Grantham Research Institute on Climate Change and the Environment

Centre for Climate Change Economics and Policy

Does choice of drought index influence estimates of drought-induced cereal losses in India?

Francisco Fontes, Ashley Gorst and Charles Palmer

August 2017

Centre for Climate Change Economics and Policy Working Paper No. 310
ISSN 2515-5709 (Online)

Grantham Research Institute on Climate Change and the Environment
Working Paper No. 274
ISSN 2515-5717 (Online)

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council. Its second phase started in 2013 and there are five integrated research themes:

1. Understanding green growth and climate-compatible development
2. Advancing climate finance and investment
3. Evaluating the performance of climate policies
4. Managing climate risks and uncertainties and strengthening climate services
5. Enabling rapid transitions in mitigation and adaptation

More information about the Centre for Climate Change Economics and Policy can be found at: www.cccep.ac.uk.

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and the Global Green Growth Institute. It has nine research programmes:

1. Adaptation and development
2. Carbon trading and finance
3. Ecosystems, resources and the natural environment
4. Energy, technology and trade
5. Future generations and social justice
6. Growth and the economy
7. International environmental negotiations
8. Modelling and decision making
9. Private sector adaptation, risk and insurance

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: www.lse.ac.uk/grantham.

DOES CHOICE OF DROUGHT INDEX INFLUENCE ESTIMATES OF DROUGHT-INDUCED CEREAL LOSSES IN INDIA?

Francisco Fontes,[1] Ashley Gorst,[2] and Charles Palmer[3]

Department of Geography and Environment & Grantham Research Institute on Climate Change and the Environment, London School of Economics (LSE)

DRAFT

August 2017

ABSTRACT

Drought events have critical impacts on agricultural production yet there is little consensus on how these should be measured and defined. This has implications for drought research and policy, which tends to either define droughts purely based on rainfall or focus uniquely on 'hot' droughts when temperature is considered. We develop a flexible, rainfall-temperature drought index that captures all dry events, including a previously overlooked class of drought events that we term 'cold' droughts. Our index is applied to a panel dataset of Indian districts over the period 1966-2009. Results suggest a statistically significant relationship between the index and agricultural production. Cold droughts are found to have consistent, negative marginal impacts that are comparable to those of hot droughts. Estimates of average yield losses due to hot droughts are reduced by as much as 33% when cold droughts are omitted. The associated economic costs are even more severely underestimated, by up to 107%.

Keywords: Agriculture, Cereals, Climate, Drought, India, Rainfall, Temperature

JEL classification: Q10, Q19, Q54, Q56

Acknowledgements: We wish to thank participants at the Conference of Shocks and Development (Dresden) as well as various colleagues in the Grantham Research Institute on Climate Change and the Environment (GRI) for useful comments, in particular Ganga Shreedhar**.** Francisco Fontes and Ashley Gorst acknowledge support from the Centre for Climate Change Economics and Policy, which is funded by the UK Economic and Social Research Council.

[1] F.Pereira-Fontes@lse.ac.uk

[2] A.gorst@lse.ac.uk

[3] C.palmer1@lse.ac.uk

1. INTRODUCTION

Extended periods of low rainfall and high temperature that reduce the availability of moisture relative to normal climate conditions broadly constitute drought events (Mishra and Singh, 2010). A number of low- and middle-income countries in the world, including those located in Sub-Saharan Africa and the Indian sub-continent, are particularly vulnerable to the impacts of such events. The human and economic costs of drought can be considerable. In India, the setting for our paper, Gadgil and Gadgil (2006) estimate that severe drought lowered annual GDP by around two to five percent between 1951 and 2003, while Pandey et al. (2007) show that drought was accompanied by a 12 to 33% increase in the poverty headcount ratio and a 25 to 60% decline in household income. The onset of drought in India has also been empirically linked to conflict, rural wages and human capital accumulation (Jayachandran, 2006; Sarson, 2015; Shah and Steinberg, forthcoming).

Against a backdrop of rising temperatures and drier conditions, droughts are projected to become more common with critical implications for agricultural production (IPCC, 2012). How drought is defined plays a central role in policymakers' responses, not only in the agricultural sector but also in the water sector and in early-warning systems. Yet, in the academic and policy literatures there is presently little consensus on how droughts might be measured and hence, defined. Indeed, there is no universal definition of the conditions constituting a drought (Wilhite, 2000). A range of indices attempt to quantify the severity of a drought, ranging from simple rainfall measures to complex indices that account for rainfall, temperature and estimates of potential evapotranspiration (Mishra and Singh, 2010). Different criteria of what constitutes a 'drought' therefore imply that a drought in one index may not constitute a drought in another. The implication is that, depending on the index used, there are classes of dry events which may simply be overlooked both in empirical analyses and by policymakers.

In this paper, we develop a simple rainfall-temperature index that allows for a flexible characterisation of drought events. It captures every dry event, in which cumulative (growing season) precipitation is below average, long-term cumulative (growing season) precipitation,[1] while accounting for temperature. The novelty of our index is to include both the type of dry

[1] Cumulative growing season precipitation is defined as total rainfall between June and September. Long-term cumulative precipitation is average cumulative growing season precipitation between 1956 and 2009.

events typically captured by indices that account for temperature, i.e. characterised by above-average temperatures ('hot' droughts), as well as ones characterised by *below-average* temperatures. To our knowledge, the latter, which we term ‘cold’ droughts, have not been explicitly studied before. Our index is then applied to a panel dataset of Indian districts over the period 1966-2009 in order to estimate the marginal and total effects of drought on cereal productivity. These estimates are then used to calculate changes in yield and associated economic impacts. In a country where over two-thirds of total land area is vulnerable to drought (Ministry of Agriculture, 2009), and rain-fed agriculture covers approximately 60% of cropped area (Sharma, 2011), our analysis contributes to an important body of research on the impacts of droughts on Indian agriculture (e.g. Pandey et al., 2007; Sarkar, 2011).

After motivating our analysis in the context of the relevant literature in Section 2, we present Indian weather data underlying hot and cold droughts, in Section 3. In Section 4, we propose an extension to a multiplicative index originally developed by Yu and Babcock (2010). This extension allows for a more flexible characterization of drought events while retaining a key strength of their index, namely the inclusion of temperature and the capacity to capture the interaction between rainfall and temperature. Applied to our panel dataset of Indian districts in Section 5, we find a statistically significant relationship between the index and agricultural production. We also find that cold droughts consistently display large negative marginal and total effects, comparable to those of hot droughts, and that omitting cold droughts leads to a large underestimation of total drought impact. Yield and economic losses are shown in Section 6 to be underestimated by up to 33% and 107%, respectively. Section 7 concludes.

2. DEFINING ‘DROUGHT’

Simple drought indices often rely solely on precipitation measures and are typically preferred by policy-makers, including the Indian Meteorological Department (IMD), over more complex indices. Until 2016, the IMD recorded a ‘drought event’ when seasonal rainfall was below 75% of its long-term average value (between 1950 and 2000), and a ‘severe drought’ when rainfall was below 50% of this value. Simple metrics of precipitation deficiency, which have the advantage of being easily interpretable, are also used to evaluate drought impacts on agricultural production. For example, to estimate drought impact in the rice-growing regions of Asia, Pandey et al. (2007) define a drought as ‘moderate’ if rainfall is 70-80 percent of normal levels, and ‘severe’ if rainfall is 70 percent below normal. Auffhammer et al. (2012)

use a similar definition to study the effect of monsoon rainfall on rice yields in Indian states. The strength of these indices lies in their simplicity.

However, simple definitions of drought are problematic for our understanding of drought impacts for two reasons. First, they impose arbitrary thresholds in order to define drought, evaluating drought impacts only after a given level of precipitation, when the agronomic or empirical basis of such thresholds is unclear (Wilhite and Glantz, 1985). Second, variables in addition to precipitation, in particular temperature, help determine the physical severity of a drought.[2] Given temperature increases driven by climate change (Hatfield et al., 2011), a growing literature suggests critical turning points at which higher temperatures cease to have positive impacts on agricultural yield. Schlenker and Roberts (2009) find that higher temperatures in the US reduce county-level yields for corn (above 29°C), soy-beans (30°C), and cotton (32°C). Guiteras (2009) and Burgess et al. (2014) both show that, on average, daily temperatures above 34°C in India reduces agricultural productivity at the district scale. Lobell et al. (2012) identify the same threshold as harmful for Indian wheat yields.

High temperatures have particularly acute effects on crop growth during periods of low precipitation since the rate of evapotranspiration, i.e. the combined process of water evaporated from land surfaces and plants, increases as temperatures rise (Prasad et al., 2008; Lobell and Gourdji, 2012). In general, this increases a plant's demand for water at a time when water availability is already low due to deficient precipitation. Recent research has documented that droughts in a range of settings have increased in severity as mean temperatures have risen. Higher temperatures, rather than the increased intensity of low rainfall events, have been held responsible for these drying trends (Vicente-Serrano et al., 2014; Diffenbaugh et al., 2015). As such, not considering the effect of temperature on the severity of a drought event could underestimate drought impact, in turn giving misleading information about the likelihood of future production losses driven by climate change.

More complex indices tend to rely on data that are not readily available in most economic datasets, e.g. for soil moisture levels and estimates of potential evapotranspiration. The lack of data for deriving such measures, which can depend on factors such as wind, radiation and humidity, limits their applicability in empirical analysis of drought impacts. In an attempt to

[2] Such variables include, for example, access to irrigation.

bridge the gap between simple and complex indices, Yu and Babcock (2010) propose a drought index that neatly captures the interaction between two of the most important factors: temperature and precipitation. Applied to the study of drought tolerance of soybean and corn yields in the US, it takes a non-zero value only for years of below-average precipitation and above-average values of heat temperature ('cooling degree-days'). The authors find that soybeans and corn have become increasingly drought-tolerant over time.

This index has since been applied in a number of other settings, for example, to the assessment of drought impact on soybean in Missouri (Purcell and Caine, 2013). Of particular relevance is research by Birthal et al. (2015), who use the index to study the resilience of rice yields to drought in India. Their results indicate that rice yields have become more tolerant to drought over time. While this approach has the advantage of being a relatively simple way to account for both temperature and precipitation, the index restricts the definition of drought to events characterised by low rainfall accompanied by higher-than-average temperatures. It does not consider events characterised by below-mean rainfall as well as below-mean temperature. Such cold droughts are common in many settings, although their impacts on agricultural production remain unknown, due to either being omitted altogether (as in Birthal et al. 2015) or joined with hot droughts in arbitrarily-defined rainfall indices. This is an important gap in the literature that our paper aims to fill.

We argue that cold droughts should not be omitted *a priori* for two reasons. First, a large number of potentially destructive droughts are not considered, which can lead to a serious underestimation of their total impact. Second, the classification of these events as non-droughts could lead to biased estimates of drought impact. Thus, where cold droughts have a significant negative impact on productivity, the application of Yu and Babcock's (2010) index potentially underestimates drought impacts due to the inclusion of cold drought events in the 'no drought' control group.

3. DROUGHT IN INDIA

According to the definition used by the IMD (at least until 2016), 13 'All-India drought years' have been recorded since the beginning of the Green Revolution in 1966 (Birthal et al., 2015). Four of these occurred between 2000 and 2012. A 'drought year' was recorded when the total area affected by a moderate or severe drought covered 20-40% of the total land area of the country and seasonal rainfall during the monsoon season exceeded 10%. When more than

40% of the total land area was affected by drought, this was known as an 'All India Severe Drought Year'.

Weather data on daily rainfall and daily average temperature at the district level are sourced from the IMD to create Figures 1 and 2.[3] Panel (a) of Figure 1 shows the proportion of districts in every given year limited to events characterized by both below-average rainfall and above-average temperature, i.e. the events considered by Birthal et al. (2015) using Yu and Babcock's (2010) index. The vertical blue lines indicate the years defined by India's government as All-India droughts. Panel (b) of Figure 1 shows the proportion of districts in years characterised by below-average rainfall and below-average temperature; a large proportion of districts are clearly affected by this type of drought event. Figure 2 shows why the omission of these events is problematic. For each year, we estimate the number of districts affected by hot droughts net of the number of those affected by cold droughts, with a positive number (in red) denoting a year in which the former exceeds the latter. A negative number (in blue) indicates a year in which the latter exceeds the former. Overall, hot droughts are slightly more prevalent than cold droughts (roughly a split of 55% hot and 45% cold drought). In the 1990s, most of the drought-affected districts were affected by hot droughts. Since 1999, the number of cold droughts has increased, with the number of districts affected by cold droughts outnumbering districts affected by hot droughts in seven out of 11 years.[4]

FIGURE 1 HERE

FIGURE 2 HERE

4. INTRODUCING A NEW DROUGHT INDEX

In this section, we build on Yu and Babcock's (2010) drought index, incorporating both rainfall and temperature. Their index is based on the following:

$$DI_{i,t} = \left[-\max\left(0, CLDD_{i,t}^{stand}\right)\right] * \left[\min\left(0, TPCP_{i,t}^{stand}\right)\right] \qquad (1)$$

[3] The rainfall data are available in gridded format at a resolution of 0.25°x 0.25° (Pai et al., 2014). Gridded temperature data are at a resolution of 1°x1° (Srivastava et al., 2009). District-level weather data are then obtained by taking a weighted average of gridded weather observations from grid cells that fall within a district's boundary based on the proportion of the grid cell that falls in each district.
[4] This pattern, however, is slightly less pronounced when we examine an alternatively-defined growing season (May-December) (Figures A1 and A2 in the Appendix).

where: *DI* denotes the drought index, for a given unit of observation, *i,* in year *t*; $CLDD_{i,t}^{stand}$ is standardized Cooling Degree Days (above 65°F, or 18.33°C); and, $TPCP_{i,t}^{stand}$ is standardized total monthly precipitation between the months of June and August.

This index gives a value of zero to drought whenever either the temperature is below average or the rainfall is above average. As such, a drought occurs in a year when temperature is uncommonly high and precipitation is low, relative to the long-term average of these variables. A strength of this simple index lies in its capacity to capture the potential of high temperatures to exacerbate the effects of low rainfall on crop production.

One weakness of the index described in (1) is that it defines as a drought only those years when an area suffers both low rainfall and high temperatures. Omitted are years when rainfall is low but temperatures are not particularly high. Defining drought events by low rainfall and high temperature restricts the measure of drought to the lower-right quadrant of Figure 3. Events in the lower-left quadrant, where both precipitation and temperature are below-average, would not be counted as droughts according to the index in (1).

FIGURE 3 HERE

We consider a wider set of drought events by defining six variables. First, using weather data from the IMD, we calculate district-specific average long-term cumulative rainfall, $\mathrm{LTAR_i}$, for the growing season (June-September) over the period 1956-2009.[5] This variable is standardized by estimating $ZTR_{it} = \frac{TR_{it} - LTAR_i}{sdTR_i}$, where TR_{it} is total cumulative rainfall over the growing season for a given year and $sdTR_i$ is the standard deviation of TR_{it}. Analogously, we calculate the district-specific average cumulative growing season cooling degree days, $\mathrm{LTACDD_i}$, for the June-September growing season as the average cumulative number of degree-degree days above the mean daily growing season temperature over the period 1956-2009.[6] Similar to rainfall, this variable is standardized by estimating $ZCDD_{it} = \frac{CDD_{it} - LTACDD_i}{sdCDD_i}$,

[5] There are two main reasons driving our choice of growing season. First, the majority of India's cereal production is cultivated in the *kharif* season, between June and September. Second, according to Jain and Kumar (2012), the majority of total yearly rainfall (approximately 80%) occurs between June and September. Authors such as Prasana (2014) also highlight that, while there is a strong and positive response to *kharif* production and June-September rainfall, the same is not necessarily true for *rabi* production and post-monsoon rainfall (October-December). This partly relates to the fact that *rabi* crops rely on available moisture from the June-September rains.

[6] The growing season cooling degree days measure is calculated as follows. First, we obtain the average growing season temperature. Second, for each day we subtract the average temperature from the observed

where CDD_{it} is total cumulative daily degree days over the growing season for a given year and $sdCDD_i$ is the standard deviation of CDD_{it}.

Let $MTR_{it} = -TR_{it}$, i.e. the negative of total cumulative rainfall. We then obtain the normalized version of this variable, NTR_{it}, by estimating $NTR_{it} = \frac{MTR_{it} - MTR_i^{min}}{MTR_i^{max} - MTR_i^{min}}$, where MTR_i^{min} denotes the minimum observed value for district *i* (i.e. the maximum rainfall observed), and MTR_i^{max} denotes its maximum observed value (i.e. lowest rainfall). Normalizing the negative of rainfall, rather than rainfall directly, allows us to generate a variable bounded between 0 and 1, with higher values signalling a more severe precipitation deficiency. Similarly, for normalizing the degree days measure, we estimate $NCDD_{it} = \frac{CDD_{it} - \mathrm{CDD}_i^{min}}{CDD_i^{max} - CDD_i^{min}}$, where CDD_i^{min} denotes the minimum observed value for district *i* (i.e. the minimum number of degree days observed), and CDD_i^{max} denotes its maximum observed value (i.e. highest number of degree days observed in a given district).

A multiplicative relationship is generated between the two normalized variables, which we use to define three different drought indices. First, hot droughts can be classified as $D1_{\mathrm{it}}$, corresponding to the classification of Yu and Babcock (2010) where rainfall is below normal and temperature above normal. Second, $D2_{\mathrm{it}}$ corresponds to low rainfall in the *absence* of abnormally high temperatures. Third, we combine $D1_{\mathrm{it}}$ and $D2_{\mathrm{it}}$ to get $D12_{\mathrm{it}}$, thus accounting for both hot and cold droughts. Formally, we have:

$$Drought = \begin{cases} D1_{\mathrm{it}} = \mathrm{NTR} * \mathrm{NCDD} \;\; if\; \mathrm{ZTR_{it}} < 0 \;\; and\; \mathrm{ZCDD_{it}} > 0; \;\; 0\; otherwise \\ D2_{\mathrm{it}} = \mathrm{NTR} * \mathrm{NCDD} \;\; if\; \mathrm{ZTR_{it}} < 0 \;\; and\; \mathrm{ZCDD_{it}} < 0; \;\; 0\; otherwise \\ D12_{\mathrm{it}} = \mathrm{NTR} * \mathrm{NCDD} \;\; if\; \mathrm{ZTR_{it}} < 0\,; \;\; 0\; otherwise \end{cases} \quad (2)$$

As such, $D1_{\mathrm{it}}$ can be interpreted as a normalized version of Yu and Babcock's (2010) index. It captures all events in the lower-right quadrant of Figure 3, taking a strictly positive value for all events characterized by below-average precipitation and above-average temperatures. The second index, $D2_{\mathrm{it}}$, only takes non-zero values for events with below-average rainfall and below-average temperature, the category Yu and Babcock omit. Constructing these two indices separately allows us to test their respective statistical significance in the yield

temperature and obtain the number of degrees above the average temperature for each day. Finally, we sum all the positive temperature deviations for each day of the growing season and obtain cumulative daily-degree days.

regressions in Section 5. Finally, a third index, $D12_{\mathrm{it}}$, simply combines $D1_{\mathrm{it}}$ and $D2_{\mathrm{it}}$ and hence, captures all the events in the lower half of Figure 3. More detail on how the indices are constructed is presented in the Technical Appendix (A).

Our indices are increasing in temperature but decreasing in precipitation since both higher temperatures and lower precipitation are expected to contribute to drought severity. A maximum value of one is obtained for the most severe droughts, and is only possible for the restricted set of drought events considered by Yu and Babcock. The similarity of their index to our own is illustrated in Table 1, which shows the correlation coefficients and the spearman correlation coefficient. As expected, our index D1 is highly correlated with Yu-Babcock, displaying a correlation coefficient of 0.776 and a spearman correlation coefficient in excess of 0.99. Our second index, on the other hand, has a negative correlation coefficient. Since Yu-Babcock is invariant with a value of zero for these events, this result is also as anticipated.

TABLE 1 HERE

Figure 4 shows how our indices change over time, for all districts (panels (a) and (b)) and for drought-affected districts only (panels (c) and (d)). Hot (D1) and cold (D2) droughts are denoted Types 1 and 2, respectively. There are clear spikes in the values of the index for a number of All-India drought years. In recent years, 2002 and 2009 are associated with the largest deviations in rainfall; spikes correspond to these two years. Similarly, 1972, 1979, 1987 are also considered years with particularly high deviations and our index rises in these years. Throughout the 1990s, however, it is striking that, despite relatively modest deviations of rainfall from trend, our index still records high values. On average, the negative deviations from long-term average rainfall were smaller throughout the 1990s. A possible explanation for this could be the fact that, as highlighted by Pai et al. (2012), overall, land surface air temperatures have increased over time. This pattern was particularly pronounced in the 1990s and 2000s.

FIGURE 4 HERE

5. IMPACT OF DROUGHT ON CEREAL PRODUCTIVITY

5.1 Data and Methodology

To investigate drought impacts on aggregate cereal productivity at the district level, we obtain agricultural data from the ICRISAT Meso-level Database.[7] For the period 1966-2009, the dataset contains detailed agricultural and socioeconomic information (ICRISAT, 2012). For most if not all districts, data are available for annual crop production and area under crop production for a range of crops. We create a balanced panel, which implies that, out of the 311 available in the dataset, only 275 districts are used in our empirical analysis due to missing weather and/or production data. Six cereals are considered, namely rice, wheat, maize, barley, sorghum, and millet.[8] Yields for each are estimated along with a simple cereal yield variable, obtained by dividing total cereal production by total cereal area. Table 2 summarises the variables used in our analysis.

TABLE 2 HERE

To model the relationship between yield and our drought index, we estimate the following fixed-effects model:

$$\ln(y_{itc}) = \alpha_i + \gamma_t + \delta_{i1} * t + \delta_{i2} * t^2 + \beta_q \, DI_{itq} + \epsilon_{it} \qquad (3)$$

where for district *i* in year *t*: *In(*y_{it}*)* denotes the natural logarithm of cereal yield (or crop *c*); α_i and γ_t represent the district and year fixed effects, respectively; δ_{i1} and δ_{i2} are the coefficients on the district-specific quadratic trend. The coefficient associated with a type *q* (i.e. Type 1 - hot or Type 2 - cold) drought index, which captures the marginal impact of a type *q* drought, is denoted β_q. Finally, ϵ_{it} represents the error term. After estimating (3), we include dummy variables for each of the drought types - Type 1, Type 2, Types 1 and 2 – in order to account for a potential intercept shift and the convergence of marginal effects.

In the following section, we perform a number of tests on the sensitivity and robustness of our main results. First, we cluster standard errors at the district level and second, consider two alternative growing seasons: May-December and annual (January-December). The former allows for the fact that, in some states, there may be substantial amounts of rain

[7] Since 1966, a number of districts have split into smaller districts. To maintain spatial consistency over time district splits are dealt with by returning split districts to their ‘parent’ districts as of 1966.

[8] For millet we add data on quantities of pearl millet and finger millet to create an aggregate quantity of millet.

outside of the June-September period. The latter is chosen to capture rainfall later in the year, which may help explain *rabi* production, despite a current lack of evidence for post-monsoon rainfall impacts on *rabi* production (see footnote 5 and Prasana, 2014).

In a third test, we adopt an alternative specification for cooling degree-days – 30 degrees rather than the long-term district average temperature over the growing season – for two reasons. First, according to research on the effects of temperature on crop yield in India (e.g. Guiteras 2009), crop yields typically start decreasing above 30 degrees Celsius. Since we use mean temperature, which in some cases is below 30 degrees C, it could be argued that an increase in temperature should not necessarily result in a decline in production. However, using 30 degrees C as an absolute cut-off point implies a value of zero in our drought index for many drought events (when rainfall is very low), which makes our index less comprehensible. Also, it is possible that in some cases, negative impacts of temperature on production might be observed at levels below 30 degrees C, e.g. in 'colder' districts, where the crops cultivated might be more sensitive to temperature deviations.

Consistent with Yu and Babcock (2010), we do not include controls in our main specifications. This is also the norm in the broader weather and climate literature. In a fourth test, we thus examine the sensitivity of all of our results to the inclusion of controls to ensure that they are robust to the inclusion of variables which are also likely to affect production, such as irrigation and the use of modern inputs. Fifth, we test alternative functional forms of the index. In particular, since the impacts of drought may not be linear, we include a squared term. Finally, we derive results using an additive index instead of a multiplicative index. An additive relationship may be relevant in a number of extreme cases. For example, in a year where rainfall is close to zero and where temperatures have also been low, our index would have a low value, which may be misleading.

5.2 Regression results

We run a regression of the natural logarithm of yield on a set of district-specific quadratic trends and the drought indices. Specifically, for both the main model described in (3) and all of the robustness and sensitivity checks we run the regression using the full sample and by crop. For the full sample and for each crop we then run three regressions. First, we include only Type 1 (hot) drought events. Second, we estimate separate coefficients for Type 1 and Type 2 (cold) drought events. Finally, we run a regression where we only include the drought index that combines Type 1 and Type 2 events, i.e. hot and cold. The results for the full sample can be seen in Table 3, with columns 1-3 showing results in the absence of dummy variables and 4-6 those with the inclusion of dummies. The results by crop are in Tables 4-5 (without dummies) and 6-7 (with dummies).

Table 3 highlights three main points. First, both types of drought have significant and negative effects when considered separately, as shown in columns 2 and 5. Thus, Type 2 events, i.e. those omitted by Yu and Babcock (2010) and Birthal et al. (2015), have large and statistically significant, negative impacts on yield.

TABLE 3 HERE

Second, by comparing the specifications where we omit Type 2 events (columns 1 and 4) with those where this type of event is included (columns 2 and 5), we note that the estimated marginal coefficient of Type 1 droughts is smaller in the former. When we include a dummy variable, the difference in magnitude is negligible. But when all dummy variables are excluded, the coefficient of Type 1 events is substantially smaller - and outside the 95% confidence interval of the estimated coefficient - when Type 2 events are also included (for a graphical representation, see Figure A3 in the Appendix). Thus, a failure to account for Type 2 events can lead to an underestimate of the marginal impacts of drought, with this underestimate being more severe when a dummy is not included.

Third, we find that in contrast to hot drought, cold droughts have a larger marginal, but lower total, effect on agricultural production. Although both excess heat and reduced moisture have negative impacts on production, reduced precipitation carries greater weight in the cold drought index than in the hot drought index since values of temperature are, by definition, higher in the latter than in the former. As a result, yields are likely to respond (more) negatively to changes in the cold drought index than in the hot drought index. A value of 0.5 in our cold drought index represents approximately the same precipitation deficiency as a value of one in our hot drought index, which could help explain larger marginal impacts.

The results by crop (Tables 4-7) corroborate the patterns found across the whole sample using the cereal index. The estimated coefficients for Type 2 events are consistently large, negative and significant for all crops except for maize (when dummies are excluded). This provides further evidence that such events have a large negative impact on production and hence, should not be excluded from analyses of drought impact on cereal yield in India. Similar to our findings for the whole sample, in most cases (maize again being the exception) when a dummy is not included to account for the intercept shift, the omission of Type 2 events leads to a smaller, estimated coefficient of Type 1 droughts. This effect is especially large in the case of rice, the crop analysed by Birthal et al. (2015), thus implying that they may have underestimated the impact of drought on rice. We estimate the potential scale of underestimation, in terms of yield and its economic value, below.

TABLE 4 HERE

TABLE 5 HERE

TABLE 6 HERE

TABLE 7 HERE

We perform a number of sensitivity checks on our results, in which we consider: (i) standard errors clustered at the state level; (ii) two alternative growing seasons (May-December and annual); (iii) an alternative specification for our degree-days variable (30 degrees C rather than the long-term district average temperature over the growing season); (iv) controls; (v) alternative functional forms of the index (including the square of the index); and, (vi) an additive index instead of a multiplicative index. For the full sample, the coefficients from each of these specifications are summarized in Tables A1-A2. In Figure A3, we present a graphical summary of the coefficient values and their confidence intervals for each specification using the full sample.[9] Tables A3-A14 summarize all the robustness checks by crop.

In summary, while different specifications unsurprisingly generate different coefficients, our overarching conclusions are quite robust, especially in the case of aggregate cereal production. One key result, that of cold droughts driving larger marginal effects than hot droughts, is, in contrast to our other main results less robust in a number of alternative specifications. Of all the checks, our results are especially robust to alternative index specifications (Tables A4, A6, A8, A10, A12, A14).

6. ESTIMATING YIELD AND ECONOMIC LOSSES

[9] Tables with the additional robustness checks are available from the authors upon request.

In a back-of-the-envelope attempt to gauge first, how important both types of droughts are in the Indian context and second, how serious the omission of Type 2 droughts is for estimating Type 1 drought impacts, we run simple simulations using our estimated regressions. This allows us to generate predictions of yields with and without droughts. Specifically, we estimate the: (i) average yield loss for an affected district over the sample period; (ii) average total production loss for an affected district over the sample period; (iii) average total value of production for an affected district; (iv) average unweighted yearly total production loss in our sample of Indian districts; and, (v) the average yearly total cost across sampled districts. A summary of estimates is presented in Tables 8 and 9, including crop-specific results.[10] Details of how we generated these estimates can be found in the Technical Appendix (B).

TABLE 8 HERE

TABLE 9 HERE

From Tables 8 and 9, we note the following. Despite a higher estimated coefficient, total yield and economic losses from cold droughts are smaller than those from hot droughts. This is due to the index values for cold droughts being substantially lower (approximately half) for affected districts. For our aggregate cereal measure, we estimate the average yield loss per district at 160 kg/ha and 110 kg/ha for hot and cold droughts, respectively. These smaller impacts on yields translate into lower total economic costs. Whereas we estimate that, in a given year, the total economic cost of a hot drought is, on average, approximately USD 1.02 billion (using 2008 crop prices; column 2 in Table 8),[11] this falls to USD 650 million for a cold drought (column 3 in Table 8).

To our knowledge, there is only one study in the literature that attempts to estimate drought costs for the whole of the country. Using data for all Indian districts, Sarkar (2011) estimate cereal losses of 27.6 million tonnes due to drought in 2002.[12] In our sample of 275 districts (out of 311 apportioned districts), we estimate total production losses of about 16.3-16.9

[10] For the crop-specific results, these were obtained using the crop-specific regressions.
[11] Crop prices in Indian rupees are converted into USD using the average monthly exchange rate obtained from http://www.x-rates.com/average/?from=USD&to=INR&amount=1&year=2008. More details on how prices are computed are available in the Technical Appendix (B).
[12] They value these losses at around 1.3 trillion rupees. An error in their calculations, however, suggests a loss closer to 130 billion rupees

million tonnes in 2002, which we value at 103 billion rupees using nominal prices. In addition to covering fewer districts, the differences in estimates are also likely to stem from methodological differences as well as the fact that we do not take into account a potential reduction in cultivated area during a drought year.[13]

Omitting cold droughts can lead to a lower estimate of hot drought impact, especially when the dummy variables are excluded. Including dummy variables allows for a convergence in the marginal effect. However, if the change in intercept is taken into account when estimating costs, the divergences in the costs persist despite the inclusion of dummy variables. These effects are quantifiably large as we illustrate by comparing the first two columns of Tables 8 and 9 for the full sample. When dummies are excluded (Table 8), this effect is large. Average yield losses are estimated to be 33% higher (from 120kg/ha to 160 kg/ha) when cold droughts are included. These estimates have a substantial effect on the estimated average annual cost. This is USD 805 million (Table 8, column 1) when cold droughts are omitted compared to USD 1.02 billion (Table 8, column 2) when they are included, which represents a 27% increase. Thus, if estimating the economic cost of hot droughts without accounting for cold droughts, the average yearly total costs of drought would approximate USD 805 million. Including cold droughts raises this total cost by 107% to USD 1.67 billion (column 4 in Table 8). The difference can be broken down as follows: USD 216 million can be attributed to the lower coefficient of hot droughts, and USD 649 million to the inclusion of cold droughts. A similar difference exists when dummy variables are included, in Table 9.[14]

The impacts derived using the separate crop-specific specifications generate patterns of yields and costs similar to those that emerge from our aggregate cereal specification. Also, the patterns that emerge in the full sample specification are present in the crop-specific regressions. Cold droughts have a particularly large impact on rice, leading to substantive physical and economic losses. Indeed, about half of the total average economic cost from cold

[13] A more detailed explanation is given in the Technical Appendix (B). There are two further studies that attempt to estimate drought costs in India, which are less relevant for the purpose of comparison with our estimates. Pandey et al. (2007) estimate costs from yield losses for three states in eastern India and the EM-DAT database bases its national-level cost estimates on the basis of losses in housing, agriculture and livestock. Within range of our estimates, the former estimate a cost of USD 900 million for the 2002 drought. However, a large number of droughts do not have associated costs. Specifically, for our sample period, only 12 drought events were recorded for India in the EM-DAT database and, out of these 12 events, only five have had their associated costs estimated.

[14] However, in the case of Table 9, this difference arises from accounting for the intercept change, rather than the underestimation of the marginal impact.

droughts can be explained by losses in rice yields. In the absence of dummies, we estimate rice yield losses of around 160kg/ha (column 5 under 'Rice' in Table 8). Excluding Type 2 drought, i.e. comparing column 1 and 2 under 'Rice' in Table 8, suggests underestimates in the region of about 33%. Thus, it is highly likely that Birthal et al. (2015) underestimated the impacts of droughts on rice yields in their analysis.[15]

7. CONCLUSION

Overall, there are three main findings that emerge from our analysis. First, after proposing an index which extends one developed by Yu and Babcock (2010), we show that both hot and cold droughts have significant impacts on agricultural productivity in India. Thus, it is important to include the latter category of droughts, especially in a setting where there has been a clear increase in the number of such events in recent years. Moreover, if an assessment of economic impacts is performed solely based on hot droughts alone, approximately half of all potential dry events would be overlooked. Our results strongly suggest that these events have had quite a severe impact on cereal yields.

Second, the omission of cold droughts leads to a smaller estimated coefficient of hot droughts, especially when a dummy variable is not included to account for a potential intercept shift. Effectively, this implies that, if cold droughts have a large negative effect on productivity, estimating the coefficient for hot droughts without accounting for cold droughts could lead to underestimates of the marginal effect of hot drought thus further downward biasing the overall impact of drought in empirical analyses. This result does not challenge the central findings of Yu and Babcock (2010) and Birthal et al. (2015) that impacts of drought have declined over time. Yet, it does question the size of the marginal impacts estimated in both of these studies, and implies that a focus on hot droughts alone does not tell the whole drought story.

Third, the quantitative implications of our results are likely to be large, particularly given the fact that our cost estimates are based purely on yield losses. Since we do not take any potential changes in the cultivated area into account, we are likely to underestimate true

[15] Birthal et al. (2015) estimate rice yield losses due to drought ranging from 187 to 200 kg/ha. Differences in estimated impact are also likely to stem from the fact that they use a different sub-sample of districts and estimate a specification that differs from the one used in our analysis, e.g. we adopt a district-specific quadratic trend whereas they adopt a linear trend, as well as including interaction terms and irrigation as a control variable.

production losses. The economic value of production loss attributable to cold droughts is illustrated to be approximately 60% of the total economic value of production losses attributable to hot droughts in our main specification. Also, omitting cold droughts and a dummy variable can lead to an underestimation of the economic value of production losses due to hot droughts, which in our simulations amounted to a difference of about 27%. While we acknowledge that our back-of-the envelope estimates are based on a number of assumptions regarding prices and so forth, they do suggest that we have found sufficient empirical evidence and an economic rationale to justify the inclusion of cold drought in analyses of drought impact.

Our results have clear implications for public policy. Since cold droughts have measurable impacts on agricultural production that are severe yet not as severe as those resulting from hot droughts, policymakers should seek to distinguish between the two types of drought defined in this paper. Simple metrics of precipitation deficiency will obviously capture both types but since temperature plays a critical role in determining the extent of dry conditions at the local scale, it still needs to be explicitly accounted for. Detecting cold drought and tracking their impacts over time can serve as an early-warning response for when cold droughts transform into hot droughts, i.e. during periods when temperatures are predicted to be above the average of long-term trends. With global warming expected to continue to contribute to rising temperatures as well as potentially influencing patterns of extreme rainfall events, our index can thus help to shape the appropriate policy response to drought, particularly with respect to climate adaptation and agricultural production in more climate vulnerable locations.

8. REFERENCES

M. Auffhammer, V. Ramanathan, and J. Vincent. Climate Change, the Monsoon, and Rice Yield in India. *Climatic Change*, 111:411-424, 2012.

P. Birthal, D. Negi, M. Khan, and S. Agarwal. Is Indian Agriculture Becoming Resilient to Droughts? Evidence from Rice Production Systems. *Food Policy*, 56:1-12, 2015.

R. Burgess, O. Deschenes, D. Donaldson, and M. Greenstone. The Unequal Effects of Weather and Climate Change: Evidence from Mortality in India. 2014.

N. Diffenbaugh, D. Swain, and D. Touma. Anthropogenic Warming Has Increased Drought Risk in California. *Proceedings of the National Academy of Sciences of the United States of America*, 112(13):3931-3936, 2015.

S. Gadgil and S. Gadgil. The Indian Monsoon, GDP and Agriculture. *Economic and Political Weekly*, 41(47):86-91, 2006.

GoI. Manual for drought management. Technical report, Ministry of Agriculture, Department of Agriculture and Cooperation, Government of India, New Delhi, 2009.

R. Guiteras. The Impact of Climate Change on Indian Agriculture. *Mimeo*, 2009.

J.L. Hatfield, K.J. Boote, B.A. Kimball, L.H. Ziska, and R.C. Izaurralde. Climate Impacts on Agriculture: Implications for Crop Production. *Agronomy Journal*, 103:351-370, 2011.

ICRISAT. District Level Database Documentation. Technical report, ICRISAT-ICAR-IRRI Collaborative Research Project, 2012.

IPCC. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2012.

Indian Metereological Department, Government of India (undated). Frequently Asked Questions: FAQ. Available at: http://imd.gov.in/section/nhac/wxfaq.pdf

S. K. Jain and V. Kumar. Trend analysis of rainfall and temperature data for India. *Current Science (Bangalore)* 102(1): 37-49, 2012.

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

A. Mishra and V. Singh. A Review of Drought Concepts. *Journal of Hydrology*, 391(1-2): 202-216, 2010.

D. Lobell, A. Sibley, and J.I. Ortiz-Monasterio. Extreme Heat Effects on Wheat Senescence in India. *Nature Climate Change*, 2:186-189, 2012.

D. Lobell and S. Gourdji. The Influence of Climate Change on Global Crop Production. *Plant Physiology*, 160(4):1686-1697, 2012.

A. Mishra and V. Singh. A Review of Drought Concepts. *Journal of Hydrology*, 391(1-2): 202-216, 2010.

D.S Pai, L. Sridhar, M. Rajeevan, O.P. Sreejith, N.S. Satbhai, and B Mukhopadyay. Development of a New High Spatial Resolution (0.25°x 0.25°) Long Period(1901-2010) Daily Gridded Rainfall Data Set over India and its Comparison with Existing Data Sets over the Region. *Quarterly Journal of Meteorology, Hydrology and Geophysics*, 65(1):433-436, 2014.

S. Pandey, H. Bhandari, and B. Hardy. *Economic Costs of Drought and Rice Farmers' Coping Mechanisms: A Cross-Country Comparative Analysis*. International Rice Research Institute, Los Banos (Philippines), 2007.

J.L. Parcell and J. Caine. Drought Tolerance of Soybean Crops in Missouri. *Paper presented at the Southern Agricultural Economics Association (SAEA) Annual Meeting*, 2013.

P.V.V. Prasad, S.A. Staggenborg, and Z. Ristic. Impacts of drought and/or heat stress on physiological, developmental, growth, and yield processes of crop plants. In L. H. Ahuja and S.A. Saseendran, editors, *Response of Crops to Limited Water: Understanding and modelling Water Stress Effects on Plant Growth Processes*, pages 301-355. ASA-CSSA, Madison, WI, USA, 2008.

V. Prasanna. Impact of monsoon rainfall on the total foodgrain yield over India. *Journal of Earth System Science*, Jun 1;123(5):1129-45, 2014.

J. Sarkar (20111). Drought, its impacts and management: Scenario in India. In R. Shaw & H. Nguyen (Eds.), Droughts in Asian Monsoon Region (pp.67-86). Bingley, UK: Emerald Group Publishing Limited.

H. Sarsons. Rainfall and conflict: A cautionary tale. *Journal of Development Economics*, 115: 62-72, 2015.

W. Schlenker and M. Roberts. Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change. *Proceedings of the National Academy of Sciences of the United States of America*, 106(37):15594-15598, 2009.

M. Shah and B. Steinberg. Drought of Opportunities: Contemporaneous and Long Term Impacts of Rainfall Shocks on Human Capital. *Journal of Political Economy*, forthcoming.

K.D. Sharma. Rain-fed Agriculture Could Meet the Challenges of Food Security in India. *Current Science*, 100(11):1615-1616, 2011.

A.K. Srivastava, M. Rajeevan, and S.R. Kshirsagar. Development of a High Spatial Resolution Daily Gridded Temperature Data Set (1969-2005) for the Indian Region. *Atmospheric Science Letters*, 2009.

S. Vicente-Serrano, J. Lopez-Moreno, S. Begueria, J. Lorenzo-Lacruz, A. Sanchez-Lorenzo, J. Garcia-Ruiz, C. Azorin-Molina, E. Moran-Tejeda, J. Revuelto, R. Trigo, F. Coelho, and F. Espejo. Evidence of Increasing Drought Severity Caused by Temperature Rise in Southern Europe. *Environmental Research Letters*, 9:044001, 2014.

D. Wilhite. *Drought: A Global Assessment*. Routledge, London, 2000.

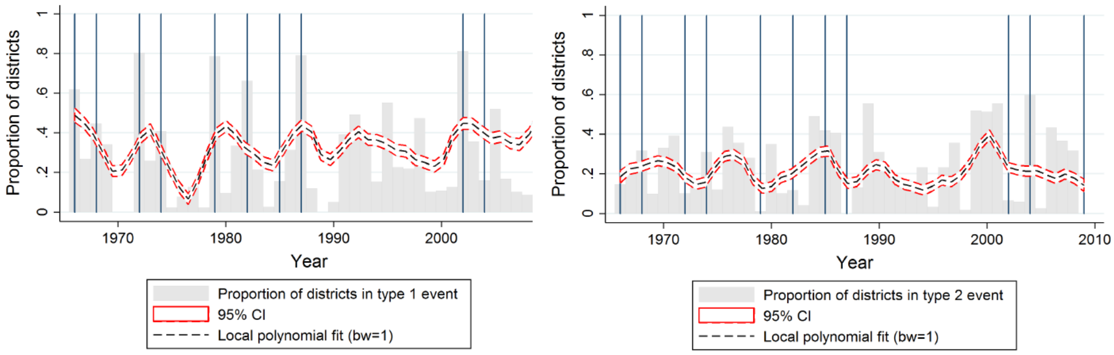
D. Wilhite and M. Glantz. Understanding the Drought Phenomenon: The Role of Definitions. *Water International*, 10(3):111-120, 1985.

T. Yu and B. Babcock. Are U.S. Corn and Soybeans Becoming More Drought Tolerant? *American Journal of Agricultural Economics*, 92(5):1310-1323, 2010.

9. FIGURES AND TABLES

1 Drought in India

Figure 1: Proportion of drought-affected districts (by type) (June-September only)

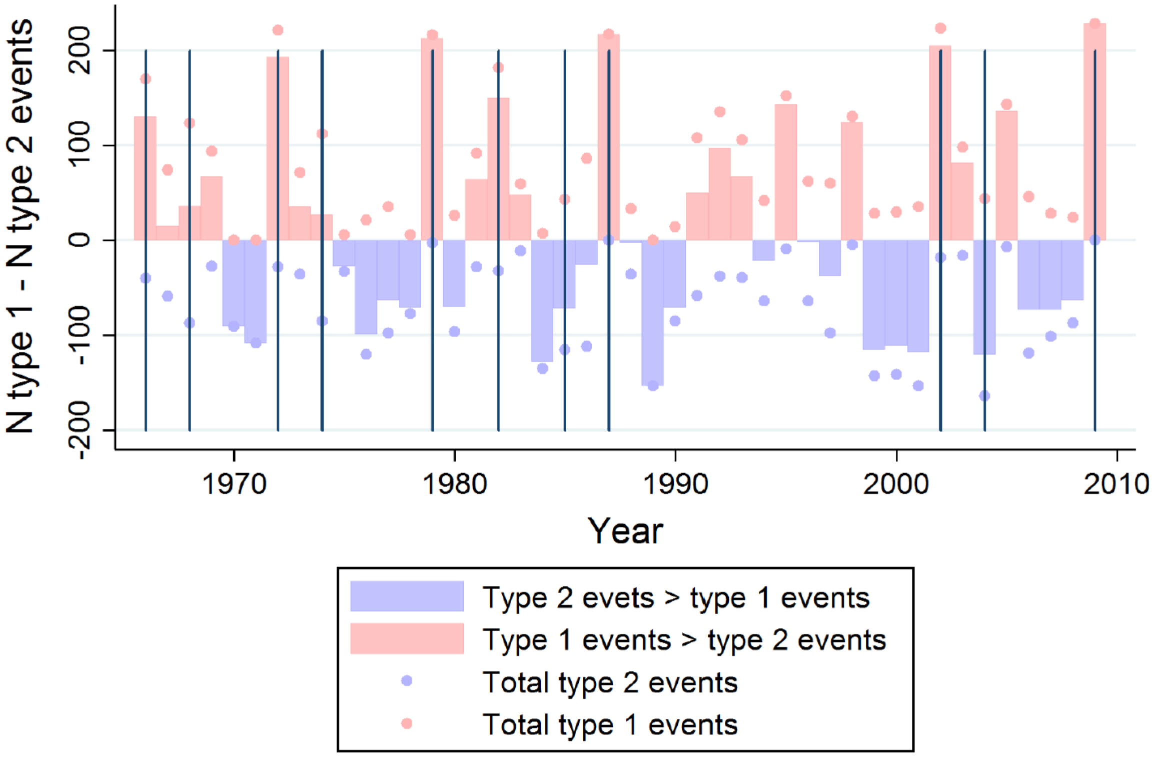


(a) **Type 1 events ("hot droughts")**
(below-average rain & above-average temperature)

(b) **Type 2 events ("cold droughts")**
(below-average rain & below-average temperature)

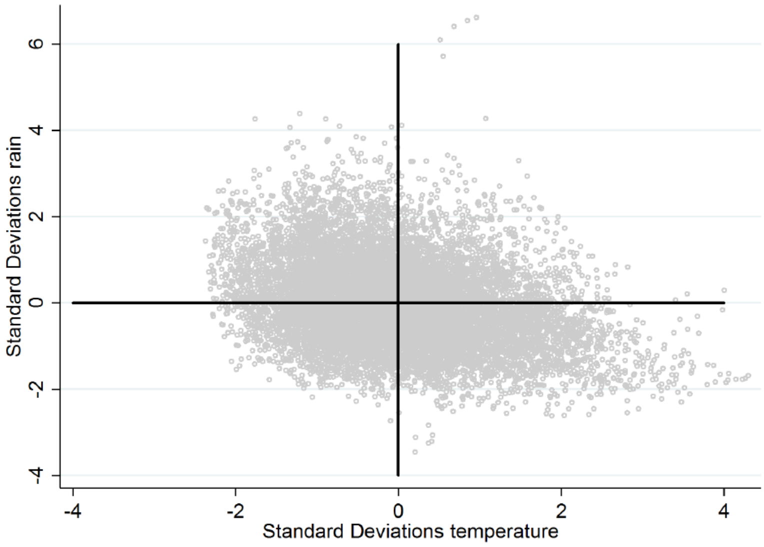
Notes: Type 1 events denote events where rainfall was below-average and temperature was above average. Conversely, Type 2 events refer to events where both rainfall and temperature were below-average. The rainfall average variable is calculated as the district mean cumulative rainfall from June-September from 1956-2009. The average temperature variable is calculated as the average degree days above the mean season temperature from June-September.
The solid vertical lines represent the years considered by the Indian Government as All-India drought years.
Source: Authors' own calculations

Figure 2: Type 1 droughts in excess of Type 2 droughts (June-September only)



Notes: The scatter points highlight the total number of droughts (by type) in a given year. In the case of Type 1 droughts (red scatter points) these can be interpreted directly (i.e. 200 means that 200 districts were affects by a Type 1 drought). However, in the case of the Type 2 droughts, these should be interpreted as the negative of the number (i.e. if the observed value is -100, this means there were 100 districts affected by Type 2 droughts).
Bar graphs show the number of affected districts affected by Type 1 droughts in excess of the number affected by Type 2 droughts. As a result a result a value of 50 would mean that there were 50 more districts affected by a Type 1 drought than affected by a Type 2 drought in a given year. The converse applies to a negative number, which highlights a higher number of districts affected by cold droughts in a given year.
The solid vertical lines represent the years considered by the Indian Government as All-India drought years.
Source: Authors own calculations

Figure 3: Rainfall-temperature quadrants

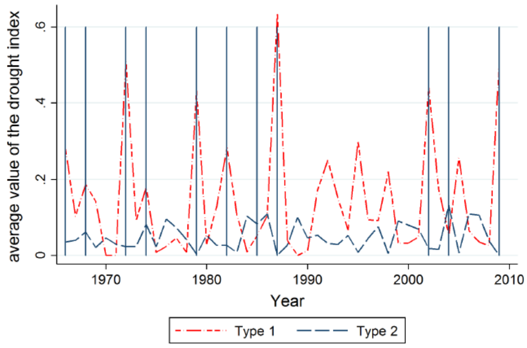


Notes: Long-term average rainfall is calculated as the average cumulative rainfall for the June-September period for the 1956-2009 period. Average temperature is defined as the number of degree-days above the mean daily June-September temperature for the 1956-2009 period.

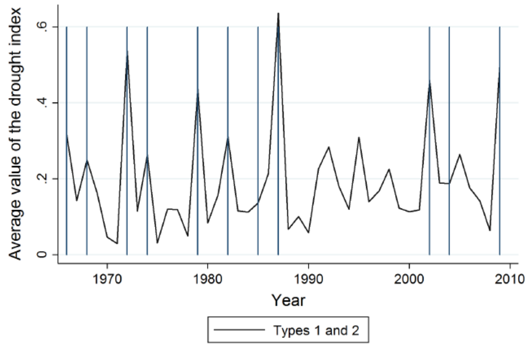
Table 1: Correlation coefficients and Spearman correlation coefficients

Correlation coefficients				
	Babcock-Yu	DI (q1)	DI (q2)	DI (q1 and q2)
Babcock-Yu	1.000			
DI (q1)	0.776	1.000		
DI (q2)	-0.181	-0.303	1.000	
DI (q12)	0.736	0.919	0.097	1.000
Spearman correlation coefficients				
	Babcock-Yu	DI (q1)	DI (q2)	DI (q1 and q2)
Babcock-Yu	1.000			
DI (q1)	0.992	1.000		
DI (q2)	-0.359	-0.359	1.000	
DI (q12)	0.821	0.830	0.217	1.000

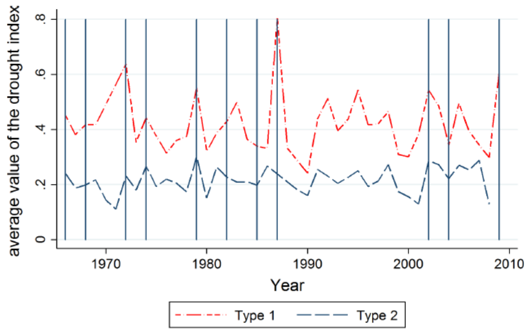
Figure 4: Average drought index value



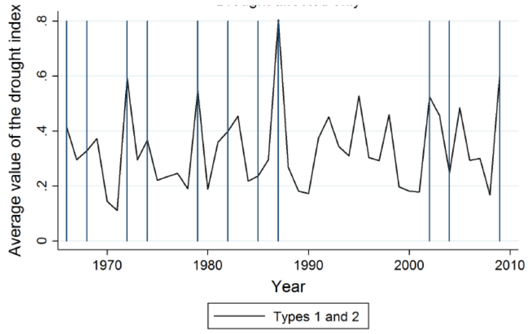
(a) Average index value, by type



(b) Average index value, all



(c) Average index value, by type (affected only)



(d) Average index value, all (affected only)

Table 2: Summary Statistics of observations in the sample

Variables	N	Mean	S.D	Min	Max
Cereal yield (t/ha)	12100	1.463	0.787	0.006	4.775
Cereal Area (1,000,000 ha)	12100	0.332	0.195	0.001	1.334
Barley yield (t/ha)	5842	1.383	0.688	0.048	5.400
Cereal area under barley production (%)	12031	0.015	0.036	0.000	0.320
Maize yield (t/ha)	10621	1.431	0.929	0.003	9.739
Cereal area under maize production (%)	12099	0.065	0.113	0.000	0.838
Millet yield (t/ha)	9852	0.798	0.439	0.000	4.000
Cereal area under millet production (%)	12100	0.131	0.220	0.000	1.000
Rice yield (t/ha)	11398	1.492	0.853	0.009	5.542
Cereal area under rice production (%)	12100	0.401	0.357	0.000	1.000
Sorghum yield (t/ha)	9694	0.774	0.434	0.001	9.836
Cereal area under sorghum production (%)	12066	0.148	0.225	0.000	0.929
Wheat yield (t/ha)	10275	1.643	0.878	0.046	6.324
Cereal area under wheat production (%)	12093	0.240	0.246	0.000	0.972
Proportion of net irrigated area (%)	12095	0.355	0.270	0.000	1.467
Rural population density (by gross cereal area)	11787	3.566	2.142	0.428	17.907
Fertiliser (t/1,000 ha)	11889	60.571	61.406	0.000	614.493
Cumulative rainfall (mm) (June-September)	12100	863.837	529.348	13.125	5313.428
Hot Degree-Days (HDD, June-September)	12100	94.422	47.204	2.697	278.413
Babcock-Yu index, June-September	12100	0.270	0.752	0.000	7.998
Drought index (quadrant 1)	12100	0.146	0.245	0.000	1
Drought index (quadrant 2)	12100	0.049	0.097	0.000	0.544
Drought index (quadrants 1 and 2)	12100	0.196	0.237	0.000	1

Notes: Rural population density is calculated by dividing total rural population by gross cropped area. Our hot degree-days measure is calculated based on average daily district temperature in the months of June-September for the period 1956-2009.

Table 3: Full sample results

	No dummies			Dummies		
Variables	1	2	3	4	5	6
Dummy (Type 1)				0.041**	0.019	
				(0.018)	(0.018)	
Drought index (Type 1)	-0.192***	-0.238***		-0.267***	-0.271***	
	(0.019)	(0.021)		(0.041)	(0.042)	
Dummy (Type 2)					0.022**	
					(0.011)	
Drought index (Type 2)		-0.391***			-0.471***	
		(0.036)			(0.052)	
Dummy (2 types)						-0.011
						(0.009)
Drought index (2 types)			-0.255***			-0.232***
			(0.021)			(0.028)
Constant	-0.361***	-0.326***	-0.313***	-0.359***	-0.323***	-0.316***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)
Time trends	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓
Controls						
Number of observations	12100	12100	12100	12100	12100	12100
Number of districts	275	275	275	275	275	275
R-squared a	0.705	0.712	0.711	0.705	0.713	0.711
R-squared w	0.719	0.726	0.725	0.72	0.727	0.725

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic district-specific trend.

Table 4: Results by crop - rice, wheat and maize (no dummies)

	Rice			Wheat			Maize		
Variables	1	2	3	4	5	6	7	8	9
Drought index (Type 1)	-0.249***	-0.303***		-0.128***	-0.161***		-0.138***	-0.118***	
	(0.020)	(0.023)		(0.016)	(0.017)		(0.026)	(0.029)	
Drought index (Type 2)		-0.469***			-0.263***			0.156***	
		(0.041)			(0.032)			(0.056)	
Drought index (2 types)			-0.323***			-0.174***			-0.084***
			(0.024)			(0.017)			(0.029)
Constant	-0.234***	-0.197***	-0.180***	-0.339***	-0.311***	-0.299***	-0.201***	-0.218***	-0.247***
	(0.034)	(0.034)	(0.034)	(0.027)	(0.028)	(0.029)	(0.050)	(0.050)	(0.050)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls									
Number of observations N	10560	10560	10560	8756	8756	8756	7656	7656	7656
Number of districts	240	240	240	199	199	199	174	174	174
R-squared a	0.532	0.542	0.541	0.713	0.716	0.716	0.398	0.398	0.396
R-squared w	0.555	0.565	0.563	0.727	0.731	0.73	0.428	0.429	0.426

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic district-specific trend.

Table 5: Results by crop - millet, sorghum and barley (no dummies)

	Millet			Sorghum			Barley		
Variables	1	2	3	4	5	6	7	8	9
Drought index (Type 1)	-0.240***	-0.287***		-0.159***	-0.188***		-0.013	-0.032	
	(0.039)	(0.044)		(0.031)	(0.035)		(0.027)	(0.029)	
Drought index (Type 2)		-0.381***			-0.239***			-0.152***	
		(0.078)			(0.066)			(0.045)	
Drought index (2 types)			-0.297***			-0.195***			-0.051*
			(0.045)			(0.036)			(0.027)
Constant	-0.724***	-0.692***	-0.684***	-0.728***	-0.710***	-0.703***	-0.349***	-0.328***	-0.315***
	(0.041)	(0.041)	(0.041)	(0.051)	(0.050)	(0.048)	(0.040)	(0.040)	(0.039)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls									
Number of observations	7172	7172	7172	6908	6908	6908	3432	3432	3432
Number of districts	163	163	163	157	157	157	78	78	78
R-squared a	0.429	0.434	0.434	0.335	0.337	0.337	0.767	0.768	0.767
R-squared w	0.459	0.463	0.463	0.369	0.371	0.371	0.78	0.781	0.78

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic district-specific trend.

Table 6: Results by crop - rice, wheat and maize (with dummies)

	Rice			Wheat			Maize		
Variables	1	2	3	4	5	6	7	8	9
Dummy (Type 1)	0.047	0.014		0.01	-0.006		0.157***	0.185***	
	(0.033)	(0.033)		(0.017)	(0.017)		(0.032)	(0.031)	
Drought index (Type 1)	-0.336***	-0.330***		-0.145***	-0.150***		-0.424***	-0.443***	
	(0.062)	(0.061)		(0.039)	(0.039)		(0.068)	(0.068)	
Dummy (Type 2)		-0.023*			0.01			0.169***	
		(0.013)			(0.011)			(0.027)	
Drought index (Type 2)		-0.380***			-0.300***			-0.439***	
		(0.072)			(0.057)			(0.107)	
Dummy (2 types)			-0.033***			-0.016**			0.169***
			(0.010)			(0.008)			(0.018)
Drought index (2 types)			-0.257***			-0.143***			-0.417***
			(0.029)			(0.024)			(0.046)
Constant	-0.232***	-0.199***	-0.191***	-0.339***	-0.309***	-0.305***	-0.198***	-0.187***	-0.184***
	(0.034)	(0.035)	(0.035)	(0.027)	(0.028)	(0.029)	(0.050)	(0.051)	(0.051)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls									
Number of observations	10560	10560	10560	8756	8756	8756	7656	7656	7656
Number of districts	240	240	240	199	199	199	174	174	174
R-squared a	0.532	0.542	0.541	0.713	0.716	0.716	0.4	0.406	0.406
R-squared w	0.556	0.565	0.564	0.727	0.731	0.73	0.431	0.436	0.436

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic district-specific trend.

Table 7: Results by crop - millet, sorghum and barley (with dummies)

	Millet			Sorghum			Barley		
Variables	1	2	3	4	5	6	7	8	9
Dummy (Type 1)	0.075**	0.062*		0.099**	0.093**		-0.005	-0.017	
	(0.031)	(0.031)		(0.043)	(0.045)		(0.024)	(0.024)	
Drought index (Type 1)	-0.378***	-0.392***		-0.340***	-0.351***		-0.003	0	
	(0.078)	(0.079)		(0.088)	(0.088)		(0.058)	(0.058)	
Dummy (Type 2)		0.087***			0.097***			0.004	
		(0.020)			(0.021)			(0.018)	
Drought index (Type 2)		-0.699***			-0.571***			-0.174**	
		(0.117)			(0.096)			(0.087)	
Dummy (2 types)			0.038**			0.059***			-0.022
			(0.017)			(0.022)			(0.014)
Drought index (2 types)			-0.371***			-0.310***			-0.004
			(0.058)			(0.050)			(0.046)
Constant	-0.727***	-0.690***	-0.677***	-0.726***	-0.692***	-0.679***	-0.350***	-0.327***	-0.324***
	(0.041)	(0.040)	(0.041)	(0.052)	(0.050)	(0.051)	(0.040)	(0.041)	(0.040)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls									
Number of observations	7172	7172	7172	6908	6908	6908	3432	3432	3432
Number of districts	163	163	163	157	157	157	78	78	78
R-squared a	0.43	0.435	0.434	0.336	0.339	0.338	0.767	0.767	0.767
R-squared w	0.459	0.465	0.463	0.37	0.373	0.372	0.78	0.781	0.781

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic district-specific trend.

Table 8: Cost estimates - No dummies

	Main				
	Type 1 (only)	Type 1 (sep)	Type 2 (sep)	2 types (sep)	2 types (Together)
	Full sample				
Av. yield loss (district) (t/ha)	0.12	0.16	0.11	0.14	0.12
Av. production loss (district) (1,000t)	40.49	50.95	39.11	45.57	41.28
Av. production cost (district) (mil usd)	9.78	12.30	9.44	11.00	9.97
Av. yearly total production loss (1,000 t)	3,332.87	4,229.59	2,689.97	6,919.56	6,270.95
Av. yearly total cost (mil usd)	804.79	1,021.32	649.55	1,670.87	1,514.25
	Rice				
Av. yield loss (district) (t/ha)	0.16	0.20	0.14	0.17	0.16
Av. production loss (district) (1,000t)	24.56	30.55	21.41	26.38	24.11
Av. production cost (district) (mil usd)	7.36	9.15	6.41	7.90	7.22
Av. yearly total production loss (1,000 t)	1,745.70	2,177.00	1,283.33	3,460.33	3,185.97
Av. yearly total cost (mil usd)	522.79	651.95	384.32	1,036.28	954.11
	Wheat				
Av. yield loss (district) (t/ha)	0.10	0.13	0.09	0.11	0.10
Av. production loss (district) (1,000t)	10.99	14.04	11.32	12.77	11.56
Av. production cost (district) (mil usd)	2.46	3.14	2.53	2.86	2.58
Av. yearly total production loss (1,000 t)	655.71	822.57	580.93	1,403.51	1,272.08
Av. yearly total cost (mil usd)	146.62	183.93	129.90	313.83	284.45
	Sorg				
Av. yield loss (district) (t/ha)	0.06	0.07	0.04	0.06	0.06
Av. production loss (district) (1,000t)	5.30	6.34	3.47	5.08	4.91
Av. production cost (district) (mil usd)	1.00	1.20	0.66	0.96	0.93
Av. yearly total production loss (1,000 t)	262.99	311.16	134.30	445.46	430.35
Av. yearly total cost (mil usd)	49.65	58.75	25.36	84.10	81.25
	Millet				
Av. yield loss (district) (t/ha)	0.09	0.11	0.06	0.09	0.08
Av. production loss (district) (1,000t)	5.96	7.23	4.40	5.92	5.61
Av. production cost (district) (mil usd)	1.04	1.27	0.77	1.04	0.98
Av. yearly total production loss (1,000 t)	289.48	351.53	185.12	536.65	508.72
Av. yearly total cost (mil usd)	50.74	61.61	32.45	94.06	89.16
	Barley				
Av. yield loss (district) (t/ha)	0.01	0.02	0.05	0.04	0.03
Av. production loss (district) (1,000t)	0.12	0.28	0.57	0.42	0.32
Av. production cost (district) (mil usd)	0.02	0.05	0.11	0.08	0.06
Av. yearly total production loss (1,000 t)	2.57	6.09	12.00	18.08	13.78
Av. yearly total cost (mil usd)	0.49	1.16	2.29	3.45	2.63
	Maize				
Av. yield loss (district) (t/ha)	0.08	0.07	- 0.04	0.02	0.04
Av. production loss (district) (1,000t)	2.13	1.81	- 1.02	0.52	0.95
Av. production cost (district) (mil usd)	0.34	0.29	- 0.16	0.08	0.15
Av. yearly total production loss (1,000 t)	111.05	94.82	- 44.45	50.36	91.88
Av. yearly total cost (mil usd)	17.91	15.29	- 7.17	8.12	14.82

Notes:Individual cereal prices used represent the 2008 weighted cereal prices converted into USD using the average monthly exchange rate over this year. Aggregate cereal prices also use the 2008 cereal-specific prices All numbers were rounded to two decimal places.

Table 9: Cost estimates - Dummies

	Main				
	Type 1 (only)	Type 1 (sep)	Type 2 (sep)	2 types (sep)	2 types (Together)
	Full sample				
Av. yield loss (district) (t/ha)	0.12	0.15	0.10	0.13	0.13
Av. production loss (district) (1,000t)	39.02	49.85	35.90	43.51	42.99
Av. production cost (district) (mil usd)	9.42	12.04	8.67	10.51	10.38
Av. yearly total production loss (1,000 t)	3,212.20	4,136.44	2,468.96	6,605.39	6,530.82
Av. yearly total cost (mil usd)	775.65	998.83	596.18	1,595.00	1,577.00
	Rice				
Av. yield loss (district) (t/ha)	0.16	0.20	0.15	0.17	0.17
Av. production loss (district) (1,000t)	23.79	30.38	22.52	26.79	26.34
Av. production cost (district) (mil usd)	7.12	9.10	6.74	8.02	7.89
Av. yearly total production loss (1,000 t)	1,691.33	2,165.23	1,351.88	3,517.11	3,480.88
Av. yearly total cost (mil usd)	506.51	648.43	404.85	1,053.28	1,042.43
	Wheat				
Av. yield loss (district) (t/ha)	0.10	0.13	0.09	0.11	0.11
Av. production loss (district) (1,000t)	10.87	14.08	10.88	12.59	12.51
Av. production cost (district) (mil usd)	2.43	3.15	2.43	2.81	2.80
Av. yearly total production loss (1,000 t)	637.44	824.96	557.78	1,382.74	1,377.14
Av. yearly total cost (mil usd)	142.54	184.47	124.72	309.19	307.94
	Sorg				
Av. yield loss (district) (t/ha)	0.05	0.06	0.02	0.04	0.04
Av. production loss (district) (1,000t)	5.06	5.83	1.85	4.07	3.96
Av. production cost (district) (mil usd)	0.95	1.10	0.35	0.77	0.75
Av. yearly total production loss (1,000 t)	247.75	286.11	71.40	357.51	347.50
Av. yearly total cost (mil usd)	46.78	54.02	13.48	67.50	65.61
	Millet				
Av. yield loss (district) (t/ha)	0.08	0.10	0.04	0.07	0.07
Av. production loss (district) (1,000t)	5.37	6.52	2.62	4.71	4.80
Av. production cost (district) (mil usd)	0.94	1.14	0.46	0.82	0.84
Av. yearly total production loss (1,000 t)	260.74	316.69	110.08	426.78	435.03
Av. yearly total cost (mil usd)	45.70	55.51	19.29	74.80	76.25
	Barley				
Av. yield loss (district) (t/ha)	0.01	0.03	0.05	0.04	0.04
Av. production loss (district) (1,000t)	0.12	0.30	0.56	0.43	0.42
Av. production cost (district) (mil usd)	0.02	0.06	0.11	0.08	0.08
Av. yearly total production loss (1,000 t)	2.46	6.39	11.96	18.35	18.06
Av. yearly total cost (mil usd)	0.47	1.22	2.28	3.50	3.45
	Maize				
Av. yield loss (district) (t/ha)	0.07	0.05	- 0.10	- 0.02	- 0.02
Av. production loss (district) (1,000t)	1.65	1.09	- 2.53	- 0.56	- 0.56
Av. production cost (district) (mil usd)	0.27	0.18	- 0.41	- 0.09	- 0.09
Av. yearly total production loss (1,000 t)	86.38	57.36	- 111.07	- 53.71	- 53.72
Av. yearly total cost (mil usd)	13.93	9.25	- 17.91	- 8.66	- 8.66

Notes:Individual cereal prices used represent the 2008 weighted cereal prices converted into USD using the average monthly exchange rate over this year. Aggregate cereal prices also use the 2008 cereal-specific prices All numbers were rounded to two decimal places.

10. TECHNICAL APPENDICIES

9.1. (A) – Data and Variables

Building the dataset:

We start with the raw data file, which already includes cumulative monthly rainfall data at the district level.

Generating rainfall variables

We generate three rainfall variables, which represent cumulative rainfall over three distinct periods, namely:

- A short Monsoon period (June-September), used in our main results
- An extended Monsoon period (May-December), used as a robustness check.
- An annual rainfall measure (January-December), used as a robustness check.

For each of the three cumulative rainfall variables we then define a long-term average rainfall measure for each district. To do this, we take the average total cumulative rainfall over the growing season for each district over the period 1956-2009. For our main specification, we define the growing season as June-September (see above).

For a given district, the general formula used is the following:

$$TR_{it} = \sum_{m=1}^{N} R_{mit}$$

Where the total rainfall in a given growing season for a given district *i* in a given year *t*, is equal to the sum of the monthly cumulative rainfall over the months (*m* to *M*) included in the growing season. To calculate the long-term average rainfall, we use the following formula:

$$LTAR_i = \frac{1}{54} \sum_{t=1956}^{T=2009} TR_{it}$$

Where the long-term average rainfall for a given district *i* is simply calculated as the average total rainfall in that district over the 1956-2009 period.

Generating temperature variables

We opt for a measure of cooling degree days (*CDD*) to capture accumulated heat over the growing season (June-September, in our main specification). This captures the number of degree-days above a reference temperature over a given time period. We use two alternative specifications for generating this variable.

In our main specification, we compute *CDD* based on the average temperature over the monsoon period (June-September). We test the sensitivity of results to alternative growing periods (May-December and Annual, respectively).

Our first step is to define the average daily temperature over the growing season for each district between 1956 and 2009. For any given district, *CDD* is estimated as:

$$CDD_{it} = \sum_{m=1}^{M} \sum_{d=1}^{D} (DT_{imd} - DTA_i)$$

Our long-term average *CDD* is then calculated as follows:

$$LTACDD_i = \frac{1}{54} \sum_{t=1956}^{T=2009} CDD_{it}$$

where *d* and *m* represent a given day and month included in the growing season and *N* and *M* respectively represent the total numbers of days in a given month and the total number of months in the growing season; *DT* denotes the average daily temperature in district *i* in day *d* of month *m*; and, *DTA* represents the average growing season daily temperature for a given district over the 1956-2009 period. Next, we create $LTACDD_i$, which is simply the average cumulative degree days above the mean daily temperature experienced by district *i* over the 1956-2009 period.

In addition, we create alternative degree day variables, where 30 degrees is used as a base temperature (instead of the average temperature over the growing season). We do this for the three alternative growing seasons. We have:

$$CDD30_{it} = \sum_{m=1}^{M} \sum_{d=1}^{D} (DT_{imd} - 30)$$

There are two main reasons why we use the district average as our reference temperature. The first has to do with the fact that we often pool a number of cereals together in our regressions. Given that there is no unique reference temperature for 'cereals', we prefer to use the district mean temperature. The second reason relates to sample size. Some districts may not have many days with an average daily temperature above 30 degrees. Thus, these districts have to be dropped from our sample since they have an invariant drought index.

Data inputting for controls used in the robustness check

In an attempt to gauge the sensitivity of our results to the inclusion of additional controls, we include a number of variables which could ostensibly be related to observed yields. These include area under cultivation, fertilizer, rural population density and proportion of irrigated area. For the latter three variables, there are a number of missing observations. For instance, in the case of rural population, the values are only recorded every 10 years when a census is performed. In other cases, there are also a number of missing observations. It should be stressed, however, that the imputations discussed below do not affect our main results since we use them only as a robustness check. We now discuss the assumptions underlying the imputation of each variable.

In the case of rural population we have data for 1971, 1981, 1991 and 2001. As such, between two waves of the census (e.g. between 1981 and 1991), we assume an exponential growth in between two waves. In the case of the pre-1971 and post-2001 data, we assume the growth rate of the subsequent and previous period, respectively, i.e. for the pre-1971 data, we assume the growth rate witnessed between 1971 and 1981. Similarly, for the post-2001 data, we assume the growth rate witnessed between 1991 and 2001. There are five occurrences where this process predicts impossible values (i.e. negative population). For these five we replace the impossible value by a missing value. Since we confine our sample to a balanced sample for all of our robustness checks, these observations are not used.

For the remaining variables, fertilizer and irrigation, we use linear interpolation (using the Stata command *ipol*) when the missing observation is between two values. In order to reduce the number of missing observations, we use linear extrapolation when they are either before the first or after the last observation (by using the Stata option *epolate* within the *ipolate* command). In the case of fertilizer we initially have 2,691 missing observations. However, the majority of these occur outside of our sample period (1966-2009). Only 270 observations occur within our sample period and are either interpolated or extrapolated. Out of these, the method produces 30 impossible (negative) values. In these cases, we simply replace the variable by a missing observation. This ensures that these impossible values do not affect our robustness checks.

We then carry out a similar exercise for the proportion of area under irrigation. First, we define the proportion of net irrigated area as the ratio of net irrigated area to net cropped area. Generating this variable results in five observations which have impossible (above 1) values. In three of these cases, since the values are below 1.05 we replace them by 1. We replace the two values that exceed 1.05 with a missing observation. Yet, after creating this variable, there are still 1,090 missing observations. Most of these occur due to missing data on net irrigated area. As such, we use the interpolation command with the extrapolation option and replace the (38) impossible (negative) values by missing values. For the remaining missing values we interpolate (with the extrapolate option) the proportion of net irrigated area and replace the negative values with a missing observation.

Generating additional variables (excluding drought indices)

We generate a number of additional variables:

- Rural population per hectare; obtained by dividing the rural population by gross cropped area
- Total cereal quantity produced; obtained by summing the quantity produced of each individual type of cereal
- Total cereal area; obtained by summing the areas devoted to each individual cereal
- Cereal yield; obtained by dividing the total quantity of cereal by total cereal area
- Individual cereal yield; obtained for each cereal by dividing total production by the total area devoted to a particular cereal

Generating drought indices

Crucial to our analysis is the construction of a novel drought index. For our purposes, we develop four drought indices. Below we describe the steps we carry out for each one.

Yu-Babcock index

We denote: total rainfall over the growing season TR_{it}; the mean of total rainfall over the growing season over 1956-2009 $LTAR_i$; and, the standard deviation of TR_{it} as $sdTR_i$. We then obtain the standardized variable using the following formula:

$$ZTR_{it} = \frac{TR_{it} - LTAR_i}{sdTR_i}$$

We proceed analogously for our CDD_{it} measure. Let: CDD_{it} be cumulative degree days above the long-term mean temperature of a district during the growing season; $LTACDD_i$ be long-term average cumulative degree days in the growing season; and, $sdCDD_i$ be the standard deviation of CDD_{it}. We compute the standardized variable:

$$ZCDD_{it} = \frac{CDD_{it} - LTACDD_i}{sdCDD_i}$$

Following this, we use the following to compute the Yu-Babcock index:

$$BYU_{it} = -\max(0, ZHDD_{it}) * \min(0, ZTR_{it})$$

Normalized indices

For the remaining indices, we use a normalized variable between 0 and 1, rather than a standardized value. We construct a variable, MTR_{it}, which is simply the negative of TR_{it} (i.e. $MTR_{it} = -TR_{it}$). The following is estimated to obtain NTR_{it} and $NCDD_{it}$:

$$NTR_{it} = \frac{MTR_{it} - MTR_i^{min}}{MTR_i^{max} - MTR_i^{min}}$$

$$NCDD_{it} = \frac{CDD_{it} - CDD_i^{min}}{CDD_i^{max} - CDD_i^{min}}$$

We differ from Yu and Babcock (2010) in creating a normalized version of the rain and temperature variables such that they vary strictly between 0 and 1, with 1 indicating the most extreme value (the highest temperature and lowest rainfall) and 0 indicating the lowest value. From these two variables, we then create a normalized rainfall-temperature index $NRTI_{it}$, which is simply a product of these variables:

$$NRTI_{it} = NTR_{it} * NCDD_{it}$$

From this, we obtain three additional indices. First:

$$DI1_{it} = \begin{cases} NRTI_{it} \ if \ ZTR_{it} < 0 \ \ and \ ZHDD_{it} > 0 \\ 0 \ otherwise \end{cases}$$

This is equivalent to a normalized version of the Yu-Babcock (2010) index. It only takes a non-zero value for events where rainfall deficiency and temperature are above average.

Second, we create our cold drought index analogously, using the following:

$$DI2_{it} = \begin{cases} NRTI_{it} \ if \ ZTR_{it} < 0 \ \ and \ ZHDD_{it} < 0 \\ 0 \ otherwise \end{cases}$$

This is the category omitted by Yu and Babcock. It only takes a non-zero value for events where rainfall deficiency is above-average and temperature is below average.

Finally, we create a third index, which combines $DI1_{it}$and $DI2_{it}$:

$$DI12_{it} = \begin{cases} NRTI_{it} \ if \ ZTR_{it} < 0 \\ 0 \ otherwise \end{cases}$$

This index takes non- zero values for all events where precipitation is below average.

The three main strengths of this index are that it: (i) defines all potential dry events; (ii) takes into account both temperature and rainfall; and, (iii) is increasing in temperature and rainfall deficiency. The main strength vis-à-vis the Yu and Babcock's index is that, by using the normalized negative of rainfall, we are able to construct an index that accounts for all potential types of drought without running into the problem of negative values that emerges from the interaction of the standardized variables.

Our index, however, still has four potential drawbacks. First, the normalization process is bounded between 0 and 1 which means that if a given district has a very large outlier in a given year but records lower values in other years then this would indicate a low value in the drought index thus masking what might have been a very bad drought year. This problem is less likely to arise in the standardized index. The second potential weakness arises from the multiplicative nature of the index. Thus, whenever temperature is close to 0 this can lead to a very low value of the drought index despite very deficient rainfall. Note, however, that this problem is only likely to arise for the $DI2_{it}$ index, and also applies to Yu and Babcock's index. Third, similar to their index, our index does not take into account intra-seasonal deficiencies in rainfall, which have been shown to have important impacts on agricultural productivity, e.g. Fishman (2016). Finally, similar to most drought indices, our index does not take into account (rare) multi-year droughts because this would require an index with 'memory' that takes into account soil moisture conditions. That said, since drought in India is mainly driven by variation in the annual monsoon, we argue that using an annual measure of monsoon rainfall is of greater relevance when estimating drought impact in our setting.

To ensure that the multiplicative relationship is not driving our results we also create a series of additive indices, which conceptually are very similar to our multiplicative indices. Specifically, the normalized additive rainfall temperature index, $NARTI_{it}$, is constructed as:

$$NARTI_{it} = NTR_{it} + NCDD_{it}$$

The only difference is that $NRTI_{it}$ is the product of rainfall and temperature while $NARTI_{it}$ is additive. From the latter, we have the following:

$$ADDI1_{it} = \begin{cases} NARTI_{it} \;\; if \; ZTR_{it} < 0 \;\; and \; ZCDD_{it} > 0 \\ 0 \; otherwise \end{cases}$$

$$ADDI2_{it} = \begin{cases} NARTI_{it} \; if \; ZTR < 0 \;\; and \; ZCDD_{it} < 0 \\ 0 \; otherwise \end{cases}$$

$$ADDI12_{it} = \begin{cases} NARTI_{it} \; if \; ZTR_{it} < 0 \\ 0 \; otherwise \end{cases}$$

where $ADDI1_{it}$, $ADDI2_{it}$and $ADDI12_{it}$are the additive equivalents to $DI1_{it}$, $DI2_{it}$ and $DI12_{it}$.

As a robustness check, we construct each of these indices for three distinct periods (i.e. June-September, May-December, and January-December) for our rain and temperature variables. We also inspect the correlations among all the indices. They are very high.

Determining the sample and generating trends

After developing the drought indices, we create a data file which includes only the observations between 1966 and 2009, i.e. our sample period. This choice is purely driven by data availability. Prior to 1966, our dependent variables (production and yields) are missing from the ICRISAT dataset and hence, would have resulted in districts being dropped. Prior to starting our analysis, we also dropped any districts for which at least one observation is missing in order to keep a balanced panel. We then generate district-specific quadratic trends using the following:

$$trend = t - 1965$$

$$trend\_sq = trend^2$$

where *t* denotes the year.

9.2. (B) – Estimating Economic Impact

As is made clear in the main text, the cost estimates generated in this paper are based purely on yield losses, without taking into account any potential changes in the cultivated area. Specifically, our cost estimates are derived using a series of seven steps. We detail all the assumptions and steps used throughout and discuss their relative strengths and weaknesses.

<u>Step 1 - Obtain a national estimate of crop prices and aggregate cereal price for each year</u>:

Crop prices: We generate a national weighted average of crop price by year (using the *egen* command and the user-written option *wtmean*), where the weight is determined by area of land under cultivation. For millet the process is slightly different since there are two kinds of millet in our sample (pearl millet and finger millet). As a result, we first generate, for each year, a weighted average of millet prices at the district-level. We then sum the total area under millet production (area under pearl millet + area under finger millet) and use this to establish a national weighted average millet price for a given year. Note that the *egen* command automatically adjusts the computation for missing data. As a result, districts that do not report prices for a given year are not included in the weighted average.

We then use 2008 crop prices to estimate prices (and costs) in USD: aggregate cereal price index (24.147 USD/quintal); rice (29.947 USD/quintal); wheat (22.360 USD/quintal); maize (16.125 USD/quintal); barley (19.089 USD/quintal); sorghum (18.88 USD/quintal); and, millet (17.5 USD/quintal). These prices are obtained by obtaining the weighted average of crop-specific prices in India for 2008 (in Rupees) and converting this by the averages of the 2008 monthly exchange rates extracted from: http://www.x-rates.com/average/?from=USD&to=INR&amount=1&year=2008.

All of the tables are also constructed using nominal yearly prices in Rupees and are available from the authors upon request.

Weaknesses and strengths of the assumptions:

National prices. For any given year, there are large differences in prices across districts. It could be argued that prices at the district- or state-level may be more appropriate. However, there are issues with missing price data at the district-level and, to a lesser extent, at the state-level even for cases where there is a non-zero quantity reported. This is the main reason why we opt for national prices.

Using fixed prices in USD: Using a fixed price throughout the sample period implies that the estimates of costs will vary depending on the chosen year since the choice of the year will, by definition drive both the exchange rate and the price level. Yet, output losses in the early periods are made comparable to losses in later periods since they are given the same value. Using nominal prices could lead to the economic cost of drought artificially increasing over time as nominal prices in most cereals have trended upwards over the sample period. In any case, we have also performed this exercise using nominal prices in rupees and the results are available from the authors upon request.

Step 2 - Estimate the regression of interest:

We estimate the following linear regression in which the coefficients for the two types of droughts are estimated separately, and there are: (i) no dummy variables; (ii) no controls; (iii) district-specific quadratic trends; (iv) district fixed effects; and, (v) year fixed effects.

$$\ln(y_{itc}) = \boldsymbol{\alpha_i} + \boldsymbol{\beta_t} + \boldsymbol{\delta_{i1}} * t + \boldsymbol{\delta_{i2}} * t^2 + \theta_1\, DI1_{it} + \theta_2\, DI2_{it} + \epsilon_{it}$$

Step 3 - Estimate the yield losses:

After Stata has generated the output for the regression in Step 2, we operationalise the following steps:

- Step 3.1 – Predict the natural logarithm of yield for drought when $DI12_{it} > 0$ (i.e. when the given district is drought affected). We do this by using the *predict* command following the estimation of the regression before replacing observations not affected by drought with an empty observation. We denote this variable $lyhat_d$**.** Note, to limit potential biases in the estimates of overall costs, we remove districts with implausible predicted yields, which we define as yields below 100 kg/ha and above 5 tonnes/ha). This assumption, however, affects very few observations, specifically 25 predicted values out of over 6,000 events in the full sample.
- Step 3.2 – Predict the natural logarithm of the yield variable under no drought (i.e. $DI12_{it} = 0; or\ DI1_{it} = 0\ and\ DI2_{it} = 0$). We rename the variables $DI1_{it}$and $DI2_{it}$ (e.g. they temporarily become $DI1original_{it}$ and $DI2original_{it}$) and create two new temporary variables: $DI1_{it} = 0$ and $DI2_{it} = 0$. We then use the *predict* command to

obtain predicted yield and a variable denoted $lyhat_{nd}$. The temporary $DI1_{it}$ and $DI2_{it}$ variables are deleted, and $DI1original_{it}$ and $DI2original_{it}$ are, respectively, renamed $DI1_{it}$ and $DI2_{it}$. We replace $lyhat_{nd}$ with an empty observation for every case where $DI12_{it} = 0$ (non-drought affected case). Note that when we estimate the predicted values using dummy variables, we also switch the dummy variable equal to zero when we replace the index coefficient with zero.

- Step 3.3 – Obtain yield values. All our predicted values are in logs. We thus convert these variables into levels and denote these variables $yhat_{nd}$ and $yhat_{d}$.
- Step 3.4 – Obtain predicted yield losses by simply subtracting the predicted yield under no drought (Step 3.2) by the actual predicted yield (Step 3.1) for all cases where $DI12_{it} > 0$. Formally, we calculate $ylosses = yhat_{nd} - yhat_{d}$.
- Step 3.5 – Obtain predicted yield losses by drought type by simply subtracting the predicted yield under no drought by the actual predicted yield for each type of drought separately. Thus, we estimate: $ylosses_1 = yhat_{nd} - yhat_{d}\ if\ DI1_{it} > 0$; and, $ylosses_2 = yhat_{nd} - yhat_{d}\ if\ DI2_{it} > 0$. Note the two types of drought are mutually exclusive (i.e. it is impossible for a district to simultaneously have a hot and a cold drought).

Step 4 - Estimate district-level production losses:

This requires three further steps:

- Step 4.1 - Convert land area to ha. As highlighted in the supporting documentation,[16] the land-use data is in 000's of ha. As a result we simply multiply cereal area by 1,000 to derive the cereal area in ha. Note that we do not model potential land-use changes as a response to a drought event, which is likely to happen. In other words, we exclude the possibility that land under cereal production might decline in a drought year given our focus on developing an inclusive drought index and estimating the marginal effects of drought on agricultural production.
- Step 4.2 - Convert yield losses to 1,000t/ha. Currently, our yield losses are in t/ha. We thus convert the yield losses to 1,000t/ha by dividing $ylosses$ by 1,000.

[16] See: http://vdsa.icrisat.ac.in/Include/document/all-apportioned-web-document.pdf

- Step 4.3 – Get the total district production losses (in 1,000t). Obtain the product of the variable obtained in Step 4.1 by that obtained in Step 4.2.

Step 5 - Estimate the district-level cost of production losses:

To do this we perform two further steps:

- Step 5.1 – Convert price data to million USD/1,000t**.** For the results shown in the paper, our price data are in USD per quintal (as explained in Step 3.1) and our production loss data (estimated in Step 4.3) are in 1,000t. To obtain the price data in million USD per 1,000t we divide our price level by 100. Note, a quintal is 100kg. To convert it into 1,000t (1,000,000kg) we multiply the price data by 10,000. However, since we want the data in million USD rather than USD, we divide this by 1,000,000. Thus, price*10,000/1,000,000 = price/100
- Step 5.2 – Obtain total value of production losses. After obtaining prices in million USD/1,000t we multiply the variable derived in Step 5.1 by the variable derived in Step 4.3 to obtain the total value of production losses in USD millions. Note, for our estimates in rupees, we apply the exact same procedure using yearly nominal prices.

Step 6 - Estimate total yearly production losses:

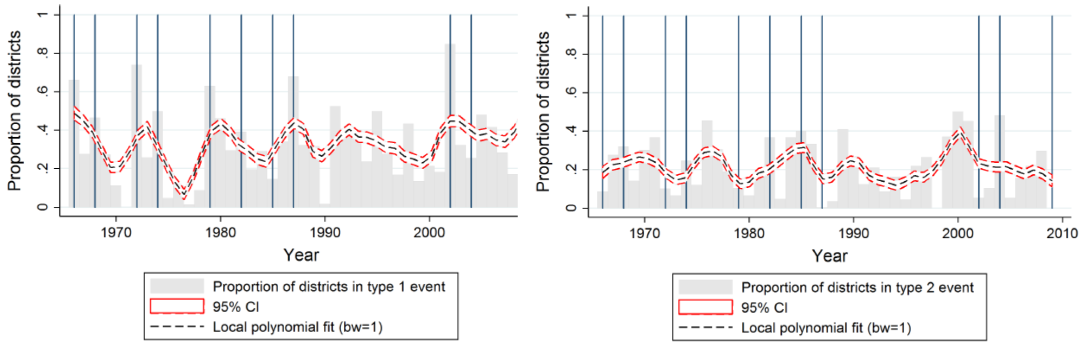
To obtain this measure in 1,000t we sum estimated total production losses of each affected district in a given year. We use the *total* function of the *egen* command. Note that the value in the table represents the unweighted average yearly loss.

Step 7 - Estimate total yearly production costs:

To obtain this measure in millions of rupees, we simply sum the estimated total value of the production losses of each affected district in a given year. We use the *egen* command with the *total* function. Note again that the value in the table represents the unweighted average yearly loss.

11. APPENDIX FIGURES AND TABLES

Figure A1: Proportion of drought-affected districts (by type, May-December growing season)



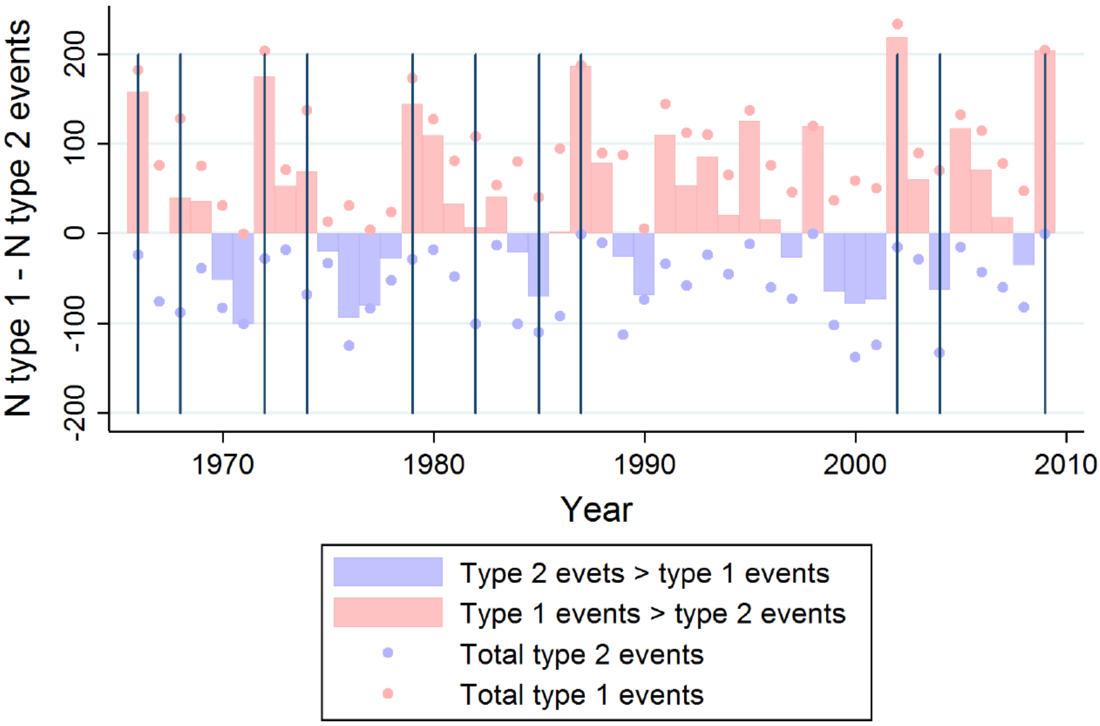
(a) **Type 1 events ("hot droughts")** **(below-average rain & above-average temperature)**

(b) **Type 2 events ("cold droughts")** **(below-average rain & below-average temperature)**

Notes: Type 1 events denote events where rainfall was below-average and temperature was above average. Conversely, Type 2 events refer to events where both rainfall and temperature were below-average. The rainfall average variable is calculated as the district mean cumulative rainfall from May-December from 1956-2009. The average temperature variable is calculated as the average degree days above the mean season temperature from May-December.

The solid vertical lines represent the years considered by the Indian Government as All-India drought years.

Figure A2: Type 1 droughts in excess of Type 2 droughts, May-December growing season



Notes: The scatter points highlight the total number of droughts (by type) in a given year. In the case of Type 1 droughts (red scatter points) these can be interpreted directly (i.e. 200 means that 200 districts were affects by a Type 1 drought). However, in the case of the Type 2 droughts, these should be interpreted as the negative of the number (i.e. if the observed value is -100, this means there were 100 districts affected by Type 2 droughts).
Bar graphs show the number of affected districts affected by Type 1 droughts in excess of the number affected by Type 2 droughts. As a result a result a value of 50 would mean that there were 50 more districts affected by a Type 1 drought than affected by a Type 2 drought in a given year. The converse applies to a negative number, which highlights a higher number of districts affected by cold droughts in a given year.
The solid vertical lines represent the years considered by the Indian Government as All-India drought years.

Table A1: Full sample robustness checks 1

	Cluster (state)			Squares			Controls		
Variables	1	2	3	4	5	6	7	8	9
Drought index (Type 1)	-0.192***	-0.238***		-0.068	-0.159***		-0.112***	-0.148***	
	(0.049)	(0.056)		(0.042)	(0.045)		(0.016)	(0.018)	
Drought index^2 (Type 1)				-0.201***	-0.126*				
				(0.073)	(0.074)				
Drought index (Type 2)		-0.391***			-0.239**			-0.269***	
		(0.078)			(0.098)			(0.038)	
Drought index^2 (Type 2)					-0.510*				
					(0.296)				
Drought index (2 types)			-0.255***			-0.252***			-0.161***
			(0.055)			(0.040)			(0.018)
Drought index^2 (2 types)						-0.004			
						(0.064)			
Constant	-0.372***	-0.346***	-0.347***	-0.363***	-0.328***	-0.313***	0.612***	0.558***	0.591***
	(0.031)	(0.033)	(0.032)	(0.022)	(0.022)	(0.022)	(0.202)	(0.197)	(0.198)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls							✓	✓	✓
Number of observations	12100	12100	12100	12100	12100	12100	8888	8888	8888
Number of districts	275	275	275	275	275	275	202	202	202
R-squared a	0.705	0.712	0.711	0.706	0.713	0.711	0.762	0.766	0.765
R-squared w	0.719	0.726	0.725	0.72	0.727	0.725	0.775	0.778	0.777

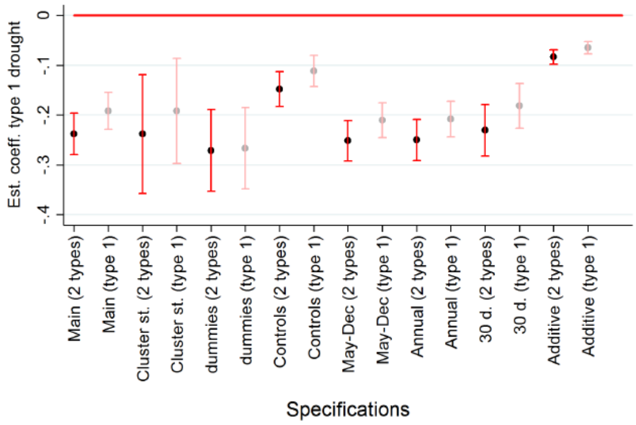
Notes: Values in parentheses denote clustered standard errors at the district level for columns 4-9. For columns 1-3 they denote clustered standard errors at the state level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a district-specific quadratic trend. In columns 7-9, we include 4 controls and their squares. These include the proportion of net irrigated area, the total cereal area, the total rural population density (rural population divided by gross cropped area) and fertilizer intensity. In columns 7-9 all districts for which at least one observation is missing for the control variables are dropped and this is the reason behind the decrease in the number of districts.

Table A2: Full sample robustness checks 2

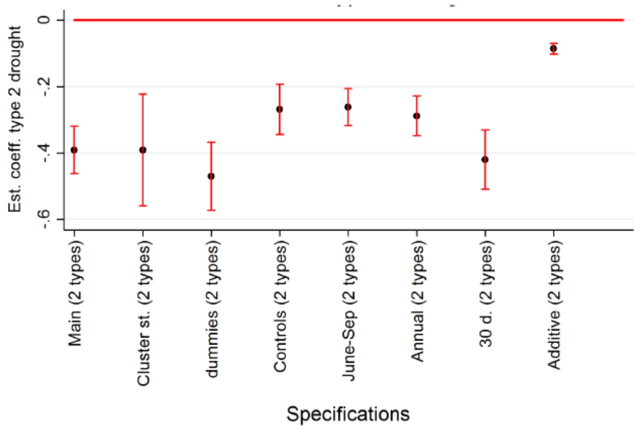
	May-December			Annual			Degree days (30 degrees)			Additive		
Variables	1	2	3	4	5	6	7	8	9	10	11	12
Drought index (Type 1)	-0.210***	-0.251***		-0.208***	-0.250***		-0.182***	-0.230***		-0.065***	-0.083***	
	(0.018)	(0.021)		(0.018)	(0.021)		(0.023)	(0.026)		(0.006)	(0.007)	
Drought index (Type 2)		-0.262***			-0.288***			-0.420***			-0.086***	
		(0.028)			(0.030)			(0.045)			(0.008)	
Drought index (2 types)			-0.253***			-0.254***			-0.264***			-0.084***
			(0.021)			(0.021)			(0.024)			(0.007)
Constant	-0.346***	-0.329***	-0.327***	-0.330***	-0.319***	-0.313***	-0.422***	-0.382***	-0.354***	-0.375***	-0.348***	-0.346***
	(0.023)	(0.022)	(0.021)	(0.023)	(0.023)	(0.022)	(0.024)	(0.024)	(0.023)	(0.021)	(0.021)	(0.021)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls												
Number of observationsN	12100	12100	12100	12100	12100	12100	11352	11352	11352	12100	12100	12100
Number of districts	275	275	275	275	275	275	258	258	258	275	275	275
R-squared a	0.711	0.716	0.716	0.71	0.716	0.716	0.704	0.71	0.712	0.704	0.711	0.711
R-squared w	0.725	0.73	0.73	0.724	0.73	0.73	0.718	0.724	0.726	0.719	0.725	0.725

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic trend. In columns 1-6 the drought index is constructed in the same way as the drought index in the paper. The only difference is the growing season used to construct the index. In column 1-3 the May-December period is used, whereas in columns 4-6 annual data is used. In columns 7-9, the index is built over the June-September period, but instead of using the average seasonal temperature for the construction of the hot degree-days variable we use an absolute threshold of 30 degrees. As a result, given that some districts do not experience daily temperatures this high, some districts drop from the sample since in a number of districts the harmful degree-days variable is always 0, which implies the drought index is also equal to 0 and does not vary. Finally, in columns 10-12, instead of multiplying the normalized negative rainfall by the normalized harmful degree-days variable, these two variables are added. In this specification, the average hot degree-days is based on the average June-September daily temperature between 1956-2009, rather than the absolute value of 30 degrees.

Figure A3: Estimated coefficients and 95% confidence intervals



(a) Coefficient - Type 1 events ("hot droughts")



(b) Coefficient Type 2 events ("cold droughts")

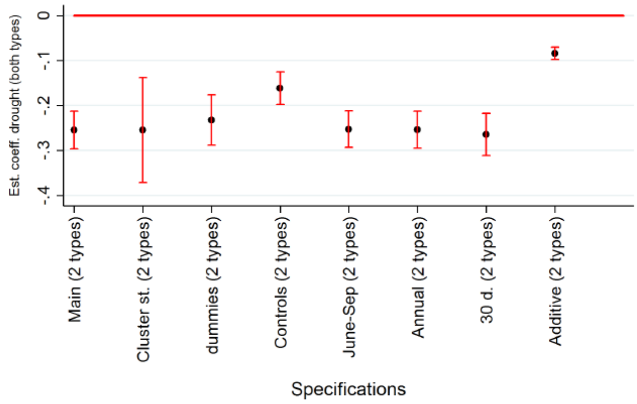


Table A3: Rice sample robustness checks 1

	Cluster (state)			Squares			Controls		
Variables	1	2	3	4	5	6	7	8	9
Drought index (Type 1)	-0.249***	-0.303***		-0.123	-0.235***		-0.205***	-0.255***	
	(0.045)	(0.052)		(0.078)	(0.080)		(0.020)	(0.023)	
Drought index$\hat{2}$ (Type 1)				-0.206*	-0.109				
				(0.122)	(0.122)				
Drought index (Type 2)		-0.469***			-0.383***			-0.386***	
		(0.080)			(0.112)			(0.039)	
Drought index$\hat{2}$ (Type 2)					-0.281				
					(0.405)				
Drought index (2 types)			-0.323***			-0.336***			-0.271***
			(0.054)			(0.061)			(0.023)
Drought index$\hat{2}$ (2 types)						0.024			
						(0.095)			
Constant	-0.362***	-0.334***	-0.333***	-0.235***	-0.197***	-0.180***	1.193***	1.134***	1.179***
	(0.060)	(0.056)	(0.056)	(0.034)	(0.034)	(0.035)	(0.243)	(0.239)	(0.240)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls							✓	✓	✓
Number of observations	10560	10560	10560	10560	10560	10560	7832	7832	7832
Number of districts	240	240	240	240	240	240	178	178	178
R-squared a	0.532	0.542	0.541	0.533	0.542	0.541	0.564	0.57	0.57
R-squared w	0.555	0.565	0.563	0.556	0.565	0.563	0.587	0.593	0.592

Notes: Values in parentheses denote clustered standard errors at the district level for columns 4-9. For columns 1-3 they denote clustered standard errors at the state level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a district-specific quadratic trend. In columns 7-9, we include 4 controls and their squares. These include the proportion of net irrigated area, the total cereal area, the total rural population density (rural population divided by gross cropped area) and fertilizer intensity. In columns 7-9 all districts for which at least one observation is missing for the control variables are dropped and this is the reason behind the decrease in the number of districts.

Table A4: Rice sample robustness checks 2

	May-December			Annual			Degree days (30 degrees)			Additive		
Variables	1	2	3	4	5	6	7	8	9	10	11	12
Drought index (Type 1)	-0.250***	-0.300***		-0.259***	-0.308***		-0.207***	-0.276***		-0.086***	-0.109***	
	(0.020)	(0.022)		(0.020)	(0.023)		(0.021)	(0.026)		(0.008)	(0.009)	
Drought index (Type 2)		-0.317***			-0.346***			-0.602***			-0.112***	
		(0.028)			(0.031)			(0.068)			(0.009)	
Drought index (2 types)			-0.302***			-0.312***			-0.323***			-0.110***
			(0.021)			(0.023)			(0.026)			(0.008)
Constant	-0.238***	-0.218***	-0.216***	-0.204***	-0.195***	-0.188***	-0.336***	-0.287***	-0.242***	-0.250***	-0.220***	-0.218***
	(0.034)	(0.035)	(0.034)	(0.035)	(0.035)	(0.034)	(0.037)	(0.037)	(0.037)	(0.034)	(0.034)	(0.033)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls												
Number of observations	10560	10560	10560	10560	10560	10560	9812	9812	9812	10560	10560	10560
Number of districts	240	240	240	240	240	240	223	223	223	240	240	240
R-squared a	0.538	0.545	0.545	0.538	0.546	0.546	0.528	0.54	0.54	0.531	0.542	0.542
R-squared w	0.561	0.567	0.567	0.561	0.569	0.568	0.552	0.563	0.563	0.554	0.565	0.565

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic trend. In columns 1-6 the drought index is constructed in the same way as the drought index in the paper. The only difference is the growing season used to construct the index. In column 1-3 the May-December period is used, whereas in columns 4-6 annual data is used. In columns 7-9, the index is built over the June-September period, but instead of using the average seasonal temperature for the construction of the hot degree-days variable we use an absolute threshold of 30 degrees. As a result, given that some districts do not experience daily temperatures this high, some districts drop from the sample since in a number of districts the harmful degree-days variable is always 0, which implies the drought index is also equal to 0 and does not vary. Finally, in columns 10-12, instead of multiplying the normalized negative rainfall by the normalized harmful degree-days variable, these two variables are added. In this specification, the average hot degree-days is based on the average June-September daily temperature between 1956-2009, rather than the absolute value of 30 degrees.

Table A5: Wheat sample robustness checks 1

	Cluster (state)			Squares			Controls		
Variables	1	2	3	4	5	6	7	8	9
Drought index (Type 1)	-0.128***	-0.161***		-0.123***	-0.185***		-0.100***	-0.126***	
	(0.028)	(0.031)		(0.039)	(0.039)		(0.018)	(0.018)	
Drought index^2 (Type 1)				-0.007	0.04				
				(0.064)	(0.064)				
Drought index (Type 2)		-0.263***			0.041			-0.183***	
		(0.078)			(0.084)			(0.035)	
Drought index^2 (Type 2)					-1.066***				
					(0.293)				
Drought index (2 types)			-0.174***			-0.226***			-0.133***
			(0.032)			(0.032)			(0.017)
Drought index^2 (2 types)						0.087			
						(0.054)			
Constant	-0.442***	-0.423***	-0.422***	-0.339***	-0.309***	-0.300***	-0.380*	-0.348*	-0.342*
	(0.062)	(0.062)	(0.063)	(0.027)	(0.028)	(0.029)	(0.197)	(0.202)	(0.202)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls							✓	✓	✓
Number of observations	8756	8756	8756	8756	8756	8756	6776	6776	6776
Number of districts	199	199	199	199	199	199	154	154	154
R-squared a	0.713	0.716	0.716	0.713	0.717	0.716	0.761	0.763	0.763
R-squared w	0.727	0.731	0.73	0.727	0.731	0.73	0.774	0.775	0.775

Notes: Values in parentheses denote clustered standard errors at the district level for columns 4-9. For columns 1-3 they denote clustered standard errors at the state level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a district-specific quadratic trend. In columns 7-9, we include 4 controls and their squares. These include the proportion of net irrigated area, the total cereal area, the total rural population density (rural population divided by gross cropped area) and fertilizer intensity. In columns 7-9 all districts for which at least one observation is missing for the control variables are dropped and this is the reason behind the decrease in the number of districts.

Table A6: Wheat sample robustness checks 2

	May-December			Annual			Degree days (30 degrees)			Additive		
Variables	1	2	3	4	5	6	7	8	9	10	11	12
Drought index (Type 1)	-0.141***	-0.162***		-0.143***	-0.167***		-0.112***	-0.142***		-0.045***	-0.058***	
	(0.014)	(0.015)		(0.014)	(0.016)		(0.014)	(0.015)		(0.006)	(0.006)	
Drought index (Type 2)		-0.124***			-0.161***			-0.251***			-0.059***	
		(0.021)			(0.022)			(0.032)			(0.007)	
Drought index (2 types)			-0.157***			-0.166***			-0.171***			-0.058***
			(0.014)			(0.015)			(0.016)			(0.005)
Constant	-0.323***	-0.312***	-0.319***	-0.315***	-0.308***	-0.309***	-0.361***	-0.334***	-0.312***	-0.345***	-0.323***	-0.323***
	(0.028)	(0.027)	(0.027)	(0.028)	(0.027)	(0.027)	(0.028)	(0.028)	(0.028)	(0.027)	(0.027)	(0.028)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls												
Number of observations	8756	8756	8756	8756	8756	8756	8712	8712	8712	8756	8756	8756
Number of districts	199	199	199	199	199	199	198	198	198	199	199	199
R-squared a	0.716	0.717	0.717	0.715	0.717	0.717	0.712	0.715	0.716	0.712	0.716	0.716
R-squared w	0.73	0.731	0.731	0.73	0.732	0.732	0.727	0.729	0.73	0.727	0.73	0.73

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic trend. In columns 1-6 the drought index is constructed in the same way as the drought index in the paper. The only difference is the growing season used to construct the index. In column 1-3 the May-December period is used, whereas in columns 4-6 annual data is used. In columns 7-9, the index is built over the June-September period, but instead of using the average seasonal temperature for the construction of the hot degree-days variable we use an absolute threshold of 30 degrees. As a result, given that some districts do not experience daily temperatures this high, some districts drop from the sample since in a number of districts the harmful degree-days variable is always 0, which implies the drought index is also equal to 0 and does not vary. Finally, in columns 10-12, instead of multiplying the normalized negative rainfall by the normalized harmful degree-days variable, these two variables are added. In this specification, the average hot degree-days is based on the average June-September daily temperature between 1956-2009, rather than the absolute value of 30 degrees.

Table A7: Maize sample robustness checks 1

	Cluster (state)			Squares			Controls		
Variables	1	2	3	4	5	6	7	8	9
Drought index (Type 1)	-0.138**	-0.118		0.181***	0.250***		-0.094***	-0.057*	
	(0.057)	(0.072)		(0.063)	(0.065)		(0.028)	(0.031)	
Drought index^2 (Type 1)				-0.520***	-0.585***				
				(0.102)	(0.102)				
Drought index (Type 2)		0.156			1.140***			0.264***	
		(0.147)			(0.155)			(0.060)	
Drought index^2 (Type 2)					-3.322***				
					(0.461)				
Drought index (2 types)			-0.084			0.336***			-0.017
			(0.082)			(0.056)			(0.031)
Drought index^2 (2 types)						-0.712***			
						(0.087)			
Constant	-0.141	-0.152	-0.154	-0.209***	-0.224***	-0.241***	0.182	0.242	0.13
	(0.094)	(0.090)	(0.087)	(0.050)	(0.051)	(0.051)	(0.183)	(0.182)	(0.184)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls							✓	✓	✓
Number of observations	7656	7656	7656	7656	6248	6248	6248		
Number of districts	174	174	174	174	174	174	142	142	142
R-squared a	0.398	0.398	0.396	0.4	0.405	0.403	0.414	0.417	0.413
R-squared w	0.428	0.429	0.426	0.431	0.436	0.433	0.446	0.448	0.444

Notes: Values in parentheses denote clustered standard errors at the district level for columns 4-9. For columns 1-3 they denote clustered standard errors at the state level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a district-specific quadratic trend. In columns 7-9, we include 4 controls and their squares. These include the proportion of net irrigated area, the total cereal area, the total rural population density (rural population divided by gross cropped area) and fertilizer intensity. In columns 7-9 all districts for which at least one observation is missing for the control variables are dropped and this is the reason behind the decrease in the number of districts.

Table A8: Maize sample robustness checks 2

	May-December			Annual			Degree days (30 degrees)			Additive		
Variables	1	2	3	4	5	6	7	8	9	10	11	12
Drought index (Type 1)	-0.113***	-0.085***		-0.108***	-0.081***		-0.113***	-0.094***		-0.038***	-0.025**	
	(0.022)	(0.026)		(0.022)	(0.026)		(0.029)	(0.033)		(0.009)	(0.010)	
Drought index (Type 2)		0.168***			0.177***			0.154**			0.058***	
		(0.044)			(0.042)			(0.068)			(0.014)	
Drought index (2 types)			-0.053**			-0.050*			-0.081**			0.004
			(0.027)			(0.026)			(0.032)			(0.010)
Constant	-0.219***	-0.233***	-0.276***	-0.217***	-0.224***	-0.277***	-0.231***	-0.248***	-0.254***	-0.231***	-0.252***	-0.328***
	(0.047)	(0.047)	(0.046)	(0.049)	(0.049)	(0.047)	(0.051)	(0.050)	(0.049)	(0.051)	(0.051)	(0.049)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls												
Number of observations	7656	7656	7656	7656	7656	7656	7612	7612	7612	7656	7656	7656
Number of districts	174	174	174	174	174	174	173	173	173	174	174	174
R-squared a	0.397	0.399	0.395	0.397	0.399	0.395	0.397	0.397	0.396	0.396	0.399	0.395
R-squared w	0.428	0.43	0.426	0.428	0.429	0.426	0.427	0.428	0.427	0.427	0.429	0.425

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic trend. In columns 1-6 the drought index is constructed in the same way as the drought index in the paper. The only difference is the growing season used to construct the index. In column 1-3 the May-December period is used, whereas in columns 4-6 annual data is used. In columns 7-9, the index is built over the June-September period, but instead of using the average seasonal temperature for the construction of the hot degree-days variable we use an absolute threshold of 30 degrees. As a result, given that some districts do not experience daily temperatures this high, some districts drop from the sample since in a number of districts the harmful degree-days variable is always 0, which implies the drought index is also equal to 0 and does not vary. Finally, in columns 10-12, instead of multiplying the normalized negative rainfall by the normalized harmful degree-days variable, these two variables are added. In this specification, the average hot degree-days is based on the average June-September daily temperature between 1956-2009, rather than the absolute value of 30 degrees.

Table A9: Millet sample robustness checks 1

	Cluster (state)			Squares			Controls		
Variables	1	2	3	4	5	6	7	8	9
Drought index (Type 1)	-0.240**	-0.287**		-0.011	-0.091		-0.151***	-0.183***	
	(0.098)	(0.111)		(0.075)	(0.082)		(0.033)	(0.038)	
Drought index^2 (Type 1)				-0.370***	-0.309**				
				(0.123)	(0.125)				
Drought index (Type 2)		-0.381*			0.035			-0.243***	
		(0.200)			(0.154)			(0.087)	
Drought index^2 (Type 2)					-1.416***				
					(0.444)				
Drought index (2 types)			-0.297**			-0.175**			-0.190***
			(0.112)			(0.079)			(0.040)
Drought index^2 (2 types)						-0.205*			
						(0.115)			
Constant	-0.671***	-0.647***	-0.647***	-0.737***	-0.704***	-0.687***	0.712**	0.691**	0.702**
	(0.049)	(0.057)	(0.056)	(0.041)	(0.041)	(0.041)	(0.306)	(0.302)	(0.306)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls							✓	✓	✓
Number of observations	7172	7172	7172	7172	7172	7172	5588	5588	5588
Number of districts	163	163	163	163	163	163	127	127	127
R-squared a	0.429	0.434	0.434	0.431	0.435	0.434	0.502	0.504	0.504
R-squared w	0.459	0.463	0.463	0.46	0.465	0.463	0.529	0.531	0.531

Notes: Values in parentheses denote clustered standard errors at the district level for columns 4-9. For columns 1-3 they denote clustered standard errors at the state level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a district-specific quadratic trend. In columns 7-9, we include 4 controls and their squares. These include the proportion of net irrigated area, the total cereal area, the total rural population density (rural population divided by gross cropped area) and fertilizer intensity. In columns 7-9 all districts for which at least one observation is missing for the control variables are dropped and this is the reason behind the decrease in the number of districts.

Table A10: Millet sample robustness checks 2

	May-December			Annual			Degree days (30 degrees)			Additive		
Variables	1	2	3	4	5	6	7	8	9	10	11	12
Drought index (Type 1)	-0.262***	-0.295***		-0.256***	-0.295***		-0.244***	-0.275***		-0.078***	-0.095***	
	(0.036)	(0.041)		(0.037)	(0.042)		(0.047)	(0.055)		(0.013)	(0.015)	
Drought index (Type 2)		-0.205***			-0.266***			-0.278***			-0.074***	
		(0.049)			(0.053)			(0.099)			(0.017)	
Drought index (2 types)			-0.286***			-0.292***			-0.311***			-0.088***
			(0.041)			(0.042)			(0.051)			(0.014)
Constant	-0.709***	-0.692***	-0.702***	-0.691***	-0.679***	-0.683***	-0.756***	-0.729***	-0.696***	-0.743***	-0.716***	-0.732***
	(0.039)	(0.039)	(0.039)	(0.040)	(0.040)	(0.040)	(0.041)	(0.041)	(0.042)	(0.042)	(0.041)	(0.042)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls												
Number of observations	7172	7172	7172	7172	7172	7172	6952	6952	6952	7172	7172	7172
Number of districts	163	163	163	163	163	163	158	158	158	163	163	163
R-squared a	0.436	0.438	0.438	0.434	0.438	0.438	0.429	0.431	0.435	0.428	0.431	0.431
R-squared w	0.465	0.467	0.467	0.463	0.467	0.467	0.459	0.461	0.464	0.457	0.461	0.46

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic trend. In columns 1-6 the drought index is constructed in the same way as the drought index in the paper. The only difference is the growing season used to construct the index. In column 1-3 the May-December period is used, whereas in columns 4-6 annual data is used. In columns 7-9, the index is built over the June-September period, but instead of using the average seasonal temperature for the construction of the hot degree-days variable we use an absolute threshold of 30 degrees. As a result, given that some districts do not experience daily temperatures this high, some districts drop from the sample since in a number of districts the harmful degree-days variable is always 0, which implies the drought index is also equal to 0 and does not vary. Finally, in columns 10-12, instead of multiplying the normalized negative rainfall by the normalized harmful degree-days variable, these two variables are added. In this specification, the average hot degree-days is based on the average June-September daily temperature between 1956-2009, rather than the absolute value of 30 degrees.

Table A11: Sorghum sample robustness checks 1

	Cluster (state)			Squares			Controls		
Variables	1	2	3	4	5	6	7	8	9
Drought index (Type 1)	-0.159*	-0.188*		0.119	0.082		-0.097***	-0.114***	
	(0.074)	(0.086)		(0.094)	(0.105)		(0.028)	(0.031)	
Drought index$\hat{2}$ (Type 1)				-0.453***	-0.429***				
				(0.149)	(0.156)				
Drought index (Type 2)		-0.239			0.392**			-0.122**	
		(0.135)			(0.185)			(0.060)	
Drought index$\hat{2}$ (Type 2)					-2.063***				
					(0.527)				
Drought index (2 types)			-0.195*			0.008			-0.115***
			(0.089)			(0.099)			(0.031)
Drought index$\hat{2}$ (2 types)						-0.344**			
						(0.146)			
Constant	-0.652***	-0.637***	-0.636***	-0.737***	-0.716***	-0.699***	-0.022	-0.059	-0.056
	(0.060)	(0.066)	(0.067)	(0.050)	(0.049)	(0.049)	(0.207)	(0.205)	(0.205)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls							✓	✓	✓
Number of observations	6908	6908	6908	6908	6908	6908	6380	6380	6380
Number of districts	157	157	157	157	157	157	145	145	145
R-squared a	0.335	0.337	0.337	0.337	0.34	0.338	0.366	0.367	0.367
R-squared w	0.369	0.371	0.371	0.371	0.374	0.373	0.4	0.4	0.4

Notes: Values in parentheses denote clustered standard errors at the district level for columns 4-9. For columns 1-3 they denote clustered standard errors at the state level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a district-specific quadratic trend. In columns 7-9, we include 4 controls and their squares. These include the proportion of net irrigated area, the total cereal area, the total rural population density (rural population divided by gross cropped area) and fertilizer intensity. In columns 7-9 all districts for which at least one observation is missing for the control variables are dropped and this is the reason behind the decrease in the number of districts.

Table A12: Sorghum sample robustness checks 2

	May-December			Annual			Degree days (30 degrees)			Additive		
Variables	1	2	3	4	5	6	7	8	9	10	11	12
Drought index (Type 1)	-0.198***	-0.215***		-0.214***	-0.228***		-0.132***	-0.158***		-0.050***	-0.059***	
	(0.030)	(0.036)		(0.033)	(0.038)		(0.036)	(0.042)		(0.011)	(0.013)	
Drought index (Type 2)		-0.102**			-0.105*			-0.232***			-0.042***	
		(0.047)			(0.054)			(0.085)			(0.016)	
Drought index (2 types)			-0.201***			-0.214***			-0.197***			-0.053***
			(0.036)			(0.038)			(0.039)			(0.013)
Constant	-0.703***	-0.697***	-0.717***	-0.671***	-0.671***	-0.699***	-0.772***	-0.753***	-0.720***	-0.750***	-0.741***	-0.758***
	(0.049)	(0.048)	(0.047)	(0.048)	(0.048)	(0.046)	(0.049)	(0.048)	(0.047)	(0.049)	(0.048)	(0.045)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls												
Number of observations	6908	6908	6908	6908	6908	6908	6732	6732	6732	6908	6908	6908
Number of districts	157	157	157	157	157	157	153	153	153	157	157	157
R-squared a	0.34	0.341	0.34	0.341	0.342	0.341	0.338	0.339	0.341	0.334	0.335	0.335
R-squared w	0.374	0.375	0.374	0.375	0.376	0.375	0.372	0.374	0.375	0.368	0.369	0.369

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic trend. In columns 1-6 the drought index is constructed in the same way as the drought index in the paper. The only difference is the growing season used to construct the index. In column 1-3 the May-December period is used, whereas in columns 4-6 annual data is used. In columns 7-9, the index is built over the June-September period, but instead of using the average seasonal temperature for the construction of the hot degree-days variable we use an absolute threshold of 30 degrees. As a result, given that some districts do not experience daily temperatures this high, some districts drop from the sample since in a number of districts the harmful degree-days variable is always 0, which implies the drought index is also equal to 0 and does not vary. Finally, in columns 10-12, instead of multiplying the normalized negative rainfall by the normalized harmful degree-days variable, these two variables are added. In this specification, the average hot degree-days is based on the average June-September daily temperature between 1956-2009, rather than the absolute value of 30 degrees.

Table A13: Barley sample robustness checks 1

	Cluster (state)			Squares			Controls		
Variables	1	2	3	4	5	6	7	8	9
Drought index (Type 1)	-0.013	-0.032		-0.046	-0.085*		-0.007	-0.025	
	(0.035)	(0.048)		(0.048)	(0.049)		(0.027)	(0.028)	
Drought index$\hat{2}$ (Type 1)				0.054	0.087				
				(0.076)	(0.075)				
Drought index (Type 2)		-0.152			-0.111			-0.137***	
		(0.078)			(0.132)			(0.045)	
Drought index$\hat{2}$ (Type 2)					-0.175				
					(0.455)				
Drought index (2 types)			-0.051			-0.136***			-0.042
			(0.049)			(0.043)			(0.027)
Drought index$\hat{2}$ (2 types)						0.149**			
						(0.070)			
Constant	-0.291***	-0.280***	-0.280***	-0.348***	-0.325***	-0.313***	-0.223**	-0.242**	-0.210**
	(0.048)	(0.052)	(0.049)	(0.040)	(0.040)	(0.038)	(0.095)	(0.095)	(0.093)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls							✓	✓	✓
Number of observations	3432	3432	3432	3432	3432	3432	3432	3432	3432
Number of districts	78	78	78	78	78	78	78	78	78
R-squared a	0.767	0.768	0.767	0.767	0.768	0.767	0.768	0.768	0.768
R-squared w	0.78	0.781	0.78	0.78	0.781	0.781	0.781	0.782	0.782

Notes: Values in parentheses denote clustered standard errors at the district level for columns 4-9. For columns 1-3 they denote clustered standard errors at the state level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a district-specific quadratic trend. In columns 7-9, we include 4 controls and their squares. These include the proportion of net irrigated area, the total cereal area, the total rural population density (rural population divided by gross cropped area) and fertilizer intensity. In columns 7-9 all districts for which at least one observation is missing for the control variables are dropped and this is the reason behind the decrease in the number of districts.

Table A14: Barley sample robustness checks 2

	May-December			Annual			Degree days (30 degrees)			Additive		
Variables	1	2	3	4	5	6	7	8	9	10	11	12
Drought index (Type 1)	-0.071***	-0.067***		-0.070***	-0.073***		-0.011	-0.033		-0.006	-0.013	
	(0.022)	(0.023)		(0.022)	(0.023)		(0.027)	(0.028)		(0.009)	(0.009)	
Drought index (Type 2)		0.019			-0.016			-0.173***			-0.032***	
		(0.031)			(0.038)			(0.045)			(0.010)	
Drought index (2 types)			-0.055**			-0.067***			-0.047*			-0.020***
			(0.022)			(0.022)			(0.026)			(0.008)
Constant	-0.288***	-0.292***	-0.304***	-0.287***	-0.285***	-0.292***	-0.351***	-0.327***	-0.319***	-0.348***	-0.331***	-0.313***
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.041)	(0.041)	(0.039)	(0.040)	(0.039)	(0.037)
Time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls												
Number of observations	3432	3432	3432	3432	3432	3432	3432	3432	3432	3432	3432	3432
Number of districts	78	78	78	78	78	78	78	78	78	78	78	78
R-squared a	0.768	0.768	0.767	0.768	0.768	0.768	0.767	0.768	0.767	0.767	0.767	0.767
R-squared w	0.781	0.781	0.781	0.781	0.781	0.781	0.78	0.781	0.78	0.78	0.781	0.781

Notes: Values in parentheses denote clustered standard errors at the district level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. District trends denote a quadratic trend. In columns 1-6 the drought index is constructed in the same way as the drought index in the paper. The only difference is the growing season used to construct the index. In column 1-3 the May-December period is used, whereas in columns 4-6 annual data is used. In columns 7-9, the index is built over the June-September period, but instead of using the average seasonal temperature for the construction of the hot degree-days variable we use an absolute threshold of 30 degrees. As a result, given that some districts do not experience daily temperatures this high, some districts drop from the sample since in a number of districts the harmful degree-days variable is always 0, which implies the drought index is also equal to 0 and does not vary. Finally, in columns 10-12, instead of multiplying the normalized negative rainfall by the normalized harmful degree-days variable, these two variables are added. In this specification, the average hot degree-days is based on the average June-September daily temperature between 1956-2009, rather than the absolute value of 30 degrees.