ESSAYS ON FIRMS, CLIMATE CHANGE AND FOOD SYSTEMS TRANSFORMATION

By

Ahmed Salim Nuhu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Agricultural, Food, and Resource Economics – Doctor of Philosophy
Economics – Dual Major

2022

ABSTRACT

ESSAYS ON FIRMS, CLIMATE CHANGE AND FOOD SYSTEMS TRANSFORMATION

By

Ahmed Salim Nuhu

This dissertation broadly examines how climate change, and the rapidly transforming agrifood value chains are impacting firms, workers, and farmers in developing countries. Chapter 1 examines the effect of ignoring adaptation when estimating the short-run impacts of temperature shocks on workers and firms in developing regions. To do this, we first obtain naïve estimates of the short-run impacts of extreme temperatures (shocks) on workers' wages and firm output in Sub-Saharan Africa using the standard panel fixed effects approach. We then obtain the pure short-run effects by adjusting the naïve estimates by conditioning the effects of the temperature shocks on the historical local temperature information held by firms and workers prior to the occurrence of the temperature shocks. The difference between the naïve and the pure short-run estimates provide evidence of adaptation. We find evidence of temperature shock effects on wages and output that similar to other studies from our naïve estimates. However, the estimated effects are much higher when we condition on firms' prior knowledge of the local temperature. This finding indicates the importance of accounting for firms and workers' knowledge of local temperature patterns (when estimating the impacts of temperature shocks) and provides evidence of incomplete adaptation (of up to 50% of the original effects). Evidence of incomplete adaptation suggests the presence of barriers to adaptation that need to be addressed to prevent a locking-in of vulnerability to climate change impacts.

In a further application to the United States in Chapter 2, we find that accounting for the historical local temperature information is less relevant in the presence of more complete

adaptation that may be aided by established institutional capacity for dealing with extreme weather. Taken together, these findings reveal (1) the importance of accounting for adaptation in estimating the impacts of short-term temperature shocks in developing regions with more barriers to adaptation and (2) that policies aimed at adaptation should not ignore local institutional and environmental contexts in which adaptation occurs.

Chapter 3 examines the effects of the recent rise of numerous midstream agri-food firms and their authorized agents on smallholder soybean farmers in Zambia. Specifically, I examine the implications of non-contractual sale of soybean output to midstream firms and processors for the welfare of smallholder farmers. Using fixed effects and instrumental variables estimation techniques to address the endogeneity of the smallholder decision to sell to large-scale firms, I find significant positive crop income effects of selling to soybean large-scale firms on all smallholders. However, the observed effects only translate into higher total household incomes and poverty reduction for medium-scale smallholders (operating 5 ha- 20 ha) but not for small-scale smallholders operating less than five hectares. The positive crop income effects are mainly driven by the opportunity to sell more although small-scale smallholders also receive a price premium from selling to large buyers. These results suggest that the recent rise in purchasing activity by firms in the soybean industry in Zambia is benefiting smallholder farmers but not necessarily enough to move the smallest of these farmers out of poverty.

This dissertation is dedicated to my mother and the memory of my late father.

ACKNOWLEDGEMENTS

All praise be to Allah for the favors He has bestowed on me.

Firs, I would like to express my profound gratitude to my advisor, Professor Saweda O. Liverpool-Tasie for her selfless dedication towards my progress in graduate school, my development as a scientist and the completion of this dissertation. I am also indebted to Professors Titus Awokuse and David Tschirley for the initial opportunity and continued financial support to study at Michigan State. I would also like to thank the rest of my guidance committee, Prof. Christian Ahlin and Prof. Jeffrey Wooldridge for their invaluable inputs towards improving the theoretical and empirical components of my dissertation. During my time at Michigan State, I benefited immensely from the opportunity to work closely with and receive mentorship from several professors especially Professors Mywish Maredia, Eduardo Nakasone and Maria Porter of the University of Arizona. Special thanks again to Prof. Titus Awokuse and Prof. Pero Dagbovie for the love they showed during my lowest moment in graduate school. My time in graduate school was also made a lot easier by the community of friends I found in East Lansing namely, friends and colleagues in both AFRE and the economics departments, the AFRE soccer team, the Islamic Society of Greater Lansing, the Spartan Village pick-up soccer family, and the African community in Lansing.

Finally, I am eternally grateful to my mother, the charcoal seller whose sacrifices made it possible for me to come this far. And to all my mentors including but not limited to Dr. Sefa A. Churchill, Dr. Samuel Acheampong, the late Dr. George Adu, Daniel N. Womborobe, Solomon Manu Sarfo, Joseph Kankam, Boateng Frimpong, Ami Korkor Adenu, Owusu Afriyie Prempeh, I am forever grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: LONG-RUN ADAPTATION TO EXTREME TEMPERATURES AMONG WORKERS AND FIRMS IN SUB-SAHARAN AFRICA

1.1 Introduction

Extreme temperatures pose a substantial threat to economic activity globally, and there is general recognition that the failure to mitigate greenhouse emissions has locked-in a certain degree of future warming with which firms and workers will need to contend. A growing body of literature documents the negative impacts of extreme temperatures on labor productivity and firm outcomes.[1] The literature on the effects of extreme temperatures on firm outcomes has often exploited unanticipated weather shocks (Dell et al., 2014; Diaz & Moore, 2017; Graff Zivin & Neidell, 2014; Heal & Park, 2016; LoPalo, 2020; Seppanen et al., 2006; Somanathan et al., 2015; Traore & Foltz, 2017; Zhang et al., 2018). But the effects of even unexpected short-term fluctuations alone may not accurately reveal the short-run impact of extreme temperatures, given that climate change is inherently a long and gradual process (Bento et al., 2020; Burke & Emerick, 2016; Carter et al., 2018; Mérel & Gammans, 2018). We know that firms adjust to changes in the business environment; it is therefore at least possible that they also adjust to the dynamics of the physical environment, even if slowly. For example, a firm deciding to establish or expand a production plant based in Accra, Ghana, might consider investing in a cooling technology for its workers if future temperatures are expected to exceed average historical local temperatures. In this case, the short-run impact of temperature shocks on the firm's performance will depend on the extent to which historical information informed the firm's previous adaptive behavior (investing in cooling technology).[2] In this case, estimates of the short-run effects that do not take into account previous adaptive behavior will capture the combined effects of the weather (possibly negative)

[1] LoPalo (2020) studies the effects of temperature on the productivity of survey enumerators in developing countries. Zhang et al. (2018) examines the impact of temperature on productivity of Chinese manufacturing plants. Traore and Foltz (2017) examines the effect of temperature on firm competitiveness in Africa.

[2] Adaptation as used in this paper refers to the process by which firms adjust to extreme temperatures, in order to moderate its negative impacts. A forward-looking investor may use future projections about the local weather. In the absence of such information, the investor might use some weighted average of past historical weather data.

and the moderating effect of the cooling technology (possibly positive) on the workers' productivity and the firm's output. Failing to account for this adaptive component will lead to incorrect estimates – likely biased downwards - of the short-run impact of temperature shocks, causing policy makers to pay less attention to these impacts and ultimately, the effectiveness of mitigation policies based on such estimates.

Understanding whether firms are adapting and the extent to which adaptive behavior mitigates the short-run impacts of extreme temperatures has several implications for development policy. First, the extent to which adaptive behavior, where present, varies systematically among specific groups of firms is relevant in assessing the distributional and welfare consequences of adaptation policy interventions. Understanding firms' adaptation will assist policymakers in designing and optimally targeting firms with different adaptation interventions, based in part on the extent and effectiveness of adaptations they have already made to extreme temperatures.

Second, understanding systematic differences in adaptation might shed some light on potential behavioral barriers such as firm inability to incorporate all available weather information into decision making that hinder effective adaptation. Such understanding could also shed light on economic obstacles such as transaction costs, and market failures driven by incomplete financial and insurance markets that prevent effective adaptation. It should be noted that such barriers can only be uncovered with further data on the characteristics or behavior of these groups firms and workers which is beyond the scope of this paper. The knowledge of these barriers will help in designing policies aimed at facilitating adaptation.

Third, understanding which firms are adapting may provide guidance to policy makers in effectively designing and using regulatory tools such as taxation, subsidies and emission caps aimed at reducing emission of greenhouse gases.

To understand adaptation, the literature has often used investments in adaptive technology such as heating, ventilation and air conditioning (HVAC) and electricity generators (Heal & Park, 2016; LoPalo, 2020; Somanathan et al., 2015; Traore & Foltz, 2017; Zhang et al., 2018). However, not all investments in adaptation technology or changes in behavior are observed in data, leading to potentially biased estimates of adaptation.

In response to the identified gap, this paper investigates the short-run impacts of temperature shocks on firms and obtains a measure of the extent to which firms are adapting to extreme temperatures using data on a panel of firms in Sub-Saharan Africa. To do this, we first obtain what we call naïve estimates of the short-run impacts of extreme temperatures (shocks) on workers' wages and firm output in Sub-Saharan Africa using the standard panel fixed effects approach. We then obtain the pure short-run effects by conditioning the effects of the temperature shocks on the historical local temperature information held by firms and workers prior to the occurrence of the temperature shocks. The difference between the naïve and pure short-run estimates provide evidence of adaptation and is a measure of the extent of adaptation. In the case of perfect adaptation based on knowledge of local temperature, the naïve and pure effects (i.e., longer-term trend-adjusted short-run estimates) are both statistically equal to zero and the long-term trends are statistically significant. In the case of zero adaptation, the naïve and adjusted estimates are statistically significant and equal, and the long-term trends are not statistically significant. Finally, in the case of incomplete adaptation, the naïve estimates and pure effects are statistically different and different from zero. Our approach to measuring adaptation does not require knowledge of how firms process and use the historical local temperature information. That is, it does not require us to know whether firms invest in cooling or heating technology, adjust the production technology and operational times, or incentivize a change in the workers' wardrobe.[3] To the best of our knowledge, this is the first paper to simultaneously examine the short-run impacts of temperature shocks with the goal to measure the existence and extent of adaptation among firms in developing countries.

We focus on Sub-Saharan Africa because the region is currently experiencing significant economic transformation, rapid urbanization, and an associated rise in carbon emissions. Moreover, workers and firms in the region are highly exposed to extremely hot temperatures (particularly, along the equator) which is likely to compound the impacts of climate change. These effects are compounded by weak regulatory capacity and widespread failure of credit and insurance markets, and binding liquidity constraints which may hinder effective adaptation.

This paper adds to a growing body of work which documents the effects of extreme temperatures on workers and firms. Zhang et al. (2018) examine the effects of temperature shocks on firm output and productivity among manufacturing plants in China. They find that a temperature above 90°F reduces productivity and output that day by about 0.4% relative to a day with temperatures between 50-60°F. In a related paper, Traore and Foltz (2017) examine the impact of temperature shocks on wage and firm productivity in Cote d'Ivoire. They find that an extra day with temperature above 80°F reduces productivity by about 3.6% relative to a day with temperature between 77 and 80°F. LoPalo (2020) examines the impact of daily temperature variation on the productivity of the Demographic and Health Surveys (DHS) enumerators in developing countries. They find that on the hottest days, enumerators tend to complete 14% fewer interviews per hour and become less productive on tasks that require little or no supervision. Behrer and Park (2017)

[3] A limitation is that we are unable to tell the difference between zero adaptation and limited need for adaptation that arises due to the capacity of governance institutions to mitigate the collective impacts of extreme temperatures. Further, we are unable to examine the optimal level of adaptation.

compare the effects of temperature shocks on wages in historically hotter counties of the United States to counties in historically colder regions and find evidence of more long-run adaptation to hotter temperature shocks in historically hotter regions. While the approach of Behrer and Park (2017) provides evidence of adaptation by comparing how economic agents in different climatic regions respond to similar temperature shocks, it does not tell us how much adaptation is occurring even among firms in the same climatic region.

We therefore extend the extant literature by explicitly quantifying the extent of adaptation to extreme temperatures by firms and workers even within the same region. Our strategy also addresses the omitted variable bias found in parts of the literature, induced by the use of observable investments in adaptation technology as a measure of adaptation. Finally, our paper is related to the recent work of Bento et al. (2020) who developed a 'unifying approach' for estimating both adaptation and the effects of temperature shocks by simultaneously exploiting variation in weather and climate variables. However, our work differs from the work of Bento et al. (2020) in that while we focus on the effects of temperature shocks on firms and workers, Bento et al. (2020) examine the impacts of variations in temperature on pollutant accumulation.

Specifically, we examine the short-run effects of temperature shocks on firm output and wages (for workers) and quantify the extent of adaptation to extreme temperatures by these workers and firms. Using firm-level data from multiple waves of the World Bank Enterprise Surveys (WBES) between 2002 and 2018, we find an inverted U-shaped relationship between temperature shocks and our outcomes (i.e., wages and output) which is consistent with the literature. Both wages and output decline in response to an additional day with an extremely low or high temperature relative to a day with moderate temperature. The effect sizes are also comparable to those found in the extant literature. But our main finding is that the short-run effects from the naïve estimates are on average half the size of the pure effects. In other words, it suggests that long-run adaptation has helped to mitigate about 50% of the short-run impacts. We also find that workers and firms tend to be better adapted the longer they have known about the local temperature. In addition, we find that the short-run impacts of temperature shocks are still much higher for small and labor-intensive firms than they are for large and capital-intensive firms—suggesting that these groups of firms are more vulnerable to temperatures shocks.

Finally, we examine whether our findings hold in a high-income setting where institutional capacity (i.e., public disaster and emergency response systems, workplace heat-safety regulation etc.) and markets for managing climate risks are more firmly established. In an application to annual wage data from the United States, we do not find evidence of adaptation among workers based on historical local temperature as is the case in Sub-Saharan Africa. The finding from the application to the United States shows that adaptation is local and as such the institutional, economic and environmental context in which it happens is vital to any attempt to understand adaptation. Compared to Sub-Saharan Africa, firm/worker-specific adaptation might occur less in the United States as institutional capacity for dealing with thermal stress (e.g., emergency response systems) may have been built into governance systems leaving little room for firm/worker-specific adaptation based on knowledge of the local temperature.

Our paper makes two important contributions to the literature. First, it is the first study to explicitly quantify long-run adaptation to temperature shocks among workers and firms in a developing country.[4] While several studies have investigated the short- and long-run impacts of temperature shocks on firm performance, no study has focused on understanding the extent to which firms are adapting to the changing patterns of climate. Second, this paper greatly extends

[4] Our estimates should be interpreted with the knowledge that we do not know the upper bound on adaptation.

the geographic coverage of the current literature on firms in developing regions to the African continent. So far, the developing country literature has been mostly focused on China (Cai et al., 2018; Zhang et al., 2018), India (Somanathan et al., 2015) and Indonesia (Xie, 2018). While Traore and Foltz (2017) shed some light on the situation in Ivory Coast, our work extends to twenty-two other African countries, thus enhancing the external validity of our findings.[5] The remainder of the paper is organized as follows; section two provides the theoretical framework which underpin our empirical analysis, section three discusses the data and estimation strategy; section four presents the results while section five concludes the paper.

1.2 Theoretical framework

This section outlines a basic theoretic representation of the idea which underlies our empirical analysis. The model focuses on the effects of temperature shocks on firms and workers and outlines how the use of short-term fluctuations alone underestimates the short-run impacts of temperature shocks in the presence of adaptation. This framework builds on the work of Deryugina and Hsiang (2014). In what follows, I consider a price-taking profit-maximizing firm, which uses a single input, labor given by l.[6] Define as π , the profit of the firm which is a function of output $Y = \psi(T)f(l)$, where $\psi(T)$ is total factor productivity which is influenced by temperature T, with $\frac{d\psi}{dT} < 0$. On the other hand, f is a twice-differentiable concave production function which depends on labor input, l. The exogeneous cost of labor (i.e., wage) is given by w which means that the firm maximizes the following profit function:

$$\max_{l} \pi = \psi(T)f(l) - wl \qquad (1)$$

[5] See Figure 1.2 for the geographical scope of the sample of firms in our study
[6] This representation abstracts away from the effects of temperature shocks on capital such as heat-slowing effects on computers or other machinery. Empirically, we do not have the data to assess temperature-capital nexus.

Define as $f_l = \frac{df}{dl}, f_{ll} = \frac{d^2 f}{dldl},\ \psi_T = \frac{d\psi}{dT}$. In equilibrium, the firm chooses, l such that

$$w = \psi(T) f_l \qquad (2)$$

In words, the firm pays each worker their marginal productivity. Totally differentiating (2) with respect to T gives $\frac{dw}{dT} = \psi_T f_l + \psi f_{ll} \frac{dl}{dT} = 0$ which implies that

$$\frac{dl}{dT} = -\frac{\psi_T f_l}{\psi f_{ll}} \leq 0 \qquad (3)$$

where the inequality in (3) follows from $\psi_T, f_{ll} \leq 0,\ \psi,\ f_l \geq 0$ and $\frac{dw}{dT} = 0$. Define as $Y_T = \frac{dY}{dT}$, the effect of temperature on output in the short run. Then,

$$Y_T = \psi_T f(l) + \psi f_l \frac{dl}{dT}$$

$$Y_T = \psi_T f(l) - \psi f_l [\frac{\psi_T f_l}{\psi f_{ll}}] \leq 0 \qquad (4)$$

Where the inequality in (4) follows from $\psi_T \leq 0\ and\ f_l \geq 0$ and $\frac{dl}{dT} < 0$. Suppose after experiencing $Y_T \leq 0$, the firm can make a costly effort to mitigate the size of Y_T. Specifically, we define as $\boldsymbol{\alpha} \in [0,1]$ the set (or extent) of adaptation strategies that the firm can take to mitigate the impacts of temperature shocks. These strategies include both observed investments in technology and unobserved adjustments such as encouraging workers to change their wardrobes. With a constantly changing climate, the firm may rely on expert forecasts for future weather patterns to make such adaptive adjustments. Preferably, a perfect weather forecast over a period long enough to justify the costs involved in investing in adaptation strategies. In the absence of perfect expert forecasts however, the firm may rely on historical weather data to engage in adaptive behavior. Define as $\bar{\tau} = \frac{1}{t} \sum_{t=1}^{n} \tau_t$, the average temperature over the period $t = 1\ to\ n$ where t is a calendar year. By using historical weather data, $\bar{\tau}$, for adaptive behavior, we can redefine adaptation as $\boldsymbol{\alpha}(\bar{\tau})$. Then define, the long-run impact as $Y_{T(LR)}$

$$Y_{T(LR)} = \frac{dY}{dT} = \left(1 - \boldsymbol{\alpha}(\bar{\tau})\right) \left[\psi_T f(l) - \psi f_l [\frac{\psi_T f_l}{\psi f_{ll}}] \right] \tag{5}$$

Which reflects both the short-run impact obtained in (3) and the extent of adaptation. Not accounting for the adaptation underestimates the short-run impact by $-\boldsymbol{\alpha}(\bar{\tau}) \left[\psi_T f(l) - \psi f_l [\frac{\psi_T f_l}{\psi f_{ll}}] \right] \geq 0$. This is not a problem if adaptation based on knowledge of historical temperatures does not occur (i.e., $\boldsymbol{\alpha}(\bar{\tau}) = 0$) in which case $Y_{T(LR)} = Y_T$. This situation could occur when institutional capacity for mitigating the impacts of adverse temperatures is more firmly established regardless of the variation in long-term normals as is the case in many societies with strong governance, emergency response and disaster management systems. In such situations, incorporating long-term normals might not provide much information about worker/firm-level adaptation because the firm does not have a need to use this information as institutional capacities for dealing with thermal stress might have been built into governance systems, leaving little room for individual level adaptation. Further, under a theoretic perfect adaptation, (i.e., the extreme and unlikely case that $\boldsymbol{\alpha}(\bar{\tau}) = 1$, and $Y_{T(LR)} = 0$) the use of short-term temperature fluctuations will correctly identify the impact of any true weather shocks (i.e., unanticipated by the firm).[7] For any $\boldsymbol{\alpha} \in (0,1)$, however, the use of short-term fluctuations alone, will reveal both the combined short-run effects and long-run adaptation component.

[7] Theoretically, we can think of a case where a factory floor is perfectly air-conditioned to $70 - 80$°F and insulated such that the impacts of extreme heat are redundant, and that productivity is just a good as on a $70 - 80$°F day.

1.3 Data and Estimation Strategy

To analyze the impact of rising temperatures on firm and workers, we pool data from different sources. Firm-level data comes from the World Bank Enterprise Surveys, macroeconomic data from the World Development Indicators database, and climate data from the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA).

1.3.1 Firm Panel Data

The firm-level data comes from the World Bank Enterprise Surveys (WBES). The WBES covers about 164,000 firms in 144 countries around the world. These data are available for about 25 African countries. We focus on the countries with two or more waves of the WBES survey. This leaves a final sample of 6,442 firm-year observations, with each firm observed at least twice between the period 2001 and 2018. Figure 1.1 shows the unique number of firms in the panel sample in each country. The firms in our sample come from the West, East and Southern subregions of the African continent with Nigeria having the most and Ghana fewest firms. We extract information on output (sales), annual wage bill, total number of employees (permanent and part-time) at the end of the year, the four-digit industry classification code, the main product of the firm, whether the firm exports directly to foreign markets, and the size of the firm.[8]

We winsorize the data on total output and wages at the 5th and 95th percentile to minimize the influence of outliers. We deflate all nominal values in local currencies to 2010 levels using the local GDP deflator and then convert the real values to U.S dollars using the World Bank's Purchasing Power Parity (PPP) exchange rate for the respective fiscal years in order to make the monetary variables comparable across countries. A summary of the variables used in the empirical

[8] The WBES collects information on the main product of the firm. The product is then assigned a four-digit code from the United Nations Industrial Classification Codes (ISIC) Rev 3.1. We use the first two digits of the four-digit code to group firms into broader industrial groups for our analysis.

analysis is presented in Table 1.1. The median firm is small, with less than 20 employees, operating for about 13 years and not an exporter. However, 30% of the sample are medium-sized firms (between 20 and 100 employees) and 14% large firms (more than 100 employees).[9] The median annual wage is about \$3,491 per annum, corresponding to a median monthly wage of \$298 with significant variation across firms and countries.

1.3.2 Climate Data

We obtain weather data from the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA). The NOAA provides Global Telecommunications System (GTS) daily temperature data that is gridded using Shepard Algorithm at a $0.5° \times 0.5°$ resolution. We extract the daily total precipitation (in mm), maximum temperature for all the grids in all the study countries. We also extract daily relative humidity data for all countries from the Physical Science Laboratory (PSL) of the NOAA at a larger resolution of $2.5° \times 2.5°$ grids. Since the city in which the firm is located spans multiple grids with varying values for the weather variables, we take the following steps to generate a single value for each firm's location (i.e., city). First, we measure the distance from the centroid of each firm's region to each of the available temperature grids within the first administrative boundary within which the city is located. We then compute the weather value (i.e., temperature, rainfall or humidity) for the firm's region as the weighted average of all grids within a 200km radius of the city using the inverse of the distance to the centroid as weights.[10] This generates a value per day for the maximum

[9] The variation in the sizes of the firms provides an opportunity to examine potential heterogeneity in the adaptative behavior of firms of different sizes.
[10] The estimates are not sensitive to the choice of alternative radius—, 100km and 300km.

temperature, relative humidity and total rainfall for all 365 days in a fiscal year (366 days for leap years) in each location.[11]

1.3.3 Explanatory Variables

The key explanatory variables to estimate the impact of short-term fluctuations on firms and workers controlling for adaptation are the maximum daily temperature (shock) and the maximum temperature trend over a given period of years preceding the shock. To compute maximum daily temperature, we use a semi-parametric approach to construct flexible temperature bins in line with the literature (Deschênes & Greenstone, 2011; Zhang et al., 2018). Specifically, we group all 365 days in a fiscal year (366 days for leap years) into ten temperature bins. Each bin except the first and last temperature bins have a width of ten degrees Fahrenheit such that for each location-year pair, each bin represents the number of days in a year during which the temperature falls within the given temperature range. We combine the five lowest temperature bins: $[-\infty, 10)$, $[10, 20°\text{F})$, $[20, 30°\text{F})$, $[30, 40°\text{F})$ and $[50, 60°\text{F})$ which are mostly empty for most locations in West, East and Central Africa in order improve the precision of the estimates on these bins. This leaves us with five temperature bins for the empirical specification; $[-\infty, 60°\text{F})$, $[60, 70°\text{F})$, $[70, 80°\text{F})$ $[80, 90°\text{F})$ and $[90°\text{F}, \infty)$. This approach allows us to exploit the full distribution of the temperature that firms and workers are exposed to in a calendar year. To obtain the local historical temperature information held by firms prior to the occurrence of the temperature shocks we are interested in, we compute the average number of days in each temperature bin over the ten years preceding the firm survey. we then match the temperature bins to the firm data at the year-location level. Figure 1.2 shows the distribution of maximum temperature days to which

[11] The NOAA defines relative humidity as a dimensionless ratio, expressed in percent, of the amount of atmospheric moisture present relative to the amount that would be present if the air were saturated. Relative humidity is a function of both moisture content and temperature.

firms in the sample were exposed in the survey year benchmarked against the average number of days in each temperature bin over the previous ten years. The number of days during which the temperature exceeds 80°F are generally higher than the ten-year average. This is further confirmed by Figure 1.3 and 1.4 which show that the maximum temperature experienced by firms in our sample are exceeding historical ten-year averages. To measure the impact of the changing temperature conditions on the workers and firms, I focus on two measures: labor wages and total output. We measure wages as the average annual wage in the firm, while total output is measured as the total sales in the year under consideration.[12]

1.3.4 Empirical Specification

The goal of this paper is to measure the short-run impact of temperature shocks on firm performance and how these short-run impacts are influenced by adaptative behavior. To do this, we first obtain naïve estimates of the short-run impacts of temperature shocks on wage and firm output (not accounting for adaptation) by implementing the equation below:

$$Y_{idt} = \sum_{b=1}^{5} \beta^b T_{dt}^b + \gamma D_{dt} + \delta_t + \eta_i + \varepsilon_{idt} \tag{6}$$

In this specification, i indexes the firm, d is the location of the firm and t is the year of the survey. Y_{idt} is the outcome variable of interest namely log of average labor wage and log of output (i.e., sales) and T_{dt}^b is an index for the temperature bin b in location d at time t. We choose a moderate temperature bin, $70\ to\ 80$°F $(21\ to\ 27$°C$)$ as the omitted category to make our interpretation of the coefficients on the hottest and coldest temperature bins intuitive. D_{dt} are the time-varying observables at the location of the firm such as precipitation and relative humidity that are correlated with temperature, wages and output. η_i accounts for time invariant firm characteristics such as the location and sector of the firm, δ_t measures time-varying observables

[12] This measure has been used in the literature see (Behrer & Park, 2017; Traore & Foltz, 2017; Zhang et al., 2018).

while ε_{idt} is the idiosyncratic error term. The standard errors are clustered at the location-sector-year level to account for clusters in the population of firms induced by the sampling design that we do not directly observe in our sample (Abadie et al., 2017).[13] Under the plausible assumption that there are no time-varying unobservables that are correlated with the exogenous current temperature (shocks) and firm outcomes, β^b identifies the semi-elasticity of wages or output with respect to an additional day within temperature bin b.

However, firms may use historical temperature data to engage in adaptive behavior. Therefore, historical temperature which is naturally correlated with current temperature will also be correlated with firm performance. And the estimates on the transitory weather shocks not accounting for this adaptation may under-estimate the short-run impacts of extreme temperatures. To investigate this, we include a measure of the long-term temperature averages as an additional explanatory variable in specification (7). Specifically, we include variables, $\bar{T}_{dt}^{b}$, which measure the (moving) average number of days over the last ten years during which the temperature fell within the temperature bin, b in location d in year t as shown in equation (7). The long-term normal is lagged by one year to reflect all the temperature data available to the firms up to the year preceding the occurrence of the shock.[14]

$$Y_{idt} = \sum_{b=1}^{5} \tilde{\beta}^b T_{dt}^b + \sum_{b=1}^{5} \alpha^b \bar{T}_{dt}^b + \gamma D_{dt} + \delta_t + \eta_i + \mu_{idt} \tag{7}$$

Perfect adaptation occurs when the coefficients on $\tilde{\beta}$ (from equation 6) are equal to β (from equation 5) and both are not statistically significant as the firm or worker is completely insulated from temperature shocks. In this case, $\boldsymbol{\alpha^b}$ is statistically significant. But zero adaptation occurs when the on $\tilde{\beta}$ (from equation 7) are equal to β (from equation 6) and both are statistically

[13] The WBES uses a two-stage stratified random sampling design where firms in each year are randomly selected from a stratum of defined by location, size and sector.
[14] We examine the sensitivity of the estimates to the choice of lags in section 4.3

significant while the $\boldsymbol{\alpha^b}$ are not statistically significant as the long-term averages have not been used by the worker/firm. Under incomplete adaptation, $\widetilde{\beta} \neq \beta$ and the difference between the two provides a measure of the extent of adaptation.

1.4 Results

Figure 1.5 shows the estimates from equations (6) and (7) for the log of average wage. Our first observation is the perceptible differences in the short-run estimates from using temperature shocks alone and the estimates from the model which accounts for the long-term trend. The vertical distance in the estimates (when statistically significant) at the extreme (cold and hot) temperatures provides evidence of incomplete adaptive behavior.[15] The full set of estimates are also presented in Table 1.2. Columns 1-2 of Table 1.2 present our fixed effects estimates on log of average wage while columns 3-5 present the estimates on the log of output. In columns 1 and 3, we present the standard naïve estimates from equation (6). In columns 2 and 4 we present the adaptation-adjusted estimates from equation (7).

Our results consistently demonstrate that failing to account for firm adaptation to climate change underestimates the short-run impact of temperature shocks on firms and workers. The differences in estimates arise because by focusing on the contemporaneous effect of temperature shocks, we ignore the important role played by the ability of firms and workers to learn and adapt based on learning in the long-run which can mitigate the short-run impacts of temperature shocks. Specifically, the use of transitory shocks alone suggests that an additional day in the year with temperature higher than 90°F decreases wages by 0.4% relative to an extra day with "normal" temperature (i.e., 70-80°F) in the short-run. Our naïve estimates are similar to the estimates of

[15] At extremely low temperatures (in the African context, i.e., below 60°F), Figure 1.4 shows that uncertainity around the estimates are extremely large as reflected in the large confidence intervals. This comes from the few number of days in the temperature bin across our sample as shown in Figure 1.2

Zhang et al. (2018) who find 0.5% reduction in productivity for the marginal day with temperature above 90°F in China. Similarly, Traore and Foltz (2017) find a 0.1% reduction in average wage in Ivory Coast while Behrer and Park (2017) find a 0.2% reduction average payroll in the United States. However, when we account for adaptation, our estimates of the temperature shock are almost double the naïve estimates. More specifically, we find that an additional day with temperature above 90°F reduces wages by 0.7% in the short run as shown in Table 1.2. The naïve estimate (with transitory shocks alone) translates into a 37-dollar decrease while the pure effects (when longer-term normals are included as an additional control) is a 65-dollar decrease in annual wages for the average worker in the short-run. The differences in the two estimates provides evidence of incomplete adaptation. The gap between the naïve to the pure short-run estimates for output is much smaller for output than for wages, suggesting that output may be better adapted to temperature shocks on average, than wages. Specifically, we find that a marginal day with temperature above 90°F reduces output by 0.7% when we do not account for adaptation and by 0.8% when we account for adaptation. This is possible if firms substitute away from labor towards capital in the long run. By increasing the share of capital in production, the firm can maintain a relatively stable output whiles reducing wages which can explain our earlier finding on wages.

1.4.1 Mechanisms of Temperature Effects and Adaptation

We examine some potential mechanisms through which extreme temperatures affect firms. First, we consider some of the physiological channels through which extreme temperatures affect firms and workers which have been noted in the literature (Heal & Park, 2013; LoPalo, 2020). Extreme temperatures can impact labor productivity and thus wages by affecting a combination of total amount of hours worked and physical and cognitive performance on extreme temperature days. However, our data does not contain information on the number of hours worked and the average time spent on tasks. Therefore, we examine the differential impacts that extreme temperatures have on labor- and capital-intensive industries. If extreme temperatures affect productivity and output through its physiological impacts on labor, we will expect the impacts of the same temperature bin to be higher for labor-intensive firms than for capital-intensive firms, all else equal. We measure labor intensity as the total number of employees per output (Dewenter & Malatesta, 2001; Zhang et al., 2018).[16] We classify a firm as labor intensive if its measure of labor intensity exceeds the country's median labor intensity and as capital intensive, otherwise. Figures 1.7 and 1.8 plot the estimates from equations (6) and (7), estimated effects of temperature shocks on wages for labor- and capital-intensive firms, respectively while the complete set of estimates for both productivity and output are reported in Table 1.3. We find that losses in wages and output induced by extreme temperatures are more statistically significant and larger in magnitude for labor intensive firms compared to capital intensive firms. This suggests that the effects of temperature shocks on wages and output may be driven through its physiological impacts on labor. Next, we consider agricultural input linkages as a potential channel through which extreme temperatures may impact firm productivity. The impacts of temperature shocks on agricultural output is widely documented in the literature.[17] If temperature shocks operated through agricultural input linkages, we would expect the impacts of temperature shocks to be larger for firms whose inputs are directly sourced from agricultural output in and around that firm's location. We classify firms as agriculture-dependent using their 4-digit ISIC code which describes the main product of the firm. The firms classified as agriculture-dependent include food and beverage manufacturing, tobacco manufacturing, textiles, leather, wood and paper products manufacturing.

[16] We also use total wages as a share of total output as an alternative measure of labor intensity and find similar effects.

[17] See Dell et al. (2012), Blakeslee and Fishman (2018), Letta et al. (2018) and Zhang et al. (2017) for discussions on the effects of temperature on agricultural productivity

Non-agriculture dependent firms include all other manufacturing, construction, mining, wholesale and retail trade and services. Table 1.4 presents both the naïve and adaptation-adjusted estimates of the effects of temperature shocks on wages and output. The estimates reveal that firms in agriculture-dependent industries are better adapted to temperature shocks than non-agriculture dependent firms. Specifically, we find that an extra day with temperature above 90°F (relative to a normal temperature day of 70-80°F) does not have a statistically different impact on the productivity of a firm just because it depends directly on agricultural output for production. The results are similar in both the naïve scenario and the adaptation-adjusted estimates, providing evidence of adaptation. However, we do find statistically significant negative impacts of an extra day with temperature above 90°F (relative to a normal temperature day of 70-80°F) on output in non-agriculture-dependent firms. For the groups of firms that may not directly depend on agricultural output for production, we find that an extra day with temperature above 90°F (relative to a normal temperature day of 70-80°F) reduces wages by 0.5% although this is down from 0.8% on average as a result of adaptation. Similarly, an extra day with temperature above 90°F (relative to a normal temperature day of 70-80°F) reduces output by 0.6% (down from 0.9% as a result of adaptation). In our classification of agricultural dependence, we do not distinguish between dependence on local agricultural output. Hence, these results could arise if the firms do not depend on local agricultural output or have diversified their input sources away from their own regions to other areas where inputs can be more sustainably guaranteed. Incidentally, the firms that do not directly depend on agricultural output for production are also firms that are highly exposed to the effects of heat (e.g., construction and mining), further confirming the physiological effects through which temperature shocks impact productivity and wages.

1.4.2 Heterogeneous Adaptation by Size of Firm and Regional Location

We also examine the heterogeneous adaptation and thus short-run impacts of temperature shocks for firms by their size and location. We classify firms into two size groups namely, small and medium sized firms (i.e., SMEs), and large firms. SMEs are firms with 5-100 employees while large firms have more than 100 employees. SMEs constitute 86% of our sample. The estimates of the effects of temperature shocks on wages and output are presented in Table 1.5. We find that larger firms are better adapted to temperature shocks than SMEs. Specifically, an extra day with temperature above 90°F (relative to 70-80°F) causes wages to decline by about 0.5% in SMEs but the adaptation-adjusted short-run impact is about 0.7%, suggesting evidence of incomplete adaptation among SMEs. Similarly, an extra day with temperature above 90°F (relative to 70-80°F) causes output to decline by about 0.8% in SMEs but the adaptation-adjusted short-run impact on output is about 0.9%. On the other hand, we do not find statistically significant impact of temperature shocks on large firms across all specifications. The differential responses of both output and wages in SMEs may reflect constraints such as access to credit and insurance options which prevent effective adaptation among small and medium sized firms.

We also examine the data for regional heterogeneity in the impacts of temperature shocks and its implications for adaptation among firms and workers. As Figure 1.1 shows, firms in the west and southern Africa regions make up about 77 percent of our sample.[18] But as Figures 1.3 and 4 show, the west is historically warmer than the south. The west African region features a combination of a tropical and semi-arid climate in the Sahel which cuts across Senegal, Mali, Niger, Chad and the northern parts of Burkina Faso and Nigeria. This contrasts with the mostly

[18] The western Africa designation comprises firms in Benin, Burkina Faso, Cameroon, Chad, Ghana, Ivory Coast, Liberia, Mali, Niger, Nigeria and Senegal. The southern Africa designiation comrpises firms in Botswana, Lesotho, Malawi, South Africa, Zambia and Zimbabwe

tropical and temperate climates of the southern African region comprising Botswana, Lesotho, Malawi, South Africa, Zambia and Zimbabwe. Given the semi-arid and relatively hotter climate in the west, we would expect firms in west Africa to be better adapted to higher temperatures than they are to colder temperatures. Therefore, the impact of cold temperature shocks should be higher in western Africa relative to hot temperature shocks. Conversely, we would expect firms in southern Africa to be better adapted to cold temperature shocks than to hot temperature shocks.

Table 1.6 presents estimates of the effects of temperature shocks on labor productivity and output by region. Consistent with our hypothesis, we find that wages in firms in the colder parts of the continent (Southern Africa) tend to reduce more in response to an extra day with temperature above 90°F (relative to 70-80°F) as compared to firms in the historically warm western Africa. Specifically, an extra day with temperature above 90°F (relative to one with 70-80°F) reduces wages by about 2.5% in the south compared to 1.3% in the west. However, we do not observe similar effects on total output. On the other extreme, we find that firms in the west are less adapted to colder temperatures (below 60°F) than firms in Southern Africa. An extra day with temperature below 60°F (relative to one with 70-80°F) reduces workers' wages by as much as 42% and output by about 31% in the west. However, a similar day with temperature below 60°F (relative to one with 70-80°F) in southern Africa is only associated with about a 0.5% decline in wages and 1.3% decline in output but the effects are not statistically significant. This suggests that policies aimed at addressing adaptation need to be optimally weighted to address the impacts of temperature shocks at both extreme ends of the temperature distribution depending on the region under policy focus as well as the extent of adaptation in the given region.

1.4.3 Robustness

Our empirical strategy is based on several key assumptions. We confirm the internal validity of our estimates by examining the sensitivity of our estimates to these assumptions. First, we assume that firms use a simple arithmetic moving average of the last ten years' weather data in making adaptation investments. This is a reasonable assumption given that the median firm in our sample has formally been in existence for only 13 years. However, the standard deviation of 15 years shows that significant variations exists in the length of time for which the firm has been in existence. This approach also implicitly assumes that firms place equal weights on information from all years. However, firms may only use the most recent information (less than ten years of temperature data) when investment decisions regarding adaptation and may place bigger weights on information from the most recent years. To examine what happens when these two assumptions breakdown, we relax the assumption of equal weights placed on data from all of the ten years preceding the occurrence of the temperature shock. Specifically, we allow firms and workers to linearly assign lower weights to older temperature data and assign higher weights to more recent data. Thus, the weights are given by $\omega_t = 1/t$ for $t = 1, 2, \dots 10$ and $t = 1$ is the most recent year. We use the weighted data to compute the average number of days in each temperature bin over the ten years preceding the survey and re-estimate equation (7) using the new ten-year temperature bins. The estimates are presented in Table 1.7 and 1.8 of Appendix B. Table 1.7 presents the estimates of the effects of the temperature shocks on labor wages for the entire sample, and then for the various subgroups of interest. Table 1.8 reports the estimates for a similar analysis for total output. The estimates are comparable to the estimates obtained in Tables 1.2-1.6 of Appendix A. This suggests that our approach is insensitive to the way historical information about the local weather is processed and used by firms.

We also use a five-year moving average of the temperature bins with equal weights to examine the sensitivity of our estimates to the choice of length of the long-term normal. Moore et al. (2019) argue that shorter-term trends (i.e., 2-5 moving averages) tend to better reflect the information set that economic agent use in forming their beliefs over the current and future climate projections and thus, it is better suited for measuring adaptive behavior over the medium term. The estimates from using a 5-year moving average are presented in Table 1.9 and 1.10 of Appendix B. Table 1.9 presents the adaptation-adjusted estimates (using the five-year temperature trends) of the effects of temperature shocks on wages. We present each set of estimates for both the entire sample and the subgroup of firms discussed earlier in the section on heterogeneity. Table 1.10 reports the adaptation-adjusted estimates of the effects of temperature shocks on total output.

The estimated results follow the pattern observed in the original estimates using the ten-year trends with the exception of the adaptation-adjusted estimate of the effect of the Above 90°F bin for wages which is not statistically significant. In general, the adaptation-adjusted short-run estimates are higher than the naïve estimates and the coefficients on the temperature shocks are consistent with our initial findings.

Next, we examine the sensitivity of our estimates to the choice of lags for the long-term temperature trend variables. In all the results presented in section four, we use a one-year lag. The choice of lag reflects the amount of time the economic agents are assumed to have and process information about the local weather to inform adaptive behavior. In our analysis so far, firms and workers are assumed to have all the temperature data one year before the occurrence of the temperature shocks (with the choice of a one-year lag). Therefore, we examine the sensitivity of our estimates to the choice of lags using two, three, four, and five-year lags. We would expect that the more time agents have to respond to the changing patterns of the climate, the higher the magnitude of adaptation.[19] Therefore, the size of the coefficient on the temperature shock variables should increase with the number of lags. Table 1.11 in Appendix B presents the estimates of equation (7) for both wages and firm output using alternative lags. In columns 1 and 6, we present our benchmark specification with one-year lag. We then increase the number of lags gradually from two years to five years before the occurrence of the shock (columns 2-5 and columns 7-10). Consistent with our hypothesis, we find that the size of the coefficient on the Above 90°F temperature bin increases consistently with the number of lags for both wages and output. We find similar effects for the other extreme of the temperature distribution (i.e., 60°F temperature bin). While this finding confirms that the magnitude of our estimates are sensitive to the choice of lags, our finding is also consistent with the notion that the more time economic agents have to process information, the better that information can be used for adaptive behavior.

Next, we attempt to validate our results using an approach for measuring adaptation proposed by Bento et al. (2020) who suggests using a deviations from the long-term temperature normal (to measure changes in extreme temperature events) along with the long-term temperature normal as specified. Thus, the effect of the so-called "surprise weather days". Formally, we estimate equation (7)

$$Y_{idt} = \sum_{b=1}^{5} \tilde{\beta}^{b} \, [Temp - \bar{T}]_{dt}^{b} \; + \sum_{b=1}^{5} \alpha^{b} \bar{T}_{dt}^{b} + \; \gamma D_{dt} + \delta_t + \; \eta_i + \mu_{idt} \qquad (8)$$

In this specification, $\tilde{\beta}$ is the coefficient on the number of days in the current year, t during which the temperature in the temperature bin ($Temp$) deviated from the long-term normal $\overline{(T^{b}})$ temperature bin, b in location d. Suppose the average number of days over the period between 2001 and 2010 with temperature above 90°F is 20 in Lagos, Nigeria. That is, on average 20 days

[19] Given a constant naïve estimate, an increase in the adaptation-adjusted estimate implies higher adaptation, measured as the difference between the two estimates.

out of 365 calendar days experience temperature above 90°F. However, Lagos experiences 30 days of above 90°F in 2011. Then, the variable $[Temp - \bar{T}]_{dt}^{b}$ for Lagos will equals 10 surprise weather days for the above 90°F bin in 2011. In our original specification in equations (7), $\tilde{\beta}$ is the coefficient on the changes in temperature occuring in the current year while in this new set-up, it measures extreme weather shocks. All other notations is as used in previous specifications.

Adaptation in this specification (8) can be measured as the difference between $\tilde{\beta}^{b} - \alpha^{b}$. This generates a testable hypothesis for measuring adaptation. Under perfect adaptation to temperature shocks $\alpha^{b} = 0$ (Null effect of the long-term normal) and the magnitude of α^{b} will be equal to the magnitude of the temperature shock effect $\tilde{\beta}^{b}$ which should not be statistically different from zero. Under incomplete adaptation $\tilde{\beta}^{b} - \alpha^{b} \neq 0$ (Bento et al., 2020). The estimates from this specification are presented in Table 1.12 (for wages) and Table 1.13 (for output) in Appendix B. Each column reports the adaptation-adjusted estimates obtained from estimating equation (8). Each column also reports the difference between $\tilde{\beta}^{b}$ and α^{b} (i.e., the measure of adaptation) for the highest temperature bin (i.e., Above 90°F). As column 1 shows, the full impact of a day of temperature above 90°F shock (relative to 70-80°F shock) is associated with about 0.9% reduction in wages and 1.5% reduction in total output. This compares favorably with our original adaptation-adjusted estimates of 0.7% and 0.8% respectively. However, the difference between the shock and the long-term normal variables, which measures adaptation is about 0.5% for productivity and 9% and are statistically significant at the 6% and the 1% level respectively. These findings are consistent with our initial along various slices of the data presented in columns 2-7 of Tables 1.12 and 1.13 of Appendix B, further confirming the internal validity to our approach.

1.4.3 Placebo Tests

Finally, we conduct placebo tests to examine the likelihood of obtaining estimates similar to the adaptation-adjusted estimates obtained in section 4 under falsified scenarios. First, we use lead variables of the long-term temperature variables instead of lagged variables. The lead variables represent information that is not yet available to firms and should therefore not have similar effects on productivity and output like the effects observed in Section Four. The estimates from using the lead variables are reported in Table 1.14 of Appendix B. We report estimates for two-year leads (i.e., two years before the occurrence of the shock), three year and four-year leads. Due to unavailability of future data for certain dates (2021 onwards), we limit the long-term temperature trends to five-year averages.[20] Columns 1-3 report the adaptation-adjusted short-run estimates of the effects of temperature shocks on wages while columns 4-6 report the adaptation adjusted estimates for total output. The estimates are mostly not statistically significant and have opposite signs to the estimates obtained in Section Four and largely counterintuitive. This provides some assurance that our original estimates have not occurred by random chance.

Next, we conduct another placebo test to examine the effects of temperature shocks on a variable that should not be affected by temperature shocks. That is, the firm's age. The firm's age is constantly changing and as long as the firm is in existence, year-on-year variations in the local temperature should not affect the firm's age. Table 1.15 of Appendix B tests this hypothesis and confirms that indeed temperature shocks do not affect the firm's age. This provides further evidence that the estimates on productivity and output obtained so far is not by random occurrence.

[20] These estimates should be interpreted with caution since the analysis is not conducted on the full sample. Firms surveyed in the year 2016 and beyond are excluded from our analysis because of the unavailability of future data.

1.5 Conclusion and Policy Recommendations

This chapter examines the short-run impacts of temperature shocks on firms (output) and workers (wages) to provide evidence of adaptation to extreme temperatures among firms and workers in Sub-Saharan Africa. We achieve this by comparing naïve estimates of the impacts of temperature shocks to estimates obtained by the controlling for the information held by firms and workers on the local temperature prior to the occurrence of the temperature shocks. We demonstrate the importance of accounting for adaptation to extreme temperatures in such impact estimates for developing countries, particularly Sub-Saharan Africa. Our approach to measuring adaptation presents three main advantages. First, it does not require knowledge or data on the information that firms hold and how they use it to mitigate the impacts of temperature shocks. Thus, the approach is not dependent on knowing whether firms and their workers make behavioral changes, adjust production times, conduct factor reallocation or invest in climate-mitigation technology in order to reduce the impacts of temperature shocks.

Another advantage of this approach is that it is useful for identifying the impact of not accounting for such adaptation, which we find to be extremely important in developing regions such as Sub-Saharan Africa. In such regions (where climate shocks have been documented to have much larger impacts) such considerations is extremely critical for the design and implementation of strategies to deal with climate shocks by government (national and sub-national), development practitioners and the private sector.

We find statistically significant and economically meaningful impacts of temperature shocks on both firms' output and workers' wages in SSA. On average, an extra day with maximum temperature above 90°F (32°C) reduces workers' wages by about 0.4% relative to an extra day with normal temperature of 70-80°F. Our estimated impacts of temperature shocks is within the ballpark of estimates in the literature which range from 0.1% to 0.5% reduction in wages and productivity for temperatures above 90°F (Behrer & Park, 2017; Traore & Foltz, 2017; Zhang et al., 2018). However, we find that the short-run impacts of temperature shocks are much higher when we adjust the estimates for adaptation by conditioning the impact of temperature shocks on the firms and worker' knowledge of the local temperature. In line with the theoretical model, this provides evidence of incomplete adaptation in SSA in a context where firm specific adaptation is likely more necessary.

Second, we find significant heterogeneity in the impacts of temperature shocks and the extent of adaptation among firms. First, although we find evidence of incomplete adaptation to hot temperature shocks among firms in general, the impacts are higher for labor intensive firms relative to capital intensive firms. Similarly, the impacts of temperature shocks are more pronounced for small-and medium-sized firms (i.e., firms with 100 or less employees) relative to large size firms. These differences in adaptation by scale of firm could be explained by the challenges that small and medium-sized firms face in developing countries including but not limited to access to credit and insurance options and limited managerial know-how which may limit effective adaptation.

Finally, we find evidence of better adaptation (i.e., marginal differences between naïve and adjusted estimates) to hot temperature shocks among firms in the temperate regions of southern Africa than adaptation to colder temperatures (70°F and below) in the relatively warmer regions of western Africa. However, we do not find evidence that the impacts of temperature shocks are driven through agricultural input channels. We validate our estimates by examining the sensitivity of our estimates to potential violations of the assumptions that underlie our estimates.

While we recognize that this approach is unable to inform how exactly firms and workers are adapting (a separate but important issue), it is particularly useful where such information on adaptation is absent. It allows us to circumvent the omitted variables bias associated with using observed investments in climate-mitigation technology such as heating, ventilation and air conditioning (HVAC) for measuring adaptation (Bento et al., 2020). Another limitation of our approach is that we are unable to distinguish zero adaptation from a situation of limited need for worker/firm-level adaptation due to institutional support as well as established local-government capacity to mitigate effects of temperature shocks.

The findings in this paper yields several important policy considerations. First, the evidence of significant short-term impacts and incomplete adaptation among SMEs and labor-intensive firms reveal the existence of significant barriers to firm level adaptation (in a context where adaptation is necessary) that policy needs to address. Efforts to address adaptation should pay particular attention to these groups of firms to prevent a further locking-in of their vulnerability to climate change impacts. The absence of effects through agricultural-input linkages also suggests that the impacts on productivity and output may be driven through physiological impacts. Further research will be needed to understand the exact channels and the appropriate policies for addressing them.

Second, our findings on regional heterogeneity of the impacts and extent of adaptation suggests that a one-size-fits-all adaptation policy may not be effective. While there is growing emphasis on extremely hot temperature shocks due to climate change, the regional heterogeneity in the impacts of and adaptation to colder temperatures suggests that adaptation policy will need to be balanced in addressing the impacts of shocks at both extreme ends of the temperature distribution. In western Africa, where colder temperature shocks have substantial impacts on wages and output this will be particularly important. Third, our finding have shown that institutional contexts matter and that adaptation policy should not only focus on the economic agents impacted by extreme temperatures but also the institutions that aid and streamline adaptive efforts.

Finally, and perhaps most importantly, our paper shows that the use of transitory weather shocks alone as is standard in the extant literature may not reveal the full extent of damages caused by climate change in the short run in SSA.[21] Such estimates (that don't account for adaptation) underestimate the impacts of short-term temperature shocks which could undermine the level of urgency given to these shocks by policy makers and development practitioners. Ultimately, the use of transitory shocks alone without adjusting the estimates for adaptation, limits the extent to which such estimates can be used to inform policy in the developing country context.

[21] Bento et al. (2020) have discussed this issue in their analysis of the effects of temperature shocks on ozone concentrration. However, this approach is yet to gain traction in the literature on weather shocks and adaptation.

APPENDICES

APPENDIX A

Tables and Figures

Table 1.1 Summary statistics

	Mean	Std. Dev	Median	Min	Max
Average annual wage (PPP $)	9,367.3	29,574.3	3,491.2	154.1	681,109.5
Average monthly wage (PPP $)	780.6	2,464.5	290.9	12.8	56,759.1
Total annual sales ('000 PPP $)	7,445	27,297.7	422.3	6.2	28,7249.4
Total number of employees	58.2	145.7	15	5	1,220
Small or medium sized firm (0/1)	0.9	0.3	1	0	1
Firm exports directly	0.1	0.3	0	0	1
Firm's years since registration	17.6	15.7	13	1	83
Average max temperature(°F)	37.1	3.8	37	26.6	45.6
10-year average max temperature(°F)	29.1	3.9	29.8	18.1	37.2
Total annual rainfall (mm)	920.4	587.1	745.4	0	2,916.3
Relative humidity (%)	64.6	13.8	64.3	26.4	91.2
Number of observations					6442

Table 1.2 Fixed effects estimates of the effect of temperature on wages and firm output

	(1)	(2)	(3)	(4)
	Log of average wage		Log of output	
Above 90°F	-0.004***	-0.007***	-0.007***	-0.008***
	(0.001)	(0.002)	(0.002)	(0.002)
80-90°F	-0.004***	-0.009***	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)
60-70°F	-0.001	-0.000	-0.007**	-0.012***
	(0.002)	(0.003)	(0.003)	(0.005)
Below 60°F	-0.012**	-0.009	-0.017***	-0.019***
	(0.006)	(0.006)	(0.006)	(0.007)
Above 90°F trend		0.001		0.002
		(0.002)		(0.002)
80-90°F trend		0.006***		0.001
		(0.001)		(0.002)
60-70°F trend		-0.006		0.004
		(0.004)		(0.005)
Below 60°F trend		0.042		0.084*
		(0.040)		(0.045)
Relative humidity	0.073***	0.083***	0.082***	0.091***
	(0.018)	(0.018)	(0.023)	(0.022)
Rainfall	0.000	0.000	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Size= Large (0/1)	0.092	0.101	1.295***	1.280***
	(0.069)	(0.068)	(0.085)	(0.084)
Size= Medium (0/1)	-0.000	0.013	2.739***	2.734***
	(0.110)	(0.108)	(0.137)	(0.137)
Exporter (0/1)	0.004	-0.012	0.073	0.073
	(0.062)	(0.063)	(0.065)	(0.065)
R-squared	0.0317	0.0530	0.203	0.207
10-year temperature trend	NO	YES	NO	YES
Firm FE	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES
p-value of joint test on current bins	0.0001	0.0001	0.0001	0.0001
F-stat	4.065	6.598	55.46	39.01
N	6442	6442	6442	6442

Notes: *** p<0.01, ** p<0.05, * p<0.1 Columns 1 and 3 represent a regression of log of average wage on temperature shocks without controls for adaptation. Columns 2 and 4 include a ten-year temperature trend for each of the temperature bins as a control adaptation. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Table 1.3 Effects of temperature on wages and output by labor intensity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log of average wage				Log of total output			
	Naïve estimates		Adaptation-adjusted		Naïve estimates		Adaptation-adjusted	
	Labor intensive	Capital intensive	Labor intensive	Capital intensive	Labor intensive	Capital intensive	Labor intensive	Capital intensive
Above 90°F	-0.001	-0.001	-0.009***	-0.004**	-0.004***	-0.001	-0.016***	0.001
	(0.002)	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)
80-90°F	-0.004**	-0.001	-0.014***	-0.002	0.001	0.001	-0.009***	0.002
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
60-70°F	0.002	0.002	0.002	0.004	0.004	-0.005**	0.004	-0.011***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)
Below 60°F	-0.016*	-0.002	-0.019**	0.003	-0.036***	0.005	-0.042***	0.000
	(0.009)	(0.005)	(0.009)	(0.006)	(0.008)	(0.005)	(0.008)	(0.006)
R-squared	0.0155	0.0240	0.0563	0.0320	0.352	0.354	0.375	0.363
10-year temperature trend	NO	NO	YES	YES	NO	NO	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES	YES
F-stat	1.889	3.556	4.024	3.670	38.22	45.67	33.45	34.81
N	1603	2250	1603	2250	1603	2250	1603	2250

Notes: *** p<0.01, ** p<0.05, * p<0.1 A firm is classified as labor intensive if share of total employees in total output exceeds the median share of employees in total output in the country (Dewenter & Malatesta, 2001; Zhang et al., 2018). The naïve estimates do not account for adaptation while adaptation-adjusted estimates include long-term temperature bins to account for adaptation. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Table 1.4 Effects of temperature on wages and output by firm dependence on agricultural output

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log of average wage				Log of total output			
	Naïve estimates		Adaptation-adjusted		Naïve estimates		Adaptation-adjusted	
	Agriculture	Non-ag	Agriculture	Non-ag	Agriculture	Non-ag	Agriculture	Non-ag
Above 90°F	-0.001	-0.005***	-0.003	-0.008***	-0.006**	-0.006***	-0.007**	-0.009***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)
80-90°F	-0.004*	-0.006***	-0.009***	-0.010***	0.000	-0.003*	0.003	-0.005**
	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
60-70°F	0.001	-0.002	0.001	-0.001	0.000	-0.009**	-0.006	-0.013**
	(0.003)	(0.003)	(0.006)	(0.004)	(0.005)	(0.004)	(0.008)	(0.006)
Below 60°F	-0.007	-0.012*	-0.003	-0.011	-0.015	-0.017**	-0.023**	-0.017**
	(0.009)	(0.007)	(0.010)	(0.007)	(0.009)	(0.007)	(0.011)	(0.008)
R-squared	0.0368	0.0647	0.0589	0.0807	0.181	0.268	0.187	0.270
10-year temperature trend	NO	NO	NO	NO	NO	NO	NO	NO
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES	YES
F-stat	2.883	9.221	3.603	8.552	16.24	58.60	12.79	41.83
N	1591	4022	1591	4022	1591	4022	1591	4022

Notes: *** p<0.01, ** p<0.05, * p<0.1 Agriculture dependent firms include food and beverage manufacturing, tobacco manufacturing, textiles, leather, wood and paper products manufacturing. Non-agriculture dependent firms include all other manufacturing, construction, mining, wholesale and retail trade and services. The naïve estimates do not account for adaptation while adaptation-adjusted estimates include long-term temperature bins to account for adaptation. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Table 1.5 Heterogeneous effects of temperature on wages and output by size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log of productivity				Log of total output			
	Naïve estimates		Adaptation-adjusted		Naïve estimates		Adaptation-adjusted	
	SME	Large	SME	Large	SME	Large	SME	Large
Above 90°F	-0.005***	-0.002	-0.007***	-0.005	-0.008***	-0.003	-0.009***	-0.011
	(0.002)	(0.003)	(0.002)	(0.006)	(0.002)	(0.003)	(0.002)	(0.007)
80-90°F	-0.005***	-0.002	-0.010***	-0.002	-0.002	0.000	-0.001	-0.002
	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.004)
60-70°F	-0.002	0.007	-0.002	0.003	-0.006	-0.011	-0.011**	-0.019**
	(0.002)	(0.011)	(0.004)	(0.013)	(0.004)	(0.008)	(0.005)	(0.009)
Below 60°F	-0.013**	-0.014	-0.008	-0.019	-0.024***	-0.010	-0.027***	-0.032**
	(0.006)	(0.011)	(0.006)	(0.020)	(0.007)	(0.009)	(0.007)	(0.015)
R-squared	0.0544	0.0382	0.0792	0.0358	0.196	0.129	0.199	0.134
10-year temperature trend	NO	NO	YES	YES	NO	NO	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES	YES
F-stat	9.104	2.607	9.550	1.856	50.22	7.237	35.78	5.085
N	5256	640	5256	640	5256	640	5256	640

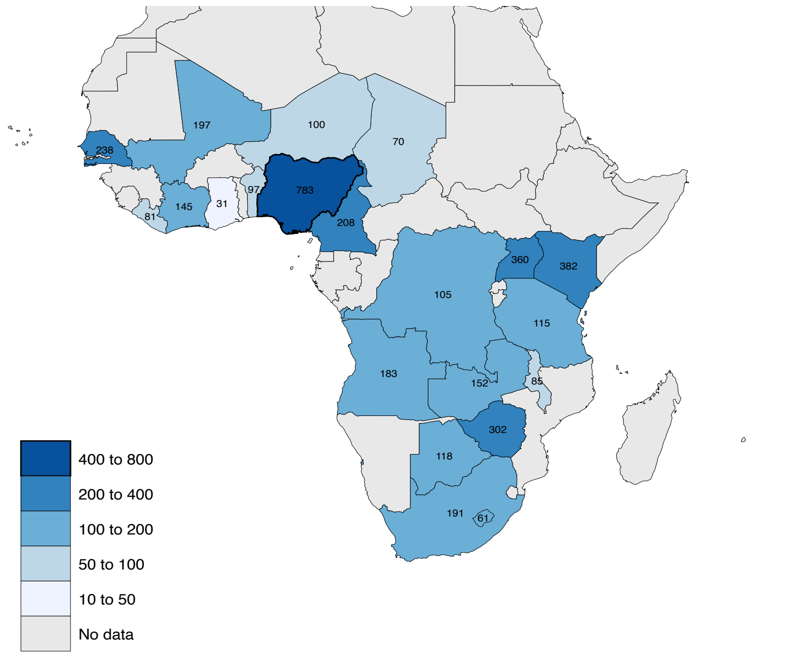
Notes: *** p<0.01, ** p<0.05, * p<0.1 SMEs are firms with 100 or less employees while large firms are those with more than 100 employees. The naïve estimates do not account for adaptation while adaptation-adjusted estimates include long-term temperature bins to account for adaptation. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Table 1.6 Heterogeneous effects of temperature on wages and output by region

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log of average wage				Log of total output			
	Naïve estimates		Adaptation-adjusted		Naïve estimates		Adaptation-adjusted	
	West	East/South	West	East/South	West	East/South	West	East/South
	-0.013***	-0.025**	-0.017***	-0.055**	-0.016***	0.016	-0.036***	-0.012
Above 90°F	(0.002)	(0.012)	(0.004)	(0.023)	(0.002)	(0.017)	(0.004)	(0.026)
	-0.009***	-0.002	-0.016***	0.004	-0.007***	-0.011*	-0.022***	0.001
80-90°F	(0.002)	(0.004)	(0.003)	(0.009)	(0.002)	(0.005)	(0.004)	(0.014)
	-0.025*	-0.020**	-0.038**	-0.034**	-0.000	-0.018	-0.003	-0.002
60-70°F	(0.015)	(0.009)	(0.017)	(0.014)	(0.020)	(0.012)	(0.021)	(0.017)
	-0.421***	-0.005	-0.469***	-0.021	-0.208**	0.007	-0.341***	-0.013
Below 60°F	(0.076)	(0.005)	(0.080)	(0.018)	(0.081)	(0.008)	(0.082)	(0.032)
R-squared	0.132	0.0116	0.136	0.0132	0.209	0.338	0.228	0.337
10-year temperature trend	YES	YES	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Time Trend	NO	NO	NO	NO	NO	NO	NO	NO
F-stat	12.25	2.771	9.017	2.292	33.35	34.29	28.09	24.02
N	3507	1622	3507	1622	3507	1622	3507	1622

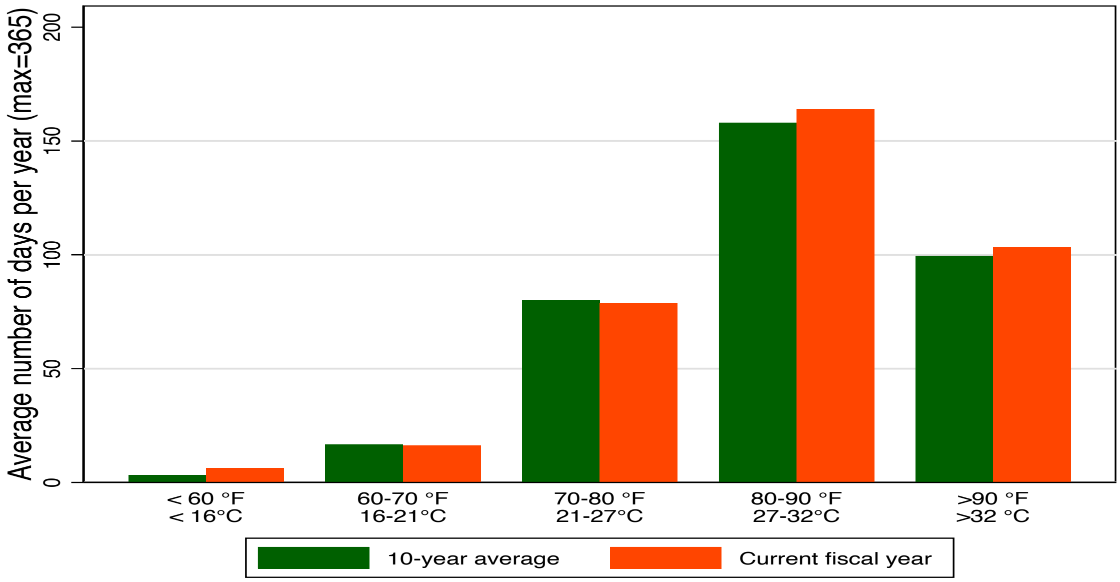
Notes: *** p<0.01, ** p<0.05, * p<0.1 The Western Africa designation comprises firms in Benin, Burkina Faso, Cameroon, Chad, Ghana, Ivory Coast, Liberia, Mali, Niger, Nigeria and Senegal. The Southern Africa designation comprises firms in Botswana, Lesotho, Malawi, South Africa, Zambia and Zimbabwe. The naïve estimates do not account for adaptation while adaptation-adjusted estimates include long-term temperature bins to account for adaptation. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Figure 1.1 Geographical location of firms in the sample



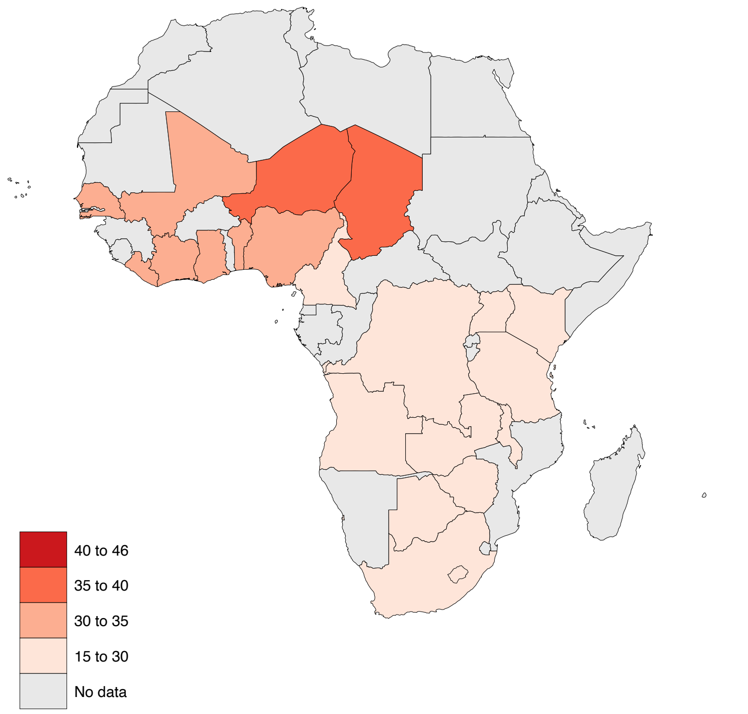
Notes: Each value represents the number of unique firms in the sample used in this study. Data source: Author's computation from the World Bank Enterprise Surveys (WBES)

Figure 1.2 Distribution of days in a year across temperature bins



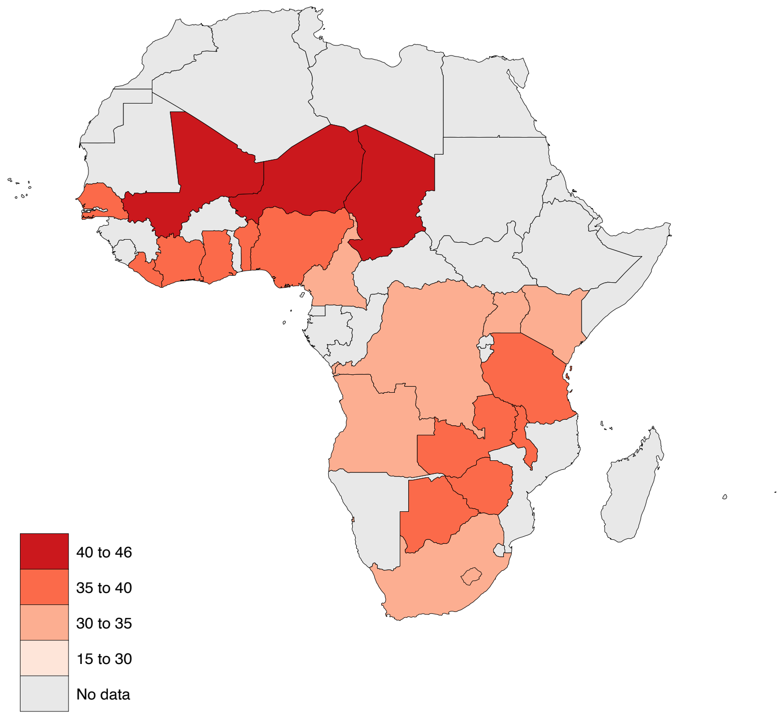
Notes: Each column represents the average number of days during which the maximum temperature falls within the specified temperature bin.

Figure 1.3 Temperature distribution in the ten years preceding the most recent firm survey year



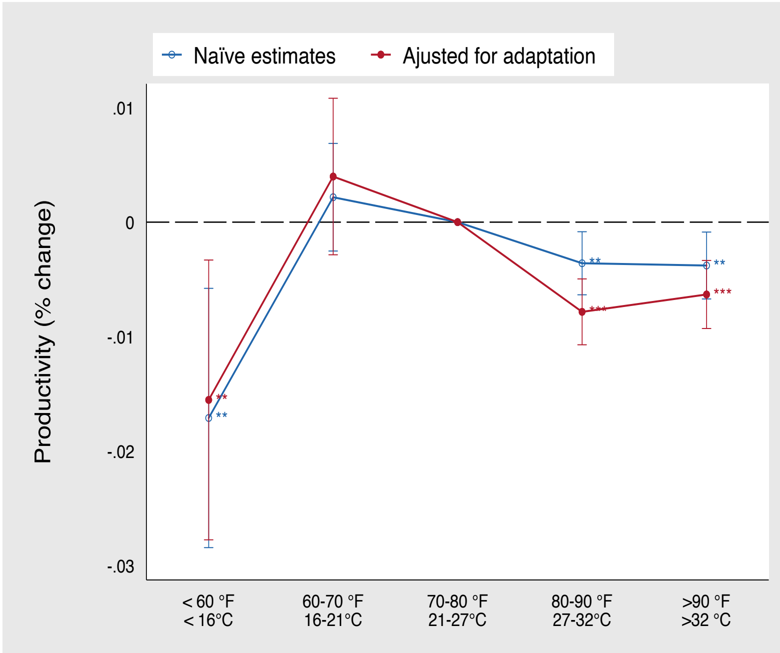
Notes: The map plots the average maximum temperature across all locations within the country of study in the ten years preceding the most recent year in which firms in those locations are observed in our data. Temperature bins are specified in degrees Celsius.

Figure 1.4 Temperature distribution in the most recent firm survey year



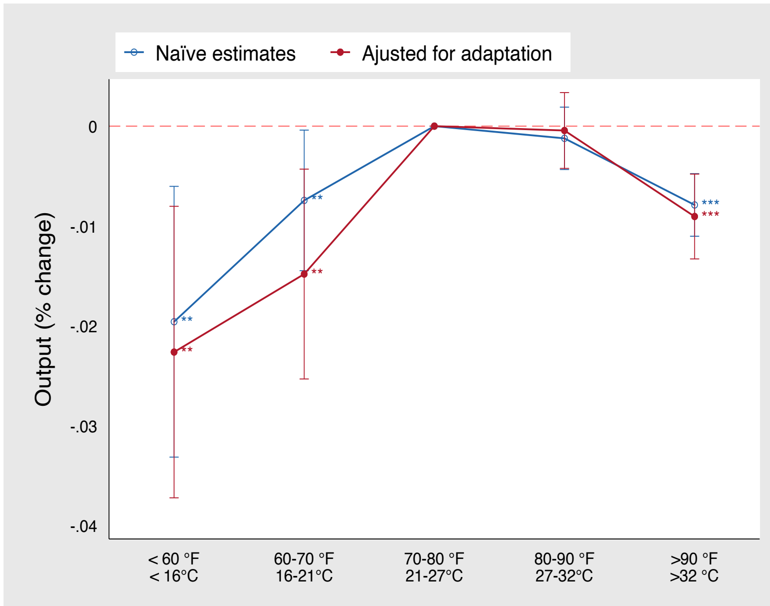
Notes: The map plots the average maximum temperature across all locations within the country of study in the most recent year in which the firms in those locations are observed in our data. Temperature bins are specified in degrees Celsius.

Figure 1.5 Effects of temperature shocks on wages



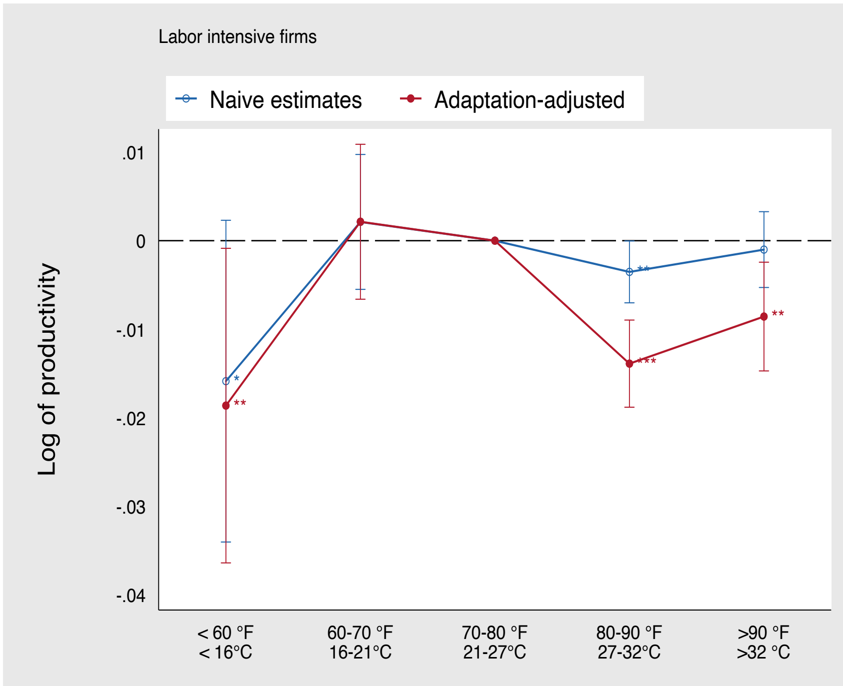
Notes: *** p<0.01, ** p<0.05, * p<0.1 Each point represents the point estimate from a fixed regression of the log of average wage (productivity) on the specified temperature bins with other weather and firm level controls. The spikes around the points are the 95% confidence intervals. The gap between the blue and red lines represent adaptation.

Figure 1.6 Effects of temperature shocks on output



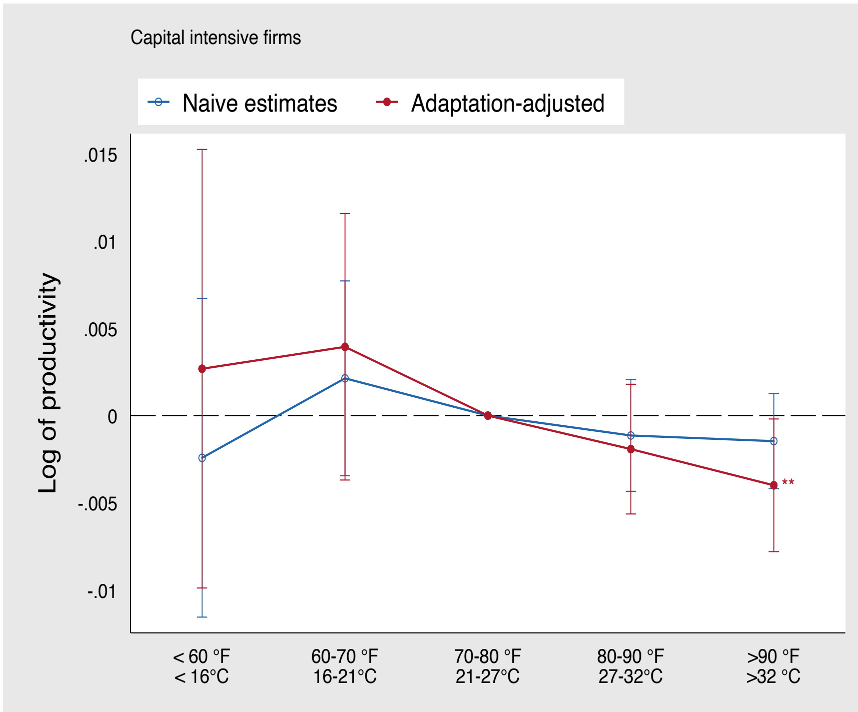
Notes: *** p<0.01, ** p<0.05, * p<0.1 Each point represents the point estimate from a fixed regression of log of total sales (output) on the specified temperature bins with other weather and firm level controls. The spikes around the points are the 95% confidence intervals. The gap between the blue and red lines represent adaptation.

Figure 1.7 Effects of temperature shocks on wages for labor-intensive firms



Notes: *** p<0.01, ** p<0.05, * p<0.1 Each point represents the point estimate from a fixed regression of the log of average wage (productivity) on the specified temperature bins with other weather and firm level controls for labor-intensive firms only. The spikes around the points are the 95% confidence intervals. The gap between the blue and red lines represent adaptation.

Figure 1.8 Effects of temperature shocks on wages for capital-intensive firms



Notes: *** p<0.01, ** p<0.05, * p<0.1 Each point represents the point estimate from a fixed regression of the log of average wage (productivity) on the specified temperature bins with other weather and firm level controls for capital-intensive firms only. The spikes around the points are the 95% confidence intervals. The gap between the blue and red lines represent adaptation.

APPENDIX B

Supplemental Tables and Figures

Table 1.7 Effects of temperature shocks on wages using weighted-temperature trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log of average wage						
	All sample	Labor intensive	SME	Agricultural	Non-ag	West Afr.	Southern Afr.
Above 90°F	-0.009***	-0.007**	-0.009***	-0.008**	-0.008**	-0.007*	-0.016
	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)	(0.012)
80-90°F	-0.009***	-0.012***	-0.011***	-0.010***	-0.010***	-0.013***	-0.007*
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
60-70°F	-0.004	0.003	-0.005	-0.004	-0.004	-0.019	-0.031**
	(0.004)	(0.005)	(0.004)	(0.007)	(0.007)	(0.015)	(0.012)
Below 60°F	-0.003	-0.014	-0.001	-0.001	-0.001	-0.414***	-0.002
	(0.005)	(0.011)	(0.006)	(0.009)	(0.009)	(0.081)	(0.007)
Above 90°F 10-year trend	0.005	0.006	0.004	0.009	0.009	-0.050***	0.041*
	(0.004)	(0.005)	(0.005)	(0.007)	(0.007)	(0.014)	(0.024)
80-90°F 10-year trend	0.013***	0.015***	0.015***	0.015***	0.015***	-0.037***	-0.045
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.011)	(0.030)
60-70°F 10-year trend	-0.016***	-0.022**	-0.014**	-0.007	-0.007	-0.289***	0.012
	(0.006)	(0.010)	(0.007)	(0.010)	(0.010)	(0.073)	(0.030)
Below 60°F 10-year trend	0.202***	0.103	0.178***	0.163*	0.163*	-4.862***	-0.272
	(0.051)	(0.071)	(0.058)	(0.093)	(0.093)	(1.493)	(0.193)
R-squared	0.0584	0.0408	0.0699	0.0387	0.0387	0.150	0.00955
Firm FE	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES
F-stat	6.188	3.627	6.763	2.317	2.317	11.30	2.171
N	6442	1603	4750	1591	4022	3507	1622

Notes: *** p<0.01, ** p<0.05, * p<0.1 Columns 1 reports the adaptation-adjusted estimates of the effects of temperature shocks on the log of average wage. Columns 2-7 reports estimates from similar regressions for the specified subgroup of firms. The 10-year temperature trends are computed by scaling the temperature data for each year by 1/t where t=1 for the most recent year and t=10 for the latest year in the period over which the ten-year trend is computed and is lagged by one year. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Table 1.8 Effects of temperature shocks on total output using weighted-temperature trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Log of total output			
	All sample	Labor intensive	SME	Agricultural	Non-ag	West Afr.	Southern Afr.
Above 90°F	-0.010***	-0.010***	-0.011***	-0.013**	-0.013**	-0.028***	0.007
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.016)
80-90°F	-0.003	-0.005*	-0.004	-0.001	-0.001	-0.022***	-0.010
	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)	(0.003)	(0.006)
60-70°F	-0.017***	0.003	-0.017***	-0.014	-0.014	0.008	-0.012
	(0.005)	(0.005)	(0.006)	(0.010)	(0.010)	(0.015)	(0.013)
Below 60°F	-0.011*	-0.032**	-0.021***	-0.010	-0.010	-0.349***	0.006
	(0.007)	(0.015)	(0.008)	(0.009)	(0.009)	(0.077)	(0.008)
Above 90°F 10-year trend	0.006	0.014**	0.007	0.015	0.015	0.007	0.006
	(0.005)	(0.006)	(0.006)	(0.011)	(0.011)	(0.015)	(0.035)
80-90°F 10-year trend	0.007**	0.013***	0.007*	0.004	0.004	0.007	0.013
	(0.004)	(0.005)	(0.004)	(0.007)	(0.007)	(0.013)	(0.044)
60-70°F 10-year trend	0.014*	-0.013	0.022**	0.013	0.013	-0.167*	-0.003
	(0.009)	(0.008)	(0.010)	(0.014)	(0.014)	(0.090)	(0.043)
Below 60°F 10-year trend	0.112*	0.100	0.046	0.179*	0.179*	-5.276***	0.021
	(0.063)	(0.093)	(0.071)	(0.095)	(0.095)	(1.247)	(0.299)
R-squared	0.209	0.262	0.155	0.140	0.140	0.238	0.336
Firm FE	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES
F-stat	39.75	18.59	25.24	8.754	8.754	29.59	24.19
N	6442	1603	4750	1591	4022	3507	1622

Notes: *** p<0.01, ** p<0.05, * p<0. Columns 1 reports the adaptation-adjusted estimates of the effects of temperature shocks on the log of total output (sales) on the transitory temperature shocks for the entire sample. Columns 2-7 reports estimates from similar regressions for the specified subgroup of firms. The 10-year temperature trends are computed by scaling the temperature data for each year by 1/t where t=1 for the most recent year and t=10 for the latest year in the period over which the ten-year trend is computed and is lagged by one year. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Table 1.9 Effects of temperature shocks on wages using 5-year (moving average) trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log of average wage						
	All sample	Labor intensive	SME	Agricultural	Non-ag	West Afr.	Southern Afr.
Above 90°F	-0.000	-0.008**	0.000	0.002	0.002	0.001	-0.036**
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.006)	(0.017)
80-90°F	-0.005***	-0.012***	-0.006***	-0.005*	-0.005*	-0.007*	-0.004
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)	(0.008)
60-70°F	-0.007*	0.004	-0.007	-0.005	-0.005	-0.011	-0.036***
	(0.004)	(0.005)	(0.004)	(0.007)	(0.007)	(0.016)	(0.014)
Below 60°F	-0.002	-0.018	-0.004	0.003	0.003	-0.320***	-0.004
	(0.006)	(0.011)	(0.007)	(0.009)	(0.009)	(0.076)	(0.008)
Above 90°F 5-year trend	-0.004	0.005*	-0.005*	-0.003	-0.003	-0.029***	0.001
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.008)	(0.023)
80-90°F 5-year trend	0.004**	0.009***	0.005**	0.006*	0.006*	-0.019***	-0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.006)	(0.020)
60-70°F 5-year trend	0.001	-0.009	0.002	0.004	0.004	-0.159***	0.007
	(0.004)	(0.006)	(0.004)	(0.006)	(0.006)	(0.034)	(0.012)
Below 60°F 5-year trend	0.048**	0.043	0.028	0.036	0.036	-1.465***	-0.029
	(0.023)	(0.035)	(0.026)	(0.037)	(0.037)	(0.444)	(0.106)
R-squared	0.0411	0.0319	0.0533	0.0222	0.0222	0.153	0.0163
Firm FE	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES
F-stat	3.647	3.356	4.644	1.542	1.542	11.88	2.469
N	6442	1603	4750	1591	4022	3507	1622

Notes: *** p<0.01, ** p<0.05, * p<0.1 Columns 1 reports the adaptation-adjusted estimates of the effects of temperature shocks on log of average wage for the entire sample. Columns 2-7 reports estimates from similar regressions for the specified subgroup of firms. The temperature trends are computed over the most recent five-year period preceding the firm survey and is lagged by one-year. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Table 1.10 Effects of temperature shocks on wages using 5-year (moving average) trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log of average wage						
	All sample	Labor intensive	SME	Agricultural	Non-ag	West Afr.	Southern Afr.
Above 90°F	-0.009***	-0.012***	-0.010***	-0.012*	-0.012*	-0.016***	0.005
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.005)	(0.021)
80-90°F	-0.003	-0.006**	-0.004	-0.001	-0.001	-0.016***	-0.010
	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.013)
60-70°F	-0.013***	0.005	-0.012**	-0.010	-0.010	0.021	-0.014
	(0.005)	(0.005)	(0.005)	(0.009)	(0.009)	(0.015)	(0.014)
Below 60°F	-0.012	-0.037**	-0.025**	-0.010	-0.010	-0.255***	0.007
	(0.008)	(0.016)	(0.010)	(0.009)	(0.009)	(0.074)	(0.010)
Above 90°F 5-year trend	0.002	0.009***	0.002	0.006	0.006	-0.005	0.001
	(0.003)	(0.003)	(0.003)	(0.007)	(0.007)	(0.007)	(0.036)
80-90°F 5-year trend	0.005**	0.008***	0.005*	0.002	0.002	0.000	0.009
	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.006)	(0.031)
60-70°F 5-year trend	0.011*	-0.005	0.013**	0.009	0.009	-0.113***	-0.001
	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.040)	(0.017)
Below 60°F 5-year trend	0.009	0.028	-0.023	0.035	0.035	-2.170***	0.029
	(0.033)	(0.046)	(0.039)	(0.045)	(0.045)	(0.477)	(0.164)
R-squared	0.210	0.261	0.158	0.137	0.137	0.244	0.361
Firm FE	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES
F-stat	38.65	19.13	25.85	8.036	8.036	28.99	24
N	6442	1603	4750	1591	4022	3507	1622

Notes: *** p<0.01, ** p<0.05, * p<0.1 Columns 1 reports the adaptation-adjusted estimates of the effects of temperature shocks on log of total output (sales) on the transitory temperature shocks for the entire sample. Columns 2-7 reports estimates from similar regressions for the specified subgroup of firms. The temperature trends are computed over the most recent five-year period preceding the firm survey and is lagged by one-year. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Table 1.11 Sensitivity of estimates to the choice of lags

	1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Log of average wage					Log of total output		
					Number of lags:					
	1-year	2-years	3-years	4-years	5-years	1-year	2-years	3-years	4-years	5-years
Above 90°F	-0.007***	-0.010***	-0.011***	-0.009***	-0.015***	-0.008***	-0.012***	-0.013***	-0.017***	-0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
80-90°F	-0.009***	-0.010***	-0.010***	-0.002	-0.003	-0.001	-0.003	-0.004*	-0.001	-0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
60-70°F	-0.000	-0.001	-0.002	0.000	0.002	-0.012***	-0.016***	-0.022***	-0.007	-0.009
	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)
Below 60°F	-0.009	-0.013**	-0.006	-0.016**	-0.018**	-0.019***	-0.019**	-0.004	-0.019**	-0.030***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.009)	(0.009)	(0.009)	(0.010)
R-squared	0.0530	0.0675	0.0692	0.0780	0.147	0.207	0.210	0.215	0.243	0.270
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
F-stat	6.598	6.682	6.910	5.800	7.603	39.01	37.06	37.44	38.72	25.14
N	6240	5379	5318	4185	2614	6442	5906	5843	4710	2844

Notes: *** p<0.01, ** p<0.05, * p<0. Columns 1-5 report the adaptation-adjusted estimates of the effects of temperature shocks on log of average wage with an increasing order of lags on the long-term temperature trend. Columns 6-10 report the adaptation-adjusted estimates of the effects of temperature shocks on log of output (sales) with an increasing order of lags from a 1-year lag to a two-year lag. Each regression includes ten-year temperature trends lagged as specified. The omitted temperature bin in each regression is i70-80°F (21-27°C). Each regression includes control for precipitation, relative humidity, the size of the firm, the log of the value of all fixed assets owned by the firm, and an indicator for whether the firm is an exporter. The standard errors clustered at the sector-location-year level are in parenthesis.

Table 1.12 Effects of temperature shocks on wages: alternative approach to measuring adaptation, Bento et. al., (2020)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log of average wage						
	All sample	Labor intensive	SME	Agricultural	Non-ag	West Afr.	Southern Afr.
Above 90°F shock	-0.009***	-0.008**	-0.008***	-0.003	-0.009***	-0.017***	-0.040*
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.024)
80-90°F shock	-0.009***	-0.014***	-0.010***	-0.007**	-0.010***	-0.016***	-0.008
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.010)
Above 90°F Trend	-0.004***	-0.004*	-0.004***	-0.001	-0.003**	-0.030***	-0.056**
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.010)	(0.024)
80-90°F Trend	0.001	-0.002	0.001	0.001	-0.001	-0.029***	0.030**
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.011)	(0.012)
R-squared	0.0454	0.0428	0.0561	0.0253	0.0588	0.136	0.0137
Firm FE	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES
Implied adaptation (above 90°F)	0.00512	0.00466	0.00410	0.00212	0.00571	-0.0124	-0.0162
P-value on adaptation	0.0676	0.170	0.168		0.104	0.202	0.177
F-stat	5.019	3.563	5.577	1.610	4.906	9.017	2.356
N	6442	1603	4750	1591	4022	3507	1622

Notes: *** p<0.01, ** p<0.05, * p<0: The dependent variable in columns 1-7 is the log of average wage. Each column reports the adaptation-adjusted estimates of temperature shocks for the two highest temperature bins as well as the long-term temperature bin. The temperature shock variables are computed as the deviation of the current year from the long-term temperature trend following Bento et. al. (2020). Implied adaptation is measured as the difference between the shock variable and the long-term variable (Bento et. al, 2020). The omitted temperature bin $70 - 80°F$ $(21 - 27°C)$. Tthe standard errors are clustered at the sector-location-year level. Each regression includes a control for the total annual rainfall and relative humidity in the firm's location, a indicator for whether the firm exports directly and the size of the firm (small, medium and large)

Table 1.13 Effects of temperature shocks on output: alternative approach to measuring adaptation, Bento et. al., (2020)

Log of output	(1) All sample	(2) Labor intensive	(3) SME	(4) Agricultural	(5) Non-ag	(6) West Afr.	(7) Southern Afr.
Above 90°F shock	-0.015***	-0.014***	-0.016***	-0.019***	-0.016***	-0.036***	-0.010
	(0.003)	(0.003)	(0.003)	(0.006)	(0.003)	(0.004)	(0.033)
80-90°F shock	-0.005**	-0.009***	-0.005**	-0.005	-0.008***	-0.022***	0.004
	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)	(0.004)	(0.016)
Above 90°F Trend	-0.006***	-0.003	-0.006***	-0.005	-0.004**	-0.014	-0.005
	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.012)	(0.031)
80-90°F Trend	0.001	0.002	0.000	-0.000	0.000	-0.008	0.004
	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.014)	(0.016)
R-squared	0.211	0.269	0.157	0.147	0.214	0.228	0.337
Firm FE	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES
Implied adaptation (above 90°F)	0.00976	0.0112	0.00990	0.0140	0.0119	0.0212	0.00556
P-value on adaptation	0.00213	0.000551	0.00298		0.00145	0.0774	0.765
F-stat	39.97	20.89	25.74	8.490	26.98	28.09	23.97
N	6442	1603	4750	1591	4022	3507	1622

Notes: *** p<0.01, ** p<0.05, * p<0: The dependent variable in columns 1-7 is the log of total output. Each column reports the adaptation-adjusted estimates of temperature shocks for the two highest temperature bins as well as the long-term temperature bin. The temperature shock variables are computed as the deviation of the current year from the long-term temperature trend following Bento et. al. (2020). Implied adaptation is measured as the difference between the shock variable and the long-term variable (Bento et. al., 2020). The omitted temperature bin $70-80°\mathrm{F}$ $(21-27°\mathrm{C})$. Tthe standard errors are clustered at the sector-location-year level. Each regression includes a control for the total annual rainfall and relative humidity in the firm's location, an indicator for whether the firm exports directly and the size of the firm (small, medium and large)

Table 1.14 Effects of temperature shocks on wages and output using lead variables of the long-term temperature trends

	(1)	(2)	(3)	(4)	(5)	(6)
	Log of average wage			Log of total output		
	2-year lead	3-year lead	4-year lead	2-year lead	3-year lead	4-year lead
Above 90°F	-0.006	-0.002	0.003	0.040***	0.044***	0.052***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.009)
80-90°F	-0.008**	-0.009***	-0.009**	0.015***	0.016***	0.017***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
60-70°F	-0.024***	-0.008	0.006	-0.018**	0.004	0.023**
	(0.008)	(0.006)	(0.008)	(0.009)	(0.007)	(0.009)
Below 60°F	0.041***	0.052***	0.044***	0.042***	0.039***	0.034***
	(0.011)	(0.010)	(0.009)	(0.013)	(0.012)	(0.011)
R-squared	0.146	0.156	0.150	0.203	0.216	0.212
Firm FE	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES
F-stat	9.834	12.12	13.27	17.02	21.70	18.92
N	1859	1859	1859	1864	1864	1864

Notes: *** p<0.01, ** p<0.05, * p<0: Each regression reports the adaptation-adjusted estimates of the effects of temperature shocks on labor productivity (columns 1-3) and total output (columns 4-6). Each regression includes lead variables of the five-year temperature trends (instead of lags). The omitted temperature bin $70-80°F$ $(21-27°C)$. Tthe standard errors are clustered at the sector-location-year level. Each regression includes a control for the total annual rainfall and relative humidity in the firm's location, an indicator for whether the firm exports directly and the size of the firm (small, medium and large)

Table 1.15 Effects of temperature shocks on firm age

	(1)	(2)	(3)	(4)
		Age of firm		
Above 90°F	0.002	0.002	-0.017*	-0.016
	(0.008)	(0.008)	(0.010)	(0.010)
80-90°F	0.004	0.009	0.009	0.013
	(0.008)	(0.008)	(0.008)	(0.009)
60-70°F	-0.002	0.005	-0.006	0.003
	(0.015)	(0.016)	(0.017)	(0.019)
Below 60°F	-0.032	-0.047	-0.039	-0.054
	(0.044)	(0.047)	(0.043)	(0.045)
Above 90°F 10-year trend			0.035***	0.034***
			(0.011)	(0.011)
80-90°F 10-year trend			-0.008	-0.009
			(0.007)	(0.007)
60-70°F 10-year trend			0.022	0.016
			(0.029)	(0.029)
Below 60°F 10-year trend			-0.302	-0.249
			(0.300)	(0.304)
R-squared	-0.000516	0.00761	0.00183	0.00951
Other controls	NO	YES	NO	YES
Firm FE	YES	YES	YES	YES
Time Trend	N	YES	YES	YES
F-stat	0.501	3.339	1.666	3.058
N	6442	6442	6442	6442

Notes: *** p<0.01, ** p<0.05, * p<0: Each regression regresses the firm's age on temperature shocks. The omitted temperature bin $70 - 80$°F $(21 - 27$°C$)$. Tthe standard errors are clustered at the sector-location-year level. Each regression includes a control for the total annual rainfall and relative humidity in the firm's location, a indicator for whether the firm exports directly and the size of the firm (small, medium and large)

REFERENCES

REFERENCES

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering? NBER Working Papers* 24003*,* National Bureau of Economic Research, Inc.

Behrer, A. P., & Park, J. (2017). Will we adapt? temperature, labor and adaptation to climate change. *Harvard Project on Climate Agreements Working Paper*, 16-81.

Bento, A., Miller, N. S., Mookerjee, M., & Severnini, E. R. (2020). *A Unifying Approach to Measuring Climate Change Impacts and Adaptation* (0898-2937). *NBER Working Papers* 27247, National Bureau of Economic Research, Inc.

Blakeslee, D. S., & Fishman, R. (2018). Weather shocks, agriculture, and crime evidence from India. *Journal of Human Resources, 53*(3), 750-782.

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

Cai, X., Lu, Y., & Wang, J. (2018). The impact of temperature on manufacturing worker productivity: evidence from personnel data. *Journal of Comparative Economics, 46*(4), 889-905.

Carter, C., Cui, X., Ghanem, D., & Mérel, P. (2018). Identifying the economic impacts of climate change on agriculture. *Annual Review of Resource Economics, 10*, 361-380.

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics, 4*(3), 66-95.

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature, 52*(3), 740-798.

Deryugina, T., & Hsiang, S. M. (2014). *Does the environment still matter? Daily temperature and income in the United States*. *NBER Working Papers* 20750*,* National Bureau of Economic Research, Inc.

Deschênes, O., & Greenstone, M. (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics, 3*(4), 152-185.

Dewenter, K. L., & Malatesta, P. H. (2001). State-owned and privately owned firms: An empirical analysis of profitability, leverage, and labor intensity. *American Economic Review, 91*(1), 320-334.

Diaz, D., & Moore, F. (2017). Quantifying the economic risks of climate change. *Nature Climate Change, 7*(11), 774-782.

Graff Zivin, J., & Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics, 32*(1), 1-26.

Heal, G., & Park, J. (2013). *Feeling the heat: Temperature, physiology & the wealth of nations*. *NBER Working Papers ,* National Bureau of Economic Research, Inc.

Heal, G., & Park, J. (2016). Reflections—temperature stress and the direct impact of climate change: a review of an emerging literature. *Review of Environmental Economics and Policy, 10*(2), 347-362.

Letta, M., Montalbano, P., & Tol, R. S. (2018). Temperature shocks, short-term growth and poverty thresholds: Evidence from rural Tanzania. *World Development, 112*, 13-32.

Linnerooth-Bayer, J., Warner, K., Bals, C., Höppe, P., Burton, I., Loster, T., & Haas, A. (2009). Insurance, developing countries and climate change. *The Geneva Papers on Risk and Insurance-Issues and Practice, 34*(3), 381-400.

LoPalo, M. (2020). Temperature, worker productivity, and adaptation: evidence from survey data production. *Available at SSRN.*

Mérel, P., & Gammans, M. (2018). Climate econometrics: Can the panel approach account for long-run adaptation?

Moore, F. C., Obradovich, N., Lehner, F., & Baylis, P. (2019). Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change. *Proceedings of the National Academy of Sciences, 116*(11), 4905-4910.

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

Seppanen, O., Fisk, W. J., & Lei, Q. (2006). Effect of temperature on task performance in office environment.

Somanathan, E., Somanathan, R., Sudarshan, A., & Tewari, M. (2015). The impact of temperature on productivity and labor supply: Evidence from Indian manufacturing. *Indian Statistical Institute, New Delhi, India*.

Traore, N., & Foltz, J. (2017). Temperatures, Productivity, and Firm Competitiveness in Developing Countries: Evidence from Africa. *Unpublished manuscript, University of Wisconsoin–Madison. https://nouhoumtraoresite. files. wordpress. com/2017/12/jmp_ntraore. pdf.*

Wise, R. M., Fazey, I., Smith, M. S., Park, S. E., Eakin, H., Van Garderen, E. A., & Campbell, B. (2014). Reconceptualising adaptation to climate change as part of pathways of change and response. *Global Environmental Change, 28*, 325-336.

Xie, V. W. (2018). Heterogeneous firms under regional temperature shocks: exit and reallocation, with evidence from Indonesia. *ASSA Paper Session, Department of Economics, University of California, San Diego, CA, USA*.

Zhang, P., Deschenes, O., Meng, K., & Zhang, J. (2018). Temperature effects on productivity and factor reallocation: Evidence from a half million Chinese manufacturing plants. *Journal of Environmental Economics and Management, 88*, 1-17.

Zhang, P., Zhang, J., & Chen, M. (2017). Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation. *Journal of Environmental Economics and Management, 83*, 8-31.

CHAPTER 2: INSTITUTIONAL CONTEXT AND ADAPTATION: EVIDENCE FROM THE UNITED STATES

2.1 Introduction

This section builds on the methods and the tools developed in section 3.4 of chapter 1 to examine the short-run impacts of temperature shocks on wages in the United States (U.S.) and examine whether adaptation-based on longer-term normals is relevant to workers (and firms) in the United States. In the preceding sections, we demonstrated that we could control for the adaptation component of the impacts of extreme temperatures by accounting for time-varying longer-term temperature normals in the panel fixed-effects regressions. However, we do not know how this argument performs across different institutional, economic and environmental contexts. Understanding how our argument works across different contexts is important for two main reasons. First, adaptation does not happen in a vacuum. It is not separable from the institutional, environmental and economic contexts in which it occurs (Wise et al., 2014). In other words, most adaptation is local.

Governmental and private institutions via formal and informal mechanisms can play a major role in directly or indirectly bolstering the resilience of firms and workers to temperature shocks. First, they can provide a policy framework for streamlining adaptation by for example, enforcing heat-related safety standards at the workplace. In addition, they can provide incentives such as subsidies or low-interest loans to encourage the adoption of cooling or heating technologies all of which can directly encourage firm-specific adaptation. It is also possible that government and private institutions provide logistical support to address extreme weather situations that may inhibit the functioning of firms and the movement of labor on such days which can reduce the need for firm-specific adaptation based on learning. Therefore, the extent to which such institutions are directly or indirectly involved in bolstering resilience can lead to differential outcomes in the rates at which firms and workers themselves have to rely on historical local temperature data to engage in adaptive behavior. This could either enhance or reduce the relevance of longer-term temperature averages in revealing information about adaptation.

For example, county governments in Texas may be relatively less equipped to facilitate snow removal and handle extremely cold days than county governments in Michigan although the Texan governments may have had data on these "occasional snowstorms" for the last century if the cost of investing in a Michigan-like adaptive response is not justified by the frequency of the events in Texas. All else equal, the differences in institutional and environmental capacity to handle the same snowstorm will lead to more firm closures, and lower wages in Texas in response to a snowstorm than in Michigan in the short-run.[22] Similarly, governments in colder areas may be relatively less equipped to handle a heat wave that threatens productivity than governments in relatively warmer areas if the frequency of heat waves (and the damages caused by them) do not justify the cost of public investments in cooling centers.

What does this mean for our strategy in equations (6) and (7)? The preceding argument suggests that differences in institutional and environmental preparedness to handle extreme temperatures might lead to differential short-run impacts of extreme temperatures on firms and workers even if they have the same amount of historical data on the local weather. Specifically, better institutional capacity in directly managing thermal stress will mitigate short-run impacts and ultimately reduce the need for firm-specific adaptation based on learning.

Accordingly, this section of the paper seeks to examine the short-run impacts of extreme temperatures on workers' wages and the extent to which workers in the United States are adapting to extreme temperatures based on climate memory. This analysis makes an important contribution

[22] This argument assumes that workers might take days off on extremely hot days for hourly-wage workers. A further assumption is that continuous closures on days of extreme heat might lower output in the long run, leading to layoffs and thus lower wages/compensation even for salaried employees.

to the literature. It sheds light on how institutional, economic and environmental factors interact with firms and workers' adaptive responses based on historical local temperature data. Comparing these results to the findings from the previous sections (on Africa) further highlight the costs of failing to account for adaptation in estimates of short-term temperature shocks in developing regions where the need for and barriers to adaptation are more prevalent.

2.2 Data

We use data from two main sources. First, the data on wages used in this section comes from the County Business Patterns (CBP) database from 1990 to 2011 from the United States Census Bureau. The CBP dataset contains information on total annual compensation paid by firms to all their employees as well as the total number of employees by these firms aggregated at county-year-industry level for all 3,006 US counties. We divide the total annual compensation in U.S. dollars by total number of employees to obtain the average compensation at the county-year-industry level. The industries in each county are classified by their two-digit North American Industry Classification System (NAICS) code. The industries in our study include mining, utilities, construction, manufacturing, retail and wholesale trade, transportation, information, finance and insurance, real estate, management of companies, administrative support, educational services, health care and social assistance, entertainment and recreation as well as accommodation and food services. The second source of data is the weather data from the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA). We extracted the daily total precipitation (in mm), maximum temperature and relative humidity data for all the grids in all counties of the United States from 1980 to 2019.

2.3 Estimation Strategy

To identify the short-run impacts of extreme temperatures and the extent of adaptation induced by climate memory, I first obtain naïve estimates of the impact of short-term temperature shocks on average wage by implementing the equation below:

$$Y_{cit} = \sum_{b=1}^{10} \beta^b T_{ct}^b + \gamma \boldsymbol{D}_{ct} + \delta_t + \eta_{ci} + \varepsilon_{cit} \qquad (9)$$

In this specification, c indexes the county, i is the two-digit NAICS sector code, while t is the year. The data is stacked at the county-industry-year level. Therefore, our outcome variable in this section, Y_{cit} is defined as the log of average compensation (i.e., per worker) in industry i in county c in year t. As before T_{ct}^b is an index for the temperature bin b in year county c at time t. In this specification, we use all ten temperature bins, each of size $10°\text{F}$ in order to exploit the full variation in the distribution of temperature days to which workers are exposed. We choose a moderate temperature bin, $70\ to\ 80°\text{F}\ (21\ to\ 27°\text{C})$ as the omitted category to make our interpretation of the coefficients on the hottest and coldest temperature bins more intuitive. $\boldsymbol{D}_{ct}$ are the other time-varying observables at the county such as precipitation and relative humidity that may be correlated with both temperature and workers' compensation. δ_t is the linear time trend, η_{ci} accounts for all-time invariant county-industry characteristics such as the geographical location of the county and, observables while ε_{cit} is the idiosyncratic error term. The standard errors are clustered at the state-year level to account for the effects of weather shocks that may be correlated within a state in a given year.

Under the plausible assumption that there are no time-varying unobservables that are correlated with current temperature and firm outcomes, β^b identifies the combined short and long-run semi-elasticity of average worker compensation with respect to an additional day within temperature bin b compared to an additional day of moderate temperature.

To isolate the long-run component and identify the short-run impact of extreme temperatures to workers, we add a measure of the long-term temperature averages as an additional explanatory variable in specification (10). Specifically, we include variables, $\bar{T}_{dt}^{b}$, which measure the (moving) average number of days over the last ten years during which the temperature fell within the temperature bin, b in county c in year t as shown in equation (10). The long-term normal is lagged by one year to reflect all the temperature data available to the firms up to the year preceding the survey.

$$Y_{cit} = \sum_{b=1}^{5} \tilde{\beta}^{b} T_{ct}^{b} + \sum_{b=1}^{5} \alpha^{b} \bar{T}_{ct}^{b} + \gamma D_{sct} + \delta_t + \eta_{ci} + \mu_{cit} \quad (10)$$

The definitions of perfect adaptation, zero adaptation and incomplete adaptation are identical to those in equations (5) and (6).

2.4 Results and Discussion

We report our estimates from equation (8) and (9) in Table 2.1. The dependent variable in Table 2.1 is the log of average annual worker compensation. Column 1 reports the estimates from equation (9) for the entire sample while column 2 reports the estimates from equation (10) which accounts for the long-term trend. We find in the naïve regression (equation 9) that an extra day with temperature above 90°F reduces wages by about 0.04% relative to an extra day with moderate temperature (i.e., $70 - 80$°F). This finding is similar to other estimates in the literature. Behrer and Park (2017) find that an extra day with temperature above 95°F reduces wages in the United States by about 0.04% relative to an extra day with temperature between $70 - 80$°F. However, we do not find a statistically significant difference between the naïve and adaptation-adjusted estimates for days with temperature above 60°F although both group of estimates are statistically significant, suggesting that accounting for time-varying longer-term normals does not tell us about the extent of adaptation. This contrasts with our findings in Sub-Saharan Africa. However, similar to Sub-Saharan Africa, we find that workers in labor-intensive industries, on average, experience statistically significantly higher reductions in wages compared to capital-intensive industries in the short run (columns 4 and 6). Similarly, workers in hot areas (defined as counties with historical 80°F and above days exceeding the fourth quartile of the historical 80°F+ day distribution) experience smaller reductions in wages in response to an extra day with 80°F+ (relative to a day with temperature between 70-80°F) in the short run than workers in cooler areas.

Several factors might explain why the use of long-term normals does not sufficiently tease out the long-run adaptation component in the USA. First, most adaptation is local and differences in the quality and the capacity of local institutions (between low and high-income countries) to streamline and mobilize resources to aid adaptation can differentially impact the extent to which firms and workers will need to use historical local climate data to engage in adaptive behavior.[23] To the extent that such built-capacity exists (in high-income country contexts such as the USA) regardless of the inter-temporal changes in long-term weather patterns, the degree to which the inclusion of time-varying long-term normals can be informative about adaptation will be limited. Second, considerable differences in the presence and functioning of insurance and other risk pricing and transfer instruments between low and high-income countries can affect the extent to which longer-term normals can reveal information about adaptation. For example, in low-income countries, economic losses as a share of GDP from natural disasters (many of which are induced by climate change) are estimated to be twice as high as the losses in high-income countries. These differences in climate impacts are not only due to investment in adaptive technology or behavior in high-income countries but also the presence and functioning of insurance and risk transfer markets. For instance, between 1980 and 2004, it is estimated that 30 percent of all losses in high-

[23] It should be noted that the capacity of local institutions to streamline and make firms and workers more resilient may itself be built by prior experience with extreme weather events (i.e., historical data)

income countries were insured which compares to just about 1 percent of all losses in low-income countries (Linnerooth-Bayer et al., 2009). Therefore, a firm in the United States facing the same extreme-temperature risk as a firm in Sub-Saharan Africa, might have less need for historical data on time-varying long-term temperature normals than its counterpart in SSA. Moreover, infrastructure in the United States may be more climate-resilient than in many countries in Sub-Saharan Africa due to stronger environmental and civil (construction) regulatory capacity, further suggesting that even when faced with the same extreme temperature risk with the same amount of information on historical temperatures, adaptation rates (based on climate memory) may differ and in the case of the United States, incorporating long-term normal might not be as effective in telling us about the firm's adaptive behavior. The foregoing arguments suggest that adaptive responses to extreme temperatures may be better streamlined in the United States than in Sub-Saharan Africa. This provides a more robust risk management mechanism for dealing with extreme temperature impacts, thus limiting the degree to which workers rely on historical weather data.

In summary, we note that incorporating long-term averages when trying to understand the impacts of short-term temperature shocks is extremely important in developing countries, particularly SSA. Failing to do so underestimates the true impact of these shocks (which could undermine the urgency attached to impacts of short-term climate shocks) could lead to misguided climate adaptation policy design and implementation. However, our findings from the US suggest that incorporating long-term normal when trying to understand the impacts of short-term temperature shocks is less important in high-income countries where built institutional capacity and stronger contingency markets for handling thermal stress might limit the need for firm specific adaptation. Our theoretical model suggest that this is possible when firm level adaptation is complete or absent/not necessary both of which depend on environmental and institutional factors as discussed.

Together these results imply that when incorporating longer-term normals with the goal to understanding adaptation, researchers should pay attention to the institutional and environmental contexts that might necessitate the use of longer-term normals by economic agents to inform adaptive behavior. Perhaps in the case of high-income countries, the (naïve) panel fixed-effects (exclusive of long-term nomals) might be sufficient in revealing adaptation (Deschênes & Greenstone, 2011; Schlenker & Roberts, 2009). The naïve panel fixed-effects approach exploits the variation in climate within regions to identify the short-run impacts exclusive of adaptation. This approach might be sufficient because institutional capacities and markets for sharing climate risks have been established regardless of the inter-temporal variations in long-term normals. Therefore, the adaptation component is captured in the fixed effect and in so far as adaptive behavior does not change dramatically (as is the case in many developing countries now experiencing inflow of adaptation investments and aid), the approach of Deschênes and Greenstone (2011) should suffice. But in the case of Sub-Saharan Africa, not incorporating time varying longer-term normals (and the use of the panel fixed-effects approach alone) underestimates the short-run impacts of extreme temperatures to workers and firms.

2.5 Conclusion and Policy Recommendations

This chapter builds on the tools developed in the preceding chapter to examine the impacts of temperature shocks on workers and for evidence of adaptation to extreme temperatures among workers in the United States. Specifically, we compare the estimates of the naïve short-run effects to the pure (adaptation-free) short-run effect to infer the level of adaptation to extreme temperatures among workers in the United States.

Similar to what we found in Sub-Saharan Africa, we find that workers in labor-intensive industries, on average, experience statistically significantly higher reductions in wages in response to extremely hot days compared to capital-intensive industries in the short run. Similarly, workers in hot areas (defined as counties with historical 80°F and above days exceeding the fourth quartile of the historical 80°F+ day distribution) experience smaller reductions in wages in response to an extra hot day in the short run than workers in cooler areas. However, in contrast with Sub-Saharan Africa, we do not find statistically distinguishable differences between the naïve and the pure short-run impacts of extreme temperatures on wages. This finding suggests that there may be a limited-need for worker-specific adaptation or zero recent adaptation to (1-year shocks) temperature shocks. While our approach does not allow us to distinguish zero recent adaptation from a situation of limited need for worker/firm-level adaptation, it suggests that the institutional, economic and environmental contexts within which workers operate, and where adaptation occurs must be understood when implementing our proposed strategy. In high-income countries, where established capacity for mitigating extreme weather impacts exists, the inclusion of time-varying longer term normal is perhaps less important and thus might not tell the researcher much about firm specific adaptive behavior. Therefore, researchers may need to exploit other alternatives for understanding adaptation in such contexts. Further, it reinforces the importance of accounting for adaptation when estimating the short-run impacts of extreme temperatures in low-income or weak institutional contexts.

APPENDIX

Table 2.1 Effects of temperature shocks on wages in the United States

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Full sample		Labor Intensive		Non-Labor Intensive		Moderate Climate		Hot Climate	
	Naïve	Adapt	Naïve	Adapt	Naïve	Adapt	Naïve	Adapt	Naïve	Adapt
Above 90°F	-0.0004***	-0.0003**	-0.0006***	-0.0004***	-0.0003**	-0.0002	-0.0005***	-0.0005***	-0.0004***	-0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)
80-90°F	-0.0004***	-0.0003***	-0.0005***	-0.0003***	-0.0003***	-0.0002**	-0.0004***	-0.0003**	-0.0002	-0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
60-70°F	-0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000	-0.0004**	-0.0003*	0.0002	0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
50-60°F	-0.0002**	-0.0001	-0.0002	0.0000	-0.0002	-0.0001	-0.0004***	-0.0004***	0.0000	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
40-50°F	0.0001	0.0002	-0.0000	0.0001	0.0002	0.0003**	-0.0002	-0.0002	0.0003*	0.0005***
	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
30-40°F	0.0003*	0.0004***	0.0003	0.0005***	0.0004**	0.0004**	-0.0001	-0.0001	0.0008***	0.0010***
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
20-30°F	0.0003	0.0005***	0.0001	0.0004*	0.0006***	0.0007***	0.0003	-0.0000	0.0003	0.0007**
	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)
10-20°F	-0.0005	0.0004	-0.0007*	0.0006*	-0.0003	0.0003	-0.0002	-0.0000	0.0010**	0.0015***
	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0006)
R-squared	0.000120	0.000416	0.000170	0.000572	8.90e-05	0.000333	7.81e-05	0.000633	0.000142	0.000559
Long-term trend	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
County Sector FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean dep-var ($)	32,375	32,375	30,122	30,122	34,628	34. 628	30,648	30,648	30,648	30,648
p-value of joint test	2.49e-05	0.000639	8.23e-05	0.00134	7.61e-05	0.00103	0.000946	0.00961	1.03e-05	2.02e-05
N	452,140	452,140	224,205	224,205	227,930	227,930	188,153	188,153	263,986	263,986

Table 2.1 (Cont'd)

Notes: *** p<0.01, ** p<0.05, * p<0.1 The dependent variable in this regression is the average payroll in a sector. The naïve estimates do not account for adaptation while adaptation-adjusted estimates include long-term temperature bins to account for adaptation. The omitted temperature bin in each regression is 70-80°F (21-27°C). Each regression includes control for precipitation and relative humidity. The standard errors clustered at the sector-county-year level are in parenthesis.

REFERENCES

REFERENCES

Behrer, A. P., & Park, J. (2017). Will we adapt? temperature, labor and adaptation to climate change. *Harvard Project on Climate Agreements Working Paper*, 16-81.

Bento, A., Miller, N. S., Mookerjee, M., & Severnini, E. R. (2020). *A Unifying Approach to Measuring Climate Change Impacts and Adaptation* (0898-2937).

Deschênes, O., & Greenstone, M. (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics, 3*(4), 152-185.

Linnerooth-Bayer, J., Warner, K., Bals, C., Höppe, P., Burton, I., Loster, T., & Haas, A. (2009). Insurance, developing countries and climate change. *The Geneva Papers on Risk and Insurance-Issues and Practice, 34*(3), 381-400.

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

Traore, N., & Foltz, J. (2017). Temperatures, Productivity, and Firm Competitiveness in Developing Countries: Evidence from Africa. *Unpublished manuscript, University of Wisconsoin–Madison. https://nouhoumtraoresite. files. wordpress. com/2017/12/jmp_ntraore. pdf.*

Zhang, P., Zhang, J., & Chen, M. (2017). Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation. *Journal of Environmental Economics and Management, 83*, 8-31.

CHAPTER 3: DO BENEFITS OF EXPANDED MIDSTREAM ACTIVITIES IN CROP VALUE CHAINS ACCRUE TO SMALLHOLDER FARMERS? EVIDENCE FROM ZAMBIA

Nuhu, A. S., Liverpool-Tasie, L. S. O., Awokuse, T., & Kabwe, S. (2021). Do benefits of expanded midstream activities in crop value chains accrue to smallholder farmers? Evidence from Zambia. *World Development*, 143. doi: https://doi.org/10.1016/j.worlddev.2021.105469

3.1 Introduction

In recent decades, many developing countries have experienced significant changes on both the demand and supply sides of the agri-food industry. For sub-Saharan African (SSA), the transformation in the food system on the demand-side is driven by dramatic changes in consumer tastes, preferences, and dietary choice (Tschirley et al., 2015). Similarly, though less notable, the landscape has also been changing on the supply-side where a significant transformation is taking place in the midstream (i.e., food processing, distribution logistics, wholesale marketing) and downstream (i.e., retail marketing, and services).

This rapid transformation of food systems in developing countries creates market and employment opportunities for farmers. One key driver of such opportunities is the increased demand for crops and as inputs for agrifood processing industries in response to changing consumption patterns of a rapidly urbanizing population (Ehrich & Mangelsdorf, 2018; Michelson et al., 2012; Reardon, 2015; N. J. Sitko et al., 2018). This reflects the situation in Zambia, where the recent rise in local and regional demand for animal feed and edible oil, as part of the ongoing transformation in the region has led to an expansion of soybean cultivation (Davids et al., 2017; Ferdinand et al., 2018; Foyer et al., 2019).[24]

The extent to which the ongoing transformation in food systems is beneficial to smallholder farmers within the broader group of farmers is uncertain. Although processing firms tend to purchase larger quantities, offer better prices and provide access to improved technologies and training, they may exclude smallholder farmers if they require high volumes of output and quality standards that many smallholders are unable to meet (Ghezán et al., 2002; Louw et al., 2008; Rao & Qaim, 2011; N. J. Sitko et al., 2018; Van Der Meer, 2006). Consequently, the extent to which

---

[24] 89% of soybean output in Zambia is estimated to go into the animal feed processing industry (Siamabele, 2019)

smallholders can benefit from these fast-evolving value chains remains an empirical question in need of more investigation.[25]

Existing studies that have investigated how ongoing dietary changes and their associated food systems transformation in developing countries impact smallholders have tended to do so through the lens of contract farming (Barrett et al., 2012; Bellemare & Novak, 2017; Euler et al., 2017; Gatto et al., 2017; Ton et al., 2017). Contract farming as used in this paper refers to formal contractual arrangements between farmers and the immediate recipient of the final soybean output prior to the production and harvesting of the soybean. Thus, non-contract farming as used in this paper refers to cultivation of soybean without any prior formal purchase agreement with the immediate buyer of output.

There are several reasons why formal contractual arrangements are likely to have significantly different implications for smallholders' welfare compared to purchases by processing firms or traders who purchase without contracts. First, the documented channels through which contract farming affects smallholder outcomes such as provision of modern inputs to improve productivity and guaranteed output markets may or may not be available to smallholder farmers in the absence of formal contracts.[26] Therefore, welfare outcomes for smallholder farmers engaged in non-contract farming arrangements may not necessarily be driven by direct provision of modern inputs and a guaranteed market but through higher prices and the opportunity to sell more output

---

[25] The Ministry of Agriculture in Zambia defines smallholder households as cultivating less than 20 hectares of land. Within this group, those who cultivate less than 5ha are classified as small-scale farmers while those who cultivate between 5-20ha of land are classified as medium-scale smallholders (N. Sitko & Chamberlin, 2015).

[26]While often felt that provision of modern inputs is largely restricted to contract farming arrangements, a recent scoping review of studies across Africa, Asia and Latin America by Liverpool-Tasie, Wineman, et al. (2020) finds that the provision of complementary services such as modern inputs, training and logistics services is also very common in non-contract arrangements with agribusiness firms though some firms are limited in their capacity to provide these services.

that is offered by these processing firms.[27] For example, Nuthalapati et al. (2020) find that in the absence of contracts, supermarkets tend to offer higher prices to farmers in order to guarantee high quality vegetable supply.[28] Second, the focus on contract farming implies that the literature inadvertently ignores a large proportion of poor smallholder farmers since contract farming opportunities are often less available to poor smallholder farmers and account for less than 10% of their sales (Reardon et al., 2019; Ton et al., 2017).

Hence, understanding if the rise of non-contract farming opportunities within the ongoing expansion of food value chains is inclusive of smallholder farmers and the channels through which such non-contract farming arrangements impacts smallholder farmers becomes an important policy consideration. It is particularly pertinent for any effort to reduce poverty and/or promote sustainable economic development since the share of agricultural land under smallholder farmers, who constitute a significant share of the world's poor is still rising (Burney & Naylor, 2012; Ogutu & Qaim, 2019).

This paper contributes to this limited literature by examining the effects of the recent rise in non-contract based purchasing activities of large soybean processing firms and traders (referred to as large buyers) on the welfare of smallholder farmers in Zambia.[29] We use longitudinal data from the Zambia Rural Agricultural Livelihoods Survey (RALS) covering the period 2012 to 2019 to examine the effects of selling soybeans to large-scale buyers on the welfare outcomes of

[27] These opportunities for higher price or the opportunity to sell could incentivize smallholder farmers to invest in the use of modern inputs even in the absence of their direct provision as is the case in contract farming arrangements.
[28] It is worth noting that soybean which is largely standardized is very different from horticultural products that tend to be highly differentiated and as such, the expansion of the two value chains may have differential impacts on farmers as well.
[29] Large buyers in this paper are defined as large-scale food processing, milling and trading firms based on three defining characteristics as described in Burke et al. (2019) and in section 3 of this paper. Our classification of selling to large buyers exclude the few outgrowers and contract-based sales. See Table 2b in appendix B for a partial list of large-scale soybean processing firms.

smallholders. The soybean value chain in Zambia offers an interesting case study for demonstrating how the ongoing dietary transformation changes and associated food systems transformation in Africa is providing commercialization opportunities for farmers outside of contract farming arrangements. The production of soybean (a key input in animal feed production) has expanded in response to growing domestic and regional demand for meat and cooking oil: a key feature of the African food systems transformation (Ferdinand et al., 2018; Foyer et al., 2019; Lubungu et al., 2013; Opperman & Varia, 2011). This rise in animal feed demand has resulted in the expansion of feed processing firms whose demand for soybean from farmers via non-contract farming arrangements provides an apt opportunity for understanding if and how the ongoing transformation of food systems in Africa includes smallholders and impacts their welfare. Furthermore, the rapid growth of the soybean industry in Zambia is a recent phenomenon. The share of households engaged in soybean cultivation in Zambia has more than doubled between 2012 and 2019.[30] These factors jointly imply that an assessment of the ongoing transformation of the soybean value chain in Zambia and its impact on smallholder farmers is both timely and relevant to the research and policy debate on the inclusiveness of Africa's food systems transformation.

This paper makes three important contributions to the literature. This is the first study (to the best of our knowledge) that empirically explores the welfare effects of the rapid expansion of non-contract opportunities in the midstream of commodity value chains in Africa on smallholder farmers. Thus far, the literature on midstream activities has focused on contract opportunities—which have been shown to predominantly benefit larger farmers (Ton et al., 2017). Second, this

[30] The rise in smallholder participation is soybean production is also said to be driven by the efforts of organizations such as the Zambian National Farmers Union and USAID funded programs to promote soybean production amongst smallholders (Chisanga & Sitko, 2013)

paper contributes to the thin empirical evidence on whether the ongoing expansion of commercialization opportunities due to Africa's changing dietary patterns benefits smallholder farmers beyond the supermarket revolution literature (Andersson et al., 2015; Ogutu & Qaim, 2019; Rao & Qaim, 2011). Third, in addition to considering the direct welfare effects of smallholder engagement in commercialization opportunities in the midstream, we also explore some of the mechanisms through which the welfare effects could occur.

Many studies either consider only the welfare effects of improved commercialization opportunities or focus largely on intermediate outcomes such as prices, sale decisions, technology adoption or farmer yields (Burke et al., 2019; Nuthalapati et al., 2020). However, higher yields might not be sufficient to cover the cost of engagement in the market and higher prices might not translate into higher profits if corresponding costs rise proportionately or at a higher rate (Meemken & Bellemare, 2020; Ragasa & Mazunda, 2018). Thus, empirical evidence on the extent to which smallholder farmers are benefitting from Africa's ongoing transformation (with insights on the mechanism explaining these effects) is important for African governments and development practitioners. Such evidence could aid decisionmakers as they design programs and policies to ensure that Africa's ongoing food system transformation is inclusive of smallholder farmers and benefits them. Inclusivity considerations that enhance the ability of smallholder farmers to meet participatory requirements of modern value chains (i.e., fixed costs of technology adoption, minimum sale volumes and safety standards) are imperative. Since smallholder farming remains a key livelihood source for many poor people in Africa, the benefits of higher level of inclusion via targeted government programs and policies could be welfare-enhancing (Christiaensen et al., 2011; Euler et al., 2017; Larson et al., 2012).

Using fixed effects and instrumental variable estimation techniques to address unobserved factors correlated with both our outcome variables and the household's decision to sell to large scale buyers, we find significant positive welfare effects of selling soybean to large buyers. Selling to large buyers is statistically significant and positively associated with higher crop incomes for the average smallholder household. This includes both small-scale farmers (operating less than 5 hectares of land) and medium-scale farmers (operating 5-20 hectares of land). However, the observed effects translate into higher total household incomes and poverty-reducing effects for medium-scale smallholders but not for small-scale smallholders. We also find that although small-scale farmers receive a price premium from selling to large buyers, the estimated income effects for all smallholders are largely driven by the opportunity to sell more output provided by large buyers. This indicates that while smallholders are included in and benefitting from the rapidly expanding soybean value chain in Zambia, additional efforts are needed to ensure that the higher crop incomes from participation translate into increased total household income and reduced poverty for small-scale smallholders.

The remainder of the paper is organized as follows. Section two provides a brief overview of the midstream sector of the Zambian soybean value chain and reviews relevant empirical literature. Section three discusses our data and defines the key variables used in the analysis. Section four contains a discussion of the empirical methods estimation and strategy while section five presents the estimated results. Finally, section six contains our concluding remarks and recommendations for policy.

3.2 A Description of the Midstream of the Zambian Soybean Value Chain

At the center of Africa's rapidly transforming food systems is a burgeoning middle segment which has largely been ignored until recently (Reardon, 2015; N. J. Sitko et al., 2018).

This midstream segment, sometimes referred to as the *"hidden middle"* is a key determinant of the prices that farmers receive as well as the price and quality of products for industry and consumers (Liverpool-Tasie & Parkhi, 2020; Richards et al., 2016; Schoonhoven-Speijer & Vellema, 2020). The hidden middle includes processors, wholesalers and numerous logistics companies that operate along food supply chains and account for about 30-40% of the value added and costs in the value chain (Reardon, 2015). Recently, there has been a rise in the number and scale of purchasing activities of midstream actors and their agents seeking to meet the rising demand of an urbanizing population (Deininger & Byerlee, 2011; N. J. Sitko et al., 2018). One implication of the expansion of the midstream is that in sectors previously dominated by atomistic small-scale traders and households, the presence of large-scale traders and processors increases competition and thus provides more market opportunities for farmers. However, it has also been documented that in many agricultural markets where midstream agents dominate, markets tend to be oligopolistic. This allows midstream agents to exploit smallholder farmers; particularly in places where farmers are spatially dispersed (Kikuchi et al., 2016; Montalbano et al., 2018; Sexton, 2013). Thus, the activities of these midstream agents are critical for many smallholder farmers as they influence farmers' incentives through the price farmers receive and the quantities farmers are able to sell (Burke et al., 2019; Nuthalapati et al., 2020).

This paper specifically focusses on large scale processing firms and traders purchasing from farmers without contracts in Zambia; henceforth referred to as large buyers. Figure 3.1 describes the soybean value chain in Zambia and highlights the critical role midstream agents play in linking smallholder farmers with consumers. In addition to linking farmers to markets, the diversity of agents within the midstream sector induces competition thus yielding potentially improved outcomes for smallholder farmers. As soybean demand in Zambia has increased over the last decade, the country has seen an expansion in soybean procurement by large buyers. However, the growth in the processing capacity and associated demand for soybean in Zambia has not been matched by production (Lubungu et al., 2013). The expansion of these buyers into areas where they can find large surpluses indicates that smallholder farmers may be able to benefit from added market opportunities.

On one hand, large buyers may be able to leverage on their economies of scale to drive out competition and thus reduce market and high-price opportunities for smallholder farmers (Burke et al., 2019). However, competition with other buyers could induce these large buyers to offer competitive prices, thus improving the welfare of smallholder farmers. Over the study period, farmers in our sample who sold to large buyers and processors received 2.81 Kwacha per kg on average, compared to 2.67 Kwacha per kg received by those who sold to small-scale traders. This is in line with Lubungu et al. (2013) who noted that large-scale buyers tend to buy soybean at non-negotiable prices which then act as the benchmark price for other buyers including small-scale traders. While the downstream markets for small-scale traders are unclear, it is argued that they often store these grains to take advantage of spatio-temporal arbitrage. Moreover, while evidence (though limited) exists in the literature on the impacts of the activities of midstream actors in traditional grain markets such as maize, there is no evidence (the authors are aware of) on the impacts of their activities in soybean markets. This is not surprising as the expansion of soybean production is a recent phenomenon in Zambia. For example, Burke et al. (2019) study the effects of the rise in medium-scale farms (MSFs) on market opportunities for smallholder maize farmers and find that smallholders in areas with growing MSFs tend to be more likely to participate in the private sector maize market. In particular, they tend to be more likely to sell to large-scale maize traders. Although Burke et al (2019) find significant positive effects on expected sales, they do not focus on welfare outcomes for smallholder farmers.

Fung et al. (2019) examine the other dimension of the Zambian grain market by investigating the impacts of the Zambian Food Reserve Agency's (FRA) maize purchasing activities on smallholder welfare in Zambia. They find sales to the FRA to be positively associated with most household-level poverty outcomes among net maize sellers. However, their analysis is restricted to a government program. N. J. Sitko and Jayne (2014) examine the activities of these midstream actors in the Zambian grain markets and find that the purchasing activities of grain assemblers benefit smallholder farmers by reducing search and transport costs. However, they do not econometrically estimate the magnitudes of the impacts of the grain purchasing activities of midstream agents. Similarly, Montalbano et al. (2018) examine the effects of market chain participation on food security for Ugandan farmers of a staple crop, maize.

While it is understandable that the literature has tended to focus on traditional grains and staples, rising incomes and the diversification of consumption away from staples to animal-sourced foods (ASF) warrants more attention to non-staple value chains in the literature. This is because farmers across Africa are increasingly including non-staple crops or non-crop cash products in their portfolio of economic activities. In Zambia, farmers are increasingly incorporating soybean in the portfolio of crops which they grow as the number of soybean processing firms continue to increase— changing the market dynamics for smallholder farmers (Ferdinand et al., 2018; Foyer et al., 2019). Thus, the goal of this paper to examine the implications of these changing market dynamics, particularly, in the midstream segment of the soybean value chain, on smallholder welfare outcomes.

3.3 Data

Our paper uses data from the Zambian Rural and Agricultural Livelihoods Survey (RALS) collected by the Indaba Agricultural Policy Research Institute (IAPRI) in collaboration with the Zambian Central Statistical Office (CSO) and the Ministry of Agriculture (MoA) in Zambia. The RALS is a three-wave nationally representative panel from 2012 to 2019 with the second wave occurring in 2015. The survey contains plot-crop level data for 7,241 households cultivating less than 20 hectares of land. Our paper focuses on households who produced and sold crop output. This leaves us with 16,858 observations over the three waves of the survey.[31] The attrition between surveys implies that our analyses in this paper may be affected by attrition bias. Accordingly, we test for attrition bias using the regression-based test proposed by (Wooldridge, 2010). The methods and results from this test are presented in Table 3.10 of Appendix B. As shown in Table 3.10 of Appendix B, we fail to reject the null of no attrition bias— suggesting that attrition is random and not a major concern for our analyses.

In 2012, 9% of all households engaged in soybean production, allocating about 14% of cultivated land to soybean production. In 2015, the percent of households engaged in soybean production only rose to 11% and the share of land allocated to soybean production by these households rose to 16%. By 2019, both the share of households cultivating, and land allocated to soybean had increased significantly. About 19% of all households were involved in soybean production and the share of cultivated land allocated by these households rose to 22%. This implies that between 2012 and 2019 the number of households cultivating soybean more than doubled. Given our focus on the differential impacts of selling to different actors in the middle segment of

[31] As a robustness check, we redo our analysis by including households that produced but did not sell any output by assuming that they sold all of their agricultural output to a small-scale buyer at the prevailing median district price. The main study findings remain, and these results are available from the authors upon request.

the value chain, we will briefly discuss the classification of soybean buyers and the choice of sales outlets by farmers.

3.3.1 Classification of Large-Scale Buyers

While large-scale processing firms such a Mt. Meru Oils and the National Milling Corporation could easily be identified, the classification of traders into large-scale and small-scale is not straightforward.[32] To ensure accuracy, several measures were taken at the survey stage to train enumerators to easily classify traders as large-scale or small-scale. Burke et al. (2019) provide a detailed description of how large buyers are classified in the data used in this paper; we provide a summary of this description as relevant for this paper. Enumerators were trained to classify buyers of soybean as large if they met the following criteria:

a) The trader's volume of purchase is greater than those bought by the average trader in the area.
b) The trader aggregates at a village center rather than traveling to individual households to collect grains.
c) And to differentiate between small-scale traders and agents of large-scale traders (who are still classified large traders), respondents were asked if they sold to someone who was formally affiliated with a company since large-scale traders are often formally registered firms.

Conditional on all three criteria being met, the buyer was classified as a large-scale buyer. That said, it is still possible that mis-categorization of buyers may occur (if for example a farmer overestimates the average quantity purchased by a typical buyer) causing measurement error. This potential mis-categorization means that we would underestimate the effects of selling to large

[32] See Table 3.11 of Appendix B for an incomplete list of large soybean processing firms in Zambia. See Lubungu et al. (2013) for highlights on the activities of some of these soybean firms.

buyers if some of the small-scale buyers are consistently classified as large buyers especially if this mis-categorization is correlated with time-varying unobserved factors that affect welfare outcomes. However, we would overestimate the effects of selling to large buyers if large buyers are mistakenly classified as small-scale buyers. Thus, as a robustness check, we use a Two Stage Least Squares-Fixed Effects (2SLS-FE) technique to deal with the potential measurement error.

Among soybean producing households, the share selling to small-scale buyers (e.g., small-scale spatio-temporal arbitrageurs and retailers in the market) decreased consistently from 2012 to 2019 as the share sold to large-scale buyers rose from about 6% to 17% (see Figure 3.3). This represents a 283% increase in the share of farmers selling to large buyers in just over six years. While we do not have precise data on the intensity of large-scale trader/milling activities, aggressive investment activities of soybean processing firms such as Mt. Meru, Zamanita (Cargill) and Global Industries Limited are consistent with the rising share of output sold to large buyers.[33] Thus, as households gravitate towards engagement with large buyers, it becomes important to examine how this engagement impacts their welfare. In our analysis, we also distinguish between small-scale smallholders (operating less than 5 hectares) and medium-scale smallholders (operating between 5 and 20 hectares) to explore potentially heterogeneous welfare effects among smallholders.

It is also worth noting that the distribution of soybean production in Zambia is not geographically uniform, with the "soybean provinces" of Central, Eastern and Northern provinces currently making up over 70% of the soybean producing sample. In these provinces, the increasing share of farmers selling to the different actors within the value chain mimics that of the entire sample. As a robustness check we replicate our analysis restricting attention to these provinces.

[33] Global industries limited invested $20 million in soybean processing plant in Zambia in 2018 to double their processing capacity to 100,000 tons per day while Cargill also acquired Zamanita soybean crushing plant.

3.3.2 Key Study Variables

The household welfare outcomes we consider include (a) gross annual household income (b) gross crop income (c) the probability of being poor, (d) the household's poverty gap, (e) the household's poverty severity and (f) household food security. Gross annual household income (in 2010 Kwacha) is measured as the sum of household income from all observed sources, while gross crop income is a measure of the total value of all crops sold including soybean. To reduce the effect of outliers, we transform all income variables used in our estimation using a logarithmic function. The poverty measure is a categorical variable which equals one if the household's real per capita income (measured as gross annual household income divided by household size) is less than the \$1.90/day international poverty line and zero otherwise (Fung et al., 2019; Roser & Ortiz-Ospina, 2013).[34] The poverty gap (i.e., the amount needed to bring the household above the poverty line for all poor households as a proportion of the poverty line of \$1.90 and zero for non-poor households) measures the depth of poverty among the poor while poverty severity (square of the poverty gap) captures the severity of poverty among the poor (Foster et al., 1984). We include the poverty gap and severity measures because while participation in these new opportunities may not impact the likelihood of a household being poor in the short term, it may impact the depth of poverty, which subsequently affects the probability that the average household will move out of poverty in the longer run. This is particularly important in a context such as Zambia with such a high poverty rate.[35]

[34] To create the poverty measure, we first convert household income from nominal to real Kwacha using 2010 as the base year. We then convert the real Kwacha to United States Dollars (USD) using the Purchasing Power Parity (PPP) exchange rate for the given year.

[35] In this paper, we are unable to estimate longer-run effects due to data constraints

Finally, we include a subjective poverty measure equal to the reported number of months the household was food insecure over the last year.[36] While we do not have a priori expectation regarding the exact impact of participation in the large buyer segment of the value chain, we will expect participation to improve household welfare if it enhances households' ability to increase income and thus consumption expenditure (Timmer, 1988). However, selling to processors and traders can worsen welfare, especially, food insecurity as it exposes farmers to ex-post risks of substantial decrease in demand or commodity prices. The impacts of lower demand and prices can be severe for households in Zambia since soybean is not a staple food crop. In addition, household food security could be negatively affected when farmers invest more in cash crop than food production and food prices are high (Anderman et al., 2014).

Table 3.1 (panel A) presents the key study outcome variables. For the average household in our sample, the real annual income (as well as per capita income) decreased over time due to inflation, thus increasing the poverty rate and the amount required to bring these households above the poverty line. Thus, widening the poverty gap. Table 3.1 (panel B) presents the other key control variables. The average respondent is a middle-aged married male with about six years of education. Though maize remains the most important crop (accounting for 46% of cultivated land over the study period), the importance of soybean has significantly increased over the last decade. Among soybean producers, soybean income (as a share of total crop income) rose from 24% to 49% between 2012 and 2019.

[36] Due to data constraints, this was an attempt to generate a crude measure of food insecurity

3.4 Empirical Strategy

To estimate the impact of selling soybean to a large buyer on smallholder farmer's welfare, we start with the unobserved panel data model specification in (11)

$$y_{it} = \beta_1 S_{it} + \boldsymbol{\beta_2} X_{it} + c_i + \delta_t + \varepsilon_{it} \tag{11}$$

where y_{it} is the welfare outcome for household i in time t and S_{it} (the main variable of interest in this study) measures whether household i at time t sold soybean to large buyer in their largest soybean transaction. X_{it} is a vector of other covariates that could affect household welfare while c_i is time-invariant unobserved household-specific heterogeneity that could be correlated with the observed covariates. δ_t are year fixed effects which we control for using time dummies. β_1 is the average effect of a household's sale of soybean to a large buyer and $\boldsymbol{\beta_2}$ is a vector of parameters (associated with the various covariates) to be estimated.

A key challenge in obtaining unbiased and consistent estimates of β_1 in equation (1) is the potential endogeneity of household decision to sell soybeans to large buyers. The decision to sell soybeans to large buyers could be correlated with unobserved time invariant household characteristics, c_i, that could affect both the decision to sell to large buyers and household welfare. This could include farmer entrepreneurship or connections that could make them more likely to take advantage of particular marketing opportunities but also more likely to be productive and non-poor. Since our data on large buyers are self-reported by farmers, a natural question that arises is the potential for measurement error in reporting as farmers may misreport their market channel.[37]

To address these two likely sources of endogeneity, we estimate two versions of equation (1) using (i) household fixed effects and (ii) 2 Stage Least Squares-Fixed Effects (2SLS-FE). In

[37] At the survey collection point, several measures were taken by enumerators to minimize the potential for errors in the classification of trader as defined in section 3.1.

each of these estimations, we always include a rich set of observed time varying covariates (i.e., X_{it}) to account for time varying factors that could jointly affect household soybean sale decisions and welfare. The addition of these covariates effectively removes them from the error term to reduce the likelihood of omitted variables bias. Our choice of observed covariates is based on the agricultural household model (that recognizes the dual role of agricultural households as both producers and consumers of numerous agricultural products (Singh et al., 1986), supplemented by previous work on the determinants of poverty and food security in Zambia as well as determinants of farmer participation in modern value chains (Burke et al., 2019; Dumas et al., 2016; Fung et al., 2019). These include total land area allocated to soybean production, the total cultivated land area, distance from the household to the nearest market, distance to a hammer mill, distance to a tarmac road, household livestock measured in tropical livestock units, household head's educational attainment, age, gender, household size and a measure of shock proxied by whether a household member died in the last year. We also include district level variables such as median price of fertilizer in the district, the mean rainfall during the growing season, the total land area cultivated in the district to account for infrastructure and other district level advantages that may draw in large buyers.

The household fixed effects specification leverages on the panel nature of the data to control for time-invariant household-level unobservables, c_i, which are correlated with our outcome variables and the decision to sell to large buyers. We cluster our standard errors at the district level to account for any intra-cluster correlation in the decision to sell to a large buyer that might be induced by the intensity of processing firms' activities in a given district.[38] To account for the fact that there might still be some time varying unobservable factors driving the household

[38] For robustness, we redo our analysis by assuming the errors are correlated over time at the household by clustering standard errors at the household level. The estimates are similar and are available upon request.

sale decision and measurement error due to farmer mis-categorization of large buyers, we also adopt, the Two stage least squares Fixed effects (2SLS-FE) approach. The 2SLS-FE exploits the panel nature of the data to address household time-invariant unobservables while simultaneously exploiting the exogenous variation in our instrument (which we discuss next) to address time varying unobservables that may be correlated with both the household's decision to sell to a large buyer and the outcome variables. We treat this as a robustness check of our preferred fixed-effects estimates.

A valid instrument needs to satisfy two main conditions: the necessary relevance condition and the exclusion restriction. To satisfy the relevance condition, the instrument (in this case) must be strongly correlated with a household's decision to sell to a large buyer. In Zambia, large buyers of grains have often been attracted to areas where large surpluses are easily available for large buyers to purchase (Burke et al., 2019; N. J. Sitko et al., 2018). While these are often areas with significant shares of medium-scale farms (operating on 5 to 20ha of land) for staple grains such as maize (N. J. Sitko et al., 2018), the share of land allocated to soybean cultivation per farmer is relatively small with production spread on small plots of land across many smallholder households.[39] As a result, large buyers may be attracted to areas where a significant share of households are engaged in soybean production. Thus, allowing the buyers to aggregate output from these smallholder farmers. Our first stage estimates confirm that the relevance condition holds and is statistically significant at the 1% level (see Table 3.6 in Appendix B).[40] We also test whether our instrument improves market access in general or only improves the probability of selling to a large buyer. Our estimates on the probability of selling to a small-scale buyer is negative

[39] The average area of total cultivated land allocated to soybean among soybean producers is about 0.8ha (Table 1.1).
[40] While for the entire sample, the F>stat is slightly under 10, the F stat is significantly higher than 10 for the small and medium sale farmer samples and the coefficient on the instrumental variable are always significant at 1%.

confirming our argument that the instrument is not capturing market access in general but the opportunity to sell to large buyers. That is the argument that large buyers are attracted to areas where they can reduce per unit transaction cost.

To satisfy the second condition— the exclusion restriction, which is not testable but arguable, we argue that the share of other households engaged in soybean production in itself should not impact the household's welfare directly except through its ability to attract large buyers and offer the household's opportunity to sell to these large buyers. This is conditional on a rich set of farmer and district level controls which capture other factors that might encourage soybean production in a community (making it attractive to large buyers) and determine farmer wealth.

One could argue that as the share of households engaged in soybean production increase, off-farm employment opportunities may increase thereby affecting smallholder welfare. However, since soybean cultivation is still spread on relatively small plots of land across many smallholder farmers, we argue that its general equilibrium effects on agricultural wage employment and labor demand will be minimal. In addition, soybean has been documented to need very little fertilizer and other farm inputs because of its ability to absorb atmospheric nitrogen— thus reducing its indirect impact through other channels such as increased non-labor input demand (Albareda et al., 2009; Foyer et al., 2019). We also argue that reverse causality (another possible concern) is unlikely as it is improbable that household i's welfare derived from soybean income at the end of period t (the harvesting period) will impact the share of other households producing soybean at the beginning of period t. Additionally, since all our estimations also control for time-invariant household unobservables and district level socio-economic and agroecological variables that may affect the share of households engaged in soybean production, we argue that conditional on all of these control variables, our instrument is reasonably exogenous.

3.5 Results

Columns 1-3 of Tables 3.2 to 3.4 of Appendix A present the estimates from our main and preferred fixed effects (FE) specification on the three sets of welfare measures (i.e., income, poverty and food security).[41] In each table we also present the 2SLS-FE estimates in columns 3-6 (see Table 3.6 of Appendix B for the first stage results for the 2SLS-FE specification). For each outcome variable, we first estimate the effects of selling to a large buyer for the full sample of all farms. We then estimate heterogeneous effects separately for small-scale smallholders (operating on less than 5 ha) and medium-scale smallholders operating on 5-20 hectares of land (Burke et al., 2019; N. Sitko & Chamberlin, 2015). In the 2SLS-FE specification, we report the first-stage F-statistic as well as the Anderson and Rubin (1949) weak instrument test statistic and the confidence intervals under the null that the coefficients on the endogenous regressor is not statistically different from zero.[42]

3.5.1 Effects on Crop and Gross Household Income

Since gross crop income and gross annual household income for the household are log transformed, Table 3.2 of Appendix A presents both the estimated coefficients as well as the semi-elasticities computed from these coefficients. The estimated elasticity is the percentage change in income when a household sells soybean to a large buyer.[43] The preferred FE estimates in panel A show that selling to a large buyer has a large and statistically significant positive effect on crop income for all households. Specifically, selling soybean to large buyers is associated with about

---

[41] Full set of estimates with control variables can be found in Tables 1a to 4a of appendix A. The FE estimates are largely consistent with the 2SLS-FE estimates.

[42] In general, we find that in the few cases where our instrument passes the weak instrument test, the 2SLS-FE coefficients tend to be larger than the FE estimates. This suggests that our FE estimates are attenuated downwards by a systematic misclassification of large buyers as small-scale buyers.

[43] The semi-elasticity is given by $e^{(\hat{\beta})} - 1$, where $\hat{\beta}$ is the estimated coefficient on the "sold to large buyers" variable. The mean values of the dependent variable are reported in each table.

34% increase in crop income in the fixed effects estimates. The positive effects of selling to large buyers holds for both small and medium-scale smallholders although the magnitude of the effects are larger for medium-scale smallholders (32%) than for small-scale smallholders (20%). For small-scale farmers, this means that participation in the large buyers' segment of the soybean value chain is associated with about ZK 340 increase in average crop revenues while for medium-scale farms, this represents about ZK 1,680 in average crop revenues.

Similarly, our fixed effects estimates show a positive and statistically significant relationship between gross annual household income and selling soybean to a large buyer. Selling to a large buyer is associated with a 9% increase in gross annual household income for the average household which translates to an increase of about ZK 944. Although we find positive heterogeneous effects for both small-scale (4%) and medium-scale smallholders (7%) the estimated effect is not statistically significantly different from zero for the small-scale farmers (columns 1-3 of panel B). The effect for medium-scale farms could be driven by lower per unit sales transaction cost that comes with selling in bulk to large buyers, which allow households to allocate time and resources to other income-generating activities. It could also work through a higher crop income effect which offsets reduction in other income due as the household allocates relatively more resources to soybean production.

Where our instrument passes the weak instrument test (i.e., the full sample and medium-scale households), we find that the estimated coefficients from the 2SLS-FE on crop income and gross household income are qualitatively similar to the FE estimates. However, the magnitude of the coefficients tend to be larger. This suggests that the FE estimates are attenuated downwards by the measurement error. This could arise if small-scale traders are consistently misclassified as large-scale buyers by farmers; reducing the size of the difference in incomes between households who sold to large buyers' and those who did not.

3.5.2 Effects on Poverty and Food Security

Panels A, B and C of Table 3.3 of Appendix A present the estimates of the effects of selling to large buyers on the probability of being in poverty, household poverty gap and household poverty severity respectively. The FE estimates (columns 1-3 of panel A) show that selling soybean to large buyers is associated with a 5% decrease in the likelihood of being poor for the average household. However, this result seems to be largely restricted to medium-scale farms among whom selling to a large buyer is associated with a 9% decrease in the likelihood of being poor. This differential effect could be partly explained by the relatively high incidence of poverty (about 77%) among small-scale farmers as compared to medium-scale farmers operating on more than 5ha of land (47%).

Columns 1-3 of panel B present the FE estimates of the impact of selling to large buyers on poverty gap. Similar to the results in panel A, selling soybean to large buyers is associated with a statistically significant decrease in the poverty gap, though only significant for medium-scale farmers. Specifically, selling soybean to a large buyer reduces the poverty gap by about 3% for the average household and about 4% for medium-scale farmers. Similarly, we find that the decline in poverty severity is restricted to medium-scale farms among whom selling to large buyers is associated with a 2% decline in the severity of poverty.

We also present the 2SLS-FE estimates of the effects of selling to large buyers on the poverty outcomes in columns 4-6 of Table 3.3. In general, the estimated coefficients are similar to the FE estimates in terms of the direction of the signs and statistical significance although the magnitudes tend to be larger— following the trend observed in the income estimates. Our instrument passes the weak instrument test only in the case of the full sample and the medium-scale farmers in the poverty and poverty gap regressions. In these instances, the direction of impact is similar to the FE estimates although the magnitudes of the effects are larger for the 2SLS-FE—highlighting the argument that our FE estimates may be attenuated downwards by the measurement error. We also find a statistically significant decline in poverty severity for the full sample using the 2SLS-FE approach. However, we do not find any heterogenous effects by farm size.

Table 3.4 presents the estimates of the effects of selling to large buyers on the number of months in the previous year during which the household did not have enough food to meet family needs. While the coefficients are negative, they are not statistically significant in both the FE and 2SLS-FE estimates. A potential reason why we may be unable to uncover any effects on food insecurity is that our measure of food insecurity by construction allows us to only estimate the lower bound effects of food insecurity. This is because, the variable captures the number of months during which the household was without enough food to meet family needs. One way to deal with this problem is to measure the calorie availability per capita in the household. This, however, requires data on both retained calories from own production and calories from purchased foods which is not available in the data. In particular, because soybean is not a staple crop in Zambia, one can expect that households will barter it for food or money for food.[44]

3.5.3 Mechanisms

Finally, we attempt to understand the mechanisms through which the observed positive impacts on income and poverty occur. Hypothetically, there are several channels through which selling to large buyers (mills, processing firms and their authorized trading agents) may impact

[44] However, we explored the impact of selling to large buyers on total value of household production (both sold and unsold) and find a positive and significant effect of selling to large buyers for both small and medium-scale farms. These results were not included for space considerations but are available upon request.

household welfare. One major channel discussed in the literature is the provision of complimentary services. Large buyers of smallholder output may provide modern agricultural inputs to farmers at subsidized prices (or on credit), or training on input use and other agricultural practices (Liverpool-Tasie, Nuhu, et al., 2020; Liverpool-Tasie, Wineman, et al., 2020). This can directly impact the crop yield and subsequently incomes. In addition, large buyers, especially processing firms may require/support smallholder farmers to meet certain requirements in terms of input use and output characteristics by guaranteeing to purchase at a premium. Without details on farmer transactions with large buyers (beyond the sale decision), one can only speculate on these mechanisms. However, with data on prices and sale quantities, one can still determine if large buyers offer any price premiums or improved sale opportunities.

Thus, in the absence of detailed information on farmer transactions with different buyers in our data, we empirically test for any price effects and/or effects in terms of market opportunity. More specifically, we attempt to explain whether the observed effects are driven by an opportunity to sell more output, or a higher sale price and thus higher soybean revenue offered by large-scale buyers. We restrict our attention to households who sold soybean to any type of buyer. This is because, our outcome variables (kilogram of output sold, soybean income and the price received per kilogram) are only available for households producing soybean. We only use the household fixed effects estimation to identify these mechanisms since our instrument does not satisfy the relevance condition in this sample, potentially because the model is underpowered from reducing the sample. This implies that the mechanisms we discuss here are only in relation to the fixed effects estimates.

Table 3.5 presents these results, while the full set of estimates are reported in 5a of appendix A. We find that selling to large-scale buyers is associated with higher prices for small-scale farmers. Small-scale farmers who sold to a large buyer received a 15% price premium over those who did not. This is an important finding as it implies that small-scale farmers in non-contractual arrangements may be compensated for the risks involved in producing a cash crop without ex-ante contracts. In addition, the price premium might reflect small-scale producers selling to large buyers produce higher quality products compared to their counterparts who do not sell to large buyers. This might occur if these large buyers offer training or other complementary inputs (e.g., seeds) to these farmers as was found by Liverpool-Tasie et al. (2020) for smallholder farmers selling to medium scale farms in Nigeria. It is also possible given the intensive efforts by NGOs such as the USAID to encourage soybean cultivation (via similar mechanism of training and input provision) among smallholder farmers as a means to diversify their crop portfolio. Without detailed data on training and input provision by market channel, we are unable to test this hypothesis but highlight this as an important area for future research. We also find that selling to large buyers is associated with being able to sell more output and this effect holds for both small-scale and medium-scale farmers. The estimated quantity and price effects also translate into higher soybean income for both small and medium-scale farms. Specifically, selling to large buyers was associated with about 23% increase in soybean income for the average smallholder farmer.

Taken together, these results imply that being able to sell soybean output to large buyers is associated with higher crop income for the average Zambian smallholder. However, the effects on gross income are mostly statistically and economically significant only for medium-scale smallholders. Medium-scale farmers who sold their soybean output to medium-scale farms had higher total incomes that translated into statistically significant reductions in poverty and poverty severity. While the crop income effects of selling to large buyers did not translate to increased gross income and poverty reduction for small-scale households, the ability to sell to large buyers enables these small-scale farmers to sell more output at a higher price. Thus, efforts to increase the quantity produced and sold by these small-scale farmers might increase the likelihood of the crop income effect translating to an increase in total household income. Our findings of improved market access are consistent with Burke et al. (2019) who find that medium-scale farms are attracting large scale buyers into the vicinity of smallholder farmers and thus increasing market opportunities for them.

3.5.4 Robustness Checks and Additional Considerations

As a robustness check, we repeat the FE analysis for the soybean provinces— the Eastern, Central and Northern provinces of Zambia. Together, these provinces make up over 70% of soybean output in our sample and also in Zambia (Lubungu et al., 2013). Thus, they may possess agro ecological conditions better suited for soybean cultivation and thus attract more large-scale buyers relative to other provinces. If this is true, then our estimates may also be capturing the opportunity to grow soybean rather than just selling output to large buyers. Thus, we restrict our analysis to the soybean provinces to test whether our results hold among the subset of households who face the same agroecological conditions for growing soybean. Our results for these soybean provinces (reported in Table 3.8 in Appendix B) are similar to the main results on income and poverty presented in Tables 3.3-3.5.[45]

We also test the sensitivity of our findings to the inclusion of average transport cost to a large buyer as an extra control. This variable was omitted from the main specification because we lose some of our sample (29%) since the variable is not available for all districts. The estimates from including average transport (while holding soybean cultivation fixed) are also consistent with

[45] We also re-estimate Tables 3.2-3.4 (but do not report) by including all households that produced but did not sell any output by assuming that they sold all of their agricultural output to a small-scale buyer at the median district price. Although the estimated elasticities are smaller, they are generally consistent with our findings

the main results in Tables 3.3-3.5 although our estimates are underpowered by a loss of sample in some specifications. Next, we examine whether the effects are driven by just the production of soybean or the opportunity to sell soybean to large buyers. To do this, we restrict our attention to only households that are engaged in soybean production at any point within the study period. These results (reported in Table 3.9 in Appendix B) confirm that even among only households that cultivated soybeans at any point within the study period, the effects (while holding soybean cultivation fixed) are driven by the opportunities presented by large buyers. In other words, the sheer cultivation of soybean is not what drives the observed effects but the opportunity to also sell output to large buyers.

3.6 Conclusion and Policy Recommendations

For smallholder farmers to benefit from the rapidly evolving agrifood value chains in developing countries, research will have to keep pace with the rapid transformation of food systems. There is still a dearth of data and empirical studies on the activities of the rapidly expanding middle segment of the food system which links farmers to the rest of the modern food supply chains. This gap in the literature implies that existing policy debates and recommendations for new public and private initiative regarding the impact of this aspect of the food supply chain on smallholder farmers will remain largely descriptive and speculative. This paper contributes to the limited literature on this important issue with empirical evidence from the Zambian soybean value chain; a rapidly growing industry due to rising domestic and regional demand for meat and animal feed. This paper uses a unique panel survey dataset (the Zambia Rural Agricultural Livelihoods Survey) to investigate the potential impact of the recent rise in non-contract-based purchasing activities of large soybean processing firms and traders on the welfare of smallholder farmers in Zambia.

We find that many large buyers (comprising mostly large-scale traders alongside numerous soybean processing firms) have increased the intensity of their activities in Zambia. Using both fixed effects and instrumental variables approaches (to address the endogeneity of farmers' market channel choices), we find statistically and economically significant positive effects of selling to large buyers in non-contract arrangements on crop incomes for smallholders in Zambia. However, for small-scale households, the increase in crop income does not translate into higher total household income (and thus lower probability of being poor) as it does for medium-scale smallholders. We find that the observed effects of selling to large buyers are largely driven by the 'opportunity to sell' although small-scale farmers receive a price premium from selling to these large buyers.

Our findings have significant implications for policy. First, we were able to show that the rise of large-scale processing firms and their agents in the midstream of the soybean value chain can be beneficial to rural smallholders even in the absence of contracts. Thus, government policies to support their efficient operation are likely to benefit smallholder farmers especially as they continue to allocate more land to soybean production. This includes policies that reduce transaction costs for both smallholder farmers and large buyers who are expanding into rural areas. Second, our study's finding that the poverty-reducing effects largely accrue to medium-scale smallholders (not small-scale households) implies that while small-scale households are able to benefit from improved commercialization opportunities, additional efforts might be necessary to ensure that the increase in their crop incomes translates to increased total household income. This could include policies and programs that support expanded soybean output and sale by small-scale farmers.

Because of data constraints, we are only able to estimate a small subset of potential mechanisms of improved welfare via interactions with large buyers. We were not able to explore other mechanisms such as improvements in smallholder yield due to complementary services (such as training or production inputs) received from large buyers. Thus, while our study reveals some positive welfare impact of the rise of large processing firms and their trading agents in the Zambian soybean value chain, further research on other mechanisms through which this occurs is needed. Thus, more research is needed to understand why higher crop incomes from selling to large buyers (and the price premium it offers) does not translate into significantly higher incomes and reduction in poverty levels for small-scale smallholders.

APPENDICES

APPENDIX A

Tables and Figures

Table 3.1 Summary statistics for key variables

	2012	2015	2019
Panel A: Outcome variables			
Gross annual household income (2010 ZK)	12,226.41	11,990.81	10,464.34
	(11,901.86)	(12,223.50)	(11,505.88)
Gross annual crop income (2010 ZK)	2,919.07	2,672.75	1,822.09
	(2,960.18)	(2,847.46)	(2,308.47)
Soyabean income (soybean producers only, 2010 ZK)	747.87	934.51	1,250.97
	(1,392.87)	(1,508.10)	(3,266.85)
Soybean income as a share of crop income (producers only)	0.24	0.28	0.49
	(0.26)	(0.29)	(0.42)
Gross off-farm income (2010 ZK)	6,101.94	7,686.04	7,043.44
	(26,598.61)	(26,155.59)	(19,480.53)
Household is below the poverty line (0/1)	0.61	0.72	0.83
	(0.49)	(0.45)	(0.38)
Poverty gap	0.30	0.41	0.54
	(0.30)	(0.33)	(0.33)
Poverty severity	0.18	0.27	0.40
	(0.23)	(0.28)	(0.30)
Number of months food insecure	1.03	1.25	1.13
	(1.66)	(1.87)	(1.81)
Kilograms (kg) of soybean sold	27.77	54.27	170.30
	(219.00)	(276.74)	(1,101.10)
Price per kg of soybean sold (2010 ZK)	2.37	2.43	3.03
	(1.34)	(0.97)	(1.11)
Panel B: Other household characteristics			
Share of cultivated land allocated to maize	0.47	0.47	0.43
	(0.28)	(0.28)	(0.25)
Share of cultivated land allocated to soybean (all sample)	0.01	0.02	0.04
	(0.06)	(0.07)	(0.11)
Share of cultivated land allocated to soybean (soybean producers)	0.14	0.16	0.22
	(0.13)	(0.13)	(0.14)
Hectares allocated to soybean (soybean producers)	0.51	0.68	1.00
	(0.64)	(0.88)	(1.94)
Hectares allocated to soybean (all sample)	0.05	0.07	0.19
	(0.24)	(0.35)	(0.94)
Household grew soybean in the year (0/1)	0.09	0.11	0.19

Table 3.1 (Cont'd)

	(0.28)	(0.31)	(0.39)
Total land size (hectares)	3.38	3.49	3.62
	(3.33)	(3.47)	(4.53)
Household has off-farm agricultural wage job (0/1)	0.07	0.13	0.13
	(0.26)	(0.33)	(0.33)
Household has non-farm employment (0/1)	0.12	0.16	0.16
	(0.32)	(0.36)	(0.37)
Tropical livestock unit	2.81	2.91	3.06
	(8.13)	(8.35)	(7.73)
Household head is male (0/1)	0.84	0.82	0.79
	(0.37)	(0.39)	(0.40)
Household head is married (0/1)	0.72	0.62	0.67
	(0.45)	(0.49)	(0.47)
Household head years of education	6.27	6.22	6.23
	(3.78)	(3.79)	(3.70)
Age of household head	45.98	48.31	51.25
	(14.40)	(14.29)	(13.68)
Household size	6.16	6.42	6.65
	(2.71)	(2.74)	(2.82)
Death of household member in the last year (0/1)	0.16	0.02	0.03
	(0.37)	(0.15)	(0.18)
Mean rainfall (mm)	124.40	113.33	115.81
	(23.60)	(22.31)	(35.88)
Coefficient of variation in rainfall (mm)	0.18	0.07	0.19
	(0.07)	(0.03)	(0.07)
Household distance to output market (in KM)	26.63	24.99	24.61
	(32.09)	(30.28)	(30.46)
Number of observations	5,594	5,921	5,343

Table 3.2 FE and 2SLS-FE estimates of the effect of selling to large buyers on farmer incomes

Panel A:	**FE Estimates**			**2SLS-FE Estimates**		
Log of crop income	(1) All	(2) Small	(3) Medium	(4) All	(5) Small	(6) Medium
Sold to large buyer (0/1)	0.29***	0.18**	0.28***	0.63***	0.67	0.56*
	(0.06)	(0.07)	(0.07)	(0.19)	(0.68)	(0.33)
Semi-elasticity (%)	33.69	20.00	32.10	88.58	95.55	75.38
F-statistic	34.80***	69.81***	11.09***	57.40***	4.801***	39.24***
A-R Wald statistic				11.00***	1.39	3.01*
A-R Confidence interval				0.25, 1.03	-0.75, 2.31	-0.08, 1.31
R-squared	0.19	0.24	0.04	0.17	0.21	0.20
Mean of income (ZK)	2,202	1,240	7, 342	2,202	1,240	7, 342
Panel B:						
Log of total income	**FE Estimates**			**2SLS-FE Estimates**		
Sold to large buyer (0/1)	0.09**	0.04	0.07*	0.38**	-0.48	0.09
	(0.04)	(0.04)	(0.04)	(0.15)	(0.57)	(0.27)
Semi-elasticity (%)	9.12	3.68	7.09	45.50	-38.37	9.36
F-stat	47.75***	73.08***	18.51***	57.39***	4.800***	39.24***
A-R Wald statistic				6.50***	0.64	0.12
A-R Confidence interval				0.08, 0.68	-2.01, 0.57	-0.08,1.31
R-squared	0.32	0.30	0.13	0.13	0.16	0.14
Mean of income (ZK)	10,489	8,095	20,532	10,489	8,095	20,532
Number of observations	16,853	13,478	3,375	16,853	13,478	3,375

Note: Robust standard errors clustered at the district level in parenthesis. In each panel, each column represents a separate regression. Each regression includes household and district level time-varying controls, as well as household and year fixed effects. Controls used in the regressions include total area of land allocated to soybean, total land area cultivated, off-farm agricultural wage employment dummy, non-agricultural wage employment dummy, distance to nearest market, tarmac road and hammer mill from household, a measure of tropical livestock unit, male household head dummy, married household head dummy, household head years of education, household size, age of household head, death of a household member in the last year, mean rainfall during the crop growing period, median district fertilizer price, total land area allocated to soybean in the district and total land area cultivated in district. . ** * p<0.01, ** p<0.05, * p<0.1

Table 3.3 FE and 2SLS-FE estimates of the effect of selling to large buyers on poverty

PANEL A:	**FE Estimates**			**2SLS-FE Estimates**		
Household is poor (0/1)	(1)	(2)	(3)	(4)	(5)	(6)
	All	Small	Medium	All	Small	Medium
Sold to large buyer (0/1)	-0.05***	-0.03	-0.09***	-0.41**	-0.53	-0.39**
	(0.02)	(0.03)	(0.03)	(0.187)	(0.597)	(0.185)
Mean of dependent variable	0.71	0.77	0.47			
A-R Wald statistic				8.107***	1.605	10.04***
A-R Confidence interval				-0.78, -0.12	-	-0.75, -0.02
					2.331*	
F-statistic	130.50***	43.98***	36.79***	7.73***	*	13.60***
PANEL B:						
Poverty Gap						
Sold to large buyer (0/1)	-0.03**	-0.004	-0.04***	-0.24**	-0.11	-0.10**
	(0.01)	(0.02)	(0.01)	(0.114)	(0.257)	(0.052)
Mean of dependent variable	0.41	0.46	0.21			
A-R Wald statistic				6.97***	0.20	3.82**
A-R Confidence interval				-0.46, -0.02	-	-0.30,0.27
F-statistic	188.50***	390.4***	49.43***	7.73***	2.33**	13.60***
PANEL C:						
Poverty Severity						
Sold to large buyer (0/1)	-0.02	0.00	-0.02**	-0.18*	0.07	-0.03
	(0.01)	(0.01)	(0.01)	(0.097)	(0.239)	(0.045)
Mean of dependent variable	0.28	0.32	0.13			
A-R Wald statistic				6.155***	0.0815	0.443
A-R Confidence interval				-0.38, 0.03	-	-0.15, 0.12
F-statistic	115.30***	202.30***	32.05***	7.73***	2.33**	13.60***
Number of observations	16,858	13,482	3,376	16,858	13,482	3,376

Note: Robust standard errors clustered at the district level in parenthesis. In each panel, each column represents a separate regression. Each egression includes household and district level time-varying controls, as well as household and year fixed effects. Controls used in the regressions include total area of land allocated to soybean, total land area cultivated, off-farm agricultural wage employment dummy, non-agricultural wage employment dummy, distance to nearest market, tarmac road and hammer mill from household, a measure of tropical livestock unit, male household head dummy, married household head dummy, household head years of education, household size, age of household head, death of a household member in the last year, mean rainfall during the crop growing period, median district fertilizer price, total land area allocated to soybean in the district and total land area cultivated in district. ** * p<0.01, ** p<0.05, * p<0.1

Table 3.4 FE and 2SLS-FE estimates of the effect of selling to large buyers on food security

Panel A: Number of months being food insecure	**FE Estimates**			**2SLS-FE Estimates**		
	(1) All	(2) Small	(3) Medium	(4) All	(5) Small	(6) Medium
Sold to large buyer (0/1)	-0.05	-0.02	-0.07	-0.19	-1.44	-0.24
	(0.07)	(0.13)	(0.11)	(0.61)	(2.28)	(0.63)
Mean of dependent variable	1.12	1.28	0.72	1.12	1.28	0.72
A-R Wald statistic				4.19***	0.05	0.34
A-R Confidence interval				-1.67,1.25	-5.92, 3.54	-1.50,1.87
F-statistic	6.51***	6.26***	3.27***	21.88***	3.88***	17.31***
Number of observations	16,858	13,482	3,376	16,858	13,482	3,376

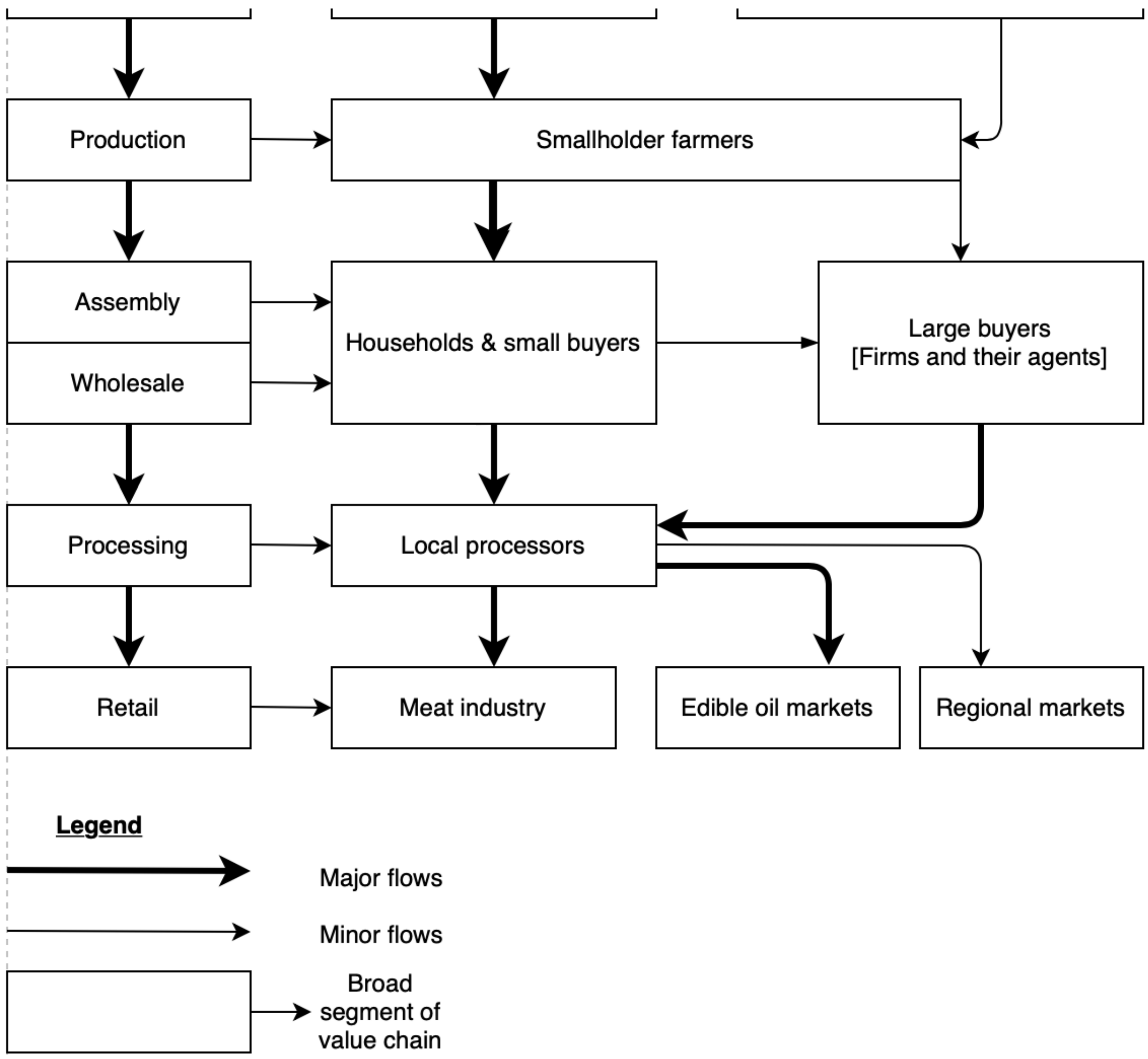
Note: Robust standard errors clustered at the district level in parenthesis. In each panel, each column represents a separate regression. Each egression includes household and district level time-varying controls, as well as household and year fixed effects. Controls used in the regressions include total area of land allocated to soybean, total land area cultivated, off-farm agricultural wage employment dummy, non-agricultural wage employment dummy, distance to nearest market from household, a measure of tropical livestock unit, male household head dummy, married household head dummy, household head years of education, number of children [5-16], number of adults[17+] on the household, age of household head, death of a household member in the last year, coefficient of variation in district rainfall (rainfall shocks) during the crop growing period, share of farms that are medium-scale in district and total land area cultivated in district. ** * $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.5 FE estimates of mechanisms through which selling to a large buyer affects welfare

	Log price/kg (ZK)			Log quantity sold (Kg)			Log soybean income (ZK)		
	All	small	medium	All	small	medium	All	Small	medium
Sold to large buyer (0/1)	0.05	0.15**	0.03	0.21***	0.16*	0.23*	0.21***	0.16*	0.24*
	(0.03)	(0.06)	(0.07)	(0.06)	(0.09)	(0.12)	(0.06)	(0.09)	(0.13)
Mean of dep. variable	2.7	2.7	2.8	578	395	981	1,033	591	2,013
R-squared	0.08	0.06	0.04	0.26	0.22	0.15	0.31	0.28	0.20
Number of observations	1,862	1,282	580	1,862	1,282	580	1,862	1,282	580

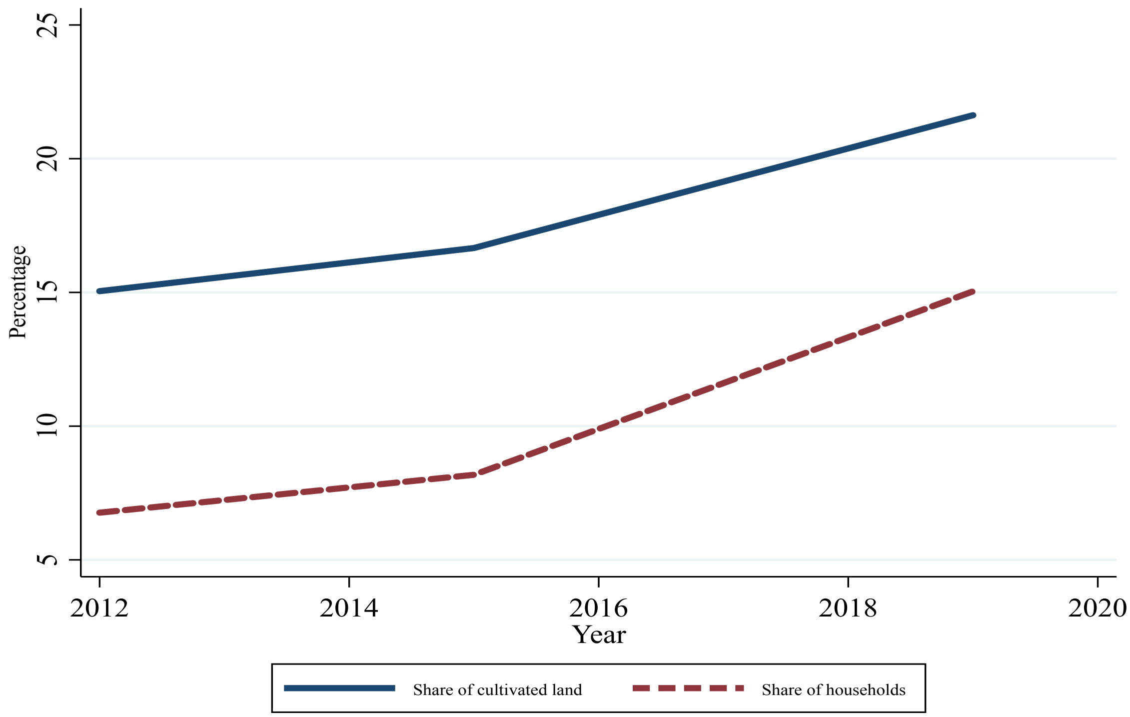
Note: Robust standard errors clustered at the district level in parenthesis. Each column represents a separate regression. Each egression includes household and district level time-varying controls, as well as household and year fixed effects. Controls used in the regressions include total area of land allocated to soybean, total land area cultivated, off-farm agricultural wage employment dummy, non-agricultural wage employment dummy, distance to nearest market from household, a measure of tropical livestock unit, male household head dummy, married household head dummy, household head years of education, household size, age of household head, death of a household member in the last year, coefficient of variation in district rainfall (rainfall shocks) during the crop growing period, share of farms that are medium-scale in district and total land area cultivated in district. ** * p<0.01, ** p<0.05, * p<0.1

Figure 3.1 Soybean value chain in Zambia



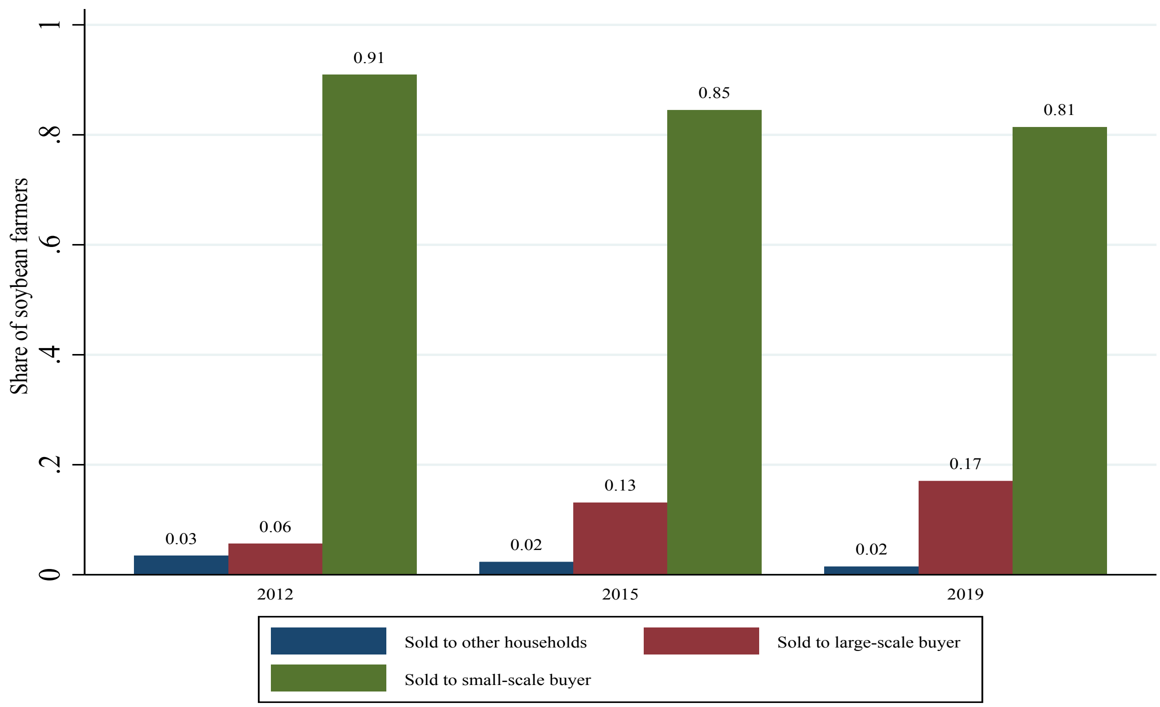
Source: Author's adapted version of the Soybean value chain in Lubungu et al., (2013).

Figure 3.2 Growth of soybean cultivation in Zambia



Notes: This figure documents the growth of soybean cultivation between 2012 and 2019. The dashed shows the share of households who cultivate soybean in the sample during the period under study while the solid line shows the share of cultivated land dedicated by soybean producers to soybean production.

Figure 3.3 Level of engagement in the midstream among producing households



Note: This figure presents the share of soybean producers who sold soybeans through the three main outlets: direct consumers (blue), large scale buyers(red) and small-scale buyers(green). The share of soybean output sold to large scale buyers more than tripled from 6% to 17% in just over six years as the share sold to small-scale buyers reduced from 91 to 81%.

APPENDIX B

Supplemental Tables

Table 3.6 First stage estimates for two stage least squares fixed effects (2SLS-FE) estimation

Dependent variable	Sold output to:			
		Large buyer		Small buyer
Sample	All	Small	Medium	All
	(1)	(2)	(3)	(4)
Share of other households producing soybean	0.43***	0.89***	0.89***	-0.52***
	(0.14)	(0.31)	(0.24)	(0.16)
Land allocated to soybean	0.04***	0.25***	0.02***	-0.04***
	(0.01)	(0.03)	(0.01)	(0.01)
Total land size	-0.00	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Household is engaged in off farm ag employment (0/1)	0.00	0.00	0.01	-0.01**
	(0.00)	(0.00)	(0.02)	(0.00)
Household is engaged in non-ag employment. (0/1)	-0.01	-0.01	-0.00	0.01
	(0.01)	(0.01)	(0.02)	(0.01)
Tropical livestock unit	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Household head is male (0/1)	-0.00	0.00	0.05	0.00
	(0.01)	(0.01)	(0.03)	(0.01)
Household head is married	-0.00	-0.00	0.00	-0.00
	(0.01)	(0.01)	(0.02)	(0.01)
Household head years of education	0.00	-0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Household head age	0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Household size	-0.00	0.00	-0.01	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Death in the last year	0.01	0.00	0.05*	-0.01
	(0.01)	(0.00)	(0.02)	(0.01)
Household distance to market	-0.00	0.00	-0.00*	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Coefficient of variation in rainfall	-0.02	-0.04	0.08	0.01
	(0.04)	(0.03)	(0.12)	(0.04)
Total land holding in district	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share of medium scale farms in district	0.04	0.01	-0.09	-0.05
	(0.06)	(0.03)	(0.11)	(0.06)

Table 3.6 (Cont'd)

Household distance to hammer mill	-0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Household distance to tarmac road	0.00	-0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Mean annual rainfall in district (mm)	-0.00***	-0.00*	-0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Total soybean area in district	-0.00	0.00	-0.00**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Median fertilizer price in district (2010 ZK)	-0.02**	-0.00	-0.05**	0.02
	(0.01)	(0.01)	(0.02)	(0.01)
year = 2015	0.02***	0.00	0.07***	-0.01*
	(0.01)	(0.01)	(0.03)	(0.01)
year = 2019	0.03***	0.00	0.06**	-0.01
	(0.01)	(0.01)	(0.03)	(0.01)
Constant	0.16***	0.06*	0.41***	0.11*
	(0.05)	(0.03)	(0.13)	(0.06)
Observations	16,858	13,482	3,376	16,858
R-squared	0.11	0.15	0.19	0.11
F-stat	7.97	34.72	46.59	10.07

Table 3.7 First stage estimates for two stage least squares fixed effects (2SLS-FE) estimation for 'Soybean Provinces' only

Soybean Provinces Only	Sold to large buyer	Sold to small buyer
	(5)	(6)
Share of other households producing soybean	0.44***	-0.47***
	(0.04)	(0.04)
Land allocated to soybean	0.03***	-0.03***
	(0.01)	(0.01)
Total land size	-0.00	0.00*
	(0.00)	(0.00)
Household is engaged in off farm ag employment (0/1)	0.00	-0.01
	(0.01)	(0.01)
Household is engaged in non-ag employment. (0/1)	-0.02	0.02
	(0.01)	(0.01)
Tropical livestock unit	0.00	-0.00
	(0.00)	(0.00)
Household head is male (0/1)	-0.01	0.01
	(0.02)	(0.02)
Household head is married	0.01	-0.01
	(0.01)	(0.01)
Household head years of education	0.00	0.00
	(0.00)	(0.00)
Household head age	0.00	-0.00
	(0.00)	(0.00)
Household size	-0.00	0.00
	(0.00)	(0.00)
Death in the last year	0.02	-0.02
	(0.01)	(0.01)
Household distance to market	-0.00	0.00
	(0.00)	(0.00)
Coefficient of variation in rainfall	-0.03	0.02
	(0.09)	(0.10)
Total land holding in district	0.00	-0.00
	(0.00)	(0.00)
Share of medium scale farms in district	0.08	-0.10
	(0.07)	(0.07)
Household distance to hammer mill	-0.00	-0.00
	(0.00)	(0.00)
Household distance to tarmac road	0.00	0.00
	(0.00)	(0.00)

Table 3.7 (Cont'd)

Mean annual rainfall in district (mm)	-0.00***	0.00***
	(0.00)	(0.00)
Total soybean area in district	0.00	-0.00
	(0.00)	(0.00)
Median fertilizer price in district (2010 ZK)	-0.15***	0.14***
	(0.03)	(0.03)
year = 2015	0.06***	-0.06***
	(0.02)	(0.02)
year = 2019	0.16***	-0.15***
	(0.03)	(0.03)
Constant	0.76***	-0.36***
	(0.12)	(0.13)
Observations	8,215	8,215
R-squared	0.12	0.12
F-stat	11.52	12.12

Note Columns 5 and 6 restrict the first stage to the Eastern, Central and Northern provinces which account for over 80% of the soybean production.
** * p<0.01, ** p<0.05, * p<0.1

Table 3.8: FE estimates of selling to large buyers on welfare outcomes for households in Eastern, Central and Northern Provinces

Panel A	Poverty status (0/1)			Poverty gap			Poverty severity		
	All	small	medium	All	small	medium	All	small	medium
Sold to large buyer (0/1)	-0.05**	-0.02	-0.09***	-0.03*	0.00	-0.03***	-0.02	0.01	-0.02***
	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
R-squared	0.22	0.16	0.15	0.32	0.33	0.16	0.32	0.35	0.15
Number of observations	8,215	6,605	1,610	8,215	6,605	1,610	8,215	6,605	1,610

Panel B	Crop income (ZK)			Gross annual income (ZK)		
	All	small	medium	All	small	medium
Sold to large buyer (0/1)	0.27***	0.11	0.22***	0.10**	0.02	0.07*
	(0.05)	(0.07)	(0.08)	(0.04)	(0.04)	(0.04)
R-squared	0.28	0.31	0.31	0.35	0.28	0.28
Number of observations	8,215	6,605	1,610	8,215	6,605	1,610

Panel C	Number of months food insecure					
	All	small	medium			
Sold to large buyer (0/1)	-0.02	0.01	-0.03			
	(0.07)	(0.15)	(0.10)			
R-squared	0.03	0.03	0.06	0.03	0.04	0.06
Number of observations	8,215	6,605	1,610	8,215	6,605	1,610

Note: Robust standard errors clustered at the district level in parenthesis. In each panel, each column represents a separate regression. Each egression includes household and district level time-varying controls, as well as household and year fixed effects. Controls used in the regressions include total area of land allocated to soybean, total land area cultivated, off-farm agricultural wage employment dummy, non-agricultural wage employment dummy, distance to nearest market from household, a measure of tropical livestock unit, male household head dummy, married household head dummy, household head years of education, household size/number of children [5-16], number of adults[17+] on the household, age of household head, death of a household member in the last year, coefficient of variation in district rainfall (rainfall shocks) during the crop growing period, share of farms that are medium scale in district and total land area cultivated in district. ** * p<0.01, ** p<0.05, * p<0.1

Table 3.9 FE estimates of selling to large buyers on welfare outcomes for soybean-producing households only

PANEL A									
	Poverty status (0/1)			Poverty gap			Poverty severity		
	All	Small	medium	All	small	medium	All	small	medium
Sold to large buyer (0/1)	-0.06***	-0.03	-0.09***	-0.02**	-0.00	-0.03**	-0.01	0.00	-0.01
	(0.020)	(0.026)	(0.027)	(0.011)	(0.015)	(0.013)	(0.010)	(0.013)	(0.010)
R-squared	0.20	0.12	0.10	0.29	0.29	0.12	0.29	0.32	0.13
Number of observations	4,044	2,917	1,127	4,044	2,917	1,127	4,044	2,917	1,127
PANEL B	Crop income (ZK)			Gross annual income (ZK)					
	All	small	medium	All	small	medium			
Sold to large buyer (0/1)	0.26***	0.17**	0.24***	0.08**	0.03	0.06			
	(0.047)	(0.065)	(0.073)	(0.035)	(0.045)	(0.036)			
R-squared	0.27	0.29	0.29	0.29	0.21	0.21			
Number of observations	4,042	2,916	1,126	4,044	2,917	1,127			
PANEL C	Number of months food insecure								
	All	small	medium						
Sold to large buyer (0/1)	-0.03	0.02	-0.07						
	(0.093)	(0.146)	(0.133)						
R-squared	0.02	0.03	0.06						
Number of observations	4,042	2,916	1,126						

Notes: Robust standard errors clustered at the district level in parenthesis. In each panel, each column represents a separate regression. Each egression includes household and district level time-varying controls, as well as household and year fixed effects. Controls used in the regressions include total area of land allocated to soybean, total land area cultivated, off-farm agricultural wage employment dummy, non-agricultural wage employment dummy, distance to nearest market from household, a measure of tropical livestock unit, male household head dummy, married household head dummy, household head years of education, number of children [5-16], number of adults[17+] on the household, age of household head, death of a household member in the last year, coefficient of variation in district rainfall (rainfall shocks) during the crop growing period, share of farms that are medium scale in district and total land area cultivated in district. ** * p<0.01, ** p<0.05, * p<0.1

Table 3.10 Regression-based attrition bias test

	Log crop income	Log household income	Below poverty line (0/1)	Poverty gap	Poverty severity
S_{it+1}	0.059	-0.045	0.048**	0.024*	0.012
	(0.052)	(0.035)	(0.021)	(0.013)	(0.011)
Number of observations	7,532	7,532	7,532	7,532	7,532

	Num. of months food insecure	Ever food insecure
S_{it+1}	-0.118	-0.024
	(0.084)	(0.021)
Number of observations	7,532	7,532

Note: Robust standard errors clustered at the district level in parenthesis. In each panel, each column represents a separate regression. Each egression includes household and district level time-varying controls, as well as household and year fixed effects. Controls used in the regressions include attrition rate in each cluster, total area of land allocated to soybean, total land area cultivated, off-farm agricultural wage employment dummy, non-agricultural wage employment dummy, distance to nearest market from household, a measure of tropical livestock unit, male household head dummy, married household head dummy, household head years of education, number of children [5-16], number of adults[17+] on the household, age of household head, death of a household member in the last year, coefficient of variation in district rainfall (rainfall shocks) during the crop growing period, share of

Table 3.11 Incomplete list of large-scale soybean purchasers in Zambia

List of stock feed companies		
Province	**District**	**Feeder producer**
Lusaka	Lusaka	Tiger Animal Feeds
Lusaka	Lusaka	Novatake
Lusaka	Lusaka	Olympic Stockfeed
Lusaka	Lusaka	Yielding Feeds Ltd
Lusaka	Lusaka	Animal Protein
Lusaka	Lusaka	Meadow feeds
Lusaka	Chongwe	Nutri feed (Z) Ltd
Southern	Choma	Chilumbwe Agri feed and Foods
List of edible oil companies		
Province	**Districts**	**Name of edible oil company**
Central	Chibombo	Mt Meru Oils
Copperbelt	Ndola	Global Industries Ltd
Central	Kabwe	Supa Oil Industry
Lusaka	Lusaka	Unified Chemicals
Lusaka	Lusaka	Parrogate Ginnery
Lusaka	Lusaka	Seba Foods Ltd
Eastern	Chipata	China Africa Cotton
List of large-scale grain traders who also buy soybeans		
Province	**District**	**List of Grain Traders**
Lusaka	Lusaka	CHC Commodities
Lusaka	Lusaka	Zdenakie Limited
Lusaka	Lusaka	AFGRI Corporation
Lusaka	Lusaka	Exporting Trading Company
Lusaka	Lusaka	NWK/LDC (Z) Ltd
Lusaka	Lusaka	Cargill (Z) Ltd
Lusaka	Lusaka	Quality Commodities
Lusaka	Lusaka	Nachel Distributors
Lusaka	Lusaka	Nyimba Filling Station
Lusaka	Lusaka	Olam/Continental Ginneries Zambia
Lusaka	Lusaka	Farmers Builders Suppliers
Lusaka	Lusaka	Bulero General Dealers

REFERENCES

REFERENCES

Albareda, M., Rodríguez-Navarro, D. N., & Temprano, F. J. (2009). Soybean inoculation: Dose, N fertilizer supplementation and rhizobia persistence in soil. *Field Crops Research, 113*(3), 352-356.

Anderman, T. L., Remans, R., Wood, S. A., DeRosa, K., & DeFries, R. S. (2014). Synergies and tradeoffs between cash crop production and food security: a case study in rural Ghana. *Food Security, 6*(4), 541-554.

Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics, 20*(1), 46-63.

Andersson, C. I., Chege, C. G., Rao, E. J., & Qaim, M. (2015). Following up on smallholder farmers and supermarkets in Kenya. *American Journal of Agricultural Economics, 97*(4), 1247-1266.

Barrett, C. B., Bachke, M. E., Bellemare, M. F., Michelson, H. C., Narayanan, S., & Walker, T. F. (2012). Smallholder participation in contract farming: comparative evidence from five countries. *World Development, 40*(4), 715-730.

Bellemare, M. F., & Novak, L. (2017). Contract farming and food security. *American Journal of Agricultural Economics, 99*(2), 357-378.

Burke, W. J., Jayne, T. S., & Sitko, N. J. (2019). Do Medium-scale Farms Improve Market Access Conditions for Zambian Smallholders? *Journal of Agricultural Economics*, 71(2) 517-533

Burney, J. A., & Naylor, R. L. (2012). Smallholder irrigation as a poverty alleviation tool in sub-Saharan Africa. *World Development, 40*(1), 110-123.

Chisanga, B., & Sitko, N. (2013). The National Consumption Requirement Study for Edible Oils in Zambia. *Indaba Agricultural Policy Research Institute (IAPRI), Lusaka*.

Christiaensen, L., Demery, L., & Kuhl, J. (2011). The (evolving) role of agriculture in poverty reduction—An empirical perspective. *Journal of Development Economics, 96*(2), 239-254.

Davids, T., Meyer, F., & Westhoff, P. (2017). Impact of trade controls on price transmission between southern African maize markets. *Agrekon, 56*(3), 223-232.

Deininger, K., & Byerlee, D. (2011). Rising global interest in farmland: can it yield sustainable and equitable benefits? *The World Bank. Retrieved from:*

https://openknowledge.worldbank.org/bitstream/handle/10986/2263/594630PUB0ID1810Box358282B01PUBLIC1.pdf?sequence=1&isAllowed=y Accessed on: 29th December 2020

Dumas, S. E., Lungu, L., Mulambya, N., Daka, W., McDonald, E., Steubing, E., . . . Lucio-Martinez, B. (2016). Sustainable smallholder poultry interventions to promote food security and social, agricultural, and ecological resilience in the Luangwa Valley, Zambia. *Food Security, 8*(3), 507-520.

Ehrich, M., & Mangelsdorf, A. (2018). The role of private standards for manufactured food exports from developing countries. *World Development, 101*, 16-27.

Euler, M., Krishna, V., Schwarze, S., Siregar, H., & Qaim, M. (2017). Oil palm adoption, household welfare, and nutrition among smallholder farmers in Indonesia. *World Development, 93*, 219-235.

Ferdinand, M., Ndibongo, T. L., Tracy, D., Brian, C., Richard, K., Emilio, T., . . . Pierre, B. (2018). Modelling soybean markets in Eastern and Southern Africa. *Regional Network of Agricultural Policy Research Institutes (ReNAPRI) Working Paper*, June 2018.

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica: journal of the econometric society*, 761-766.

Foyer, C. H., Siddique, K. H., Tai, A. P., Anders, S., Fodor, N., Wong, F. L., . . . Considine, M. J. (2019). Modelling predicts that soybean is poised to dominate crop production across A frica. *Plant, Cell & Environment, 42*(1), 373-385.

Fung, W., Liverpool-Tasie, L. S. O., Mason, N. M., & Oyelere, R. U. (2020). Do crop purchase programs improve smallholder welfare? The case of Zambia's Food Reserve Agency. *Agricultural Economics*, 51(4) 519-533.

Gatto, M., Wollni, M., Asnawi, R., & Qaim, M. (2017). Oil palm boom, contract farming, and rural economic development: Village-level evidence from Indonesia. *World Development, 95*, 127-140.

Ghezán, G., Mateos, M., & Viteri, L. (2002). Impact of supermarkets and fast–food chains on horticulture supply chains in Argentina. *Development Policy Review, 20*(4), 389-408.

Kikuchi, M., Haneishi, Y., Tokida, K., Maruyama, A., Asea, G., & Tsuboi, T. (2016). The structure of indigenous food crop markets in sub-Saharan Africa: The rice market in Uganda. *The Journal of Development Studies, 52*(5), 646-664.

Larson, D. F., Otsuka, K., Matsumoto, T., & Kilic, T. (2012). Should African rural development strategies depend on smallholder farms? An exploration of the inverse productivity hypothesis. *Agricultural Economics, 45(3), 355-367.*

Liverpool-Tasie, L. S. O., Nuhu, A. S., Awokuse, T., Jayne, T., Muyanga, M., Aromolaran, A., & Adelaja, A. (2020). Spillover Effects of Medium-Scale Farms on Smallholder Behaviour and Welfare: Evidence from Nigeria. *APRA Working Paper, No. 38, September 2020..*

Liverpool-Tasie, L. S. O., Wineman, A., Young, S., Tambo, J., Vargas, C., Reardon, T., . . . Bizikova, L. (2020). A scoping review of market links between value chain actors and small-scale producers in developing regions. *Nature Sustainability, 3*(10), 799-808.

Liverpool-Tasie, L. S. O., & Parkhi, C. M. (2020). Climate Risk and Technology Adoption in the Midstream of Crop Value Chains: Evidence from Nigerian Maize Traders. *Journal of Agricultural Economics*, Early View.

Louw, A., Jordaan, D., Ndanga, L., & Kirsten, J. F. (2008). Alternative marketing options for small-scale farmers in the wake of changing agri-food supply chains in South Africa. *Agrekon, 47*(3), 287-308.

Lubungu, M., Burke, W. J., & Sitko, N. J. (2013). Analysis of the soya bean value chain in Zambia's Eastern Province. *IAPRI Working Papers, No. 74.*

Meemken, E.-M., & Bellemare, M. F. (2020). Smallholder farmers and contract farming in developing countries. *Proceedings of the National Academy of Sciences, 117*(1), 259-264.

Michelson, H., Reardon, T., & Perez, F. (2012). Small farmers and big retail: trade-offs of supplying supermarkets in Nicaragua. *World Development, 40*(2), 342-354.

Montalbano, P., Pietrelli, R., & Salvatici, L. (2018). Participation in the market chain and food security: The case of the Ugandan maize farmers. *Food Policy, 76*, 81-98.

Nuthalapati, C. S. R., Sutradhar, R., Reardon, T., & Qaim, M. (2020). Supermarket procurement and farmgate prices in India. *World Development, 134*, 105034. doi:https://doi.org/10.1016/j.worlddev.2020.105034

Ogutu, S. O., & Qaim, M. (2019). Commercialization of the small farm sector and multidimensional poverty. *World Development, 114*, 281-293.

Opperman, C., & Varia, S. (2011). Technical Report: Soybean Value Chain. *Southern Africa Trade Hub Technical Report. Retrieved from https://www.satradehub.org*

Ragasa, C., & Mazunda, J. (2018). The impact of agricultural extension services in the context of a heavily subsidized input system: The case of Malawi. *World Development, 105*, 25-47.

Rao, E. J., & Qaim, M. (2011). Supermarkets, farm household income, and poverty: insights from Kenya. *World Development, 39*(5), 784-796.

Reardon, T. (2015). The hidden middle: the quiet revolution in the midstream of agrifood value chains in developing countries. *Oxford Review of Economic Policy, 31*(1), 45-63.

Reardon, T., Echeverria, R., Berdegué, J., Minten, B., Liverpool-Tasie, S., Tschirley, D., & Zilberman, D. (2019). Rapid transformation of food systems in developing regions: highlighting the role of agricultural research & innovations. *Agricultural Systems, 172*, 47-59.

Richards, P., Reardon, T., Tschirley, D., Jayne, T., Oehmke, J., & Atwood, D. (2016). Cities and the future of agriculture and food security: A policy and programmatic roundtable. *Food Security, 8*(4), 871-877.

Roser, M., & Ortiz-Ospina, E. (2013). World Population Growth. *OurWorldInData.org.* Retrieved from: https://ourworldindata.org/world-population-growth.

Schoonhoven-Speijer, M., & Vellema, S. (2020). How institutions governing the economic middle in food provisioning are reinforced: The case of an agri-food cluster in northern Uganda. *Journal of Rural Studies*, 80, 34-44.

Sexton, R. J. (2013). Market power, misconceptions, and modern agricultural markets. *American Journal of Agricultural Economics, 95*(2), 209-219.

Siamabele, B. (2019). Soya Beans Production in Zambia: Opportunities and Challenges. *American Journal of Agricultural and Biological Sciences,* 14(1) 55-60

Singh, I., Squire, L., & Strauss, J. (1986). Agricultural household models: Extensions, applications, and policy. *World Bank,* Retrieved from https://documents.worldbank.org.

Sitko, N. J., Burke, W. J., & Jayne, T. (2018). The quiet rise of large-scale trading firms in East and Southern Africa. *The Journal of Development Studies, 54*(5), 895-914.

Sitko, N. J., & Jayne, T. (2014). Exploitative briefcase businessmen, parasites, and other myths and legends: assembly traders and the performance of maize markets in eastern and southern Africa. *World Development, 54*, 56-67.

Timmer, C. P. (1988). The agricultural transformation. *Handbook of development economics, 1*, 275-331.

Ton, G., Desiere, S., Vellema, W., Wytse, S., & D'Haese, M. (2017). The effectiveness of contract farming for raising income of smallholder farmers in low-and middle-income countries: a systematic review. *Campbell Systematic Reviews, 13(1) 1-131.*

Tschirley, D., Reardon, T., Dolislager, M., & Snyder, J. (2015). The rise of a middle class in East and Southern Africa: Implications for food system transformation. *Journal of International Development, 27*(5), 628-646.

Van Der Meer, C. (2006). Exclusion of small-scale farmers from coordinated supply chains. *Agro-food Chains and Networks for Development, Amsterdam, 14*, 209-218.

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