Essays on Environment and Development Economics

by

Sayahnika Basu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2021 by the
Graduate Supervisory Committee:

Nicolai Kuminoff, Co-Chair
Kelly Bishop, Co-Chair
Valerie Mueller
Berthold Herrendorf
Alvin Murphy

ARIZONA STATE UNIVERSITY

May 2021

ABSTRACT

There is a growing consensus that environmental hazards and changing weather patterns disproportionately affect the poor, vulnerable, minority communities. My dissertation studies the nature of risk faced by vulnerable groups of individuals, how these risks affect their labor choice, income, consumption, and migration patterns.

In Chapter 1, I study how seniors of different racial and income groups respond to information about hazardous waste sites in their neighborhood and their cleanup process. I find white seniors tend to move out at a higher rate when informed about the presence of a waste site as well as when the site is cleaned up compared to non-white seniors. This suggests that neighborhood gentrification exhibits inertia in the manifestation after the cleanup of Superfund sites. I find an assortative matching of seniors to neighborhoods based on their race and income, reinforcing findings in the environmental justice literature.

Chapter 2 documents the effect of drought on labor choices, income, and consumption of rural households in India. I find that household consumption, as well as agricultural jobs, declines in response to drought. Further, I find that these effects are mediated by job skills and land ownership. Specifically, I find that households with working members who have completed primary education account for most of the workers who exit the agricultural sector. In contrast, I find that households with farmland increase their agricultural labor share post-drought. Cultural norms, relative prices, and land market transaction costs provide potential explanations for this behavior.

Chapter 3 builds a simple model of household labor allocation based on reduced-form evidence I find in chapter 2. Simulation of the calibrated model implies that projected increases in the frequency of droughts over the next 30 years will have a net effect of a 1% to 2% reduction in agricultural labor. While small in percentage terms,

this implies that 2.5 to 5 million individuals would leave agriculture. An increase in drought will also increase the size of the manufacturing wage subsidy needed to meet the goals of ‘Make in India policy by 20%. This is driven by the need to incentivize landowners to reduce farm labor.

DEDICATION

To my most honest critic and greatest ally, my brother.

ACKNOWLEDGMENT

I am indebted to my advisors Professor Nicolai Kuminoff and Professor Kelly Bishop for their patience, understanding, encouragement, and support. I am also thankful to my committee members, Professor Valerie Mueller, Professor Berthold Herrendorf, and Professor Alvin Murphy who have also guided me in writing this dissertation. In particular, Professor Kuminoff's research guidance, regular constructive feedback, and thoughtfulness have made this research possible. I have not only benefited from the research expertise of Professor Bishop and Professor Mueller but also learned a great deal about the profession. I also thank Professor Murphy for his helpful suggestions and comments. Apart from his research mentorship, I am thankful to Professor Herrendorf for believing in me and offering words of encouragement at every opportunity.

I have received insightful comments during seminars and workshops at Arizona State University. I am grateful to the faculty at Arizona State University including Manjira Datta, Alejandro Maneli, Edward Schlee, Natalia Kovrijnykh, Gustavo Ventura, Domenico Ferraro, Hector Chade, Esteban Aucejo, Basit Zafar, Kevin Reffett, Rajnish Mehra, Alex Bick, Jonathan Ketcham, Stephie Fried, and Glenn Sheriff. I have worked closely with many other faculty of the department including Nancy Roberts, Joana Girante, Kelvin Wong, Jose Mendez, Cara McDaniel, and Richard Cox. I have learned a lot in the process. It has instilled in me the discipline and rigor needed to complete the Ph.D.

My colleagues and friends at Arizona State University have always been welcoming, supportive, and have provided valuable comments and feedback on my research. I express my gratitude to Sophie Mathes, Giorgi Tsutskiridze, Diana MacDonald, Luis Armando Fernández Intriago, Nirman Saha, Santiago García Couto, Nazim Tamkoç, Juhee Bae, Marco Mangini, Paola Ugalde, Jacob French, Zach Tobin, and Tomás

Sanguinetti. I thank the staff at ASU's Department of Economics including Laura J Talts, Callie Harriman and Tamra Eaton. They have been helpful, warm, and kind.

I have been lucky to get the opportunity to build some incredible friendships during this time. Deepasmita Bose, Parijat Basu, Sailik Sengupta, Sriloy De, Tiyasa Ray, Sumitava Ghosh, Karolina Jańczak, and Sefane Çetin have made this journey memorable. I will forever cherish the time I spent with Angelica Martinez Leyva and Ambika Athreya. I would also like to thank my childhood friends, Saveri Pal and Alolika Ray, for always being eager to listen to me from miles away. I thank Sayon Ghosh, Diya Choudhury, Shalini Rao, Debdipto Banerjee, Arani Dhar, and Sumedha Guha for being my greatest cheerleaders.

To a great extent, my family has contributed to making me the person I am today. In particular, I want to thank Tuba for never forgetting to check-in and Robo for always making that call on the way home from work. Lastly, the mental strength that my parents provided throughout my graduate school is unparalleled. I am grateful to my father who taught me the art of thinking, perseverance and discipline and my mom who taught me to be optimistic against all odds. To them, I owe everything.

TABLE OF CONTENTS

LIST OF TABLES

Table | Page

LIST OF FIGURES

Chapter 1

ENVIRONMENTAL JUSTICE FOR SENIORS: EVIDENCE FROM THE SUPERFUND PROGRAM

Note: Parts of the research described in this essay were done in collaboration with Jonathan Ketcham and Nicolai Kuminoff.

1.1 Introduction

Racial discrimination has been a concern throughout US history. Even though the US government has worked to improve the socio-economic conditions of minority populations, racial segregation remains a pertinent social issue. Economists have worked to understand why minorities, people of color, and low-income households bear a disproportionate amount of burden or risk from environmental pollution. There is evidence of both demand-side and supply-side causes. On the supply side, environmental justice literature notes that there is disproportionate siting of landfills, waste sites, and toxic emissions in communities with higher minority populations (Been (1994), Been and Gupta (1997)).[1] On the demand side, there is evidence that residential sorting contributes to market dynamics after a new waste site is identified (Depro *et al.*, 2015). People who have the means to move away and leave the neighborhood are more likely to be replaced by minorities. Therefore, sorting on public goods including environmental quality may contribute to racial segregation.

[1] The US Environmental Protection Agency defines Environmental Justice as the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies. Fair treatment means that no population is forced to bear a disproportionate share of the negative human health or environmental consequences resulting from industrial, municipal, and commercial operations or execution of federal, state, local, and tribal programs and policies.

The goal of this paper is to understand how the cleanup of hazardous waste sites across the US affected residential sorting and the Environmental Protection Agency's (EPA) success in meeting its "goals". In particular, the paper aims to quantify the migration patterns of seniors in response to information about the existence and cleanup of hazardous waste sites targeted by the US EPA's Superfund Program. About 15% of the US population is over age 65. These individuals are mostly retired and therefore more likely to choose a residential location based on public amenities.[2] I use administrative data from the US Centers for Medicare and Medicaid Services to track seniors who move out of neighborhoods hosting hazardous waste sites in order to understand how their new location choices modify their pollution exposures and how exposure changes differ among racial and income groups. This allows me to provide new insights into the consequences of place-based investments such as Superfund cleanups.

To inform the empirical research design, first I introduce a simple conceptual model for how individuals sort into neighborhoods based on local public goods and amenities including a neighborhood's demographic composition. Then I describe how I developed a novel panel data on seniors that track their exposure to EPA's Superfund sites over a 15 year period. Individual-level data on seniors provides information about their demographic characteristics, if and when they are diagnosed with common chronic conditions and their annual residential locations up to the zip+4 code. EPA's Superfund data provides information about the location of the site, date of listing to be cleaned up, various stages of cleanup, and the end date of the process. I define treatment and control areas as 2-mile and 2-4 mile radi around a Superfund site.[3]

[2] 18.8% seniors work in some part-time or full-time job.

[3] The methodology used to determine the research design is similar to Muehlenbachs *et al.* (2015).

Panel data regressions comparing the migration of different racial and income groups between treatment and control areas reveal the following. First, I find that seniors living within 2 miles of a Superfund site are 5.1% more likely to move out compared to seniors living within 2 to 4 miles of the site in response to the new information identifying the existence of a Superfund site in the neighborhood. I find similar effects when the neighborhood is informed that the site has been cleaned up. I find that race plays an important factor in determining migration patterns. White seniors are 7.1% more likely to move in response to the presence of Superfund site compared to non-white seniors. I find similar effects of smaller magnitude post-cleanup. Additionally, leveraging the geographically refined data, I find that higher-income movers who move out of existing or cleaned up Superfund site neighborhoods on an average move to neighborhoods with higher median household income, higher median house value, and higher rates of owner-occupancy. On the contrary, lower-income non-white seniors who move out of cleaned-up Superfund site neighborhoods tend to move to neighborhoods characterized by lower median household income, lower median house value, and higher rates of renter-occupancy. Focusing on how pollution levels differ across an old and new location for movers, I find that white seniors tend to move to neighborhoods with lower levels of pm2.5 and pm10 compared to non-white seniors.

Overall, this paper advances the environmental justice literature in economics by recovering the effect of information about the existence and cleanup of Superfund sites on different racial and income groups within the US population of seniors. In particular, this study contributes to the literature in several ways. First, I focus on an understudied population in the environmental economics field, individuals aged 65 and above. Second, I provide evidence using individual-level data that allows me to observe changes in neighborhood amenities that are experienced by movers. Lastly, unlike prior studies that have generally focused on a particular region, I incorporate information on all US Superfund sites that were listed on the National Priority List and cleaned up between 1999 and 2013.

1.2 Background

The Environmental Protection Agency's Superfund Program, established in 1980, is a federal program aimed to protect human health and the environment by managing the cleanup of hazardous waste sites. It is also responsible for significant local and national environmental emergencies. In the 1970s, the discovery of harmful toxic waste sites such as the 'Love Canal Emergency' in New York, 'Valley of Drums' in Kentucky and the impact it had on human health led to the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) being passed. It was responsible for mitigating environmental danger caused by the unregulated expansion of hazardous waste landfills in 1980. This later came to be known as the Superfund Program. Around the same time, the 'Environmental Justice Movement' was born out of protests in Warren County, North Carolina, against the uneven distribution of environmental threats in disadvantaged and minority communities. EPA defines Environmental Justice as the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income, with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies. The cleanup of hazardous waste sites is one way to address the aims of the Environmental Justice movement.

The cleanup of Superfund sites is a prolonged process, involving four stages. The first stage of cleanup is "proposal". Primary inspection of the site is carried out following the proposal. In order to make the cleanup process more trackable, in 1983, the EPA devised the National Priority List (NPL) based on a scoring system to determine the riskiness of each site for human health. The scoring system came to be known as the Hazardous Ranking Score (HRS). Constrained by the available funding, the first sites to get scheduled for cleanup had a cutoff HRS score of 28.5 and were put on the NPL. The second stage is called the "listing". The site is put on the NPL if its HRS score is above a cutoff. The cutoff is flexible based on the funding available at that moment. After the site is put on the NPL, the site undergoes construction to clean it. "Construction completion" denotes the third and penultimate stage. Thereafter, the site gets rid of the toxic waste and is available for productive use. "Deletion" is the last stage where the site is removed from the NPL. The duration of the entire process ranges between five to twenty years. Federal funding has been the primary source of funding to clean up Superfund sites. Annual federal appropriations have ranged between $ 1.1 to $ 2 billion over the years from 1999 to 2013. The other sources of funds for the clean-up are the potentially responsible parties who contributed to the existence of the hazardous waste site and the state in which the site is located. The state is responsible to pay 10% of the cleanup cost if the site is located within its boundaries.

The Superfund Program informs residents near a waste site about the various stages of the process. Federal Register notices and local newspaper articles are the two modes of communication with local residents. The Federal Register notice provides information about the decisions on how to clean up the waste sites. It also informs localities about the various stages of clean-up. Public comments on Federal Register notices are encouraged within an allocated period after notice submission. For every minor update, the EPA uses the Federal Register to notify the public. A copy of the Federal Register is available at the local library or depository. Public Notices are also issued in local newspapers to inform people of the listing and deletion stages of the clean-up process.[4]

1.3 Related Literature

This paper contributes to two branches of the environmental economics literature. The first branch studies the effects of the Superfund program on economic outcomes. The second branch focuses on environmental justice.

1.3.1 Superfund Program

The Superfund Program is an expensive federal program aimed at cleaning hazardous waste sites. Economists have studied the effects of such cleanups on various economic outcomes. Hazardous waste sites are infamous for dispersing harmful chemicals that worsen human health. One branch of literature focuses on how cleanup affects the health of residents near the sites. Currie *et al.* (2011) finds that mothers living within 2000 meters of a Superfund site have a 20-25% increase in congenital defect risk in their children. Additionally, children exposed to toxic waste while gestating have substantially worse cognitive and behavioral outcomes than do their unaffected siblings (Persico *et al.*, 2020).

Another branch focuses on housing market outcomes. There is heterogeneity across geographical locations in how cleanups affect local property values. For example, deletion is associated with an increase in housing values relative to the proposal in specific markets, such as northern New Jersey (Gamper-Rabindran and Timmins, 2011). Kiel and Williams (2007) finds another dimension of heterogeneity; larger sites

[4]For examples of public notices issued by the EPA in local newspapers refer to Section A.1 of the Appendix. The notices declare dates of such actions taken by the EPA along with other information.

in areas with fewer blue-collar workers are more likely to have a negative impact on property values.

1.3.2 Environmental Justice

There are two broad hypotheses for what causes the correlation between race, income, and environmental quality. The first hypothesis states that there is disproportionate siting of waste sites in neighborhoods populated with poor minority communities. The second hypothesis states that post-siting market dynamics cause people who can afford a cleaner environment to move out and minorities to move in. A considerable part of the literature has concentrated on the market dynamics hypothesis. The data used in the literature has primarily been geographically aggregated data.

One way to study nuisance-driven residential sorting is to observe whether significant demographic changes result after the siting of a hazardous waste site or other disposal facilities (Oakes *et al.*, 1996; Been and Gupta, 1997; Shaikh and Loomis, 1999; Pastor *et al.*, 2001; Morello-Frosch *et al.*, 2002). Improvement in environmental quality in California is associated with increases in population, housing density, income, and an increase in racial segregation (Banzhaf and McCormick, 2020; Banzhaf and Walsh, 2008, 2013).

Although the studies using aggregate geographic data make an important contribution to the literature, they fail to distinguish broader migration patterns from nuisance-driven migration. Observing how the demographic composition changed over some time (for example - a decade) does not differentiate nuisance-driven migration from other confounding factors resulting in migration within that period. Crowder and Downey (2010) was the first study to use individual-level data in the Environmental Justice literature that finds persistence in neighborhood choices of

Black and Latino individuals close to pollution sources. People of color are more likely to move into neighborhoods close to hazardous waste sites and are found to bear disproportionate exposure to toxins (Depro *et al.*, 2012, 2015).

Against this background, my paper contributes to the literature in the following ways. First, I focus on an understudied population, individuals aged 65 and above. Seniors form 15% of the US population and are known to have increased vulnerability to pollution exposures due to their advanced ages and morbidities (Deryugina *et al.*, 2019; Bishop *et al.*, 2018). Second, this paper makes an important contribution to the environmental justice literature by providing evidence based on individual-level data spanning over fifteen years. The administrative Medicare data from the US Centers for Medicare and Medicaid Services (CMS) tracks seniors each year including their location choices up to zip+4 level. Third, I am able to leverage the geographic resolution and panel structure of the CMS dataset to observe the new neighborhood choices of seniors who move out of neighborhoods hosting Superfund sites and what those choices imply for changes in neighborhood amenities. Lastly, this is a national study, incorporating the universe of Superfund sites that were listed on the NPL and cleaned up between 1999 and 2013.

1.4 Theoretical Framework

In this section, I adapt the model from Banzhaf and Walsh (2008) in order to provide a model of residential sorting. The model depicts agents who decide on a residential location based on its environmental quality and demographic composition. This allows me to derive testable implications of Superfund site discoveries and cleanups on sorting behavior.

1.4.1 Setup

Consider a model of two communities, each with an identical and fixed housing stock of measure 0.5. The price of housing in a community j is P^j. The population is composed of two types of individuals, $r \in \{b, w\}$. Type b is the minority, which is of measure $\beta < 0.5$ and type w of measure $1 - \beta$. In the context of the paper, type-b residents are regarded as people of color while type-w individuals are white residents.

There is heterogeneity in income, Y within each type r which is given by the continuous distribution function $F_r(Y)$. I make two assumptions about the income distribution, shown in equations (1) and (2).

1. $F_w(Y) \leq F_b(Y) \quad \forall$ Y

   This implies that the income distribution of type-w population first-order stochastically dominates the income distribution of the type-b population. On average, richer individuals are likely to locate in the community with the better public good.

2. $Y_b^l = Y_w^l < F_w^{-1}[\frac{0.5(1-2\beta)}{1-\beta}] < Y_b^h \leq Y_w^h$

   This technical condition restricts the difference in income distribution between the two groups. Y_r^l and Y_r^h are the lower and upper bounds of the income distribution of type-r individuals respectively. There is some positive measure of each group that falls in the top half of the pooled income distribution. This condition ensures that the sorting of both types of individuals is not solely due to income differences. $F_w^{-1}[\frac{0.5(1-2\beta)}{1-\beta}]$ is the boundary income of the type-w individuals when community 1 is valued more than community 2 and is only occupied by type-w individuals.[5] Given that in this extreme case, the

[5]Putting $s_w^1 = 1$ in the first equation of (1.9), the boundary income is obtained.

upper bound of the type-b income distribution is higher than the boundary income of type-w individuals, it does not restrict sorting only based on income stratification but allows for more channels through which the sorting across the two communities could work.

Each individual has a demand for a unit of housing in each of the communities and has preferences over a numeraire good, x. Each community is characterized by an exogenous public good, g_j and endogenous demographic characteristics, s_r^j. s_r^j is the share of type-r individuals in community j. The exogenous public good, g_j is the environmental quality of community j.

The utility function of a type-r individual, located in community j is given by:

$$\begin{aligned} U_r^j &= U[x, V_r^j] \\ &= U[Y - P_j, V(g_j, D(s_r^j))] \end{aligned} \tag{1.1}$$

The function $U(.)$ is continuous and increasing at a decreasing rate in both arguments. $V(.)$ is increasing in both arguments. However, it is non-monotonic in s_r^j, implying that complete segregation ($s_r^j = 1$) does not yield the highest utility for type-r individual in location j.

Within each type-r, there is ordering of community pairs (P_1, V_r^1), (P_2, V_r^2) such that higher value of V_r^j implies higher value of P_j. This, in turn, ensures stratification by income within each type.[6] Because people can be continuously stratified with respect to income and there are two communities, some individuals of each type are indifferent between the two communities. The indifference condition between the two communities for each type provides a "boundary" income for that type. The

[6] A proof of this statement can be found in Epple and Sieg (1999).

boundary indifference incomes of the two groups of people are given by $\bar{Y_w}$ and $\bar{Y_b}$. Assume that the price of housing in community 1, $P^1 = 0$. This normalization helps in solving the model and finding the willingness to pay to live in community 2. Since the communities differ in their environmental quality and the demographic composition of the two types of people, the willingness to pay to live in community 2 might differ between the two types of people. This implies that the willingness to pay for type-b might be greater than type-w and vice-versa. The type-r which has a higher willingness to pay in community 2 sets the price of housing in community 2 and is denoted by $Bid_{\bar{Y_r}}$.

1.4.2 Equilibrium

An equilibrium is defined as an allocation of individuals across the two communities and a price level in community 2 such that no one wants to re-sort. In such an equilibrium, the following conditions have to hold:

1. **Boundary Indifference:** Since stratification exists within each group of individuals, the boundary income condition for the two types are:

$$U_w(\bar{Y_w}, V(g_1, D(s_w^1))) = U_w(\bar{Y_w} - Bid_{\bar{Y_w}}, V(g_2, D(s_w^2))) \tag{1.2}$$

$$U_b(\bar{Y_b}, V(g_1, D(s_b^1))) = U_b(\bar{Y_b} - Bid_{\bar{Y_b}}, V(g_2, D(s_b^2))) \tag{1.3}$$

2. **Housing Market Clearance:** The measure of individuals choosing each community must be 0.5 as each community is characterized by a fixed stock of measure 0.5.

3. **Demographic Composition:** Since there is a fixed identical stock of housing in both locations, feasibility requires that the sorting of individuals across the two communities is determined as follows:

$$s_w^1 = \nu(S_w^1)/0.5 \tag{1.4}$$

$$s_b^1 = \nu(S_b^1)/0.5 = 1 - s_w^1 \tag{1.5}$$

$$s_w^2 = \nu(S_w^2)/0.5 = 2(1-\beta) - s_w^1 \tag{1.6}$$

$$s_b^2 = \nu(S_b^2)/0.5 = 1 - s_w^2 = s_w^1 + 2\beta - 1 \tag{1.7}$$

where $\nu(S_r^j)$ is the measure of individuals of type r in total population choosing community j. Notice that demographic compositions can be expressed in terms of the proportion of type w individuals living in community 1 i.e. s_w^1 [7].

1.4.3 Characterizing Equilibrium Sorting Behavior

Individuals sort into the two communities given their incomes and the relative rankings of the two communities (computed using V^j for all j). After people sort themselves into two communities, one can find the share of the white individuals in community 1, s_w^1 (or analogously any other racial group in either neighborhood) and subsequently all the racial composition of both communities following equations (1.5),

[7] Measure of type-w in total population is $1-\beta$. This implies $s_w^1 + s_w^2 = \nu(S_w^1)/0.5 + \nu(S_w^2)/0.5 = 2(\nu(S_w^1) + \nu(S_w^2)) = 2(1-\beta)$, as $\nu(S_w^1)$ is the measure of type-w individuals in the total population locating in community 1.

(1.6) and (1.7). Given the value of exogenous environmental quality, the composition of type-w individuals in community 1, V_r^j can be easily computed for each type-location combination. Therefore, given the income distributions, boundary income of each type can be expressed as a function of s_w^1.

$$\bar{Y_b} = \begin{cases} F_b^{-1}[\frac{0.5*(s_w^1+2\beta-1)}{\beta}], & \text{if } V_b^1 > V_b^2. \\ F_b^{-1}[\frac{0.5*(1-s_w^1)}{\beta}], & \text{if } V_b^2 > V_b^1. \end{cases} \tag{1.8}$$

$$\bar{Y_w} = \begin{cases} F_w^{-1}[\frac{0.5(2-2*\beta-s_w^1)}{1-\beta}], & \text{if } V_w^1 > V_w^2. \\ F_w^{-1}[\frac{0.5*s_w^1}{1-\beta}], & \text{if } V_w^2 > V_w^1. \end{cases} \tag{1.9}$$

Thus any sort of re-sorting results in a change in the value of s_w^1, thereby causing the boundary incomes to change.

$Bid_{\bar{Y_r}}$ determines the willingness to pay to reside in community 2 of type-r boundary individuals. Recall that the $Bid_{\bar{Y_r}}$ can be expressed as a function of the endogenous variable s_w^1 (refer to equation (1.2) and (1.3)).

$$Bid_{\bar{Y_r}} = Bid(s_w^1, g_1, g_2) \tag{1.10}$$

For every pair of values of (g_1, g_2), there is a one-to-one mapping from s_w^1 onto each $Bid_{\bar{Y_r}}$. Thus, any change in the either value of (g_1, g_2) causes the people to re-sort as they change the 'bid' function as well as s_w^1.

1.4.4 Implications of New Information about Environmental Quality

Consider two communities with the same level of environmental quality i.e. $g_1 = g_2$. For the purpose of exposition, assume $(s_w^1, s_w^2) = (0.8, 0.2)$. Given that individuals have preferences for their own demographic group, this implies that type-w individuals value community 1 more than community 2 and the opposite is true for the type-b individuals. Thus, the type-w individuals are ready to pay less to live in community 2 compared to type-b individuals. This implies $Bid_{\bar{Y_w}} < Bid_{\bar{Y_b}}$. $Bid_{\bar{Y_b}}$ sets the price of housing in community 2 i.e. P_2.

Now suppose that one of the communities experiences an information shock about the environmental quality of the neighborhood. Individuals are informed about the improvement of the environmental quality of community 2 i.e $g_1 < g_2$. Neighborhoods differ in two aspects: environmental quality, and the share of each demographic group residing in each neighborhood. The valuation of each community by each type of individual depends on the environmental quality and composition of the demographic groups of the neighborhood. Thus, the difference in environmental quality causes individuals to change their valuation of the neighborhood. Recall that the 'bid function' depends on the environmental quality and demographic composition. Since the type-w individuals are on average richer than type-b individuals, the type-w individuals are ready to pay more for the better quality neighborhood. Thus, type-w individuals sort themselves into neighborhood 2, causing the type-b individuals to leave, thereby changing the demographics of the communities each time an individual leaves or arrives at a location. This continues until a new equilibrium is reached. Therefore, conditional on which racial and income groups leave and arrive in a community, it is possible to comment on which direction the demographic composition of the neighborhood would change.

The following hypotheses are tested in the data:

1. Given two neighborhoods, one that is informed about the degradation of the environmental quality and another that does not perceive any change in environmental quality, the individuals that leave the bad environmental quality

neighborhood are likely to be of type-w and those that move in are likely to be of type-b.

2. Given two neighborhoods, one that is informed about the improvement of the environmental quality and another that does not perceive any change in environmental quality, the individuals that leave the improved environmental quality neighborhood is likely to be of type-b and those that move into those neighborhoods are likely to be of type-w.

The next section describes the sources of the data that were used, followed by the identification strategy and basic summary statistics.

1.5 Data & Descriptive Statistics

1.5.1 Data

I use three main sources of data. The first one is the individual-level, administrative, and survey data from the Centers for Medicare and Medicaid Services (CMS). The second one is information on nationwide Superfund sites from the Environmental Protection Agency (EPA). Lastly, American Community Survey (ACS) provides the data on neighborhood characteristics.

Individual Data

The Centers for Medicare and Medicaid Services (CMS) records administrative data containing information about millions of individuals' address histories, medical claims, and demographics. This study focuses on a random sample of 735,647 Medicare beneficiaries. These data describe individuals aged 65 and above for the years 1999 through 2013. Any demise within that duration is recorded.

Each individual has a unique beneficiary ID. The dataset also records information on one's race, birth year, gender, medical history, and geographic location. I group values of 'race' into two broad categories: whites, taking the value 1 and non-whites, value 0.[8] Non-white individuals are considered minorities. The medical history of individuals in the dataset includes the name of the chronic condition and the date of the first diagnosis after that individual started receiving Medicare benefits (which typically happens at age 65). Additionally, it not only records an individual's location up to the zip+4 code but tracks the individual's location over the span of fifteen years. I use geolytics dataset to map the zip+4 into latitude-longitude coordinates. One limitation of the dataset was the absence of individual income. However, I use Medicaid eligibility as a binary measure of whether a person's income fell below the eligibility threshold.[9]

Superfund Data

The Environmental Protection Agency (EPA) provides information about hazardous waste sites across the country. Any landfill that had been recognized by the EPA as a Superfund site since the 1980s is recorded on the website. The data reports the day on which the site was proposed for clean-up, the day it was put on the National Priority List (NPL) for cleanup, and also the day it was removed from the NPL after the cleanup. Each site is identified by a unique ID. For every ID, there is a de-

[8]The values that the variable 'race' takes follow the Research Triangle Institute's convention on coding. The algorithm classifies the race of an individual into 7 categories; 0: Unknown, 1: Non-Hispanic White, 2: Black (African-American), 3: Others, 4: Asian/Pacific Islander, 5: Hispanic, and 6: American Indian/Alaska Native.

[9]Medicaid services are provided to individuals who fall below a cutoff income threshold and people with disabilities (might be under the age of 65). The benefits received by dual-eligible individuals vary across various income groups. The income groups are decided based on the federal poverty level. The Medicaid webpage provides more information on the dual eligibility criteria. I leverage the information about individual eligibility for Medicaid services from 1999 to 2013 which is available in the CMS dataset.

tailed geographic location. The location data shows the associated latitude-longitude coordinates for every site.[10]

Neighborhood data

The Census Bureau publishes American Community Survey (ACS) Block Group data. The Block Group data for the year 2012 contains block group aggregates indicating the characteristics of each block group. The characteristics include household income, house value, percentage of owner-occupied and renter-occupied properties, gross rent, percentage of seniors, percentage of different racial groups. This dataset is instrumental in understanding the characteristics of neighborhoods hosting *listed* Superfund sites, cleaned up Superfund sites as well as the neighborhoods of movers from Superfund neighborhoods.

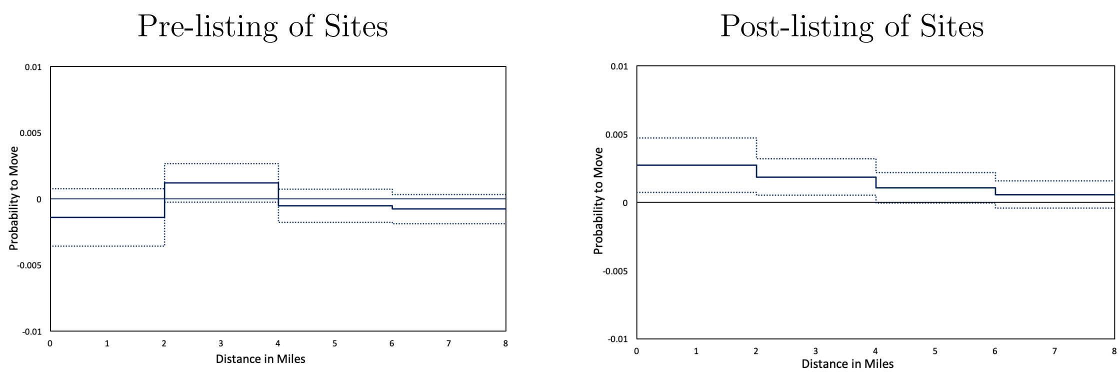
1.5.2 Research Design

I identify treatment and control groups based on prior literature. Prior studies exploring the effects of Superfund clean-up on property values have restricted their treatment group up to 3 miles from the Superfund site (Greenstone and Gallagher (2008), Gamper-Rabindran and Timmins (2011)). However, I identify the treatment group following the procedure used in Muehlenbachs *et al.* (2015). The procedure helps to identify the distance threshold within which residents respond to the infor-

[10]Figure A.3 and A.4 in Appendix Section A.2 shows the locational distribution of the sites listed on and deleted from the NPL between 1999 and 2013 respectively. The listed and deleted sites over the fifteen years (1999-2013) are mostly concentrated in the northeast region of the country, showing the uneven spatial distribution of the Superfund sites across the country. Figure A5 shows the number of sites that were listed and deleted between the years 1999 and 2013 in Appendix Section A.2. The prolonged cleanup process, taking five to twenty years coupled with the limited fund for clean up accounts for the temporal variation of the listed and deleted sites.

mation about *listed* Superfund sites and *deleted* Superfund sites. In this context, I am interested in particular in the migration response of residents.[11]

Figure 1.1: Effect of the Distance from Superfund Sites on the Probability to Move



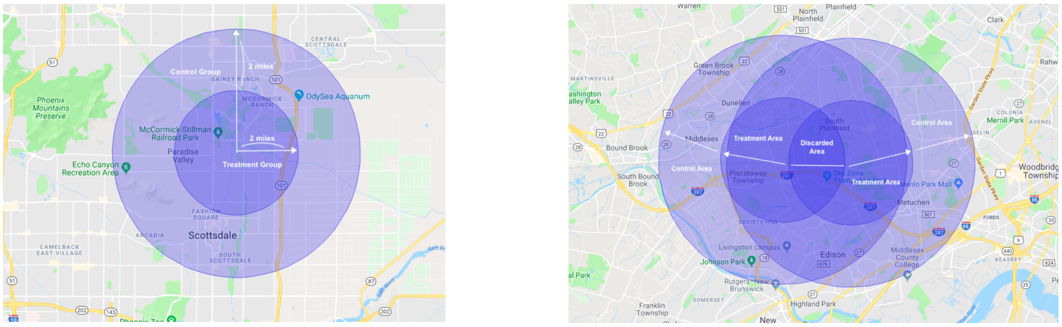
Regression controls for county times year fixed effects, block group, and individual characteristics. The dependent variable is a binary variable denoting whether individual moves or not. The coefficient on the distance dummy and their confidence intervals are plotted in the figure.

The information about the listing of Superfund sites in a vicinity affects residents' migration decisions. In Figure 1.1, I find that the probability to move is insignificant within the 2 miles from a Superfund site before the listing takes place. However, the listing of the Superfund site affects the migration decision of residents within 2 miles of the site. Based on this analysis, I assign individuals living within 2 miles of a site as being in the treatment group and those within 2-4 miles of a site as being in the control group.[12]

[11]Appendix section A.2 explains the procedure to determine the distance threshold, econometric specification, and the results.

[12]For similar analysis with respect to *deleted* Superfund sites, please refer to Appendix Section A.2

Figure 1.2: Assignment of Treatment and Control Groups



For the figure on the left, the inner 2 miles is the treatment area and the outer two miles is the control area. For the figure on the right, due to proximity of two Superfund sites, the double treated area is discarded.

The treatment and control areas form a donut-shaped structure as depicted in Figure 1.2. In general, residents located within 2 miles of a Superfund site are assigned treatment and those located within 2 to 4 miles are assigned control. However, in certain parts of the country, such as the northeast, some sites are located close to one another. In such cases, I discard areas that are within 2 miles of multiple sites.

1.5.3 Summary Statistics

Neighborhood Demographics

In this section, neighbourhood characteristics are tabulated for the control and treated neighbourhood for the *deleted* and *listed* sites.[13] In the study period i.e. 1999 to 2013, there were 310 sites scheduled for a clean-up and hence listed on the National Priority List (NPL); and there were 190 sites that were cleaned up and consequently deleted from the NPL. However, the number of sites used in this study consisted of 294 listed and 184 deleted sites. The other 22 sites were not used in my analysis

[13] The ‘deleted’ sites refer to those sites that were cleaned up and hence deleted from the NPL whereas the ‘listed’ sites mean that these sites are listed on the NPL for cleaning up.

because none of the seniors in the CMS dataset could be traced within the 6-mile radius of the dropped sites.

Table 1.1: Neighborhood Demographics of Deleted and Listed Sites

	Listed Sites		**Deleted Sites**	
Neighborhood Characteristics	Treatment 0-2 miles	Control 2-4 miles	Treatment 0-2 miles	Control 2-4 miles
# Sites	294		184	
% over 65	0.15	0.15	0.16	0.18
% White	0.60	0.53	0.61	0.66
Median Household Income (in 2012 $)	60,558	62,771	63,186	72,312
Median House Value (in 2012 $)	277,368	323,578	284,575	336,365
% Renter Occupied	0.35	0.38	0.34	0.31
% Owner Occupied	0.56	0.53	0.57	0.60
Gross Rent (Median)	1024	1111	1079	1162
Year Built	1961	1959	1963	1962

The treatment area for the deleted sites consists of the 2-mile radius around the sites whereas the control area is the 2-4 mile radius surrounding them. The values are averaged over all the sites deleted between 1999 and 2013. Similarly, the values for the listed sites could be interpreted.

Table 1.1 displays the average values of the neighborhood characteristics for the treatment and control groups for all deleted and listed sites between 1999 and 2013 using the ACS block group data for 2012. The key assumption is that the block group characteristics do not vary considerably over the years. The year 2012 block group characteristics are used as an approximation for the other years in the study. For example, the numbers in the first column report averages of block group characteristics that were within the 2-mile radius of all relevant sites that were listed on the NPL between the years 1999 and 2013. The second column reports the same for the control group (2-4 mile radius) of all sites listed on NPL. Similarly, the next two columns contain the treatment and the control neighborhood characteristics for all the *deleted* sites.

The table serves two main purposes. First, it demonstrates the difference between the neighborhoods that host a listed Superfund site and the one that hosts a deleted Superfund site. The percentage of white individuals, as well as the percentage of seniors, are higher in the block groups that have a cleaned-up site compared to the block groups that would undergo the clean-up of a site in the subsequent years. The median household income and median house value are higher in the cleaned-up neighborhood compared to the one that is scheduled for clean-up. Neighborhoods experiencing improvement in environmental conditions have a higher proportion of owner-occupied properties compared to renter-occupied; the gross rents are also higher. Thus, these trends motivate the idea that Superfund clean-up tends to increase the value of the housing (Gamper-Rabindran and Timmins (2011)) and this could drive out individuals who are not able to afford housing in those areas. A similar argument could be proposed for the listed sites' neighborhoods. As neighborhoods are declared to host a Superfund site, the housing becomes attractive to some demographic groups that move into these areas.

Secondly, it establishes the difference between the treatment and control groups for both listed and deleted sites. Observe that the treatment and control areas have differences in the neighborhood characteristics. The percentage of seniors living in the control area is the same or higher compared to the treatment area. For *listed* sites, the percentage of white individuals is higher in the treatment area compared to the control. I find the opposite is true for *deleted* sites. Household income, house value, gross rent are higher in the control area compared to the treated areas for neighborhoods hosting both *listed* and *deleted* sites. In neighborhoods with *listed* sites, the percentage of renter occupied properties are higher in the control compared to the treatment whereas, in the case of neighborhoods with *deleted* sites, I find a higher share of renter occupied properties in the treatment area.

Characteristics of Residents

Table 1.2 tabulates the characteristics of seniors living near *listed* and *deleted* Superfund sites. I compare their characteristics to the national average. This allows me to understand how the residents near Superfund sites differ from the average senior.

Table 1.2: Charactertistics of Residents near Listed and Deleted Superfund Sites

	Listed Sites		Deleted Sites		
	0-2 miles Treatment	2-4 miles Control	0-2 miles Treatment	2-4 miles Control	National Average
Demographics					
% Whites	76.0	72.7	78.0	82.0	83.0
% Low Income	20.0	20.0	19.0	16.4	16.0
% Male	40.0	40.0	40.0	40.0	42.0
% Move	4.3	4.3	4.3	4.3	4.0
Mean Age	77.0	77.0	77.0	77.0	76.0
Common Diseases					
Cancer	14.2	15.6	15.2	15.4	12.0
COPD	19.6	19.6	22.3	20.0	18.0
Asthma	7.0	8.0	7.0	8.0	6.0
Chronic Kidney Disease	13.7	13.7	13.9	12.9	10.0
Dementia	13.5	13.9	14.3	13.9	12.0
# Obs	69,280	197,355	43,818	116,462	59,300,000
# Sites	294		184		-

There are more white seniors living within 2 miles of a *listed* Superfund site compared to between 2 to 4 miles of the site. However, once the sites are cleaned up, I find more white seniors to be living in the control area compared to the treatment area. One of the factors could be that the migration decision after the information about listing or deletion of Superfund site may take more than a year. I find low-income households to be living within 2 miles to *deleted* Superfund sites compared to between 2 to 4 miles of the site. I notice no difference in the percentage of low-income seniors between the treatment and control areas of *listed* Superfund site. About 40% of the seniors are males, consistent with lower mortality among females. The average

age in the study sample is 77 years. Comparing these statistics to the national average, I find that more non-white, low-income seniors are located near Superfund sites compared to the national average. Additionally, I find that the probability to move is higher among seniors near these sites.

Since this is an aging population, certain common diseases are prevalent. About 14-15% of the seniors are affected by cancer, compared to 12% for an average senior in the country. Chronic Obstructive Pulmonary Disease (COPD) rate is at least the same or higher for seniors in the treatment area compared to the control areas for *listed* and *deleted* Superfund sites which are higher than the national average. Asthma, Chronic Kidney Disease, and Dementia are higher among seniors within 4 miles of a *listed* or *deleted* Superfund site, compared to the national average. Therefore, on average, I find that seniors living near Superfund sites have higher occurrences of chronic conditions compared to the national average.

1.6 Empirical Framework and Results

1.6.1 Effect of Listing and Deletion of Superfund Sites on Migration

In this section, I lay out the empirical specification I use to test the probability of migration of different racial and income groups residing in close proximity to Superfund sites. I follow up with the results.

$$\begin{aligned} m_{ijt} &= \alpha_j + \delta_t + \beta_0 + \beta_1 w_{ijt} + \beta_2 I_{ijt} + \beta_3 trt_{ijt} + \beta_4 w_{ijt} \times I_{ijt} \\ &\quad + \beta_5 w_{ijt} \times trt_{ijt} + \beta_6 I_{ijt} \times trt_{ijt} + \beta_7 w_{ijt} \times I_{ijt} \times trt_{ijt} + \beta_8 R_{ijt} + \epsilon_{ijt} \end{aligned} \tag{1.11}$$

In specification (1.11), the dependent variable, m_{ijt} is a binary indicator variable that takes value 1 if individual i located in county j in year t moves out, zero otherwise. w_{ijt} is the indicator variable which is equal to 1 if the individual is white, zero otherwise. I_{ijt} is the indicator variable denoting whether an individual i located in county j in year t is Medicaid eligible. trt_{ijt} is the treatment variable taking value 1 if the resident is located within 2 miles of a Superfund site, zero if located within 2 to 4 miles of the site. R_{ijt} controls for the presence of chronic conditions such as COPD, Asthma, Dementia, etc. To take into account spatial and temporal variation that may affect migration decisions, I include county (α_j) and year (δ_t) fixed effects.

From the above specification, I am primarily interested in the effect of treatment on the migration probability which is given by the following expression.

$$\frac{\Delta P(m)}{\Delta trt} = \beta_3 + \beta_5 \bar{w} + \beta_6 \bar{I} + \beta_7 \bar{w} \times \bar{I} \tag{1.12}$$

Apart from the quantification of treatment effect, the other interesting exercise is to understand how the treatment effect varies across different income and racial groups.

A. Income Differential:

$$\frac{\Delta P(m)}{\Delta trt}|_{I=0} - \frac{\Delta P(m)}{\Delta trt}|_{I=1} = \begin{cases} -\beta_6 - \beta_7 \bar{w}, & \text{if w} = avg(w) \\ -\beta_6, & \text{if w} = 0 \\ -\beta_6 - \beta_7, & \text{if w} = 1 \end{cases}$$

B. Race Differential:

$$\frac{\Delta P(m)}{\Delta grp}|_{w=1} - \frac{\Delta P(m)}{\Delta grp}|_{w=0} = \begin{cases} \beta_5 + \beta_7 \bar{I}, & \text{if I} = avg(I) \\ \beta_5, & \text{if I} = 0 \\ \beta_5 + \beta_7, & \text{if I} = 1 \end{cases}$$

In part (A), I first provide the expression for the treatment effect on different income groups, which I further categorize for the average population, white and non-white seniors. In part (B), I perform a similar exercise to part (A), I find the difference in treatment effect across white and non-white seniors. I further quantify the difference in treatment across racial groups based on different income levels.

Table 1.3: Effect of Listing and Deleting Superfund Sites on Probability to Move

	Listing	Deletion
Effect of treatment	0.005**	0.006**
	(0.002)	(0.003)
Income Differential in Treatment	0.005	0.003
	(0.007)	(0.008)
non-white	-0.001	-0.004
	(0.007)	(0.012)
white	0.006	0.005
	(0.009)	(0.01)
Race Differential in Treatment	0.007**	0.004***
	(0.003)	(0.001)
high-income	0.012***	0.006
	(0.004)	(0.008)
low-income	0.006	0.003
	(0.005)	(0.013)
mean dep var	0.098	0.076
	(0.298)	(0.265)
R^2	0.16	0.18
# obs	266,635	160,280

Table 1.3 summarizes the results for *listed* and *deleted* Superfund sites. I find that the probability to move out increases upon receiving the information about listing or deleting a site for the residents within 2 miles of a site compared to within 2 to 4 miles of the site. In particular, the information about the listing of a site causes residents within 2 miles of the site to move out with a probability of 0.005 more compared to residents within 2 to 4 miles. This 0.005 increase in probability to move is equivalent to 5.1% increase in the probability to move. In the case of *deleted* site, the probability to move out for the treated individuals is 0.006 higher compared to individuals in the control area.

Additionally, the second panel of Table 1.3 quantifies the difference in treatment effects across individuals of different income groups. I do not find any significant different in the treatment effect for both *listed* and *deleted* Superfund sites. In the last panel, I quantify the difference in treatment for individuals of different racial groups. I find that white seniors respond to the information about listing or deletion of Superfund site with a higher probability to move out compared to non-white seniors. In particular, I find the effect of information about listing causes white individuals to move out with a 0.007 higher probability than non-white individuals. In the case of *deleted* Superfund sites, I find the same qualitative effect but the difference in the probability to move out between white and non-white seniors is 0.004.

A Superfund site newly listed on the NPL is a signal for the residents nearby to view the neighborhood to be of lower environmental quality. This information leads to a higher probability of white higher-income seniors moving out compared to non-white lower-income seniors. This is observable in Table 1.3. On the contrary, the information about a Superfund site being deleted from the NPL is a signal for the residents nearby to view the neighborhood to be of higher environmental quality. This could also lead to an increase in property values. While the conceptual model would lead me to expect low-income, non-white seniors to have a higher probability of moving out of these neighborhoods. I find the opposite effect. This could potentially be related to the fact that while the conceptual model abstracts from transition dynamics between equilibria the gentrification of neighborhoods that host hazardous waste sites takes considerable time, partly because of the stigma associated with these poor environmental quality neighborhoods.

1.6.2 New Neighborhood Characteristics of Movers

My findings thus far are consistent with prior evidence that the discovery of hazardous waste sites tends to cause higher-income white people to move out and be replaced by minorities (Been and Gupta, 1997). However, there is little knowledge about where the movers from the Superfund site neighborhood move to. In this section, I analyze the new neighborhood choices made by different demographic groups.

In the CMS data, the implications of neighborhood choices that each demographic group makes are observable by comparing the new and old neighborhood characteristics of the movers. Table 1.4 summarizes the key statistics. Across all income and racial groups, movers on an average move into neighborhoods with a higher fraction of seniors of their own race. Low-income individuals move into neighborhoods with lower median household income and lower house value, irrespective of their race. Low-income white seniors move to neighborhoods with higher fractions of owner-occupied properties and higher gross rent compared to low-income non-white seniors. High-income white seniors move into neighborhoods with higher median household income, higher house value, more owner-occupied properties, and higher rents compared to their old neighborhood. I find the same pattern for high-income non-whites, except they move to neighborhoods with lower median household income and a higher share of renter-occupied properties.

Table 1.4: Neighborhood Characteristics of Movers out of Listed Superfund Sites

Neighborhood	Listed	Low-Income		High-Income	
Characteristics	Sites	White	Non-white	White	Non-white
# Movers	-	176	164	733	119
# Block Groups	64	68	250	48	
% Whites	0.60	0.73	0.29	0.76	0.44
% Over 65 yrs	0.15	0.20	0.15	0.23	0.17
Median Household Income	60,558	59,454	49,544	71,145	59,304
Median House Value	277,368	256,418	272,759	326,975	284,465
% Renter Occupied	0.35	0.29	0.46	0.25	0.39
% Owner Occupied	0.56	0.58	0.42	0.63	0.50
Median Gross Rent	1024	1044	996	1170	1122
Year Built	1961	1971	1964	1975	1968

I do the same exercise for *deleted* Superfund sites in Table 1.5. The table illustrates a few facts relating to the dynamics of neighborhood characteristics to investigate assortative matching. Firstly, irrespective of race and income, seniors that move out of neighborhoods with cleaned-up Superfund sites, on average, move to neighborhoods where more seniors of their race reside. Secondly, the high-income individuals, irrespective of their race move on average to a neighborhood that is characterized by higher median household income, higher median house value, and a higher fraction of owner-occupied properties. Low-income individuals move to neighborhoods with lower median household income and house value, irrespective of their race. Low-income white seniors move to neighborhoods with more owner-occupied properties where low-income non-white seniors move to neighborhoods with more renter-occupied properties. Lastly, the gross rent is higher in the new neighborhood for the low-income white movers and is lower for low-income non-white movers.

In summary, the neighborhood choices of movers from neighborhoods with *deleted* or *listed* site imply that, on average, high-income individuals, irrespective of their

Table 1.5: Neighborhood Characteristics of Movers out of Deleted Superfund Sites

Neighborhood Characteristics	Deleted Sites	Low-Income White	Low-Income Non-white	High-Income White	High-Income Non-white
# Movers	-	97	70	476	60
# Block Groups	-	39	31	169	16
% Whites	0.61	0.68	0.32	0.75	0.58
% Over 65 yrs	0.16	0.16	0.14	0.19	0.16
Median Household Income	63,186	64,945	47,183	73,166	83,042
Median House Value	284,575	250,918	199,628	320,133	360,550
% Renter Occupied	0.34	0.27	0.43	0.25	0.27
% Owner Occupied	0.57	0.66	0.48	0.64	0.63
Median Gross Rent	1079	991	1035	1225	1486
Year Built	1963	1970	1966	1972	1976

race, move into neighborhoods characterized by higher median household income, higher median house values, and higher rates of owner-occupied housing. However, the low-income non-whites moving out of *listed* and *deleted* sites neighborhood tend to move into worse neighborhoods if we use income and house values as proxy measures for neighborhood quality.

1.6.3 Pollution Levels Across Old and New Locations

One possible way to understand how the Superfund Program addresses the goals of the Environmental Justice movement is to measure the changes in pollution experienced by movers. In this section, I focus on two common air pollution measures: pm2.5 and pm10. Fine particulates smaller than 2.5 microns in diameter (pm2.5) and coarser particulates smaller than 10 microns in diameter (pm10) are well-known to be a health risk for seniors (Bishop *et al.*, 2018). With this in mind, I use the following specification to estimate movers' changes in exposure to pollution between their old and new locations.

$$\Delta p_{ijkt} \quad = \quad \beta_0 + \beta_1 w_{ijkt} + \beta_2 I_{ijkt} + \beta_3 w_{ijkt} \times I_{ijkt} + R_{ijkt}\beta_4 + \epsilon_{ijkt} \qquad (1.13)$$

The dependent variable is the difference in pollution levels (pm2.5 and pm10) measured in $\mu g/m^3$ between old zip+4 j and new zip+4 k of emigrant i moving in year t. The rest of the variables are the same as in specification (1.11). I calculate the average effect of one's race and income on the difference in the pollution levels experienced by them.

Table 1.6: Effect of Race and Income on Difference in Pollution Levels

	Δpm2.5		Δpm10	
	Listing	Deletion	Listing	Deletion
white	-0.105	-0.41***	-0.325*	-0.52*
	(0.107)	(0.098)	(0.199)	(0.320)
low-income	0.12	0.068	0.13	0.218
	(0.092)	(0.087)	(0.127)	(0.244)
mean dep var	-0.21	-0.13	-0.38	-0.45
	(1.51)	(1.73)	(3.91)	(4.38)
R^2	0.03	0.04	0.01	0.02
# obs	6559	2811	6559	2811

Table 1.6 summarizes the results. I track the difference in pollution levels for seniors who move after the listing and deletion of Superfund sites. I find that white emigrants tend to move to cleaner neighborhoods compared to non-white emigrants. For example, white emigrants that move out of *deleted* Superfund site neighborhood tend to move to places that have lower levels of pm2.5 and pm10 compared to non-white seniors. The -0.41 ug/m3 relative reduction in pm2.5 experienced by white

movers is equivalent to a 3.4% reduction relative to the levels in their initial neighborhoods. I find similar qualitative results for seniors moving out of neighborhoods hosting *listed* Superfund sites with respect to pm10. I find that income of seniors does not play a significant role in how the level of pollutants differs across racial and income groups. These results suggest that the residential sorting could undermine the effects of hazardous waste site cleanup by nudging non-whites to move to more polluted neighborhoods.

1.7 Conclusion

This article investigated Tiebout (1956) hypothesis of voting with one's feet in the context of Superfund Clean up. Different demographic groups are found to react differently at various stages of the cleanup process. Individual migration decisions of seniors living within a 2-mile radius of Superfund sites listed on or deleted from the NPL between 1999 and 2013 are compared to migration patterns of individuals residing within a 2 to 4 miles radius of the site. I find seniors living near Superfund sites have a 0.5 percentage point higher probability to move out compared to seniors living farther away. Further, I find that white seniors respond to the existence and clean-up of Superfund sites by exiting such neighborhoods at a higher rate compared to non-white seniors.

In addition, I find that higher-income movers moving out of existing or cleaned up Superfund site neighborhoods tend to move to neighborhoods with higher median household income, higher median house value, and higher owner-occupied properties. In contrast, lower-income, non-white seniors moving out of cleaned-up Superfund site neighborhoods move to neighborhoods characterized by lower median household income, median house value, and higher renter-occupied properties. Focusing on how the new location choices differ in the pollution levels, I find that white seniors tend to move to neighborhoods with lower levels of pm2.5 and pm10 compared to non-white seniors. These findings reinforce prior conclusions from the economics literature that race-based and income-based sorting patterns can work against the US EPA's Environmental Justice objectives.

Chapter 2

DROUGHT SHOCKS AND HOUSEHOLD OCCUPATION CHOICES

2.1 Introduction

Developing countries primarily engaging in agricultural activities form one of the most vulnerable groups in the face of climate change. More than 50% of the Indian workforce is employed in agriculture, contributing to 17-18% of the total GDP (World Bank, 2015). In India, climate change manifested as drought is likely to be more frequent and severe in the future (Bisht *et al.*, 2019). Droughts are detrimental to agriculture. In recent times, bad harvest, farmers suicide, and protests are rampant in India (Carleton, 2017). However, 60% of agricultural land remains rainfall-dependent. In this context, it is important to understand the extent of vulnerability of rural households engaging in agriculture and the degree to which behavioral responses by them could help moderate the consequences of weather. One important margin is to what degree rural households can diversify their occupational choices to manage the consequences of weather changes. In this paper, I study particularly the effect of drought on household occupational diversification from agriculture, consumption risk associated with droughts, and the potential mechanisms underlying such diversification.

The goal of this paper is to recover the effect of drought on household labor allocation, to understand how household consumption levels are affected, and to explore potential mechanisms that could modify the impact of drought on household occupation choices in rural India. In particular, the mechanisms that are considered are possession of non-agricultural skills among household members, switching costs involving ownership of farmland and farm equipment, the difference in sectoral wages, and risk-sharing among household members. Firstly, household members with a certain level of education may be better equipped to find a job in the non-agriculture sector. Secondly, landowning households that invest in farm equipment could face some switching costs regarding labor allocation choices. Thirdly, the difference in sectoral wages is a reason why workers may be attracted to the non-agriculture sector. Lastly, economists have studied how rural households engage in risk-sharing activities informally (Rosenzweig and Stark, 1989; Udry, 1994; Morten, 2019). I consider the possibility that households may diversify occupations to engage in risk-sharing within the household. Apart from exploring these mechanisms, I also quantity the effect drought has on household consumption. This helps me to understand how drought affects household well-being.

I combine high-resolution climate data with a highly detailed survey of households, spanning almost twenty years. Panel regressions with household and year fixed effects show that households reduce their share of agricultural jobs by 2.9%. Households where the members possess primary education exit agriculture at a higher rate in response to drought. Landowning households increase their share of agricultural jobs by 3.14%. This result follows from the fact that there are cultural norms involving land ownership that hinder households to move easily from the agriculture sector (Fernando, 2020). I find that hindu landowning households where these cultural norms prevail tend to allocate more labor to agriculture in response to drought. I do not find any significant changes in sectoral wages in response to drought. I also find no evidence in terms of consumption changes in households where members have diversified from agriculture. Studies have found substantial misreporting of income data in developing countries (Ravallion, 2003). Consumption is a better measure to understand how the welfare of these rural households is affected. I find that consumption reduces by 7.3% following a drought over the non-monsoon term last year.

Additionally, I conduct a series of robustness checks to address identification concerns. One important issue is attrition bias. I check whether migration is correlated with drought occurrence and find that it is not. Splitting of households is common and is considered non-random (Foster and Rosenzweig, 2002; Thomas *et al.*, 2012). I consider split households as a single unit. I find they show the same qualitative response to drought. Adding region times year fixed effects to address diverse labor markets does not alter the effect of drought on labor allocation. Lack of irrigation facilities makes drought a major problem in agriculture. Dropping households with access to irrigation at baseline, I find households to exit at a faster rate.

Mahatma Gandi National Rural Employment Guarantee Act (MGNREGA) is an employment scheme, guaranteeing employment for 100 days a year to each unemployed household. It was rolled out between 2006 and 2008 across Indian districts. I find that the employment scheme did not alter household labor choices in response to drought. Additionally, I find that there are gender differences in labor choices in response to drought. I find women leaving agriculture at a higher rate compared to men within a household following a drought. This may follow from the fact that non-agricultural jobs available are traditionally women-centric. 77.4% of the total workforce in the manufacturing and textile industry is comprised of women (Shazli and Munir, 2014).

This paper contributes to the literature studying the impact of weather changes on the sectoral labor movement in many ways. First, I use a household-level panel that spans twenty years to provide evidence of sectoral labor reallocation in response to drought. Prior studies have focused on individual (rather than household) labor outcomes and/or used cross-section (rather than a panel) data. Second, I consider jobs of all household members instead of just an indicator for the head of the household. This helps me to advance knowledge on diversification within households. Third, I use a larger and more nationally representative sample than earlier studies. My household sample covers approximately 8,000 households in 18 states in India. Lastly, I focus on short-term fluctuations in weather as opposed to long-run changes. This makes it easier to identify the effects of weather patterns separately from other confounding changes in technology and institutions.

This paper also advances knowledge on the role of barriers to the sectoral labor movement in three ways. First, I provide evidence with new household-level panel data and complement earlier work on labor market frictions related to land ownership in rural India. Second, this is the first study that attempts to understand how the weather impacts of sectoral mobility are augmented by the frictions related to land ownership. Lastly, I quantify the effect of non-agricultural skills among household members in mediating the effect of weather on labor reallocation.

2.2 Related Literature

This paper contributes to the growing literature at the intersection of Development and Environmental Economics called "Envirodevonomics" (Greenstone and Jack, 2015). Rural households earn a considerable part of their income from the weather-dependent agriculture sector. The agriculture sector is particularly vulnerable to productivity shocks caused by abnormal weather. Extreme weather has led to poor agricultural output (Taraz, 2018) and even an increase in farmer suicide in India (Carleton, 2017). Negative effects of abnormal weather on agricultural productivity have been found for both positive and negative deviations from historical norms (Kochar, 1999; Ito and Kurosaki, 2009; Emerick *et al.*, 2016).

Households in rural areas can adapt to changing weather patterns in multiple ways. First, households facing credit constraints can create an informal risk-sharing community within their villages based on shared social attributes such as caste (Rosenzweig, 1988; Munshi and Rosenzweig, 2016; Ferrara, 2003; Rosenzweig and Stark, 1989; Townsend *et al.*, 1994). Second, household members can migrate to the nearest urban areas in bad years for economic opportunities (Meghir *et al.*, 2017; Morten, 2019; Banerjee and Duflo, 2007; Jessoe *et al.*, 2018; Mueller *et al.*, 2014; Bohra-Mishra *et al.*, 2014; Maystadt *et al.*, 2016). A final form of adaptation is labor market reallocation. Households can reallocate labor from more weather-dependent sectors to more weather resilient sectors (Emerick, 2018; Colmer, 2018; Kochar, 1999; Rose, 2001; Bandyopadhyay and Koufias, 2012; Skoufias *et al.*, 2017; Noack *et al.*, 2019).

Prior studies of how weather affects sectoral labor reallocation can be divided into those focusing on temperature changes and those focusing on changes in precipitation. Rising temperature and unpredictable rainfall are both detrimental to agricultural productivity (Pachauri *et al.*, 2014). Extreme temperature affects agricultural productivity (Welch *et al.*, 2010) and in turn leads to sectoral reallocation (Colmer, 2018; Liu *et al.*, 2020). Deviations in precipitation from historical norms have also been found to affect labor market participation (Rose, 2001) and household labor allocation (Emerick *et al.*, 2016; Kochar, 1999; Skoufias *et al.*, 2017; Bandyopadhyay and Koufias, 2012). All of these studies focus on labor markets in rural India except for Bandyopadhyay and Koufias (2012) which focuses on Bangladesh.

My paper contributes to the literature on the impact of weather changes on the sectoral labor movement in several ways. First, I provide evidence of sectoral labor reallocation in response to drought, using a household-level panel that spans twenty years. This differentiates my work from prior studies that focused on individual (rather than household) labor outcomes and/or used cross-section (rather than a panel) data. Second, I use jobs of all household members instead of just an indicator for the head of the household. This allows me to advance knowledge on diversification within households. Third, my household sample covers approximately 8,000 households in 18 states in India providing a larger and more nationally representative sample than earlier studies. Lastly, I focus on short-term fluctuations in weather as opposed to long-run changes. This makes it easier to identify the effects of weather patterns separately from concomitant changes in technology and institutions.

This paper also contributes to the branch of literature studying labor market frictions in developing countries. Prior literature has focused on barriers to credit, insurance, information, transportation, and frictions arising due to possession of assets including land that restricts allocative efficiency in labor markets (Blattman *et al.*, 2013; Bianchi and Bobba, 2013; Gollin and Rogerson, 2016; Liu *et al.*, 2020; Fernando, 2020). Low-income households have limited access to credit and insurance (Townsend, 2011; Banerjee and Duflo, 2005; Karlan and Zinman, 2009). Further, infrastructure including transportation is underdeveloped in developing countries, particularly in rural areas that influence labor market movements, and can be further worsened by bad weather conditions (Burgess and Donaldson, 2010; Viswanathan and Kumar, 2015; Dallmann and Millock, 2017). Lastly, even though possession of illiquid assets such as cattle and land may serve as collateral in the presence of credit constraints, the inability to easily sell these assets may constrain occupational transitions (Das *et al.*, 2013; Fernando, 2020).

In India, there are additional barriers to the sectoral labor movement related to low levels of non-agricultural skills and land ownership. There are more skilled workers in the non-agriculture sector as opposed to the agriculture sector (Herrendorf and Schoellman, 2018). This leads to natural barriers for agricultural workers to enter the non-agriculture sector. Laws regarding buying and selling of farmland are very restrictive (Deininger *et al.*, 2007a), leading to limited sales and rental markets (Deininger *et al.*, 2007b; Morris and Pandey, 2007; Skoufias, 1995). There is also a strong prevalence of patrilineal land inheritance customs (Agarwal and Bina, 1994). This may be because the land is deemed as a mark of identity in Indian rural societies (Jodhka, 2006; Sharma, 2007) and also because individuals take pride in taking care of inheritance as it is a sacred duty in Hindu culture (Bhat and Dhruvarajan, 2001). These barriers matter as Fernando (2020) finds that individuals who inherit land are significantly less likely to enter non-agricultural work.

This paper advances knowledge on the role of barriers to the sectoral labor movement in three ways. First, I provide evidence with new household-level panel data and complement earlier work on labor market frictions related to land ownership in rural India. Second, this is the first study that investigates how the weather impacts of sectoral mobility are mediated by the frictions related to land ownership. Lastly, I recover the effect of non-agricultural skills among household members in mediating the effect of weather on labor reallocation.

2.3 Data and Summary Statistics

I use two main sources of data. The first is household-level data from the India Human Development Survey (IHDS). The second is a series of gridded temperature and precipitation datasets from Willmott and Matsuura (2001).

2.3.1 Household Data

The IHDS is a nationally representative survey of urban and rural households in 33 states and 372 districts across mainland India. The first round of interviews was completed during 2004-05 and the second round during 2011-12. Part of the IHDS sample is linked to data from an earlier study conducted by researchers at the University of Maryland and the National Council of Applied Economic Research, India. The 1993-94 survey, known as the Human Development Profile of India (HDPI), consists of 33,230 rural households living in 16 states and 184 districts.[1]

I merge the two datasets and use them to construct a long panel for approximately 20 years from 1993 to 2012. This reduces the sample size to 7,999 rural households that were interviewed during all three rounds. These households remained in the same house for almost 20 years.[2] The sample is restricted to households living in rural areas because agriculture is primarily concentrated in the rural parts of the country.[3] Households that did not have any adult members (15-65 years old) were dropped from the final sample.

The earliest survey wave in 1993-94 (henceforth 'Wave 1') contains information on occupational categories but not labor hours worked. In contrast, the following two waves in 2004-05 ('Wave 2') and 2011-12 ('Wave 3') report both labor hours and occupational categories. Every job is categorized into two groups: agricultural or non-agricultural.[4] Household members may work in more than one job. Incorporating such possibilities, the key outcome variable is defined as the number of agricultural

[1] Data collection in Appendix section B.1 elaborates on the sample collection of the three waves. Table B.1 summarizes the three waves of data.

[2] For households that split and moved to a different house but remained in the same neighborhood were considered for robustness checks.

[3] A rural area in India is defined as an area with population density up to 150 per square miles and a minimum of 75% of the male working population involved in agriculture and allied activities. However, in the current context, I use code for urbanization used by IHDS.

[4] Occupational categorization differs slightly across the three waves. In Wave-1, agricultural jobs included cultivation, allied agricultural activities, agricultural wage worker, and cattle tending. Non-agricultural jobs included non-agricultural wage workers, artisan/independent work, petty shop/other small business, organized business/trade, salaried employment/pension, qualified profession/not classified, and domestic servants. In subsequent waves, the categorization was coarser. Agricultural jobs included farmworkers, agricultural wage workers, and animal tending; and non-agricultural jobs include non-agricultural wage workers, salaried workers, and business.

jobs within the household divided by the total number of household jobs, expressed in percentage terms.[5] The final sample encompasses 18 states, covering 184 districts.

2.3.2 Rainfall and Temperature Data

Terrestrial Precipitation: 1900-2017 Gridded Monthly Time Series (V 5.01) data archives precipitation measures in mm for each month from 1900 to 2017 for every 0.5-degree by 0.5-degree latitude/longitude grid node. Terrestrial Air Temperature: 1900-2017 Gridded Monthly Time Series (Version 5.01) data archives temperature measures in degree celsius for each month from 1900 to 2017 for every 0.5-degree by 0.5-degree latitude/longitude grid node. To compute the rainfall and temperature measures for a latitude-longitude node, they combine data from 20 nearby weather stations using an interpolation algorithm based on the spherical version of Shepards distance-weighting method.

I compute district-level rainfall as the monthly average of the precipitation levels of each 0.5 degrees by 0.5 degrees latitude/longitude grid cell within the boundaries of the district using the Terrestrial Precipitation: 1900-2017 Gridded Monthly Time Series (V 5.01) data archives. I convert these district-level precipitation measures into z-scores for the lagged monsoon and non-monsoon periods. Then the z-scores are re-coded to represent a drought in the following way: districts with positive z-scores are re-coded as zero (i.e., no drought) and districts with negative z-score are re-coded as the absolute value of the z-score (i.e., below average rainfall). Therefore, a higher z-score implies a more severe drought. India receives 90% of its annual rainfall within

[5] Table B.2 summarizes the panel households across three waves and the households that are unique to each wave of data, categorized into urban and rural. The number of states in the panel HH sample (18) is greater than Wave-1 (16). There has been division and renaming of administrative units (states) during that period. In Wave-1, Uttaranchal, Jharkhand, and Chattisgarh were part of Uttar Pradesh, Bihar, and Madhya Pradesh respectively. Later, by Wave-3, Uttaranchal, Jharkhand, and Chhattisgarh were separate states.

the monsoon months of June, July, August, and September. Agriculture is heavily dependent on the monsoon. I account for the possibility for multiple cropping cycles by using the yearly z-score in my primary specification.[6] Then I investigate sensitivity to instead defining the drought variable over the monsoon period only (June-Sep) and non-monsoon (Oct-May) period to explore the potentially heterogeneous effects of rainfall during the rainy and dry seasons (Mueller *et al.* (2014)). To disentangle the effect of low precipitation from temperature in a district, I control for temperature over the period for which the drought variable is defined. I include two temperature variables: minimum and maximum temperature calculated from the Terrestrial Air Temperature: 1900-2017 Gridded Monthly Time Series (Version 5.01) dataset. This allows my model to account for the evidence that variation in minimum and the maximum temperature have opposite effects on crop yields, particularly rice yields in tropical countries like China and India (Welch *et al.*, 2010). Therefore, instead of average monthly temperature, I include minimum and maximum monthly temperature (over the term for which the drought variable is defined).[7]

2.3.3 Summary Statistics

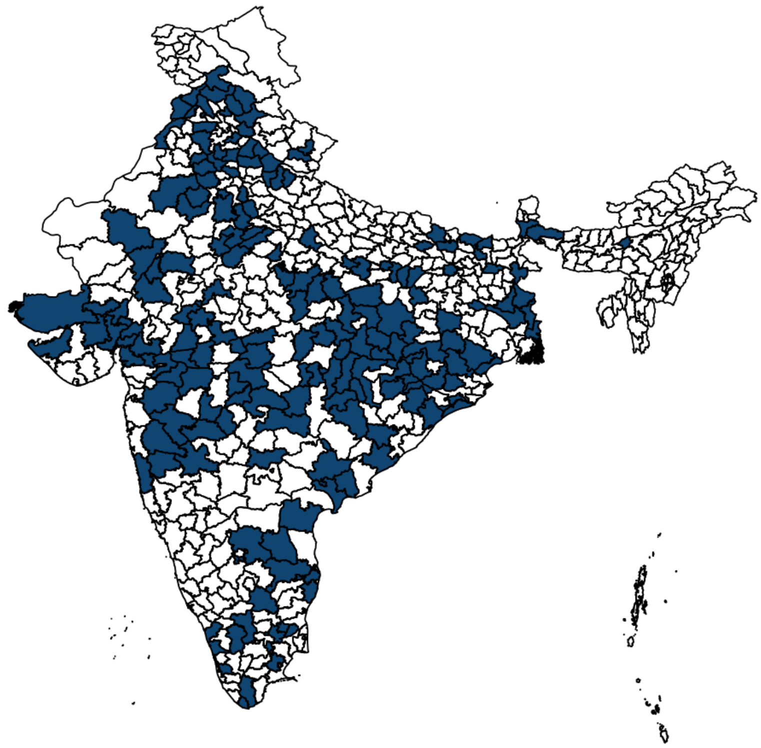
The study sample is formed by combining household data with climate data. Appendix Table B.3 shows the geographical distribution of 7999 households across 184 districts in 18 Indian states. Figure 2.1 highlights the districts (in dark blue) in which the sample households are located. Madhya Pradesh, Maharashtra are the states with the maximum number of districts, housing a sizeable part of the sample.

[6] There are three cropping cycles in India: Kharif (July-October), rabi (October-march) and summer (march-June). Kharif is the main cropping season, significantly affected by monsoon rainfall (Prasanna, 2014) Rabi season depends on the moisture retained in the soil from the monsoon rainfall.

[7] For more details on weather data, please refer to Appendix section B.1, 'Rainfall and Temperature Data'.

Assam is the state with the lowest number of households. The study sample is more or less evenly distributed across the country (refer to Figure 2.1).

Figure 2.1: Sample Districts in the Study



The blue shaded districts are the districts included in the sample. 184 in-sample districts are spread across 18 states (North: Himachal Pradesh, Punjab, Uttaranchal, Haryana, Rajasthan, Uttar Pradesh; East: Bihar, Assam, West Bengal, Jharkhand, Orissa; Central: Chhatishgarh, Madhya Pradesh; West: Gujarat, Maharashtra; South: Andhra Pradesh, Kerala, Tamil Nadu

Table 2.1 summarizes some of the key statistics for the study sample. The number of household members has decreased from 6 members to 4 members, reflected through the decrease in the number of male (3 to 2 members), female (3 to 2 members), and child (2 to 1 member) because of household splits and migration. However, aging over the last twenty years led to an increase in the number of seniors and a decrease in the number of children in the household. There is a greater number of male working members than female working members in each wave.

Table 2.1: Summary Statistics Across Three Waves

Characteristics	Wave-1	Wave-2	Wave-3
State Count	18	18	18
District Count	184	184	184
HH count	7,999	7,999	7,999
HH members	6.13	6.17	4.29
HH female members	2.94	3.01	2.02
HH male members	3.18	3.16	2.27
HH child	2.37	2.00	0.72
HH seniors	0.23	0.35	0.40
HH working age members	3.52	3.81	3.17
HH working age female members	1.68	1.88	1.48
HH working age male members	1.84	1.93	1.69
Head's education (secondary)	0.22	0.04	0.04
Head's education (college)	0.04	0.01	0.006
Land Owners	0.68	0.67	0.69
Land Cultivators	0.70	0.60	0.63
Availed Irrigation	0.40	0.36	0.39
HH job count	2.63	3.41	2.90
HH members working in Agriculture only	1.42	2.40	1.84
Female HH members working in Agriculture only	0.48	1.25	0.96
Male HH members working in Agriculture only	0.94	1.15	0.89
Working age HH members employed in Agriculture	1.70	2.63	2.26
Working age female HH members employed in Agriculture	0.49	1.24	1.06
Working age male HH members employed in Agriculture	1.20	1.39	1.20
Fraction of Agricultural Jobs	0.69	0.73	0.65
Mean HH Income (in 2012 rupees)	32,069	92,246	1,18,784
Med HH Income (in 2012 rupees)	20,576	53,805	67,200
Mean HH Agricultural Income	17,964	49,823	57,547
Med HH Agricultural Income	7,089	24,000	21,450
HH w/ MGNREGA worker(s)	0	0	0.37
HH MGNREGA workers (conditional on availing MGNREGA)	0	0	1.47
Lagged year z-score	0.79	0.49	0.29
Lagged monsoon z-score	0.70	0.41	0.28
Lagged non-monsoon z-score	0.53	0.54	0.42
Min temp last year	17.73	17.03	17.19
Max temp last year	32.30	32.92	33.26
Min monsoon temp last year	26.91	27.18	27.32
Max monsoon temp last year	31.59	31.66	31.65
Min non-monsoon temp last year	17.73	17.03	17.19
Max non-monsoon temp last year	31.45	32.33	33.16

The upper panel of the table summarizes the mean of the household characteristics across the three waves. The lower panel tabulates the lagged z-scores and temperature variables for each wave.

Education indicators for the household head show that wave-1 households had more educated heads. In wave-1, the head of the household was secondary educated in about 22% of households which fell to 4% of households by wave-3. Aging (consequently death) and splitting of households could be the reasons for falling education levels of household heads. Approximately 70% of the households own land and 60-70% of those households cultivate their farmland. Irrigation facilities are limited. Across the three waves, about 35-40% of households availed irrigation.

The number of household jobs varies between two to four jobs with at least half of the jobs in the agriculture sector. Among household members employed in agriculture only, female adults outnumber male adults except in the first wave. Working-age household members employed in agriculture are between two to three members on average as household members have multiple jobs. Among the working-age members, male members are more likely to be employed in agriculture compared to female members. In rural areas, male members are also more likely to have multiple jobs and engage in market work. Incorporating all households jobs across all members, the fraction of agricultural jobs vary between 0.65 to 0.73, implying rural households are primarily employed in the weather-dependent agriculture sector.[8]

Income variables show that there has been an increase in household income across the three waves from 32,906 rupees in wave-1 to 1,18,784 rupees in wave-3 (expressed in 2012 rupees). The mean and median income statistics vary vastly, reflecting on the fact that bigger land-owning households pull up the average income despite the heterogeneity across households. Average agriculture income in a rural household form around 48% to 56% of the total income. 37% of the households (approximately

[8]The composition of non-agricultural jobs within a typical rural household, with or without farmland is tabulated in Appendix Table B.4.

3000) availed MGNREGA. For these households, 1-2 members were employed in an MGNREGA job.

The last panel of the table summarizes the climate variables. The lagged z-score measures show a higher value for yearly and monsoon terms in wave-1 and it decreases across the later waves. This implies that more households experienced drought or that the drought that households faced in wave-1 was severe compared to the next two waves or both. The minimum and maximum monthly temperatures vary between 17-degree Celsius and 33-degree Celsius across the three waves. The monsoon term has a milder minimum temperature because of heavy precipitations.

2.4 Empirical Specifications and Results

2.4.1 Diversification from Agriculture

The baseline empirical model is a reduced-form regression of annual household allocation of agricultural jobs expressed as the percentage of agricultural jobs in household i located in district d in year t (Y_{idt}):

$$Y_{idt} = \beta_0 + \beta_1 D_{dt-1} + \mathbf{T_{dt-1}}\beta_2 + \mathbf{X_{idt}}\beta_3 + \alpha_i + \delta_t + \epsilon_{idt} \qquad (2.1)$$

The above regression includes a vector of time-varying household characteristics $\mathbf{X_{idt}}$ (number of household members, number of adult female members, number of adult male members (between the age of 14 to 65 years), two indicators for household head's education level: secondary and above, college and above) for each household i located in district d in year t. This captures the altering demographics of the household composition over almost twenty years. D_{dt-1} is the lagged drought variable which vary by district d and year *t-1*. Higher z-scores indicate a more severe drought. The spatial and temporal variations of drought occurrence are key to the identification of the effect of drought on household labor allocation. I also include additional temperature-related variables defined in Table 2.1, $\mathbf{T_{dt-1}}$. Household (α_i) and year fixed effects (δ_t) capture the effect of unobservable household characteristics and time effects that may influence job allocation within a household. I cluster the standard errors at a geographical level for which drought is defined (Bertrand *et al.*, 2004; Wooldridge, 2003; Abadie *et al.*, 2017).

Table 2.2: Effect of Drought on Percentage of Agricultural Jobs

	(1)	(2)	(3)
Lagged z-score (year)	-0.556		
	(0.846)		
Lagged z-score (monsoon)		-1.995**	
		(0.911)	
Lagged z-score (non-monsoon)			-1.073
			(0.848)
HH FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
R-squared	0.029	0.029	0.029
Observations	23,997	23,997	23,997
mean dep var	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, household controls and temperature controls for appropriate period corresponding to the drought definition are used. Standard errors in parentheses.
* $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

In equation (2.1) the effect of drought on household labor allocation is identified by spatial and temporal variation in the timing of drought experienced by individual households. The parameter of interest, β_1 is identified by the within-household variation of the timing of last year's drought shock, conditional on weather and household covariates. I expect the sign of the coefficient to be negative.

The summary of the results of the agricultural diversification is presented in Table 2.2. Each column defines drought over a different period. Column (1) defines drought as z-scores over the last year. Columns (2) and (3) defines drought into monsoon z-score and non-monsoon z-score. The coefficient on lagged z-score (year) is negative but insignificant as also the coefficient on lagged z-score (non-monsoon). But, the coefficient on the lagged z-score (monsoon) is negative and significant. Every one-unit increase in z-score (one standard deviation higher), leads to 1.995 percentage points decrease in the percent of agricultural jobs in a household. Given that the average household agricultural jobs percentage is 68.98, one standard deviation increase in z-score (implying a more severe drought) leads to a 2.89% decrease in the agricultural job share within the household.[9]

The above result aligns with previous findings. For example, weather-driven reductions in agricultural labor demand cause people to move to the manufacturing sector (Colmer, 2018). Kochar (1999) and Rose (2001) also show increased off-farm employment when hit by idiosyncratic shocks affecting agriculture. Drought is seen as one of the major reasons for poor agricultural output (Colmer, 2018) and consequently, farmer suicides (Carleton, 2017). Labor market adaptation through migration (Meghir *et al.*, 2017; Gray and Mueller, 2012) and reallocation to non-agricultural sector (Kochar, 1997, 1999; Rose, 2001; Bandyopadhyay and Koufias, 2012; Skoufias *et al.*, 2017; Noack *et al.*, 2019) have been well documented. The results of the above specification are expected to closely resonate with the findings of Skoufias *et al.* (2017). One exception is how the definition of diversification varies. Skoufias *et al.* (2017) defines occupational diversification for each household member as the probability that non-head member 'i' in household 'j' in district 'd' has the same occupation

[9] Appendix Table B.5 shows the coefficients and standard errors for the other covariates in the regression.

or employment characteristics as the household head. Their sample is restricted to households whose head works in agriculture. I do not restrict the head to be employed in agriculture. I observe how the proportion of agricultural jobs within a household is changing ex-post a drought. This could potentially imply I am capturing a more diverse population of rural households, not restricting based on the household head's occupation.

The following subsections explore the mechanisms underlying the diversification from agriculture and how consumption is affected by drought.

2.4.2 Mechanisms

Ex-post household labor allocation due to short-run weather fluctuations has been previously established. However, less is known about what household labor characteristics drive such reallocation. Changes in the composition of jobs across household members could be driven by a couple of factors: skill transferability, switching cost, sectoral wages, and risk-sharing. I explore the relative merits of each mechanism using the following reduced-form regression specification:

$$Y_{idt} = \beta_0 + \beta_1 D_{dt-1} + \beta_2 H_{id1} D_{dt-1} + \mathbf{T_{dt-1}}\beta_3 + \mathbf{X_{idt}}\beta_4 + \alpha_i + \delta_t + \epsilon_{idt} \quad (2.2)$$

Here, I introduce the variable H_{id1} interacted with the lagged drought variable. The baseline variation in any household factor across rural households (that influences labor reallocation) identifies the heterogeneous treatment effect. H_{id1} is the variable denoting heterogeneity in household i located in district d at baseline (1993-94). This variable captures one of the factors driving labor reallocation within the household. I describe each factor and the underlying mechanism in the following sub-sections.

Observe that there is no un-interacted H_{id1} term because it is time-invariant and is implicit in the household fixed effect, α_i.

The parameter of interest is β_2 that captures the additional influence of the household characteristics on the effect of drought on household labor allocation. $\beta_1 + \beta_2$ measures the effect of drought on the fraction of agricultural jobs for households characterized by the presence of the factor compared to those that do not. I test whether the term $\beta_1 + \beta_2$ is significantly different from zero using an F-statistic.

Non-Agricultural Skill

Skill transferability between agricultural sector jobs and non-agricultural sector jobs could be one of the important drivers of sectoral reallocation of household labor. Rural households primarily engaging in agriculture may face hindrance in finding a non-agricultural job because of the lack of required skill set.[10] To measure the effect of skill on the ease of labor reallocation from agricultural to non-agricultural sectors, I use education as a proxy for the skill level of household members. I use the median education level of workers in the non-agricultural sector at baseline to proxy for the skill required to work in that sector. To be more precise, I use two definitions for H_{id1} to test for skill transferability in two separate regressions. I define H_{id1} as 1 if the household head's educational attainment is higher than the median education level of the workers in the non-agricultural sector, zero otherwise. Note that I assume a higher education level of the head signifies a higher education level

[10] At this point, it is important to acknowledge that there are differences in the skill set of (low-skilled) contract workers and relatively (high-skilled) permanent workers (Colmer, 2018). However, I do not distinguish whether household members move into the non-agricultural sector as contract workers or as permanent workers. I study the reallocation of labor at an extensive margin. The intensive margin is important in the context of labor adaptation within the household and could be a topic of future research. There is a data limitation that does not allow me to distinguish between contract and permanent workers in all three waves. Information about whether wage workers are employed in contractual or permanent positions are only available for waves 2 and 3.

of other members. I also define H_{id1} as 1 if the highest educational attainment of the household's working members is higher than the median education level of the workers in the non-agricultural sector, zero otherwise. This definition allows for the possibility that the education level of the head may not reflect the education level of the working members perfectly. Labor reallocation decisions within the household ideally would be taken after incorporating non-agricultural skills for all working members. The median education level of the workers in the non-agricultural sector happens to be the completion of primary education.[11] Finding a negative coefficient for β_2 would imply that a higher education level leads to higher chances of leaving an agricultural job and landing a non-agricultural job.

Table 2.3 reports the results. The effect of monsoon drought on the percentage of household agricultural jobs is significant. One unit increase in the z-score leads to a 2 percentage point reduction in the percentage of agricultural jobs. Reinforcing the idea that monsoon term experiences 90% of the annual rainfall, deviations from the historical precipitation mean has serious consequences for the agricultural sector. Additionally, I find that the education level of working-age members matters. Households where the highest education level among working-age members is higher than primary education (the median education level of workers in the non-agricultural sector at baseline) decrease their percentage of agricultural jobs by 1.992 percentage points compared to those households where the working-age members do not have primary education. The average percentage of household agricultural jobs is 68.89%. The households where the highest education level of working members is higher than primary education reduce the percentage of their agricultural jobs by 2.9% approxi-

[11]Note that even though education variable is categorical, the ordering of the numerical value is in ascending order of education level. Hence the usage of median as a statistic for central location is reasonable.

mately.[12] Therefore, skill transferability is one of the mechanisms aiding the switch from agricultural to non-agricultural sector.

Table 2.3: Heterogeneous Treatment Effect: Skill for Non-Agricultural Sector

	(1)	(2)	(3)
Lagged z-score (year)	-0.470		
	(1.125)		
Non-ag skill X Lagged z-score (year)	-0.141		
	(0.789)		
Lagged z-score (monsoon)		-2.000*	
		(1.163)	
Non-ag skill X Lagged z-score (monsoon)		0.00745	
		(0.897)	
Lagged z-score (non-monsoon)			-1.304
			(1.202)
Non-ag skill X Lagged z-score (non-monsoon)			0.356
			(0.984)
Non-ag skill X Lagged z-score + Lagged z-score	-0.610	-1.992**	-0.948
	(0.782)	(0.899)	(0.818)
R-squared	0.0293	0.0290	0.0289
Observations	23,997	23,997	23,997
mean dep var	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98

The dependent variable is the percentage of agricultural jobs in household 'i' located in district 'd' in year 't'. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, household controls and temperature controls for the appropriate period corresponding to the drought definition are used. Non-ag skill is an indicator which takes value 1 if the highest education level of the working-age members within the household is greater than median education level of the non-agricultural sector at baseline, 0 otherwise. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Option Value/Ties to Land

High capital investment in agriculture is one of the drivers of sectoral labor reallocation (Matshe and Young, 2004; Ahituv and Kimhi, 2002). High fixed cost hinders

[12] I perform the same regression where I look at the education level of the household head, instead of the highest education among working members. I find similar results with varying magnitude (refer to Appendix Table B.6). The other coefficients for Table 2.3 are also reported in Appendix Table B.6.

moving out of agriculture. The evolution of agriculture in India is characterized by improved technology (Emerick *et al.*, 2016). Rural households own farm equipment that facilitates agricultural activities. However, these equipment are costly and hard to sell. The common farm equipment that the IDHS households own are tubewells, electric pumps, diesel pumps, bullock carts, tractors, threshers, biogas plants. Even though ownership of farm tools is an important component of fixed cost, but most households in the sample own basic farming equipment. There is not much variation to test for switching cost in terms of farm equipment ownership.[13]

Another source of cost is the ownership of farmland. Although land inheritance is potentially responsible for households to engage in agriculture, it could also limit mobility across sectors. Cultural obligations to retain land coupled with land market transaction costs influence occupational choices in rural India (Fernando, 2020). Data on ownership of land help to obtain empirical evidence for the abovementioned mechanism. I test how baseline farmland ownership influences household labor allocation decisions. I use three possible definitions of land ownership. The first is an indicator of household land ownership. In this case, the variable H_{id1} takes value 1 if household i in district d at baseline 1 owns farm-land, zero otherwise. This definition allows me to identify the effect of ownership alone. It does not indicate that farmland is used for agricultural activities. To understand how ownership alone differs from ownership of farmland used for cultivation, I use another definition. The second definition is land cultivation. H_{id1} takes value 1 if household i in district d at baseline 1 owns farmland and uses for cultivation, zero otherwise. The variation of land cultivation across households helps to identify the effect of drought on labor allocation for households where the farmland is cultivated. The third and the last

[13]There are only 5 households at baseline that do not possess any of the farm equipment. Look at Appendix Table B.7 and Table B.8 for specific and total tool count of households in wave-1.

one is the fraction of agricultural income. H_{id1} is the fraction of agricultural income in household i in district d at baseline 1. A higher fraction of baseline agricultural income would imply a greater investment in agriculture. A positive sign of β_2 would imply land attachment prohibits sectoral mobility of labor. The magnitude of β_2 is likely to be the lowest when the land ownership indicator is used, followed by when the land cultivation indicator is used and highest when the fraction of agricultural income is used.

Table 2.4: Hetereogenous Treatment Effect: Land Ownership

	(1)	(2)	(3)
Lagged z-score (year)	-6.079***		
	(1.305)		
Land ownership X Lagged z-score (year)	8.249***		
	(1.455)		
Lagged z-score (monsoon)		-8.390***	
		(1.311)	
Land ownership X Lagged z-score (monsoon)		9.744***	
		(1.559)	
Lagged z-score (non-monsoon)			-2.464
			(1.532)
Land ownership X Lagged z-score (non-monsoon)			2.012
			(1.757)
Land ownership X Lagged z-score + Lagged z-score	2.170**	1.354	-0.452
	(0.860)	(0.968)	(0.971)
R-squared	0.0371	0.0375	0.0292
Observations	23,997	23,997	23,997
mean dep var	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98

The dependent variable is the percentage of agricultural jobs in household 'i' located in district 'd' in year 't'. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. Land Ownership is an indicator which takes value 1 if household owns land at baseline, 0 otherwise. Standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Switching cost is proxied by baseline land ownership, baseline land cultivation (for those who own land), and baseline fraction of agricultural income. Table 2.4 summarizes the results for specification (2.2) where land ownership is the source of heterogeneity.[14] Yearly drought (lagged z-score (year)) significantly affects the percentage of agricultural jobs within households, after accounting for land ownership. Monsoon drought (lagged z-score (monsoon)) has a stronger effect after controlling for land ownership. Land ownership seems to be an important factor for the rural landscape in India and an important determinant for labor reallocation decisions. The average land ownership at baseline in the sample is 0.68. Additionally, landowners increase their fraction of agricultural jobs by 2.17 percentage points with a unit increase in last year's z-score compared to households who do not own land. The average fraction of agricultural jobs in the sample is 68.98%, therefore landowners increase the fraction of agricultural jobs by 3.14 % when hit by drought last year, manifested as an increase in one unit of z-score.[15] This result at the least follows from the fact that there are Hindu laws governing land ownership that restricts sectoral labor reallocation.

I also consider heterogeneity in religion. In particular, I investigate whether Hindu households tend to stick to agriculture following a drought compared to non-hindu households. I find that Hindu households owning farmland show a strong affinity towards agriculture. Table 2.5 illustrates this. I find that Hindu households in general do not show affinity towards working in agriculture. There is no evidence that they move towards agriculture in response to drought. However, Hindu households that own land tend to allocate more workers in the agricultural sector. I find landowning

[14]Appendix Table B.10 shows the results when baseline land cultivation and baseline fraction of household agricultural income is the source of heterogeneity in rural households respectively.

[15]Appendix Table B.9 summarizes the other coefficients and standard errors for the specification reported in Table 2.4.

Table 2.5: Heterogenous Treatment Effect – Hindu Indicator

	(1)	(2)	(3)	(4)	(5)	(6)
Lagged z-score (year)	-1.606			-4.407***		
	(1.472)			(1.148)		
Hindu Indicator X Lagged z-score (year)	1.232			6.580***		
	(1.684)			(1.307)		
Lagged z-score (monsoon)		-2.367			-6.055***	
		(1.725)			(1.287)	
Hindu Indicator X Lagged z-score (monsoon)		0.480			7.331***	
		(1.949)			(1.487)	
Lagged z-score (non-monsoon)			-1.859			-2.287*
			(1.732)			(1.259)
Hindu Indicator X Lagged z-score (non-monsoon)			0.965			2.046
			(1.867)			(1.433)
Hindu Indicator X Lagged z-score + Lagged z-score	-0.374	-1.887*	-0.893	2.172**	1.275	-0.241
	(0.940)	(1.019)	(0.915)	(0.916)	(1.019)	(0.979)
R-squared	0.0291	0.0287	0.0287	0.0345	0.0340	0.0291
Observations	24,009	24,009	24,009	24,009	24,009	24,009
mean dep var	68.98	68.98	68.98	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98	33.98	33.98	33.98

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. In columns (1)-(3), Hindu Indicator takes value 1 if religion of the household head is hindu at baseline, zero otherwise. In columns (4)-(6), Hindu Indicator takes value 1 if the religion of the household head is hindu and own farmland. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Hindu households increase their agricultural labor share by 3.14% following a drought last year.

Sectoral Wages

Higher pay in any particular sector leads to potential wage difference between the two sectors, attracting household to the higher-paying sector (Perloff, 1991; Moretti, 2000; Liu, 2017). The difference in sectoral wages could arise for a couple of reasons. Under the assumption that the regional climate shock affects the agricultural sector alone, the difference could be driven by how the labor market in the two sectors interplay. In other words, how the labor demand and labor supply in the two sectors respond to agricultural productivity loss. To provide an example, the productivity loss in agriculture would lower labor demand in the agricultural sector, thereby reducing agricultural wages while non-agricultural wages remain the same. This, in turn, creates a difference in sectoral wages in the short run along with a rise in unemployment.

To test how sectoral wages vary with drought, I utilize detailed information about household income and occupational choices of members. Household income is divided by the source of income: agricultural wage income and non-agricultural wage income. Dividing the total household agricultural wage income by the number of household agricultural wage workers, I compute the average household agricultural wage per worker. Similarly, I compute the non-agricultural wage income per household worker. Note that the agricultural wage income is different from farm income and non-agricultural wage income is different from household business income. Using specification (2.1), I test whether the drought significantly affects agricultural and non-agricultural wages per worker. The coefficient on the drought variable D_{dt-1} is the estimate of interest here. The trend in the wages of the two sectors at the district level assist in understanding the factors that influence the household labor diversification process. It is difficult to apriori gauge into the sign of β_1 because it depends on the interplay of labor markets across the two sectors. However, Jayachandran (2006) showed that the prevalence of landlessness among Indian agricultural workers reduces the elasticity of the wage with respect to productivity. In light of these findings, I assume agricultural wage would be responsive to drought occurrence.

Sectoral wages do not show any significant change with lagged z-score defined over the last year, the last monsoon term, and the last non-monsoon term.[16] Recall that the agricultural (and non-agricultural) wages were calculated by dividing household agricultural wage (non-agricultural wage) income by the number of workers employed in agricultural (non-agricultural) wage jobs within the household. The labor hours of the workers are not taken into consideration here. Every worker irrespective of the

[16] Appendix Table B.12 and Table B.13 show the results for the other coefficients.

hours worked, is treated as fully contributing to household income from that sector. This could potentially cause errors in the sectoral wage computation.

Table 2.6: Effect of Drought on Agricultural and Non-Agricultural Wages

	(1)	(2)	(3)	(4)	(5)	(6)
Lagged z-score (year)	752.1			435.7		
	(790.8)			(935.4)		
Lagged z-score (monsoon)		499.4			1668.5	
		(916.7)			(1169.1)	
Lagged z-score (non-monsoon)			490.7			1783.6
			(756.4)			(1104.6)
N	8,647	8,647	8,647	7,235	7,235	7,235
R^2	0.332	0.325	0.331	0.426	0.427	0.420
mean dep var	11,529	11,529	11,529	1,66,401	1,66,401	1,66,401
sd dep var	12,598	12,598	12,598	19,186	19,186	19,186

Each column records results for a different definition of drought. In column (1) and (4), drought is defined as the z-score for last year, in column (2) and (5), it is defined as the z-score for the last monsoon term, and in column (3) and (6) for the last non-monsoon term. For each column, temperature controls are used for the appropriate period corresponding to the drought definition. The dependent variable for columns (1)-(3) is the agricultural wage for each household which is calculated by dividing total household agricultural wage income by the number of agricultural wage workers in the household. The dependent variable for the next three columns (4)-(6) is the non-agricultural wage for each household which is calculated by dividing total household non-agricultural wage income by the number of non-agricultural wage workers in the household.
* $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

To get a better sense of all the changes in sectoral wages with drought occurrence, I winsorize the sectoral wages data at 1% at each tail. Since income and wage are variables that are more likely to have outliers, winsorizing is a standard tool. Table B.13 and Table B.14 in the Appendix section exhibit the results for sectoral wages winsorized at 1%. The agricultural wages do not change significantly with changes in lagged z-score. However, non-agricultural wages do respond to changes in lagged z-scores. In Table B.13 (1% winsorized wages), I observe that the monsoon and non-monsoon lagged z-score affect non-agricultural wages significantly. One unit increase in the lagged z-score over the monsoon term increases non-agricultural wages by 1984 rupees. Given that the average non-agricultural wage in the sample is 16322 rupees, one unit increase in the lagged monsoon z-score leads to a 12.15% increase in the non-agricultural wages. Lagged non-monsoon z-score also significantly affects the non-agricultural wage. One unit in the z-score (non-monsoon term) increases the non-agricultural wage by 2127 rupees, which translates to a 13.03% increase in non-agricultural wages.

Summarizing the results for the sectoral wages, I find that agricultural wages do not respond to drought occurrence, whereas non-agricultural wages seem to increase. I also documented that households on average move from agricultural to non-agricultural sectors. This could be driven by a couple of factors. Firstly, in the agricultural labor market, there could a simultaneous reduction of the agricultural labor supply and labor demand (recall that I define my drought variable at the district level). On the contrary, the non-agricultural labor market could experience an increase in labor demand because markets are localized and there is negligible trade across geographic regions because of poor transportation (which could potentially be exacerbated because of drought occurrence). Secondly, another situation could be that the household keeps the same number of workers in agricultural jobs but more members start working and work in the non-agricultural sector. This keeps the labor demand and supply in the agricultural sector unchanged. The increase in labor supply and labor demand (because of localized markets) in the non-agricultural sector increases the wages in that sector.

Risk Sharing

Households in developing countries face different risk environments because of the lack of a social security system, incomplete or missing financial and credit markets. Rural households depend on informal arrangements to insure against risk. There is extensive literature on risk-sharing in developing countries. Since the seminal paper by Townsend *et al.* (1994) that focuses on village as an insurance group for rural households, there has been substantial work related to risk sharing within a village (Udry, 1994; Morduch, 1995; Ravallion *et al.*, 1997; Morten, 2019). Insurance groups are not restricted by geographic location. Households within same caste or ethnic groups pool risk (Munshi and Rosenzweig, 2006; Grimard, 1997; Rosenzweig and Stark, 1989). Family ties are also important factors for risk pooling. Risk-sharing across extended families is another insurance tool (Park, 2006; Altonji *et al.*, 1992; Fafchamps and Gubert, 2007; Stark and Lucas, 1988; Witoelar, 2013).

Table 2.7: Risk Sharing: Effect of Diversification on Household Consumption

	(1)	(2)	(3)
Lagged z-score (year)	-6001.9**		
	(2804.2)		
Diversify from Ag X Lagged z-score (year)	3402.3		
	(3990.1)		
Lagged z-score (monsoon)		-9143.7**	
		(4454.6)	
Diversify from Ag X Lagged z-score (monsoon)		2679.0	
		(4098.3)	
Lagged z-score (non-monsoon)			-5112.0
			(4788.8)
Diversify from Ag X Lagged z-score (non-monsoon)			-5592.1
			(6196.3)
Diversify from Ag	-804.2	-937.9	3481.9
	(3261.0)	(3435.7)	(3157.1)
N	10,056	10,056	10,056
R^2	0.063	0.060	0.064
mean dep var	1,03,728	1,03,728	1,03,728
sd dep var	1,03,063	1,03,063	1,03,063

The dependent variable is the total consumption (in 2012 rupees) in household 'i' located in district 'd' in year 't'. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for the appropriate period corresponding to the drought definition. Diversify from Ag is an indicator which takes value 1 if household reduces their share of agricultural jobs compared to the last survey wave, 0 otherwise. Standard errors in parentheses. * $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

To smooth consumption during periods of income shock, apart from the above-mentioned insurance groups, rural households have resorted to other insurance mechanisms. Temporary migration is one of the main tools for risk-sharing in developing countries (Meghir *et al.*, 2017). Permanent migration is low in developing countries like India (Munshi and Rosenzweig, 2016; Topalova, 2010). The foremost reason is that migration permanently could be risky and individuals may be excluded from their native insurance networks (Bryan *et al.*, 2014; Tunali, 2000). Morten (2019) studies the dynamics of temporary migration and risk-sharing within villages in India. Another tool to smooth consumption is by selling durable production assets (Rosenzweig and Wolpin, 1993). Kochar (1999) shows that households shift labor from farm to off-farm employment and to what extent this shift explains the observed lack of correlation between consumption and idiosyncratic crop shocks.

In the current context, I hypothesize that the households are sharing risk (in times of drought occurrence) by reallocating labor from agricultural to non-agricultural sector jobs that probably help them to smooth consumption. Empirically, the test would involve testing whether consumption patterns across households that do reallocate labor to non-agricultural jobs and those that don't differ significantly. Consumption data is available for waves 2 and 3. I construct the indicator variable 'Diversify from Ag' which takes value 1 if the household share of agriculture jobs has fallen compared to the previous wave, zero otherwise. I interact this term with the lagged z-score to quantify the additional effect of diversification on total household consumption. Table 2.7 records the results. I find no significant influence of diversification from agriculture on the effect of drought on total household consumption.[17]

2.4.3 Consumption

Consumption changes in response to drought are indicative of the effect of drought on household well-being. A negative correlation between drought and household consumption could imply imperfect adaptation. On the other hand, no or positive cor-

[17] Appendix Table B.14 tabulates the other coefficients for Table 2.7.

relation would imply some kind of consumption smoothing because of diversification from agriculture. Kochar (1999) shows that the unobserved correlation between consumption and idiosyncratic crop shocks could be explained by an increase in market hours of work from farm work. I carry out the exercise in the same vein.

The consumption module in the IHDS survey wave-2 and wave-3 record consumption expenditure under 'consumption expenditure in the past 30 days' and 'consumption expenditure in the past 365 days'. The items included in the 'consumption expenditure in the past 30 days' include food items, regular day-to-day expenditure like fuel, telephone, cable, personal care, and other household expenditure like rents, medical expenditures.[18] The items under the 'consumption in the past 365 days' comprises of medical expenditure, school expenditure, more durable household expenses.[19]

I computed the total monthly expenditure by adding the expenditure on each item on the 'consumption in the past 30 days' list. I also added the expenditure on individual items on the 'consumption in the past 365 days' list to find the annual expenditure for those items. To find the total annual expenditure of the household, I multiplied the total monthly expenditure by 12 and added the annual expenditure. This variable could be constructed for 7999 households in wave-2 and wave-3.[20] I use specification (2.1) to test whether and how the consumption expenditure in household

[18] The items namely are rice, wheat, sugar, kerosene, cereal products, pulses, meat, fish, sweeteners, edible oil, eggs, milk, milk products, vegetables, salt, spices, tea, coffee, biscuits, paan, tobacco, intoxicants, fruits, nuts, eating out expenses, fuel, entertainment cost, telephone, cable, internet, personal care, toilet articles, household items, transportation cost, house rent, consumer taxes and fees, services, medical expenditure

[19] The items are medical expenditure, school tuition fees, books, and other educational articles, clothing, bedding, footwear, furniture and fixtures, crockery and utensils, cooking and other household appliances, goods for recreation, jewelry and ornaments, personal vehicles, therapeutic appliances, repair and maintenance, insurance premiums, vacations, social functions.

[20] Wave-1 does not have consumption data recorded under similar modules. So it could not be constructed.

i located in district d in year t is affected by drought. Every variable in the specification remains the same except the dependent variable is changed to total annual consumption expenditure of household i located in district d in year t. The sample for the regression is restricted to wave-2 and wave-3.

Table 2.8: Effect of Drought on Total Household Consumption

	(1)	(2)	(3)
Lagged z-score (year)	-4535.2		
	(2962.4)		
Lagged z-score (monsoon)		-6120.8	
		(4258.4)	
Lagged z-score (non-monsoon)			-7607.2**
			(3790.4)
N	15,998	15,998	15,998
R^2	0.058	0.057	0.059
mean dep var	1,03,840	1,03,840	1,03,840
sd dep var	1,03,189	1,03,189	1,03,189

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. The dependent variable is total consumption expenditure in the last year for household 'i' located in district 'd' in year 't'. The sample is restricted to wave-1 and wave-2 of the survey.
* $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

Table 2.8 shows how drought affects the total consumption expenditure. I find that an increase in z-score over all three periods leads to a reduction in consumption expenditure, however, it is only significant for the non-monsoon drought term. One unit increase in lagged z-score (defined over the non-monsoon term) leads to a 7,607 rupees reduction in consumption expenditure. Given the average annual expenditure is 1,03,840 rupees, this implies consumption expenditure reduces by 7.3% when there is a one-unit increase in lagged z-score. This reflects that households are negatively affected by drought.[21]

2.4.4 Threats to Identification

There are some potential threats to the above identification. The following subsections explore some of the threats and their implications. I also suggest robustness checks to test whether those are detrimental to the identification.

Omitted Variable Bias

The first threat is omitted variable bias. There could be some omitted time-varying factors at the household level that are correlated with the drought shock and contribute to the reallocation of jobs. Some of them are an increase in the number of household dependents, the household head being sick in the last year, the number of working-age members. I check how the coefficients in the main regression change when I control for the above variables. Location-specific demand for labor in the agricultural or non-agricultural sector in a particular year could mask the sign and magnitude of the main coefficient. To address that concern, I include region-by-year fixed effects in the main specification and check whether the results alter.

There are two more important considerations here. Access to irrigation facilities could modify the effect of drought on household labor reallocation decisions. To understand how the absence of irrigation, strengthens the effect of drought, I test

[21] I include the results related to the impact of drought on household income in the appendix. I find that drought increases total household income, caused primarily by increases in non-agricultural household income. However, household consumption decisions are not so linear in rural households in developing countries. Consumption choices and expenditures could be different for male and female members within a household (Duflo and Udry, 2004). This could lead to different income and consumption patterns. Apart from this, the increase in income and the decrease in consumption patterns associated with drought could also be explained by the precautionary savings of households for future shocks. Even though this evidence is mention-worthy but it is not the focus of the paper and is left for future research.

how drought affects the fraction of agricultural jobs in households without irrigation facilities. Any major policy during or in between the survey years could majorly change the household allocation of jobs. Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) was implemented in India in three phases: 2006, 2007, and 2008. The policy guarantees any willing worker 100 days of employment at the minimum wage level in manual or unskilled jobs in rural India. Drought management in India in recent years highlights the importance of this policy. This policy could particularly be beneficial for households experiencing drought. It is crucial to check whether such a policy alters the effect of drought on the household allocation of jobs. Since all households were exposed to the policy between wave 2 and wave 3, I interact survey wave dummy with the drought variable to understand whether drought has any differential effect based on the period of occurrence. The thought exercise here is that if there is any differential effect of drought between the years 2005 and 2011, it could potentially be driven by the policy implementation.

Attrition Bias

A second threat to identification is attrition bias. Attrition in the survey occurs in the form of migration, death, or splitting of households. To comprehend how these affect the main variable of interest, I briefly recap the survey strategy. IHDS survey does not track a household (or a part of a household) if it moves out of the neighborhood (primary survey unit). However, it does track households that move out of the original household but remain in the same neighborhood. It is assigned a split identifier. The potential reason for a household missing in the following wave is either migration or death. The composition of the neighborhood is changed because of households moving in and out of different labor markets that could potentially be related to drought occurrence.

To eliminate such possibilities, I check whether household attrition is correlated with drought occurrence last year. I implement specification (2.1) and change the dependent variable to the fraction of missing household members. To address the second concern that some households split and moved to a different house but remained in the same neighborhood, I consider an alternative sample. The alternative sample is where all split households are included. I run the main specification (2.1) using the alternative sample. This helps to understand whether the splitting of households is a threat to the study. Previous studies have shown that splitting of households is non-random and could affect results significantly if not taken into account (Foster and Rosenzweig, 2002; Thomas *et al.*, 2012). The section on robustness checks includes the results and discussions about the above threats to the identification.

2.5 Robustness Checks

In this section, I address identification issues related to the primary specification. I conduct the following robustness checks.

2.5.1 Incorporating MGNREGA

Mahatma Gandhi National Rural Employment Guarantee Act, Indian labor law, and social security measure that aims to guarantee the 'right to work' was introduced in 2005. The program guarantees 100 days of unskilled manual work at a minimum wage payment. The program is implemented by the local village government (called the gram panchayats). There is no formal eligibility except that the candidate needs to be a rural resident of at least 18 years of age. The program was rolled out in three phases. The first phase was rolled out in 2006 in 200 districts, followed by 170 districts in 2007 and the other remaining districts in 2008.

Table 2.9: Availability of MGNREGA Program

	(1)	(2)	(3)
Lagged z-score (year)	-1.775		
	(1.298)		
Lagged z-score (year) X Year Indicator (2005)	1.038		
	(1.525)		
Lagged z-score (year) X Year Indicator (2011)	3.030*		
	(1.589)		
Lagged z-score (monsoon)		-2.547**	
		(1.196)	
Lagged z-score (monsoon) X Year Indicator (2005)		0.622	
		(1.544)	
Lagged z-score (monsoon) X Year Indicator (2011)		2.145	
		(1.713)	
Lagged z-score (non-monsoon)			-2.745
			(1.734)
Lagged z-score (non-monsoon) X Year Indicator (2005)			2.178
			(2.001)
Lagged z-score (non-monsoon) X Year Indicator (2011)			2.563
			(2.258)
Lagged z-score X Year indicator (2005) + Lagged z-score	-0.737	-1.925*	-0.567
	(1.035)	(1.092)	(1.350)
Lagged z-score X Year indicator (2011) + Lagged z-score	1.255	-0.401	-0.182
	(1.214)	(1.613)	(1.392)
Observations	23,997	23,997	23,997
R-squared	0.030	0.029	0.029
mean dep var	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98

The dependent variable is the percentage of agricultural jobs in household 'i' located in district 'd' in year 't'. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. Year indicator (2005) [Year indicator (2011)] takes value 1 if the observation corresponds to wave-2 [wave-3], zero otherwise.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In the context of the current study, MGNREGA was implemented in all rural areas throughout the country between wave-2 and wave-3. To understand the effect of the policy on labor diversification, I introduce two additional interactions with the lagged z-score in the specification (2.1). The two interaction terms are 'Lagged z-score X Year indicator (2005)' and 'Lagged z-score X Year indicator (2011)'. The variables 'Year indicator (2005)' and 'Year indicator (2011)' take value 1 if the observation is for wave-2 and wave-3 respectively. Therefore, the interaction term allows me to understand if there is any differential effect of drought on the percentage of agricultural jobs based on the year. The year indicator (2011) includes any change that could alter the effect of drought on labor reallocation. This also includes the availability of MGNREGA.

Table 2.9 summarizes the key coefficients of the regression. I find that the effect of drought does not have a differential effect for the households in wave-3 (refer to the coefficient for Lagged z-score X Lagged z-score + Year indicator (2011)). This could plausibly imply that the exposure to MGNREGA did not affect labor allocation decisions significantly in response to drought. However, the drought before wave-2 seems to affect the labor allocation decision marginally significantly. One unit increase in the z-score (for households in wave-2) significantly reduces the percentage of agricultural jobs by 1.925 percentage points.

2.5.2 Migration of members in the sample households

Differential migration across households based on their exposure to drought would have caused concern over the selection problem. I compute the number of missing household members in each household in the sample (7999) between wave-1 and wave-2 and wave-2 and wave-3. I divide the number of missing members in each household by the total number of household members present in the last wave and express that variable in percentage terms. I use specification (2.1) to test whether drought significantly affects the percentage of missing members within a household. However, in Table 2.10, I find that there is no significant effect of an increase in the lagged z-score on the percentage of household members that migrated between survey waves.

Table 2.10: Effect of Drought on the Percentage of Missing Members Among Households in the Study Sample

	(1)	(2)	(3)
Lagged z-score (year)	-0.116		
	(0.450)		
Lagged z-score (monsoon)		-0.506	
		(0.483)	
Lagged z-score (non-monsoon)			0.0185
			(0.393)
N	15,998	15,998	15,998
R^2	0.351	0.350	0.351
mean dep var	16.18	16.18	16.18
sd dep var	19.26	19.26	19.26

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for the appropriate period corresponding to the drought definition. The dependent variable is the fraction of missing members in the household from the last survey wave. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.5.3 Other Robustness Checks

60% of agricultural land is dependent on precipitation. The major part of the annual precipitation in India occurs in June, July, August, and September (also known as the monsoon showers). Irrigation facilities are still not wide-spread across the country. Access to irrigation facilities alter the effect of climate change on agricultural crop yields and revenue (Benonnier *et al.*, 2019; Schlenker *et al.*, 2005). One unit increase in the lagged monsoon z-score leads to 3.54 percentage points decrease in the percentage of household agricultural jobs. The average agricultural job percentage is 62.67. This implies that a one-unit increase in monsoon z-score leads to a 5.65% reduction in the percentage of household agricultural jobs. Comparing these results to the ones in Table 2.2, I find that the effect is worse for households without irrigation facilities. No access to irrigation facilities makes the household more vulnerable to fluctuations in monsoon precipitation.

Table 2.11: Robustness Checks - Effect of Drought on Percentage of Agricultural Jobs

	Coefficient (Std Error)	N (R^2)	mean (sd) dep var
(1) Dropping households with irrigation at baseline	-3.538***	14340	62.67
	(1.190)	0.04	(36.28)
(2) Accounting for diverse labor market	-1.862*	23997	68.98
	(0.958)	0.03	(33.98)
(3) Accounting for households splits	-1.982**	23997	68.49
	(0.919)	0.03	(33.44)

Each row (1) through (4) represent a separate regression. The dependent variable is the percentage of agricultural jobs in a household located in district 'd' in year 't'. The key independent variable is the drought measure (lagged z-score defined for the monsoon term), reported in the table. The rest of the covariates included are same as in specification (2.1). Regressions in rows (1) and (2) includes household and year fixed effects. In row (3), I include district, household and region X year fixed effects. In row (1), households with access to irrigation at baseline (wave-1) are dropped. In row (2), I add region fixed effects to account for diverse labor markets. In row (3), we treat the households that split from the original house in later survey waves as a single household.

The study sample comprises rural households located across 184 districts in 18 states of India. Figure 2.1 (in the Data section) shows the distribution of the districts across India. The labor market in different parts of the country could be different (Debroy, 2013). Additionally, these markets could look vastly different over twenty years. Even though I add district and household fixed effects, to account for the evolving labor markets across India, I include region-year fixed effects.[22] Comparing the results of Table 2.11 to Table 2.2, I find that the sign remains the same i.e. an increase in lagged z-score reduces the percentage of household agricultural jobs. However, the effect is marginally significant for monsoon drought and the magnitude of the coefficient also falls. Given that the data is a panel of three years and that the

[22]The sample is divided into two regions. Region 1 includes the states of Himachal Pradesh, Punjab, Uttaranchal, Haryana, Rajasthan, Uttar Pradesh, Bihar, Assam, West Bengal, Jharkhand, and Orissa. South includes. Region 2 includes Chhatisgarh, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Kerala, and Tamil Nadu.

specifications already include household and district fixed effects, the region could not be more geographically redefined.

Household splits are common. Younger members grow up and move into a different house. The IHDS survey was designed in such a way that the households that split but stayed in the same neighborhood (primary survey unit (PSU)) but re-interviewed in the next round of the survey. The split households were assigned an identifier which helped to link the split households to the original households. In the current study, I use the sample of households that remained in the same physical house in the three waves of the survey. I refer to the sample as the "original" household sample. However, previous studies have documented that household splitting is non-random and could affect results significantly (Foster and Rosenzweig, 2002; Thomas *et al.*, 2012). To understand how the diversification of labor would have been if the households did not split, I construct an additional sample "imagine no splits" households sample.[23] I consider all the households that split in wave-2 and wave-3 from the original households in wave-1 to be part of the original household. For example, HH 1 (in 1993-94) split into HH11 and HH12 (in 2004-05). Again HH11 split into HH111 and HH112 (in 2011-12) and HH22 split in HH221 HH222 and HH223 (in 2011-12). For the "imagine no split" sample, the household in the three waves will look like: HH1 (1993-94) , HH11 + HH12 (2004-05) and HH111 + HH112 + HH221 + HH222 + HH223 (in 2011-12). I find the results are similar to the "original" households sample. Even though drought defined over any term reduces the percentage of agricultural jobs, monsoon drought significantly affects labor diversification. One unit increase in the monsoon z-score decreases the percentage of agricultural jobs by 1.982 percentage

[23]There were no households that did not split between wave-1 and wave-2. Between wave-2 and wave-3, 6,513 households did not split. One possible reason being the survey was conducted by HDPI for wave-1 and IHDS for wave-2 and wave-3. Therefore, I could not construct the sample of households that never split across the three waves.

points. Given the average percentage of agricultural jobs in the sample is 68.49%, one unit increase in monsoon z-score translates to a 2.89% reduction in the percentage of within household agricultural jobs.

2.5.4 Different definitions of diversification

To understand the labor diversification among different groups within the households (for example males vs females), I re-define the dependent variables in several ways.

Table 2.12 records the results for drought defined over last year's monsoon (lagged z-score). In Table 2.12, the dependent variable in column (1) is the percentage of household members employed in agriculture, in column (2) is the percentage of female members employed in agriculture, and in column (3) is the percentage of male members employed in agriculture. One unit increase in the lagged z-score leads to a 2.5 percentage points reduction in the percentage of household members employed in agriculture. The average percentage of household members working in agriculture is 42%, which implies one unit increase in the z-score leads to a 6% reduction in the percentage of members working in an agricultural job. Observing columns (2) and (3) and performing similar calculations, I find that one unit increase in lagged z-score leads to a 3.19 and 2.41 percentage points decrease in the percentage of female and male members in agricultural jobs. The average percentage of females (males) employed in agriculture in a rural household is 39 % (48%). Therefore, an increase in z-score by one unit leads to a reduction in the percentage of females (males) in agricultural jobs by 8.16% (5.06%). This shows that females are more likely to move out of agricultural jobs when the household experiences drought.[24] This could mean

[24]The sample size differs in columns (1)-(3). Column (2) and (3) has a smaller sample size. This is because some households do not have any female and male members respectively.

that the non-agricultural job available in the year is traditionally women-oriented. This includes manufacturing jobs in the textile industry where 77.4% of the total workforce is women (Shazli and Munir, 2014). On the other hand, agriculture has been traditionally a male-oriented job (Census, 2001).

Table 2.12: Effect of Monsoon Drought on Different Diversification Variables

	(1)	(2)	(3)	(4)	(5)	(6)
Lagged z-score (monsoon)	-2.523***	-3.187**	-2.414***	-3.476***	-2.036**	-5.567***
	(0.760)	(1.243)	(0.710)	(1.012)	(0.905)	(1.697)
N	23,997	23,689	23,635	23,997	22,987	23,363
R^2	0.224	0.222	0.085	0.094	0.080	0.189
mean dep var	42.01	39.07	47.67	57.98	65.62	51.76
sd dep var	28.33	38.10	34.57	34.37	40.66	45.02

Each column denotes a different dependent variable denoting diversification away from agriculture. (1): Percentage of household members employed in agriculture (2): Percentage of female household members employed in agriculture (3): Percentage of male household members employed in agriculture (4): Percentage of working-age household members (14-65 age) employed in agriculture (5): Percentage of working-age male household members (14-65 age) employed in agriculture (6): Percentage of working-age female household members (14-65 age) employed in agriculture
Standard errors in parentheses. * p < 0.1 , ** p<0.05, *** p<0.01

The latter three columns: column (4)-(6) observes how the diversification occurs within the working-age members only. Even though child labor is rampant in India, it is not absurd to assume that working-age members are likely to be employed. In column (4), the dependent variable is the percentage of working-age members employed in agriculture. I find that one unit increase in the z-score leads to a 3.48 percentage points reduction in the percentage of working members employed in agriculture. The average percentage of working members in agriculture is 58%, which implies that the reduction in the percentage of working members in agriculture is almost 6% with a unit increase in the z-score. Similarly, columns (5) and (6) show that the percentage of male (female) working members employed in agriculture fall by 2.04 and 5.57 percentage points with an increase in the z-score. The average percentage of male (female) working members employed in agriculture is 65.6% (51.8%). This implies one unit increase in the z-score leads to a 3.1% (10.76%) reduction in the percentage of male (female) working-age members in agricultural jobs. The results of the last three columns are in a similar vein to the first three columns. The diversification is stronger for the female than the male members of the household.

2.6 Conclusion

In this paper, I study how drought affects the labor choices of rural households in India. I further explore possible mechanisms that may modify the household response to drought. In particular, I consider switching cost, non-agricultural skill, differential sectoral wages, and risk-sharing. This paper contributes to the 'envirodevonomics' literature by complimenting earlier findings on how households respond to deviations in rainfall. It extends the literature, in particular, by providing evidence on how land-ownership modifies the effect of drought on household labor allocation.

I combine high-resolution climate data with detailed survey panel data on households to address the research questions. I find that households reduce their share of agricultural jobs by 2.9% following a drought. Additionally, I find that households with members who have greater than primary education are more likely to move to the non-agriculture sector in response to drought. Cultural norms associated with owning farmland as well as long bureaucratic processes associated with buying and selling of farmlands have restricted land market transactions. Land-owning households respond to drought by increasing their share of agricultural jobs by 3.14%, reinforcing the frictions associated with owning farmland. This result has important policy implications governing land ownership.

These findings have important implications on the structural transformation of the Indian economy. In the following chapter of my dissertation, I build a simple household labor allocation model incorporating land ownership to conduct imporant policy experiments related to future structural transformation.

Chapter 3

THE IMPACT OF DROUGHT ON STRUCTURAL TRANSFORMATION IN INDIA

3.1 Introduction

Developing countries with large agricultural sectors are vulnerable to climate change. Rural households engaged in the agricultural sector have an incentive to insure themselves against climate risk. For example, informal arrangements can allow households to insulate themselves against income shocks in the absence of formal insurance markets (Rosenzweig and Stark, 1989; Udry, 1994; Townsend *et al.*, 1994; Morten, 2019). Another less studied way of insuring against income risk is sectoral labor reallocation (Colmer, 2018). However, labor market frictions can limit the extent of adjustment. In particular, the land market in India is notoriously rigid due in part to cultural norms concerning land ownership and 65-70% of rural households are landowners (Fernando, 2020). These households' attachment to their land may reduce the rates at which they choose to reallocate their labor to the manufacturing and service sectors.

In India, climate change is expected to increase the frequency and severity of drought (Bisht *et al.*, 2019). This raises serious concerns because 60% of agricultural land remains rainfall dependent. While India has witnessed a falling agricultural employment share since the 1970s, more than 50% of the Indian workforce is still employed in agriculture, contributing to 17-18% of the total GDP (World Bank, 2015). Figure 3.1 shows the percentage of drought districts within each state, the percentage of agricultural land, and the percentage of agricultural land that is rain-fed for

2014. The considerable overlap between locations with a high occurrence of drought and agricultural land that is rainfed underscores rural households' vulnerability to droughts. Additionally, Pachauri *et al.* (2014) predicts agricultural losses of $7 billion USD by 2030 due to droughts. In this context, it is important to understand the degree to which household reallocation of labor is possible given the labor market barriers and the extent to which it could help moderate the consequences of droughts.

Figure 3.1: Drought, Agriculture and Rainfed Land Proportion

(1) (2) (3)

(1): More than 50% of the districts experienced drought in dark brown shaded states, more than 25% of the districts experienced drought in light brown shaded states (2): Green shaded states have more than 50% of land used for agriculture (3): Orange shaded states have more than 50% of agricultural land dependent on rainfall

The goal of this paper is to understand how drought affects household labor allocation and the extent to which land ownership modifies the reallocation of household labor in rural India. In chapter 2, I combined a rich household survey, spanning almost twenty years, with a high-resolution climate dataset. Panel data regressions revealed that rural households reduce their share of agricultural labor by 3% in response to one standard deviation decrease in district-level rainfall. This 3% reduction is equivalent to 110 annual household labor hours. I investigated heterogeneity in this response across landed and landless households and found that landless households reduce agriculture labor hours by 10.2% whereas households that own land do not reduce their share of agricultural labor.

Motivated by these empirical facts developed in chapter2, I build a model of a household labor allocation with two sectors - agriculture and non-agriculture and use the model to formalize hypotheses about how land ownership modifies the effect of drought on labor allocation. I model attachment to agriculture following Zimmermann (2020). For landless households, the attachment to agriculture reflects adjustment costs and barriers to entering the non-agricultural sector, and for landed households, the attachment to agriculture additionally includes the psychological cost of leaving one's land, for example, due to religious and cultural norms. I calibrate the model to match the average labor hours for landed and landless households across drought and non-drought years and I use the calibrated model to analyze how the projected near-future increase in the frequency of droughts will affect structural transformation in India. Finally, I use the model to predict how climate change will affect the size of subsidy payments needed for the government to achieve its future targets for increasing employment in the manufacturing and service sectors.

My counterfactual scenarios simulate the effects of increasing the frequency of drought over the next thirty years as predicted by the IPCC 5th assessment report (Pachauri *et al.*, 2014). For landed households, I find that the average share of agriculture labor increases marginally as they move from the status quo to an extreme regime where droughts occur every year. This counterintuitive result follows from the fact that land ownership hinders sectoral movement and requires additional labor during inclement weather. On the contrary, for landless households, I find that the agriculture labor share falls as droughts increase in frequency. Combining the effects predicted for landed and landless households, I find that there is a net reduction of 1-2% in average agriculture labor hours from the status quo to a constant-drought regime. This relatively small percentage change, however, is equivalent to 2.5 to 5 million individuals moving out of agriculture.

I use these results to consider the 'Make in India' initiative, which was launched in September 2014 with the aim of transforming India into a manufacturing hub, raising manufacturing to 25% of GDP, and creating 100 million manufacturing jobs. Green (2014, 2015) analyze the efficacy of the 'Make in India' campaign and predict the sectoral labor shares needed to achieve these targets. They conclude that to achieve the 25% manufacturing share, the agricultural labor share must decline to 38%, 33%, and 25% in 2022, 2025, and 2035, respectively. I use my model to calculate how large wage subsidies would have to be to achieve these sectoral targets. I find that to achieve the 38% agricultural labor share by 2022, the non-agricultural wage would need to be subsidized by 3.26 rupees per hour (a 9.8% increase from 2011 levels), by 2025 it would need to be subsidized by 4.96 (14.9%) rupees per hour, and by 2035 it would need to be subsidized by 9.36 (28.1%) rupees per hour. Under the climate change regime with increased drought, the non-agriculture wage would need to be subsidized by; 5.37 (16.1%) rupees per hour by 2022, 8.37 (25.0%) rupees per hour by 2025, and 16.27 (48.7%) rupees per hour by 2035, reinforcing the importance of sectoral barriers.

This paper contributes to the growing literature at the intersection of environmental economics and development economics, also known as "envirodevonomics" (Greenstone and Jack, 2015). First, I contribute to the strand of literature that studies the impact of changes in weather and climate on sectoral labor movements. Second, I contribute to a strand of the literature seeking to understand frictions and barriers to the movement of labor out of the agricultural sector. My framework builds on this literature by modeling how the effect of weather on sectoral labor allocation is modified by frictions in labor movements in India. Lastly, I build a model of household labor allocation that incorporates attachment to agriculture arising from sectoral barriers such as land ownership. This model allows me to analyze how the effect of climate change on labor allocation is modified by land ownership.[1]

3.2 Data, Summary Statistics and Empirical Specification

There are two main sources of data. The first one is household-level panel data from the India Human Development Survey (IHDS). The second is a series of gridded weather datasets from Willmott and Matsuura (2001). Section 2.3 in chapter 2 describes the data sources and summary statistics in detail.

I build the model of household labor allocation based on the reduced form evidence I find which is outlined in section 2.4 of chapter 2. In chapter 2, the dependent variable is the share of agricultural jobs in a household. The earliest wave of the household-level dataset did not record labor hours. However, I imputed the labor hours of the earliest wave based on the latter two waves of household-level data. Appendix section C.1 describes the imputation of labor hours, tabulates the summary statistics including share of labor hours and records the empirical results where the dependent variable is the share of agricultural labor hours instead of the share of agricultural jobs.

3.3 Model

3.3.1 Setting

Consider an economy with two types of households indexed by j = {f,nf}. Those indexed by 'f' own farmland and those indexed by 'nf' do not own farmland. I will continue to refer to these two groups as "landed" and "landless". There are two

[1] For a detailed review of the literature refer to section 2.2 in chapter 2.

sectors of production: agriculture, a, and non-agriculture, m. Every household j is endowed with τ units of labor. Households face drought realizations at the beginning of each period $z \in \{0, 1\}$ that affects wages $(w^a(z), w^m(z))$ and farm income $(\pi(z))$. After drought shocks are realized, household j decides how to allocate labor to the agriculture sector, l_z^{aj}, to the non-agriculture sector, l_z^{mj}, and to leisure, l_z^j. They receive the associated wages and farm income at the end of the period, which in turn, determines household consumption c_z^j. The landed households, 'f' only receive the farm income. The timeline is as follows:

Figure 3.2: Timeline

Weather Realization	Decision	Payoffs
Drought	Ag Labour : l_z^a	Wages
$z \in \{0, 1\}$	Non-Ag Labour : l_z^m	$w^m(z)$
$\downarrow$	Leisure : l_z	$w^a(z)$
$\{w^m, w^a, \pi\}$	Consumption: c_z	$\pi(z)$

Preferences

Households receive utility from consumption, c_z^j, and from working in agriculture, l_z^{aj}. They experience disutility from working, L_z^j. The utility function is as follows:

$$U_z^j(c_z^j, l_z^{aj}, L_z^j) = \frac{(c_z^j)^{1-\sigma}}{1-\sigma} + \psi_z^j(l^{aj})^{\delta} - k_z^j(L_z^j)^{\alpha} \tag{3.1}$$

The first part of the utility function is a standard CARA specification. The second term represents the attachment to agriculture, similar to Zimmermann (2020). The last term represents the preference for leisure. This follows from Greenwood *et al.*

(1988) and is standard in the literature. The parameter σ is the curvature parameter on consumption. δ is the curvature parameter on the utility benefit of working in agriculture. ψ_z^j is the heterogeneous land attachment parameter. It is allowed to differ across landed and landless households and across drought and non-drought states. A higher value of ψ_z^j implies a stronger preference for agricultural work. k_z^j is the heterogeneous leisure parameter that also varies across household types and drought states. A higher value of k_z^j implies a higher preference for leisure.

Constraints

Agriculture wages, w_z^a, and non-agriculture wages, w_z^m, are both potentially affected by the weather realization, z. Additionally, households that own land receive farm income, $\pi_(z)$, which is also affected by drought occurrence z. Total household time, τ, is divided into labor supplied, L_z^j, and leisure, l_z^j. Labor hours are then divided between agricultural work, l_z^{aj}, and non-agricultural work, l_z^{mj}. The budget and time constraints for the landed households are thus as follows:

$$c_z^j = w_z^a l_z^{a,j} + w_z^m l_z^{m,j} + \pi_z \tag{3.2}$$

$$L_z^j = l_z^{a,j} + l_z^{m,j} \tag{3.3}$$

$$\tau = L_z^j + l_z^j \tag{3.4}$$

Landless households face the same set of constraints except that their budget constraint excludes farm income, as they do not own land.

3.3.2 Household Optimization Problem

The optimization problem for a landed household can be represented as follows:

$$\max_{l_z^{a,f},l_z^{m,f}} \quad \frac{(c_z^f)^{1-\sigma}}{1-\sigma} + \psi_z^f (l_z^{a,f})^{\delta} - \frac{k_z^f}{2}(L_z^f)^{\alpha} \tag{3.5}$$

$$\text{s.t.} \quad c_z^f = w_z^a l^{a,f} + w_z^m l^{m,f} + \pi_z \tag{3.6}$$

$$L_z^f = l_z^{a,f} + l_z^{m,f} \tag{3.7}$$

$$\tau = L_z^f + l_z^f \qquad \forall z \in \{0,1\} \tag{3.8}$$

Given the intital weather realizaton, the wages in the agriculture and non-agriculture sectors, and the total time endowment, landed households solve the constrained optimization problem in (3.5)-(3.8) by choosing their optimal labor allocation and consumption in both drought (i.e., z= 1) and non-drought (i.e., z=0) states i.e. by choosing $\hat{l}_1^{a,f}$, $\hat{l}_1^{m,f}$, $\hat{L}_1^f$, $\hat{c}_1^f$, $\hat{l}_0^{a,f}$, $\hat{l}_0^{m,f}$, $\hat{L}_0^f$, and $\hat{c}_0^f$.

Rearranging the first-order conditions yields two key equations that summarize households' decision rules.[2]

$$(c_z^f)^{-\sigma} w_z^m = k_z^f \alpha (l_z^{a,f} + l_z^{m,f})^{\alpha-1} \tag{3.9}$$

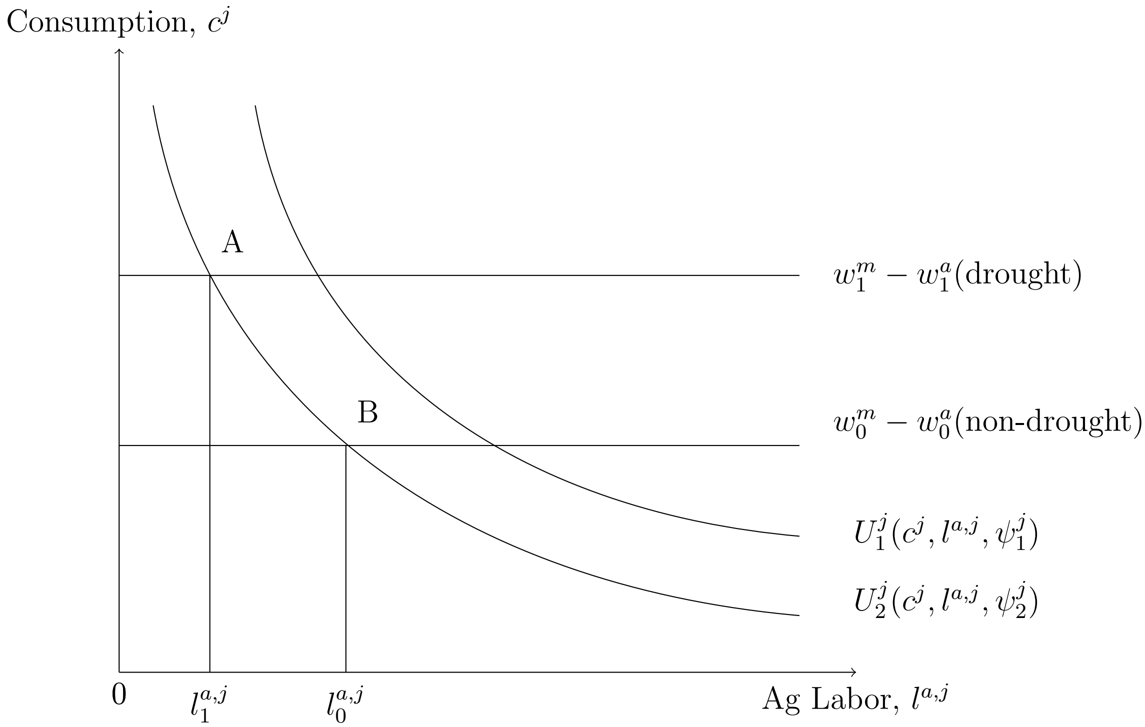
$$(w_z^m - w_z^a) = \frac{\psi_z^f \delta (l_z^{a,f})^{\delta-1}}{(c_z^f)^{-\sigma}} = \frac{MU_{l_z^a}}{MU_{c_z}} \tag{3.10}$$

The comparative statics of equation (3.9) imply that landed households decide how much labor to supply (L_z^f) across drought and non-drought based on the wages in that state. If wages fall, households supply less labor. Total labor supply (L_z^f)

[2]For the detailed first-order conditions, refer to the Appendix section C.2.

is allocated to the agriculture ($l_z^{a,j}$) and non-agriculture sectors ($l_z^{m,j}$), governed by the first-order condition shown in equation (3.10). Specifically, the amount of labor allocated to agriculture is governed by the difference in wages between the two sectors and the ratio of the marginal utility of working in agriculture to the marginal utility of consumption.

Figure 3.3: Household Labor Allocation Decision



The above figure shows the allocation of labor across drought and non-drought states. A higher value of ψ^j implies greater attachment to land and a higher allocation of labor into the agriculture sector.

Figure 3.3 provides the graphical intuition for the sectoral labor allocation decision. The utility function parameters are set to make utility concave. The slope of the utility function in the consumption-agriculture labor space is given by the rightmost part of equation (3.10). The smaller the difference in sectoral wages, the higher the allocation of labor to agriculture (labeled B). The wage difference during drought

$(w_1^m - w_1^a)$ is higher compared to the wage difference during non-drought $(w_0^m - w_0^a)$. Therefore, households allocate less labor to agriculture during a drought (labeled A). It is important to note that the total labor supplied is different in each state. Thus, the reduction in agricultural labor does not necessarily imply a decrease in the share of agriculture labor hours. All else constant, a higher value of ψ implies higher attachment to agriculture and hence a higher utility level. Therefore irrespective of the wage difference in each state, households with higher ψ will allocate more labor to the agriculture sector.

The optimization problem for a landless household can be represented as follows:

$$\max_{l_z^{a,nf}, l_z^{m,nf}} \quad \frac{(c_z^{nf})^{1-\sigma}}{1-\sigma} + \psi_z^{nf}(l_z^{a,nf})^{\delta} - \frac{k_z^{nf}}{2}(L_z^{nf})^{\alpha} \tag{3.11}$$

$$\text{s.t.} \quad c_z^{nf} = w_z^a l^{a,nf} + w_z^m l^{m,nf} \tag{3.12}$$

$$L_z^{nf} = l_z^{a,nf} + l_z^{m,nf} \tag{3.13}$$

$$\tau = L_z^{nf} + l_z^{nf} \qquad \forall z \in \{0,1\} \tag{3.14}$$

Given the intital weather realizaton and taking wages, w_z^a and w_z^m and the total time endowment τ as given, landed households maximize their utility function (3.11) given the constraints (3.12)-(3.14) by choosing the optimal labor allocation and consumption in both drought (i.e., z= 1) and non-drought (i.e., z=0) states by choosing $\hat{l}_1^{a,nf}$, $\hat{l}_1^{m,nf}$, $\hat{L}_1^{nf}$, $\hat{c}_1^{nf}$, $\hat{l}_0^{a,nf}$, $\hat{l}_0^{m,nf}$, $\hat{L}_0^{nf}$, and $\hat{c}_0^{nf}$.

The tradeoffs faced by landless households are similar to the landed households with two distinctions. First, landless households may have different values for the attachment to land parameter ψ_z^{nf} and the leisure parameter k_z^{nf} compared with landed households. Second, as noted earlier, the landless households do not have farm income, π. Apart from these differences, the landed and landless households face similar tradeoffs in deciding how to allocate labor across the agriculture and non-agriculture sectors in drought and non-drought states.

3.4 Results

I calibrate the model to match the labor allocation decisions of rural Indian households across drought and non-drought periods. First, I calculate wages using the National Sample Survey (NSS) data of India. Next, I calibrate a subset of the parameters that are not identifiable in my data by fixing them at values estimated in prior studies of developing countries. Lastly, I use the method of moments to recover the remaining parameters to match data moments.

3.4.1 Calibration

The Employment and Unemployment surveys of the National Sample Survey (NSS) provides national and state data on wages paid to labor. I use three rounds of the NSS dataset (that correspond to the survey years of the IHDS dataset) to calculate hourly wages in rural India for 2011-12, 2004-05, and 1993-94. The NSS does not follow the same households over time. However, the NSS income module is more detailed than IHDS. The NSS records industry code, hours worked in any job in the last week, and the total payment at the end of the week. I assume that a full day of work is equivalent to eight hours of work. I convert the days of work into hours worked last week. The hourly wage is the total weekly payment divided by the number of hours worked. Based on the National Industrial Classification (NIC), the wages are classified as agriculture or non-agriculture. I further stratify agriculture and non-agriculture wages into drought and non-drought periods. For instance, agriculture wage during drought is measured as the average agriculture wage among districts that experienced a drought. I use the most recent round of 2011-12 wages for the calibration exercise.

Table 3.1: Calibration Summary

Parameter		Source	Value
Hourly ag wage (drought)	w_1^a	Data	17.09
Hourly ag wage (non-drought)	w_0^a	Data	20.82
Hourly non-ag wage (drought)	w_1^m	Data	33.43
Hourly non-ag wage (non-drought)	w_0^m	Data	33.34
Farm Income (drought)	π_1	Data	32,648
Farm Income (non-drought)	π_0	Data	51,215
Curvature parameter on consumption	σ	Literature	0.7
Curvature parameter on leisure	α	Literature	2
Curvature paramter on ag attachment	δ	Literature	0.5
Ag attachment for Landed HH (drought)	ψ_1^f	Method of Moments	0.4480
Ag attachment for Landed HH (non-drought)	ψ_0^f	Method of Moments	0.2883
Leisure parameter for Landed HH (drought)	k_1^f	Method of Moments	2.3250
Leisure parameter for Landed HH (non-drought)	k_0^f	Method of Moments	2.0260
Ag attachment for Landless HH (drought)	ψ_1^{nf}	Method of Moments	0.4018
Ag attachment for Landless HH (non-drought)	ψ_0^{nf}	Method of Moments	0.3065
Leisure parameter for Landless HH (drought)	k_1^{nf}	Method of Moments	2.6370
Leisure parameter for Landless HH (non-drought)	k_0^{nf}	Method of Moments	2.5400

Note: The leisure parameter is measured in units of 10^{-6}. Hourly wages are measured in rupees.

The top part of Table 3.1 reports sectoral wages across drought and non-drought periods for 2011-12. Agriculture wages are lower by 3 rupees for the drought-affected districts. However, the wages in the non-agriculture sector differ only by 0.1 rupees. I calculate the average farm profit among landowners from Wave-3 of the IHDS dataset. IHDS records farm income separately for each household that owns land. Using the drought status for the year 2011, I calculate the average farm income for drought and non-drought districts. Farm income is 36% higher for the non-drought period, compared to the drought period, reflecting lower productivity during drought.

The middle part of Table 3.1 summarizes how I choose $\{\sigma, \alpha, \delta\}$ from the prior literature since they are not easily identified by the available data. First, I set the consumption curvature parameter to 0.7 which falls toward the middle of the range of estimates from prior studies. For example, an experimental study conducted on rural households in India concluded that they are moderately risk-averse (Binswanger, 1980) with a partial risk aversion coefficient is between 0.316 and 0.812. More recently, Sengupta (2011) finds that rural households exhibit intermediate-risk aversion with a partial risk aversion coefficient between 0.81 and 1.74. I estimate the model with sigma=0.7,

Next, I set the value for α, the curvature parameter on total household labor supplied (L), to be 2, following work on real business cycle models. Lastly, I set δ, the curvature parameter on the preference for agriculture work, to be 0.5. Zimmermann (2020) proposes there is an affinity to work on one's farm in rural India. The additively separable part of the utility function denoting affinity to work on land exhibits diminishing returns to scale. Technically, the logical range for δ is (0,1).

I jointly calibrate the remaining eight parameters $\{\psi_1^f, \psi_0^f, k_1^f, k_0^f, \psi_1^{nf}, \psi_0^{nf}, k_1^{nf}, k_0^{nf}\}$ to match the eight data moments constructed from the three waves of the IHDS dataset. These parameters are the attachment to agriculture, ψ, and leisure preference, k, across drought and non-drought periods for landed and landless households. The four data moments for the landed households are: average total household labor hours in agriculture during drought periods, average total household labor hours in agriculture during non-drought periods, average total household labor hours in non-agricultural work during drought periods, and average total household labor hours in non-agricultural work during non-drought periods. I use these moments to calibrate the attachment to agriculture across drought, ψ_1^f, and non-drought states, ψ_0^f, and the leisure parameter across drought, k_1^f, and non-drought states, k_0^f. I follow an analogous procedure for landless households. Table 3.2 shows how well the model moments match the data moments for landed and landless households.

Table 3.2: Model Fit: Targeted Moments

	Data Moments	Model Moments	% Difference
Landed			
Ag labor hours (drought)	2,514	2,513	-0.04
Ag labor hours (non-drought)	2,370	2,368	-0.08
Non-ag labor hours (drought)	1,419	1,420	0.07
Non-ag labor hours (non-drought)	1,522	1,525	0.20
Landless			
Ag labor hours (drought)	1,598	1,599	0.06
Ag labor hours (non-drought)	1,691	1,691	0.00
Non-ag labor hours (drought)	2,300	2,297	-0.13
Non-ag labor hours (non-drought)	2,214	2,215	0.04

To see the intuition for identifying sources of variation in the data that support the joint calibration, first consider the calibration for landed households. The leisure parameter, k_z^f, affects the total labor supplied in any state, i.e., the total labor hours in agriculture and non-agriculture sectors. A lower value of k_z^f implies more labor supplied, ceteris paribus. The total agriculture labor hours and non-agriculture labor hours moments help to identify the leisure parameter. Further, the attachment to land parameter, ψ_z^f, plays a key role in the decision of how much labor to allocate to the agricultural sector. A higher value of ψ_z^f implies a higher allocation of agriculture labor hours, holding everything else constant. Therefore, the agriculture labor hours moment helps to pin down the attachment to agriculture parameter. The calibration for landless households follows the same logic.

3.4.2 Interpretation of Parameters

The results of the calibration exercise are presented in the bottom part of Table 3.1. For the landed households, the attachment to agriculture parameter is 0.448 during drought periods and 0.2883 during non-drought periods. A higher value of the attachment to agriculture parameter denotes a higher preference for agricultural work. Drought is characterized by low rainfall and more generally, very dry weather. Rural landed households in India with attachment to their land require more workers in the farm to produce during the dry season (Fernando, 2020). This is reflected through the higher value of ψ_1^{nf}. Landless households show the same pattern for the attachment to agriculture parameter even though the magnitude differs; it is 0.4018 for the drought state and 0.3065 for the non-drought state. Taken together, these results suggest that moving out of agriculture entails adjustment costs or barriers to entry in the non-agriculture sector. A higher value for the attachment to agriculture parameter for landless households captures these abovementioned costs. These are higher during drought periods compared to non-drought periods implying a greater impediment to switching sector during unfavorable weather.

A higher value for the leisure parameter, k, implies a stronger preference for leisure. All else constant, a higher value for k implies less labor supplied. Since labor is measured in annual hours, the value of the parameter is very small. The value of leisure is higher during drought for landed and landless households. The unfavorable weather conditions in a tropical country during drought leads to a higher preference for leisure.

3.4.3 Non-Targeted Moments

Table 3.3 reports the non-targeted moments. Given that the model is a simple static model of labor allocation, the choice for non-targeted moments is limited. Total income and consumption are the most natural options for non-targeted moments. Household surveys are found to underestimate income data (Ravallion, 2003). Therefore, I compare consumption across drought and non-drought states for each type of household.

Table 3.3: Non-Targeted Moments

	Data Moments	Model Moments	% Difference
Landed HH			
Consumption (drought)	98,296	1,20,050	-18.12
Consumption (non-drought)	1,25,203	1,62,010	-22.72
Landless HH			
Consumption (drought)	78,657	84,112	-6.48
Consumption (non-drought)	95,869	1,00,990	-5.07

There is a non-trivial difference between the model's predicted consumption and data on consumption. This may be because the model does not include savings or investment. In the model, income is consumed in each period. Landed households exhibit a sizeable difference compared to landless households. This follows from the fact that landed households have a greater potential to accumulate wealth and, hence, save and invest more than the landless households.

3.5 Counterfactual Analysis

In this section, I use my calibrated model to conduct two quantitative exercises. First, using drought projections reported by the Intergovernmental Panel on Climate Change's (IPCC) 5th assessment report, I predict the agriculture labor share across landed and landless households for the next 30 years. Second, I calculate the government subsidy to non-agricultural labor that would be needed to achieve the goals of the well-publicized 'Make in India' campaign that was launched in September 2014 in response to one of the slowest growth periods in the history of the Indian economy.

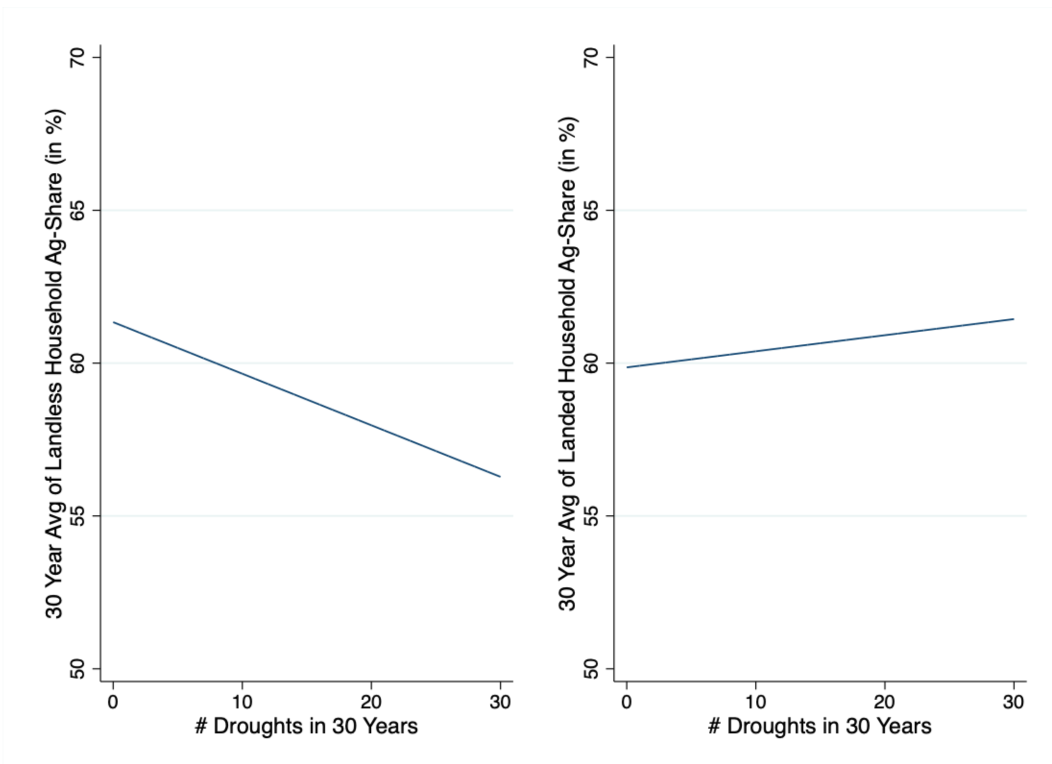
3.5.1 Drought Projections

My analysis is motivated by the fact that the IPCC's fifth assessment report (Pachauri *et al.*, 2014) predicts an increased risk of drought in Asia in both the near term (2030-2040) and the long term (2080-2100). The IPCC determined the extent of risk based on the following factors: magnitude; high probability or irreversibility of impacts; timing of impacts; persistent vulnerability or exposure contributing to risks; or limited potential to reduce risks through adaptation or mitigation. Based on these factors, the IPCC concluded that in the near term, drought risk is low to medium. In the long term, drought risk is a function of how much mean temperatures rise above the pre-industrial levels. For a 2°C increase in the mean global temperature, there could be low to medium risk and for a 4°C increase in mean global temperature, the risk could be medium to high, depending on the current level of adaptation.

Additionally, Gupta and Jain (2018) provides drought predictions through the end of the 21st century for India specifically using precipitation and temperature data obtained from Regional Climate Models (RCMs). Their analysis suggests that Northern and North-western India may face increasing frequency of drought in the near future (2021-2050). In the more distant future (2071-2100), most parts of India are expected to face increasing frequency of drought except for south-eastern regions such as Odisha, Chhattisgarh, and parts of Maharashtra, Madhya Pradesh, and Telangana. These projects match the IPCC report in predicting that India will witness more frequent droughts in the future.

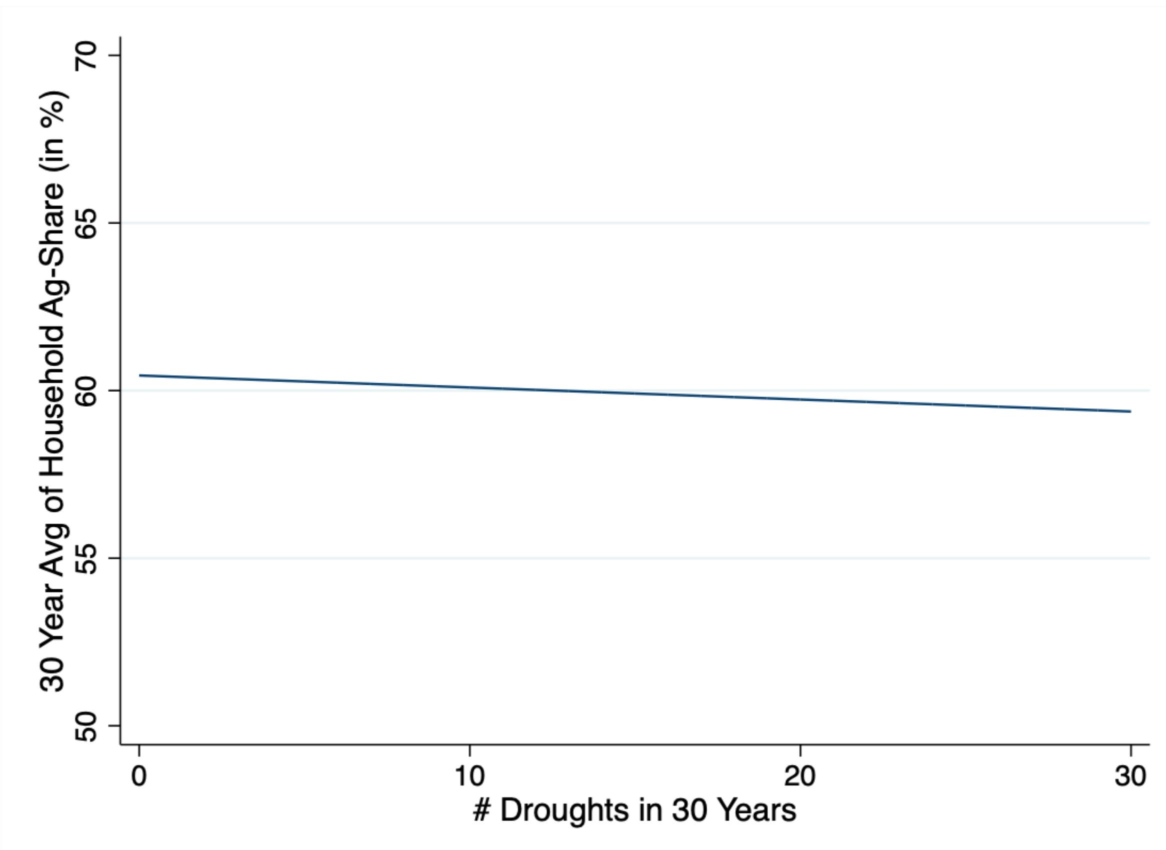
In order to understand how I use my model to simulate the effects of drought projections, it is important to first reiterate that my model matches the labor hours of landed and landless households across drought and non-drought states. It also matches the share of agricultural labor across household type and drought status. In a drought (non-drought) year, the share of agriculture labor in a landed household is 61.44% (59.86%) and in a landless household is 56.28% (61.34%). I use these results to define the baseline share of agriculture labor for the future. Then I use the model to predict how labor allocation would adjust to increased frequency of droughts in India.

Figure 3.4: Increase in Drought Frequency and Household Agriculture Labor Share



I consider a near-future period as a span of 30 years. I find that there were zero to nine yearly droughts between 1981-2010 across the districts in India. I use my model to calculate the average share of agriculture labor in rural households over the next 30 years, as the number of droughts experienced increases. Since I abstract away from changes in the severity of drought, which is also predicted to increase under climate change, the results of this experiment are best interpreted as the lower bounds on future changes in the labor shares.

Figure 3.5: Increase in Drought Frequency and Average Agriculture Labor Share



Note: The weights used in the weighted average are the proportion of each type of household. In the IHDS dataset, there were 70% landed households and 30% landless households.

Figure 3.4 shows the model-predicted relationship between drought and average agriculture labor share across landed and landless households. Given that the attachment to agriculture is stronger for the landed households during droughts, I find that the landed households will continue to allocate more household members to agriculture as the number of droughts increase. I find that the opposite is true for the landless households, who are more likely to move away from agriculture in response to drought. Focusing on the scaling of the y-axis shows that the share of agriculture labor does not change drastically with an increase in drought.

Given that droughts have opposite effects on the agricultural labor share for landed and landless households, I calculate the net effect on the average share of agriculture labor. Figure 3.5 plots the weighted average of agriculture labor share. The weights used are the proportion of landed (70 percent) and landless (30 percent) households in the sample. My model predicts that the share of agriculture labor falls marginally. Specifically, the difference in the 30-years average household agriculture share is only around 1-2% even when droughts occur every year. While the percentage change is small, it implies that 2.5 to 5 million individuals would exit the agricultural sector.

3.5.2 Targeting Structural Transformation

To encourage manufacturing in India, the Indian government launched the 'Make in India' initiative in 2014 during one of the slowest growth periods in modern history. The main objective of the initiative was to increase the manufacturing share of GDP to 25% in the near future (from 15% in 2013). Other objectives of the campaign were to facilitate investment, foster innovation, enhance skill development, protect intellectual property, and build manufacturing infrastructure. Most East Asian and Southeast Asian countries that have achieved high sustained growth rates experienced industrialization before the rise of the service sector. However, India faces different circumstances. The agriculture sector employs approximately half of the labor force, but contributes the least to the GDP. The service sector produces most of the output, around 60% of the GDP, and ranks second after agriculture in terms of employment. Manufacturing ranks behind services in terms of GDP share and employs the fewest individuals, playing a relatively minor role in the economy. Thus, in principle, a re-alignment of the labor force toward more productive activities could yield large benefits in terms of economic growth.

Table 3.4: Agriculture Employment Share Predictions

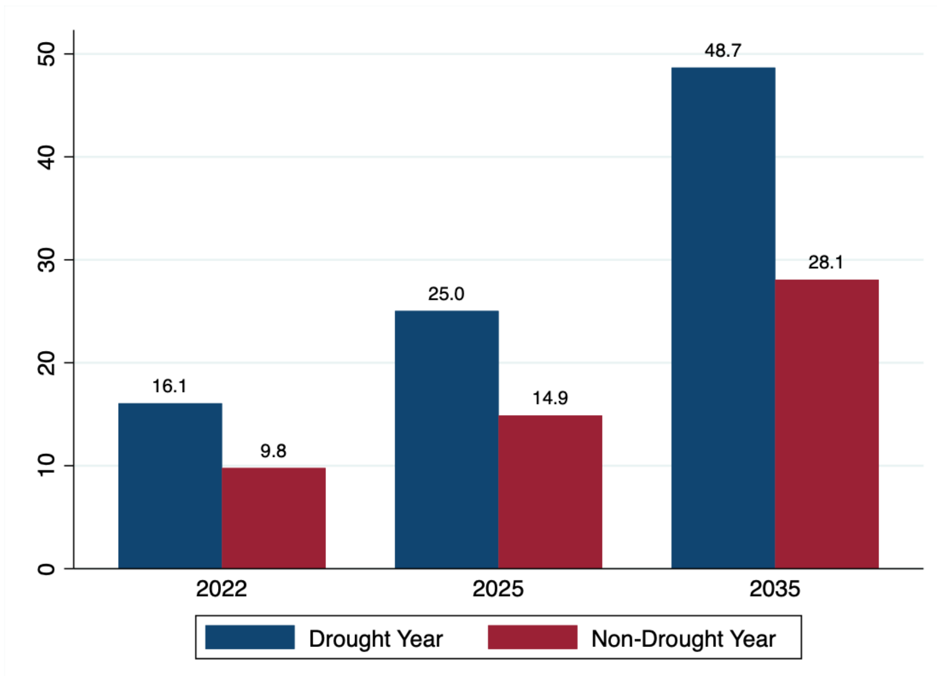
Year	Agriculture Employment Share	
	With Reform	Without Reform
2022	38	-
2025	33	37
2035	25	35

Source: Green (2015)

Green (2014, 2015) study the effectiveness of the 'Make in India' campaign and argue that the status quo would continue the prevalence of labor in low-productivity sectors and suggest that labor reforms are needed to achieve the targets. In Table 3.4, I report the agricultural labor shares suggested by Green (2015) under 'no reform' and 'with reform' scenarios. The no-reform scenario, predicts that the share of agricultural employment would be around 35% by 2035. The GDP share of manufacturing in 2035 would be 13%, way below the target of 25%. Some of the reforms suggested include reducing labor regulations, facilitating land acquisition, improving business-government relations, and providing public goods (judicial reforms, institutional reform, education, and infrastructure). Furthermore, Green (2015) predicts that structural transformation is necessary to achieve the 25% manufacturing share of GDP. Increasing the share of manufacturing in GDP to 25% from 15% in 2013, would imply reducing the agriculture employment share to as low as 25% by 2035.

The baseline share of agriculture labor in an average rural Indian household is 58.81% in a drought year and 60.65% in a non-drought year. Assuming the employment share of the country is reflected through the employment share of the average rural household, I quantify the non-agriculture wage increase required to achieve the structural transformation target needed for manufacturing to increase to 25% of GDP. This quantitative exercise determines the wage growth needed to achieve those targets. Alternatively, it can be interpreted as defining the government subsidy needed to finance the structural transformation target.

Figure 3.6: Percentage Increase in Non-Ag Wage to Achieve Ag-Share Targets



Note: Non-ag wage in 2011 was 33.43 for drought districts and 33.34 for non-drought districts. The target ag employment shares are 38%, 33%, and 25% in the years 2022, 2025 and 2035, respectively.

The nominal wage rates in 2011 for drought and non-drought districts are 33.43, and 33.34, respectively. I treat 2011 as the baseline year and solve for the wage subsidies required to meet the agricultural employment share targets under the current climate regime. Figure 3.6 shows the percentage increase in the non-agriculture wages needed to achieve the target agricultural employment shares for 2022, 2025, and 2035 for drought and non-drought years. To achieve a 38% of agriculture employment share in 2022, my model predicts that the non-agriculture wages would need to increase by 16.1% and 9.8% in drought and non-drought affected regions, respectively; to achieve a 33% agriculture employment share in 2025, the non-agriculture wage would need to increase by 25% and 14% in drought and non-drought affected regions, respectively; and to achieve a 25% agriculture employment share in 2035, the non-agriculture wage would need to increase by 49% and 28% in drought and non-drought affected regions, respectively.[3]

The climate-change scenario for increased drought frequency would require the growth of non-agriculture wages to be higher relative to the current climate regime. In particular, landed households would require higher non-agricultural wages to induce them to move out of agriculture in a drought scenario. For example, meeting the 2035 target would require increasing non-agriculture wages by 28.1% or 9.36 rupees. For 75% of the working population of 500 million individuals to be employed in non-agriculture, the government would need to pay a wage subsidy of 7.4 trillion (2011) rupees (101.25 billion USD). In the case of a drought, the increase in non-agriculture wages needs to be 48.7%, which is an increase of 16.27 rupees, equivalent to subsidies of 12.9 trillion rupees (175.77 billion USD). In terms of GDP, the required wage subsidy is around 5.5%, and is as high as 9.64% under the climate-change scenario.

[3] Appendix Table C.9 shows the level of non-agriculture wages needed to achieve the agriculture labor share targets across drought and non-drought years.

3.5.3 *Implications of Land Ownership for the Make in India Policy*

In the previous counterfactual simulation, the attachment to farmland exhibited by landowners gives them a strong incentive to remain in the agricultural sector, even when the non-agricultural wage rises. Recall that the attachment to agriculture parameter in the landowner's utility function represents the psychological cost of leaving their farmland along with other barriers to entry in rural labor markets. In order to disentangle the effects of land ownership from the other barriers to entry, I repeat the simulation from the last section after first assigning the attachment to agriculture parameters of the landless households to the landed households. Then I recompute the non-ag wages that would be required to achieve the target agricultural labor share needed to meet the goals of the 'Make in India' policy. The following table compares two cases: the default case where each type of household has its respective attachment to agriculture parameters shown in Table 3.1 and the counterfactual case where landed households are assigned the attachment to agriculture parameters of the landless households.

Table 3.5: Non-Ag Wages Needed to Achieve 'Make in India' Policy Goals in the Absense of Attachment to Farmland

		Non-Agricultural Wages			
Year	Target	Ag parameter: default		Ag parameter: counterfactual	
	Ag Share	Drought	Non-Drought	Drought	Non-Drought
2022	38%	38.80	36.60	37.00	37.30
2025	33%	41.80	38.80	39.70	39.20
2035	25%	49.70	42.70	46.50	43.50

This exercise yields two interesting results as shown in Table 3.5. First, I find that the non-agricultural wages needed to achieve the target agricultural labor share for each of the years 2022, 2025, and 2035 without climate change increases marginally compared to the default scenario. This is because the attachment to agriculture for landless households is higher compared to landed households in a non-drought year. This could possibly represent barriers to labor market entry such as non-agricultural skill. Second, the non-agricultural wages are lower in a drought year in the counterfactual case compared to the default case. This is true because drought causes landless households to move out of agriculture at a higher rate. As both types of households now have the same level of attachment to land during drought, the non-agricultural wages needed to achieve the target agricultural employment share is lower.

To achieve the target share for 2035, the government would need to pay the non-agricultural workers 46.50 rupees (in 2011 terms) per hour under climate change and 43.50 rupees (in 2011 terms) per hour without climate change. To implement this policy for a working population of 500 million individuals, the government would need to pay a wage subsidy of 8.1 trillion (2011) rupees (110.97 billion USD). In case of drought, the wage subsidy is equivalent to 10.4 trillion (2011) rupees (142.48 billion USD). In comparison, in the default case, the wage subsidy was 7.4 trillion (2011) rupees without climate change and 12.9 trillion (2011) rupees under climate change. In a situation where the country may experience an equal frequency of drought and non-drought years, the average cost for the wage subsidy would be 10.15 trillion (2011) rupees under the default situation and 9.25 trillion (2011) rupees under the counterfactual situation. These findings suggest that attachment to agriculture exhibited by landed households would substantially increase the government's cost of meeting the "Make in India" policy goals.

3.6 Conclusion

This chapter extended the "envirodevonomics" literature by investigating how land ownership influences the effects of drought on occupation choices by rural Indian households. In chapter 2, I used high-resolution climate data combined with a survey panel data on households to conclude that drought reduces the share of agricultural labor hours in rural households overall, but this effect is driven by landless households that exit agriculture at a faster rate. Landed households increase agricultural labor only slightly in response to drought. These findings imply that land ownership is an important barrier to sectoral labor reallocation.

In chapter 3, I developed a partial equilibrium model of labor allocation across agriculture and non-agriculture sectors. I calibrated the model to match the labor hours across landed and landless households across drought and non-drought states and used the model to predict how landed and landless households will respond to future climate change that increases drought frequency. My results imply that landless households will tend to reduce their agricultural labor share whereas landed households will allocate marginally higher labor to agriculture as drought becomes more frequent. The net effect is a 1% to 2% decline in agricultural labor, but this marginal change is equivalent to 2.5 to 5 million people leaving agriculture.

I also predicted the non-agriculture wage subsidy that would be needed to achieve the structural transformation targets of the 'Make in India' campaign. My findings suggest that to achieve an agriculture labor share of 25% by the year 2035 in the absence of climate change, the non-agriculture wage would need to increase by 28.1% relative to 2011. However, this subsidy would need to be increased to as much as 48.7% in the presence of increased droughts under climate change. Overall, the results of this paper highlight the importance of institutional and cultural barriers in the adaptation process of rural households exposed to climate change.

REFERENCES

Abadie, A., S. Athey, G. W. Imbens and J. Wooldridge, “When should you adjust standard errors for clustering?”, Tech. rep., National Bureau of Economic Research (2017).

Agarwal, B. and A. Bina, *A field of one’s own: Gender and land rights in South Asia*, vol. 58 (Cambridge University Press, 1994).

Ahituv, A. and A. Kimhi, “Off-farm work and capital accumulation decisions of farmers over the life-cycle: the role of heterogeneity and state dependence”, Journal of Development Economics **68**, 2, 329–353 (2002).

Altonji, J. G., F. Hayashi and L. J. Kotlikoff, “Is the extended family altruistically linked? direct tests using micro data”, The American Economic Review pp. 1177–1198 (1992).

Bandyopadhyay, S. and E. Koufias, *Rainfall variability, occupational choice, and welfare in rural Bangladesh* (The World Bank, 2012).

Banerjee, A. V. and E. Duflo, “Growth theory through the lens of development economics”, Handbook of economic growth **1**, 473–552 (2005).

Banerjee, A. V. and E. Duflo, “The economic lives of the poor”, Journal of economic perspectives **21**, 1, 141–168 (2007).

Banzhaf, H. S. and E. McCormick, *2. Moving beyond Cleanup: Identifying the Crucibles of Environmental Gentrification* (Stanford University Press, 2020).

Banzhaf, H. S. and R. P. Walsh, “Do people vote with their feet? an empirical test of tiebout”, American economic review **98**, 3, 843–63 (2008).

Banzhaf, H. S. and R. P. Walsh, “Segregation and tiebout sorting: The link between place-based investments and neighborhood tipping”, Journal of Urban Economics **74**, 83–98 (2013).

Been, V., “Locally undesirable land uses in minority neighborhoods: Disproportionate siting or market dynamics?”, Yale Law Journal pp. 1383–1422 (1994).

Been, V. and F. Gupta, “Coming to the nuisance or going to the barrios? a longitudinal analysis of environmental justice claims”, Ecology law quarterly **24**, 1, 1–56 (1997).

Benonnier, T., K. Millock and V. Taraz, “Climate change, migration, and irrigation”, (2019).

Bertrand, M., E. Duflo and S. Mullainathan, “How much should we trust differences-in-differences estimates?”, The Quarterly journal of economics **119**, 1, 249–275 (2004).

Bhat, A. K. and R. Dhruvarajan, "Ageing in india: drifting intergenerational relations, challenges and options", Ageing and Society **21**, 621 (2001).

Bianchi, M. and M. Bobba, "Liquidity, risk, and occupational choices", Review of Economic Studies **80**, 2, 491–511 (2013).

Binswanger, H. P., "Attitudes toward risk: Experimental measurement in rural india", American journal of agricultural economics **62**, 3, 395–407 (1980).

Bishop, K. C., J. D. Ketcham and N. V. Kuminoff, "Hazed and confused: Air pollution, dementia, and financial decision making", NBER Working Paper No **24970** (2018).

Bisht, D. S., V. Sridhar, A. Mishra, C. Chatterjee and N. S. Raghuwanshi, "Drought characterization over india under projected climate scenario", International Journal of Climatology **39**, 4, 1889–1911 (2019).

Blattman, C., N. Fiala and S. Martinez, "Credit constraints, occupational choice, and the process of development: long run evidence from cash transfers in uganda", (2013).

Bohra-Mishra, P., M. Oppenheimer and S. M. Hsiang, "Nonlinear permanent migration response to climatic variations but minimal response to disasters", Proceedings of the National Academy of Sciences **111**, 27, 9780–9785 (2014).

Bryan, G., S. Chowdhury and A. M. Mobarak, "Underinvestment in a profitable technology: The case of seasonal migration in bangladesh", Econometrica **82**, 5, 1671–1748 (2014).

Burgess, R. and D. Donaldson, "Can openness mitigate the effects of weather shocks? evidence from india's famine era", American Economic Review **100**, 2, 449–53 (2010).

Carleton, T. A., "Crop-damaging temperatures increase suicide rates in india", Proceedings of the National Academy of Sciences **114**, 33, 8746–8751 (2017).

Colmer, J., "Weather, labor reallocation and industrial production: evidence from india", (2018).

Crowder, K. and L. Downey, "Interneighborhood migration, race, and environmental hazards: Modeling microlevel processes of environmental inequality", American Journal of Sociology **115**, 4, 1110–1149 (2010).

Currie, J., M. Greenstone and E. Morettia, "Superfund cleanups and infant health", The American economic review **101**, 3, 435–441 (2011).

Dallmann, I. and K. Millock, "Climate variability and inter-state migration in india", CESifo Economic Studies **63**, 4, 560–594 (2017).

Das, N., R. Yasmin, J. Ara, M. Kamruzzaman, P. Davis, J. Behrman, S. Roy and A. R. Quisumbing, "How do intrahousehold dynamics change when assets are transferred to women? evidence from bracs challenging the frontiers of poverty reductiontargeting the ultra poor program in bangladesh", (2013).

Debroy, B., "Indias segmented labour markets, inter-state differences, and the scope for labour reforms", Economic Freedom of the States of India (2013).

Deininger, K., S. Jin and H. K. Nagarajan, *Determinants and consequences of land sales market participation: Panel evidence from India* (The World Bank, 2007a).

Deininger, K., S. Jin and H. K. Nagarajan, *Efficiency and equity impacts of rural land rental restrictions: Evidence from India* (The World Bank, 2007b).

Depro, B., C. Timmins and M. ONeil, "White flight and coming to the nuisance: can residential mobility explain environmental injustice?", Journal of the Association of Environmental and resource Economists **2**, 3, 439–468 (2015).

Depro, B. M., C. Timmins and M. O'Neil, "Meeting urban housing needs: Do people really come to the nuisance?", Working Paper 18109, National Bureau of Economic Research (2012).

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor and J. Reif, "The mortality and medical costs of air pollution: Evidence from changes in wind direction", American Economic Review **109**, 12, 4178–4219 (2019).

Duflo, E. and C. Udry, "Intrahousehold resource allocation in cote d'ivoire: Social norms, separate accounts and consumption choices", Working Paper 10498, National Bureau of Economic Research (2004).

Emerick, K., "Agricultural productivity and the sectoral reallocation of labor in rural india", Journal of Development Economics **135**, 488–503 (2018).

Emerick, K., A. de Janvry, E. Sadoulet and M. H. Dar, "Technological innovations, downside risk, and the modernization of agriculture", American Economic Review **106**, 6, 1537–61 (2016).

Epple, D. and H. Sieg, "Estimating equilibrium models of local jurisdictions", Journal of political economy **107**, 4, 645–681 (1999).

Fafchamps, M. and F. Gubert, "The formation of risk sharing networks", Journal of development Economics **83**, 2, 326–350 (2007).

Fernando, A. N., "Shackled to the soil? inherited land, birth order, and labor mobility", Journal of Human Resources pp. 0219–10014R2 (2020).

Ferrara, E. L., "Kin groups and reciprocity: A model of credit transactions in ghana", American economic review **93**, 5, 1730–1751 (2003).

Foster, A. D. and M. R. Rosenzweig, "Household division and rural economic growth", The Review of Economic Studies **69**, 4, 839–869 (2002).

Gamper-Rabindran, S. and C. Timmins, "Hazardous waste cleanup, neighborhood gentrification, and environmental justice: Evidence from restricted access census block data", The American Economic Review **101**, 3, 620–624 (2011).

Gollin, D. and R. Rogerson, *2. Agriculture, Roads, and Economic Development in Uganda* (University of Chicago Press, 2016).

Gray, C. and V. Mueller, "Drought and population mobility in rural ethiopia", World development **40**, 1, 134–145 (2012).

Green, R., "Structural change forecasts for india: How big of a bang can a big bang have?", (2015).

Green, R. A., "Can make in india make jobs? the challenges of manufacturing growth and high-quality job creation in india", (2014).

Greenstone, M. and J. Gallagher, "Does hazardous waste matter? evidence from the housing market and the superfund program", The Quarterly Journal of Economics **123**, 3, 951–1003 (2008).

Greenstone, M. and B. K. Jack, "Envirodevonomics: A research agenda for an emerging field", Journal of Economic Literature **53**, 1, 5–42 (2015).

Greenwood, J., Z. Hercowitz and G. W. Huffman, "Investment, capacity utilization, and the real business cycle", The American Economic Review pp. 402–417 (1988).

Grimard, F., "Household consumption smoothing through ethnic ties: evidence from cote d'ivoire", Journal of development Economics **53**, 2, 391–422 (1997).

Gupta, V. and M. K. Jain, "Investigation of multi-model spatiotemporal mesoscale drought projections over india under climate change scenario", Journal of Hydrology **567**, 489–509 (2018).

Herrendorf, B. and T. Schoellman, "Wages, human capital, and barriers to structural transformation", American Economic Journal: Macroeconomics **10**, 2, 1–23 (2018).

Ito, T. and T. Kurosaki, "Weather risk, wages in kind, and the off-farm labor supply of agricultural households in a developing country", American journal of agricultural economics **91**, 3, 697–710 (2009).

Jayachandran, S., "Selling labor low: Wage responses to productivity shocks in developing countries", Journal of political Economy **114**, 3, 538–575 (2006).

Jessoe, K., D. T. Manning and J. E. Taylor, "Climate change and labour allocation in rural mexico: Evidence from annual fluctuations in weather", The Economic Journal **128**, 608, 230–261 (2018).

Jodhka, S. S., "Beyond'crises': rethinking contemporary punjab agriculture", Economic and Political Weekly pp. 1530–1537 (2006).

Karlan, D. and J. Zinman, "Observing unobservables: Identifying information asymmetries with a consumer credit field experiment", Econometrica **77**, 6, 1993–2008 (2009).

Kiel, K. A. and M. Williams, "The impact of superfund sites on local property values: Are all sites the same?", Journal of urban Economics **61**, 1, 170–192 (2007).

Kochar, A., "Does lack of access to formal credit constrain agricultural production? evidence from the land tenancy market in rural india", American Journal of Agricultural Economics **79**, 3, 754–763 (1997).

Kochar, A., "Smoothing consumption by smoothing income: hours-of-work responses to idiosyncratic agricultural shocks in rural india", Review of Economics and Statistics **81**, 1, 50–61 (1999).

Liu, M. Y., Y. Shamdasani and V. Taraz, "Climate change, structural transformation, and infrastructure: Evidence from india", (2020).

Liu, Y., "Pushed out or pulled in? participation in non-farm activities in rural china", China Agricultural Economic Review **9**, 1, 111–129 (2017).

Matshe, I. and T. Young, "Off-farm labour allocation decisions in small-scale rural households in zimbabwe", Agricultural Economics **30**, 3, 175–186 (2004).

Maystadt, J.-F., V. Mueller and A. Sebastian, "Environmental migration and labor markets in nepal", Journal of the Association of Environmental and Resource Economists **3**, 2, 417–452 (2016).

Meghir, C., M. Mobarak, C. Mommaerts and M. Morten, "Migration and consumption insurance in bangladesh", in "2015 Meeting Papers", No. 190 (2017).

Morduch, J., "Income smoothing and consumption smoothing", Journal of economic perspectives **9**, 3, 103–114 (1995).

Morello-Frosch, R., M. Pastor Jr, C. Porras and J. Sadd, "Environmental justice and regional inequality in southern california: implications for future research.", Environmental Health Perspectives **110**, Suppl 2, 149 (2002).

Moretti, E., "Do wages compensate for risk of unemployment? parametric and semiparametric evidence from seasonal jobs", Journal of Risk and Uncertainty **20**, 1, 45–66 (2000).

Morris, S. and A. Pandey, "Towards reform of land acquisition framework in india", Economic and Political Weekly pp. 2083–2090 (2007).

Morten, M., "Temporary migration and endogenous risk sharing in village india", Journal of Political Economy **127**, 1, 1–46 (2019).

Muehlenbachs, L., E. Spiller and C. Timmins, "The housing market impacts of shale gas development", American Economic Review **105**, 12, 3633–59 (2015).

Mueller, V., C. Gray and K. Kosec, "Heat stress increases long-term human migration in rural pakistan", Nature climate change **4**, 3, 182 (2014).

Munshi, K. and M. Rosenzweig, "Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy", American Economic Review **96**, 4, 1225–1252 (2006).

Munshi, K. and M. Rosenzweig, "Networks and misallocation: Insurance, migration, and the rural-urban wage gap", American Economic Review **106**, 1, 46–98 (2016).

Noack, F., M.-C. Riekhof and S. Di Falco, "Droughts, biodiversity, and rural incomes in the tropics", Journal of the Association of Environmental and Resource Economists **6**, 4, 823–852 (2019).

Oakes, J. M., D. L. Anderton and A. B. Anderson, "A longitudinal analysis of environmental equity in communities with hazardous waste facilities", Social Science Research **25**, 2, 125–148 (1996).

Pachauri, R. K., M. R. Allen, V. R. Barros, J. Broome, W. Cramer, R. Christ, J. A. Church, L. Clarke, Q. Dahe, P. Dasgupta *et al.*, *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change* (Ipcc, 2014).

Park, C., "Risk pooling between households and risk-coping measures in developing countries: Evidence from rural bangladesh", Economic Development and Cultural Change **54**, 2, 423–457 (2006).

Pastor, M., J. Sadd and J. Hipp, "Which came first? toxic facilities, minority move-in, and environmental justice", Journal of urban affairs **23**, 1, 1–21 (2001).

Perloff, J. M., "The impact of wage differentials on choosing to work in agriculture", American Journal of Agricultural Economics **73**, 3, 671–680 (1991).

Persico, C., D. Figlio and J. Roth, "The developmental consequences of superfund sites", Journal of Labor Economics **38**, 4, 1055–1097 (2020).

Prasanna, V., "Impact of monsoon rainfall on the total foodgrain yield over india", Journal of earth system science **123**, 5, 1129–1145 (2014).

Ravallion, M., "Measuring aggregate welfare in developing countries: How well do national accounts and surveys agree?", Review of Economics and Statistics **85**, 3, 645–652 (2003).

Ravallion, M., S. Chaudhuri *et al.*, "Risk and insurance in village india: Comment", ECONOMETRICA-EVANSTON ILL- **65**, 171–184 (1997).

Rose, E., "Ex ante and ex post labor supply response to risk in a low-income area", Journal of Development Economics **64**, 2, 371–388 (2001).

Rosenzweig, M. R., "Risk, implicit contracts and the family in rural areas of low-income countries", The Economic Journal **98**, 393, 1148–1170 (1988).

Rosenzweig, M. R. and O. Stark, “Consumption smoothing, migration, and marriage: Evidence from rural india”, Journal of political Economy **97**, 4, 905–926 (1989).

Rosenzweig, M. R. and K. I. Wolpin, “Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in india”, Journal of political economy **101**, 2, 223–244 (1993).

Schlenker, W., W. M. Hanemann and A. C. Fisher, “Will us agriculture really benefit from global warming? accounting for irrigation in the hedonic approach”, American Economic Review **95**, 1, 395–406 (2005).

Sengupta, K., “Risk aversion in rural india”, (2011).

Shaikh, S. L. and J. B. Loomis, “An investigation into the presence and causes of environmental inequity in denver, colorado”, The Social Science Journal **36**, 1, 77–92 (1999).

Sharma, A., “The changing agricultural demography of india: evidence from a rural youth perception survey”, International Journal of Rural Management **3**, 1, 27–41 (2007).

Shazli, T. and A. Munir, “Female work participation in handloom industry-a case study of mubarakpur town, azamgarh district, up”, Journal of Education & Social Policy **1**, 1, 76–83 (2014).

Skoufias, E., “Household resources, transaction costs, and adjustment through land tenancy”, Land Economics pp. 42–56 (1995).

Skoufias, E., S. Bandyopadhyay and S. Olivieri, “Occupational diversification as an adaptation to rainfall variability in rural india”, Agricultural Economics **48**, 1, 77–89 (2017).

Stark, O. and R. E. Lucas, “Migration, remittances, and the family”, Economic development and cultural change **36**, 3, 465–481 (1988).

Taraz, V., “Can farmers adapt to higher temperatures? evidence from india”, World Development **112**, 205–219 (2018).

Thomas, D., F. Witoelar, E. Frankenberg, B. Sikoki, J. Strauss, C. Sumantri and W. Suriastini, “Cutting the costs of attrition: Results from the indonesia family life survey”, Journal of Development Economics **98**, 1, 108–123 (2012).

Tiebout, C. M., “A pure theory of local expenditures”, Journal of political economy **64**, 5, 416–424 (1956).

Topalova, P., “Factor immobility and regional impacts of trade liberalization: Evidence on poverty from india”, American Economic Journal: Applied Economics **2**, 4, 1–41 (2010).

Townsend, R. M., *Financial systems in developing economies: Growth, inequality and policy evaluation in Thailand* (Oxford University Press, 2011).

Townsend, R. M. *et al.*, "Risk and insurance in village india", ECONOMETRICA-EVANSTON ILL- **62**, 539–539 (1994).

Tunali, I., "Rationality of migration", International Economic Review **41**, 4, 893–920 (2000).

Udry, C., "Risk and insurance in a rural credit market: An empirical investigation in northern nigeria", The Review of Economic Studies **61**, 3, 495–526 (1994).

Viswanathan, B. and K. K. Kumar, "Weather, agriculture and rural migration: evidence from state and district level migration in india", Environment and Development Economics **20**, 4, 469 (2015).

Welch, J. R., J. R. Vincent, M. Auffhammer, P. F. Moya, A. Dobermann and D. Dawe, "Rice yields in tropical/subtropical asia exhibit large but opposing sensitivities to minimum and maximum temperatures", Proceedings of the National Academy of Sciences **107**, 33, 14562–14567 (2010).

Witoelar, F., "Risk sharing within the extended family: evidence from the indonesia family life survey", Economic Development and Cultural Change **62**, 1, 65–94 (2013).

Wooldridge, J. M., "Cluster-sample methods in applied econometrics", American Economic Review **93**, 2, 133–138 (2003).

Zimmermann, L., "Why guarantee employment? evidence from a large indian public-works program", GLO Discussion Paper 504, Essen (2020).

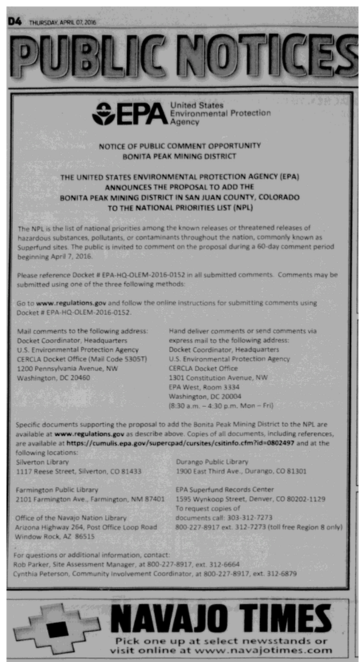
APPENDIX A

ENVIRONMENTAL JUSTICE FOR SENIORS: EVIDENCE FROM THE SUPERFUND PROGRAM

A.1 Background

The Environmental Protection Agency (EPA) announces stages of the clean-up process of a hazardous site and assigns a period for public comments about the proposed plan. The following two subsections provide examples of those announcements. The following public notice in the *Navajo Times* is an example of an information shock about the environmental quality of a neighborhood.[1] People reading this information decides to move out or stay in the neighborhood with the Superfund site.

Figure A.1: Example of Negative Information Shock



D4 THURSDAY, APRIL 07, 2016

PUBLIC NOTICES

EPA United States Environmental Protection Agency

NOTICE OF PUBLIC COMMENT OPPORTUNITY
BONITA PEAK MINING DISTRICT

THE UNITED STATES ENVIRONMENTAL PROTECTION AGENCY (EPA)
ANNOUNCES THE PROPOSAL TO ADD THE
BONITA PEAK MINING DISTRICT IN SAN JUAN COUNTY, COLORADO
TO THE NATIONAL PRIORITIES LIST (NPL)

The NPL is the list of national priorities among the known releases or threatened releases of hazardous substances, pollutants, or contaminants throughout the nation, commonly known as Superfund sites. The public is invited to comment on the proposal during a 60-day comment period beginning April 7, 2016.

Please reference Docket # EPA-HQ-OLEM-2016-0152 in all submitted comments. Comments may be submitted using one of the three following methods:

Go to **www.regulations.gov** and follow the online instructions for submitting comments using Docket # EPA-HQ-OLEM-2016-0152.

Mail comments to the following address:
Docket Coordinator, Headquarters
U.S. Environmental Protection Agency
CERCLA Docket Office (Mail Code 5305T)
1200 Pennsylvania Avenue, NW
Washington, DC 20460

Hand deliver comments or send comments via express mail to the following address:
Docket Coordinator, Headquarters
U.S. Environmental Protection Agency
CERCLA Docket Office
1301 Constitution Avenue, NW
EPA West, Room 3334
Washington, DC 20004
(8:30 a.m. – 4:30 p.m. Mon – Fri)

Specific documents supporting the proposal to add the Bonita Peak Mining District to the NPL are available at **www.regulations.gov** as describe above. Copies of all documents, including references, are available at **https://cumulis.epa.gov/supercpad/cursites/csitinfo.cfm?id=0802497** and at the following locations:

Silverton Library
1117 Reese Street, Silverton, CO 81433

Durango Public Library
1900 East Third Ave., Durango, CO 81301

Farmington Public Library
2101 Farmington Ave., Farmington, NM 87401

EPA Superfund Records Center
1595 Wynkoop Street, Denver, CO 80202-1129
To request copies of documents call: 303-312-7273
800-227-8917 ext. 312-7273 (toll free Region 8 only)

Office of the Navajo Nation Library
Arizona Highway 264, Post Office Loop Road
Window Rock, AZ 86515

For questions or additional information, contact:
Rob Parker, Site Assessment Manager, at 800-227-8917, ext. 312-6664
Cynthia Peterson, Community Involvement Coordinator, at 800-227-8917, ext. 312-6879

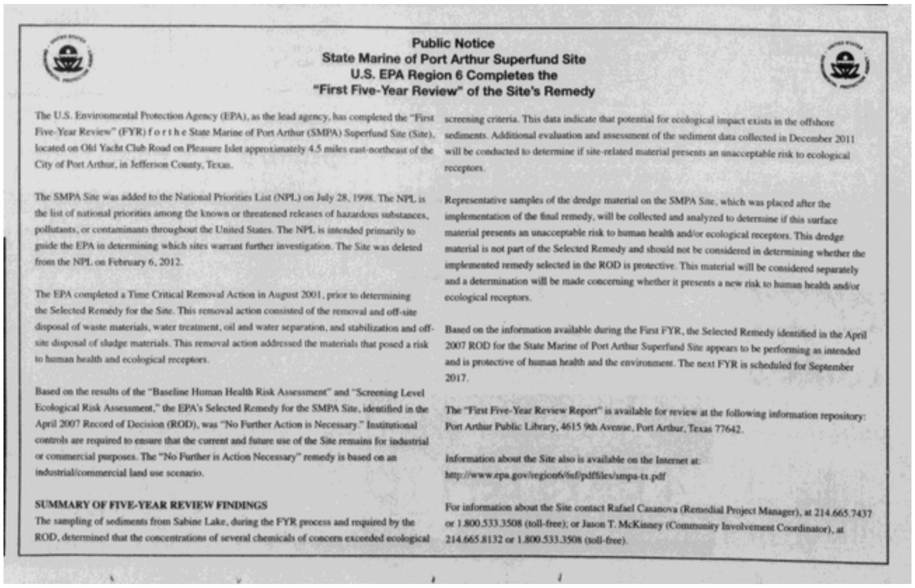
NAVAJO TIMES
Pick one up at select newsstands or visit online at www.navajotimes.com

Once, the site is cleaned up, EPA also announces the end of the cleaning process by declaring the date of removal of the site from the National Priority List. This illustrates positive information shock about the environmental quality of the place.[2]

[1] https://newspaperarchive.com/ provided the public notices in local newspapers.

[2] Disclaimer: The Notice about deletion does provide some additional information other than site deletion from NPL.

Figure A.2: Example of Positive Information Shock



Public Notice
State Marine of Port Arthur Superfund Site
U.S. EPA Region 6 Completes the
"First Five-Year Review" of the Site's Remedy

The U.S. Environmental Protection Agency (EPA), as the lead agency, has completed the "First Five-Year Review" (FYR) f o r t h e State Marine of Port Arthur (SMPA) Superfund Site (Site), located on Old Yacht Club Road on Pleasure Islet approximately 4.5 miles east-northeast of the City of Port Arthur, in Jefferson County, Texas.

The SMPA Site was added to the National Priorities List (NPL) on July 28, 1998. The NPL is the list of national priorities among the known or threatened releases of hazardous substances, pollutants, or contaminants throughout the United States. The NPL is intended primarily to guide the EPA in determining which sites warrant further investigation. The Site was deleted from the NPL on February 6, 2012.

The EPA completed a Time Critical Removal Action in August 2001, prior to determining the Selected Remedy for the Site. This removal action consisted of the removal and off-site disposal of waste materials, water treatment, oil and water separation, and stabilization and off-site disposal of sludge materials. This removal action addressed the materials that posed a risk to human health and ecological receptors.

Based on the results of the "Baseline Human Health Risk Assessment" and "Screening Level Ecological Risk Assessment," the EPA's Selected Remedy for the SMPA Site, identified in the April 2007 Record of Decision (ROD), was "No Further Action is Necessary." Institutional controls are required to ensure that the current and future use of the Site remains for industrial or commercial purposes. The "No Further is Action Necessary" remedy is based on an industrial/commercial land use scenario.

SUMMARY OF FIVE-YEAR REVIEW FINDINGS
The sampling of sediments from Sabine Lake, during the FYR process and required by the ROD, determined that the concentrations of several chemicals of concern exceeded ecological screening criteria. This data indicate that potential for ecological impact exists in the offshore sediments. Additional evaluation and assessment of the sediment data collected in December 2011 will be conducted to determine if site-related material presents an unacceptable risk to ecological receptors.

Representative samples of the dredge material on the SMPA Site, which was placed after the implementation of the final remedy, will be collected and analyzed to determine if this surface material presents an unacceptable risk to human health and/or ecological receptors. This dredge material is not part of the Selected Remedy and should not be considered in determining whether the implemented remedy selected in the ROD is protective. This material will be considered separately and a determination will be made concerning whether it presents a new risk to human health and/or ecological receptors.

Based on the information available during the First FYR, the Selected Remedy identified in the April 2007 ROD for the State Marine of Port Arthur Superfund Site appears to be performing as intended and is protective of human health and the environment. The next FYR is scheduled for September 2017.

The "First Five-Year Review Report" is available for review at the following information repository: Port Arthur Public Library, 4615 9th Avenue, Port Arthur, Texas 77642.

Information about the Site also is available on the Internet at:
http://www.epa.gov/region6/6sf/pdffiles/smpa-tx.pdf

For information about the Site contact Rafael Casanova (Remedial Project Manager), at 214.665.7437 or 1.800.533.3508 (toll-free); or Jason T. McKinney (Community Involvement Coordinator), at 214.665.8132 or 1.800.533.3508 (toll-free).

A.2 Summary Statistics and Research Design

Figure A.3: Location - Listed Superfund Sites

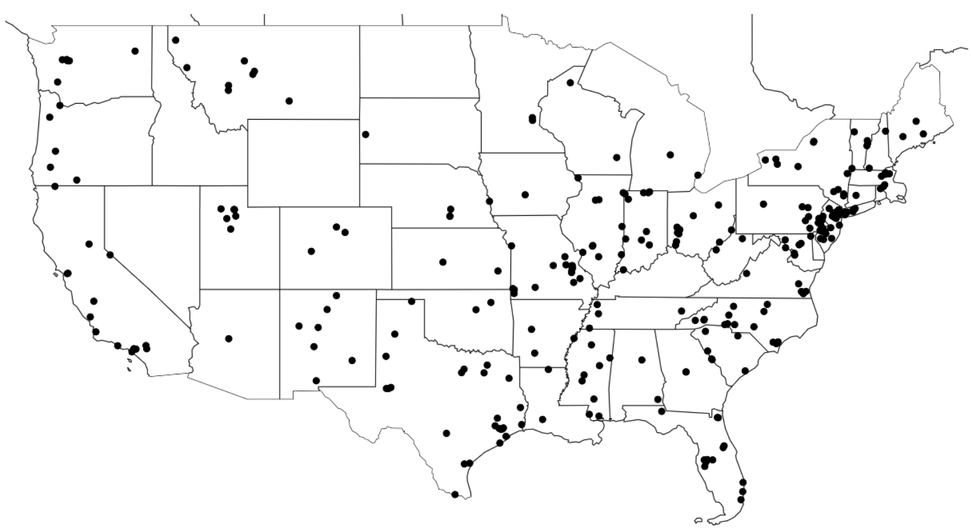


Figure A.4: Location - Deleted Superfund Sites

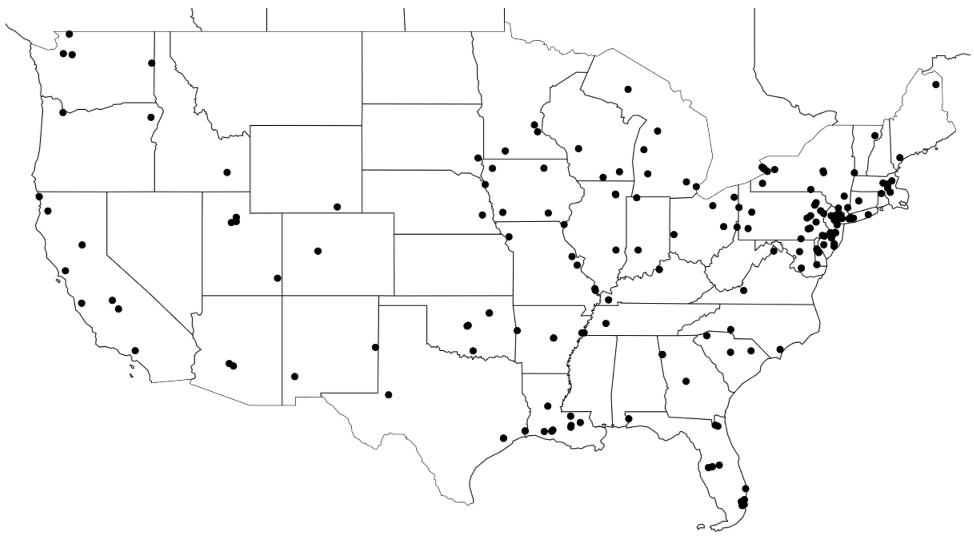
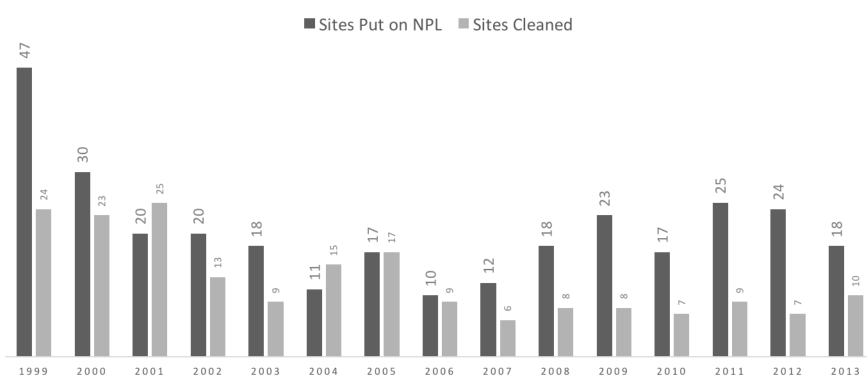


Figure A.5: Temporal Variation in Listing and Deleting Superfund Sites



Research Design: The following empirical specification is used to ascertain the boundary separating treatment and control groups (Muehlenbachs *et al.*, 2015).

$$move_{ijt} = \mathbf{B_{ijt}}\beta_1 + \mathbf{I_{ijt}}\beta_2 + \alpha_j * \delta_t + \sum_{k=1}^{k=4}(\beta_{3k} * D_{kijt}) + \epsilon_{ijt} \tag{A.1}$$

In the above specification, $move_{ijt}$ equals 1 if individual i living in county j in year t moves, zero otherwise. I control for block-level characteristics, $\mathbf{B_{ijt}}$, such as house value, income, percentage renter-occupied properties, percentage of seniors, racial groups, and for individual-level characteristics, $\mathbf{I_{ijt}}$ such as gender, birth year, presence of chronic conditions, etc. To account for unobservable spatial and temporal changes that may affect moving decisions, I include county (α_j) and year (δ_t) fixed effects. D_{1ijt} is a dummy variable taking the value 1 if individual i is located within 2 miles of a site, 0 otherwise. Similarly, D_{2ijt} takes value 1 if the individual is located within 2 to 4 miles of a site, zero otherwise, and so on. The key parameters of interest here are β_{3k}s.

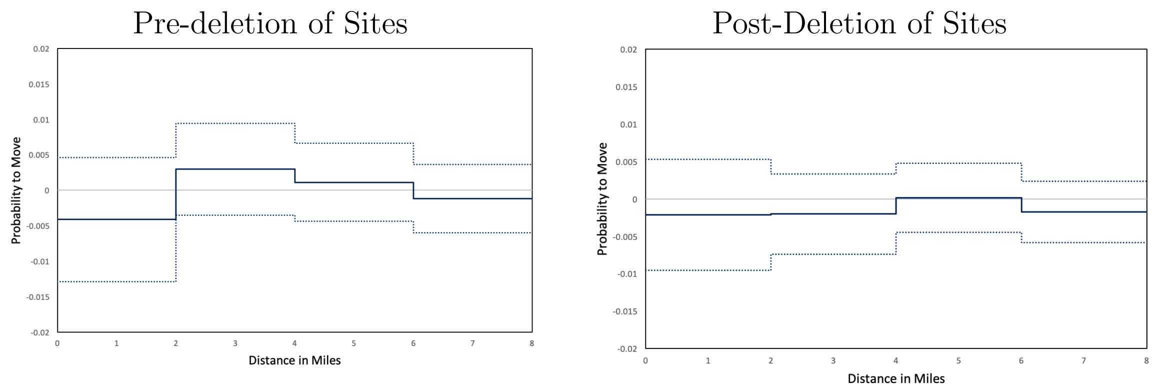
In order to identify the individual's location with respect to all the Superfund sites that were listed or deleted in a particular year, I used the Haversine Great Circle formula. The CMS dataset provided annual zip-codes for every beneficiary. The Geolytics dataset mapped every zip-code to a geographic co-ordinate. The latitude and longitude of every Superfund site are documented by the EPA. The calculation of distances between every possible site and every individual in the data over the fifteen years of study was facilitated by the availability of geo-coordinates for every Superfund site as well as for the location of individuals. For example, consider two pairs of latitude and longitude: (ϕ_1, ψ_1) and (ϕ_2, ψ_2), expressed in radians. The distance (d) in miles between the two points is given by:

$$
\begin{aligned}
a &= \sin^2(\frac{\triangle\phi}{2}) + \cos(\phi_1) * \cos(\phi_2) * \cos(\frac{\triangle\psi}{2}) \\
c &= 2 * \text{atan2}(\sqrt{a}, \sqrt{1-a}) \\
d &= R * c
\end{aligned}
$$

where, R = 3959 miles, the mean radius of the earth.

Using the distances computed, individuals were assigned to the treatment and control groups. To elaborate, consider the year 1999. 47 sites were listed on the NPL in 1999. The distance between every individual's location in 1999 and the 47 listed sites was calculated. The individuals whose distances happen to be less than equal to 2 miles were considered to be in the treatment group and whose distances were between 2 to 4 miles were put in the control group. 2 sites were dropped as no individual could be traced within 4 miles of those sites in the year 1999. This exercise is repeated for the rest of the years (2000 to 2013). The data files for every year from 1999 to 2013 were appended to form the estimation sample. Therefore, each observation in the estimation sample is a year-person. A single observation would read the beneficiary ID of the senior, the year of location within the 4 miles of a site along with other demographic characteristics mentioned earlier. Similarly, the estimation sample for the deleted sites was calculated.

Figure A.6: Effect of the Distance from Superfund Sites on the Probability to Move



Regression controls for county times year fixed effects, block group, and individual characteristics. The dependent variable is a binary variable denoting whether individual moves or not. The coefficient on the distance dummy and their confidence intervals are plotted in the figure.

APPENDIX B

DROUGHT SHOCKS AND HOUSEHOLD OCCUPATION CHOICES

B.1 Data

The sample collection for the IHDS wave 2004-05 and 2011-12 was complex. The sample for the 2004-05 IHDS wave is a composite of several separate sub-samples. The two broad categories of the sub-samples are the re-interview sample of the households previously interviewed in 1994-95 for the Human Development Profile of India (HDPI) and new households. The idea was to be able to conduct panel studies across individuals and households. In the 2004-2005 sample, 15,000 households out of the original 33,230 households in the HDPI survey, were eligible for reinterview. There was attrition based on household splitting, migration, death or other reasons. HDPI villages were randomly ranked within each HDPI district and households were sampled within each stratum according to this ranking. In the event that, some households could not be contacted, a replacement household was selected for re-interview within that village to reach a quota. For the state of Karnataka, all records were lost. The entire sample was new but was in similar proportions of the sampling of HDPI households. IHDS extended the original sample. In addition to the 16 states, it added households in randomly selected villages across 17 states and union territories. For all states, villages were sorted by a random number within randomly sorted districts. To add urban households to the existing sample, towns were sampled from the 2001 Census list. With the Census maps as a guide, once a town was selected, a sample of 15 households was drawn from the towns selected. In the later survey of 2011-12, 40,018 households were re-interviewed. 6,911 households were lost due to migration or death. 2,134 additional households were added because, in some towns or villages, original households could not be contacted.

Table B.1 summarizes the characteristics of households in the HDPI, IHDS 1 and IHDS 2 survey waves, categorized by rural and urban households.

Table B.1: Summmary Statistics of HDPI (Wave-1), IDHS 1 (Wave-2) and IHDS 2 (Wave-3)

Characteristics	Wave -1	Wave-2		Wave-3	
	Rural	Rural	Urban	Rural	Urban
# States	16	32	31	32	32
# Districts	195	277	218	273	234
# HH Count	33,230	26,734	14,820	27,579	14,573
# HH with landownership	0.65	0.60	0.08	0.62	0.10
# no adult HH	333	519	159	898	304
– only senior HH	312	498	151	855	288
– only child HH	1	1	0	1	0
– senior-child HH	20	20	8	42	16
# HH members	5.85	5.36	4.88	4.90	4.76
# adults (15-65 years)	3.41	3.27	3.25	3.07	3.28
– female	1.64	1.63	1.61	1.57	1.65
– male	1.77	1.64	1.64	1.50	1.63
# seniors (above 65 years)	0.25	0.31	0.22	0.36	0.31
# child (below 15 years)	2.18	1.78	1.40	1.46	1.17
# members with a job	0.93	2.85	1.58	2.24	1.65
# members with only agri jobs	1.34	1.72	0.19	1.30	0.17
– female	0.44	0.90	0.10	0.70	0.10
– male	0.90	0.82	0.08	0.60	0.06
# members with only non-agri jobs	0.47	0.45	1.28	0.54	1.42
– female	0.07	0.08	0.22	0.11	0.29
– male	0.40	0.37	1.06	0.43	1.13
# members with agri and non-agri jobs	0.34	0.34	0.05	0.39	0.06
– female	0.04	0.06	0.01	0.10	0.14
– male	0.30	0.28	0.04	0.29	0.04
# HH jobs	2.50	2.85	1.58	2.64	1.70
# HH agri jobs	1.68	2.06	0.25	1.70	0.23
# HH non-agri jobs	0.81	0.79	1.33	0.94	1.48
# MGNREGA jobs	-	-	-	0.32	0.02
HH total income	32,004	83,306	1,53,967	1,02,273	1,78,753
HH agri income	17,157	39,608	5,730	44,102	6,542
HH non-agri income	8,427	37,714	1,35,512	45,592	1,49,545
HH non-work income	943	5,984	12,725	12,578	22,666
HH MGNREGA income	-	-	-	1,261	86
HH consumption (last month)	-	2,756	3,979	5,809	8,448
HH consumption (last year)	-	12,468	19,034	31,341	48,673

The table reports characteristics of the households in each wave, categorized into urban and rural regions. HH total consumption data was not available for wave-1. MGNREGA policy was made available between waves 2 and 3, mean of MGNREGA-related variables are conditional on the households availing the policy.

Weather Data: Terrestrial Precipitation: 1900-2017 Gridded Monthly Time Series (V 5.01) data archives precipitation measures in mm for each month from 1900 to 2017 for every 0.5-degree by 0.5-degree latitude/longitude grid node. To compute the rainfall measure for a latitude-longitude node, they combine data from 20 nearby weather stations using an interpolation algorithm based on the spherical version of Shepards distance-weighting method.[1] I compute district-level rainfall as the average of the precipitation levels of each 0.5 degrees by 0.5 degrees latitude/longitude grid note within the geographical boundaries of the district. District-level precipitation z-scores are calculated for the lagged monsoon and non-monsoon periods. The z-scores are re-coded to represent a drought in the previous year in the following way: districts with greater than zero z-scores are re-coded as zero z-score, districts with less than zero z-score are re-coded as the absolute value of the z-score. Therefore, a higher z-score implies a more severe drought. India receives 90% of the annual rainfall within the monsoon months of June, July, August, and September. Agriculture is heavily dependent on the monsoon rainfall. However, to take into account the possibility for multiple cropping cycles, the yearly z-score is used in the primary specification.[2] Additional results are calculated for drought variable defined over the monsoon term (June-sep) to see an aggravated effect. Drought defined over the non-monsoon (oct-may) term is used to explore the additional effect of the dry season (Mueller *et al.* (2014)).

Terrestrial Air Temperature: 1900-2017 Gridded Monthly Time Series (Version 5.01)the dataset records monthly temperature data in degree celsius for each month from 1900 to 2017 for every 0.5-degree by 0.5-degree latitude/longitude grid node. They compute the temperature data in the same way as the precipitation data. They combine data from 20 nearby weather stations using an interpolation algorithm based on the spherical version of Shepards distance-weighting method.[3] I calculate the temperature variable for the district by averaging the temperature measurements for every 0.5 degrees by 0.5 degrees latitude/longitude grid note within the geographical boundaries of the district. To disentangle the effect of low precipitation levels from temperature changes in a district, I control for temperature in that district over the term for which the drought variable is defined. Instead of the average monthly temperature, I include two temperature variables: minimum and maximum temperature. It has been previously documented that variation in minimum and the maximum temperature has opposite effects on crop yields (Welch *et al.*, 2010) which is not captured through the average measure.

---

[1] Willmott, C. J. and K. Matsuura (2001) Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950 - 1999)

[2] There are three cropping cycles in India: Kharif (July-October), rabi (October-march) and summer (march-June). Kharif is the main cropping season, significantly affected by monsoon rainfall (Prasanna, 2014). Rabi season depends on the moisture retained in the soil from the monsoon rainfall.

[3] Willmott, C. J. and K. Matsuura (2001) Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950 - 1999)

Table B.2: Summmary Statistics of Only HDPI (Wave-1), IDHS 1 (Wave-2), IHDS 2 (Wave-3) and Panel Households in the Study Sample

Characteristics	Wave -1	Wave-2		Wave-3		Panel
	Rural	Rural	Urban	Rural	Urban	HH
# States	16	32	31	32	32	18
# Districts	195	277	218	273	234	184
# HH Count	25,231	18,735	14,820	19,811	14,342	7,999
# Landownership HH	0.63	0.57	0.08	0.60	0.10	0.68
# no adult HH	333	519	159	898	304	0
– only senior HH	312	498	151	855	288	0
– only child HH	1	1	0	1	0	0
– senior-child HH	20	20	8	42	16	0
# HH members	5.76	5.02	4.88	4.70	4.76	5.53
# adults (15-65 years)	3.38	3.04	3.25	2.88	3.27	3.50
– female	1.63	1.52	1.61	1.48	1.64	1.68
– male	1.75	1.51	1.64	1.40	1.63	1.82
# seniors (above 65 years)	0.26	0.29	0.22	0.34	0.31	0.33
# child (below 15 years)	2.13	1.69	1.40	1.48	1.17	1.70
# members with a job	2.12	2.30	1.52	2.09	1.64	2.26
# members with only agri jobs	1.31	1.53	0.19	1.17	0.17	1.68
– female	0.43	0.80	0.10	0.64	0.10	0.78
– male	0.89	0.73	0.08	0.53	0.29	0.90
# members with only non-agri jobs	0.48	0.45	1.28	0.54	1.42	0.47
– female	0.08	0.08	0.22	0.12	0.29	0.07
– male	0.40	0.37	1.06	0.43	1.13	0.36
# members with agri and non-agri jobs	0.32	0.31	0.05	0.37	0.06	0.39
– female	0.04	0.06	0.01	0.09	0.02	0.07
– male	0.28	0.25	0.04	0.27	0.04	0.39
# HH jobs	2.45	2.61	1.58	2.46	1.70	2.98
# HH agri jobs	1.64	1.85	0.25	1.54	0.22	2.10
# HH non-agri jobs	0.81	0.76	1.33	0.91	1.48	0.88
# MGNREGA jobs	-	-	-	0.29	0.02	0.37
HH total income	31,984	79,489	1,53,967	96,322	1,78,996	81,033
HH agri income	16,901	35,246	5,730	38,322	6,422	41,778
HH non-agri income	8,724	38,133	1,35,512	45,747	1,49,833	30,639
HH non-work income	999	6,109	12,725	12,253	22,740	6,668
HH MGNREGA income	-	-	-	1,189	82	1,414
HH consumption (last month)	-	2,656	3,979	5,680	8,465	4,582
HH consumption (last year)	-	12,108	19,034	30,026	48,615	24,258

The last column tabulates the means (across the three waves) of the characteristics of the households in the study sample. Columns (2) through (6) summarizes the means for households (not included in the study sample) unique to each of the survey wave. Income is reported in 2012 rupees.

Table B.3: States, Districts and Household Count in the Sample

Region	**State**	**District Count**	**HH Count**
North	Himachal Pradesh	7	428
North	Punjab	7	420
North	Uttaranchal	3	83
North	Haryana	10	570
North	Rajasthan	14	767
North	Uttar Pradesh	12	324
East	Bihar	10	434
East	Assam	1	6
East	West Bengal	8	585
East	Jharkhand	4	155
East	Orissa	17	600
Central	Chhatishgarh	14	419
Central	Madhya Pradesh	22	893
West	Gujarat	12	404
West	Maharashtra	19	963
South	Andhra Pradesh	10	420
South	Kerala	3	121
South	Tamil Nadu	11	407
Total Count	18	184	7999

Table B.4: Composition of Non-Agricultural Jobs in the Three Waves

	Wave-1			Wave-2			Wave-3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All	Land	No land	All	Land	No land	All	Land	No land
Non-agricultural wage job	0.43	0.42	0.43	0.46	0.44	0.47	0.55	0.54	0.58
Salaried job	0.26	0.28	0.23	0.31	0.32	0.30	0.26	0.26	0.25
Business	0.31	0.29	0.34	0.23	0.23	0.23	0.19	0.20	0.17
# HH	4,383	2,516	1,867	4,333	2,699	1,634	4,962	3,089	1,873

Column (1), (4) and (7) records the mean fraction of each of non-agricultural job category among the total non-agricultural jobs in the household for all rural households in wave-1, wave-2 and wave-3 respectively. Column (2), (5) and (8) considers rural households that own land at baseline. Column (3), (6) and (9) considers rural households that do not own land at baseline. This is a subsample of the study sample as some households have zero non-agricultural labor.

B.2 Additional Result Tables

Table B.5: Effect of Drought on Percentage of Agricultural Jobs

	(1)	(2)	(3)
Lagged z-scoore (year)	-0.556		
	(0.846)		
Lagged min temperature (year)	-1.060**		
	(0.463)		
Lagged max temperature (year)	1.699***		
	(0.626)		
Lagged z-score (monsoon)		-1.995**	
		(0.911)	
Lagged min temperature (monsoon)		2.257**	
		(0.923)	
Lagged max temperature (monsoon)		0.756	
		(0.608)	
Lagged z-score (non-monsoon)			-1.073
			(0.848)
Lagged min temperature (non-monsoon)			-1.060**
			(0.432)
Lagged max temperature (non-monsoon)			1.245**
			(0.565)
Year dummy (2005)	1.493	2.194**	1.639
	(0.915)	(0.877)	(0.998)
Year dummy (2011)	-6.689***	-6.047***	-6.998***
	(1.259)	(1.073)	(1.492)
# HH members	-0.300**	-0.313**	-0.283*
	(0.150)	(0.147)	(0.150)
# adult female HH members	1.788***	1.822***	1.791***
	(0.371)	(0.369)	(0.370)
# adult male HH members	-1.978***	-1.939***	-1.991***
	(0.268)	(0.265)	(0.268)
Head's education (higher secondary)	-2.170	-2.349	-2.078
	(1.660)	(1.632)	(1.667)
Head's education (college)	-2.214	-2.044	-2.320
	(2.188)	(2.171)	(2.179)
Constant	36.00	-11.64	51.76***
	(22.29)	(24.76)	(18.50)
N	23,997	23,997	23,997
R^2	0.029	0.029	0.029
mean dep var	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. Standard errors in parentheses.
* $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

Table B.6: Heterogeneous Treatment Effect – Skill Transferability

	Working age members			Household Head		
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged z-score (year)	-0.470			-0.183		
	(1.125)			(0.960)		
Non-ag skill X Lagged z-score (year)	-0.141			-1.077		
	(0.789)			(0.796)		
Lagged min temperature (year)	-1.059**			-1.056**		
	(0.462)			(0.463)		
Lagged max temperature (year)	1.698***			1.695***		
	(0.627)			(0.625)		
Lagged z-score (monsoon)		-2.000*			-1.719*	
		(1.163)			(1.023)	
Non-ag skill X Lagged z-score (monsoon)		0.00745			-0.790	
		(0.897)			(0.883)	
Lagged min temperature (monsoon)		2.257**			2.258**	
		(0.925)			(0.924)	
Lagged max temperature (monsoon)		0.756			0.752	
		(0.608)			(0.609)	
Lagged z-score (non-monsoon)			-1.304			-1.159
			(1.202)			(0.956)
Non-ag skill X Lagged z-score (non-monsoon)			0.356			0.241
			(0.984)			(0.860)
Lagged min temperature (non-monsoon)			-1.058**			-1.061**
			(0.433)			(0.432)
Lagged max temperature (non-monsoon)			1.238**			1.242**
			(0.566)			(0.566)
2005.year	1.501	2.194**	1.633	1.528*	2.212**	1.637
	(0.908)	(0.872)	(0.995)	(0.909)	(0.875)	(0.997)
2011.year	-6.685***	-6.047***	-6.992***	-6.663***	-6.034***	-6.997***
	(1.261)	(1.074)	(1.493)	(1.258)	(1.074)	(1.493)
# HH members	-0.300**	-0.313**	-0.283*	-0.300**	-0.313**	-0.283*
	(0.150)	(0.147)	(0.150)	(0.150)	(0.147)	(0.150)
# adult female HH members	1.789***	1.822***	1.792***	1.782***	1.820***	1.792***
	(0.370)	(0.368)	(0.371)	(0.371)	(0.369)	(0.371)
# adult male HH members	-1.976***	-1.939***	-1.991***	-1.982***	-1.940***	-1.990***
	(0.269)	(0.266)	(0.268)	(0.268)	(0.265)	(0.268)
Head's education (higher secondary)	-2.148	-2.350	-2.086	-1.911	-2.176	-2.078
	(1.655)	(1.625)	(1.668)	(1.673)	(1.638)	(1.668)
Head's education (college)	-2.216	-2.044	-2.319	-2.240	-2.052	-2.323
	(2.189)	(2.171)	(2.180)	(2.190)	(2.173)	(2.180)
Constant	35.99	-11.63	51.95***	36.05	-11.52	51.85***
	(22.27)	(24.76)	(18.52)	(22.27)	(24.74)	(18.53)
Non-ag skill X Lagged z-score	-0.610	-1.992**	-0.948	-1.260	-2.509***	-0.918
	(0.782)	(0.899)	(0.818)	(0.855)	(0.963)	(0.920)
R-squared	0.0293	0.0290	0.0289	0.0294	0.0291	0.0289
Observations	23,997	23,997	23,997	23,997	23,997	23,997
Mean Dep Var	68.98	68.98	68.98	68.98	68.98	68.98
SD Dep Var	33.98	33.98	33.98	33.98	33.98	33.98

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, household controls and temperature controls for appropriate period corresponding to the drought definition are used. In columns(1)-(3): Non-ag skill is an indicator which takes value 1 if the highest education level of the working age members within the household is greater than median education level of the non-agricultural sector at baseline, 0 otherwise. In columns (4)-(6): Non-ag skill is an indicator which takes value 1 if the education level of the household head is greater than median education level of the non-agricultural sector at baseline, 0 otherwise. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.7: Distribution of Households Across (Each) Equipment Count at Baseline (Wave-1)

	Equipment Count		
Equipment Name	0	1	2
Tubewell	6	891	7102
Generator	7	62	7930
Thresher	6	236	7757
Winnower	6	108	7885
Bullock cart	7	1350	6642
Biogas plant	7	118	7874
Tractor	6	269	7724

The column (2)-(4) tabulates the number of households possessing zero, one or two of each of the equipment named in column (1)

Table B.8: Distribution of Households Across Total Equipment Count at Baseline (Wave-1)

Total Equipment count	Household Count
0	5
2	2
7	1
8	1
9	14
10	49
11	126
12	399
13	1,575
14	5,827

The total equipment count is the total the quantity of the following equipments: tubewell, generator, thresher, winnower, bullock cart, biogas plant and tractor

Table B.9: Hetereogenous Treatment Effect – Switching Cost (Land Ownership)

	(1)	(2)	(3)
Lagged z-score (year)	-6.079***		
	(1.305)		
Land ownership X Lagged z-score (year)	8.249***		
	(1.455)		
Lagged min temperature (year)	-1.101**		
	(0.456)		
Lagged max temperature (year)	1.864***		
	(0.630)		
Lagged z-score (monsoon)		-8.390***	
		(1.311)	
Land ownership X Lagged z-score (monsoon)		9.744***	
		(1.559)	
Lagged min temperature (monsoon)		2.144**	
		(0.886)	
Lagged max temperature (monsoon)		0.686	
		(0.610)	
Lagged z-score (non-monsoon)			-2.464
			(1.532)
Land ownership X Lagged z-score (non-monsoon)			2.012
			(1.757)
Lagged min temperature (non-monsoon)			-1.065**
			(0.430)
Lagged max temperature (non-monsoon)			1.291**
			(0.566)
2005.year	1.269	2.160**	1.569
	(0.916)	(0.873)	(0.993)
2011.year	-6.930***	-6.082***	-7.072***
	(1.241)	(1.048)	(1.481)
# HH members	-0.329**	-0.344**	-0.284*
	(0.147)	(0.144)	(0.150)
# adult female HH members	1.735***	1.773***	1.775***
	(0.367)	(0.367)	(0.370)
# adult male HH members	-2.026***	-1.990***	-1.987***
	(0.268)	(0.265)	(0.267)
Head's education (higher secondary)	-2.228	-2.448	-2.110
	(1.672)	(1.639)	(1.668)
Head's education (college)	-2.248	-2.166	-2.257
	(2.188)	(2.163)	(2.180)
Constant	31.78	-5.991	50.45***
	(22.52)	(23.51)	(18.62)
Land ownership X Lagged z-score + Lagged z-score	2.170**	1.354	-0.452
	(0.860)	(0.968)	(0.971)
R-squared	0.0371	0.0375	0.0292
Observations	23,997	23,997	23,997
mean dep var	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. Land Ownership is an indicator which takes value 1 if household owns land at baseline, 0 otherwise. Standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.10: Heterogenous Treatment Effect – Switching Cost (Ties to Land)

	Land Cultivation Indicator			Fraction of Agricultural Income		
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged z-score (year)	-6.674***			-6.719***		
	(1.336)			(1.250)		
Ties to Land X Lagged z-score (year)	8.956***			13.46***		
	(1.531)			(2.038)		
Lagged min temperature (year)	-1.116**			-1.139**		
	(0.453)			(0.438)		
Lagged max temperature (year)	1.860***			1.846***		
	(0.631)			(0.629)		
Lagged z-score (monsoon)		-9.180***			-9.166***	
		(1.360)			(1.297)	
Ties to Land X Lagged z-score (monsoon)		10.71***			16.01***	
		(1.650)			(2.194)	
Lagged min temperature (monsoon)		2.132**			2.117**	
		(0.883)			(0.875)	
Lagged max temperature (monsoon)		0.680			0.724	
		(0.611)			(0.610)	
Lagged z-score (non-monsoon)			-2.547			-2.799*
			(1.604)			(1.421)
Ties to Land X Lagged z-score (non-monsoon)			2.083			3.868
			(1.860)			(2.522)
Lagged min temperature (non-monsoon)			-1.067**			-1.085**
			(0.430)			(0.423)
Lagged max temperature (non-monsoon)			1.290**			1.323**
			(0.566)			(0.559)
2005.year	1.293	2.186**	1.576	1.469	2.436***	1.530
	(0.917)	(0.873)	(0.994)	(0.897)	(0.873)	(0.995)
2011.year	-6.915***	-6.065***	-7.068***	-6.814***	-5.947***	-7.067***
	(1.242)	(1.046)	(1.482)	(1.230)	(1.064)	(1.486)
# HH members	-0.329**	-0.346**	-0.284*	-0.334**	-0.359**	-0.278*
	(0.146)	(0.143)	(0.150)	(0.144)	(0.142)	(0.149)
# adult female HH members	1.727***	1.767***	1.775***	1.752***	1.799***	1.775***
	(0.365)	(0.366)	(0.369)	(0.364)	(0.366)	(0.370)
# adult male HH members	-2.031***	-1.993***	-1.987***	-1.987***	-1.942***	-1.999***
	(0.267)	(0.264)	(0.267)	(0.263)	(0.261)	(0.268)
Head's education (higher secondary)	-2.230	-2.491	-2.107	-2.348	-2.638	-2.133
	(1.669)	(1.634)	(1.668)	(1.646)	(1.612)	(1.671)
Head's education (college)	-2.240	-2.127	-2.261	-1.737	-1.561	-2.205
	(2.186)	(2.159)	(2.180)	(2.162)	(2.136)	(2.190)
Constant	32.15	-5.501	50.50***	32.90	-6.618	49.73***
	(22.53)	(23.35)	(18.63)	(22.22)	(23.60)	(18.44)
Ties to Land X Lagged z-score + Lagged z-score	2.282***	1.526	-0.464	6.739***	6.844***	1.070
	(0.874)	(0.981)	(0.979)	(1.276)	(1.397)	(1.612)
R-squared	0.0382	0.0391	0.0292	0.0444	0.0456	0.0298
Observations	23,997	23,997	23,997	23,997	23,997	23,997
mean dep var	68.98	68.98	68.98	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98	33.98	33.98	33.98

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. In columns (1)-(3), Ties to Land is Land Cultivation variable. Land cultivation is an indicator which takes value 1 if household owns and cultivates land at baseline, 0 otherwise. In columns (4)-(6), Ties to Land is Fraction of Agricultural Income. Fraction of agricultural income is a continuous variable at baseline, higher values indicating greater affinity towards agriculture. Standard errors in parentheses.

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

Table B.11: Heterogenous Treatment Effect – Hindu Indicator

	(1)	(2)	(3)	(4)	(5)	(6)
Lagged z-score (year)	-1.606			-4.407***		
	(1.472)			(1.148)		
Hindu Indicator X Lagged z-score (year)	1.232			6.580***		
	(1.684)			(1.307)		
Lagged min temperature (year)	-1.029**			-1.054**		
	(0.465)			(0.460)		
Lagged max temperature (year)	1.716***			1.869***		
	(0.623)			(0.625)		
Lagged z-score (monsoon)		-2.367			-6.055***	
		(1.725)			(1.287)	
Hindu Indicator X Lagged z-score (monsoon)		0.480			7.331***	
		(1.949)			(1.487)	
Lagged min temperature (monsoon)		2.139**			2.075**	
		(0.926)			(0.903)	
Lagged max temperature (monsoon)		0.765			0.646	
		(0.614)			(0.618)	
Lagged z-score (non-monsoon)			-1.859			-2.287*
			(1.732)			(1.259)
Hindu Indicator X Lagged z-score (non-monsoon)			0.965			2.046
			(1.867)			(1.433)
Lagged min temperature (non-monsoon)			-1.018**			-1.022**
			(0.433)			(0.432)
Lagged max temperature (non-monsoon)			1.314**			1.363**
			(0.561)			(0.565)
2005.year	1.510	2.253**	1.601	1.311	2.244**	1.528
	(0.917)	(0.872)	(0.997)	(0.913)	(0.873)	(0.992)
2011.year	-6.645***	-5.959***	-7.026***	-6.795***	-5.931***	-7.098***
	(1.255)	(1.064)	(1.480)	(1.248)	(1.049)	(1.471)
# HH members	-0.296**	-0.306**	-0.275*	-0.318**	-0.328**	-0.276*
	(0.150)	(0.147)	(0.150)	(0.148)	(0.146)	(0.150)
# adult female HH members	1.767***	1.800***	1.772***	1.732***	1.766***	1.759***
	(0.370)	(0.368)	(0.371)	(0.367)	(0.365)	(0.372)
# adult male HH members	-1.960***	-1.921***	-1.971***	-1.998***	-1.957***	-1.968***
	(0.267)	(0.264)	(0.267)	(0.267)	(0.264)	(0.266)
Head's education (higher secondary)	-2.175	-2.337	-2.051	-2.253	-2.451	-2.079
	(1.652)	(1.624)	(1.661)	(1.666)	(1.635)	(1.659)
Head's education (college)	-2.120	-1.968	-2.232	-2.113	-2.047	-2.168
	(2.174)	(2.158)	(2.166)	(2.179)	(2.157)	(2.166)
Constant	34.85	-8.820	48.71***	30.66	-3.111	47.30**
	(22.20)	(25.05)	(18.33)	(22.40)	(24.00)	(18.57)
Hindu Indicator X Lagged z-score + Lagged z-score	-0.374***	-1.887***	-0.893***	2.172***	1.275***	-0.241***
	(0.940)	(1.019)	(0.915)	(0.916)	(1.019)	(0.979)
R-squared	0.0291	0.0287	0.0287	0.0345	0.0340	0.0291
Observations	24,009	24,009	24,009	24,009	24,009	24,009
mean dep var	68.98	68.98	68.98	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98	33.98	33.98	33.98

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. In columns (1)-(3), Hindu Indicator takes value 1 if religion of the household head is hindu at baseline, zero otherwise. In columns (4)-(6), Hindu Indicator takes value 1 if the religion of the household head is hindu and own farmland. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.12: Effect of Drought on Agricultural Wage Income

	Non-Winsorized			1% winsorized		
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged z-score (year)	752.1			641.7		
	(790.8)			(754.5)		
Lagged min temperature (year)	-1495.4*			-1481.0*		
	(828.4)			(822.2)		
Lagged max temperature (year)	-456.4			-399.0		
	(348.7)			(332.4)		
Lagged z-score (monsoon)		499.4			208.8	
		(916.7)			(839.6)	
Lagged min temperature (monsoon)		-286.0			43.20	
		(977.5)			(970.6)	
Lagged max temperature (monsoon)		124.8			86.85	
		(418.0)			(406.1)	
Lagged z-score (non-monsoon)			490.7			535.6
			(756.4)			(716.3)
Lagged min temperature (non-monsoon)			-1320.3*			-1389.2*
			(791.4)			(775.8)
Lagged max temperature (non-monsoon)			-410.0			-416.0
			(376.5)			(357.3)
2005.year	10454.4***	10305.8***	10354.9***	10191.0***	9987.4***	10159.7***
	(765.9)	(735.2)	(698.3)	(733.6)	(693.7)	(681.5)
2011.year	13706.0***	13649.8***	13593.0***	13395.8***	13160.0***	13402.5***
	(1006.9)	(1096.2)	(1068.8)	(975.4)	(1064.6)	(1037.4)
# HH members	37.41	51.03	26.07	20.46	33.37	10.42
	(100.8)	(99.64)	(103.2)	(96.63)	(96.02)	(98.84)
# adult female HH members	-144.7	-74.82	-145.7	-169.5	-107.4	-177.9
	(242.7)	(245.6)	(245.6)	(217.8)	(219.9)	(217.8)
# adult male HH members	221.2	226.0	225.3	184.4	189.1	190.7
	(181.2)	(186.2)	(182.7)	(158.5)	(163.8)	(160.5)
Head's education (higher secondary)	2568.8	2474.3	2503.1	1408.6	1288.1	1379.4
	(2812.3)	(2773.3)	(2781.2)	(2060.3)	(2014.4)	(2036.6)
Head's education (college)	-3589.1	-3134.2	-3327.0	-1894.5	-1392.5	-1717.2
	(3970.3)	(4011.0)	(3933.3)	(2750.0)	(2760.3)	(2708.1)
Constant	45349.2**	6433.6	40738.3**	43445.7**	-871.7	42325.4**
	(19489.8)	(26779.6)	(17465.1)	(18939.2)	(26691.5)	(17105.5)
N	8,647	8,647	8,647	8,647	8,647	8,647
R^2	0.332	0.325	0.331	0.379	0.370	0.379
mean dep var	11,529	11,529	11,529	11,334	11,334	11,334
sd dep var	12,598	12,598	12,598	11,352	11,352	11,352

Each column records results for a different definition of drought. In column (1) and (4), drought is defined as the z-score for last year, in column (2) and (5), it is defined as the z-score for the last monsoon term and in column (3) and (6) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. The dependent variable is agricultural wage for each household which is calculated by dividing total housheold agricultural wage income with number of agricultural wage worker in the household. The dependent variable for the columns (1)-(3) is not winsorized unlike the next three columns (4)-(6) for which the dependent variable is winsorized at 1%. * $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

Table B.13: Effect of Drought on Non-Agricultural Wage Income

	Non-Winsorized			1% winsorized		
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged z-score (year)	435.7			845.1		
	(935.4)			(864.2)		
Lagged min temperature (year)	1201.2*			979.9		
	(696.5)			(638.8)		
Lagged max temperature (year)	-2642.5***			-2378.8***		
	(689.4)			(646.5)		
Lagged z-score (monsoon)		1668.5			1983.8*	
		(1169.1)			(1061.3)	
Lagged min temperature (monsoon)		-4229.4***			-4002.6***	
		(1228.7)			(1137.5)	
Lagged max temperature (monsoon)		-394.1			-276.1	
		(549.3)			(521.5)	
Lagged z-score (non-monsoon)			1783.6			2127.2**
			(1104.6)			(1024.8)
Lagged min temperature (non-monsoon)			1283.5**			1185.3**
			(610.7)			(577.2)
Lagged max temperature (non-monsoon)			356.8			531.7
			(909.5)			(797.2)
2005.year	8521.4***	7353.8***	6519.1***	8250.4***	7360.7***	6203.0***
	(1338.4)	(974.3)	(1483.9)	(1233.0)	(941.7)	(1347.5)
2011.year	25730.8***	24778.1***	22180.0***	24698.6***	24011.1***	20952.1***
	(1979.3)	(1717.6)	(2557.7)	(1734.4)	(1604.0)	(2106.8)
# HH members	330.4*	393.3**	303.9*	238.5	299.4*	217.9
	(184.9)	(184.6)	(181.3)	(149.8)	(152.0)	(148.8)
# adult female HH members	-1356.3***	-1497.9***	-1401.2***	-1094.4***	-1222.2***	-1153.6***
	(476.7)	(493.9)	(471.2)	(365.7)	(384.0)	(362.4)
# adult male HH members	236.0	202.4	301.4	302.8	273.0	372.8
	(335.0)	(333.2)	(333.6)	(299.8)	(297.5)	(295.8)
Head's education (higher secondary)	-1429.2	-959.7	-1414.0	-1624.7	-1177.8	-1615.7
	(2225.3)	(2247.7)	(2140.0)	(2187.2)	(2215.4)	(2113.5)
Head's education (college)	3497.7	4432.6	3744.3	3409.8	4281.0	3588.5
	(3466.4)	(3019.9)	(3311.1)	(3390.6)	(2965.0)	(3253.8)
Constant	69997.3***	130898.5***	-28372.7	64971.4***	120767.5***	-32330.5
	(23748.4)	(39922.4)	(27391.6)	(22976.9)	(36426.3)	(23881.1)
N	7,235	7,235	7,235	7,235	7,235	7,235
R^2	0.426	0.427	0.420	0.465	0.467	0.461
mean dep var	1,66,401	1,66,401	1,66,401	16,322	16,322	16,322
sd dep var	19,186	19,186	19,186	17,281	17,281	17,271

Each column records results for a different definition of drought. In column (1) and (4), drought is defined as the z-score for last year, in column (2) and (5), it is defined as the z-score for the last monsoon term and in column (3) and (6) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition. The dependent variable is non-agricultural wage for each household which is calculated by dividing total housheold non-agricultural wage income with number of non-agricultural wage worker in the household. The dependent variable for the columns (1)-(3) is not winsorized unlike the next three columns (4)-(6) for which the dependent variable is winsorized at 1%. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.14: Risk-Sharing: Effect of Diversification on Household Consumption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lagged z-score (year)	-6001.9**			-767.0***			-2246.4		
	(2804.2)			(178.1)			(2067.3)		
Diversify from Ag X Lagged z-score (year)	3402.3			364.7**			-854.0		
	(3990.1)			(152.2)			(2985.7)		
Lagged min temperature (year)	-6823.8**			306.4**			-3265.1**		
	(2751.4)			(143.8)			(1504.6)		
Lagged max temperature (year)	2851.0			225.2*			550.4		
	(2912.5)			(125.5)			(1877.9)		
Lagged z-score (monsoon)		-9143.7**			-666.4***			-3198.6	
		(4454.6)			(233.2)			(2757.8)	
Diversify from Ag X Lagged z-score (monsoon)		2679.0			369.6**			-1904.5	
		(4098.3)			(160.7)			(2795.1)	
Lagged min temperature (monsoon)		5753.4			-353.0**			924.3	
		(4229.2)			(170.0)			(2638.9)	
Lagged max temperature (monsoon)		-948.4			-5.411			295.9	
		(3165.2)			(131.0)			(1878.1)	
Lagged z-score (non-monsoon)			-5112.0			-163.7			-3473.0
			(4788.8)			(247.4)			(3330.9)
Diversify from Ag X Lagged z-score (non-monsoon)			-5592.1			-344.5			-2703.0
			(6196.3)			(237.4)			(4774.6)
Lagged min temperature (non-monsoon)			-6277.3**			148.1			-2446.4
			(3044.1)			(152.9)			(1641.8)
Lagged max temperature (non-monsoon)			1332.5			334.7***			618.6
			(2672.7)			(109.2)			(1542.6)
Diversify from Ag	-804.2	-937.9	3481.9	-97.08	-111.8	201.0	433.7	509.4	1633.5
	(3261.0)	(3435.7)	(3157.1)	(120.4)	(128.7)	(135.0)	(2361.1)	(2458.6)	(2384.3)
2011.year	21745.0***	20423.4***	21704.8***	3389.4***	3607.9***	3270.9***	6838.3***	6289.6***	6583.7***
	(3247.8)	(3282.3)	(2799.0)	(136.8)	(143.3)	(144.4)	(2121.0)	(2215.0)	(1970.5)
# HH members	3131.2**	3318.7***	3174.9**	126.1**	121.0**	133.5**	-2075.1**	-1959.8**	-2030.5**
	(1283.7)	(1262.7)	(1263.3)	(59.14)	(57.84)	(59.52)	(988.0)	(987.2)	(970.8)
# adult female HH members	8258.8***	8207.2***	8318.5***	151.9*	158.7*	151.6*	5866.8***	5820.2***	5851.8***
	(2613.0)	(2627.3)	(2608.2)	(87.42)	(84.92)	(87.70)	(1944.7)	(1955.8)	(1943.2)
# adult male HH members	13350.9***	13182.9***	13219.2***	460.6***	456.1***	453.9***	7853.6***	7804.7***	7775.3***
	(2933.0)	(2942.0)	(2885.4)	(120.9)	(120.4)	(121.2)	(2033.1)	(2041.5)	(1998.3)
Head's education (higher secondary)	8100.6	9164.0	7716.7	-34.60	-52.09	42.99	16923.3	17154.8	16686.7
	(23927.3)	(24015.0)	(23693.0)	(547.7)	(539.9)	(552.8)	(20455.3)	(20401.2)	(20333.5)
Head's education (college)	1573.9	25.97	160.8	143.1	69.49	-5.618	-4339.0	-4423.7	-4672.8
	(26835.0)	(26894.0)	(26624.7)	(701.4)	(689.3)	(715.4)	(21600.6)	(21501.2)	(21464.3)
Constant	57692.9	-87725.3	99151.8	-11219.6**	11002.3**	-12252.0***	50065.3	-20712.2	35098.8
	(101103.8)	(129435.5)	(78509.2)	(4618.7)	(4790.1)	(3538.6)	(62291.5)	(86853.0)	(44881.7)
N	10056	10056	10056	10056	10056	10056	10056	10056	10056
R^2	0.063	0.060	0.064	0.357	0.355	0.355	0.019	0.018	0.021
mean dep var	1,03,840	1,03,840	1,03,840	1,03,840	1,03,840	1,03,840	1,03,840	1,03,840	1,03,840
sd dep var	1,03,189	1,03,189	1,03,189	1,03,189	1,03,189	1,03,189	1,03,189	1,03,189	1,03,189

Table B.15: Effect of Drought on Total Household Income

	(1)	(2)	(3)
Lagged z-score (year)	5042.0		
	(3718.1)		
Lagged z-score (monsoon)		2355.6	
		(3708.7)	
Lagged z-score (non-monsoon)			10972.4***
			(3943.5)
N	23,997	23,997	23,997
R^2	0.109	0.108	0.113
mean dep var	80,243	80,243	80,243
sd dep var	1,58,816	1,58,816	1,58,816

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition.
* $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

Table B.16: Effect of Drought on Household Income by Sources

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lagged z-score (year)	706.8			3731.2**			1046.0*		
	(3162.1)			(1523.2)			(566.0)		
Lagged z-score (monsoon)		-1325.1			3574.1**			894.2	
		(2942.6)			(1612.8)			(635.8)	
Lagged z-score (non-monsoon)			2917.0			4782.3***			3672.0***
			(3469.0)			(1590.5)			(873.0)
N	23,997	23,997	23,997	23,997	23,997	23,997	23,997	23,997	23,997
R^2	0.037	0.037	0.038	0.121	0.121	0.122	0.044	0.044	0.048
mean dep var	40,994	40,994	40,994	30,633	30,633	30,633	6,668	6,668	6,668
sd dep var	1,30,709	1,30,709	1,30,709	72,121	72,121	72,121	37,207	37,207	37,207

Each column records results for a different definition of drought. In columns (1),(4),(7), drought is defined as the z-score for last year, in columns (2),(5),(8) it is defined as the z-score for the last monsoon term and in columns (3),(6),(9) for the last non-monsoon term. For each column, temperature controls are used for the appropriate period corresponding to the drought definition. The dependent variable for columns (1)-(3) is household agricultural income, for columns (4)-(6) is household non-agricultural income and for columns (7)-(9) is household non-work 'other' income. * $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

Table B.17: Effect of Drought on Household Income Shares

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lagged z-score (year)	-1.400			1.410			0.901		
	(1.096)			(1.071)			(0.788)		
Lagged z-score (monsoon)		0.274			1.321			0.589	
		(1.362)			(1.178)			(1.146)	
Lagged z-score (non-monsoon)			-2.094*			1.305			1.106
			(1.228)			(1.038)			(0.705)
N	23,997	23,997	23,997	23,997	23,997	23,997	23,997	23,997	23,997
R^2	0.012	0.011	0.012	0.048	0.047	0.048	0.026	0.026	0.026
mean dep var	49	49	49	33	33	33	7	7	7
sd dep var	68	68	68	57	57	57	35	35	35

Each column records results for a different definition of drought. In columns (1),(4),(7), drought is defined as the z-score for last year, in columns (2),(5),(8) it is defined as the z-score for the last monsoon term and in columns (3),(6),(9) for the last non-monsoon term. For each column, temperature controls are used for the appropriate period corresponding to the drought definition. The dependent variable for columns (1)-(3) is the percentage of household agricultural income, for columns (4)-(6) is the percentage of household non-agricultural income and for columns (7)-(9) is the percentage of household non-work 'other' income. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B.3 Robustness Checks

Table B.18: Effect of Drought (Monsoon) on Percentage of Agricultural Jobs

	(1)	(2)	(3)
Lagged z-score (year)	-0.730		
	(1.207)		
Lagged z-score (monsoon)		-3.538***	
		(1.190)	
Lagged z-score (non-monsoon)			-0.427
			(1.134)
N	14,340	14,340	14,340
R^2	0.039	0.039	0.039
mean dep var	62.67	62.67	62.67
sd dep var	36.28	36.28	36.28

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for appropriate period corresponding to the drought definition.

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

Table B.19: Effect of a Drought of the Percentage of Agricultural Jobs

	(1)	(2)	(3)
Lagged z-score (year)	-0.737		
	(0.891)		
Lagged z-score (monsoon)		-1.862*	
		(0.958)	
Lagged z-score (non-monsoon)			-1.328
			(1.013)
N	23,997	23,997	23,997
R^2	0.030	0.030	0.030
Region X Year FE	Yes	Yes	Yes
mean dep var	68.98	68.98	68.98
sd dep var	33.98	33.98	33.98

Standard errors in parentheses

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

Table B.20: Effect of Drought on Percentage of Agricultural Jobs for "Imagine No Split" Households

	(1)	(2)	(3)
Lagged z-score (year)	-0.653		
	(0.846)		
Lagged z-score (monsoon)		-1.982**	
		(0.919)	
Lagged z-score (non-monsoon)			-1.129
			(0.856)
N	23,997	23,997	23,997
R^2	0.034	0.034	0.034
mean dep var	68.49	68.49	68.49
sd dep var	33.44	33.44	33.44

Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, household controls and temperature controls for appropriate period corresponding to the drought definition are used. The sample is "Imagine no split" household sample which considers split households to be a single unit as they were in the baseline survey. Standard errors in parentheses.
* $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

APPENDIX C

THE IMPACT OF DROUGHT ON STRUCTURAL TRANSFORMATION IN INDIA

C.1 Data, Summary Statistics and Reduced-Form Evidence

Imputation of Labor Hours for Wave-1: IHDS 2004-05 (Wave-2) and 2011-12 (Wave-3) document labor hours spent on different job categories for each household member. However, HDPI 1994-95 (Wave-1) does not have labor hours, it documents job categories only. To impute labor hours for the first wave, the following was carried out: The total household labor hours in Wave-2 and Wave-3 were categorized into agriculture and non-agriculture hours respectively. The percentage of agriculture hours is the household agriculture hours divided by the household labor hours. To figure out, what household-level variables affected the percentage of agriculture hours in Wave-2 and Wave-3, a linear lasso regression was employed using all possible independent variables for which data was available for all three waves of data. Lasso algorithm figures the relevant independent variables and thereafter the predicted value of percentage of agricultural hours for Wave-1 was calculated. The total labor hours for Wave-1 is the average total labor hours in Wave-2 and Wave-3 for that household. Agriculture labor hours for each household was calculated by multiplying the total annual household labor hours by the predicted percentage of agriculture hours for that household.

Table C.1: Summary statistics across three waves

Characteristics	Wave-1	Wave-2	Wave-3
Number of States	18	18	18
Number of Districts	184	184	184
Number of Households	7,999	7,999	7,999
Household members	6.13	6.17	4.29
Household female members	2.94	3.01	2.02
Household male members	3.18	3.16	2.27
Household working age members	3.52	3.81	3.17
Land Owners	0.68	0.67	0.69
Land Cultivators	0.71	0.60	0.64
Working age household members employed in Agriculture	1.70	2.63	2.26
Household Job Count	2.63	3.41	2.90
Fraction of Agricultural Jobs	0.69	0.73	0.65
Household Annual Labor Hrs	3500	4206	3619
Fraction of Agricultural Labor Hrs	61.50°	62.81	55.84
Mean household Income (in 2012 rupees)	32,069	92,246	118,784
Mean household Agricultural Income	17,964	49,823	57,547
Yearly z-score last year	0.79	0.49	0.29
Monsoon z-score last year	0.70	0.41	0.28
Non-monsoon z-score last year	0.53	0.54	0.42
Min temp last year	17.73	17.03	17.19
Max temp last year	32.30	32.92	33.26
Min monsoon temp last year	26.91	27.18	27.32
Max monsoon temp last year	31.59	31.66	31.65
Min non-monsoon temp last year	17.73	17.03	17.19
Max non-monsoon temp last year	31.45	32.33	33.16

The upper panel of the table summarizes the mean of the household characteristics across the three waves. The lower panel tabulates the z-scores and temperature variables for the last year for each wave. The agricultural hours for the first wave ($\circ$) is imputed based on a Lasso-type regression, elaborated in appendix section 'Imputation of Labor Hours in Wave-1'.

Table C.2: The effect of drought on the agricultural labor share

	(1)	(2)	(3)
Lagged z-score (year)	-1.833**		
	(0.707)		
Lagged z-score (monsoon)		-1.367	
		(0.897)	
Lagged z-score (non-monsoon)			-2.304**
			(0.892)
HH fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
N	22,175	22,175	22,175
R^2	0.025	0.020	0.022
mean dep var	60.05	60.05	60.05
sd dep var	37.52	37.52	37.52

The dependent variable measures the agricultural labor share, scaled from 0 to 100. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, household controls and temperature controls for appropriate period corresponding to the drought definition are used. Standard errors in parentheses.
* $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

Table C.3: Hetereogenous Treatment Effect: Land ownership

	(1)	(2)	(3)
Lagged z-score (year)	-6.121***		
	(1.250)		
Land ownership X Lagged z-score (year)	6.383***		
	(1.434)		
Lagged z-score (monsoon)		-5.871***	
		(1.461)	
Land ownership X Lagged z-score (monsoon)		6.788***	
		(1.647)	
Lagged z-score (non-monsoon)			-4.998***
			(1.397)
Land ownership X Lagged z-score (non-monsoon)			3.848**
			(1.705)
HH fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Land ownership X Lagged z-score + Lagged z-score	0.262	0.917	-1.150
	(0.766)	(0.925)	(1.071)
R-squared	0.0284	0.0233	0.0231
Observations	22,175	22,175	22,175
mean dep var	60.05	60.05	60.05
sd dep var	37.52	37.52	37.52

The dependent variable is the percentage of agricultural hours in household 'i' located in district 'd' in year 't'. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for the appropriate period corresponding to the drought definition. Land Ownership is an indicator that takes value 1 if the household owns land at baseline, 0 otherwise. Standard errors in parentheses. * $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

Table C.4: Hetereogenous Treatment Effect: Land Cultivation

	(1)	(2)	(3)
Lagged z-score (year)	-6.628***		
	(1.257)		
Land cultivation X Lagged z-score (year)	6.991***		
	(1.443)		
Lagged z-score (monsoon)		-6.556***	
		(1.481)	
Land cultivation X Lagged z-score (monsoon)		7.647***	
		(1.668)	
Lagged z-score (non-monsoon)			-5.302***
			(1.406)
Land cultivation X Lagged z-score (non-monsoon)			4.181**
			(1.676)
Land cultivation X Lagged z-score + Lagged z-score	0.364	1.091	-1.121
	(0.762)	(0.917)	(1.046)
R-squared	0.0290	0.0240	0.0232
Observations	22175	22175	22175
mean dep var	60.05	60.05	60.05
sd dep var	37.52	37.52	37.52

The dependent variable is the percentage of agricultural hours in household 'i' located in district 'd' in year 't'. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for an appropriate period corresponding to the drought definition. Land Ownership is an indicator that takes value 1 if the household owns land at baseline, 0 otherwise. Standard errors in parentheses. * $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

Table C.5: Hetereogenous Treatment Effect: Fraction of Ag Income

	(1)	(2)	(3)
Lagged z-score (year)	-6.034***		
	(1.098)		
Fraction of ag income X Lagged z-score (year)	9.183***		
	(1.687)		
Lagged z-score (monsoon)		-5.900***	
		(1.289)	
Fraction of ag income X Lagged z-score (monsoon)		10.06***	
		(1.898)	
Lagged z-score (non-monsoon)			-4.405***
			(1.247)
Fraction of ag income X Lagged z-score (non-monsoon)			4.688**
			(1.942)
Fraction of ag income X Lagged z-score + Lagged z-score	3.148***	4.160***	0.283
	(1.038)	(1.193)	(1.392)
R-squared	0.0303	0.0252	0.0231
Observations	22175	22175	22175
mean dep var	60.05	60.05	60.05
sd dep var	37.52	37.52	37.52

The dependent variable is the percentage of agricultural hours in household ‘i’ located in district ‘d’ in year ‘t’. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for an appropriate period corresponding to the drought definition. Land Ownership is an indicator that takes value 1 if the household owns land at baseline, 0 otherwise. Standard errors in parentheses.
* $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

Table C.6: Heterogenous Treatment Effect: Hindu Indicator

	(1)	(2)	(3)
Lagged z-score (year)	-6.096***		
	(0.980)		
Hindu Indicator X Lagged z-score (year)	7.729***		
	(1.037)		
Lagged z-score (monsoon)		-5.936***	
		(1.305)	
Hindu Indicator X Lagged z-score (monsoon)		8.594***	
		(1.273)	
Lagged z-score (non-monsoon)			-8.389***
			(1.165)
Hindu Indicator X Lagged z-score (non-monsoon)			10.18***
			(1.314)
Land ownership X Lagged z-score	1.633***	2.658***	
(joint_sd)	(0.766)	(0.862)	
R-squared	0.0326	0.0279	0.0316
Observations	22175	22175	22175

The dependent variable is the percentage of agricultural hours in household 'i' located in district 'd' in year 't'. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for an appropriate period corresponding to the drought definition. Land Ownership is an indicator that takes value 1 if the household owns land at baseline, 0 otherwise. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.7: Effect of Land Ownership on Farm Equipment Spending

	(1)	(2)	(3)
Lagged z-score (year)	283.3		
	(236.9)		
Land ownership X Lagged z-score (year)	-613.4		
	(452.2)		
Lagged z-score (monsoon)		63.94	
		(362.7)	
Land ownership X Lagged z-score (monsoon)		-387.0	
		(504.6)	
Lagged z-score (non-monsoon)			478.0
			(336.6)
Land ownership X Lagged z-score (non-monsoon)			-1000.2*
			(552.7)
Land ownership X Lagged z-score	-330.1	-323.0	-522.2
(joint_sd)	(401.1)	(473.5)	(539.8)
R-squared	0.00126	0.00109	0.00128
Observations	15068	15068	15068
mean dep var	1427	1427	1427
sd dep var	21300	21300	21300

The dependent variable is the amount of rupees spent on new farm equipment by household 'i' located in district 'd' in year 't'. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for an appropriate period corresponding to the drought definition. Land Ownership is an indicator that takes value 1 if the household owns land at baseline, 0 otherwise. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.8: Effect of Land Ownership on Loan

	(1)	(2)	(3)
Lagged z-score (year)	-0.00873		
	(0.0218)		
Land ownership X Lagged z-score (year)	0.0237		
	(0.0218)		
Lagged z-score (monsoon)		-0.0238	
		(0.0254)	
Land ownership X Lagged z-score (monsoon)		0.0520**	
		(0.0261)	
Lagged z-score (non-monsoon)			0.0404
			(0.0354)
Land ownership X Lagged z-score (non-monsoon)			-0.0325
			(0.0304)
Land ownership X Lagged z-score	0.0150	0.0282	0.00792
(joint_sd)	(0.0197)	(0.0273)	(0.0280)
R-squared	0.0222	0.0250	0.0234
Observations	15068	15068	15068
mean dep var	0.55	0.55	0.55
sd dep var	0.50	0.50	0.50

The dependent variable is the indicator for a loan taken by household 'i' located in district 'd' in year 't'. Each column records results for a different definition of drought. In column (1), drought is defined as the z-score for last year, in column (2), it is defined as the z-score for the last monsoon term and in column (3) for the last non-monsoon term. For each column, temperature controls are used for an appropriate period corresponding to the drought definition. Land Ownership is an indicator that takes value 1 if the household owns land at baseline, 0 otherwise. Standard errors in parentheses.
* $p < 0.1$, **$p < 0.05$, *** $p < 0.01$

C.2 Model and Results

Landed Households:

The FOCs:

$$l^{a,f} \quad : \quad (c^f)^{-\sigma} w^a + \psi^f \delta (l^{a,f})^{\delta-1} - \frac{k}{2}\alpha(l^{a,f} + l^{m,f})^{\alpha-1} = 0 \tag{C.1}$$

$$l^{m,f} \quad : \quad (c^f)^{-\sigma} w^m - \frac{k}{2}\alpha(l^{a,f} + l^{m,f})^{\alpha-1} = 0 \tag{C.2}$$

$$\tag{C.3}$$

Landless Households:

The FOCs:

$$l^{a,nf} \quad : \quad (c^{nf})^{-\sigma} w^a + \psi^{nf} \delta (l^{a,nf})^{\delta-1} - \frac{k}{2} \alpha (l^{a,nf} + l^{m,nf})^{\alpha-1} = 0 \tag{C.4}$$

$$l^{m,nf} \quad : \quad (c^{nf})^{-\sigma} w^m - \frac{k}{2} \alpha (l^{a,nf} + l^{m,nf})^{\alpha-1} = 0 \tag{C.5}$$

$$\tag{C.6}$$

Table C.9: Non-Ag Wages for 25% Manufacturing GDP Share

Year		Non-Ag Wage for 'Reform' Ag Shares					
	Ag Share	Landless		Landed		Average	
	Green (2015)	Drought	No-Drought	Drought	No-Drought	Drought	No-Drought
2011	-	33.43	33.34	33.43	33.34	33.43	33.34
2022	38	35.00	34.69	43.00	38.10	38.80	36.60
2025	33	37.09	36.35	46.55	40.23	41.80	38.30
2035	25	43.50	40.30	55.91	44.78	49.70	42.70