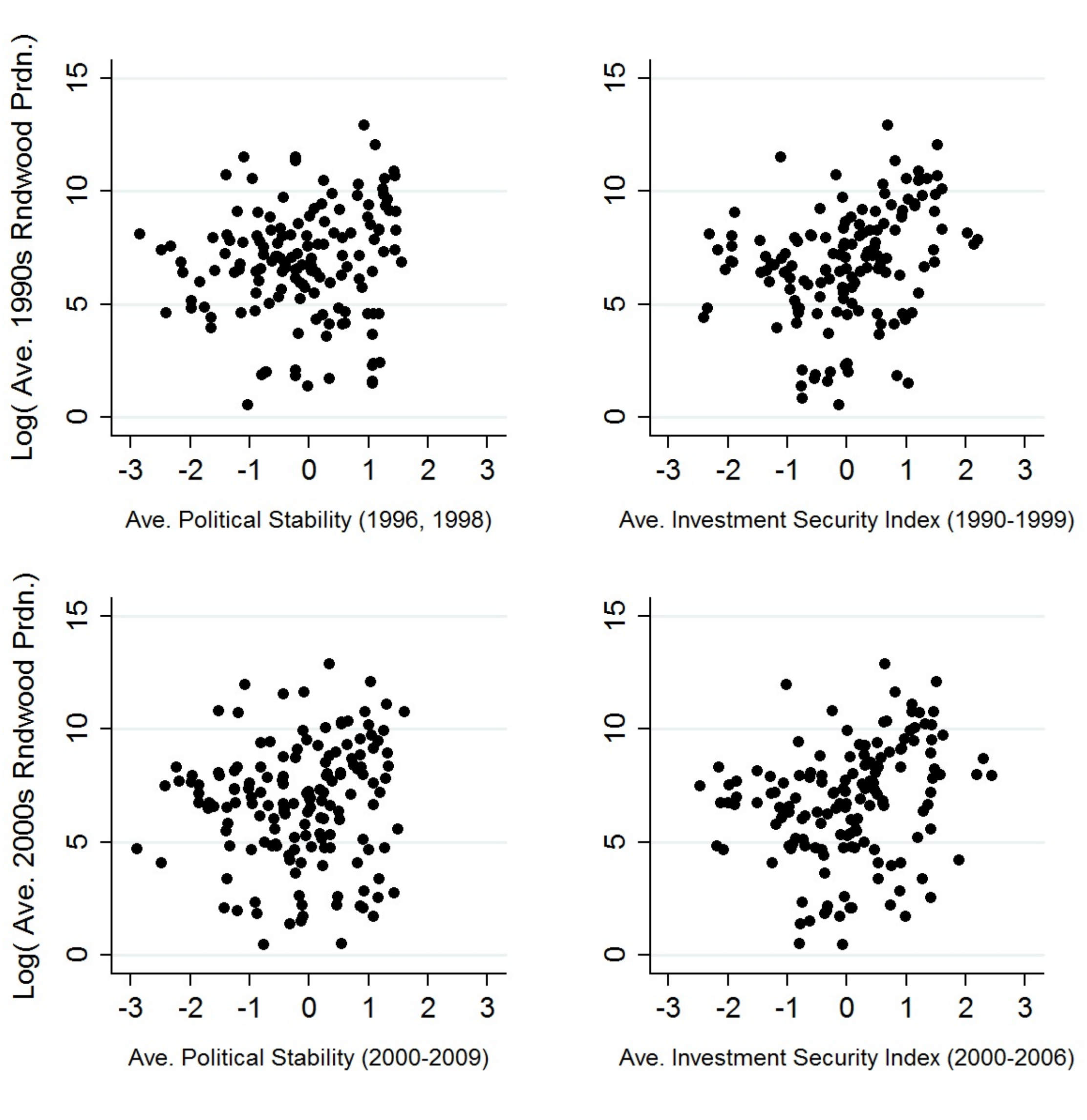
Figures and Tables

Figure 3.1: Relationship between forest area change, political stability, and investment security.



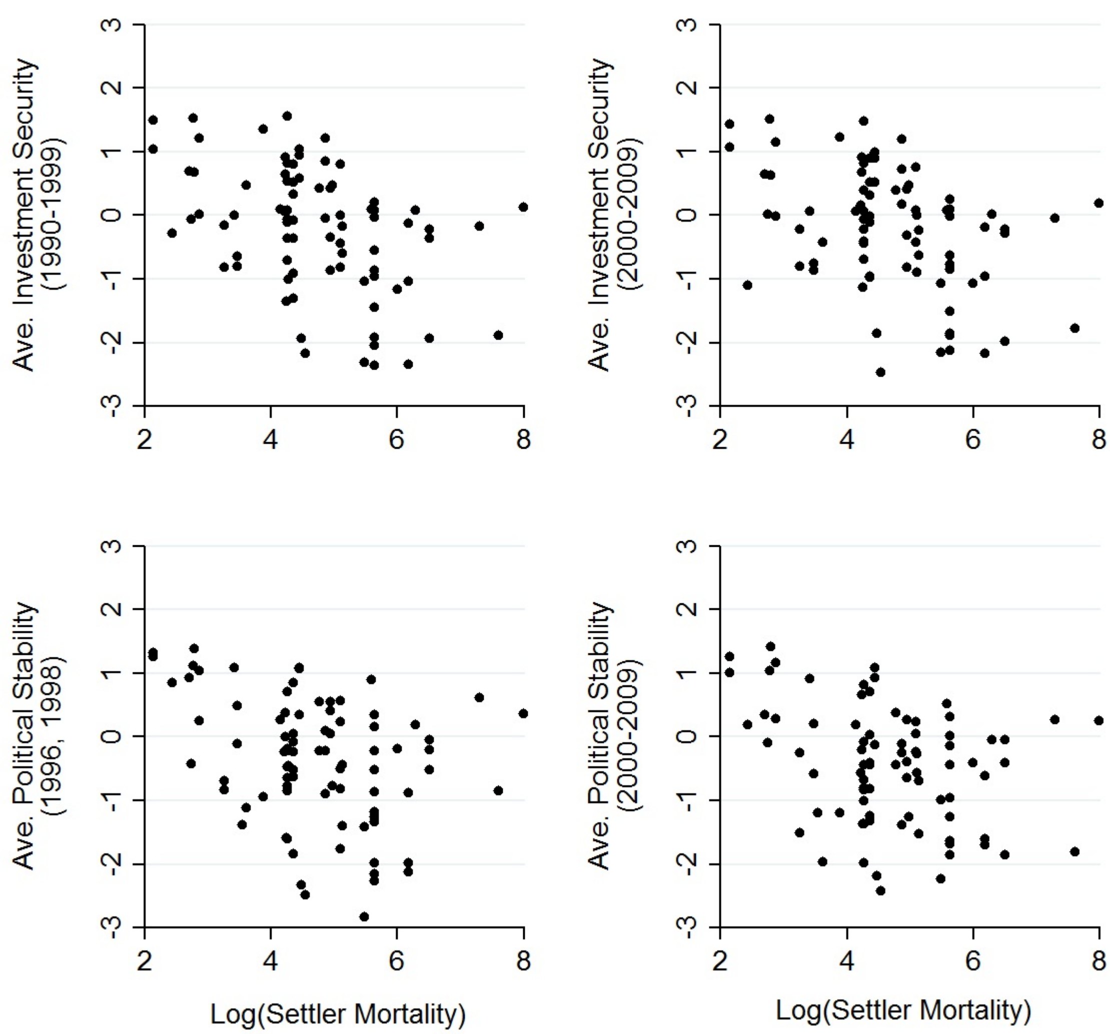
Note: Political stability and investment security expressed as z-scores. Each point is a country-level observation.

Figure 3.2: Relationship between industrial roundwood production, political stability, and investment security.



Note: Political stability and investment security expressed as z-scores. Each point is a country-level observation.

Figure 3.3: First-stage relationship between political stability, and investment security, with settler mortality.



Note: Political stability and investment security expressed as z-scores. Each point is a country-level observation.

Table 3.1: Descriptive statistics

	World			Settler Mortality Subsample		
	Mean	Std. Dev.	n	Mean	Std. Dev.	n
Change in forest area b/w 1990 and 2000 (%)	0.88	16.15	216	-3.65	12.28	86
Change in forest area b/w 2000 and 2010 (%)	-0.59	11.41	216	-3.81	10.40	86
Log ave. roundwood production (1990-1999)	6.52	2.61	161	7.02	2.21	83
Log ave. roundwood production (2000-2009)	6.43	2.81	166	7.08	2.28	83
Ave. investment security index (1990-1999)	11.55	4.01	178	10.73	3.92	84
Ave. investment security index (2000-2006)	11.67	4.10	179	10.82	3.87	84
Ave. political stability score (1996, 1998)	-0.08	1.00	182	-0.42	1.00	86
Ave. political stability score (2000, 2002-2009)	-0.02	0.98	197	-0.45	0.89	86
Log forestland area (1990)	6.74	3.30	216	8.50	2.61	86
Log forestland area (2000)	6.74	3.27	216	8.46	2.60	86
Log population density (1990)	3.93	1.54	195	3.70	1.51	86
Log population density (2000)	4.12	1.52	201	3.90	1.49	86
Ratio of coastline to land area	1.27	4.17	196	0.26	0.72	86
Ratio of fresh water resources to land area	0.53	0.58	174	0.62	0.60	85
Ratio of roads to land area	0.80	2.02	196	0.42	0.90	86
Log European settler mortality		n/a		4.62	1.29	86

Table 3.2: OLS forest area change regression results

	Forest area change, 1990 to 2000			Forest area change, 2000 to 2010		
	(1)	(2)	(3)	(4)	(5)	(6)
Whole world						
z Investment security	0.053 ***		0.054 **	0.043 ***		0.043 ***
	(0.019)		(0.021)	(0.013)		(0.016)
z Political stability		0.033 **	-0.002		0.028 ***	-0.001
		(0.014)	(0.016)		(0.010)	(0.011)
Property rights factors			[0.02]			[0.01]
Environmental factors	[0.06]	[0.13]	[0.07]	[0.10]	[0.29]	[0.12]
Demand factors	[0.49]	[0.60]	[0.53]	[0.51]	[0.38]	[0.58]
R^2	0.26	0.18	0.26	0.19	0.12	0.11
Number of obs	162	166	161	164	170	164
Settler subsample						
z Investment security	0.047 **		0.043 *	0.027 **		0.018
	(0.020)		(0.022)	(0.014)		(0.015)
z Political stability		0.037 **	0.006		0.031 **	0.015
		(0.016)	(0.016)		(0.014)	(0.014)
Property rights factors			[0.06]			[0.11]
Environmental factors	[0.80]	[0.91]	[0.81]	[0.89]	[0.70]	[0.84]
Demand factors	[0.81]	[0.98]	[0.84]	[0.53]	[0.42]	[0.60]
R^2	0.12	0.12	0.12	0.08	0.07	0.09
Number of obs	83	85	83	83	85	83

Notes: Heteroskedastic robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Wald test p-values in square brackets (joint significance, H_0: coefficients equal 0). Property rights factors consist of both the investment security and political stability variables. Environmental factors consist of the log of the beginning period forest area, the freshwater resources to land are ratio and its square, and the coast to land area ratio. Demand factors consist of the log of the population density and the roads length to land area ratio.

Table 3.3: IV forest area change regression results

	Forest area change, 1990 to 2000			Forest area change, 2000 to 2010		
	(1)	(2)	(3)	(4)	(5)	(6)
Two-stage least squares						
z Investment security	0.110 ***		0.218 **	0.090 **		0.239 **
	(0.034)		(0.093)	(0.040)		(0.116)
z Political stability		0.150 ***	-0.119		0.126 **	-0.219
		(0.058)	(0.090)		(0.064)	(0.138)
Property rights factors			[0.01]			[0.07]
Environmental factors	[0.33]	[0.30]	[0.14]	[0.34]		[0.46]
Demand factors	[0.62]	[0.40]	[0.41]	[0.99]		[0.25]
Number of obs	83	85	71	83	85	71
First stage						
F-stat investment security	25.4		10.7	19.31		12.1
Log settler mortality	[0.00]		[0.03]	[0.00]		[0.11]
Early institution factors			[0.02]			[0.01]
F-stat political stability		14.2	14.0		12.81	7.8
Log settler mortality		[0.00]	[0.53]		[0.00]	[0.35]
Early institution factors			[0.00]			[0.02]

Notes: Heteroskedastic robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Wald test p-values in square brackets (joint significance, H_0: coefficients equal 0). Property rights factors consist of both the investment security and political stability variables. Environmental factors consist of the log of the beginning period forest area, the freshwater resources to land are ratio and its square, and the coast to land area ratio. Demand factors consist of the log of the population density and the roads length to land area ratio. Early institution factors consist of the level of democracy in the first year of independence and in 1900, and the level of executive constraint in the first year of independence.

Table 3.4: OLS roundwood production regression results

	Log ave. annual roundwood production, 1990 to 1999			Log ave. annual roundwood production, 2000 to 2009		
	(1)	(2)	(3)	(4)	(5)	(6)
Whole world						
z Investment security	0.422 ***		0.203	0.365 ***		0.051
	(0.141)		(0.158)	(0.147)		(0.167)
z Political stability		0.539 ***	0.408 **		0.619 ***	0.592 ***
		(0.152)	(0.174)		(0.147)	(0.166)
Property rights factors			[0.00]			[0.00]
Environmental factors	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Demand factors	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
R^2	0.63	0.65	0.64	0.60	0.63	0.62
Number of obs	144	147	144	147	150	147
Settler subsample						
z Investment security	0.398 **		0.371	0.387 *		0.328
	(0.199)		(0.229)	(0.216)		(0.236)
z Political stability		0.292	0.043		0.337	0.099
		(0.185)	(0.201)		(0.222)	(0.235)
Property rights factors			[0.15]			[0.21]
Environmental factors	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Demand factors	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
R^2	0.65	0.66	0.66	0.66	0.65	0.65
Number of obs	80	82	80	82	82	80

Notes: Heteroskedastic robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Wald test p-values in square brackets (joint significance, H_0: coefficients equal 0). Property rights factors consist of both the investment security and political stability variables. Environmental factors consist of the log of the beginning period forest area, the freshwater resources to land are ratio and its square, and the coast to land area ratio. Demand factors consist of the log of the population density and the roads length to land area ratio.

Table 3.5: IV roundwood production regression results

	Log ave. annual roundwood production, 1990 to 1999			Log ave. annual roundwood production, 2000 to 2009		
	(1)	(2)	(3)	(4)	(5)	(6)
Two-stage least squares						
z Investment security	1.844 ***		1.163	2.260 **		1.693
	(0.600)		(1.055)	(0.899)		(1.351)
z Political stability		2.303 ***	0.858		2.957 **	0.486
		(0.837)	(1.031)		(1.378)	(1.521)
Property rights factors			[0.00]			[0.00]
Environmental factors	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Demand factors	[0.01]	[0.00]	[0.13]	[0.02]	[0.00]	[0.22]
Number of obs	80	82	70	80	82	70
First stage						
F-stat investment security	23.9		10.6	18.0		12.0
Log settler mortality	[0.00]		[0.04]	[0.00]		[0.13]
Early institution factors			[0.04]			[0.01]
F-stat political stability		12.0	14.2		9.6	7.6
Log settler mortality		[0.00]	[0.77]		[0.00]	[0.01]
Early institution factors			[0.00]			[0.02]

Notes: Heteroskedastic robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Wald test p-values in square brackets (joint significance, H_0: coefficients equal 0). Property rights factors consist of both the investment security and political stability variables. Environmental factors consist of the log of the beginning period forest area, the freshwater resources to land are ratio and its square, and the coast to land area ratio. Demand factors consist of the log of the population density and the roads length to land area ratio. Early institution factors consist of the level of democracy in the first year of independence and in 1900, and the level of executive constraint in the first year of independence.

Appendix

Data descriptions and sources

Coastline - Length of coastline, including islands. From CIA (2000).
Colonial origins - Colony of Belgian, British, Dutch, French, German, Italian, Portuguese, or Spanish origins, categorical values. From La Porta et al. (1999).
Constraint on executive - Categorical data on the level of constraint faced by the government executive in the first year of national independence. From Jaggers and Gurr (1996).
Corruption 2005 - Corruption perceptions index as reported by Transparency International. `http://www.transparency.org/research/cpi/cpi_2005`, accessed 26 May 2015.
CPI-U 1990-2009 - U.S. consumer price index, all urban consumers, city average, all items. From Bureau of Labor Statistics.
Democracy variables - Categorical data on the level of democracy in the first year of independence and in 1900. Categories comprised of measures for the competitiveness of political participation and government executive recruitment, and executive constraints. From Jaggers and Gurr (1996).
European settler mortality - Deaths per 1000 settlers as inferred and/or described on records of settler and soldier deaths. See Acemoglu, Johnson, and Robinson (2000) Appendix B for detailed information. From Acemoglu, Johnson, and Robinson (2001).
Forest area 1990, 2000, 2010 - Forested area (natural or plantation), does not include other wooded land or tree stands in agricultural production. From FAO CountrySTAT (FAO, 2010).
French legal origin - Indicator for whether the country's legal system originated from French Company Law or Commercial code. From La Porta et al. (1999)
Fresh water resources 2002 - Renewable internal freshwater from river-flows and groundwater from rainfall. From The World Bank (2015).
GDP per capita 1990-2009 - Nominal gross domestic product per capita (U.S. Dollars). From The World Bank (2015).
Humidity variables - Morning minimum, morning maximum, afternoon minimum, and afternoon maximum, in percent. From Parker (1997)
Land area 2000 - Country total area, excluding inland water bodies, continental shelf claims, and EEZs. From The World Bank (2015).
Natural resource abundance - Percent of world reserves for gold, iron, oil, silver, and zinc. From Parker (1997).
Investment security index 1990-2006 - Index measures investment as a fraction of national output attributed to political variables. From Bohn and Deacon (2000).
Political stability index 1996, 1998, 2000, 2002-2009 - Index measures the perceptions of the likelihood of political instability and/or politically motivated violence. Lower values indicate higher perceived likelihood of instability. From The World Bank (2015).
Population density, 1990, 2000 - Population per square kilometer of land area. From The World Bank (2015).
Religion compositions 1980 - Percent of population Catholic and Muslim. From La Porta et al. (1999).
Roads - Length of paved and unpaved roads, and highways. From CIA (2000).
Roundwood production 1990-2009 - Annual industrial roundwood production volume, includes sawlogs, veneer logs, and other industrial roundwood. From FAOSTAT.
Schooling - Log years of schooling of the total population over 25 years of age in 1960. From Barro and Lee (2001).
Temperature variables - Average temperature, minimum monthly high, maximum monthly high, minimum monthly low, and maximum monthly low, in centigrade. From Parker (1997).

Forestland use and GDP

Gross domestic product may or may not be an important covariate in regressions involving forest cover change and industrial roundwood production. Many question whether GDP is a driver of environmental quality, vis-a-vis an environmental Kuznets curve, with respect to deforestation. This paper takes the approach used in Deacon (1994, 1999); Bohn and Deacon (2000) and does not directly include GDP as a regressor. If GDP is used in regressions it may be endogenous to the process of deforestation and roundwood production, and if so, it is likely to be triangular in relationship to the instruments I employ and the regressors of interest. For these reasons the OLS and IV results discussed below should be viewed with a skeptical eye. In the results below I only consider GDP with respect to forest cover change.

When GDP[26] is included in the OLS regressions for forest cover change using only the settler subsample, and where I use only a single regressor of interest (either investment security or political stability), the regressors of interest remain significant at the 10% level or better. In these results GDP itself is significant only in the models where investment security is used. Investment security is significant in regressions using the whole world with GDP, but political stability is not.

Considering GDP may be endogenous I revisit the two-stage least squares regressions and instrument for investment security and GDP, or political stability and GDP, using settler mortality and early institution variables. Utilizing this approach only the investment security measure for the 1990s cross section of forest cover change is significant, while GDP is not significant in any set of results. Swapping out the settler mortality for years of schooling in 1960 as the instrument, results in the statistical significance of the investment security measure in both decade cross sections, but political stability remains non significant. Here again, GDP is also not significant in any set of results.

[26]GDP for the two decade cross-sections is formulated from annual nominal GDP per capita inflated by the U.S. CPI-U and averaged for each decade.