Essays in Development Economics

Francisco Manuel Oteiza Aguirre

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of

Royal Holloway, University of London.

Department of Economics

Royal Holloway, University of London

January 5, 2018

I, Francisco Manuel Oteiza Aguirre, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work. I declare Chapter 1 was entirely my work, that Chapter 2 was co-authored with Fernando Aragón and Juan Pablo Rud, and that Chapter 3 was co-authored with Laura Abramovsky, Britta Augsburg, Melanie Lührmann and Juan Pablo Rud. Within this partly co-authored work, I declare that the following contributions are entirely my own work:

- Chapter 2: 100% of the processing and analysis of satellite weather and soil data, 50% of of the processing and analysis of agricultural data. 33% of the drafting of the paper.

- Chapter 3: 100% of two field trips to study areas, 33% of the design of data collection instruments, 80% of the data processing, 20% of the analysis of the data, 50% of the drafting of the paper.

Abstract

This thesis aims at improving our understanding of the constraints under which poor households make decisions. The first two chapters deal with the exposure of rural poor households to extreme weather events, using the cases of Tanzania and Peru as examples. Using a panel of Tanzanian farms, I first show that a combination of satellite data, for temperature, and reanalysis data, for precipitation, successfully captures the impact of weather variables on yields. I then explore some of the margins of adjustment available to farmers when exposed to temperature shocks, and discuss my results within the existing theory and evidence. Chapter 2, co-authored with Fernando Aragón and Juan Pablo Rud, takes a closer look at farmer reactions after experiencing a temperature weather shock. Using data from a representative sample of Peruvian farmers, coupled with satellite weather data, we study how high temperature affects farmer yields. Then, using a model of producer-consumer households, we tease out the negative impact of shocks on productivity, net of farmer reactions during the course of the growing season. Chapter 3 was co-authored with Laura Abramovsky, Britta Augsburg, Melanie Lührmann and Juan Pablo Rud. Using data from a randomised controlled trial designed by the team, we study how households make sanitation investments. We show that an information campaign increased toilet ownership in the short, but not the medium term. We also find that the intervention was successful at increasing expected benefits from toilet ownership related to pride and social status, but did not affect household's perceptions of other benefits, such as private health or privacy, nor increased their awareness of health externalities.

Acknowledgements

This thesis was written between September 2013 and September 2017, in partial fulfilment of the requirements for the degree of Doctor of Philosophy of the Department of Economics, Royal Holloway, University of London. My Doctoral Studies were supported by a College Research Scholarship from Royal Holloway, for which I am immensely grateful. Part of the work presented in Chapter 1 was partially funded by a RHUL-EPSRC Global Challenges Research Fund 2016 grant. Chapter 3 presents research carried out as part of WaterAid UK's Project "Sustainable Total Sanitation Nigeria -implementation, learning, research, and influence on practice and policy", implemented by Wateraid UK and Wateraid Nigeria, and generously funded by the Bill and Melinda Gates Foundation.

I'd like to thank my colleagues at Royal Holloway Economics Department's PhD programme. The open and approachable teaching staff, friendly fellow PhDs and incredibly patient administration will be sorely missed. To the staff and students who attended the RHUL internal seminars and commented on my projects, I sincerely appreciate it, and hope I have been able to return at least a small amount of that help.

To my friends and colleagues at the IFS-EDePo, thank you for welcoming me here and allowing me to use this as my (practically) full time office. Being surrounded by your talent and generosity has transformed my PhD experience, and made me a better economist.

I have also benefitted from enormously from an exchange at the Department of Economics of the University of Oslo (UiO), between March and August

2016. I'd like to thank my colleagues and friends at UiO for the warm welcome, and in particular, Anirban Mitra and Kalle Moene for your generous invitation.

I'd like to thank Melanie Lührmann, my second PhD supervisor, for being so open and available for discussion, even on late train rides back to London. To my supervisors at the IFS, Britta Augsburg and Laura Abramovsky, thank you for recruiting me and helping me strengthen the skills I was (and still am) missing. Working together was a rewarding experience every single day.

To my PhD supervisor, Juan Pablo Rud, I thank you especially for your constant support. I feel privileged to have had such a generous and helpful mentor these years. Our discussions are always challenging and enriching. I have learned, among many other things, from you dogged resistance to simplifying answers.

To my girlfriend and travel partner, Ragnhild Schreiner, thank you for challenging me every day, and for you constant care and support. Finally, I'd like to thank my family, who tolerated my long and protracted student life stoically. Our summers together gave me the rest and perspective I needed after each year, thank you for all these years of loving support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis studies two aspects affecting the economic lives of the poor that present serious challenges in the XXI Century. The first two chapters deal with the exposure of rural poor households to extreme weather events, using the cases of Tanzania and Peru as examples. Climate change is expected to bring dramatic shifts in agricultural activities at a global scale, and affect the pattern of extreme weather, particularly around the tropics (IPCC [2014]). Quantifying the exposure of rural households to weather related risk and understanding their reactions to shocks is a timely challenge. A precise understanding of their degree of exposure, and their margins of reaction and adaptation, is needed for the design of climate and poverty alleviation policies.

A significant constraint in the analysis of weather impacts and agriculture in developing countries is data availability. As discussed by Auffhammer et al. [2012], weather monitoring stations are scarce and sparsely distributed, and their temporal coverage is patchy. In order to construct balanced panels of weather conditions over space and time, researchers must rely on alternative sources, generally in the form of gridded datasets. Chapter 1 addresses this issue. Using a panel of Tanzanian farms between 2008 and 2012, I show that a combination of satellite data, for temperature, and reanalysis data, for precipitation, successfully captures the impact of weather variables on yields. I verify the importance of precipitation for yields widely studied in the literature, and document a land surface temperature threshold of 22°C beyond which additional heat is detrimental for yields. I show that this is comparable to an average daily air temperature of 25°C and this is in line with recent findings from developing countries such as Burgess et al. [2017]. Each additional degree above this threshold during the growing season decreases output per hectare planted by 8 percentage points, controlling for precipitation. In turn, a reduction of one standard deviation in precipitation over the growing season, results in 15 percentage point lower yields.

The second section of Chapter 1 explores some of the margins of adjustment available to farmers when exposed to temperature shocks. High temperatures affect yields of staples, grown by at least 80% of the farms in our sample and mainly for own consumption, but do not significantly affect cash crops, grown by 10% of the farms. This makes it clear that a majority of the households in our sample experience drops in the production and income, when exposed to higher temperatures. However, high temperatures do not appear to affect consumption per capita, suggesting that households successfully smooth out consumption. I discuss several possible risk coping mechanisms discussed in related studies, and find evidence of 1) increased land being allocated to more heat resistant staples such as sweet potato and rice, and 2) increased areas planted with cash crops and a higher number of trees of permanent crops. Nonetheless, these reactions appear to be driven only by households who already had these crops as part of their portfolio, as there is no evidence of additional households adopting them. I also find that households are more likely to take out (mostly informal) loans for subsistence purposes after hot seasons, although the share of households who declare to have access to any type of credit is just 10%. Overall, while changing crop shares and taking up credit seem like feasible mechanisms to part of the farmers in our sample, this does not appear to apply to the majority of farms.

The final strategy I explore is that of migration. Rosenzweig and Stark [1989] showed that migration patterns induced by marriage among a sample of Indian villages are consistent with risk mitigation strategies that aim to diversify spatially correlated weather risk. In the Tanzanian context, migration has been shown to result in increases of up to 36% in individual consumption, by Beegle et al. [2011]. I find indeed that after a growing season with high temperatures, households are 1 percentage point more likely to see one of their members migrate, particularly young men and middle aged women. Nonetheless, I show that this does not translate into higher remittances received by the household of origin after a season of high temperatures, and thus reduced crop yields. This evidence complements the intuition behind Burgess et al. [2017], who suggest that insurance must be limited in India, since higher temperatures, by affecting agricultural productivity, severely affect mortality rates in rural areas.

In Chapter 2, co-authored with Fernando Aragón and Juan Pablo Rud, we take a closer look at farmer reactions after experiencing a temperature weather shock. Using data from a large and nationally representative sample of farmers from Peru, coupled with high resolution satellite weather data for the last decade, we study how high temperature affects farmer yields. Then, using a model of producer-consumer households, we tease out the negative impact of shocks on productivity, net of farmer reactions during the course of the growing season. We document how temperatures above a region-specific threshold impact productivity negatively, and find that farmers react to this shock in different ways, according to their level of income.

This Chapter presents three important contributions to the existing literature studying the effects of weather shocks on agriculture. First, as in Chapter 1, we estimate impacts at the farm level, and use publicly available gridded weather products that make our analysis replicable in most developing country contexts, as long as high quality agricultural data is available. Second, we split the analysis along two different geographical regions which have distinct production methods and access to markets, and show that the severity of impacts, as well as the adjustment margins available to farmers, are region-specific. Finally, we simulate the magnitude of impacts on yields for the end of the Century under two different scenarios from the 4th IPCC Assessment report, one 'business-as-usual' and a second in which large amounts of greenhouse gas emissions are curbed. We find that yields are predicted to fall in one region, the coast, and to increase in the other, the highlands, under both forecasts. This disparity is an important insight about the uneven effects of the increase in average temperatures expected, and for the design of future policy.

This initial focus on production is complemented in Chapter 3 by a look at how poor households invest in their own health. This chapter was co-authored with Laura Abramovsky, Britta Augsburg, Melanie Lührmann and Juan Pablo Rud. Low levels of investment on health-advancing durables at the household level often results in poor health and economic outcomes, particularly in developing countries where complementary efforts by the other agencies are scarce. Nonetheless, these investments are often low due to credit or solvency constraints, lack of appropriate information or other externalities at the community or household levels. Understanding the binding constraints that hamper such health investments is key for designing interventions that help to achieve efficient levels of household investment and provide long-run improvements in health and longevity.

We study how households make investment decisions regarding health improvements by designing and evaluating the results of a cluster-randomised controlled trial (RCT) carried out in Nigeria. Nigeria faces enormous challenges in the field of sanitation, with 34% of its population practising open defecation and slightly falling toilet ownership rates over the last decade (Unicef and WHO [2015]). We look at the constraints behind the decision to construct a private toilet and the types of households that respond positively to an intervention that provides information about the health risks surrounding the practice of open defecation. We analyse a randomly assigned information campaign called Community Led Total Sanitation (CLTS), which provided no subsidies or credit, and was originally designed to promote private toilet construction and reduce open defecation levels in rural Bangladesh, and was later rolled out to other countries in Asia and Africa (Kar [2003]).

Our study makes three main contributions to the existing literature. First, we show that the information campaign increased toilet ownership by a moderate 3 percentage points, six to twelve months after the intervention. While small, our estimates are, to the best of our knowledge, the first to show positive and statistically significant CLTS impacts among a representative sample of households. In a second part, we exploit rich household level data to investigate the different channels through which an information campaign such as CLTS could operate. We find that the intervention was successful at increasing expected benefits from toilet ownership relating to pride and social status. It did not, however, change the household's perceptions on other private benefits, such as health or privacy, nor increased awareness of the externalities deriving from toilet ownership and usage. At the same time, program impacts appear to be stronger for households that perceived toilets not to be too expensive to build: increasing the perceived benefits of toilet construction was more effective among households with low initial perceived costs. We find no evidence that the impact of CLTS on toilet construction and open defecation reduction is driven by changes in social capital nor institutional sanctions.

Finally, we investigate the effect of the information campaign on households with low initial access to sanitation: lower education or asset poor households. Because inadequate sanitation might affect women and children disproportionately, we also consider female headed households and households with children. We find no evidence of larger program impacts among households with children compared to the rest of the sample. On the other hand, we find that CLTS program impacts are concentrated among female-headed households, and households with lower levels of education and asset wealth. Treatment effects are in the neighbourhood of 5-6 percentage points among these groups, which also have lower levels of toilet ownership at baseline: between 33% and 21%. Estimated programme impacts suggest that an information-only campaign may help reduce the sanitation gap, but given the magnitude of the sanitation challenge in the Nigerian context, it might not be enough to close the gap completely.

The structure of the thesis will be as described in this introduction. Chapter 1 will address weather shocks among Tanzanian farmers, Chapter 2 will look into farmer responses to shocks within a sample of Peruvian farmers, and Chapter 3 will analyse health investment decisions in the Nigerian states of Ekiti and Enugu. Most of the Tables and Figures relevant to each Chapter are included in the main body of the thesis, but additional Tables and robustness checks are included in the Appendix, with a section for each Chapter.

Chapter 2

Satellite data and the impacts of weather shocks on Tanzanian agriculture

2.1 Introduction

Climate change is expected to bring dramatic shifts in agricultural activities at a global scale (IPCC [2014]). The rural poor in developing countries are at the front line of these impacts. Many reside near the tropics, or in already warm areas, and traditional farming is their main source of livelihood. These are the areas in which weather shocks, in the shape of extremely hot temperatures, droughts and floods are likely to become less of a rare occurrence. Quantifying the exposure of rural farmers to weather related risk and understanding their reactions to shocks is important in the design of climate and agricultural policy, as well as poverty alleviation and food security efforts.

A significant constraint in the analysis of weather impacts and agriculture in developing countries is data availability. As discussed by Auffhammer et al. [2012] weather monitoring stations are scarce and sparsely distributed, and their temporal coverage is patchy. In order to construct balanced panels of weather conditions over space and time, researchers must rely on gridded datasets, from climate reanalysis or satellite-only sources. However, each source has its own limitations (Dell et al. [2014]), and there is little agreement in the literature around about the appropriate choice. Reanalysis datasets are heavily dependent on the climate models behind them and will tend to underestimate true weather variability, particularly in areas with low station coverage, and in high temporal resolutions (e.g. daily or higher). Satellite data, on the other hand, provide snapshots of the weather recorded at fixed times of the day, according to the instrument's orbit, which may or may not coincide with daily maximums, minimums or averages.

In this Chapter, I show that gridded products can effectively be used to analyze the impact of weather shocks at the farm level when weather station data is sparse or missing. At the same time, the results are consistent with other studies that rely on monitoring station data carried out in both developed and developing countries. Using a panel of Tanzanian farms between 2008 and 2012, I show that a combination of satellite data, for temperature, and reanalysis data, for precipitation, successfully captures the impact of weather variables on yields. I verify the importance of precipitation for yields widely studied in the literature, and document a land surface temperature threshold of 22℃ beyond which additional heat is detrimental for yields. I show that this is comparable to an average daily air temperature of 25℃, and in line with recent findings from developing countries such as Burgess et al. [2017]. Each additional degree above this threshold during the growing season decreases output per hectare planted by 8 percentage points, controlling for precipitation. In turn, a reduction of one standard deviation in precipitation over the growing season, results in 15 percentage point lower yields.

Recent studies from US farms document sizeable drops in crop yields due to extreme heat, and suggest important heterogeneous effects: increments of agricultural production in temperate regions, but significant losses in hotter areas Deschenes and Greenstone [2007], Schlenker et al. [2005], Schlenker and Roberts [2009]. In this context, the case of Tanzania is particularly relevant for two reasons. First, it is a country with already high average temperatures, exposed to large losses if present warming trends continue. Second, it is a country where estimates suggest as much as 76% of the population rely on subsistence agriculture, a production method with potentially smaller margins for adaptation than modern, capital intensive farms from developed economies (ESRF [2014]). In a closely related study, Rowhani et al. [2011], analyze the impact of temperature and precipitation from interpolated weather station data, on yields for three staple crops at the regional level. They find non-linear relationships between precipitation and yields but fail to find this for the case of temperature, possibly due to the aggregate nature of their climate data. The authors stress the importance of having reliable, high-frequency weather measures to enhance our understanding. The evidence I present in this Chapter suggests that satellite and gridded products deserve a place in any development economist's toolbox. I validate my results by instrumenting satellite temperature data with a climate reanalysis source and find that the impacts found are robust in sign and significance.

In the second section of this Chapter, I explore some of the farmers reactions to temperature shocks, and shed light on some of the possible adjustment mechanisms available to them. High temperatures affect yields of staples, grown by at least 80% of the farms in the sample, and mainly for own consumption, but do not affect cash crops. Extremely hot seasons, at the same time, do not appear to affect consumption per capita, suggesting that households partially offset the impact of lower yields and can smooth out consumption. Salazar-Espinoza et al. [2015] follow a panel of farms from Mozambique and study their reactions to droughts. The authors find that farmers react by readjusting their crop choice and increasing the fraction of land devoted to staple crops, and suggest a consumption smoothing strategy based on the accumulation of stocks of staples for own consumption. I find little evidence of increased land being allocated to staples after a hot season, except among sweet potato and rice farmers, who increase their areas planted by 18% and 9% respectively. Sweet potato in particular has been suggested by

Dercon [1996] to be a suitable candidate for risk mitigation by sample farmers from Western Tanzania, given its resistance to weather shocks. The evidence I present from a nationally representative sample of farmers suggests that this is also a strategy for reducing exposure to temperature shocks, although its scope is limited: only 10% of the farmers grow sweet potato and this share is unchanged by temperature experienced in the last growing season.

At the same time, I find that farmers increased areas planted with cash crops by 10% and the number of trees of permanent crops by 6%, for each additional HDD experienced in the past season. I also find that households are more likely to take out (mostly informal) loans for subsistence purposes after hot seasons. Taken together, these findings lend credence to the idea that farming households in Tanzania are not completely credit constrained when facing spatially correlated shocks and, while asset stocks might be playing an important role, informal financial instruments are available. The share of households who do rely on formal or informal credit markets, is, however, less than 9% of the sample, so it appears that asset stocks are still the main risk coping mechanism, as found by Chaudhuri and Paxson [2002] for the case of rural India. The limited role of credit as a consumption smoothing mechanism might be explained by the fact that 49% of the subsistence loans taken in the sample came from informal sources (mainly relatives and neighbours), who are exposed to the same, spatially correlated weather shocks.

Next, I find that after a growing season with high temperatures, households are 1 percentage point more likely to see one of their members migrate, particularly young men and middle aged women. The average migration rate in the sample is 7%, so this represents an increase of 14%. This finding contributes to a growing cluster of evidence that proves migration is responsive to weather shocks, notably by Munshi [2003], and more recently by Kleemans and Magruder [2017] and Groger and Zylberberg [2016]. In the Tanzanian context, migration has been shown to result in increases of up to 36% in individual consumption, by Beegle et al. [2011]. In this Chapter, I investigate instead whether migration of one of its members results in higher remittances received by the household of origin. Burgess et al. [2017] suggest that insurance must be limited in India, since higher temperatures, by affecting agricultural productivity, severely affect mortality rates in rural areas. I confirm this by presenting evidence that hot growing seasons are not followed by higher remittances to the household of origin, challenging the importance of this channel as a consumption insurance mechanism in the Tanzanian case.

The rest of this Chapter is organized as follows. The following Section summarizes recent findings in the literature on weather shocks and agriculture, and puts this Chapter's contribution into context. Section 2.2 describes the data sources and Section 2.3 outlines the empirical strategy. Section 2.4 presents the results and Section 2.6 concludes.

2.2 Data Sources

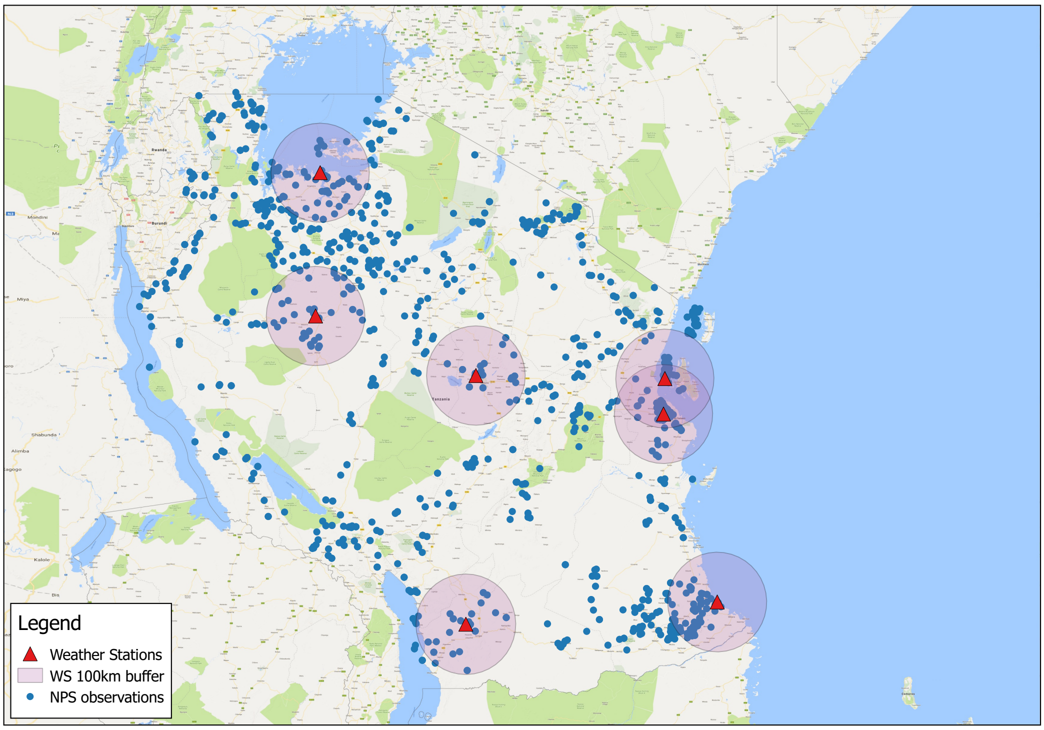
2.2.1 Weather data

Recent efforts to study the impacts of weather shocks on agricultural yields have mostly relied on monitoring station data (Deschenes and Greenstone [2007], Schlenker and Roberts [2009]) This is feasible in most advanced economies with a large number of stations covering most of their territory, but it becomes more challenging when working in developing countries. The NOAA National Center for Environmental Information provides historical daily weather measurements from monitoring stations all around the globe. The database includes information from 55,348 stations within the United States, while for Tanzania, information is available from only 14 stations. Of these, only 7 were active and have at least 50% coverage between 2007 and 2013.[1] This means that the spatial coverage of monitoring stations in Tanzania is limited: only 46% of the observations are within 100 km of their closest station, as seen in Figure 2.1. Variation in the weather by location is generally achieved by interpolating the information from multiple stations. This step

[1] Historical weather station data available here: https://www.ncdc.noaa.gov/cdo-web/

implies imposing some structure on the behaviour or weather variables, and is generally done by applying inverse distance weights, or the use of more sophisticated interpolations, such as the one used by Rowhani et al. [2011]. In this case only 21% of the sample lies within 100 km of at least two weather stations, the minimum number needed to achieve significant variation in weather outcomes across households. Furthermore, these are all located in the Dar es Salaam area, which results in an unrepresentative sample of farmers. Missing temperature and precipitation days are also common, a problem also noted by Burgess et al. [2017]: none of the stations contains information for more than 60% of the days in the sample. For all of these reasons weather station data is not a suitable source in this case. I will, however, use weather station data as a benchmark for the other sources used.

Figure 2.1: NPS observations and weather monitoring station coverage



Note: Blue dots represent NPS observations, and red triangles show the location of the seven weather monitoring stations for which data is available from NOAA. Less than 45% of the observations in the sample lie within 100 km of the nearest station.

Satellite and reanalysis data, however, are available for the whole Tanzanian territory. They can be extracted from their raw raster files directly onto the coordinate points that identify households, with no need to carry out any spatial interpolation. The use of satellite imagery in economics has been developing fast over the last decade, for uses as wide as the monitoring of droughts (Guiteras et al. [2015a]) or the prediction of poverty levels (Jean et al. [2016]). I will use land surface temperature (LST) readings from the MOD11C1 product as a proxy for air temperature. This data is collected by the Moderate Resolution Imaging Spectroradiometer (MODIS) tool aboard the Terra satellite. The product contains two daily recordings of surface temperatures on a 0.05×0.05 degree global grid, and is already cleaned of low quality readings and processed for consistency. The satellite passes over Tanzania at 8 am and 8 pm each day.[2] The high spatial and temporal resolution of the MOD11C1 product makes it a good candidate for the estimation of weather outcomes in areas with low monitoring station coverage such as Tanzania.

It is worth noting that while I will use MODIS LST values, the variable measured by weather stations and modelled by most reanalysis data products is air temperature. Evidence suggests that both variables are highly correlated, and this was confirmed for the specific case of the MODIS product by Mutiibwa et al. [2015]. Nonetheless, the reader should be aware of this distinction when comparing the results of this Chapter to other studies using re-analysis or monitoring station data.

As an additional source for robustness checks, I use a weather reanalysis dataset. The ERA-Interim gridded reanalysis product, prepared by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides daily temperature measures at 12:00 pm.[3] Interim is a reanalysis project that combines *in situ* measurements (from the few Tanzanian weather stations available), with satellite and other remotely sensed data, and inputs them into a global climate model to produce daily outputs in a 0.125×0.125 degree grid

[2] Several authors have carried out validation exercises to measure the reliability of using MODIS LST data, for example Mostovoy et al. [2005]. Data available from the USGS' Land Processes Distributed Active Archive Center (LP DAAC).

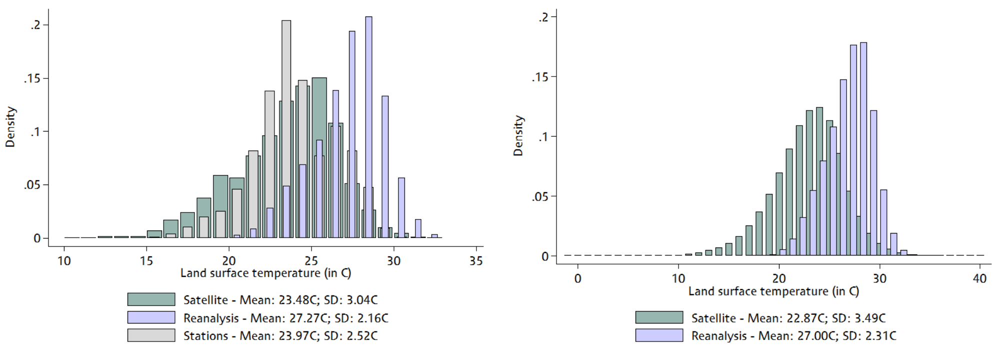
[3] Data access available at `http://www.ecmwf.int/`.

format. This is a lower spatial resolution than that of the MOD11C1 product, with squares of approximately 13 km by side at the Equator. Also, as mentioned by Dell et al. [2014], the reliability of reanalysis outputs depends on the number of different sources of primary weather data available for each particular region. Given the paucity of weather stations in Tanzania, weather reanalysis products for this area will inevitably contain a high degree of uncertainty, and will be very reliant on the assumptions of their underlying climate models used (Auffhammer et al. [2013]). This challenge questions the advantages of using reanalysis products, when high quality, satellite datasets are also available.

Precipitation data comes from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS). CHIRPS is also a reanalysis product that combines 0.05×0.05 degree resolution satellite imagery with *in-situ* monitoring station data, and builds a gridded rainfall time series (Funk et al. [2015]). This is a much smaller grid than the one provided by Interim (which also offers precipitation data), with squares of 5.6 km by side.

Figures 2.2a and 2.2b plot the distributions of average daily temperatures by source, from January 1st, 2007 to December 31st, 2013. For comparison purposes, Figure 2.2a shows the temperatures recorded by each source at the exact location of the seven Tanzanian weather stations with at least 50% coverage. Weather station average air temperature and satellite average land surface temperature have similar means, although the latter exhibits a slightly higher variance. This is reasonable, given that stations average maximum and minimum temperatures whenever they happen to occur within a day, while satellite LST takes the average of temperatures recorded daily at in its morning and evening passes over Tanzania. Interim data, which models daily temperature at midday, is on average 3°C higher, and exhibits the lowest variance. This pattern can be observed in the whole sample of NPS locations, in Figure 2.2b, as well, suggesting it is not a localized phenomenon. Given their reliance on global climate models, reanalysis data products are prone to underestimate true weather variation, while satellite sources don't. This is important because I rely on this variation to identify the impact of changing weather conditions over the growing season. For this reason, I use the Interim temperature data as a supplemental source, and take the MOD11C1 satellite product as my preferred one.

Figure 2.2: Daily temperature distributions by source



(a) Monitoring station locations **(b)** Whole sample of NPS observations

Note: Satellite data bars show the distribution of average daily recording by the MODIS instrument, that overpasses Tanzania at 8 am and 8 pm. ERA-Interim shows daily air temperature at 12 pm. Monitoring station data shows the average between maximum and minimum daily temperatures. Both Figures show temperature distributions for the whole 2007-2013 period. *Figure A:* Distribution of daily temperatures from each source at the exact location of the seven Tanzanian weather stations used, to avoid spatial interpolation errors. *Figure B:* Distribution of daily temperatures observed in all NPS locations.

On average, daily precipitation as measured by monitoring stations was of 2.3 mm, with a standard deviation of 10.2 mm. Reanalysis data sources have similar mean values, 2.6 mm from the Interim product and 2.7 from CHIRPS. These sources show significantly different variations: 4.8 mm and 8.1 mm respectively.[4] The CHIRPS dataset seems to have approached the true distribution of precipitation among the weather station locations, has a significant variation and high spatial resolution, so it will be my preferred

[4]Interim and CHIRPS mean and SDs correspond to data from the locations where the weather stations are located. Over the whole sample of NPS observations, CHIRPS has a mean 3.6 mm and a SD of 10.7 mm, while Interim has a mean of 2.6 mm and a SD of 4.5 mm. The patter of significantly lower variation within the Interim dataset is still evident.

source. [5]

2.2.2 Household and agricultural data

Tanzania's National Panel Survey was carried out biannually from 2008 to 2012, and is a nationally representative panel of Tanzanian households. The first round surveyed 3,265 households from 26 of the country's 31 regions. Rounds 2 (2010) and 3 (2012) tracked and located these same households, if they moved. The survey also followed the initial set of individuals, tracking those who set up new households, which are then included in the sample. In order to achieve a balanced panel and control for observed and unobserved heterogeneity at the farm level, I restrict the sample to those households that did not move and were successfully located and interviewed in at least two survey waves. This results in a panel of around 2,000 households for which I have complete crop level information.

The NPS includes an agricultural module with questions at the plot level regarding the type of crop and area planted, the amounts harvested, and how much of the harvest was sold. Products that were not sold but instead consumed at home were imputed with their equivalent market prices. This information refers to the Tanzanian rainy seasons (or *masika*), which runs from March to June, for the years 2008, 2010 and 2012.

Table 2.1 shows the main characteristics of farming households in the sample. Less than a third of the farms have sold part of their harvest, showing the importance of farming for domestic consumption in this context. Irrigation is extremely rare, as well as the ownership of tractors, and the use of fertilizers and pesticides. Almost all the farms in the sample grow at least one staple crop. I define staple crops as those of which farmers sell, on average, less than 50% of their production. These include: maize, rice, sorghum, millet,

[5] A third, often used source for precipitation data comes from the Tropical Measuring Mission (TRMM), launched in November 1997 and active until April 2015. Besides being no longer active, this product has a lower spatial resolution (0.25x0.25 squares) than both of the reanalysis sources compared. For comparison and robustness purposes, I ran my main specifications using TRMM data instead of CHIRPS data, and estimates are similar in magnitude, but not statistically significant due to higher measurement error.

Table 2.1: Household characteristics by NPS wave

	2008/2009	2010/2011	2012/2013
Characteristics of the Household Head			
Age	48.11	49.37	50.84
Male (%)	76.85	76.84	75.86
Completed primary education (%)	68.40	71.52	70.32
Household characteristics			
Household size	5.31	5.66	5.69
Below poverty line (%)	17.48	20.48	22.85
Agricultural production			
Sold any crops after last harvest (%)	27.03	30.62	28.91
Total value of sold crops (in Th. Shillings)	137,362	189,949	298,562
Any irrigated plots? (%)	4.03	3.49	3.40
Used inorganic fertilizer (%)	14.37	17.33	16.74
Used pesticides (%)	14.94	12.90	14.23
Hired labour (%)	43.47	36.88	43.20
Owns a tractor (%)	0.15	2.51	2.97
Grows staple crops (%)	91.37	86.68	87.62
Grows cash crops (%)	10.73	8.65	11.23
Grows permanent crops (%)	0.60	0.65	0.65
Number of different crops grown	3.59	3.72	3.85
Weather over the last rainy season			
Average daily temperature (C)	21.43	22.57	22.05
Average daily precipitation (mm/day)	414.68	441.47	394.55
Observations	1,957	2,314	2,258

Notes: Sample restricted to farming households that did not move, and were interviewed on at least two of the survey rounds.

beans, sweet potato, ground nuts and cassava. On the other hand, the main (non-permanent) cash crops (e.g. those of which more than 50% of production is sold) grown in the area are sesame, cashews, tomatoes, cotton, tobacco, onions, chickpeas, carrots and seaweed. Note that this second category includes some traditional cash crops such as tobacco and cotton, as well as others not commonly regarded as such, like seaweed, but that are relevant for the Tanzanian case. A much smaller percentage of farmers grow these market-oriented cash crops; the highest share observed is 10.73%, from the 2008 wave. Finally, permanent crops, mainly banana, mango, papaw and orange trees, are very rare and account for less than 1% of households in the NPS sample.

While most characteristics show little change over time, the share of households in the sample that are below the poverty line seems to increase with each wave of the survey. This is in line with average poverty rates in the whole NPS sample over this period.[6]

2.3 Empirical strategy

2.3.1 Agricultural production

In order to estimate the effect of weather shocks on agricultural yields, I start by assuming a Cobb-Douglas agricultural production function:

$$Y_{idt} = A_{idt} T_{idt}^{\alpha} L_{idt}^{\gamma} e^{\epsilon_{idt}}, \tag{2.1}$$

where Y_{idt} is agricultural output of farmer i in locality d at time t. T and L are quantities used of land and labor (hired and domestic), respectively. A is total factor productivity. ϵ_{idt} is a random shock that affects output after input choice is made and thus is, by definition, uncorrelated to input use.

We can express the above equation in terms of yields as:

$$y_{idt} = \frac{Y_{idt}}{T_{idt}} = A_{idt} \, L_{idt}^{\gamma} \, T_{idt}^{\alpha-1} \, e^{\epsilon_{idt}}, \tag{2.2}$$

Here T and L will be the allocations of land and labour imputed into the production function, as decided by the households at the start of growing season. The NPS asks farmers both the total area planted with each crop at the onset of the season, and total area harvested. I will use the first of these two, since the second will be endogenously determined by the weather

[6] This does not seem to coincide, though, with the decrease in Tanzania's poverty headcount ratio (at US$ 1.90 in 2011) from 53% of the population in 2007 to 47% in 2011, stated by the World Bank. The NPS uses a lower poverty line of TSh 23,933 per adult equivalent every 28 days in 2010 prices, approximately equal to 0.57 USD a day.

outcomes over the course of the growing season. For the case of L, I will use household size as a proxy of available labour in the household, which will be accounted for by household fixed effects.

I assume that A_{idt} depends on a function of local weather conditions during the growing season ($g(\beta,\omega_{dt})$), household and farm characteristics ($\mathbf{z_i}$), time-invariant local economic and environmental conditions (ρ_d), common shocks to productivity at the growing season level (ψ_t), and (potentially) other unobserved heterogeneity (υ_{idt}). In particular:

$$A_{idt} = \exp(g(\beta,\omega_{dt}) + \phi'\mathbf{z_i} + \rho_d + \psi_t + \upsilon_{idt}), \tag{2.3}$$

where $g(\beta,\omega_{dt})$ is a non-linear function to be specified later. The parameters of interest are denoted by β, which describes the relation between weather and total factor productivity.

Recent work has employed a pooled cross-section of households to tackle related questions to the ones studied here, or a panel approach at the county level (Deschenes and Greenstone [2007]). Tanzania's NPS is composed of a panel of households visited over the course of three waves, and thus allows for a more precise control for unobservable differences at the household level, as in Salazar-Espinoza et al. [2015]. I construct a panel including only those households that were interviewed on the three waves and did not move, in order to be able to control for observable and unobservable characteristics at the plot and farm levels.[7] Identification will come exclusively from weather variations from the mean at the household level, so I drop the d subscript from all but the weather variables, which are estimated at the grid level. The specification used will be the following:

$$ln(y_{it}) = g(\beta,\omega_{dt}) + \alpha_i + \psi_t + \xi_{it} \tag{2.4}$$

[7] This does not reduce the sample significantly since, as I discuss later, migration of the entire household is extremely rare: less than 1% of the households in the sample migrate entirely between any two survey waves.

The next section describes the different approaches used in the literature to model $g(\beta,\omega_{dt})$ and the alternatives used in this study.

2.3.2 Measuring the weather

The next step in the estimation of weather impacts is to choose the functional form imposed on weather variables that is the most appropriate for the purpose of this study. In line with past efforts in the literature, I model temperature outcomes using growing season degree-days. This approach allows for a flexible way to model the relationship between weather and agricultural productivity as a function of cumulative exposure to heat over the growing season, and relies on the assumption of time separability. In other words, it assumes that temperature outcomes have the same impact on output per hectare whenever they occur within a given growing season. Similar to Schlenker and Roberts [2009], I construct two measures of cumulative exposure to heat: degree days (DD) and harmful degree days (HDD). DD measures the cumulative exposure to temperatures between 8°C and τ_{high} while HDD captures the exposure to temperatures above τ_{high}, where τ_{high} is some threshold.[8] The inclusion of HDD allows for potentially different, non-linear, effects of extreme heat, compared to lower temperatures.

Formally, I define $DD = \frac{1}{n}\sum_m g^{DD}(h_m)$, with

$$g^{DD}(h_m) = \begin{cases} 0 & \text{if } h_m \leq 8°\text{C} \\ h_m - 8°\text{C} & \text{if } 8°\text{C} < h_m \leq \tau_{high} \\ \tau_{high} - 8°\text{C} & \text{if } \tau_{high} < h_m, \end{cases}$$

In this case, h_m will be the average land surface temperature recorded by the MODIS instrument, as it passes over Tanzania at 8 am and 8 pm, on day m. n is the total number of days in a growing season, which in this case is equal

[8] The lower threshold of 8°C is common in the literature, and comes from the fact that temperatures below this level do not contribute to plant development significantly. Additionally, in the Tanzanian case, average temperatures below that 8°C are extremely rare, so our estimates are unchanged by omitting this lower threshold completely.

to 122. Similarly, $HDD = \frac{1}{n}\sum_m g^{HDD}(h_m)$, with

$$g^{HDD}(h_m) = \begin{cases} 0 & \text{if } h_m \leq \tau_{high} \\ h_m - \tau_{high} & \text{if } \tau_{high} < h_m \end{cases}$$

A key issue when using a degree-day approach is where to set the value of threshold τ_{high}. The literature has usually assumed a threshold of around 29-30°C (Deschenes and Greenstone [2007], Lobell et al. [2011]) beyond which higher temperatures become detrimental for crop development, based on botanical evidence. Other authors have estimated crop specific thresholds from their data, such as Schlenker and Roberts [2009] or Tack et al. [2017]. These estimates are likely to be crop and context dependent and hence might not be transferable to the Tanzanian case. A second challenge in this case is that optimal growing temperatures are generally estimated using maximum daily temperatures. MODIS measures of land surface temperature are carried out each day at 8 am and 8 pm, when the Terra satellite overpasses Tanzania, and will presumably be lower than the actual maximum temperatures reached during each day. Therefore I estimate what that threshold is by an iterative regression method, as the one used in Schlenker and Roberts [2009] and Tack et al. [2017]. This method consists of a simple algorithm by which (1) I create DD and HDD variables with values of τ_{high} ranging from 20°C to 40°C then (2) I estimate model (2.4) using log of output per hectare as outcome, and (3) I pick the threshold that maximizes model fit (within R squared). Using this data driven approach, I identify τ_{high} for the sample under analysis. Results of this exercise are discussed in Section 2.4 below.

Another alternative often discussed in the literature is the 'binned' approach. By constructing 1°C bins that include the share of days in a growing season that experienced temperatures within each range, this method is more flexible, as it does not impose any constraints on the shape of the relationship between temperatures and yields. However, in order for these bins to contain enough variation, one needs a detailed profile of intra-day temperatures, not just one or two measures a day. Cesaraccio et al. [2001] and Schlenker and Roberts [2009] use daily maximum and minimum temperatures and impose a sinusoidal curve on them to simulate the whole trajectory of daily temperatures. With this, they then build bins for the fraction of the growing season spent in each temperature interval, and obtain a distribution with significant variation. On the contrary, day and night satellite land surface temperatures are not the maximum and minimum daily temperatures, which usually occur after midday and before sunrise. This is the case for most applications of satellite weather data, when the user does not have a choice on the timing of the recordings available to her. Modelling intra-day temperatures would require making additional assumptions on the profile of daily temperatures based on the small sample of weather stations that record them, so I disregard this approach here.

I measure exposure to rain using average daily precipitation (PP) during the growing season and its square. To facilitate interpretation of coefficients, I rescale both variables to be expressed in terms of standard deviations from the mean. The resulting function of weather outcomes is then:

$$g(\beta, \omega_{dt}) = \beta_0 DD_{dt} + \beta_1 HDD_{dt} + \beta_2 PP_{dt} + \beta_3 PP_{dt}^2. \tag{2.5}$$

Several other measures of precipitation were proposed by the literature, in recent years. Because the timing, magnitudes and distribution of rainfall might be important for crop growth, other authors have suggested measures that provide more information on this over the growing season. I will test the explanatory power of several of these measures for the Tanzanian case: the number of rainy days, the longest dry spell, and the number of heavy rainfall (>100 mm) days. The construction of these variables is straightforward.

2.3.3 Spatial correlation

Weather phenomena exhibit varying degrees of temporal and spatial correlation. If unaccounted for, this can lead to over rejection of the null hypothesis due to under-estimated standard errors. I estimate standard errors robust to spatial and serial correlation following Conley [1999].[9] I allow for correlation of any kind between observations, that decays with distance up to a cut-off at 500 km, and autocorrelation for up to six years.

2.4 Weather Impacts on Agricultural Yields

2.4.1 Degree-days and optimal temperatures

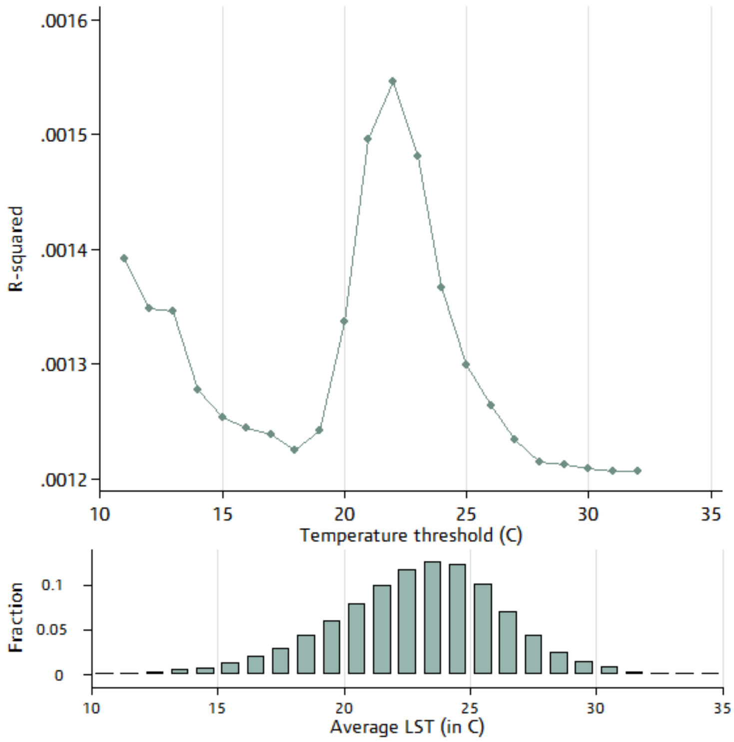
Results for the iterative approach applied in the estimation of optimal temperature thresholds using a panel specification are shown in Figure 2.3. The Figure plots the resulting R^2 coefficients of individual regressions with aggregate crop yield at the household level on the left hand side and DD/HDD, taking different temperature thresholds, on the right hand side, together with controls for precipitation, household and year fixed effects. The threshold used in each regression is plotted on the x-axis. In order to make meaningful comparisons, I express R^2 coefficients net of the fixed effects, that is, I show only the component explained by the independent variables of interest, namely DD, HDD and precipitation.

Optimal temperatures for agricultural yields can be expected to lie somewhere between 21°C, the optimal temperature for human productivity estimated by Seppanen et al. [2006]; and 33°C the threshold estimated by Tack et al. [2017] for sorghum, a highly heat resistant crop. Figure 2.3 shows that the thresholds estimated in this case are on the low end of this range. We see that model fit using satellite data is maximized at 22°C.

How does this threshold of 22°C compare to other values estimates recently in the literature? Temperatures measured at 8 am and 8 pm are usually far from the daily maximum and generally below the daily average temperature. To benchmark this estimate, recall that, as mentioned in the previous section, average land surface temperatures of 23°C were estimated from the locations where monitoring stations were located, and where these in turn registered

[9] I implement these using code from Hsiang [2010] and adapted by Thiemo Fetzer in the *reg2hdfespatial* command for Stata.

Figure 2.3: Model fit by temperature threshold



Note: R^2 coefficients of regressions of yield on DD/HDD specifications, using different thresholds to split between DD and HDD. These thresholds are plotted on the x-axis. All specifications include controls for precipitation and household and year fixed effects.

average temperatures of 24°C. Indeed, I regressed daily temperatures (from monitoring stations) on land surface temperatures (from satellite), and found that a land surface temperature of 22°C predicts an average temperature of 25°C. These estimates in line with the work by Burgess et al. [2017], who find that mortality in rural India increases during growing seasons experiencing more days with temperatures above 24°C (75 F). Other authors find higher thresholds for agricultural yields using data from the US. For the case of corn yields, for example, Schlenker and Roberts [2009] calculate the optimum at 29°C, the same as Burke and Emerick [2016] and Lusk et al. [2017]. In earlier work, Richie and NeSmith [1991] established a 32°C threshold as a reference for the whole US agricultural sector, commonly used in studies that estimate temperature impacts on aggregate yields since Schlenker and Roberts [2009].

My findings, together with the work of Burgess et al. [2017], suggest that this threshold may be lower in developing countries, where agriculture is less capital intensive and more reliant on human labour.

These first results show that optimal growing temperatures can be estimated using satellite products which have the resolution and spatial coverage needed to carry out this analysis is practically any part of the globe, irrespective of the availability and quality of *in situ* measurements. Satellite sources have the additional advantage of not relying on global climate models, used in the assimilation of data for reanalysis datasets.

2.4.2 Weather impacts on aggregate yields

Using the threshold calculated above, I estimate the impact of different weather variables on agricultural yields, measured as the quantity harvested of each crop divided by the total area planted with it at the onset of the growing season. The panel specification with year fixed effects controls for observed and unobserved heterogeneity at the farm level, so identification of effects comes from changes in weather outcomes at the household level, after controlling for nation-wide shocks in each season. Results are presented in Table 2.2. The dependent variable here is the logarithm of crop yields and I include all farm-crop combinations.

Table 2.2 presents point estimates from an OLS estimation of agricultural yields on temperature and precipitation over the course of the last, complete, rainy season. The first panel of Table 2.2 compares the coefficients of specifications modelling temperature outcomes using the degree-day approach. Column (1) shows the results of my preferred OLS specification. It shows no statistically significant effect of degree-days on yields. Harmful degree days, on the other hand, show negative and significant effects of 8 percentage points, as expected from the non-linear relationship between temperature and yields widely documented in the literature. This is consistent with what was found by Schlenker and Roberts [2009], who note that the slope of the decline above the optimum temperature is steeper than its incline before it. This is also a

Table 2.2: Weather impacts on agricultural yields

	(1)	(2)	(3)	(4)	(5)
Temperature:					
Average DD	0.02	0.33	0.01	0.00	0.02
	(0.04)	(0.30)	(0.03)	(0.03)	(0.04)
Average HDD	-0.08***	-0.16**	-0.09***	-0.09***	-0.08***
	(0.03)	(0.07)	(0.03)	(0.03)	(0.03)
Precipitation:					
Total, std	0.16*	0.22			0.17*
	(0.09)	(0.14)			(0.09)
Squared, std	-0.10	-0.09			-0.10
	(0.07)	(0.09)			(0.07)
Number of rainy days (>0.1 mm)			-0.01		
			(0.01)		
Number of rainy days, squared			0.00		
			(0.00)		
Longest dry spell in RS (days)				0.00	
				(0.00)	
Longest dry spell, squared				-0.00	
				(0.00)	
Number of heavy rain days (>100 mm)					0.02
					(0.03)
Fixed Effects:					
Growing season FE's	Yes	Yes	Yes	Yes	Yes
Household FE's	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	IV	OLS	OLS	OLS
Observations	14,486	14,001	14,486	14,486	14,486

Notes: Standard errors robust to spatial and serial correlation following Conley [1999], using code from Hsiang [2010] and adapted by Thiemo Fetzer, in parenthesis. They assume a discrete cut-off for spatial correlation of errors at 500 km and six period lags. Stars indicate statistical significance: **p <0.10, ** p <0.05, *** p <0.01.*

reasonable finding for a region in which, as can be seen from the bottom panel of 2.3, 63% of the days have average temperatures above 22°C. This impact is economically significant: the mean and standard deviation of average harmful degree days in the sample is 1.2 and 1.0 respectively, meaning that a 1°C increase is equivalent to a 1 SD increase.

The negative impacts of additional harmful degree days are confirmed and are twice as high when I instrument MOD11C1 degree day variables with average temperatures from the Interim gridded product. While both data sources contain some level of measurement error, instrumenting one with another can recover reliable estimates, as long as errors of the two datasets are uncorrelated. While the sign and statistical significance of the estimates in Column

(2) are reassuring, they also have a higher variance, possibly due to the lower spatial resolution of the Interim product. [10] For this reason, I do not use IV estimates as my preferred specification in the following estimations.

Columns (1) through (5) of Table 2.2 also test the explanatory power of several measures of precipitation over the growing season. Column (1) shows that a 1 SD increase in total precipitation results in yields 15 percentage points higher. In this sense, additional precipitation has the potential to offset increases in temperatures. The negative but statistically insignificant coefficient for the squared term suggests a non-linear relationship between precipitation and yields, although it is harder to detect given the skewed distribution of daily precipitation rates towards zero. Columns (3) to (5) study the explanatory power of other measures often discussed in the literature. None of these measures provides statistical significant results, and I will therefore not include them in my preferred specification, that will instead be equivalent to that shown in Column (1).

2.4.3 Results by crop

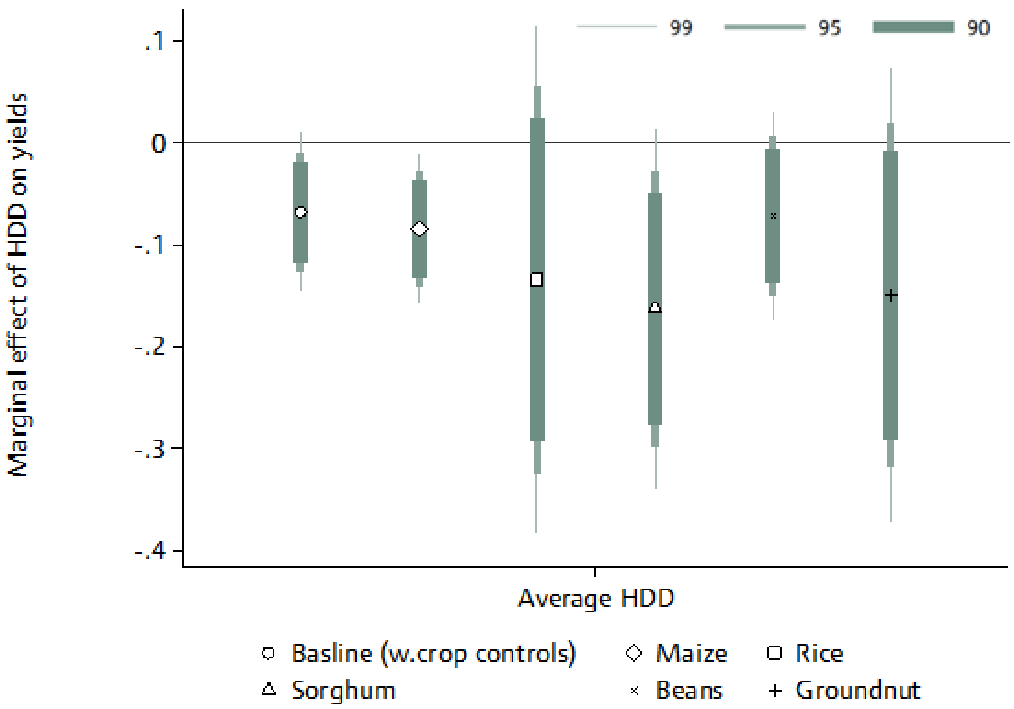
An important concern that may arise from the results discussed so far, is the possibility that the impacts estimated are driven by one particularly heat-sensitive crop variety. The HDD point estimate from Table 2.2 is averaged over all plots in the sample, and therefore over crop types so if this is indeed the case, then our conclusions would be restricted only to farmers growing this (or these) particularly vulnerable crops.

In order to check that this is indeed not the case, I first run the same regression as the one presented in Column (1) of Table 2.2 introducing crop controls. The first bar in Figure 2.4 shows that the point estimate for HDDs in this case is -0.07, in line with the baseline estimates presented above, and also significant at the 5% level. I then run the same analysis restricting the

[10] A weak correlation between MODIS and Interim weather datasets could also be behind these larger standard errors, and potentially inconsistency of the IV estimates (Wooldridge [2010]). This is rejected by the F-statistics from the first stage regressions of both instrumented variables, DD and HDD, which are $F = 9.7$ and $F = 42.6$, respectively, and by the Kleibergen-Paap rk Wald F test for weak identification.

sample to each of the 5 most popular crops in the NPS dataset. Four out of five of these crops show similar HDD coefficients as the whole sample, in both magnitude and significance. This is reassuring, as it shows that no single crop appears to be driving the results for the whole sample.

Figure 2.4: Marginal Effect of 1 Additional HDD on Yields by Crop



Note: Bars plot the coefficients of an OLS regression at the plot level with agricultural yield as the dependent variable. The leftmost bar shows the HDD point estimate of a specification identical to that from Column (1) in Table 2.2, but including crop controls. The remaining estimates come from crop-specific regressions. Bars represent confidence intervals at of 90%, 95% and 99% levels. Household and year fixed effects included. All specifications include controls for DDs, total precipitation and its square. Standard errors robust to spatial and serial correlation following Conley [1999], using code from Hsiang [2010] and adapted by Thiemo Fetzer, in parenthesis. They assume a discrete cut-off for spatial correlation of errors at 500 km and six period lags.

Overall, Table 2.2 and Figure 2.4 show that both temperature and precipitation play an important role at determining agricultural yields in Tanzania, and that satellite and reanalysis datasets can effectively capture these effects. Given the lack of weather monitoring station data in the developing world, it is hard to understate the number of potential uses of open access, high quality weather gridded products such as those from the MODIS instrument.

While the impact of droughts and floods on subsistence agriculture contexts has been discussed extensively before, the effects of high temperatures have received somewhat less attention. The precise estimation of optimal temperature thresholds for Tanzanian agriculturalists, and the initial evidence discussed above suggesting important losses from higher temperatures, motivates me to focus on these impacts in the following sections. I will aim to shed some initial light on the different margins of adjustment that farming households in the sample have available, and how high temperatures affect them.

2.5 Weather shocks and farmer reactions

Most of the farms in this sample are small, family owned plots, that focus on staple crops for own consumption. Over the three years in which the NPS was carried out, between 87% and 91% of the households declared to grow at least one staple crop. At the same time, more than 80% of the production of staple goods was not sold, but destined to self-consumption. It is of interest then to establish what the impact of the weather was on the yields of staple and cash separately, to further understand the level of exposure to weather risk of these households.

Table 2.3 presents the results of splitting the sample according to the type of crop, and repeating the analysis above. To this end, I define staple crops as those of which farmers sell, on average, less than 50% of their production. These include: maize, rice, sorghum, millet, beans, sweet potato, ground nuts and cassava. On the other hand, the main (non-permanent) cash crops (e.g. those of which more than 50% of production is sold) grown in the area are sesame, cashews, tomatoes, cotton, tobacco, onions, chickpeas, carrots and seaweed. Households may of course grow both kinds of crops, but, recall from Table 2.1, only around 10% of the households in the sample grew any cash crops in each survey wave.

Table 2.3: Impacts on crop yields and farmer decisions

	ln(Yields) (t)		ln(Area planted) (t+1)		
	(1)	(2)	(3)	(4)	(5)
Crop type:	Staple	Cash	Staple	Cash	Perm
Temperature:					
Average DD	0.05*	-0.18	0.04	-0.01	-0.10***
	(0.03)	(0.12)	(0.03)	(0.04)	(0.03)
Average HDD	-0.08***	-0.05	0.03*	0.10***	0.06***
	(0.03)	(0.07)	(0.02)	(0.03)	(0.02)
Controls:					
Growing season FE's	Yes	Yes	Yes	Yes	Yes
Household FE's	Yes	Yes	Yes	Yes	Yes
Observations	10,030	666	10,380	677	9,780

Notes: All specifications include controls for total precipitation and its square. Standard errors robust to spatial and serial correlation following Conley [1999], using code from Hsiang [2010] and adapted by Thiemo Fetzer, in parenthesis. They assume a discrete cut-off for spatial correlation of errors at 500 km and six period lags. Stars indicate statistical significance: **p <0.10, ** p <0.05, *** p <0.01.*

Columns (1) and (2) study the impact of weather variables on yields for the two types of crops. Point estimates for HDDs are negative and of similar magnitude in both cases. They are only statistically significant in the case of staple crops, for which the non-linear relationship between temperature and yields, characterised by a positive effect of DDs and a negative effect of HDDs, is estimated precisely. This is a reassuring finding, since it agrees with past studies on the impact of temperatures of yields such as Deschenes and Greenstone [2007], Schlenker and Roberts [2009]. Cash crops do not appear to respond in the same way to temperatures, although the sample size is significantly reduced. This difference in the relationship between temperature and yields for staple and cash crops, does not seem to stem from differential access to irrigation, which is extremely low for all crop types. The share of farms with irrigated plots, among those who grow cash crops is 4.2%, only slightly higher than the share among farms who grow staples, 3.6%. Nonetheless, the access to irrigation might help farmers mitigate the effect of high temperatures, so previous studies using county level data have sometimes omitted areas with high shares of irrigated land. By using a panel at the farm level, in this case these characteristics are controlled for by the farm level fixed effects.

These columns suggest that, together with the well documented negative impact of droughts, there is a meaningful effect of temperatures on yields, particularly of staple crops, that should not be omitted from future studies. This more nuanced approach to the meaning of weather shocks is relevant, particularly for the line of research that uses them to instrument for household income (for example, Hidalgo et al. [2010]) or economic growth (for example, Miguel et al. [2004]).

Weather outcomes over the course of a growing season might also condition the farmer's future decisions. A common risk coping mechanism in areas with incomplete credit markets such as rural Tanzania is the build up of stocks (Dercon [2005]). A consumption-smoothing farmer that experiences a negative shock to yields in year t, might reduce their stocks of staple foods for own consumption during that year. In year $t+1$ this farmer could decide to increase the area planted with staple crops, in order to replenish their stocks, at the cost of cash crops, and therefore cash income. Salazar-Espinoza et al. [2015] analysed a sample of farmers from Mozambique and found evidence suggesting that indeed, they operate with a buffer stock of staple crops, and reduce areas planted with permanent and cash crops after experiencing a drought, in order to replenish it. An alternative mechanism could arise if formal or informal borrowing is available and imperfectly correlated with weather shocks: farmers might increase the planted area of cash crops in the following harvest, in order to raise funds for postponed purchases or to repay loans taken up during the bad growing season. Columns (3) to (5) of Table 2.3 study how farmers react to past weather shocks by looking at their crop choice decisions. The dependent variable in this case is the logarithm of the area planted with each crop during the following growing season.

Column (3) shows the relationship between temperature in one season and the area planted with staple crops in the next. Areas planted with staples increase by 3 percentage points for each HDD experienced in the previous season, significant at the 5% level. In a study of farming households from the semi arid area of Western Tanzania, Dercon [1996] found that sweet potato provided the least risk of total crop failure when faced with droughts. The author poses that the choice of sweet potato cultivation, a highly weather resistant tuber, is indeed a risk management mechanism that comes at the cost of reduced returns, given the crop's low price. In this nationally representative sample, around 10% of the households cultivate sweet potato in any particular growing season.

In order to explore the effects of temperature on staple crop growing decisions, I ran the same analysis as the one showed in Column (3) for three staple crops: sweet potato and the two most popular staples, maize and paddy. I study both intensive (e.g. increased land assigned to these crops) and extensive (e.g. more farmers growing these crops) margins. On the extensive margin, I find that the number of households who chose to plant each of these crops is mostly unreactive to HDDs experienced in the past growing season, except for the case of paddy, where households appear to be 1 percentage point more likely to grow it for each additional HDD experienced in the previous season. On the other hand, on the intensive margin, I find the that the total area planted with paddy and sweet potato increases by statistically significant 9 and 18 percentage points respectively. These results, included in Table A.2 in the Appendix, complement Dercon's, and suggest that both rice and sweet potato play a role in risk management for farms exposed to temperature risk. It seems, however, that this role is mostly restricted to the households who already grew these crops in the past (22% for paddy, 10% for sweet potato). It becomes clear that the while crop switching seems fairly rare, farmers indeed shift the total areas dedicated to each crop as a reaction to past weather outcomes.

Going back to Table 2.3, strong reactions are observed when looking at cash and permanent crops as well. Column (4) shows that the area planted with cash crops increases by 10 percentage points with each additional HDD experienced in the previous growing season. Contrary to what was found by Salazar-Espinoza et al. [2015], this finding lends support to the hypothesis that, after a season of reduced yields, households increase cash crops area shares in order to repay loans used for consumption smoothing. I will pursue this channel further in the following subsection, where I will check whether there is evidence of higher loan take up in seasons with high temperatures. Land allocation decisions do not seem to have a clear relationship with past rainfall outcomes.

Tanzanian agriculturalists have a third source of food and income that is not included in these two categories: permanent crops, mainly banana, mango, papaw and orange trees. Calculating yields for permanent crops is complicated by the fact that farmers frequently inter-crop trees in their plots. The NPS does, however, have information on the amount of new trees planted at the start of each growing season I this as a proxy for the decision to invest more resources into the production of permanent crops, and use the logarithm of new trees as the dependent variable of Column (5). The number of trees increases as well, by 6 pp, for each additional harmful degree day experienced in the previous season. On the contrary, it seems like, as in the case of staple crops, investment into permanent crops increases after a growing season with high rainfall in a non-linear way.

To understand the implications of this reaction further, I checked whether there was any significant increase in the amount of land owned or purchased after a hot season and found no evidence of this. Given that the total area owned by farmers remains stable, this increase in areas planted with both staple and cash crops can be possible either through a more intensive use of land, via intercropping, or by the use of land otherwise left fallow. [11] In both cases, the higher intensity in the use of land comes at a cost: intercropping will provide a wider range of crops but lower yields, and fallow lands are low quality, and meant to stay out of the production function in order to recover

[11] Intercropping is a strategy, used by 68% of the farms in the sample, by which multiple crops are grown in the same plot.

the soil.

2.5.1 Lagged effects

A reasonable concern that could arise given the results presented above, is that weather impacts might have longer term effects. Increased cultivation of cash and permanent crops in subsequent rainy seasons might drain resources from the growing of staples and this could propagate the drop in yields further over time. I test for this by including lagged terms for each weather outcomes and find no evidence of lagged effects. The results, presented in Table A.1 in the Appendix, suggest instead that the impact of weather shocks on yields is only detectable in the short term.

2.5.2 Impacts on household finances

The evidence presented in Table 2.3, points towards farmers reacting in two ways to high temperature shocks. Those growing sweet potato or rice increase their land allocations for this crop. There is no evidence of other farmers taking up this crop, however. At the same time, those cultivating cash and permanent crops increase the amount of land (or number of trees) planted with these, thereby inputting more land into the production of market oriented crops. Both reactions are consistent with consumption smoothing behaviour, but of different kinds. The first group aims at insuring the future harvest against total loss, the second aims at increasing market sales at the end of the harvest, possibly to repay outstanding debts. Table 2.4 explores these possibilities in more depth.

The risk management strategies described in the previous section aim at reducing risk in subsequent campaigns. The question about what risk coping strategies households rely on once they have suffered a shock is still pending. Column (1) from Table 2.4 shows that consumption levels per capita are not affected significantly by temperature experienced during the previous growing season. The rather precisely estimated zero coefficients for temperature measures, when combined with the documented fall in yields of staple crops,

Table 2.4: Impacts on Household Consumption and Debt

	Cons./capita	Debts				
Dep var:	(1) ln(TSh)	(2) ln(TSh)	(3) Any	(4) Subs	(5) Bus	(6) Oth
Temperature:						
Average DD	-0.00	0.02	-0.01	-0.01	-0.00	0.00
	(0.01)	(0.15)	(0.01)	(0.01)	(0.01)	(0.00)
Average HDD	0.00	0.05	0.01*	0.01***	0.01	-0.00
	(0.01)	(0.06)	(0.01)	(0.00)	(0.00)	(0.00)
Controls:						
Growing season FE's	Yes	Yes	Yes	Yes	Yes	Yes
Household FE's	Yes	Yes	Yes	Yes	Yes	Yes
% HH with debts			8.42	4.35	3.60	1.40
Observations	7,497	639	7,588	7,588	7,588	7,588

Notes: Consumption: DV is the natural logarithm of total, annual, real expenditures per adult equivalent living in the household. This measure includes all consumed goods and services over the past year, both in and outside the household, and including both purchases as well as produced goods and gifts. Debts: Column (2) DV is the natural logarithm of total debts taken up, in thousands of Shillings; Columns (3) to (6) DV equals 1 if household took up loans for any, subsistence, business or other purposes. All specifications include controls for total precipitation and its square. Standard errors robust to spatial and serial correlation following Conley [1999], using code from Hsiang [2010] and adapted by Thiemo Fetzer, in parenthesis. They assume a discrete cut-off for spatial correlation of errors at 500 km and six period lags. Stars indicate statistical significance: **p <0.10, ** p <0.05, *** p <0.01.*

suggests that households do perform some smoothing, either from relying on their own stocks of staples or savings, or by taking loans. NPS data does not distinguish between consumption from contemporary own production and from accumulated stocks from previous growing seasons, and does not contain detailed enough information on savings, so I cannot test the importance of the former channel. The total size of loans, estimated from the sample of households who indeed had them, did not increase in a statistically significant way, as shown in Column (2). However, Column (3) confirms that households are 1 pp more likely to have taken up loans for each additional harmful degree day experienced in the past growing season. From an average of only 8.4%, this represents a increase of 12% in the number of households taking up debt for each additional HDD experienced.

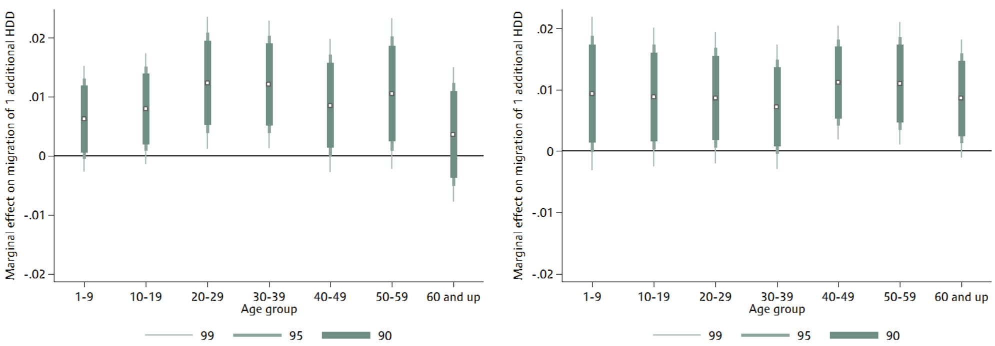
Columns (4) to (6) from Table 2.4 separates loans by main purpose. Column (4), for example, shows that households are 1 pp more likely to take up loans directed at subsistence purposes for each additional harmful degree day experienced, and this estimate is statistically significant at the 1% level. Subsistence loans include a general a category of "subsistence needs", as well as school fees and medical needs. Columns (5) shows an point estimate of similar magnitude but not statistically significant for the take up of loans directed at business related purchases. Business related loans include the purchase of land for agricultural purposes, agricultural or other inputs, and agricultural machinery.

Subsistence loans are mostly taken from neighbours and relatives, who account for 49% of these in the sample. These loans will then have to be repaid, which might explain the increase in cash and permanent crop allocations in subsequent growing seasons observed in the previous Table. Other loans, meant to pay for celebrations, purchases of dwellings and other expenses, are unaffected by weather outcomes. Table 2.4 established two important findings. First, that consumption seems to be smoothed to some degree by households. Reductions in yields from hot or dry seasons established above do not translate into lower consumption levels. It also confirmed that, at least for a small share of the farmers in the sample, credit is one available risk coping channel. The scope of this channel is limited, in theory, by the fact that since most loans are local in origin, spatially correlated shocks such as weather phenomena cannot be effectively insured against. This might explain the low take up of these loans in the sample (4% of the sample). The mechanism for consumption smoothing applied by the rest of the households in the sample remains unclear so far.

2.5.3 Migration

The final margin I explore is that of migration. Migration has been considered part of a household's risk coping mechanisms at least since Rosenzweig and Stark [1989]. Weather shocks have been shown to trigger migration as well by Groger and Zylberberg [2016], Kleemans and Magruder [2017], Munshi [2003] among others. Households might be able to attenuate the losses accruing from weather related income shocks by seeing one of their members move out, thus reducing the number of dependants, and, possibly, by increased remittances received from members living elsewhere. The NPS tracked its initial sample of households, locating and interviewing all its members in each of the two subsequent waves. Full migration, meaning the displacement of all the members of the household, is extremely rare in this sample: between 2008 and 2010, only 1% of the households moved entirely, and this was the period with the highest share. On the other hand, migration of some members of the household is more common. During the same period, 12% of households in the sample saw at least one of their members move out.

Figure 2.5: Marginal Effect of 1 Additional HDD on Propensity to Migrate by Age and Gender



(a) Men **(b)** Women

Note: Bars plot the coefficients of an OLS regression at the individual level with migration as the dependent variable, and the number of DDs/HDDs experienced in the previous growing season interacted with age group in the right hand side. Household and year fixed effects included. All specifications include controls for total precipitation and its square. Standard errors robust to spatial and serial correlation following Conley [1999], using code from Hsiang [2010] and adapted by Thiemo Fetzer, in parenthesis. They assume a discrete cut-off for spatial correlation of errors at 500 km and six period lags.

To study this, I first ask whether extreme temperatures have an effect on the propensity of household members to move. I allow for weather impacts to vary by gender and age group, since migrations rates among age extremes are low and might hide meaningful impacts. Figures 2.5a and 2.5b plot the estimated effects of an additional harmful degree day on the previous growing season, on the probability of an individual to move. It becomes clear from the Figures that economic migration patterns throughout the life cycle are different by age and gender. Hot temperatures push men between the ages of 20 and 39 to migrate by an additional 1 pp for each HDD, and this is significant at the 1% level. Figure A.1 in the Appendix shows that this 20 to 29 is the age at which men migrate the most, on average, for any reason. However in the case of women, who normally migrate when in their 20s, extreme temperatures increase their propensity to migrate the most when they are aged 40 to 59. These migration patterns may be related to family structure and different labor market conditions at towns and cities, for each of the genders. In any case, this increased rate of out-migration in the face of hot temperatures poses another channel by which the households in the sample may avoid drops in consumption per capita.

The second question of interest is whether migrants indeed help consumption smoothing of their household of origin in the face of extreme temperatures or precipitation shortages. This can be investigated by looking at the remittances that households declare to receive. Due to coding inconsistencies across waves, only the 2008 and 2012 waves can be used for this comparison. I find little evidence to support the hypothesis that households use increased remittances for migrant members as an instrument of consumption smoothing. As can be seen in Table 2.5, domestic remittances do not react in a statistically significant way to additional harmful degree days, although these do increase with beneficial degree days. This implies that domestic remittances do not serve as insurance against bad harvests, since they move in identical directions (e.g. with DD), or do not change at all. This finding is in line with recent evidence presented by Burgess et al. [2017], who show that rural households in India are indeed very imperfectly insured against weather shocks.[12]

Domestic remittances might be dominated by household members living in nearby towns or villages, and might therefore be subject to the same, spatially

[12]Domestic remittances also exhibit no relationship with precipitation levels at the location of the receiving household.

Table 2.5: Effects of the weather on remittances received by households

Dep.var.: Remittances received from:	Domestic		Foreign	
	(1) ln(Sh)	(2) Yes/No	(3) ln(Sh)	(4) Yes/No
Temperature:				
Average DD	0.38***	0.03***	-0.03	-0.00
	(0.12)	(0.01)	(0.03)	(0.00)
Average HDD	-0.00	-0.00	0.02	0.00
	(0.09)	(0.01)	(0.02)	(0.00)
Controls:				
Growing season FE's	Yes	Yes	Yes	Yes
Household FE's	Yes	Yes	Yes	Yes
Years included	2008,2012	2008,2012	2008,2012	2008,2012
Observations	7,588	7,588	7,588	7,588
Mean remittance, if received (th. Sh.)	163.57		339.05	
% HHs who received		12.65		0.33

Notes: All specifications include controls for total precipitation and its square. Standard errors robust to spatial and serial correlation following Conley [1999], using code from Hsiang [2010] in parenthesis. They assume a discrete cut-off for spatial correlation of errors at 50 km and six period lags. Stars indicate statistical significance (assuming district-level clustering): **p <0.10, ** p <0.05, *** p <0.01.* All models include household controls: age, age squared, gender, and level of education of the household head.

correlated weather shocks as the household of origin. In this sense, they might not be well suited for insuring against income shocks. While the NPS does not contain detailed enough information on the origin of these remittances, it does include a separate category for remittances received from abroad. Foreign remittances are more suitable candidates for income insurance, since they are not likely to be subject to the same shocks as the household of origin. However, Columns (3) and (4) show that this is hardly the case. Households are not more likely to receive higher remittances after experiencing more or less DD/HDDs, as seen in Column (3). More importantly, just 0.33% of the households in the sample receive remittances from abroad, and this share is not affected by weather conditions, as seen in Column (4).

All in all, it appears that while weather shocks might indeed push Tanzanian individuals to migrate, particularly men in prime-working age and middle aged women, this does not translate in detectable increases in transfers to the households members left behind. This does not mean that individual income and consumption do not benefit from migration: as Beegle et al. [2011] show, the potential gains from migration at the individual level in Tanzania are large. The evidence I present shows instead that remittances from members living elsewhere do not appear to be a significantly large mechanism for consumption smoothing among household of origin, in the face of temperature shocks.

2.6 Conclusion

This Chapter makes two main contributions to the existing literature. First, it shows that satellite data is a useful tool for capturing weather conditions in areas where monitoring stations are sparse or completely unavailable. Satellite products have a higher variation than reanalysis products in these areas, because they don't rely on *in situ* measurements nor global climate models, and have high spatial and temporal resolutions. I show that a regular finding in the literature, namely the non-linear relationship between temperature and agricultural yields, can be detected using MODIS satellite temperature and CHIRPS reanalysis precipitation data, two public datasets with almost global coverage.

The second main contribution is the analysis of farmer responses, using a panel of agricultural households in Tanzania. I find that as well as low precipitation, extreme temperatures significantly affect yields of staple crops. This has important implications for studies on rural income dynamics, as they highlight the importance of an often omitted variable. Households react to this by increasing the area planted with cash crops and the number of trees of permanent crops, during the following growing season. This is consistent with a model in which households smooth consumption during bad harvests by taking loans, which I also observe. Finally, I document the migration patterns driven by extreme temperatures and show that although young men and middle aged women are indeed more likely to migrate after increased temperatures, this does not translate in higher remittances for the household of origin. Household consumption levels do not seem to be affected by weather outcomes, however, meaning that the mechanism for consumption smoothing among the large majority of households that do not grow cash or permanent crops remains unclear, and possibly related to asset stocks.

Chapter 3

Temperature and agriculture: how do farmers respond to extreme heat?

3.1 Introduction

Extreme weather events are forecasted to become a more regular phenomenon over the course of this Century (IPCC [2014]). Subsistence farmers in developing countries will be exposed to potentially large losses, and their capacity to adapt will be put to the test. Chapter 1 showed how weather gridded products, such as satellite and reanalysis data, can be useful tools for the research community in understanding the impact of extreme weather events. Using a simple model of production and consumption households, in this Chapter we will aim to shed further light on the nature of temperature shocks, the magnitude of their impact in two very different geographical contexts, and the varying ways in which farms react to them.

This Chapter uses a unique micro-dataset of Peruvian farmers for years 2007-2015 and combines it with high resolution, daily weather data. In line with the findings from Chapter 1, we show how satellite imagery, geo-matched with existing household surveys, can help overcome shortcomings in meteorological data, a major constraint to study the economic effects of temperature in developing countries. The public availability of these data means that it is feasible to replicate our study in different countries.

Our study then provides two key contributions to the literature. First, we estimate the relationship between temperature and agricultural productivity using observational data at farm level. Most previous studies have identified this relation using controlled agronomic experiments (see for instance Abrol and Ingram [1996] or Peng et al. [2004]) or data aggregated by administrative units, such as U.S counties, or African countries (see for example, Schlenker and Roberts [2009]).[1] While informative, these approaches have a number of limitations. Agronomic experiments may fail to account for relevant human, institutional, and technological factors. Similarly, using aggregated data may hide important within-unit heterogeneity and attenuate the estimated effects.[2] This Chapter contributes to the literature by providing farm-level estimates, from a nationally representative sample of farmers.

Second, we estimate the effect of extremely hot temperatures on agricultural yields. The use of rich micro-data and a production function approach allows us to examine how this effect is driven by changes in total factor productivity, output and input use. Thus, we can examine the short-run economic responses of farmers, that help us understand farmers' ability to mitigate weather shocks. As Burke and Emerick [2016] show evidence that over recent decades farmers had a limited ability to adapt to climate change, our results are also informative about potential long term effects.

Our empirical analysis focuses on the coast and highlands of Peru. These two geographical regions have significantly different socio-economic and cli-

[1] Efforts for developed countries have flourished during recent years, as reviewed by Carleton and Hsiang [2016] and Auffhammer and Schlenker [2014] among others. A related literature focuses on the relationship between rainfall and yields has received slightly more attention. For instance, in a study of the determinants of land invasions, Hidalgo et al. [2010] estimate the economic impact of rainfall on a sample of Brazilian households and find a similar non-linear relationship as the one we observe with temperatures. Similarly, Auffhammer et al. [2012] use state-level data to show that changing rainfall patterns over the last half a century affected India's rice yields negatively.

[2] A related paper is Welch et al. [2010]. They estimate the impact of solar radiation, average maximum and minimum temperatures on rice yields in a panel of 227 irrigated farms in South East Asia.

matic characteristics. The coast is hotter with little rain, relatively richer, and with a more modern, market oriented, agriculture. In contrast, the highlands are cooler, more reliant on precipitation, and with a larger share of subsistence farmers. This makes Peru an especially interesting case study: it allows us to test how different forms of agriculture are affected by, and react to, extreme heat.

We start by documenting a non-linear relation between temperature and agricultural yields in both regions. Similar to previous studies, we find a positive effect of temperature up to certain threshold, above which extreme heat becomes harmful. The magnitude is economically significant. For instance, in the highlands, an increase of 1°C in the average growing season temperature above the optimal level would decrease output per hectare by more than 10 percent. This estimate is comparable to the 8 percent losses documented in Chapter 1 for a sample of Tanzanian farms.

In this study, we go beyond study of impacts on yields, and exploiting rich, farm level data, we shed light on the adjustments taking place at the household level. Using data on input use and endowments, we estimate a production function and document a similar non-linear effect of temperature on total factor productivity (TFP). We then examine farmers' responses to extreme heat. To guide the empirical analysis, we propose a simple model of consumer-producer households in which extreme heat enters as a negative productivity shock. At the core of the model is the observation that some farmers may be living close to subsistence levels and, hence, may be less able to reduce consumption in response to lower productivity. The model provides drastically different predictions for subsistence and non-subsistence farmers. Non-subsistence farmers respond to the drop in productivity by reducing flexible inputs (such as hired labor) and total output. In contrast, subsistence farmers respond by using their endowments (land and domestic labor) more intensively to offset the drop in agricultural output. To the extent that uncultivated land or non-working time provide a future return, for instance by increasing productivity via fallowing or facilitating investment in human capital, this short-run response could have detrimental effects in the long-term.

With this framework in mind, we examine the effect of extreme heat on input use and find evidence consistent with the model predictions. In the case of coastal farmers, we observe a significant decrease in hired workers and total agricultural output. However in the highlands, land use increases while total output remains unchanged, despite also experiencing a drop in productivity. This last result can be explained if farmers are close to a minimum consumption threshold. To further explore this interpretation, we distinguish between subsistence and non-subsistence farmers using indicators of extreme poverty. We find that the increase in land use is driven by extreme poor farmers. Furthermore, we also find that, among this group, extreme heat increases household members' work in agricultural activities, and the likelihood of child labor.

Evidence of this sort has not yet been documented for developing countries. Efforts to study farmer's reactions and adaptation strategies in this context have so far relied heavily on farmers' self-report exposure to weather shocks.[3] However, the reliability of self-reported data has been called into question (Guiteras et al. [2015a]).

Finally, we use our estimates to simulate the potential effect on yields of evenly distributed increments of 1.5℃ to 3℃ in average daily temperatures. These increments are consistent with scenarios RCP2.6 and RCP8.5 of the 4th IPCC Assessment Report. Our simulations suggest important heterogeneous impacts: while yields fall in the coast, they increase in the highlands. This result is consistent with the notion that an increase in average temperatures would have re-distributional effects across regions, with low-lying areas suffering the most and currently colder areas benefiting (see, for example, Deschenes

[3]For example, Hisali et al. [2011] study farmers' self reported exposure to different kinds of climate shocks and their stated strategies of adaptation, and find that temporary measures like reducing consumption, borrowing and relying on savings are common. Similarly Akpalu et al. [2015], Di Falco et al. [2011], Gbetibouo [2009] among others, find that access to credit and awareness of climate change are important determinants of self reported farm-level adaptation.

and Greenstone [2007] for similar results across US states). This exercise complements more sophisticated modelling efforts of global food production and trade by providing regional adaptation margins estimated at the farm level.

The rest of this Chapter is organized as follows. Section 3.2 discusses the main characteristics of Peru's coast and highlands, and develops an analytical framework to guide the empirical analysis. Section 3.3 describes our data and empirical strategy. Section 3.4 presents the main results and additional checks. Section 3.5 presents the simulations of climate change scenarios. Section 3.6 concludes.

3.2 Background

3.2.1 Peru's climatic regions

Peru has three main climatic regions: the coast to the west, the Andean highlands, and the Amazon jungle to the east.[4] In the present study we focus on the coast and the highlands (see Figure 3.2 for a location map).[5] These regions have very different climates, average socio-economic characteristics, and production methods. This makes Peru an especially interesting case study: it allows us to test how different forms of agriculture are affected by, and react to, extreme temperature.

Table 3.1 presents summary statistics for our sample of farming households. There are several relevant observations for the empirical analysis. First, farmers in the highlands are poorer and closer to subsistence levels. Their average consumption is almost half of their coastal counterparts and the poverty rates are significantly higher. For instance, the rate of extreme poverty is 20 percent in the highlands but only 4 percent in the coast.[6] They are also more

[4] This classification is based on altitude and position relative to the Andean mountains. The coast is the region from 0 to 500 meters above sea level (masl) on the west range of the Andes. Highlands range from 500 to almost 7,000 masl, while the jungle is the region of low lands (below 1000 masl) to the east of the Andes.

[5] We drop the jungle due to small sample size and poor quality of satellite data: many observations are missing due to cloud coverage.

[6] A household is considered extreme poor if its expenditure falls below the value of a minimum food basket. In contrast, the poverty line includes other essential goods and services in addition to food (INEI, 2000).

Table 3.1: Summary statistics by climatic region

	Coast (1)	Highlands (2)	p-value (1)=(2) (3)
Household Characteristics			
Daily per capita expenditure	4.040	2.347	0.000
Poor (%)	26.1	55.0	0.000
Extreme poor (%)	4.5	20.9	0.000
Has at least 1 unmet need (%)	29.5	38.2	0.000
Child labor (%)	13.2	39.0	0.000
Household size	4.406	4.312	0.001
HoH completed primary educ. (%)	59.0	49.8	0.000
Agricultural characteristics			
Value of ag. output (in constant 2007 Soles)	3052.7	682.0	0.000
Output per hectare (in constant 2007 Soles/hectare)	2323.8	1077.3	0.000
Land used (ha)	2.381	1.906	0.000
Hire workers (%)	55.8	47.1	0.000
Nr. HH members work in agric.	2.197	2.315	0.000
Fruits (% total output)	31.8	3.5	0.000
Tubers (% total output)	5.6	35.6	0.000
Cereals (% total output)	30.2	31.4	0.010
Use tractor (%)	54.1	16.4	0.000
Use fertilizer (%)	76.8	74.4	0.000
Use pesticides (%)	65.1	41.2	0.000
Uncultivated land (% of land holding)	12.1	44.8	0.000
Irrigated land (% land holding)	82.3	28.7	0.000
Weather in last growing season			
Average temperature (in C)	33.1	21.4	0.000
% days with HDD	81.9	7.6	0.000
Precipitation (mm/day)	0.907	3.514	0.000
Observations	7,969	47,026	

Notes: Sample restricted to farming households in Coast and Highlands. Columns (1) and (2) display mean values. Column (3) shows the p-value of a mean comparison test between both regions. Expenditure and agricultural output are measured in 2007 USD. HoH=household head, HH=household, HDD=harmful degree days likely to have least one basic need unmet, and experience substantially higher rates of child labor.[7]

Second, there are significant differences in agricultural practices and outcomes between both regions. Agriculture in the coast is more modern and market-oriented than in the highlands. Coastal farmers cultivate more cash crops, such as cereals (corn and rice) and fruit trees, obtain higher output (total and per unit of land) and are more likely to use modern, capital intensive, inputs such as tractors, fertilizers, and pesticides. Most of their land also has access to some form of irrigation.[8]

In contrast, highlands farmers have less access to irrigation and are more reliant on subsistence crops, such as tubers (potatoes). They also have smaller plots and use their land less intensively. For instance, the share of land left uncultivated is around 12% in the coast, but more than 40% in the highlands.[9] This last observation is consistent with crop rotation and fallowing, practices common in traditional agriculture.

Finally, climatic conditions are also very different. The coast has a subtropical climate with mild to hot temperatures and very little rainfall, while the highlands have cooler temperatures and more rain during the growing season. These climatic differences become more apparent when observing the distribution of daily temperature in these two regions (see Figure 3.1).

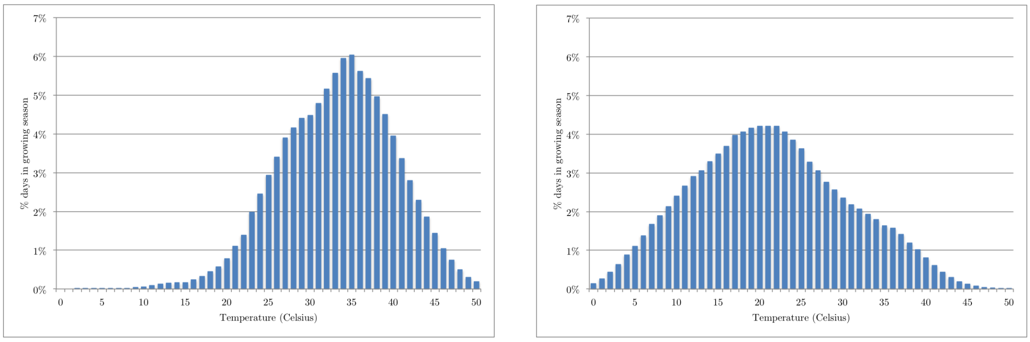
These features suggest that coast and highland production units are significantly different. This might, for example, limit the access that highland farmers have to mitigation technologies or shape the response of farmers to

[7] The measure of poverty by basic unmet needs is widespread in Latin America since the 1980s In the case of Peru, a household is considered poor if it fulfils one of the following conditions: it has dirt floors, there are more than three members living in one room, the household does not have a flush toilet (connected to sewage or septic tank), there is at least one child aged 6 to 12 not attending school, or the household head did not complete primary school.

[8] Given the potential importance of irrigation as a method to counteract the damage from high temperatures, a branch of the literature decides to exclude areas with high irrigation coverage, see for instance Schlenker and Roberts [2009]. We prefer instead to present results splitting the sample between two regions and controlling for the share of irrigated land.

[9] Uncultivated land is land without seasonal nor permanent crops. This includes land covered with grasses, shrubs or woods, or left fallow.

Figure 3.1: Distribution of daily average temperature by climatic region



(a) Coast **(b)** Highlands

Notes: Figures depict share of days in growing season in each temperature bin.

extreme weather events. We argue that the Peruvian coast and highlands serve as case studies for two different models of agriculture usually found in developing countries: one that is more modern, market-oriented, and capital intensive (coast), and another that has a smaller scale, less capital intensive, and directed towards own consumption (highlands).

3.2.2 Analytical framework

We propose a simple model to examine the effect of extreme temperature on farmers' economic decisions, such as input use and total output. At the core of our model is the observation that, in contrast to modern farming, some producers in developing countries may be close to subsistence levels. This feature creates drastically different short-run responses to extreme temperature, which we then examine in the data. Importantly, the model highlights potential channels for weather shocks to have long-term effects on farmer's well-being.

3.2.2.1 Environment

Consider a producer-consumer household that lives for two periods, or agricultural campaigns, $t = 1,2$. The household has utility in a given period t equal to $U_t(c_t - s, l_t)$, where c_t is consumption, l_t is leisure, and s is a minimum consumption threshold, below which utility is very low or undefined.[10] The farm is the only source of income of the household. The agricultural output Y is defined by the production function $f(T_t, L_t^h, L_t^d, A_t)$, where A_t is total factor productivity (TFP), T_t is the amount of land, L_t^h is the amount of labor hired (market labor), and L_t^d is the amount of labor provided by households members (domestic labor). Households are endowed with 1 unit of domestic labor and land. We also normalize the prices of consumption and agricultural goods to 1.

Market labor can be hired in perfectly competitive markets at wage w. The wage rate is defined by the opportunity cost of workers in a non-agricultural activity.[11] In contrast, domestic labor can only be used in the household farm or consumed as leisure. This assumption could reflect, for instance, imperfect labor markets, or input specificity.[12] Note that domestic labor does not need to include only adult workers, but could be interpreted, for instance, as child labor.

We assume that there are no land markets.[13] Land can be used for farm production or left uncultivated. Uncultivated land generates a return r in the next period. This return reflects the increase in future output from letting land fallow or rotating crops. This benefit has been documented in the agronomic and economic literature (Goldstein and Udry [2008]). We assume that this return is equal to $r(1-T_{t-1})$, where T_{t-1} is the amount of land cultivated in the previous period. Other than letting land fallow, the household cannot invest, save, lend, or borrow.

Total factor productivity, A_t, is determined by a host of variables such as quality of soil, water availability, and farming technology. Importantly,

[10] A particular example of this type of preferences is the Stone-Geary utility function $U = \alpha \ln(c-s) + \beta \ln l$.

[11] This assumption reduces the need to worry about endogenous local agricultural wages.

[12] Results are similar if we assume that domestic labor can be sold in local markets but can only be used in agriculture. In that case, weather shocks would still reduce the opportunity cost of leisure although not by decreasing farm's productivity but local agricultural wages.

[13] The assumption of imperfect land markets is not crucial for the model predictions, but simplifies the model analysis and exposition.

A_t is affected by local weather conditions. Farmers know the distribution of weather outcomes (or climate) in their area, but are nonetheless exposed to unanticipated weather shocks during the growing season, *after* they have made her initial choice of inputs. After being exposed to a weather shock, farmers can adjust their input choice, although limited by some inputs being fixed. Estimating this short-run response to unanticipated weather shocks is at the core of our empirical analysis.

3.2.2.2 Short-run input choice

We focus on farmers' short-run decisions in period 1. These decisions are taken after the agricultural campaign has started. At that moment, land has already been planted (used) and farmers may adjust the amount of inputs in response to the weather shock. We assume that labor (both market and domestic) is flexible. However, land is not: farmers can increase its quantity, but not decrease it. This assumption reflects a feature of Peruvian farming in which some crops, like potatoes, have flexible planting schedules and can be planted throughout the growing season.

The household problem is to choose a new level of inputs in response to the weather shock. This problem can be written as:

$$\underset{T_t, L_t^m, L_t^d}{\text{maximize}} \sum_{t=1}^{2} E[U_t(c_t - s, l_t)] \tag{3.1}$$

$$\text{subject to } c_t = f(A_t, T_t, L_t^m, L_t^d) - wL_t^m + r(1 - T_{t-1}), \tag{3.2}$$

$$l_t = 1 - L_t^d, \text{ and } T_1 \geq \tilde{T}_1, \tag{3.3}$$

where $\tilde{T}_1$ is the quantity of land already planted, i.e. the long-run optimal choice before the weather shock.

We start by assuming that the consumption level is above the minimum threshold. In that case, the short run choice of inputs in period 1 is defined by the following conditions:

$$\begin{aligned}
\frac{\partial U_1}{\partial c_1}\frac{\partial f}{\partial L_1^d} - \frac{\partial U_1}{\partial l_1} &= 0 \\
\frac{\partial f}{\partial L_1^m} - w &= 0 \\
\frac{\partial U_1}{\partial c_1}\frac{\partial f}{\partial T_1} - r\frac{\partial EU_2}{\partial c_2} + \lambda &= 0 \\
\lambda(T_1 - \tilde{T}_1) &= 0
\end{aligned}$$

There are some relevant observations. First, the amount of market labor depends of marginal productivity of labor and market wage, as in standard production models. Second, the amount of domestic labor and land depend of the opportunity cost of leisure and unused land. These opportunity costs are increasing in the marginal productivity of the input and the marginal utility of current consumption, in the case of leisure, and future consumption, in the case of land.

This last observation highlights the trade-offs faced by farmers: one between farm profits (consumption) and leisure, and another between present and future consumption. The dynamic link is created by the returns to land fallowing, which makes not using part of the land endowment similar to an investment decision.

3.2.2.3 Response to extreme heat

We analyze the effect of extreme heat as a negative shock to total factor productivity.[14] How do farmers respond to this shock? The response depends of whether consumption is above or at the minimum consumption threshold, s. We call these two types of farmers: non-subsistence and subsistence farmers, respectively.

If the level of consumption is above the minimum threshold, then lower productivity reduces use of hired labor. The effect on domestic labor and land

[14] This assumption is based on previous findings of negative effects of extreme temperature on crop yields, as described more extensively in Chapter 1. In Section 3.4, we test empirically this assumption for the Peruvian case.

is, however, ambiguous. This happens because marginal productivity falls but the marginal utility of current consumption increases. If this latter effect is not too large, then extreme heat would reduce domestic labor and keep land use unchanged.[15] This implies a reduction in agricultural output.[16]

The response is different for subsistence farmers. In that case, the drop in productivity could push consumption below the minimum threshold. This is undesirable and forces the household to direct its resources of land and domestic labor to increase farm output. However, the quantity of hired labor drops, because its marginal productivity is now smaller and it still has to be paid the non-agricultural market wage w.

Table 3.2 summarizes the model empirical predictions. Note that in both cases, subsistence and non-subsistence farmers, crop yields (Y/T) drop, as a result of the productivity shock. For non-subsistence farmers this is followed by a drop in output, while for subsistence farmers the shock triggers instead an increase in land use. The main insight is that, in contrast to non-subsistence farmers, subsistence farmers would respond to extreme temperature by using their endowments (land and domestic labor) more intensively to offset the reduction in agricultural output. To the extent that uncultivated land provides a future return, for instance by increasing productivity via fallowing or crop rotation, this short run response could have detrimental effects in the long-term.[17]

3.3 Methods

3.3.1 Data

We combine household surveys with satellite imagery to construct a comprehensive dataset with information on agricultural, socio-demographic, and weather variables at farm level. The dataset includes around 55,000 households

[15] Farmers would actually like to reduce land but this is, by assumption, not feasible in the short run.

[16] We would also observe a decrease in agricultural output even if domestic labor or land use increase, since this would happen only if consumption decreases.

[17] A similar argument could be made for leisure if it is used to invest in human capital.

Table 3.2: Main empirical predictions

Effect of extreme temperature on	Non-subsistence farmers $c > s$	Subsistence farmers $c = s$
Yields (Y/T)	-	-
Agricultural output (Y)	-	0
Hired labor (L^h)	-	-
Domestic labor (L^d)	?	+
Land (T)	?	+

Notes: Empirical predictions derived from our model of agricultural productions by household type.

located in Peru's coast and highlands, and covers years 2007-2015.[18]

3.3.1.1 Agricultural and socio-demographic data

We use repeated cross sections of the Peruvian Living Standards Survey (ENAHO), an annual household survey collected by the National Statistics Office (INEI). This survey is collected in a continuous, rolling, basis. This guarantees that the sample is evenly distributed over the course of the calendar year. Importantly, the ENAHO includes the geographical coordinates of each primary sampling unit, or survey block.[19] In rural areas, this corresponds to a village or cluster of dwellings. We use this information to link the household data to satellite imagery. Figure 3.2 depicts the location of the observations used in this study.

The ENAHO contains rich information on agricultural activities in the 12 months prior to the interview. We use this information to obtain measures of agricultural output and input use. To measure real agricultural output, we construct a Laspeyres index with quantity produced of each crop and local prices.[20] Land use is obtained by adding the size of parcels dedicated to seasonal and permanent crops. We observe the size and use of each parcel, but not which specific crops are cultivated in each one. Since most farmers cultivate several crops, this limitation of the data prevents us from calculating

[18] Our sample includes households with some agricultural activity. We drop 282 farmers reporting land holdings larger than 100 has.

[19] There are more than 3,400 unique coordinate points.

[20] As weights, we use the median price of each crop in a given department in 2007.

crop-specific yields. Instead, our analysis focuses on output per hectare (Y/T) and total factor productivity.

We use self-reported farm expenditures on external workers to obtain measures of extensive and intensive use of hired labor. Labor information on household members is available for the week prior to the interview. We use this information to obtain the number of hours spent working in agriculture, and an indicator of child labor.[21] We use these variables as proxies for domestic labor.

We complement the household survey with data on soil quality from the Harmonized World Soil Database Fischer et al. [2008]. This dataset provides information on several soil characteristics relevant for crop production on a 9 km square grid.[22]

3.3.1.2 Temperature and precipitation

Similar to other developing countries, Peru has very few and sparse weather monitoring stations. To overcome this data limitation, we use satellite imagery.[23] For temperature, we use the MOD11C1 product provided by NASA.[24] This is same source used in Chapter 1 and provides LST data on a daily basis, with a spatial resolution of 0.05×0.05 degrees.

Precipitation data comes from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) product Funk et al. [2015], also used in Chapter 1. CHIRPS is a re-analysis gridded dataset that combines satellite imagery with monitoring station data. It provides estimates of daily precipitation with a resolution of 0.05×0.05 degrees. As seen in Chapter 1, these

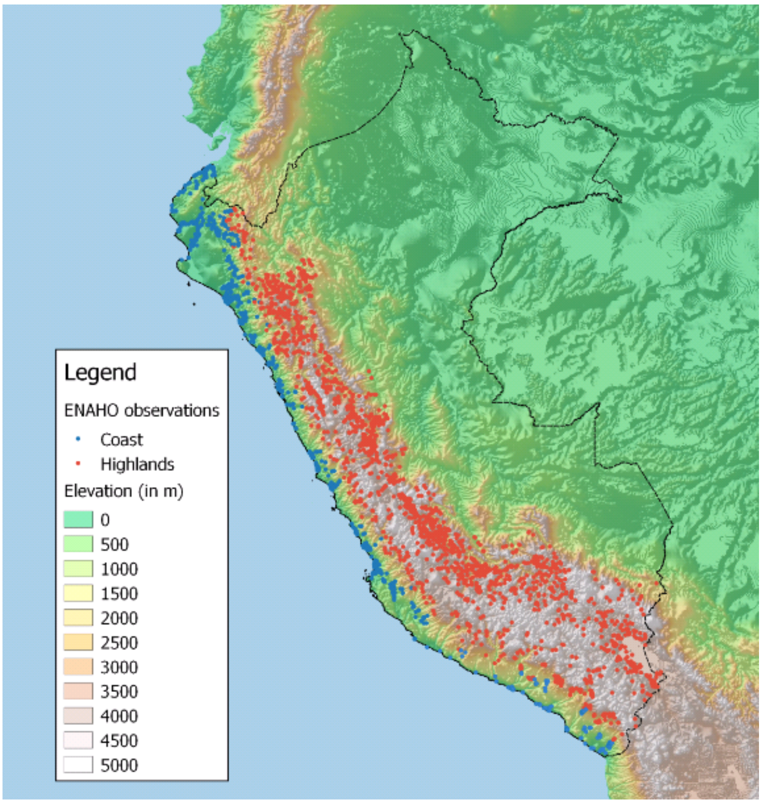
[21] Information on number of hours of work is only available for individuals 14 years and older. Child labor is an indicator equal to one if a child aged 6-14 reports doing any activity to obtain some income. This includes helping in own farm, selling services or goods, or helping relatives, but excludes household chores.

[22] The soil qualities include nutrient availability and retention, rooting conditions, oxygen availability, excess salts, toxicity and workability.

[23] A more extensive discussion on the alternative weather data sources available, as well as their respective advantages and limitations is presented in the Chapter 1.

[24] The satellite estimates are very precise. Validation studies comparing satellite estimates and ground readings find a discrepancy of only 0.1-0.4 Celsius degrees Coll et al. [2005, 2009], Wan and Li [2008].

Figure 3.2: ENAHO observations 2007-2015



two sources can successfully capture weather conditions in areas with sparse monitoring station coverage, and can be used to study weather impacts on agricultural yields.

We combine the weather data with household's location to obtain daily measures of temperature and precipitation for each farmer during the last completed growing season.[25] We fix the growing season to months October through March. This period corresponds to the southern hemisphere's Spring and Summer.

3.3.2 Empirical strategy

The aim of the empirical strategy is to estimate the effect of extreme heat on agricultural outcomes. Guided by the analytical framework laid out in Section

[25]We assign the outcomes for growing season t (October $t=1$ through March t), to any household interviewed as of April t and up to March $t+1$. We believe this approach is conservative since it only assigns weather outcomes to households once the growing seasons has finished.

3.2.2, we examine the effect of temperature on total factor productivity (TFP), and farmers' short run responses to this productivity shock in the form of changes in input use.

3.3.2.1 Identification of shocks to TFP with a Production Function approach

In order to estimate the effect of weather shocks on TFP, we start by assuming a Cobb-Douglas agricultural production function:

$$Y_{idt} = A_{idt} T_{idt}^{\alpha} L_{idt}^{\gamma} e^{\epsilon_{idt}},$$

where Y_{idt} is agricultural output of farmer i in locality d at time t. ϵ_{idt} is a random shock that affects output after input choice is made and thus is, by definition, uncorrelated to input use. T and L are quantities of land and labor (hired and domestic), and A is total factor productivity. [26]

We assume that A_{idt} depends of local weather conditions, (ω_{dt}), in locality d during growing season t, household and farm characteristics (ϕZ_i), time-invariant local economic and environmental conditions (ρ_d), a common trend or productivity (ψ_t) and (potentially) other unobserved heterogeneity (v_{idt}). In particular: $A_{idt} = \exp(g(\beta, \omega_{dt}) + \phi Z_i + \rho_d + \psi_t + v_{idt})$, where $g(\beta, \omega_{dt})$ is a non-linear function to be specified later. The parameter of interest is β which describes the relation between weather and total factor productivity.

Combining these structural assumptions, we obtain the following regression model:

$$\ln Y_{idt} = \alpha \ln T_{idt} + \gamma \ln L_{idt} + g(\beta, \omega_{dt}) + \phi Z_i + \rho_d + \psi_t + \xi_{idt}, \tag{3.4}$$

where the error term $\xi_{idt} = \epsilon_{it} + v_{idt}$. To estimate this model, we use measures of agricultural output and inputs, include a vector of household and farm characteristics as a proxy for Z_i, and replace ρ_d and ψ_t with district and growing

[26] The inclusion of other inputs such as capital and materials (e.g. fertilizers, pesticides) does not change any of the empirical results, as seen in the next section, so we omit them in this exposition.

season fixed effects.[27] The main difference between the model estimated in this case and the one used for Chapter 1 is its cross sectional nature. We cluster the standard errors at district level to account for spatial and serial correlation in the error term.[28]

To the extent that TFP determinants are fully captured by our set of controls variables and fixed effects (i.e.,$v_{idt} = 0$), we can estimate (3.4) using an OLS regression. If that is not the case, the unobserved heterogeneity would create an omitted variables problem and render OLS estimates inconsistent.

To address this concern, we complement our OLS estimates with a 2SLS model using endowments (i.e., household size and area of of land owned) as instruments for input use. The motivation to use these instrument comes from the observation that, in the absence of input markets, the quantity used of land and domestic labor would be proportional to the household endowment.[29] The validity of these instruments would require that endowments affect output only through its effect on input use, i.e., endowments should not be conditionally correlated to unobserved heterogeneity, v_{idt}.[30]

In line with previous studies, we also examine the effect of temperature on other proxies of productivity, such as agricultural yields. We are unable to calculate crop-specific yields (except for a small share of farmers) so instead we use output per hectare. The main advantage of using this outcome is that we do not need to estimate the contribution of inputs to total output which reduces endogeneity concerns. The main disadvantage is that, being a partial measure of productivity, agricultural yields capture the effect of changes in

[27] A district is the smallest administrative jurisdiction in Peru and approximately half the size of the average U.S. county. Our sample includes 1,320 districts out of a total of 1,854.

[28] Results are robust to clustering standard errors at provincial level (see Table 3.6). A third alternative often discussed in the literature is to correct spatial and serial correlation using the procedure suggested by Conley [1999]. However, this approach is not feasible in our case due to conformability errors as described in Hsiang [2016].

[29] With perfect input markets, we would obtain the standard result of separability of consumption and production decisions and there would be no correlation between endowments and input use Benjamin [1992]. Empirically, this would create a problem of weak instruments.

[30] The interpretation of this IV strategy would be as a local average treatment effect, since the coefficients would be identified from farmers subject to input market imperfections.

TFP but also changes in input use.

To estimate the effect of temperature on agricultural yields, we estimate a reduced form version of equation (3.4) without controlling for input use. We use a similar specification to examine the effect of temperature on quantity of inputs such as land, market and domestic labor. The estimated regression is:

$$y_{idt} = g(\beta, \omega_{dt}) + \phi Z_i + \rho_d + \psi_t + \xi_{idt}, \tag{3.5}$$

where y is the outcome of interest and $g(\cdot)$ is a non-linear function of weather conditions. Similar to equation (3.4) , we control for several household and farm characteristics, and include district and growing season fixed effects. However, the identification assumption is less restrictive: we need that, conditional on the included controls, stochastic deviations from average weather characteristics are uncorrelated to the error term.

3.3.2.2 Modelling the relation between weather and productivity

We model the relation between weather and agricultural productivity as a function of cumulative exposure to heat and water. This approach is based on the assumption of time separability, i.e., weather outcomes have the same impact on output per hectare whenever they occur within a given growing season. This assumption implies that what matters its the cumulative exposure to heat and water, not the time when they occur.

Similar to Schlenker and Roberts [2009], and as done in Chapter 1, we construct two measures of cumulative exposure to heat during the growing seasons: degree days (DD) and harmful degree days (HDD).[31] DD measures the cumulative exposure to temperatures between 8°C and an upper threshold τ_{high}, while HDD captures exposure to hotter temperatures (above τ_{high}). The inclusion of HDD allows for potentially different, non-linear, effects of extreme

[31] We complement this approach with a second, more flexible specification to examine the robustness of our results. In particular, we define 1°C intervals as temperature bins containing the share of days in the growing season with daily temperatures within that range. We recursively replace $g(\omega)$ in Equation 3.4 with each of the temperature bins and estimate a separate effect for each one-degree temperature threshold. Results are displayed in the bottom panel of Fig 3.3.

heat.[32]

Note that we calculate degree days for an average day not for the entire season. This is, however, simply a re-scaling and does not affect the results. Similarly, we measure exposure to precipitation using the average daily precipitation (PP) during the growing season and its square. With these definitions in mind, we parametrize the function relating weather to productivity $g(\beta, \omega_{dt})$ as:

$$g(\beta, \omega_{dt}) = \beta_0 DD_{dt} + \beta_1 HDD_{dt} + \beta_2 PP_{dt} + \beta_3 PP_{dt}^2. \tag{3.6}$$

Previous studies in U.S. have set the threshold τ_{high} between 29-32°C, respectively Deschenes and Greenstone [2007], Schlenker and Roberts [2009]. These estimates, however, are likely to be crop and context dependent and hence might not be transferable to our case.[33] For that reason, we prefer to use a data-driven approach and estimate the value of this threshold, which consists of an iterative regression algorithm described in Chapter 1 and used by Schlenker and Roberts [2009] and Tack et al. [2017]. We estimate an upper threshold for the whole country but, given the substantial climatic and agronomic differences between coast and highlands, also for each region.[34] Our estimates suggest using a value of τ_{high} equal to 32°C for the whole sample, 27°C for the coast, and 35°C for the highlands.

3.4 Results

3.4.1 Effect on agricultural productivity

Our first set of results examines the effect of temperature on agricultural productivity. Following the existing literature we focus on output per unit of land (Y/T). However, we also examine the effect on output conditional on input use. We interpret this result as the effect on total factor productivity (TFP).

Table 3.3 presents our main results pooling all the farmers and splitting

[32] For a detailed description of the construction of DD and HDD variables see Chapter 1.

[33] In addition to differences in crop mix and agricultural technology, we use a different measure of temperature (i.e. land surface temperature). These factors make previous estimates not applicable to our case study.

[34] See Appendix B.1.1 for a detailed discussion of the estimation procedure and results.

the sample in two climatic regions: coast and highlands. Figure 3.3 plots the estimated relation between temperature and agricultural yields. The top panel uses the piece-wise linear specification using DD and HDD as measures of exposure to heat, while the bottom panel uses a more flexible specification with the share of days in a growing seasons within each 1°C interval.

We find a clear non-linear relationship between measures of productivity and temperature. Heat has a beneficial effect up to a certain optimal temperature.[35] Above this level, extreme heat has a negative effect on productivity. The magnitude of the effect is economically significant. Our estimates imply. for instance, that an increase of 1°C on the average growing season temperature above the optimal level would decrease agricultural yields by more than 10% in the highlands and almost 20% in the coast.

Table 3.3: Effect of temperature on agricultural productivity

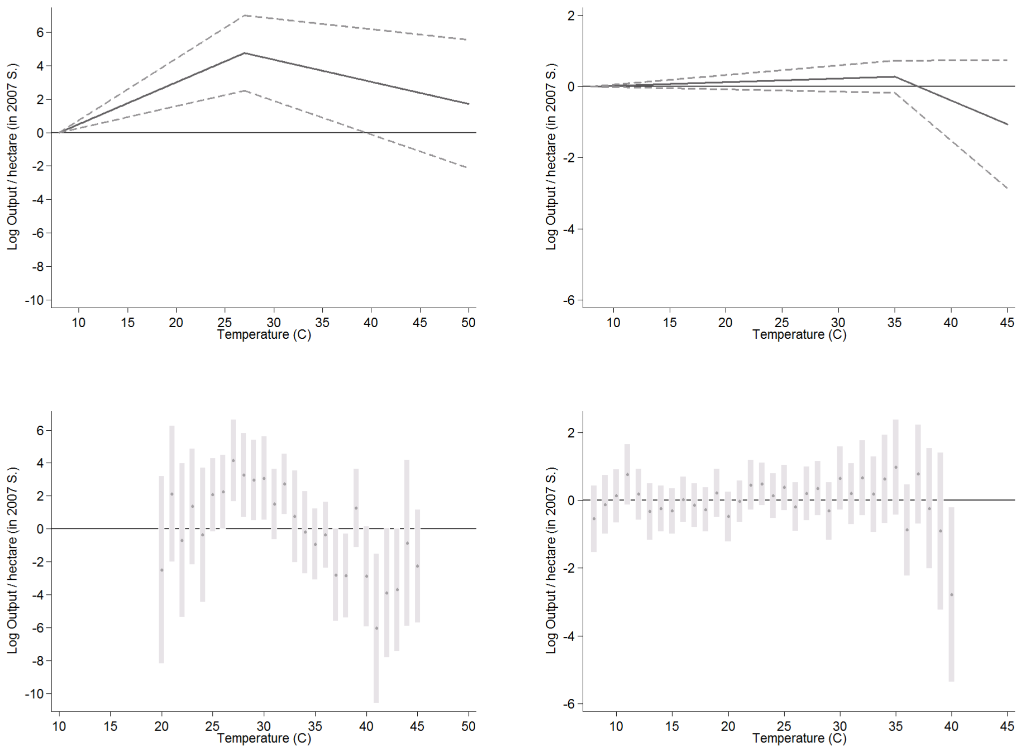
	Coast			Highlands		
	Y/T	TFP	TFP	Y/T	TFP	TFP
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature:						
Average DD	0.149**	0.093*	0.104*	0.011	0.013	0.021**
	(0.059)	(0.054)	(0.056)	(0.010)	(0.008)	(0.009)
Average HDD	-0.195***	-0.157***	-0.179***	-0.118*	-0.103*	-0.127*
	(0.065)	(0.054)	(0.054)	(0.068)	(0.061)	(0.064)
Fixed Effects:						
Input controls	No	Yes	Yes	No	Yes	Yes
Method	OLS	OLS	2SLS	OLS	OLS	2SLS
Observations	7,962	7,961	7,961	47,020	47,019	47,019
R-squared	0.195	0.349	0.334	0.267	0.446	0.409
F-stat. (first stage)			311.3			1789.5

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include as controls: district and growing season fixed effects household head's demographics (age, age^2, gender, and educational attainment), indicators of soil characteristics, average precipitation, and its square. Columns 1 and 4 use log of output per hectare as outcome, while the rest of columns use log of total output. Columns 2-3 and 5-6 include log of inputs as additional controls. Inputs include land used, household members' hours worked in agriculture , and farms' expenditure on hired workers. Columns 3 and 6 use endowments (land owned and household size) as instruments for quantities of land and domestic labor.

[35]As previously discussed, we estimate this point to be 27°C in the coast, and 35°C in the highlands.

Figure 3.3: Non-linear relationship between temperature and agricultural yields

(a) Coast **(b)** Highlands



Notes: Top panels depict piece-wise function for temperature, including degree-days and harmful degree days at each region's respective thresholds. Bottom panel shows coefficients of individual regressions of model 3.5 using 1°C bins that measure the fraction of the growing season spent in that temperature. Bars indicate 95% confidence intervals.

3.4.2 Farmers' short-run responses

Our second set of results uses the richness of our micro data to examine the responses of farmers to the productivity shock of extreme heat. This is important not only to understand sort-term adaptation strategies, but also to better assess the effect of extreme heat on total agricultural output. This effect is not necessarily the same as the change in productivity, since it could be amplified or attenuated by farmer's change in input use.

Using the analytical framework in Section 3.2.2 as a guide, we examine the effect of temperature on total output, land use, and hired labor. Table 3.4 displays the results for the coast and highlands. Our findings highlight substantial differences in the response between these two regions.

In the coast, extreme heat (HDD) reduces total output and use of hired labor, in the extensive margin. An increase of 1°C in the average HDD reduce output by 17% and the likelihood of hiring workers by almost 3 percentage points. This findings are consistent with the effects of a drop in productivity for a non-subsistence farmer.[36]

A different picture emerges in the highlands, where farmers are poorer and a larger proportion practices subsistence farming. We observe a similar reduction in the intensive use of hired labor as in the coast. However, there is an increase in land use and no significant effect on total output, despite the documented drop in productivity. This last result highlights the importance of taking into account human responses when estimating the effect of weather shocks. In this case, the increase in land use seems to offset the loss of productivity.

Where do this extra land come from? As shown in Table 3.1, around 40% of land in the highlands is not used for seasonal or permanent crops, but left uncultivated: covered with grasses, shrubs and woodlands, or fallow.[37] Leaving land idle is not necessarily sub-optimal nor indicative of low quality. Instead, it is a common practice in traditional agriculture as a way to improve future land productivity. In that sense, it is more akin to an investment decision (Goldstein and Udry [2008]). From this perspective, our results suggest a connexion between short-run responses and long term impacts: the more intensive use of land today might have negative effects on future land returns.

3.4.3 Subsistence vs non-subsistence farmers

We interpret our previous findings as evidence that subsistence farmers increase input use to offset the loss of productivity and maintain a minimum level of consumption. In this section we examine this interpretation in more detail.

To do so, we would need to observe how close a household was to a min-

[36]Interestingly, changes in DD, which are associated with increases in productivity have an effect of similar magnitude but positive.

[37]Preliminary results suggest that most of the additional land comes from reduction in areas covered with shrubs and woodlands. Note that this type of land is not necessarily barren land, but could be land in later stages of a crop rotation process.

Table 3.4: Effect of temperature on total output and input use

	ln(output)	ln(land used)	Hired workers=1	ln(expend. in hired workers)
	Y	T	L^h	L^h
	(1)	(2)	(3)	(4)
Coast:				
Average DD	0.146**	-0.004	0.029***	0.094
	(0.062)	(0.013)	(0.010)	(0.085)
Average HDD	-0.174**	0.021	-0.028***	0.002
	(0.071)	(0.013)	(0.009)	(0.052)
Observations	7,962	7,962	7,962	4,436
R-squared	0.189	0.222	0.185	0.277
Highlands:				
Average DD	0.005	-0.006	-0.000	0.028***
	(0.010)	(0.007)	(0.003)	(0.010)
Average HDD	-0.046	0.073**	0.018	-0.135*
	(0.070)	(0.036)	(0.021)	(0.080)
Observations	47,020	47,020	47,020	22,126
R-squared	0.266	0.325	0.155	0.290

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include district and growing season fixed effects, average precipitation and its square, and same household and soil controls as column 1 in Table 3.3

imum consumption threshold before the negative weather shock. We are unable to do this due to data limitations. Instead, we use an indicator of extreme poverty to distinguish between subsistence and non-subsistence farmers.[38] Extreme poverty is highly correlated with climatic regions: incidence of extreme poverty is 4% in the coast and more than 20% in the highlands. A potential problem with using extreme poverty is that it may be an outcome of weather shocks. For that reason, we complement this approach using an indicator of having at least one unmet need, a more stable proxy for chronic poverty. Guided by the analytical framework (see Table 3.2), we split the sample of highland farmers between subsistence and non-subsistence farmers and examine the effect of temperature on land use (T) and measures of domestic labor

[38] The extreme poverty line measures the value of a minimum food consumption.

(L^d), such as number of hours worked in agricultural activities and child labor.

Table 3.5 displays the results. Consistent with our interpretation, we find that the increase in land use is driven mostly by subsistence, extreme poor, farmers. Moreover, we find suggestive evidence of a more intensive use of domestic labor among this group. For instance, extreme heat seems to increase the number of hours household members work in agriculture, as well as the likelihood of child labor.[39] In contrast, there is no significant effect on non-subsistence farmers and the magnitude of the point estimates is smaller.

Table 3.5: Effect of temperature on land and domestic labor, by subsistence level

	Using extreme poverty			Using basic unmet needs		
	ln(land used)	ln(hours agric. work)	Child labor=1	ln(land used)	ln(hours agric. work)	Child labor=1
	T	L^d	L^d	T	L^d	L^d
	(1)	(2)	(3)	(4)	(5)	(6)
Subs. farmers						
Average DD	-0.010	-0.004	-0.020*	-0.008	-0.016*	-0.020**
	(0.009)	(0.009)	(0.012)	(0.008)	(0.008)	(0.009)
Average HDD	0.150*	0.120	0.197*	0.097*	0.067	0.063
	(0.080)	(0.077)	(0.101)	(0.050)	(0.052)	(0.056)
Observations	9,783	9,783	9,783	17,946	17,946	17,946
R-squared	0.382	0.251	0.235	0.377	0.225	0.208
Non-subs. farmers						
Average DD	-0.006	-0.024***	-0.014***	-0.009	-0.027***	-0.018***
	(0.007)	(0.007)	(0.005)	(0.006)	(0.007)	(0.006)
Average HDD	0.056	0.011	0.016	0.032	-0.002	0.043
	(0.036)	(0.042)	(0.029)	(0.039)	(0.051)	(0.036)
Observations	37,173	37,173	37,173	29,023	29,023	29,023
R-squared	0.331	0.220	0.184	0.326	0.240	0.209

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include district and growing season fixed effects, average precipitation and its square, and same household and soil controls as column 1 in Table 3.3. Columns 1-3 use an indicator of extreme poverty to identify subsistence farmers, while columns 4-6 use an indicator of having at least 1 unmet basic need. Sample includes only farmers located in the highlands Panel A and B split he sample between subsistence and on-subsistence farmers. We classify farmers in these two categories using an indicator of extreme poverty (columns 1-3) or an indicator of basic unmet needs (columns 4-6).

[39]Note that the reference period for these labor outcomes is the last week while exposure to weather correspond to the last completed growing seasons. This may introduce measurement error and an attenuation bias.

3.4.4 Ancillary results

3.4.4.1 Robustness checks

Table 3.6 presents several robustness checks on the effect of temperature on agricultural productivity. We focus on output per hectare, but results are similar when estimating total factor productivity. First, we examine alternative specifications. Column (1) adds to the baseline regression covariates of household endowments and farm controls (such as size of area owned, household size, and share of irrigated land), while Column (2) includes department-by-growing season fixed effects. This is a very demanding specification that flexibly accounts for local trends in agricultural productivity. Column 3 estimates the baseline model using a common HDD threshold of 32°C for both regions, while Column (4) clusters standard errors at province level (n=159) instead than at district level. In all cases, the results remain similar and suggest strong negative effects of extreme heat.

Second, we examine possible biases due to our definition of exposure to temperature. Recall that our baseline results consider exposure to temperature in the last *completed* growing season (October-March). This means, for example, that for households interviewed in March 2010, we are assigning weather variables for the period October 2008-March 2009. However, for households interviewed a month later (April 2010) we would assign weather from period October-2009-March 2010. If agricultural output is affected by the most recent weather outcomes, then by assigning households the weather of the last complete growing season we would introduce measurement error.

To examine the relevance of this issue, we first drop households interviewed during the growing season (column 5). This is the group for which this measurement error is most relevant. The magnitude of our estimates remain comparable to the baseline result, although less precisely estimated. Note that we are dropping almost half of the sample. Finally, we change the definition of exposure to weather and use the 6 months prior to the interview, not the growing season. The estimates become much smaller and statistically insignificant. These results suggest that the last completed growing season is indeed the relevant time-frame for our analysis.

Table 3.6: Robustness checks

	ln(output per hectare)					
Robustness check	Additional controls (1)	Dep-GS FE (2)	32°C threshold (3)	Clustering prov. level (4)	Drop Oct-Mar (5)	Last 6 months (6)
Coast						
Average DD	0.161***	0.110*	0.082*	0.149**	0.112	0.080
	(0.061)	(0.064)	(0.045)	(0.060)	(0.132)	(0.079)
Average HDD	-0.197***	-0.178***	-0.172***	-0.195***	-0.146	-0.021
	(0.062)	(0.058)	(0.049)	(0.069)	(0.093)	(0.031)
Observations	7,961	7,962	7,962	7,962	3,928	7,962
R-squared	0.204	0.213	0.195	0.195	0.231	0.191
Highlands						
Average DD	0.008	0.009	0.011	0.011	-0.006	-0.006
	(0.010)	(0.011)	(0.009)	(0.012)	(0.014)	(0.007)
Average HDD	-0.125*	-0.100	-0.067*	-0.118	-0.124	0.025
	(0.068)	(0.085)	(0.040)	(0.094)	(0.092)	(0.018)
Observations	47,020	47,020	47,020	47,020	23,631	47,020
R-squared	0.280	0.280	0.267	0.267	0.290	0.267

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level, except column 4. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include district and growing season fixed effects, average precipitation and its square, and same household and soil controls as column 1 in Table 3.3. Column 1 adds log of household size and land owned, and share of irrigated land. Column 2 includes department-by-growing season fixed effects. Column 3 uses a common threshold of 32°C to calculate DD and HDD for both regions. Column 4 clusters standard errors by province. Column 5 drops households interviewed during the growing season. Column 6 calculates DD and HDD using temperature in the 6 months prior to the interview date.

3.4.4.2 Changes in crop mix

Our analysis so far has focused on changes in input used as the main response of farmers to extreme heat However, as discussed in Burke and Emerick [2016], farmers may also respond to changes in temperature by changing the varieties or crops they cultivate. We examine this possible effect in this section.

We focus on the two most important types of crops in our sample: cereals (mostly maize) and tubers (mostly potatoes). The botanical literature shows that both crops suffer when exposed to extreme temperatures Hatfield and Prueger [2015]. For each crop type, we estimate equation (3.5) using as outcomes their share of total output as well as the log of total quantity produced. Crop mixes can change for two different reasons. Unpredicted temperature shocks might hurt a specific crop more than others, thus reducing this crop's yield at the end of the season to a larger degree than the rest. On the other hand, farmers are able to adjust partially their initial selection of crops planted and the hectares dedicated to each over the course of the growing season, and our impact estimates might be capturing farmer reactions, together with weather impacts on yields. Crops like potatoes have a more flexible growing season than others, such as corn, leaving farmers the option to increase the amount of potatoes planted during the course of the growing season to offset lower yields of other crops. The Peruvian agricultural calendar, published by the Peruvian Ministry of Agriculture, shows that 70% of the country's production of starchy corn is planted between the months of September and December. During that same period, only 40% of the total amounts of potatoes are planted. Potato planting is relatively stable throughout the year, with no month seeing less than 5% of the yearly production being planted. So the decision of how much of each crop to plant has a degree of endogeneity with weather outcomes over the growing season. This makes the interpretation of our estimates challenging, but might reveal interesting clues about farmer behaviour.

Table 3.7 shows the results of these regressions by region. We can see that cereal production increases with degree-days in both the coast and the highlands, and decreases significantly with harmful degree-days in terms of quantity. In the highlands, however, the exposure to extreme temperatures is associated with an increase in potato production (as a share of total production and in quantities). This suggests that in the highlands, along with the increase in marginal lands, farmers may be adjusting by increasing the production of potatoes. This would represent an important channel of damage mitigation available to highland farmers. However, their exposure to sustained changes in the climate is manifested in their increased dependence on what is already their main crop.[40]

Table 3.7: Effect of temperature on crop mix

	Cereals		Tubers	
	% output	quantity	% output	quantity
	(1)	(2)	(3)	(4)
Coast				
Average DD	0.040***	0.250***	-0.012**	-0.082
	(0.011)	(0.084)	(0.005)	(0.099)
Average HDD	-0.019*	-0.150**	0.008*	0.127**
	(0.011)	(0.067)	(0.004)	(0.054)
Observations	7,564	4,220	7,564	1,981
R-squared	0.347	0.436	0.166	0.460
Highlands				
Average DD	0.012***	0.043***	-0.036***	-0.096***
	(0.002)	(0.011)	(0.004)	(0.019)
Average HDD	-0.013	-0.106*	0.021*	0.020
	(0.014)	(0.061)	(0.012)	(0.069)
Observations	46,651	39,031	46,651	38,150
R-squared	0.396	0.342	0.475	0.377

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include district and growing season fixed effects, average precipitation and its square, and same household and soil controls as Column (1) in Table 3.3. The outcome variable in Columns (1) and (3) is the share of total agricultural value produced by each type of crop. Columns (3) and (4) use instead the log of quantity (measured in kg).

3.5 Simulations of climate change scenarios

In this section, we use our estimates from the previous section to simulate damages to yields that can be expected with higher temperatures predicted in

[40] An alternative explanation to the hypothesis described above would be that potatoes simply have a different optimal temperature threshold, and that the temperatures we record as part of HDDs are actually still beneficial for potato development and growth. To examine this explanation, we calculate crop-specific yields using a sub-sample of farmers that obtain at least 90% of their total production value from a single crop. Note that only 23% of the farmers in our sample enter this category. We then run a similar exercise as the one performed for the whole sample looking for the optimal temperature thresholds We find, however, no evidence that the potatoes' threshold is higher than the one for maize or rice.

climate change scenarios. Even though we only estimate short-term impacts, we consider that this exercise is still informative of potential long term effects. This is based in previous findings from U.S. suggesting a limited ability of farmers to adapt to climate change Auffhammer and Schlenker [2014], Burke and Emerick [2016].

We consider two possible scenarios. First, we assume a "business as usual" scenario where global GDP continues to grow at present levels and fossil fuel dependence is not curbed. This corresponds to the A1B scenario of the Special Report on Emission Scenarios, and the RCP8.5 representative concentration pathway (RCP) used in IPCC [2014]. Under this set of assumptions, Peru is expected to experience an increase of 3℃ to 3.5℃ with respect to its average temperatures during the 1990-2000 period (Gosling et al. [2011]). We assume a country-wide increase of 3℃.

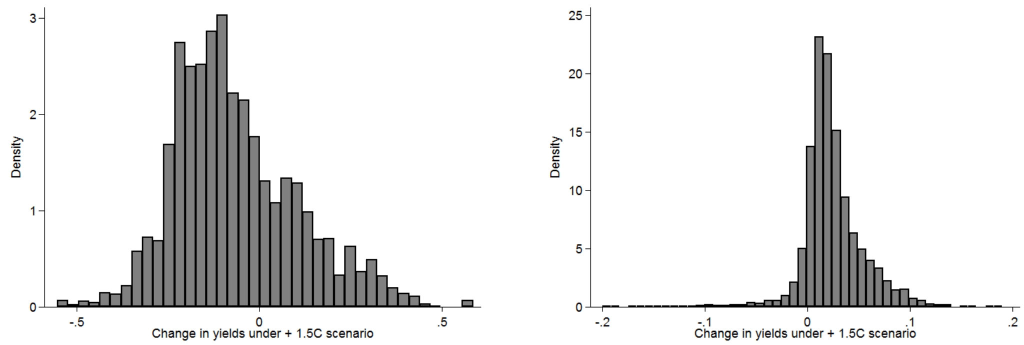
Our second scenario is a more stringent one, which assumes similar GDP and population convergence, but combined with faster adoption of green technologies and energy efficient production methods, resulting in a steep reduction of green house gases. This is in line with the RCP2.6 scenario used in IPCC [2014], in which Peru is expected to experience an increase of 1.5℃. Finally, average precipitation is predicted to stay within one standard deviation of its natural internal variability in both scenarios (IPCC [2014]), so we do not assume any change in this respect.

For each coordinate in our sample, we calculate DDs using the average daily temperature observed in period 2005-2015 and under each scenario k. We call these measures $DD_i^{2005-2015}$ and DD_i^k. We calculate HDDs using a similar procedure. Then, for each observation in our sample i, we estimate damages Δy_i using the coefficients from Columns (1) and (4) in Table 3.3, and the predicted change in DDs and HDDs under each scenario. The specification used is the following:

$$\Delta y_i = \hat{\beta}_1 \Delta DD_i + \hat{\beta}_2 \Delta HDD_i$$

Figure 3.4: Distribution of impacts on yields under 1.5℃ scenario

(a) Coast **(b)** Highlands



Notes: Distribution of impacts from an assumed increase of 1.5°C consistent with the more optimistic, B1 scenario of the IPCC 4th Assessment Report.

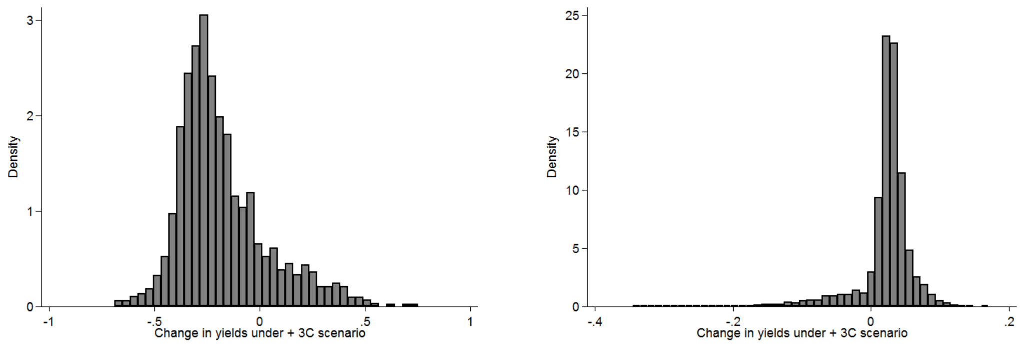
where, $\Delta DD_i = DD_i^{scenario} - DD_i^{2005-2015}$, and similarly for the remaining variables.

Figures 3.5 and 3.4 present the distribution of damages expected for each of the locations in our sample in the two climate change scenarios. We observe that the coast, initially warmer, experiences significant drops in yields, with losses in the range of 0 to 50%. In contrast, highland farmers are predicted to experience small increases in yield, of around 10%. This diverging results comes from the different initial distribution of temperatures in each region, and is consistent with other studies finding stronger negative impacts in low-lying areas, such as Auffhammer and Schlenker [2014].

A few caveats should be mentioned regarding the above simulations. First, climate change forecasts predict a mean average temperature for each region but do not explicitly model the change in temperature variability that is expected. Therefore while we chose to model each scenario as an even increase of 1.5℃ and 3℃ to the average daily temperature recorded during our period of interest, one could think of many other mean-preserving spreads that would still fit these mean predicted temperatures over the year while altering the distribution of temperatures This could have drastically different effects on our impact estimates. For example, in our "business as usual" scenario

Figure 3.5: Distribution of impacts on yields under 3°C scenario

(a) Coast **(b)** Highlands



Notes: Distribution of impacts from an assumed increase of 3°C consistent with a "business as usual" A1B-type scenario of the IPCC 4th Assessment Report.

we could increase all daily temperatures above the median by 6°C and leave the rest unchanged, resulting as well in an average daily temperature increase of 3°C. However this second option will likely result in stronger negative impacts since we would be skewing the distribution of daily temperatures towards more HDDs. While we opted for the most straightforward application of climate change forecasts, there is reason to believe that variance in temperatures will also increase over time, suggesting that our predictions could serve as a lower bound for actual impacts.

Secondly, our modelling does not account for a multitude of factors that might affect agricultural yields in the area and has therefore to be taken as the forecasted effects of temperature and precipitation change only. An important omitted factor here is the increased concentration of CO_2 in the atmosphere and its interaction with changing weather conditions. While lab experiments suggest that higher levels of CO_2 could help plant growth, there is significant uncertainty regarding its interaction with other weather variables and its impact on global agricultural yields remains hard to predict (Gosling et al. [2011]). We also do not consider any impacts from increased flooding and reduced water access due to glacial melting, nor potential changes of relative food prices. Due to these limitations, the predictions presented in this section should be interpreted with caution.

3.6 Conclusion

This Chapter documents farmers' short run responses to extreme weather in the context of a developing country. By combining detailed micro-level agricultural household surveys with daily satellite temperature and precipitation records, we create a comprehensive dataset spanning from 2007 to 2015. We first document a strong and robust non-linear relationship between daily temperatures and agricultural productivity, which are increasing up until a threshold of 27°C in the coast and 35°C in the highlands. Daily temperatures above this threshold become sharply detrimental for agricultural yields. This finding is robust to multiple estimation approaches and is consistent with the literature for developed countries.

We then examine farmers' short-run reactions to this productivity shock and find drastically different responses across regions. In the coast, farmers respond by reducing hired labor and total output. In contrast, in the highlands, farmers increase land use and domestic labor , especially among extremely poor farmers. These more intensive use of inputs offsets the loss of productivity and keeps output unchanged. This a completely new set of findings for developing countries and is specially relevant given current forecasts of extreme weather patters caused by anthropogenic climate change.

Finally, using standard assumptions in the literature we model the distribution of impacts on yields resulting from two possible global warming scenarios and find that while coastal farmers seem to be exposed to yield losses of up to 50% by the year 2080, highland farms are predicted to see moderate increases of up to 10%. Our findings support the hypothesis that low-lying areas will be the most severely affected by climate change over the course of this century.

Chapter 4

Information, Social Status and Health Investments: Evidence from an RCT in Nigeria

4.1 Introduction

Low levels of investment on health-advancing durables at the household level often result in poor health and economic outcomes in developing countries. Examples of lumpy investments are, among others, improved cooking stoves or malarial bed nets. This underinvestment may be related to liquidity or credit constraints, as found for example in the case of malarial bed nets (Cohen and Dupas [2010]). An alternative reason may be a lack of information among households about the benefits of adopting a costly new health technology. Existing evidence suggests that providing health related information to households has a positive effect on household investments and health behaviour (Dupas [2011]).[1] Investment decisions might also be affected by intra-household bargaining externalities, as shown by Miller and Mobarak [2013].[2]

[1] Information has been proven to play a role in other, non-health, investment decisions as well. For example, Jensen [2010] showed that by only providing information on market returns to education, average schooling increased by 0.20-0.35 additional years among a sample of students from the Dominican Republic.

[2] A number of studies show that intra-household bargaining power can be an important driver to realize investments preferred by women, especially those affecting child health and education (Hoddinott and Haddad [1995], Quisumbing et al. [2003], Quisumbing and

Understanding the binding constraints that hamper such health investments is key for designing interventions that help to achieve efficient levels of household investment and provide long-run improvements in health and longevity.

In the absence of water and sewerage networks, safe sanitation is one such investment. The United Nations missed its 2015 Millennium Development Goal target of halving the number of people without access to basic sanitation by almost 700 million people. Today, close to 2.4 billion people still lack improved sanitation facilities and 1 billion still defecate in the open (Unicef and WHO [2015]). While the costs of these practices in terms of child health and human capital accumulation are well understood (Prüss-Ustün et al. [2014]), improvement in sanitation coverage is still slow. In this Chapter we analyse a randomly assigned information campaign called Community Led Total Sanitation (CLTS) in Nigeria. Nigeria faces enormous challenges in the field of sanitation, with 34% of its population practising open defecation and slightly falling toilet ownership rates over the last decade (Unicef and WHO [2015]). The study was implemented in the states of Enugu and Ekiti, and accompanied by three surveys of a random sample of 4,671 households from 246 clusters, distributed evenly across the two states (covering around 9% of the population in the area). The CLTS intervention provided no subsidies or credit. It was designed to promote private toilet construction and reduce open defecation levels in rural Bangladesh, and has been adapted to the Nigerian context. The key event, a village meeting, explained visually and graphically the potential water contamination risks associated with open defecation.

Our study makes three main contributions to the existing literature. First, we show that the information campaign increased toilet ownership by 3 pp, from a baseline level of 36%.[3] These impacts are concentrated on the short term, six to twelve months after the intervention. Similar to Guiteras et al. [2015b] we conduct our analysis using a random sample of all households in

Maluccio [2003], Udry et al. [1995]).

[3]Ownership of functioning toilets of any kind. Treatment effects on other outcomes will be discussed further in Section 4.5.

the area of study. Other studies have instead focused only a selected sample of households with at least one child at the time of interview. Our estimates, while small, are to the best of our knowledge, the first to show positive and statistically significant CLTS impacts among a representative sample of households.[4]

Recent evidence from other randomised experiments provides mixed evidence on the effectiveness of information provision as a way to increase safe sanitation adoption. In Ethiopia, an information-only campaign was found to increase ownership of toilets with stable flooring by 9 percentage points (Crocker et al. [2016]) from baseline levels below 23%. Other authors find larger impacts from similar campaigns in Mali (Pickering et al. [2015]) and India (Clasen et al. [2014]). Using evidence from four cluster randomised field experiments designed to reduce the prevalence of open defecation in India, Indonesia, Mali and Tanzania, Gertler et al. [2015] show that information campaigns were effective at promoting household investment and behavioural change. On the other hand, Guiteras et al. [2015b] find improvements in toilet ownership in Bangladesh driven by subsidy provision but no effects from an information campaign nor from supply-side incentives. Our estimates show that information campaigns can indeed increase toilet ownership, but are low compared to most other studies.

This seemingly contradictory evidence suggests that it is important to identify more precisely the mechanisms that enable or constrain the effectiveness of sanitation programs and other health investment initiatives. As our second contribution, we exploit rich household level data to investigate the different channels through which an information campaign such as CLTS could operate. We analyse whether CLTS increased expected benefits from sanitation, increased social capital, or spurred institutional sanctions. We find that the intervention was successful at increasing expected emotional benefits from sanitation, relating to pride and social status. It did not, however, change the

[4] Crocker et al. [2016] also includes all types of households, but is not a randomised experiment, and is therefore not directly comparable.

household's perceptions on other private benefits, such as health or privacy, nor increased awareness of the externalities deriving from toilet ownership and usage. At the same time, program impacts appear to be stronger for households that perceived toilets not to be too expensive to build: increasing the perceived benefits of toilet construction was more effective among households with low initial perceived costs. We find no evidence that the impact of CLTS on toilet construction and open defecation reduction is driven by changes in social capital nor institutional sanctions.

Finally, we investigate the effect of the information campaign on households with low initial access to sanitation: lower education or asset poor households. Because inadequate sanitation might affect women and children disproportionately, we also consider female headed households and households with children. In contrast to Gertler et al. [2015], we find no evidence of larger program impacts among households with children compared to the rest of the sample. On the other hand, we find that CLTS program impacts are concentrated among female-headed households, and households with lower levels of education and asset wealth. Treatment effects are in the neighbourhood of 5-6 pp among these groups, which also have lower levels of toilet ownership at baseline: between 33% and 21%.

These findings have nuanced implications for sanitation policy. First of all, programme impacts suggest that an information-only campaign may help reduce the sanitation gap. In countries such as Nigeria, where toilet coverage in our study areas is below 50%, it may however not suffice on its own to close it. Second, we find that CLTS changed status concerns around sanitation, but fell short in its effort to increase households' understanding of the private benefits and externalities associated with toilet ownership. Finally, by identifying population sub-groups where CLTS was more effective, we provide input for a more precise targeting of the policy, that avoids wasteful treatment of non-responsive populations.

This Chapter is structured as follows. In the next two sections, we describe the intervention and the experimental design, respectively. Section 4 discusses the empirical methods and section 5 shows the main set of results. Section 6 explores alternative channels of behavioral change. Section 7 concludes.

4.2 Community Led Total Sanitation

4.2.1 Background

The concept of Community Led Total Sanitation (CLTS) was first developed by Kamal Kar and the Village Education Resource Centre (VERC) in Bangladesh, in 2000.[5] While carrying out an impact assessment of WaterAid's decade-old water and sanitation strategy in Bangladesh, they noticed that the existing strategies, heavily reliant on subsidies for toilet construction, fell short of their objectives. Though toilet uptake had increased, new construction was mostly concentrated among middle and high income households. Additionally, open defecation remained common practice, even among households with toilets (Kar [2003]). To tackle this problem, they developed the new 'no subsidy community empowerment approach' (Kar [2003]). It focused on asking every member of the community to first consider the sanitation situation in the village, and then agree on a collective action plan to change it. Since its first trials in Bangladesh in 2000, CLTS has been rolled out to several Asian and African countries. It has been the sanitation approach of choice for the Nigerian Government's Strategy for Scaling Up Sanitation and Hygiene since 2007.

WaterAid conducted piloting activities along with UNICEF and local government authorities (LGAs) before scale-up of CLTS in Nigeria commenced in 2008. Within the states in our study, Wateraid Nigeria has been implementing CLTS in selected, mostly rural communities since 2012, and has tailored the intervention to the local context. In the next section, we describe the intervention that we study as it was implemented in Nigeria since late 2014.

[5] See http://www.communityledtotalsanitation.org/page/clts-approach for more details.

4.2.2 The intervention

The first stage of the intervention consisted of an advocacy and sensitisation visit in which a team of facilitators met with community leaders, village chiefs or other important local decision makers.[6] In this meeting, the potential benefits of CLTS in achieving sustainable behavioural change and the health implications of open defecation were presented. Facilitators and civic leaders then arranged an appropriate date and time for the triggering meeting, which involves the whole community. Local leaders then spread the word around the community, and persuaded as many members as possible to attend.

The triggering meeting is the main component of the intervention. Facilitators engage attendees in a series of activities to inform and involve as many members of the community as possible. First, they carry out a mapping exercise of the village. Each attendee marks their household location and regular open defecation site, if any, on a stylised village map. Second, facilitators trace the community's contamination paths of human faeces into water supplies and food in a crude fashion. Facilitators are given some flexibility how to best emphasize this point, also depending on timing. They could choose from a list of possible activities to illustrate the contamination effect. For example, they carried out calculations of what each household's contribution of faeces to the village environment was. In some cases, they even relied on examples using fresh stool to contaminate a bottle of sparkling water, to make the point as graphic as possible.

As a closing task, attendees were asked to draw up a community action plan, based on the contributions of as many members as possible. The plan's objective was for the village to achieve open-defecation-free (ODF) status. It was written down by a volunteer, assisted by facilitators and village leaders.

[6] Wateraid Nigeria worked with two partner NGOs in the implementation of CLTS. Facilitating teams consisted at least two members of a partner NGO, and four government officials from the LGA's water, sanitation and hygiene (WASH) unit. Each LGA has its own WASH unit that receives support, financial or otherwise, from WaterAid Nigeria, and is responsible for the CLTS implementation and follow up. Facilitators were trained by WaterAid Nigeria staff in conducting CLTS triggering meetings, and participated in the triggering of several villages in their assigned LGA.

The plan was then posted in a public spot. Volunteers were chosen to follow up regularly on the commitments each attendee made towards implementing the plan. After the triggering meeting, WASH unit officials regularly visited the villages to follow up on their advances. If a village reached ODF status, it obtained certification by the LGA's WASH unit, the national Rural Water Supply and Sanitation Agency (RWASSA) and the National Task Group on Sanitation (NTGS).

4.2.3 Components of CLTS

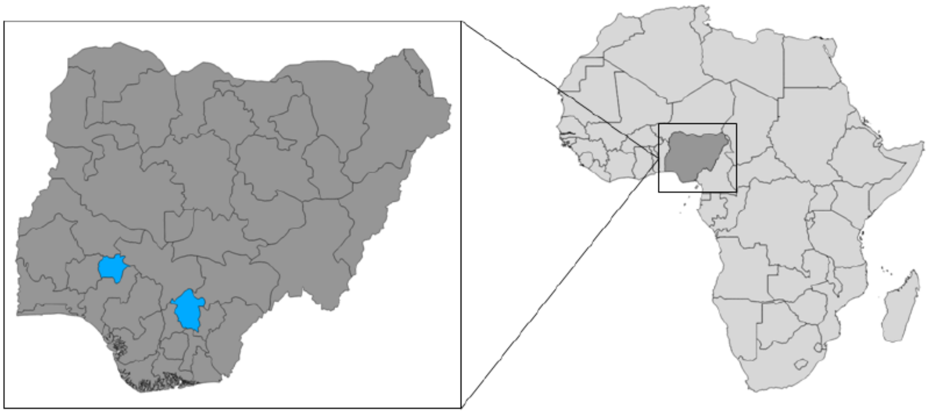
CLTS provides neither subsidy nor credit to finance toilet construction. Instead, it was designed as a tool to i) promote private toilet construction through information on the benefits of sanitation, and ii) to stop open defecation by conveying information about its health implications. Special attention is placed, for example, on correcting the widespread misconception that pit latrines are infectious, particularly to women. In this sense CLTS acts as a standard information delivery intervention. As part of its information content, CLTS emphasises that as long as a small number of people in the community continues to defecate in the open, all community members are at risk of contracting sanitation related diseases. Thus, it is delivering information regarding the importance of sanitation externalities for individual health.

CLTS also promotes a collective sense of disgust and shame around the practice of open defecation, and of pride attached to toilet ownership. CLTS activates these feelings to transform social norms and to change sanitation standards in the community. This second aspect of CLTS goes beyond simple information delivery and seeks to leverage the power of social interactions.

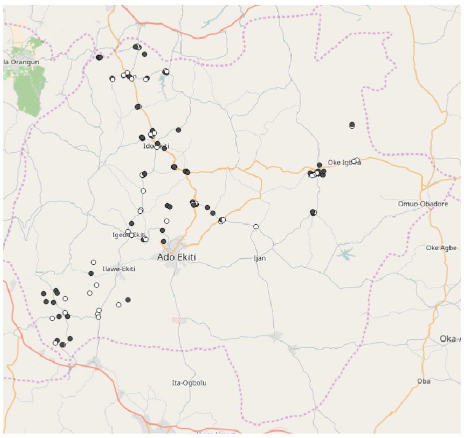
4.3 The experiment

We use evidence from a cluster-randomized trial carried out in Nigerian states of Ekiti and Enugu, implemented by WaterAid Nigeria. These states were chosen in collaboration with Wateraid Nigeria because of their relative low toilet coverage and the fact that they comprehend both urban and rural areas (see Figure 4.1). At the same time, within each state, the LGA's selected where those in which there were enough communities with no recent experience of CLTS-like interventions, by WaterAid or any other NGO, to achieve a large enough control sample.[7]

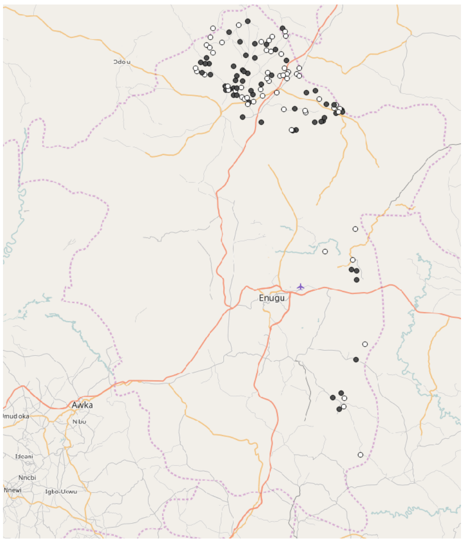
Figure 4.1: Geographical location of study areas in Enugu and Ekiti states



(a) The Nigerian states of Ekiti (left) and Enugu (right), in blue



(b) Ekiti state



(c) Enugu state

Note: CLTS (black) and control (white) clusters in the states of Ekiti and Enugu.

Below, we first describe the randomisation and implementation strategy (Section 4.3.1), then the sampling process, data collection and follow-up mea-

[7]In Enugu, the LGA's included were Igbo Eze North, Igbo Eze South, Nkanu East and Udenu. In Ekiti, the LGA's included were Ido Osi, Ikole, Moba, Irepodun Ifelodun and Ekiti South West.

surement timing (Section 4.3.2), key outcome measurements (Section 4.3.3) and finally present summary statistics from the baseline survey to demonstrate balancedness (Section 4.3.4).

4.3.1 Randomisation design and implementation

The unit of randomisation was defined by the nature of the intervention. CLTS is a village-wide intervention that invites all households within a community to the triggering meeting described above. A cluster randomized design was therefore deemed appropriate. At the same time, there are constraints on the minimum and maximum number of households that can be ‘triggered’ in a single CLTS meeting. Taking these constraints into account, our implementing partner, Wateraid Nigeria, divided the study sample into triggerable units, or clusters. These clusters are comprised of small villages, neighbourhoods or quarters, depending the type of settlement they are located in. Clusters do not match Nigerian administrative units, but are smaller or equivalent to Settlements/Autonomous Communities. On average, clusters consist of 1.7 villages or quarters, and their composition is similar in both states.[8] Triggering units consist of geographically close villages since all villages or neighbourhoods in a cluster would be triggered together. Additionally, clusters were designed to be self-contained, and not share markets or large public areas with each other, so as to avoid program spillovers. CLTS meetings were carried out once in each treatment cluster, and households from all villages in this cluster were invited to attend.

Randomisation of clusters was performed after stratifying by LGA, in order to ensure balanced treatment and control samples at this level. Within each LGA, half of the clusters were randomly assigned to receive CLTS treatment and half to remain untreated as control clusters. The result of this randomisation was then shared with Wateraid Nigeria.

[8]The median and modal number of villages or quarters within a cluster is 1. The maximum number of villages in a cluster is 7, occurring only once.

4.3.2 Sampling and data collection

The study frame was established in October 2014 by first carrying out a household census in the nine LGAs in Ekiti and Enugu. The census collected basic household information from all 50,333 households in the area (27,888 from Enugu and 22,445 from Ekiti).

To ensure representativeness within a limited budget, we randomly selected 20 households from each cluster for interview in our baseline survey. This pre-treatment survey was carried out between December 2014 and January 2015. We restricted the frame to areas with no history of past CLTS activities carried out by Wateraid or the Nigerian Government nor, to the best of our knowledge, by any other NGOs. Our final sample consists of 4,671 households from 246 clusters, distributed evenly across the two states, or around 9% of the population in the area, according to our census.[9]

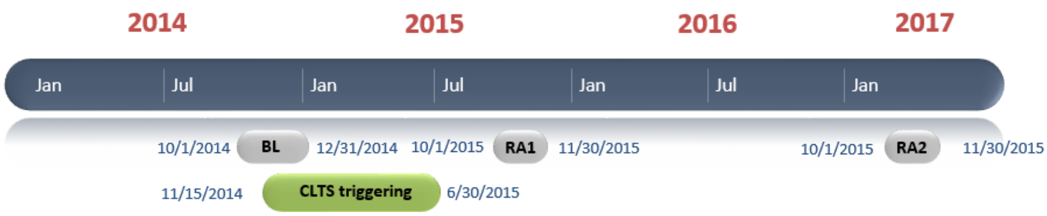
After baseline data collection was complete, households were randomly assigned to either treatment or control. CLTS was implemented in treatment areas during the first half of 2015. Follow up data was collected in two waves. The first follow-up took place between December 2015 and February 2016, in what we call the First Rapid Assessment (RA1). It measures outcomes between 6 and 12 months after the intervention, i.e. short term impacts. Medium term impacts are captured in the Second Rapid Assessment (RA2), which was carried out in March and April 2017. These two post-treatment measures allow us to study the dynamics of CLTS impacts over time. Figure 4.2 summarises intervention and data collection timings.

4.3.3 Key outcomes and covariates

At baseline, detailed information regarding all household members in our sample was collected. This includes basic demographics such as age, gender and level of education, as well as employment status, income and expenses, sani-

[9]Written consent was obtained from every household before interviews were carried out. An external partner, InDepth Precision Consult, was in charge of all data collection rounds and was blinded to treatment status. Baseline questionnaires were carried out by pen and paper, while RA1 and RA2 surveys were carried out using an electronic survey system.

Figure 4.2: Project timeline: implementation and data collection waves



Note: CLTS implementation, above, in green. Baseline (bl), first rapid assessment (RA1) and second rapid assessment (RA2) surveys in grey. CLTS implementation was not carried out in a given location until baseline data was collected.

tation practices, health status and characteristics of the dwelling. A series of questions were also included to measure respondents' expected benefits from sanitation, as well as their beliefs about social norms and their awareness of health externalities.

The main aim of CLTS is to increase ownership rates of private toilets and reduce or eliminate the practice of open defecation. Toilet ownership can be measured along dimensions of quantity and quality. The simplest outcome measure is whether a household owns or is constructing a toilet of any kind. Keeping track of construction (rather than counting finished toilets) is important as many of the households in our sample are likely credit constrained, hence construction efforts may involve several smaller investments that are spread over a longer period. Our second measure reflects quality. Since precarious pit latrines are filled frequently and require regular emptying (hence allowing for toilet divestment through lack of maintenance), it measure whether households own a functioning toilet. Since a functioning toilet is a necessary condition for a household to abandon open defecation, it will be our main outcome of interest. Given the pervasiveness of unimproved, unsafe sanitation in rural Nigeria, we include a third, stricter quality measure: whether a household owns a finished, functioning and improved toilet. A toilet will be considered improved if it satisfies the criteria used by the WHO/UNICEF Joint Monitoring Program.

Finally, we ask respondents about their sanitation practices, and whether they perform open defecation. There is a risk of systematic over-reporting of toilet ownership and under-reporting of OD habits in treatment areas when measures are self reported. We cannot rule out such measurement error with respect to our two sanitation practice measures. To validate our toilet ownership measurements, interviewers at RA1 asked households whether they could see the latrines at the end of each interview. 68% of households who declared to own a toilet allowed interviewers to inspect them - 70% in CLTS and 65% among control households. Inspection did not yield discrepancies between self-reported and actual ownership levels. We interpret both - the similar consent rate for inspection that is - if anything- higher in treatment areas, and the truthful reports of those that consented as evidence that there is no systematic measurement error in latrine ownership across groups. Our results are qualitatively and quantitatively very similar across the three measures of toilet ownership, and closely mirrored by reverse results on open defecation practice.

Together with the information on outcomes, we have gathered rich data on household characteristics, which we will use as control for increased precision in the estimation, and in the exploration of heterogeneous impacts. One important dimension we will be looking at is household wealth. We measure this using a relative wealth index, built using information on household asset ownership, at baseline, by principal component analysis. As described in McKenzie [2005], this is a useful, and accurate, way of measuring relative wealth, in developing country contexts, where income and expenses show significantly higher levels of volatility. The index was normalized to have a maximum value of 1 and a minimum of 0, with a standard deviation of 0.12. Details on the assets included in the construction of the index, and their factor loadings, are presented in Table C.1, in the Appendix.

4.3.4 Summary statistics

Our sample consists of 4,646 households from nine LGAs in Ekiti and Enugu. Table 4.1 shows that treatment and control group are balanced in terms of mean outcomes and on a series of controls. The only exception is household size which is slightly larger in the control group. We will include this variable in our regressions to confirm that they do not affect our results.

Table 4.1: Balance between Treatment and Control groups at Baseline

	Control	Treatment	P-value
Toilet Ownership			
HH has (or is constructing) a latrine (%)	37.52	37.49	0.99
HH has a functioning latrine (%)	36.19	35.87	0.92
HH has a functioning, improved toilet (%)	32.68	33.01	0.91
Toilet Usage			
All members of household use toilet (%)	34.09	33.78	0.91
Main respondent performs OD (%)	61.66	61.22	0.89
Head Characteristics			
HH head age	55.60	54.32	0.15
HH head male (%)	64.04	62.47	0.38
HH head employed (%)	76.79	76.04	0.69
Highest education level attended by HH head	1.439	1.451	0.88
HH size	3.991	3.733	0.03**
Children under the age of 6	0.486	0.472	0.69
Household Characteristics			
HH primary activity is farming (%)	45.05	48.69	0.32
HH income, past year (th. USD)	0.528	0.574	0.25
Relative asset wealth index	0.00	0.00	0.54
HH has any savings (%)	22.50	22.73	0.92
HH has any debt (%)	20.63	19.50	0.50
Home-owner (%)	62.08	64.04	0.56
Renter (%)	15.10	14.00	0.63
F-Test - All variables	F(18,245)= 1.46		0.10
F-Test - Exc. HH size	F(17,245)= 0.89		0.59
Observations	4667		

Notes: Mean values measured at baseline. Statistically significant differences between CLTS and control households appear at the expected rate and are found only for household size (at the 5% level). Improved toilets defined using the classification in Unicef and WHO [2015]. Relative wealth index constructed by principal component analysis of a series of questions regarding asset ownership, following for example McKenzie [2005]. The excluded category for household tenure is free tenure, in the form of squatting, or borrowing. *Source:* Baseline questionnaire.

Additionally, we ran a simple OLS regression taking treatment status as a dependent variable, and included all variables in Table 4.1 as controls. We then performed an F test of joint significance for the whole set of regressors, and found that we cannot reject the null hypothesis at the 10% level (as shown in the bottom row "F-Test = All variables"). Once we removed household size, the single variable for which we observe an imbalance, the explanatory power of the remaining variables falls significantly (see next row in table). This supports our claim that treatment and control samples are on average identical, except for their size.

4.4 Empirical method

We measure the program's impacts using analysis of covariance (ANCOVA) estimation. The difference between ANCOVA and a standard difference in difference (DID) approach is that in the former we introduce the outcomes measured at baseline as a control variable, instead of treating them as a pre-treatment survey wave. McKenzie [2012] describes the efficiency advantage of the ANCOVA estimator compared to both DID and simple difference (SD) estimators, and shows that ANCOVA is always preferable in experimental contexts if pre-treatment information is available. Indeed, in our case with a baseline and two post-treatment waves, the ratio of DID to ANCOVA estimator variances is equal to $\frac{3}{1+2\rho}$, where ρ is the autocorrelation of outcome variables across survey waves. For example, in the case of toilet ownership or construction, $\rho = 0.57$ between baseline and first follow-up, so that the variance of the DID estimator is 40% higher than that of ANCOVA.[10] We chose ANCOVA as our preferred estimator due to this significant increase in power.[11]

In our first specification, we do not distinguish between short- and medium-run impacts and pool both follow-up waves. We compare average

[10]Given that toilets are lumpy investments in household infrastructure, we expected to observe higher autocorrelation between outcomes. Latrine inspections by interviewers and consent rates to do so suggest that this moderate ρ is not driven by measurement error across groups.

[11] We also present DID and simple difference estimates for the main results in the appendix (see Table C.3). They are virtually identical to the ANCOVA estimates.

outcomes between CLTS and control households as follows:

$$y_{ivgt} = \gamma CLTS_v + \theta y_{ivg0} + X_i'\beta + \delta_t + \mu_g + \epsilon_{ivgt} \tag{4.1}$$

Where y_{ivgt} is the outcome variable for household i, located in LGA g in cluster v, measured at follow-up $t = \{1,2\}$. $CLTS_v$ is an indicator variable equal to 1 if the cluster is part of the CLTS group. The coefficient of interest will be γ, the causal impact of the CLTS treatment. y_{iv0} is the value of the outcome variable and X_i' is a vector of household characteristics, both measured at baseline ($t = 0$). In most of our specifications, the covariates included in the regressions will be: gender, age, age squared, employment status and education attainment of the household head, as well as household size and a dummy indicating whether farming is the household's main economic activity. Finally, we introduce a time fixed effect δ_t, which is a dummy for RA2, and for LGA fixed effects μ_g to remove level differences across LGAs.

An alternative approach to ours is the one used by Cameron et al. [2015], who use DID specifications but run it only on the sub-sample of households that did not have a toilet at baseline. We believe that our approach is a more comprehensive one, because besides persuading non-owners to construct toilets, CLTS informs households who already own a toilet, about the importance of its maintenance and usage. At baseline, more than 70% of the toilets in our sample were pit latrines of different sorts. These pits require regular emptying (annual or biannual, generally), and will sometimes collapse and become unusable. It is therefore a margin that we think should be contemplated in our estimations, which is why we include the whole sample of households and use an ANCOVA specification to control for baseline outcomes.

In a second specification, we explore how CLTS impacts evolve over time, and estimate impacts separately for the short and the medium-run. In this case we will have two coefficients of interest γ_1 and γ_2, corresponding to impacts as measured at RA1 and RA2 respectively. The specification changes to:

$$y_{ivgt} = \sum_{t=1}^{2} \gamma_t \left(CLTS_v \times I_t\right) + +\theta\, y_{ivg0} + X_i'\beta + \delta_t + \mu_g + \epsilon_{ivt} \qquad (4.2)$$

In further analysis, we look into heterogeneous program impacts. We will do this by allowing CLTS impacts from both Equation 4.1 and Equation 4.2 to vary for specific sub-populations. Given the geographical distribution of our study areas, we will estimate standard errors robust to correlation at the cluster level, which is also the level at which the treatment was randomised.

4.5 Results

Table 4.2 presents estimates of CLTS treatment effects on the four key outcomes capturing toilet ownership and open defecation. We use the two specifications described in Section 4.4. Results in panel A are based on Equation 4.1 which pools observations across the two both follow up periods. Programme impacts are allowed to vary over time in Panel B which shows results from the specification detailed in Equation 4.2. All specifications include household controls, LGA fixed effects and a dummy variable for the second follow up period.[12] In this ANCOVA specification, we drop all observations at baseline but include the value of the dependent variable measured at baseline as additional control.

Panel A shows that CLTS increased toilet ownership in all three specifications by 3 pp, significant to the 10% level. Average toilet ownership in control areas at RA1 was 45%-35% according to the outcomes used, so this represents an increase in coverage of 7%-9%. The equality of estimated coefficients across the first two measures of toilet ownership, suggests no change in toilet maintenance patterns, a common concern in areas like those in our study, where pits require regular emptying and reinforcing. By the same token, the similar estimates from Panel A, Columns (2) and (3), reject a second concern regard-

[12] For further reference, Table C.2 in the Appendix reproduces the results from Panel A and include the estimated coefficient for all control variables.

Table 4.2: CLTS impacts on toilet ownership and open defecation, ANCOVA

LHS: Toilet/OD	Cons./Finished	Functioning	Improved	OD (main resp.)
	(1)	(2)	(3)	(4)
Panel A: Pooled estimates				
CLTS (γ)	0.03*	0.03*	0.03*	-0.04**
	(0.02)	(0.02)	(0.02)	(0.02)
Panel B: Impacts by period				
CLTS x RA1 (γ_1)	0.04**	0.03*	0.02	-0.05**
	(0.02)	(0.02)	(0.02)	(0.02)
CLTS x RA2 (γ_2)	0.02	0.03	0.03	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)
HH controls	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes
Survey round FEs	Yes	Yes	Yes	Yes
Control Mean (BL)	0.38	0.36	0.33	0.61
F-test $\gamma_1 = \gamma_2$ (p-value)	0.29	0.79	0.70	0.54
No. of TUs	247	247	247	247
No. of HHs	4,555	4,555	4,555	4,555
No. of obs.	9,110	9,110	9,110	9,110

Notes: HH covariates: age, age squared, gender, education attainment level and employment status of the HoH; HH size and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* BL, RA1 and RA2 household surveys.

ing community-led sanitation interventions: that they might be effective at stimulating unimproved sanitation only. Open defecation, as reported by the main respondent in each household, followed the trend of ownership variables and fell by 4 pp, or just below 7%, significant to the 5% level.[13]

CLTS aims to reduce open defecation by a) promoting investment in toilets among non-owner households, and b) stimulating behavioural change among households that own toilets but perform open defecations anyway. In some contexts, these pathways may be hard to tease out due to the existence of public toilets that non-owning households may use, or, as in the case of India, the low levels of usage observed among toilet owners. In our study area, public toilets are rare and usage of private toilets is high. Indeed, at baseline, 95% of non-owners declared to defecate in the open, while among toilet owners, only 4% did it. At RA2, open defecation rates for households who owned

[13] All results presented here are robust to using DID or simple difference estimators instead (see Table C.3).

functioning toilets at both baseline and RA2, and for those who only had one at RA2 (e.g. new toilet owners), were 4% and 5% respectively. These two pieces of evidence as suggest that toilet construction is the main channel through which open defecation was reduced in this area.

Panel B presents the results of using specification 4.2, where impacts are estimated separately for the short and medium terms. We see stronger short-term improvements in the outcomes constructing or owning a toilet, owning a functioning toilet and performing open defecation in the short-run (compared to the pooled estimation). Medium term impacts, on the other hand, are only observed for open defecation. This suggests that two years after the intervention, when RA2 data was collected, differences in toilet construction or ownership between CLTS and control areas were no longer detectable. Nonetheless, when a Wald test for equality of coefficients is run comparing γ_1 and γ_2, we cannot reject the null hypothesis. Given that impacts are small, we do not have enough power to assert that they are indeed different.

As discussed in the introduction, evidence from other recent CLTS-like interventions has shown both extremely high impacts and no statistical impacts at all. Table 4.2 above provides a middle ground: while it appears that CLTS had some effect on toilet construction, these impacts appear to be relatively small. At the same time, we observe that, in the case of Nigeria, construction increased at similar rates than ownership of improved toilets. Furthermore, we find reductions in open defecation that mirror the increase in toilets. This rejects two concerns regarding CLTS, namely that a "no subsidy approach" might mostly stimulate the construction of unimproved toilets, and that construction and usage might be independent. Our results reject both hypotheses.

Next, we compare our results to other studies evaluating similar interventions. Our estimates of CLTS impacts are dwarfed by those published in some recent studies from other developing countries. For example, Pickering et al. [2015] report increases in toilet ownership of 30 percentage points from a cluster-randomized CLTS intervention in Mali. Clasen et al. [2014] carried out a similar trial in 100 rural villages in Odisha and report CLTS-driven impacts of 50 percentage points in toilet ownership. However, both studies are not directly comparable to ours since the researchers targeted a specific set of households to interview. While our study sample is composed of a random draw of all households in the selected LGAs, Pickering et al. [2015] surveyed just households with at least one child below the age of ten, and only included villages with toilet coverage of less than 60% in the study frame. Similarly, Clasen et al. [2014] only surveyed households with at least one child under the age of 4 or with a pregnant woman.

Two studies report smaller programme impacts, whose magnitude is closer to the results of our study. Both focus on a similarly selected sub-sample of households with small children. An evaluation of the Total Sanitation and Sanitation Marketing (TSSM) campaign in Indonesia, which contained a CLTS component, found that toilet coverage increased by only 3 pp (Cameron et al. [2013]). Their sample of households was restricted to those with at least one child under the age of 2. Briceño et al. [2015] evaluated a similar TSSM campaign in Tanzania, interviewed households with children under the age of 5, and found impacts of 8 pp. This focus on families with small children is based on the presumption that parents of small children have stronger preferences for health investments due to their lifetime returns when such investment improves childhood circumstances.

To our knowledge, the only study that also interviewed a random sample of households in a village, was carried out by Guiteras et al. [2015b] in Bangladesh. The authors find no evidence of a statistically significant impact from a CLTS-like intervention in the absence of subsidies. The positive and significant impacts of CLTS we report from the Nigerian experiences are therefore a novel finding.

To compare our findings to studies that use selected groups, we replicate the sample selection criteria used in the above studies and estimate treatment effects for each sub-group in our sample. At baseline, we identified all households with at least one child below the age of 6, and estimated CLTS treatment effects for that subset of households separately. For comparison purposes, we repeated the same exercise on households without children. Columns (1) and (2) in Table 4.3 present the results of this analysis on the outcome of ownership of a functioning toilet. CLTS treatment effects are only slightly higher for the sub-sample of households with children. Pooled treatment effects for households with and without children are both significant to the 10% level. A single regression that interacts treatment with a dummy for presence of children does not allow us to reject the null hypothesis that programme impacts are identical among the two groups, as shown in the row labelled "F-test γ Yes=No (p-value)". Not only are the effects virtually identical between the two groups, the level of ownership of functioning toilets were also similar at baseline. Tables C.6, C.7 and C.8 in the Appendix confirm these findings for our other outcome measures: the presence of children does not seem to strongly determine how households react to CLTS.

Table 4.3: Heterogeneous CLTS impacts on ownership of functioning toilets

LHS: Functioning toilet	Children <6 y/0		Female HoH		Uneducated HoH		<median wealth	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	No	Yes	No	Yes	No	Yes	No	Yes
Panel A: Pooled estimates								
CLTS x Post (γ)	0.03*	0.04*	0.02	0.05**	0.02	0.06**	0.02	0.06**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Panel B: Impacts by period								
CLTS x RA1 (γ_1)	0.03	0.04*	0.02	0.06**	0.02	0.07***	0.02	0.07**
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
CLTS x RA2 (γ_2)	0.03	0.03	0.02	0.04	0.02	0.05	0.03	0.05*
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Control Mean (BL)	0.36	0.37	0.38	0.33	0.40	0.28	0.48	0.21
F-test γ Yes=No (p-value)	0.88		0.25		0.09		0.19	
F-test γ_1 Yes=No (p-value)	0.64		0.20		0.07		0.10	
F-test γ_2 Yes=No (p-value)	0.85		0.52		0.29		0.50	
No. of Triggerable Units	247	238	246	238	244	240	232	245
No. of households	3,201	1,354	2,888	1,667	3,104	1,451	2,037	2,025
No. of observations	6,402	2,708	5,776	3,334	6,208	2,902	4,074	4,050

Notes: HH covariates: age, age squared, gender, education attainment level and employment status of the HoH; HH size and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* BL, RA1 and RA2 household surveys.

A second important target population discussed in the health investment literature are female headed households. With respect to CLTS, Kar [2003] poses that women are “one of the greatest internal forces for mobilisation and promotional activities in the villages”. The literature presents two rationales why female headed households may react stronger to CLTS and may invest more into health (here: sanitation) in general. first, there is some evidence suggesting gender-specific preferences in certain domains, such as health and children’s welfare; second, women may enjoy larger returns from sanitation in terms of personal safety and privacy. The intra-household bargaining literature has established that when preferences of household members over expenditures or investments differ, observed choices will be the result of a bargaining process. In consequence, even if the demand or health investments is higher among females (compared to males), it may only be converted into investment in combination with female decision power. Miller and Mobarak [2013] provide supportive evidence for this. They conduct a field experiment designed to separate between female preferences and health technology adoption decisions in the household.[14] They show that the adoption of a technology, that faces higher demand from women than men, is constrained by the level of female decision making authority within the household.

As a first approach to this, we estimated CLTS impacts separately for female headed households and non-female headed households, i.e. those in which women have decision versus those where they do not. As Columns (3) and (4) in Table 4.3 show, it is indeed female headed households who experience the strongest treatment effects from CLTS. Note that at baseline, female-headed households had lower toilet coverage rates, meaning that the information campaign seems to be more effective in households where women have the decision power to build a toilet and helps them catch up with male-headed households.[15] Thus, while we do not separately observe preferences from decisions,

[14] The health technology they study is an improved cooking stove that emits less indoor smoke.

[15] At baseline, toilet ownership rates in female headed households are 5 percentage points lower than in male headed households. CLTS levels this difference.

our results are consistent with both, higher sanitation preferences of females and higher sanitation investment in households where females hold decision power.

We test for the robustness of this result through two additional measures of female decision power. In the first, we define households with female decision power as those which are female-only (e.g. no adult males), female headed, households in which a women has the highest level of educational attainment, and households in which at least one adult woman is employed while no men are. At baseline, 54% of our sample entered into this category. In a second approach, we consider households with female decision power as those in which, when asked who in the household decides about major household investments, the respondent (the eldest woman in the household), answered that it was her who decided.[16]. 21% of our sample fell into this category at baseline. Table C.4 in the Appendix shows that the parameter estimates are identical to those for the more restrictively defined female headed households. To the best of our knowledge, this is the first quantitative evidence of the differential impact of CLTS according to female decision power within the household.

A common concern regarding lumpy investments in developing countries, is that underinvestment may be concentrated among households with scarce financial resources, e.g. due to liquidity constraints, leading to a prominent role of subsidies or micro-loans in interventions aimed at fostering investment. India's Total Sanitation Campaign (TSC), for example, aimed at increasing investments by providing large construction subsidies to eligible households. A unique trait of the CLTS approach is that it does *not* include any subsidies or credit for toilet construction. This has been hailed by its creators as a fundamental aspect of its design, since it avoids creating a "culture of dependence on subsidies" (IDS [2011]). At the same time, evidence has shown that this lack of financial support might undercut its effects. As mentioned above, Guiteras et al. [2015b] found no evidence of CLTS impacts on toilet

[16] Alternative answers were her partner, her and her partner jointly, or someone else in the household

ownership unless combined with subsidies, while Cameron et al. [2013] find that toilet construction in non-poor households explain most of the impacts in their evaluation.

We therefore investigate whether CLTS has heterogeneous impacts according to household wealth levels. We use two different measures to proxy for wealth: the level of education of the household head, and an index of relative wealth. Notice that, at baseline, ownership of functioning toilets was significantly lower among the least educated (28% v 40%), and among the relatively less wealthy households (21% v 48%). Columns (5) to (8) of Table 4.3 show that households whose head has not finished primary education and households with below median wealth experienced CLTS program impacts of 7 percentage points in the short term, significant at the 5% level. Given these groups' initially low rates of toilet coverage, these impacts represent an increase in toilet ownership of 25% (among low educated households) and 33% (among households with below median wealth). For below median wealth households, we also find medium term impacts that are economically and statistically significant. Their toilet coverage rates are 5 percentage points higher than those of households in the control group in RA2. On the other hand, households with educated heads and above median wealth appear to have experienced no significant programme impacts, even though their coverage rates were below 50% at baseline.

A reasonable concern at this point, might be that the different household characteristics being used to compare programme impacts so far, are correlated, and proxies for other, underlying characteristic that is not observed. While we cannot reject the latter statement, Table C.5 in the Appendix present the pairwise correlations between each group of households presented in Table 4.3 above. The presence of children is negatively correlated with the remaining three categories: female head, uneducated head and below median asset wealth. Having a female head is positively, although not strongly, correlated with the head having no primary education and being asset poor. These three, non-identical but overlapping groups, have an important characteristic in common: they all exhibit lower toilet coverage at baseline than their complements. This could explain why we observe stronger impacts among them than in the rest of sample.

This is a novel finding, given that CLTS provides no subsidies. Tables C.6, C.7 and C.8 in the Appendix confirm these findings for the remaining outcomes: while the presence of children does not seem to determine how households react to CLTS, households with below median wealth (or low education) experienced the highest impacts. A reasonable concern is that these households might construct lower quality toilets, i.e. unimproved ones, which would undermine the health benefits expected. We do not, however, find evidence supporting this concern. Table C.7 shows that CLTS increased the ownership of improved toilets by 4-5 pp in the short term and by 6-5 pp in the medium term, when considering uneducated heads of households or households with below median wealth.

4.6 Channels of impact

Having established that households with lower education, lower wealth and those with a female head react more strongly to a health information campaign such as CLTS, we proceed to study the channels through which the intervention worked. Several mechanisms have been discussed in the health investment literature, or have been suggested by anecdotal evidence from the field. The aim of this section is to shed light on the constraints dampening health investments in a developing country context, as revealed by the patterns of adoption and non-adoption brought about by CLTS.

First, we will investigate information about its costs and benefits as a channel to investment. In Section 4.6.1, we investigate whether the expected (rather than actual) cost of constructing a toilet affects investment decisions, and whether CLTS may have corrected price misconceptions. In Section 4.6.2), we explore whether expected benefits from sanitation played a role in toilet adoption, and whether these were changed through CLTS. A second channel is that CLTS may have been more effective in communities with higher social capital, as proposed by Cameron et al. [2015], and that CLTS may have affected social capital and associativity through its coordination approach (Section 4.6.3). Finally, we consider whether CLTS program impacts can be explained by institutional sanctions on open defecation that may have been imposed by village leaders following CLTS (Section 4.6.4).

4.6.1 Expected costs

Sanitation investment decisions are likely driven by expectations about its costs and benefits. (Initial) misconceptions about the cost of installing a toilet may therefore lead to suboptimal investment decisions (underinvestment by those who overestimate its costs and vice versa). Correcting misconceptions on the cost of installing a toilet could be a mechanism by which CLTS may increase toilet ownership in programme areas, as suggested in Alzua et al. [2017]. The CLTS activities do not cover any aspects of construction costs, hence misconceptions at the household level could only be corrected through CLTS' collective nature. By triggering conversations and information exchange between members of the community, CLTS could have helped households with overly high expected costs of construction to learn that it was, in fact, a relatively affordable investment.

At baseline we collected information on actual and expected prices of toilet construction. Figure 4.3 plots the distribution of both minimum and maximum expected construction costs for building a ventilated improved pit latrine. The cost question was posed to all households in the sample. The blue line plots average construction costs as reported by households who own a toilet and recalled the total cost of its construction. *Average* expected costs closely match actual costs. However, there is significant variation in expected prices which leads to a fraction of households overestimating actual costs, and another fraction underestimating them.

The CLTS intervention carried out in Nigeria did not explicitly include the provision of accurate cost of construction data. If, however, CLTS corrects cost misconceptions, and moves expected costs closer to actual ones, then we would expect to see an increase in investment among those who initially overestimated costs at baseline (and potentially, a decline in investment among treated households who underestimated them). In contrast, if CLTS does not change cost expectations, but costs play a role in the investment decision, then we should observe higher toilet construction levels among households with lower expected costs at baseline than by those who expect high construction costs, in both treatment and control groups. This is because, if CLTS successfully increased expected benefits from toilet ownership among households, these would still be compared to the expected costs of investing.

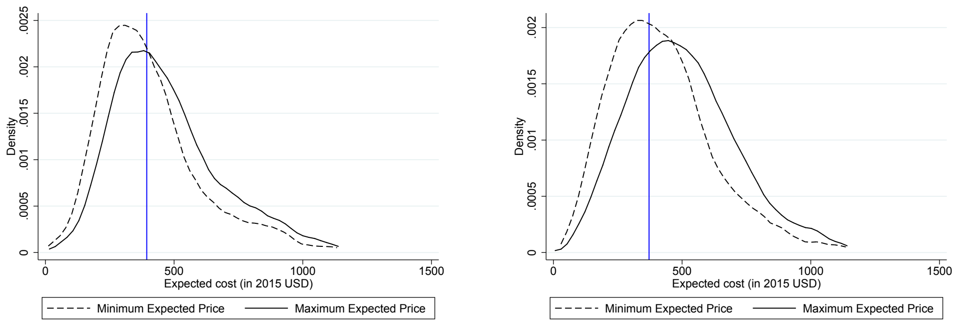
Two simple testable predictions follow: first, at baseline, households with higher cost expectations should ceteris paribus be less likely to have constructed a toilet, than those with lower cost expectations. Indeed, we find that having cost expectations equal to one standard deviation above the mean, is associated with a 4 pp lower likelihood of owning a functioning toilet.[17] Second, if CLTS corrects misconceptions of cost, then we should observe a stronger CLTS impact among those with initially high expectations of cost and a negative impact of CLTS on those whose estimate was too low at baseline (relative to the control group whose misconceptions are not corrected).

We construct a discrete variable equal to one if, at baseline, a household reported expected costs above the median actual cost of construction. Median construction costs are calculated at LGA level, to control for regional price variation that might be correlated with baseline toilet ownership rates (Augsburg and Rodriguez-Lesmes [2015]). Interacting this variable with our treatment indicator variable, we are able to estimate program impacts separately for households with expected costs above and below the median. In order to avoid extreme values, expected cost of construction answers were truncated at a maximum of USD 1,000, which affected less than 2% of the households in

[17]See Table 4.7. The standard deviation of the natural logarithm of expected prices is equal to 0.63.

Figure 4.3: Distribution of actual and expected toilet construction costs (in 2015 USD)

(a) Ekiti state **(b)** Enugu state



Note: Blue line indicates average actual construction costs, as reported by toilet owning households, for any type of toilets. Expected minimum and maximum prices are average expected prices over four different toilet models presented to respondents. *Source:* Baseline household surveys.

the sample. Table 4.4 shows the results of this analysis.

Although the question regarding expected construction costs was directed at all households, we only obtained 2,011 valid responses, from our sample of 4,600. This explains the lower number of observations in the regressions presented in Table 4.4, and the smaller precision in its estimates. Nonetheless, point estimates from Panel A in the table suggest a larger treatment effect among households with lower expected costs. This is line with the idea that, given that CLTS did not aim to correct these perceptions, high expected costs will be a barrier to adoption. While not statistically significant, treatment effects from Panel B are consistent with those observed for the whole sample: higher point estimates are found in the short term (RA1) than in the medium term (RA2). The sign and magnitude of short term point estimates also suggest a stronger CLTS response from households who underestimated construction costs. Taken together, these results suggest that there was no correction of mistaken beliefs regarding prices at baseline, but that high expected costs were an impediment that CLTS did not overcome.

Next, we split the sample into the groups for whom we found heterogeneous treatment impacts (households with children, low education, low wealth

Table 4.4: Impact of CLTS by baseline levels of Expected Cost

Dep var:	Cons./Finished	Functioning	Improved	OD
	(1)	(2)	(3)	(4)
Panel A: Pooled estimates				
CLTS x High Expected Cost	-0.01	0.02	0.03	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)
CLTS x Low Expected Cost	0.04	0.02	0.02	-0.02
	(0.03)	(0.02)	(0.02)	(0.02)
Panel B: Impacts by period				
CLTS x RA1 x High Expected Cost	0.00	0.01	0.02	0.00
	(0.03)	(0.03)	(0.03)	(0.03)
CLTS x RA1 x Low Expected Cost	0.04	0.02	0.02	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)
CLTS x RA2 x High Expected Cost	-0.01	0.02	0.04	-0.03
	(0.04)	(0.03)	(0.03)	(0.04)
CLTS x RA2 x Low Expected Cost	0.02	0.01	0.01	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)
HH controls	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes
F-test (p-value)	0.25	0.99	0.78	0.76
F-test RA1 (p-value)	0.37	0.91	0.87	0.42
F-test RA2 (p-value)	0.28	0.89	0.55	0.78
No. of TUs	238	238	238	238
No. of HHs	1,958	1,958	1,952	1,952
No. of Observations	3,916	3,916	3,904	3,904

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01.* *Source:* BL, RA1 and RA2 household surveys.

or a female-head) and repeat the estimation of interaction effects between CLTS and expected costs. Our intention is to find out whether, within the sub-samples where we observed the strongest treatment effects, high expected costs were similarly inhibiting of toilet adoption and their consequent reductions in open defecation. Each panel from Table 4.5 shows the results for each of our four main outcomes. Comparing the point estimates of households with high and low expected prices, among our four sub-samples of interest, we see that in most of the cases, the CLTS response in these groups is concentrated among those with low expected construction costs at baseline. For example, from Panel A, showing results for the construction or ownership of toilets, we see that households whose head had no primary education, saw treatment effects 14 pp larger, if they had low expected costs of construction than otherwise. The size of the sample in all these regressions is reduced due to missing responses, which affects our power and the statistical significance of our estimates. Nevertheless, these results suggest that the information campaign only managed to persuade households with initially low sanitation uptake that, at the same time, did not think that building a toilet would be extremely onerous. Secondly, they imply that there is room for a complementary intervention that provides cost information and targets households with excessively high expected costs, who might be willing to invest if their cost misconceptions were corrected.

4.6.2 Expected benefits

We next explore the role that new information about the benefits of sanitation may have in fostering sanitation. CLTS triggering meetings focused extensively on the dangers of open defecation, its consequences for the whole community, and on the individual health benefits of sanitation improvements. It also aimed at attaching feelings of disgust and embarrassment to the practice of open defecation, and of pride to the ownership of a private toilet. CLTS might have thus increased expected benefits from private sanitation investments, leading to increased demand for them. Affordability is defined by the net benefit of a good: the difference between expected benefit and cost. The information provided by CLTS facilitators might have led to an upwards revision of benefit expectations, tipping treated households at the margin towards construction.

Our rich survey data from both baseline and RA2 allows us to construct three indices of expected benefits from private sanitation and costs from open defecation. These indices separate benefits along the dimensions of i) private, mostly health-related benefits and costs, ii) a set of subjective rewards in terms of social status and pride resulting from toilet ownership and embarrassment associated with open defecation, and iii) externalities. Table 4.6 lists the ques-

Table 4.5: Heterogeneous Impact of CLTS by baseline levels of Expected Cost

LHS:	Child <6 y/0		Fem. HoH		Uned.d HoH		< med. wealth	
	(1) No	(2) Yes	(3) No	(4) Yes	(5) No	(6) Yes	(7) No	(8) Yes
Panel A: Cons/Fin. toilet								
CLTS x High Exp. Cost	-0.01	0.01	0.01	-0.04	0.00	-0.03	0.00	-0.01
	(0.04)	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)	(0.04)	(0.05)
CLTS x Low Exp. Cost	0.05*	-0.00	0.00	0.11**	0.01	0.11**	0.01	0.06
	(0.03)	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)	(0.03)	(0.04)
F-test (p-value)	0.11	0.88	0.97	0.04	0.74	0.04	0.82	0.21
Panel B: Funct. toilet								
CLTS x High Exp. Cost	0.02	0.01	0.02	0.00	0.02	0.01	0.05	0.00
	(0.03)	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)	(0.03)	(0.04)
CLTS x Low Exp. Cost	0.01	0.03	-0.01	0.07	-0.00	0.10**	0.00	0.05
	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.05)	(0.03)	(0.04)
F-test (p-value)	0.93	0.71	0.37	0.25	0.50	0.15	0.26	0.34
Panel C: Improved toilet								
CLTS x High Exp. Cost	0.04	0.00	0.03	0.04	0.03	0.02	0.07*	0.00
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
CLTS x Low Exp. Cost	0.02	0.04	-0.00	0.07	0.01	0.07	-0.00	0.04
	(0.03)	(0.03)	(0.03)	(0.05)	(0.03)	(0.05)	(0.03)	(0.04)
F-test (p-value)	0.61	0.53	0.44	0.70	0.50	0.44	0.13	0.48
Panel D: Open defecation								
CLTS x High Exp. Cost	0.00	-0.03	-0.01	-0.02	-0.01	-0.02	-0.03	-0.01
	(0.03)	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)	(0.04)	(0.04)
CLTS x Low Exp. Cost	-0.02	-0.02	-0.00	-0.05	0.00	-0.12**	0.01	-0.06
	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.05)	(0.03)	(0.04)
F-test (p-value)	0.54	0.77	0.90	0.62	0.73	0.13	0.30	0.34
HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of TUs	231	201	235	190	234	178	219	196
No. of HHs	1,329	629	1,410	548	1,494	464	1,071	707
No. of Observations	2,658	1,258	2,820	1,096	2,988	928	2,142	1,414

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* BL, RA1 and RA2 household surveys.

tions that we use for these indices.[18] We aggregate the questions listed in the

[18] The last question included in the private benefit index might also be considered as part of the externalities associated with toilet ownership. Because we were more interested in the household's understanding of the true costs and benefits of toilet ownership, we included it here and not in the externality index. However, including it in the externality index leads to identical results.

table using principal component analysis. The advantage of PCA over equally weighed indices is that the former puts more weight on questions that present a higher variance over the population, and therefore helps us create a richer ranking within our sample (McKenzie [2005]). We standardize responses so that positive values imply higher expected benefits/costs from toilet ownership/open defecation, and then extract one factor via principal component analysis. We observe most, but not all question items at baseline and at RA2, so we estimate the three indices at baseline and RA2 separately.

Table 4.6: Definition of Expected Benefit Indices

Index/Questions	BL		RA2	
	(1)	(2)	(3)	(4)
Private costs & benefits				
If a neighbour built a toilet for the first time, do you think his/her family...				
...will be healthier because of the toilet/latrine?	✓	0.45	✓	0.46
...will be more productive because of this toilet/latrine?	✓	0.35	✓	0.27
...will feel that women in the family will be safer with this toilet/latrine?	✓	0.43	✓	0.47
...will save time because they now have this toilet/latrine?	✓	0.37	✓	0.31
...will get sick more easily when using this toilet/latrine?	✓	0.43	✓	0.45
...will see the women getting infections because of this toilet/latrine due to pit heat?	✓	0.41	✓	0.42
You (most people in your community) believe that				
...defecating in the open is unhealthy			✓	0.06
...defecating in the open is dangerous			✓	0.10
...defecating in a toilet is dangerous			✓	0.09
...if a household has a toilet/latrine, neighbours will come to use it			✓	-0.03
Emotional benefits				
If a neighbour built a toilet for the first time, do you think his/her family...				
...will be happier because of the toilet/latrine?	✓	0.48	✓	0.48
...will be less embarrassed when family and friends come to visit?	✓	0.50	✓	0.45
...will feel proud because of having this toilet/latrine?	✓	0.54	✓	0.54
...will be have a higher status in the society because of the toilet/latrine?	✓	0.48	✓	0.41
You (most people in your community)				
...would feel proud of owning a toilet			✓	0.00
...would feel embarrassed to defecate in the open			✓	0.25
...believe that it is acceptable to defecate in the open			✓	0.20
It is acceptable to defecate in the open	✓	0.02		
Health externalities				
You (most people in your community) believe that				
...the use of toilets by neighbours protects you from sickness.			✓	0.64
...if your neighbours use a toilet/latrine, the environment you live in is cleaner			✓	0.65
...the use of toilet/latrines by any of your neighbours may cause you harm			✓	0.41
The use of toilets by any of your neighbours protects you from sickness	✓	0.65		
If my neighbours use a toilet/latrine, the environment I live in is cleaner	✓	0.61		
The use of toilet/latrines by any of your neighbours may cause you harm	✓	0.45		

Note: Questions included in each of the three expected benefit indices. Columns (1) and (3) indicate whether each question was included at Baseline or RA2, respectively. For each index and period, we carried out a principal component analysis including all available questions, and constructed an index using the first factor. Columns (2) and (4) indicate the loadings with which each of these questions enter the baseline or RA2 indices, respectively. Questions containing “You (most people in your community)” were randomly assigned to households with either the “You” or the “Most people in your community” formulations. There were no significant differences in the answers to these questions, on average, according to their phrasing so for this purpose, we include both types of questions indistinctly. *Source:* BL and RA2 household surveys.

The first index covers health and non-health private benefits accruing to individuals from the ownership and usage of private toilets. This index captures a household's understanding of the private costs and benefits from toilet ownership, and the dangers (to themselves only) involved in performing open defecation. The second index, termed emotional benefits, captures subjective benefits in terms of social status and pride resulting from toilet ownership and embarrassment associated with open defecation. This second index reflects the following components of CLTS: CLTS puts a strong emphasis on associating toilet ownership with a sense of pride and accomplishment, framing private sanitation as an aspirational good. Equally, the intervention seeks to associate open defecation with strong feelings of shame and disgust. In describing an activity during which attendees show CLTS implementers around the village, identifying areas where community members regularly perform open defecation, Kar [2003] states that *[t]he initial embarrassment experienced by the community during the "walk of shame" gave way to a strong desire to stop open defecation and to get rid of these areas.* Our third index captures these aspects of "emotional" benefits from toilet ownership, and also the expected status gain from toilet ownership. The third index includes questions that refer directly to negative externalities, such as the negative externalities in terms of health and soil pollution that open defecation by neighbours causes onto oneself.

Table 4.7: Correlations between Expected Costs, Benefits and Toilet Ownership at Baseline

LHS:	Cons./Finished		Functioning		Improved		OD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Private Benefit Index	0.01	0.01	0.01	0.01	0.01*	0.01	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Emotional Benefit Index	-0.02**	-0.01	-0.02***	-0.01	-0.02***	-0.01	0.02**	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Externality Index - High	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
ln(Expected construction costs)		-0.05***		-0.06***		-0.04**		0.05***
		(0.02)		(0.02)		(0.02)		(0.02)
Mean Dep.var. (BL)	0.37	0.41	0.36	0.39	0.33	0.36	0.61	0.58
HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of HHs	4,542	1,955	4,542	1,955	4,542	1,955	4,530	1,949

Note: Questions included in each of the three expected benefit indices. Columns (1) and (3) indicate whether each question was included at Baseline or RA2, respectively. For each index and period, we carried out a principal component analysis including all available questions, and constructed and index using the first factor. Columns (2) and (4) indicate the loadings with which each of these questions enter the baseline or RA2 indices, respectively. Questions containing “You (most people in your community)” were randomly assigned to households with either the “You” or the “Most people in your community” formulations. There were no significant differences in the answers to these questions, on average, according to their phrasing so for this purpose, we include both types of questions indistinctly. *Source:* BL and RA2 household surveys.

Expected benefits from sanitation at baseline did not correlate with either higher levels of toilet ownership or lower levels of open defecation, as seen in Table 4.7. In fact, Columns (1), (3), (5) and (7) show that there is statistically significant *negative* correlation, between high levels of expected emotional benefits and all our sanitation outcomes. This finding lends support to the idea that safe sanitation is marginally more valued, in aspirational or social-status terms, by those who do not have access to it. The inclusion of expected costs in these regressions eliminates the statistical significance of this correlation, as seen in even-numbered columns. This might be driven by the restricted size of the sample, which is less than half the size due to limited answers to the expected price questions, but point estimates are comparable.

Since treatment status was assigned at random, the average values of these indices at baseline were not significantly different between CLTS and Control areas, as seen in Table C.9, in the Appendix. Moreover, Table C.10 shows that average values of all three indices were mostly balanced across the different types of households discussed in the previous section. So the first question we ask is whether CLTS, which aimed to increase knowledge and perception along all three domains, changed these expected benefits indices in significant way.[19] Table 4.8 presents the result of an ANCOVA regression of each index at RA2 on a treatment indicator and host of control variables, including the index value at baseline. Note that changes in expected benefits are only observed over the medium term, not the short term.

We find no evidence of statistically significant impacts of CLTS on private benefits (Columns (1) and (2)) or on awareness of externalities (Columns (5) and (6)) 2 years after the intervention, although point estimates for this last index are quite large. Columns (3) and (4), on the other hand, show that CLTS significantly increased the average emotional benefit index of households. The magnitude of this effect is economically meaningful: 0.15 is equivalent to 10%

[19] We recorded these answers at baseline and RA2, so our post-treatment measure is from two years after the intervention.

of a standard deviation in the expected emotional benefit index.[20]

Table 4.8: CLTS Increased the Emotional Expected Benefit Index, measured at RA2

Dep var: Index of	Private benefits		Emotional benefits		Health externalities	
	(1)	(2)	(3)	(4)	(5)	(6)
CLTS	0.05	0.05	0.15*	0.15*	0.16	0.17
	(0.08)	(0.08)	(0.08)	(0.08)	(0.11)	(0.11)
HH controls	Yes	Yes	Yes	Yes	Yes	Yes
Index at BL	No	Yes	No	Yes	No	Yes
LGA FEs	Yes	Yes	Yes	Yes	Yes	Yes
No. of TUs	242	242	242	242	242	242
No. of HHs	4,125	4,118	4,125	4,115	4,125	4,117

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01.* *Source:* RA2 Household survey.

The index of emotional benefits includes questions around the aspirational nature of sanitation investments, and the shame of performing open defecation. Other sanitation campaigns that leveraged perceived status and social norms to achieve their objectives have been shown to have positive effects on toilet construction in the past. Stopnitzky [2017], for example, shows how a campaign that associated toilet construction with social status and marriageability of young men, had large impacts on toilet adoption in India. Similarly, using a panel of poor households from Madhya Pradesh and Tamil Nadu, India, Augsburg and Rodriguez-Lesmes [2015] show that 80% of the toilet owners report to have increased their status in the community after building the toilet. The results presented in Table 4.8, are, to the best of our knowledge, the first attempt at estimating program impacts on social status considerations directly. At a first glance, our results seem to suggest that CLTS was successful at attaching a higher social value to toilet ownership.

We now turn to the question of whether this increase in expected emotional benefits from sanitation, is behind the higher toilet construction rate observed in CLTS areas. A natural channel to suggest would be that CLTS in-

[20] The index has a mean of 0 and a standard deviation of 1.54.

creased the expected benefits from sanitation of a certain group of households, who then went on to construct toilets at higher rates than in control areas. But this is certainly not the only channel through which expected benefits could be operating. There is also the possibility, for example, that expected social benefits increased evenly over the whole population, but that only households with certain characteristics, such as higher purchasing power, or young children, reacted by constructing toilets. Alzua et al. [2017], for example, suggest that the changing expectations of households who already owned toilets, allowed them to put pressure on others who did not, who then proceeded built toilets.

As a first test, we check whether CLTS had differential impacts according to baseline levels of each of these indices. Perhaps an information campaign such as the one under study is more effective when the agents have lower expectations regarding the benefits of sanitation to begin with. This effect could operate at the household level, e.g. among households with lower (or higher) expectations, or the community level. Tables C.11, C.12 and C.13, in the Appendix, show no evidence of heterogeneous treatment effects along these lines. Therefore, if expected benefits indeed play a role in mediating CLTS impacts, it is not determining the correct "initial conditions" in which CLTS is successful. Instead, the CLTS might be affecting toilet construction via its ability to *change* these expectations.

Next, we study what the impact of CLTS was, on each expected benefits index, according to household type. If CLTS increased expected benefits for the same groups of households for which it also increased toilet construction, then this would be informative about the channel through which CLTS affects households' investment decisions. We again take the indices of expected benefits at the household level as dependent variables, and regress them on a treatment indicator equal to one if the household is assigned to CLTS and complies with each of our four sub-samples of households. A second indicator variable will be equal to one if the household is assigned to CLTS but is not part of this sub-sample, in order to capture CLTS impacts on the rest of the households. Results for this approach are shown in Table 4.9. In the Table, PBI stands for private benefit index, EBI for emotional benefit index, and Ext for Externalities index.

Table 4.9: Heterogeneous Impacts of CLTS on Emotional Benefit Index

Dep var: Index of	Children <6 y/0			Female HoH			Uneducated HoH			<median wealth		
	(1) PBI	(2) EBI	(3) Ext.	(4) PBI	(5) EBI	(6) Ext.	(7) PBI	(8) EBI	(9) Ext.	(10) PBI	(11) EBI	(12) Ext.
CLTS x Yes	-0.01	0.21*	0.25*	-0.01	0.06	0.17	0.06	0.05	0.19	0.10	0.16*	0.21*
	(0.12)	(0.12)	(0.14)	(0.11)	(0.10)	(0.12)	(0.11)	(0.11)	(0.13)	(0.10)	(0.09)	(0.12)
CLTS x No	0.08	0.13	0.14	0.09	0.21**	0.17	0.05	0.20**	0.16	0.04	0.15	0.08
	(0.09)	(0.08)	(0.11)	(0.10)	(0.09)	(0.11)	(0.09)	(0.09)	(0.11)	(0.11)	(0.10)	(0.12)
HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index at BL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test Yes=No (p-value)	0.48	0.44	0.32	0.43	0.12	1.00	0.96	0.14	0.76	0.59	0.88	0.26
No. of TUs	242	242	242	242	242	242	242	242	242	241	241	241
No. of HHs	4,118	4,115	4,117	4,118	4,115	4,117	4,118	4,115	4,117	3,692	3,690	3,692

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* RA2 Household survey.

The first three columns from Table 4.9 present CLTS impacts on each of the three expected benefit indices, according to the presence of children under the age of 6. We see that CLTS increased both emotional benefits and externalities indices in a statistically significant way, among households with children. A similar pattern is observed in Columns (10) to (12), where CLTS increased the values of the same two indices among households with below median wealth, but did not affect those of wealthier households. As we have seen in Section 4.5 above, households with children and with below median wealth reacted more strongly to the CLTS intervention than the rest of the sample. In these two cases then, CLTS appears to have increased expected benefits and promoted toilet construction among the same groups of households, supporting the idea that construction was triggered by increased expected benefits.

Results by level of education (Columns (4) to (6)) and gender (Columns (7) to (9)) of the household head, do not paint the same picture. In these two cases, CLTS only appears to have affected expected emotional benefits in a significant way, and this only for the population that showed the lowest impact in terms of toilet construction: households whose heads have completed primary school and who are male. This second finding opens up two important questions. First, why did the increase in expected benefits among these groups of households not lead to increased toilet construction? Toilet ownership was 40% and 38% among households with male heads and educated heads, respectively, so construction of new toilets could have been expected. However, CLTS was an information only campaign, and other constraints might have been binding in this case, such as affordability or imperfect credit markets.

In turn, the second question raised by Table 4.9 is, why did households with uneducated or female heads increase toilet construction if their expected benefits from sanitation remained the same? Barring other channels of impact, which as we will see in the following subsections, do not appear to play a significant role, one possible explanation is that these households reacted not to changes in their own beliefs, but in the beliefs of the rest of their community.

We have seen that the increased valuation of toilets by the families of would-be wives increased toilet construction among families of marriage-able men in India (Augsburg and Rodriguez-Lesmes [2015], Stopnitzky [2017]). By the same token, CLTS may have persuaded uneducated and female household heads to build toilets, by increasing their desirability among the rest of the community. This motivates a more nuanced interpretation of the impacts of CLTS and other community-led sanitation and health efforts, by highlighting the effective, but understudied role of social pressure.

4.6.3 Solving coordination problems

The community-driven nature of the CLTS intervention is an important aspect to consider when identifying the channels through which it works, and the constraints faced by the communities it was applied in. By coordinating the attention and efforts of the whole community (or at least of all those who attended the meetings), CLTS triggering meetings could be addressing sanitation bottlenecks at both community and household level. CLTS could, for example, coordinate the efforts of young and capable members of a village to help in the construction of toilets for widows or seniors. In this sense, CLTS could help coordinate efforts and shift villages away from detrimental but absorbing equilibria.

To gauge the relevance of this channel we asked all households who owned toilets at RA2 whether they had received any help, financially or otherwise, in its construction. We then offered several alternative sources of help, including family or neighbours, village officials, friends or family from outside the village, etc., to which respondents could answer yes or no in each case.[21] This question was not asked at RA1, which included a shorter questionnaire, but was directed at all toilet owning households at RA2. We find no significant difference in the

[21] The full list of options is the following: "Did any of the following people help you, financially or otherwise, in constructing your toilet?" a) Family or neighbours, b) Village officials, c) Villagers other than near neighbours or village officials, d) Friends/family members outside the village, e) Village leaders, f) Member of the same church, g) LGA officials / WASH unit, h) Other government officials, i) Members of a non-governmental organization/charity, and/or j) Others.

answers to these questions between CLTS and control households.

A second related channel previously discussed in the literature is that CLTS will be more effective in areas with higher levels of social capital. Cameron et al. [2015] show that program impacts from a CLTS intervention in Indonesia were higher in villages where on average, households were more likely to participate in religious, female or other kinds of groups. The authors pose that this higher degree of participation in social activities increases programme impacts by providing the structure with which households are able to exert social pressure on each other, and achieve the objectives agreed upon during the CLTS triggering meeting. Alternatively, social capital may also help disseminate information delivered by CLTS within the community.

Table 4.10: Definition of Social Capital Index

Index/Questions	BL		RA2	
	(1)	(2)	(3)	(4)
Social Capital				
How often, over the past 12 months, have you...				
... worked on a community project?	✓	0.31	✓	0.13
... donated blood?	✓	0.05	✓	0.01
... attended any public meeting to discussion of town or school affairs?	✓	0.29	✓	0.29
... attended a political meeting or rally?	✓	0.25	✓	0.28
... attended any club or organizational meeting (not for work)?	✓	0.32	✓	0.30
... had friends over to your home?	✓	0.36	✓	0.43
... been in the home of a friend of a different race/ethnicity or had them in your home?	✓	0.31	✓	0.34
... been in the home of someone of a different neighbourhood or had them in your home?	✓	0.36	✓	0.43
... been in the home of someone you consider to be a community leader or had one in your home?	✓	0.37	✓	0.29
... volunteered?	✓	0.32	✓	0.04
... attended religious services (not including weddings and funerals)?	✓	0.06	✓	0.24
... had relatives over to your home?			✓	0.31
... served as an official or served on a committee of any local club or community association?	✓	0.22		

Note: Questions included in the social capital benefit indices. Columns (1) and (3) indicate whether each question was included at Baseline or RA2, respectively. For each period, we carried out a principal component analysis including all available questions, and constructed an index using the first factor. Columns (2) and (4) indicate the loadings with which each of these questions enter the baseline or RA2 indices, respectively. *Source:* BL, RA1 and RA2 household surveys.

We measure social capital using detailed information on participation in social activities around the community, gathered in two separate rounds. As in the case of expected beliefs, this was measured both at baseline and RA2. So we create two separate indices using principal component analysis, each including the complete set of questions which were part of the questionnaire in each wave. The questions included and their respective factor loadings are presented in Table 4.10. We test whether programme impacts vary according to our index of social capital at both the cluster and household levels, and the results are presented in Table 4.11. Column (1) reproduces the baseline results from Table 4.2. Column (2) interacts the CLTS treatment indicator with the social capital index at the cluster level. We find no significant difference in programme impacts between clusters with high and low social capital, as can be seen from the statistically insignificant coefficient in the third row. Column (3) presents estimates using household-level variation in social capital instead. We do not find that CLTS impacts are enhanced nor diminished by household social capital. Finally, Column (4) includes both social capital indices. Overall, our results show that there is no statistically significant evidence of social capital affecting CLTS programme effects on average ownership of functioning toilets. For better comparison with Cameron et al. [2015], we repeat this analysis using an equally weighted index of social capital as used by the authors. Results, shown in Table C.14 in the Appendix, are similar and our conclusions unchanged.

Finally, given the participatory nature of the CLTS intervention, it is not unreasonable to think that CLTS might have had an impact on social capital at the household or village levels. The intervention brought together multiple members of the community and proposed a collective challenge (getting rid of OD) to them. This could have in turn triggered more community engagement on behalf of households. Exploiting the fact that social capital questions were asked both at baseline and RA2, we check whether CLTS modified the levels of social capital at the household or cluster levels. Table C.15 in the Appendix provides no evidence for this hypothesis. The treatment coefficient is small and standard errors large in all the specifications. Overall, we see no evidence that CLTS households or clusters exhibit higher levels of social capital at RA2.

Table 4.11: CLTS impacts by baseline levels of Social Capital

	(1)	(2)	(3)	(4)
CLTS (γ)	0.03*	0.03*	0.03*	0.03*
	(0.02)	(0.02)	(0.02)	(0.02)
Cluster SC (BL)		0.01		0.00
		(0.01)		(0.02)
Treated × Cluster SC (BL)		0.01		0.02
		(0.02)		(0.02)
Household SC (BL)			0.01	0.01
			(0.01)	(0.01)
Treated × Household SC (BL)			0.00	-0.01
			(0.01)	(0.01)
HH controls	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes
Control Mean (RA1)	0.40	0.40	0.40	0.40
No. of Triggerable Units	247	247	247	247
No. of households	4,555	4,555	4,113	4,113
No. of observations	9,110	9,110	8,226	8,226

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* RA2 Household survey.

4.6.4 Institutional sanctions

A final possible channel we explore is that CLTS activities in the community have influenced local authorities or traditional rulers to impose new sanctions on open defecation. Anecdotal evidence from the field suggested this had occurred in at least one village, where after the CLTS triggering meeting, traditional rulers imposed in-kind fines to any member of the community seen performing OD. In order to test whether this could be a possible mechanism of CLTS impacts, we again rely on the rich data collected from households at RA2.

We asked every household in our sample whether sanctions or fines existed in their village, for individuals found performing open defecation. Using this as our dependent variable, we estimated CLTS treatment effects and found no impacts. Both sanctions and fines appear to be a common institutional trait of the communities in our sample, with 40% of households declaring that they exist, but CLTS does not seem to have motivated an increase in

Table 4.12: Did CLTS bring about Sanctions or Fines for OD?

Dep var:	Sanctions	Fines
	(1)	(2)
CLTS	-0.04	-0.01
	(0.03)	(0.02)
HH controls	Yes	Yes
LGA FEs	Yes	Yes
Control Mean (Baseline)	0.40	0.40
No. of observations	4,555	4,555

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p* <0.10, ** *p* <0.05, *** *p* <0.01.* *Source:* RA2 Household survey.

their prevalence. If anything, point estimates, albeit statistically insignificant, point towards the contrary.

4.7 Conclusion

Sanitation remains an urgent concern for policy makers in the developing world. Effective policy design requires a nuanced understanding of the constraints that households face when deciding whether to carry out lumpy investments such as the construction of a private toilet. This Chapter contributes to this effort using findings from a cluster randomised experiment carried out in the Nigerian states of Ekiti and Enugu.

We have shown that CLTS, an information only campaign designed to curb open defecation levels, had positive but moderate effects. While toilet construction increased and open defecation fell in the short run, these effects become undetectable two years after the intervention. We find that CLTS was most effective among households with female heads, households whose heads did not finish primary school, and household who are asset poor, all of whom had below average toilet ownership levels at baseline. Asset poor households in particular, show lasting impacts of CLTS, with treated households being 5 pp more likely to own a functioning toilet than control households, two years after the intervention.

Increasing the effectiveness of such a policy will require lifting other binding constraints, not affected by this policy. CLTS successfully increased expected benefits of sanitation related to pride and social status but did not affect an index which includes health and other private expected benefits, nor households' awareness of sanitation externalities. Stronger messaging in these two aspects might improve program outcomes. Our findings also suggest that price expectations could be one of these constraints. Toilet ownership was lower among households with high price expectations at baseline. This group also appears to be less likely to construct toilets as a result of the intervention, although not in a statistically significant way. The provision of accurate cost information could be a valuable addition to the CLTS programme.

Finally, we discuss to alternative channels through which CLTS could have acted. We show that there is no evidence of CLTS effectiveness being affected by baseline levels of social capital, at the household or cluster level, and that the intervention did not affect social capital levels. Also, institutional sanctions or fines, to punish open defecation, are not more common in CLTS areas after the programme.

Chapter 5

General Conclusions

The three Chapters of my thesis aim at understanding how poor households make decisions, how they react to shocks, and what constraints they face in their production and investment decisions. A common theme appeared during the course of this research: households decisions are subject to multiple constraints, and they define the way these households react to shocks or public policies. Some of these constraints, such as access to credit or imperfect labour markets, are well studied and reasonably understood by the development economics field. Others, such as the degree of adaptation possible by subsistence farmers to weather shocks, or the factors limiting investment in health enhancing technologies, still provide challenges for researchers and policy makers. The accumulation of evidence, combining an increased availability of data with nuanced economic models, can contribute in the search for these answers.

In this thesis I have tried to provide some of this evidence. In the first Chapter, I documented the usefulness of remotely sensed weather data to study temperature shocks on a sample of Tanzanian farmers. I showed that while these households are strongly affected by seasons with hot days which reduce their yields of staple crops, this does not seem to translate into lower consumption per capita. There is evidence of some risk coping mechanisms being used, such as increasing the share of sweet potato, incurring in mostly informal subsistence loans or the migration of some of the household members. However, these adjustment mechanisms apply to a small share of the whole sample. While unobserved, I suggest that the build up (and depletion) of stocks is a likely candidate for consumption smoothing for the majority of the households.

In the second Chapter, my co-authors and I dig deeper into how farmers react to temperature shocks. Using a sample of Peruvian farmers from 2007 to 2015, we find that short run reactions are region-specific, and also highly determined by the level of consumption. Constrained households, living close to a subsistence level of consumption, react to negative productivity shocks by increasing the amount of land cultivated to buttress their overall output. On the contrary, less constrained households absorb the shock and see their output reduced. Interestingly, using projections from two global climate scenarios, we find that predicted yields for each of the two regions under analysis will on average, evolve in opposite directions: they will fall in the warm, dry coast, and increase in the high, wet highlands.

In the final Chapter, my co-authors and I analyze the results from a randomized controlled trial carried out in two Nigerian states aimed at increasing investment in safe sanitation. We find that even in a context of rural poverty, information only campaigns can effectively increase toilet ownership, albeit by a moderate amount. We explore the possible mechanisms by which the intervention worked, and find that while it did not change the expected health and safety benefits of sanitation, households exhibit higher levels of expected status-related benefits associated with toilet ownership. Additionally, we find that although CLTS did not provide any financial incentives or assistance, the intervention had the highest treatment effects among households with low asset wealth and low levels of education.

The recurrent theme in these Chapters is the importance of understanding the constraints under which households operate. This is a necessary step for the design of successful policy, and not always evident without the use of detailed data and economic theory.

Appendix A

Additional Results - Chapter 1

Table A.1: No lagged effects of weather conditions on yields

	(1)	(2)	(3)	(4)
Temperature:				
Average DDs	0.01	0.03	0.01	0.03
	(0.04)	(0.04)	(0.04)	(0.03)
Average DDs, t-1		-0.05		-0.06
		(0.07)		(0.06)
Average HDDs	-0.08***	-0.07**	-0.08**	-0.06**
	(0.03)	(0.03)	(0.03)	(0.03)
Average HDDs t-1		-0.04		-0.05
		(0.03)		(0.03)
Controls:				
Growing season FE's	Yes	Yes	Yes	Yes
Household FE's	Yes	Yes	Yes	Yes
Observations	12,621	12,621	12,621	12,621

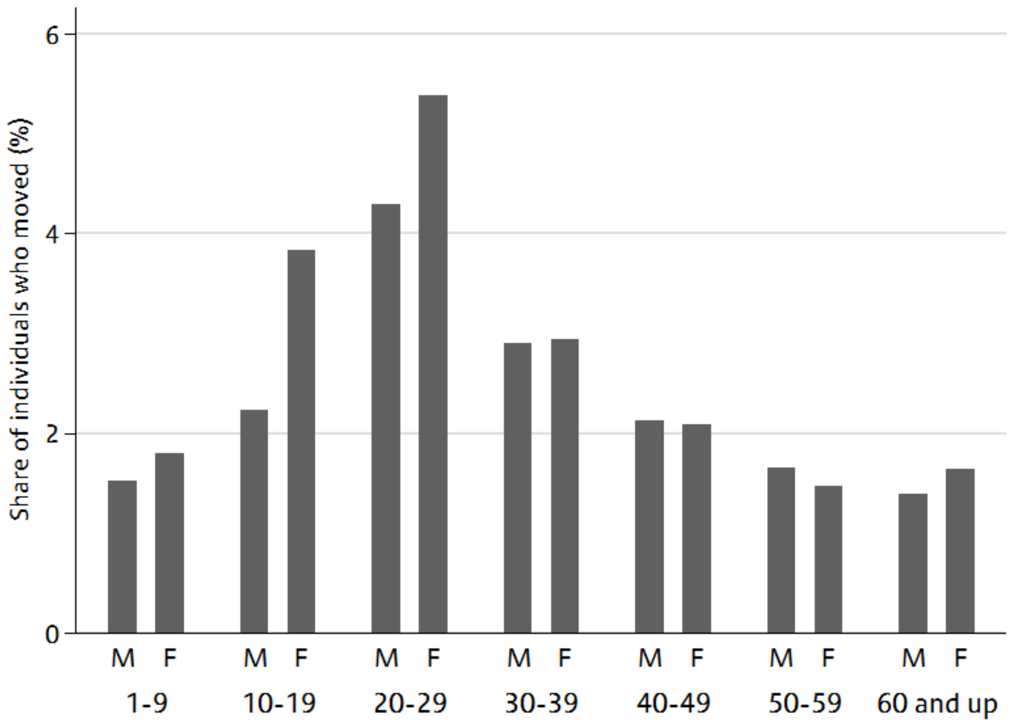
Notes: Standard errors robust to spatial and serial correlation following Conley [1999], using code from Hsiang [2010] and adapted by Thiemo Fetzer, in parenthesis. They assume a discrete cut-off for spatial correlation of errors at 500 km and six period lags. Stars indicate statistical significance: **p <0.10, ** p <0.05, *** p <0.01.*

Table A.2: Crop choice and area planted for three selected crops

Crop:	Maize		Paddy		Sweet Potato	
	(1)	(2)	(3)	(4)	(5)	(6)
Variable:	1=Yes	ln(ha)	1=Yes	ln(ha)	1=Yes	ln(ha)
Temperature:						
Average DD (t-1)	0.03**	0.01	-0.00	-0.00	-0.02**	0.11
	(0.01)	(0.03)	(0.01)	(0.10)	(0.01)	(0.10)
Average HDD (t-1)	-0.00	0.00	0.01*	0.09***	0.01	0.18***
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.06)
Controls:						
Growing season FE's	Yes	Yes	Yes	Yes	Yes	Yes
Household FE's	Yes	Yes	Yes	Yes	Yes	Yes
% HH who grow crop	66.17		22.21		9.87	
Observations	7,588	4,946	7,588	1,657	7,588	737

Notes: Standard errors robust to spatial and serial correlation following Conley [1999], using code from Hsiang [2010] and adapted by Thiemo Fetzer, in parenthesis. They assume a discrete cut-off for spatial correlation of errors at 500 km and six period lags. Stars indicate statistical significance: **p <0.10, ** p <0.05, *** p <0.01.*

Figure A.1: Average migration rates by gender and age group



Note: Individual propensities to move districts, as deduced from the migration histories detailed in the NPS questionnaire.

Appendix B

Additional Results - Chapter 2

B.1 Methodological appendix

B.1.1 Optimal temperature thresholds

Our first step in the analysis is to combine weather and agricultural output data from our sample of Peruvian farms to determine the threshold between DD and HDD. To do so, we estimate equation 3.5 varying the value of τ_{high} in 1 degree intervals from 20°C to 40°C. We record the R-square and select the threshold value the produces the best fit. We perform this analysis using the whole sample and splitting it by climatic region. Our specification uses log of output per hectare as main outcome but results are robust to using log of agricultural output, controlling for input use, or adding a richer set of fixed effects (department-by-growing season). Figure B.1 shows the results of this exercise. The best fit for the whole sample is achieved with a value of $\tau_{high} = 32$°C. This value is different by region. In the case of the coast, the best fit is achieved with a threshold of 27°C, and of 35°C in the case of the highlands.

Differences in region-specific thresholds may be due to the lower availability of heat mitigation strategies for the farmers in our sample or a host of other geography, climate and socio-economic factors. Our estimates are likely to capture not only the botanical relationship between plant growth and temperature, but also the reduction in labour productivity experienced by higher temperatures. In a meta-review of task performance studies, Seppanen et al. [2006] find that the optimal air temperature for office type work is 22℃, so our thresholds could feasibly be averaging a higher botanical threshold and lower human productivity threshold. Given the non-experimental nature of our data, we are unable to distinguish between the two. Our estimated thresholds, in any case, are the policy relevant parameters given the production function prevalent in, and mitigation strategies available to, each region.

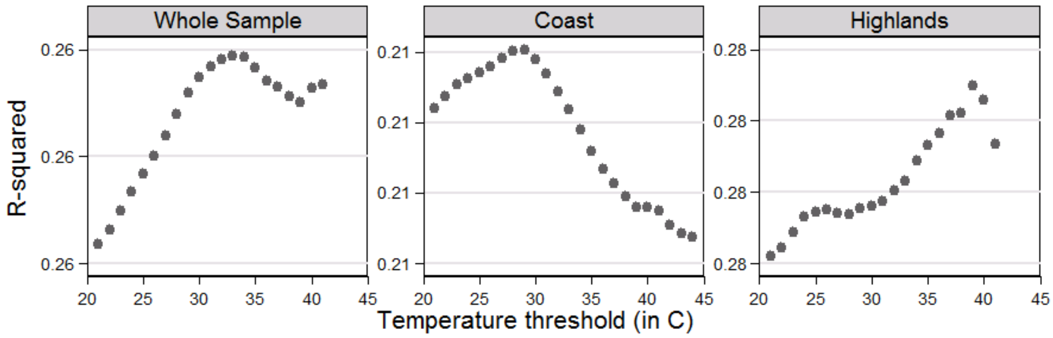
Once the optimal thresholds for each region are established it is useful to return to the distribution of temperatures over the growing season. We can observe how average daily temperatures over the growing season have a higher mean and variance in the coast than in the highlands. This in turn means that the coast experiences more average degree-days, harmful degree days, and more frequent 'hot' days (i.e. days with positive HDD) than the highlands. This differing distribution of temperatures will reflect in a different estimate of the impact of temperatures on yields, and thus must be taken into account when interpreting coefficients. Since coefficients are estimated using the average number of DDs and HDDs over the whole growing season, our estimates smooth out the real impact of temperatures on yields [Burke et al., 2015]. This will be accounted for in the interpretation of our results. At the same time, highlands experience almost four times as much rainfall as the coast so we expect to see stronger negative effects of heavy rainfall in this region.

B.2 Additional results

B.2.1 Impacts within the growing season

A closer look at the impacts within a particular growing season can potentially address important questions regarding the nature of the impacts described in the previous section. A common challenge faced by most studies focused on human-natural systems such as agriculture, is the difficulty to determine what part of the system is the one being affected. In our case, we would be interested in knowing whether the negative impact of high temperatures observed is in

Figure B.1: Model fit (R^2) of weather regressions with different temperature thresholds



Notes: Figures plot model fit (R^2) for regressions of Equation 3.5 using different values of τ_{high}, the thresholds to split between DD and HDD. Controls include household head's characteristics (age, age^2, gender and education attainment), precipitation, its square, indicators of soil quality, and district and growing season fixed effects. Each plot present the results of using a different sample.

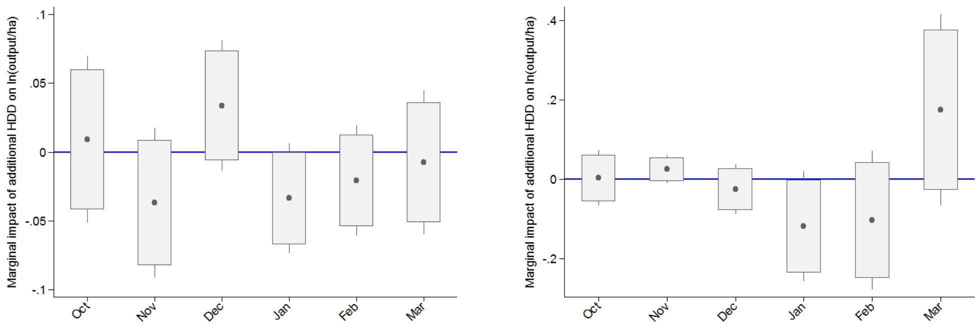
fact a negative productivity shock to human productivity, or whether it is affecting the quality of inputs or the botanical process of plant development and growth. Any of the three alternatives could be driving our results and have potentially very different consequences regarding what the appropriate policy response is.

We investigate this further by breaking down degree-days and harmful degree-days by month within a growing season. The Peruvian growing season runs from October to March, but different activities take place each month, according to region and crop grown. For example, the National Agricultural Calendar states that the sowing of crops takes place mostly between October and December, with some activities also taking place during September [MINAGRI, 2007]. By studying the months within a growing season during which temperature shocks affect farmer yields the most, we can understand what part of the agricultural cycle is most vulnerable.

Our next step is therefore to run a regression similar to our baseline specification but in which we split DDs and HDDs by month within a growing season to relax the assumption that their impact must be identical. Results are shown in Figure B.2, where we see that the strongest negative impacts

Figure B.2: Harmful degree-day impacts by month

(a) Coast **(b)** Highlands



Notes: Point estimates for the impact of additional harmful degree days by growing season month. The specification also includes degree-days by month and is otherwise identical to our baseline specifications with household and soil quality controls in columns 1 and 4 in Table 3.3). Spikes (bars) indicate 95% (90%) confidence intervals.

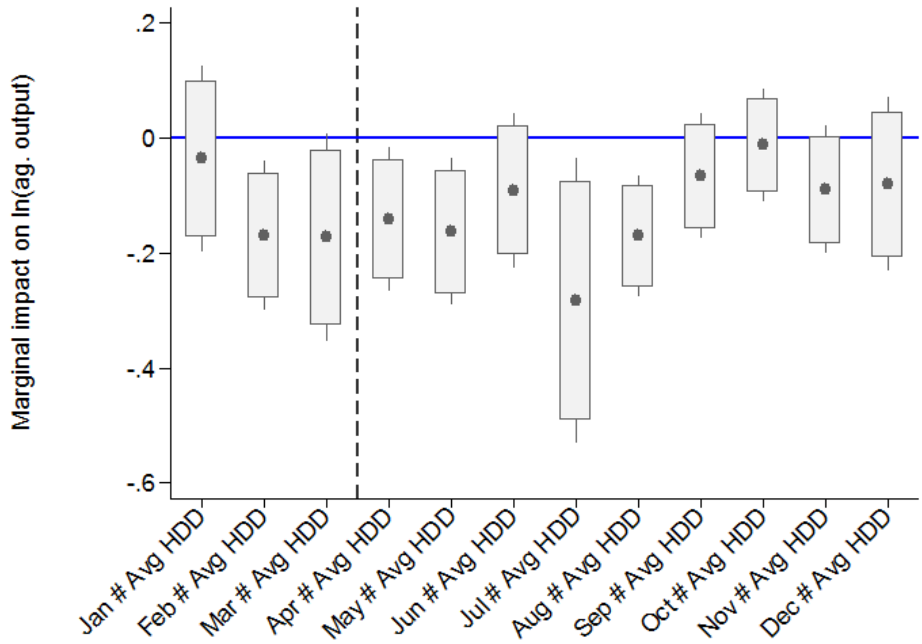
originate in the month of January for both regions. This is a month in which sowing is mostly finished and where plant growth and development is already in process. While we cannot discard impacts via reduced human productivity, this month is not one in which the hardest labour is performed which suggests that impacts are mostly botanical and/or driven by lower input quality. Nonetheless, estimation is imprecise and these results must be interpreted with caution given the large variation in agricultural practices and calendars in our sample.

B.2.2 Effects by month of interview

An argument could be made that weather impacts might not be uniform over time. For instance, one could argue that households interviewed right after the end of a growing season, could be more affected by the recent growing season's temperature than a household being interviewed six months later. This second household will have been exposed to six months of temperature and precipitation that do not enter our estimations, so their agricultural output might not be affected by weather conditions during the past growing season, or could be affected by mitigation behaviour. Similarly, households may be better able to recall agricultural production soon after a growing season. These factors could introduce measurement error and bias our estimates.

To examine the importance of this issue, we interact the degree day and harmful degree-day coefficients from Equation 3.5 with a variable indicating the month in which the interview was conducted. Figure B.3 shows the resulting coefficients for the interaction with the HDD term. Estimates suggest that a growing season's hot temperatures have the hardest impact four to five months after its end, between the months of July and August. This is the period in which all output being harvested was exposed to the hot weather season and is therefore where the low yields effectively kick in. A soft attenuation of the effects is observed thereafter, once the new growing season begins.

Figure B.3: Effect of HDD by month of interview



Notes: Point estimates for the average number of harmful degree days over the last growing season interacted with the month of interview. The specification is otherwise identical to our baseline specifications (Columns (1) and (4)) in Table 3.3). Spikes (bars) indicate 95% (90%) confidence intervals.

Appendix C

Additional Results - Chapter 3

Table C.1: Components of the Relative Wealth Index and their Factor Loadings

Asset	Loading	Asset	Loading
Bicycle	0.0174	Air conditioner	0.1048
Motorcycle/scooter/tricycle	0.124	Power generator	0.2559
Four Wheeler (Car, trucks, etc.)	0.2104	Sewing machine	0.1201
Chair(s)	0.1512	Electric iron	0.2918
Table(s)	0.1763	Pressure cooker	0.1286
Bed(s)	0.1095	Electric fans	0.2934
Cupboard(s)	0.1983	Steel and glass plates	0.1582
Other furniture	0.0766	Gold/diamond jewellery	0.1218
Refrigerator	0.2733	Other jewellery	0.1109
Washing machine	0.1569	Cow, Bullock	0.041
Micro-wave	0.1708	Calf	0.0178
Gas cooker	0.2182	Goat, Sheep	0.0731
Plasma (Flat Screen) TV	0.1943	Pig	0.0187
Other TV	0.2696	Poultry	0.0786
Satellite dish (monthly subscription)	0.2049	Fish	0.004
Other satellite dish (DSTV/ETC)	0.2134	Exotic Dogs	0.0601
Radio/CD/DVD Player	0.2145	Irrigation equipment	0.0224
Smart phones	0.1146	Other agricultural equipment	-0.0108
Other Telephone(s)	0.0904	Any other assets not listed	0.0123
Computer	0.1965		

Note: Questions included in the estimation of our index for relative wealth, measured at baseline. We carried out a principal component analysis including all available questions, and constructed and index using the first factor. Column (2) indicates the loadings with which each of these questions entered the index. *Source:* Baseline survey.

Table C.2: CLTS impacts on toilet ownership and open defecation, ANCOVA

LHS: Toilet/OD	Cons./Finished	Functioning	Improved	OD (mr)
	(1)	(2)	(3)	(4)
CLTS (γ)	0.03*	0.03*	0.03*	-0.04**
	(0.02)	(0.02)	(0.02)	(0.02)
HH head male	-0.00	0.01	0.00	-0.03**
	(0.01)	(0.01)	(0.01)	(0.01)
HH head age	0.00***	0.00***	0.00***	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)
Age squared	-0.00***	-0.00***	-0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
HH head employed	0.00	0.02	0.02	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)
HH head education: Primary school	0.05***	0.04***	0.05***	-0.02
	(0.02)	(0.02)	(0.01)	(0.02)
HH head education: Junior Secondary	0.03	0.02	0.04	0.02
	(0.02)	(0.02)	(0.02)	(0.02)
HH head education: Senior Secondary	0.09***	0.06***	0.08***	-0.06***
	(0.02)	(0.02)	(0.02)	(0.02)
HH head education: Tertiary	0.08***	0.08***	0.10***	-0.10***
	(0.02)	(0.02)	(0.02)	(0.02)
HH size	0.00	-0.00	-0.00	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
HH primary activity is farming	-0.03**	-0.03**	-0.03**	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)
Owns or is constructing toilet at BL	0.49***			
	(0.02)			
Owns a functioning toilet at BL		0.54***		
		(0.01)		
Owns an improved toilet at BL			0.49***	
			(0.01)	
Performs OD at BL				0.51***
				(0.01)
LGA FEs	Yes	Yes	Yes	Yes
Survey round FEs	Yes	Yes	Yes	Yes
Control Mean (BL)	0.38	0.36	0.33	0.61
No. of TUs	247	247	247	247
No. of HHs	4,555	4,555	4,555	4,542
No. of obs.	9,110	9,110	9,110	9,084

Notes: HH covariates: age, age squared, gender, education attainment level and employment status of the HoH; HH size and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* BL, RA1 and RA2 household surveys.

Table C.3: CLTS impacts, difference in difference and simple difference estimates

LHS: Toilet/OD	Cons./Finished		Functioning		Improved		OD (mr)	
	(1) DD	(2) SD	(3) DD	(4) SD	(5) DD	(6) SD	(7) DD	(8) SD
Panel A: Pooled estimates								
CLTS (γ)	0.03	0.03	0.03**	0.02	0.02	0.02	-0.04**	-0.04
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Panel B: Impacts by period								
CLTS x RA1 (γ_1)	0.04**	0.04	0.03*	0.03	0.02	0.02	05**	-0.05
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
CLTS x RA2 (γ_2)	0.02	0.01	0.03	0.02	0.03	0.02	-0.04	-0.04
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
HH controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey round FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean (BL)	0.38	0.45	0.36	0.40	0.33	0.353	0.61	0.5761
F-test $\gamma_1 = \gamma_2$ (p-value)	0.29	0.23	0.79	0.66	0.70	0.86	0.57	0.59
No. of TUs	247	247	247	247	247	247	247	247
No. of HHs	4,555	4,671	4,555	4,671	4,555	4,671	4,551	4,671
No. of obs.	13,665	9,342	13,665	9,342	13,665	9,342	13,652	9,342

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* Household surveys.

Table C.4: CLTS impacts among households with female decision power

LHS: Functioning toilet	Female HoH		Female DP 1		Female DP 2	
	(1) No	(2) Yes	(3) No	(4) Yes	(5) No	(6) Yes
Panel A: Pooled estimates						
CLTS x Post (γ)	0.02	0.05**	0.02	0.05**	0.03*	0.05*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Panel B: Impacts by period						
CLTS x RA1 (γ_1)	0.02	0.06**	0.02	0.05**	0.03	0.06*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
CLTS x RA2 (γ_2)	0.02	0.04	0.01	0.05**	0.03	0.04
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
Control Mean (RA1)	0.38	0.33	0.41	0.39	0.41	0.34
F-test γ Yes=No (p-value)	0.25		0.12		0.55	
F-test γ_1 Yes=No (p-value)	0.20		0.29		0.43	
F-test γ_2 Yes=No (p-value)	0.52		0.16		0.85	
No. of Triggerable Units	246	238	245	243	247	226
No. of households	2,888	1,667	2,132	2,423	3,620	935
No. of observations	5,776	3,334	4,264	4,846	7,240	1,870

Notes: Female DP 1: female-only (e.g. no adult males) or female-headed households, households in which a woman has the highest level of educational attainment, and households in which at least one adult woman is employed while no men are. Female DP 2: those in which, when asked who in the household decides about major household investments, the respondent (the eldest woman in the household), answered that it was her who decided. HH covariates: age, age squared, gender, education attainment level and employment status of the HoH; HH size and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* BL, RA1 and RA2 household surveys.

Table C.5: Pairwise correlations between household types at Baseline

Household type	Children <6 y/0	Female HoH	Uneducated HoH	<median wealth
Children <6 y/0	1.00	-	-	-
Female HoH	-0.18	1.00	-	-
Uneducated HoH	-0.20	0.37	1.00	-
<median wealth	-0.13	0.23	0.31	1.00

Note: Pairwise correlations between household types measured at baseline. *Source:* Baseline survey.

Table C.6: Heterogeneous CLTS impacts on ownership or construction of toilets

LHS: Const./Finished	Children <6 y/0		Female HoH		Uneducated HoH		<median wealth	
	(1) No	(2) Yes	(3) No	(4) Yes	(5) No	(6) Yes	(7) No	(8) Yes
Panel A: Pooled estimates								
CLTS x Post (γ)	0.03	0.04	0.03	0.04*	0.02	0.06**	0.02	0.06**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Panel B: Impacts by period								
CLTS x RA1 (γ_1)	0.05**	0.04	0.03	0.07***	0.02	0.10***	0.02	0.08***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
CLTS x RA2 (γ_2)	0.02	0.03	0.03	0.01	0.02	0.01	0.02	0.03
	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Control Mean (BL)	0.37	0.39	0.39	0.34	0.41	0.30	0.50	0.22
No. of Triggerable Units	247	238	246	238	244	240	232	245
No. of households	3,201	1,354	2,888	1,667	3,104	1,451	2,037	2,025
No. of observations	6,402	2,708	5,776	3,334	6,208	2,902	4,074	4,050

Notes: HH covariates: age, age squared, gender, education attainment level and employment status of the HoH; HH size and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* BL, RA1 and RA2 household surveys.

Table C.7: Heterogeneous CLTS impacts on ownership of improved toilets

LHS: Improved	Children <6 y/0		Female HoH		Uneducated HoH		<median wealth	
	(1) No	(2) Yes	(3) No	(4) Yes	(5) No	(6) Yes	(7) No	(8) Yes
Panel A: Pooled estimates								
CLTS x Post (γ)	0.03*	0.02	0.02	0.05**	0.01	0.05**	0.03	0.05**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Panel B: Impacts by period								
CLTS x RA1 (γ_1)	0.03	0.01	0.01	0.04*	0.01	0.04*	0.02	0.05*
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
CLTS x RA2 (γ_2)	0.04*	0.02	0.02	0.05**	0.02	0.06*	0.03	0.05*
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Control Mean (BL)	0.33	0.33	0.34	0.30	0.36	0.26	0.45	0.18
No. of Triggerable Units	247	238	246	238	244	240	232	245
No. of households	3,201	1,354	2,888	1,667	3,104	1,451	2,037	2,025
No. of observations	6,402	2,708	5,776	3,334	6,208	2,902	4,074	4,050

Notes: HH covariates: age, age squared, gender, education attainment level and employment status of the HoH; HH size and farming as the main economic activity. Errors clustered at the cluster level. *$p<0.10$, ** $p<0.05$, *** $p<0.01$. *Source:* BL, RA1 and RA2 household surveys.

Table C.8: Heterogeneous CLTS impacts on open defecation

LHS: Performs OD	Children <6 y/0		Female HoH		Uneducated HoH		<median wealth	
	(1) No	(2) Yes	(3) No	(4) Yes	(5) No	(6) Yes	(7) No	(8) Yes
Panel A: Pooled estimates								
CLTS x Post (γ)	-0.04**	-0.06**	-0.03*	-0.06**	-0.03	-0.07***	-0.02	-0.08***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Panel B: Impacts by period								
CLTS x RA1 (γ_1)	-0.04*	-0.07**	-0.04	-0.07***	-0.03	-0.09***	-0.03	-0.07**
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
CLTS x RA2 (γ_2)	-0.04*	-0.04	-0.03	-0.04	-0.03	-0.06*	-0.00	-0.08***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Control Mean (BL)	0.61	0.60	0.60	0.63	0.57	0.68	0.50	0.75
No. of Triggerable Units	247	238	246	238	244	240	232	245
No. of households	3,195	1,347	2,880	1,662	3,097	1,445	2,033	2,016
No. of observations	6,390	2,694	5,760	3,324	6,194	2,890	4,066	4,032

Notes: HH covariates: age, age squared, gender, education attainment level and employment status of the HoH; HH size and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* BL, RA1 and RA2 household surveys.

Table C.9: Balance tests for Constructed Indices, at Baseline

	Control	CLTS	P-value
Expected Costs of Construction			
Average expected cost of a VIP latrine (2015 USD)	333.84	328.91	0.70
Expected Benefit Indices			
Private Benefit Index, Household	-0.03	0.03	0.51
Private Benefit Index, Cluster	-0.03	0.03	0.51
Emotional Benefit Index, Household	0.01	-0.01	0.79
Social Benefit Index, Cluster	0.01	-0.01	0.78
Externality Index - Households	0.05	-0.05	0.33
Externality Index, Cluster	0.05	-0.05	0.33
Social capital indices			
Social Capital, Household	-0.07	0.07	0.27
Social Capital, Cluster	-0.04	0.08	0.32
Observations	2,332	2,339	

Notes: Mean expected benefit indices, by household type and measured at baseline. Indices were constructed by principal component analysis, and have a mean of 0 for the whole sample. The Table shows that at baseline, this index was mostly balanced across treatment arms. There are two exceptions: households with below median wealth appear to have slightly higher expected emotional benefit scores than wealthier households; and households whose heads had completed primary school appear to have slightly higher expected emotional and health externality benefits than households whose heads had not. For each index and definition, we performed an adjusted Wald test of equality of means comparing the average values of the index between WDP and non-WDP households, and present the p-values from that test. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* RA2 Household survey.

Table C.10: Average Expected Benefit Indices at Baseline

	Children <6 y/0		Uneducated HoH		<median wealth		Female HoH	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	No	Yes	No	Yes	No	Yes	No	Yes
Expected benefit indices (Baseline)								
Private benefits	-0.01	0.01	-0.01	0.03	0.00	0.09	0.00	-0.00
F-test Yes=No (p-value)	0.69		0.59		0.23		0.92	
Emotional benefits	-0.00	0.00	0.00	-0.01	-0.03	0.17	0.02	-0.03
F-test Yes=No (p-value)	0.98		0.90		0.01		0.29	
Health externalities	0.03	-0.07	0.03	-0.05	0.07	0.03	0.02	-0.03
F-test Yes=No (p-value)	0.05		0.10		0.45		0.24	
Treatment status								
Assigned to CLTS	0.50	0.50	0.50	0.50	0.49	0.51	0.50	0.51
F-test Yes=No (p-value)	0.83		0.78		0.47		0.39	
No. of households	3,254	1,362	3,139	1,477	2,075	2,079	2,925	1,696

Notes: Mean expected benefit indices, by household type and measured at baseline. Indices were constructed by principal component analysis, and have a mean of 0 for the whole sample. The Table shows that at baseline, this index was mostly balanced by household type. There are two exceptions: households with below median wealth appear to have slightly higher expected emotional benefit scores than wealthier households; and households whose heads had completed primary school appear to have slightly higher expected emotional and health externality benefits than households whose heads had not. For each index and definition, we performed an adjusted Wald test of equality of means comparing the average values of the index between WDP and non-WDP households, and present the p-values from that test. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01.*
Source: RA2 Household survey.

Table C.11: CLTS impacts by baseline levels of Private Benefit Index

	(1)	(2)	(3)	(4)
CLTS (γ)	0.03*	0.03*	0.03*	0.03*
	(0.02)	(0.02)	(0.02)	(0.02)
Private Benefit, Household (BL)		-0.00		-0.00
		(0.00)		(0.00)
Treated × Private Benefit Index, Household (BL)		0.00		0.01
		(0.01)		(0.01)
Private Benefit, Cluster (BL)			0.00	0.00
			(0.01)	(0.02)
Treated × Private Benefit Index, Cluster (BL)			-0.02	-0.03
			(0.02)	(0.02)
HH controls	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes
Control Mean (BL)	0.36	0.36	0.36	0.36
No. of Triggerable Units	247	247	247	247
No. of households	4,555	4,547	4,555	4,547
No. of observations	9,110	9,094	9,110	9,094

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01.* *Source:* Baseline, RA1 and RA2 Household surveys.

Table C.12: CLTS impacts by baseline levels of Emotional Benefit Index

	(1)	(2)	(3)	(4)
CLTS (γ)	0.03*	0.03*	0.03*	0.03*
	(0.02)	(0.02)	(0.02)	(0.02)
Social Benefit, Household (BL)		-0.01		-0.01
		(0.01)		(0.01)
Treated × Social Benefit Index, Household (BL)		0.00		0.01
		(0.01)		(0.01)
Social Benefit, Cluster (BL)			-0.00	-0.00
			(0.02)	(0.02)
Treated × Emotional Benefit Index, Cluster (BL)			-0.01	-0.01
			(0.02)	(0.03)
HH controls	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes
Control Mean (BL)	0.36	0.36	0.36	0.36
No. of Triggerable Units	247	247	247	247
No. of households	4,555	4,543	4,555	4,543
No. of observations	9,110	9,086	9,110	9,086

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01.* *Source:* Baseline, RA1 and RA2 Household surveys.

Table C.13: CLTS impacts by baseline levels of Externality Index

	(1)	(2)	(3)	(4)
CLTS (γ)	0.03*	0.03*	0.03*	0.03*
	(0.02)	(0.02)	(0.02)	(0.02)
Externalities, Household (BL)		-0.00		0.01
		(0.01)		(0.01)
Treated × Externality Index, Household (BL)		0.01		0.00
		(0.01)		(0.01)
Externalities, Cluster (BL)			-0.01	-0.02
			(0.01)	(0.02)
Treated × Externality Index, Cluster (BL)			0.02	0.02
			(0.02)	(0.02)
HH controls	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes
Control Mean (BL)	0.36	0.36	0.36	0.36
No. of Triggerable Units	247	247	247	247
No. of households	4,555	4,546	4,555	4,546
No. of observations	9,110	9,092	9,110	9,092

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* Baseline, RA1 and RA2 Household surveys.

Table C.14: CLTS Impacts by baseline levels of Social Capital - Equally Weighted Index

	(1)	(2)	(3)	(4)
CLTS (γ)	0.03*	0.03*	0.03*	0.03*
	(0.02)	(0.02)	(0.02)	(0.02)
Cluster SC (BL, EW)		0.00		0.00
		(0.00)		(0.00)
Treated × Cluster SC (BL, EW)		0.00		0.01
		(0.01)		(0.01)
Household SC (BL, EW)			0.00	0.00
			(0.00)	(0.00)
Treated × Household SC (BL, EW)			-0.00	-0.00
			(0.00)	(0.00)
HH controls	Yes	Yes	Yes	Yes
LGA FEs	Yes	Yes	Yes	Yes
Control Mean (RA1)	0.40	0.40	0.40	0.40
No. of Triggerable Units	247	247	247	247
No. of households	4,555	4,555	4,555	4,555
No. of observations	9,110	9,110	9,110	9,110

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01. Source:* RA2 Household survey.

Table C.15: Impact of CLTS on Social Capital

Dep var:	Household SC		Cluster SC	
	(1)	(2)	(3)	(4)
CLTS	0.05	0.03	0.04	0.03
	(0.10)	(0.10)	(0.10)	(0.10)
HH controls	Yes	Yes	Yes	Yes
Household SC (BL)	No	Yes	No	No
Cluster SC (BL)	No	No	No	Yes
No. of Triggerable Units	242	242	242	242
No. of observations	3,937	3,552	4,524	4,524

Notes: HH covariates: age, age squared, gender, employment status, literacy and ethnicity of the HoH; HH size, property tenure and farming as the main economic activity. Errors clustered at the cluster level. **p <0.10, ** p <0.05, *** p <0.01.*
Source: RA2 Household survey.

Bibliography

Abrol, Y. P. and Ingram, K. T. (1996). *Global climate change and agricultural production. Direct and indirect effects of changing hydrological, pedological and plant physiological processes*, chapter Effects of Higher Day and Night Temperatures on Growth and Yields of Some Crop Plants. FAO and John Wiley & Sons Ltd.

Akpalu, W., Hassan, R. M., et al. (2015). How can African agriculture adapt to climate change: Climate variability and maize yield in South Africa.

Alzua, M. L., Djebbari, H., and Pickering, A. J. (2017). A collective approach promotes private supply of sanitation.

Auffhammer, M., Hsiang, S., Schlenker, W., and Sobel, A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, 7(2):181–198.

Auffhammer, M., Ramanathan, V., and Vincent, J. R. (2012). Climate change, the monsoon, and rice yield in India. *Climatic Change*, 111(2):411–424.

Auffhammer, M. and Schlenker, W. (2014). Empirical studies on agricultural impacts and adaptation. *Energy Economics*, 46:555–561.

Augsburg, B. and Rodriguez-Lesmes, P. (2015). Sanitation dynamics: toilet acquisition and its economic and social implications. Technical report, Institute for Fiscal Studies.

Beegle, K., De Weerdt, J., and Dercon, S. (2011). Migration and economic mobility in tanzania: Evidence from a tracking survey. *Review of Economics and Statistics*, 93(3):1010–1033.

Benjamin, D. (1992). Household composition, labor markets, and labor demand: Testing for separation in agricultural household models. *Econometrica*, 60(2):287–322.

Briceño, B., Coville, A., and Martinez, S. (2015). Promoting handwashing and sanitation: evidence from a large-scale randomized trial in rural Tanzania.

Burgess, R., Deschenes, O., Donaldson, D., and Greenstone, M. (2017). Weather, climate change and death in India.

Burke, M. and Emerick, K. (2016). Adaptation to climate change: Evidence from US agriculture. *American Economic Journal: Economic Policy*, 8(3):106–140.

Burke, M., Hsiang, S. M., and Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*.

Cameron, L. A., Olivia, S., and Shah, M. (2015). Initial conditions matter: social capital and participatory development. *Available at SSRN 2704614*.

Cameron, L. A., Shah, M., and Olivia, S. (2013). Impact evaluation of a large-scale rural sanitation project in Indonesia.

Carleton, T. A. and Hsiang, S. M. (2016). Social and economic impacts of climate. *Science*, 353(6304).

Cesaraccio, C., Spano, D., Duce, P., and Snyder, R. L. (2001). An improved model for determining degree-day values from daily temperature data. *International Journal of Biometeorology*, 45(4):161–169.

Chaudhuri, S. and Paxson, C. (2002). Smoothing consumption under income seasonality: buffer stocks vs. credit markets. *Discussion Paper, Columbia University*, 4:E3.

Clasen, T., Boisson, S., Routray, P., Torondel, B., Bell, M., Cumming, O., Ensink, J., Freeman, M., Jenkins, M., Odagiri, M., Ray, S., Sinha, A., Suar, M., and Schmidt, W.-P. (2014). Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *The Lancet Global Health*, 2(11):645–653.

Cohen, J. and Dupas, P. (2010). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics*, 125(1):01–45.

Coll, C., Caselles, V., Galve, J. M., Valor, E., Niclos, R., Sanchez, J. M., and Rivas, R. (2005). Ground measurements for the validation of land surface temperatures derived from AASTR and MODIS data. *Remote Sensing of Environment*, 97(3):288–300.

Coll, C., Wan, Z., and Galve, J. M. (2009). Temperature-based and radiance-based validations of the V5 MODIS land surface temperature product. *Journal of Geophysical Research: Atmospheres*, 114(D20). D20102.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of econometrics*, 92(1):1–45.

Crocker, J., Geremew, A., Atalie, F., Yetie, M., and Bartram, J. (2016). Teachers and sanitation promotion: an assessment of Community-Led Total Sanitation in Ethiopia. *Environmental Science & Technology*, 50(12):6517–6525. PMID: 27211881.

Dell, M., Jones, B., and Olken, B. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*.

Dercon, S. (1996). Risk, crop choice, and savings: Evidence from Tanzania. *Economic Development and Cultural Change*, 44(3):485–513.

Dercon, S. (2005). Risk, insurance and poverty: A review. *Insurance against poverty*, pages 9–37.

Deschenes, O. and Greenstone, M. (2007). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review*, 97(1):354–385.

Di Falco, S., Veronesi, M., and Yesuf, M. (2011). Does adaptation to climate change provide food security? a micro-perspective from Ethiopia. *American Journal of Agricultural Economics*, 93(3):829–846.

Dupas, P. (2011). Health behavior in developing countries. *Annual Review of Economics*, 3(1):425–449.

ESRF (2014). Tanzania human development report 2014: Economic transformation for human development. Technical report, Economic and Social Research Foundation; United Nations Development Programme and Government of the United Republic of Tanzania.

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H., Verelst, L., and Wiberg, D. (2008). Global agro-ecological zones assessment for agriculture. Technical report, IIASA.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J. (2015). The climate hazards infrared precipitation with stations: a new environmental record for monitoring extremes. *Scientific Data*, 2.

Gbetibouo, G. A. (2009). *Understanding farmers' perceptions and adaptations to climate change and variability: The case of the Limpopo Basin, South Africa*, volume 849. International Food Policy Research Institute.

Gertler, P., Shah, M., Alzua, M. L., Cameron, L., Martinez, S., and Patil, S. (2015). How does health promotion work? evidence from the dirty business of eliminating open defecation. Technical report, National Bureau of Economic Research.

Goldstein, M. and Udry, C. (2008). The profits of power: Land rights and agricultural investment in ghana. *Journal of political Economy*, 116(6):981–1022.

Gosling, S. N., Dunn, R., Carrol, F., Christidis, N., Fullwood, J., Gusmao, D. d., Golding, N., Good, L., Hall, T., Kendon, L., et al. (2011). Climate: Observations, projections and impacts. *Climate: Observations, projections and impacts*.

Groger, A. and Zylberberg, Y. (2016). Internal labour migration as a shock coping strategy: Evidence from a typhoon. *American Economic Journal: Applied Economics*.

Guiteras, R., Jina, A., Mobarak, A. M., et al. (2015a). Satellites, self-reports, and submersion: Exposure to floods in Bangladesh. *American Economic Review*, 105(5):232–36.

Guiteras, R., Levinsohn, J., and Mobarak, A. M. (2015b). Encouraging sanitation investment in the developing world: A cluster-randomized trial. *Science*, 348(6237):903–906.

Hatfield, J. L. and Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, 10, Part A:4–10. USDA Research and Programs on Extreme Events.

Hidalgo, F. D., Naidu, S., Nichter, S., and Richardson, N. (2010). Economic determinants of land invasions. *The Review of Economics and Statistics*, 92(3):505–523.

Hisali, E., Birungi, P., and Buyinza, F. (2011). Adaptation to climate change in Uganda: evidence from micro level data. *Global Environmental Change*, 21(4):1245–1261.

Hoddinott, J. and Haddad, L. (1995). Does female income share influence household expenditures? Evidence from Côte d'Ivoire. *oxford Bulletin of Economics and Statistics*, 57(1):77–96.

Hsiang, S. (2016). Climate econometrics. *Annual Review of Resource Economics*, 8:43–75.

Hsiang, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of Sciences*, 107(35):15367–15372.

IDS (2011). *The CLTS Approach*. Institute of Development Studies.

IPCC (2014). 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. Technical report, IPCC, Geneva, Switzerland. Core writing team, R.K. Pachauri and L.A. Meyer (eds.).

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*.

Kar, K. (2003). Subsidy or self-respect? participatory total community sanitation in Bangladesh. *IDS Working Paper 184*.

Kleemans, M. and Magruder, J. (2017). Labour market responses to immigration: Evidence from internal migration driven by weather shocks. *The Economic Journal*.

Lobell, D. B., Bänziger, M., Magorokosho, C., and Vivek, B. (2011). Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature Climate Change*.

Lusk, J. L., Tack, J., and Hendricks, N. P. (2017). Heterogeneous yield impacts from adoption of genetically engineered corn and the importance of controlling for weather.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2):210–221.

McKenzie, D. J. (2005). Measuring inequality with asset indicators. *Journal of Population Economics*, 18(2):229–260.

Miguel, E., Satyanath, S., and Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy*, 112(4):725–753.

Miller, G. and Mobarak, A. M. (2013). Gender differences in preferences, intra-household externalities, and low demand for improved cookstoves. Technical report, National Bureau of Economic Research.

MINAGRI (2007). Calendario agricola.

Mostovoy, G., King, R., Reddy, K., and Kakani, V. (2005). Using MODIS LST data for high-resolution estimates of daily air temperature over Mississippi. In *Proceedings of the 3rd international workshop on the analysis of multi-temporal remote sensing images*, pages 16–18. IEEE Piscataway.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics*, 118(2):549–599.

Mutiibwa, D., Strachan, S., and Albright, T. (2015). Land surface temperature and surface air temperature in complex terrain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10):4762–4774.

Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., Centeno, G. S., Khush, G. S., and Cassman, K. G. (2004). Rice yields decline with higher night temperature from global warming. *Proceedings of the*

National academy of Sciences of the United States of America, 101(27):9971–9975.

Pickering, A. J., Djebbari, H., Lopez, C., Coulibaly, M., and Alzua, M. L. (2015). Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial. *The Lancet Global Health*, 3(11):701–711.

Prüss-Ustün, A., Bartram, J., Clasen, T., Colford, J. M., Cumming, O., Curtis, V., Bonjour, S., Dangour, A. D., De France, J., Fewtrell, L., et al. (2014). Burden of disease from inadequate water, sanitation and hygiene in low-and middle-income settings: a retrospective analysis of data from 145 countries. *Tropical Medicine & International Health*, 19(8):894–905.

Quisumbing, A. R. et al. (2003). *Household decisions, gender, and development: a synthesis of recent research.* International Food Policy Research Institute.

Quisumbing, A. R. and Maluccio, J. A. (2003). Resources at marriage and intrahousehold allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa. *Oxford Bulletin of Economics and Statistics*, 65(3):283–327.

Richie, J. and NeSmith, D. S. (1991). Temperature and crop development.

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

Rowhani, P., Lobell, D. B., Linderman, M., and Ramankutty, N. (2011). Climate variability and crop production in Tanzania. *Agricultural and Forest Meteorology*, 151:449–460.

Salazar-Espinoza, C., Jones, S., and Tarp, F. (2015). Weather shocks and cropland decisions in rural Mozambique. *Food Policy*, 53:9–21.

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

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

Seppanen, O., Fisk, W. J., and Lei, Q. (2006). Effect of temperature on task performance in office environment. *Lawrence Berkeley National Laboratory.*

Stopnitzky, Y. (2017). No toilet no bride? intrahousehold bargaining in male-skewed marriage markets in India. *Journal of Development Economics*, 127:269–282.

Tack, J., Lingenfelser, J., and Jagadish, S. V. K. (2017). Disaggregating sorghum yield reductions under warming scenarios exposes narrow genetic diversity in US breeding programs. *Proceedings of the National Academy of Sciences.*

Udry, C., Hoddinott, J., Alderman, H., and Haddad, L. (1995). Gender differentials in farm productivity: implications for household efficiency and agricultural policy. *Food policy*, 20(5):407–423.

Unicef and WHO (2015). *Progress on sanitation and drinking water: MDG assessment.* World Health Organization.

Wan, Z. and Li, Z. (2008). Radiance based validation of the V5 MODIS land surface temperature product. *International Journal of Remote Sensing*, 29(17-18):5373–5395.

Welch, J. R., Vincent, J. R., Auffhammer, M., Moya, P. F., Dobermann, A., and Dawe, D. (2010). 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.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.