CLIMATE CHANGE AND PRODUCTION RISK: EVIDENCE FROM INDIAN AGRICULTURE*$$

Shreekant Gupta#, Partha Sen and Saumya Verma

Delhi School of Economics

September 2016

Abstract

In India, agriculture accounts for about sixty percent of employment. How would climate change, that is expected to hit agriculture in poorer countries very hard, affect India’s agriculture? We study the impact of climate change on the mean and variance of yields of three food grains — rice (India’s major crop), sorghum and pearl millet — at the district level using a large panel dataset from 1966-2011. An agricultural production function is estimated with exogenous climate variables -- precipitation and temperature -- controlling for other non climate inputs. We hypothesise that climate variability increases production risk. To capture the impact of climate extremes, climate variables are modelled as anomalies. The results show that climate change adversely affects mean and variance of crop yields. Higher temperature reduces crop yields, on an average. Climate extremes, rainfall extremes in particular increase crop yield variability.

Keywords: Climate change; agricultural impacts; developing countries.
JEL Codes: O13, Q54, R11

*This paper has benefited from the comments of the participants, especially Takashi Kurosaki and Yoshihisa Godo, at the seminar at the Institute for Economic Research, Hitotsubashi University.

$ The paper was completed while Partha Sen was visiting the Institute for Economic Research, Hitotsubashi University. He thanks the institute for its hospitality.

# Corresponding author: sgupta@econdse.org
1. Introduction

Significant warming of the Earth’s surface and ocean temperatures over the past century has been attributed to anthropogenic activities (IPCC 2007). For India, annual mean temperature has increased gradually but continuously over 1901-2007, along with accelerated warming in recent years (Kothawale et al. 2010). Simulation results from global and regional climate models for India predict a significant increase in annual mean temperatures and summer monsoon rainfall, along with significant inter annual variability in both, which will manifest in increased intensity, higher frequency extreme events in the 2030’s (GoI 2010).

Developing countries, like India, which are located in lower altitudes are expected to be the worst affected with losses in agricultural production of up to 21 percent (Cline 2007). In India, agriculture alone (excluding forestry and fisheries) accounts for 12 percent of Gross Domestic Product (GDP) and is a major source of livelihood for around 69 percent of the rural population (GoI 2011). According to an Indian Planning Commission Report, around 80 percent of the poor reside in rural areas (GoI 2013), and depend on agriculture for their sustenance.

The present study looks at the effect of climate variables (in addition to other inputs) on agricultural production in India. We pay special attention (neglected in studies so far) on “weather anomalies”. We focus on three food grains grown in India, namely, rice, sorghum and pearl millet. Rice and millets have been chosen on purpose. Rice is the staple crop of Asia and is central to the food security of about half of the world’s population (FAO 2013). The crop accounts for approximately 30 percent of the total dietary intake, globally and in South Asia (Lobell et al. 2008). India accounts for approximately 67 percent of total rice production in South Asia. The crop accounts for 23.3 percent of gross cropped area and about 43 percent of total food grain production in India (Singh 2009). Rice production in the tropics is sensitive to climatic factors (temperature, rainfall, and solar radiation) which affect the crop in various ways during different stages of its growth (Yoshida 1978).

Coarse cereals like sorghum and pearl millet are the major staple food for the poor, in addition to being used as animal feed and for alcohol production (Basavraj et al. 2010). In terms of food grain production millets rank fourth in India behind rice, wheat and maize (FAO 2011). Sorghum (jowar) and pearl millet (bajra) are the two millets considered in this study. The Green Revolution, which took place in the 1960’s led to a substantial increase in rice and wheat production. Production of millets has more or less remained constant between1966-2006 whereas that of rice and wheat has increased by 125 percent and 285 percent, respectively (MNI 2009).

The present study examines the impact of climate change on mean and variance of yields of three food grains grown in India, namely, rice (India’s major crop), sorghum and pearl millet. The analysis is conducted with district-level data from 1966 to 2011, the period for which continuous data is available at the district level. Changes in climate are found to affect crop yield levels and variances in a crop specific fashion. For rice and pearl millet, higher rainfall is found to increase mean yield whereas higher temperatures adversely affect crop yields. Further, drought and flood events are found to exacerbate crop yield variability, with temperatures significantly higher than the long period average increasing yield variability of millets.
We have not considered wheat, although it is a major crop, because during the period under study, it witnessed major changes in technology—the so-called “green revolution”. It is also a crop that is cultivated in those parts of India where big land owners are dominant. So the prices of wheat and its inputs are subject to political pressure that would take us beyond what we want to explain. Therefore, we stick to India’s production of rice and the coarse cereals—pearl millet and sorghum.

The paper is organized as follows. In the next section, projected trends and regional variation in climate variables (rainfall and temperature) for India are discussed, elaborating the crucial role climate variables play in agricultural production. Section 3 elaborates our conceptual framework along with related literature on the impact of climate on agriculture, particularly with regard to India. Section 4 describes the data and econometric methodology. Section 5 presents the econometric tests performed prior to estimation with results. Section 6 presents and interprets the results of our analysis. Section 7 concludes the paper.

2. Climate Change and Agriculture in India

2.1 Trends and regional variation in rainfall and temperature

Considerable heterogeneity in precipitation trends across regions and seasons has been observed in Asia during the last century. A declining trend in average seasonal rainfall in addition to frequent rainfall deficit monsoons along with significant inter decadal variability has been observed for the South Asian economies (IPCC 2014). Results from most of the Coupled Model Inter Comparison Project Phase 5 models predict an increase in heavy rainfall events in the future (IPCC 2013).

India’s climate system is dominated by the summer or south-west monsoon (and to a lesser extent by the winter or North-east monsoon). South West monsoon rainfall is a major source of precipitation for India and parts of South Asia (Ramanathan et al. 2005) and accounts for over 80 percent of India’s rainfall (Bagla 2006,2012). Owing to the monsoon’s pivotal role in the Indian economy in general, it is one of the most studied weather phenomena. Agricultural production depends crucially not only on the amount of rainfall, but its distribution across space and time (Goswami 2006).

Studies point to a gradual and accelerated weakening of the Indian monsoon, with increased frequency and intensity of rainfall extremes (Auffhammer et al. 2012). Pal and Al-Tabbaa (2009) find an increase in the frequency and magnitude of monsoon rainfall deficit along with reduced frequency and magnitude of monsoon rainfall excess for five regions of India during 1871-2005. For the period 1871-1920, deficit monsoon rainfall years exceed excess and normal rainfall years.

A declining trend in monsoon rainfall of 0.4 mm per year was observed in India during 1871-2009 (GoI 2010). Results from a simulation exercise from 1930-2000 point to a decreasing trend in south west monsoon rainfall since 1950 (Ramanathan et al. 2005). They find a 5 percent reduction in rainfall in addition to significant inter decadal variability of around 3 percent. Further, observed reductions are highest for Central India, which is crucial for growth of the three crops considered in this study. Drought frequency has increased considerably since the 1930s, particularly during the 1990s. Lal and Aggarwal (2000) and Lal (2003) use Atmospheric General Circulation Models and find an increasing likelihood of
more intense rainfall events in Central India in the future. Goswami et al. (2006), using daily gridded rainfall data by the Indian Meteorological Department, examine the frequency and trends in extreme rainfall events for Central India during 1951-2000. They find an increasing trend in standard deviation of the daily rainfall anomalies, which is attributed to higher moisture content in the atmosphere due to global warming. They find a significant increase in extreme rainfall events from 1951-70 to 1981-2000, along with a reduced frequency of light and moderate rain events. Further, precipitation intensity during the monsoon season is found to increase at the rate of 10 percent per decade.

Studies have found decreasing trends in early and late monsoon rainfall and number of rainy days for India from 1951-2003, followed by a significant expansion of total area with monsoon rainfall one standard deviation below the mean (Ramesh and Goswami 2007). Future projections for the 2030s point to a 3 – 7 percent increase for all regions except the southern peninsula. Projections of rainfall extremes also point to a reduction in the frequency of rainy days for all regions except the Himalayas, North West and Southern plateau. Intensity of rainy days is however, expected to increase for the Himalayan, North East, West, North West and Southern East coastal regions (GoI, 2010). Studies based on Greenhouse Gas forced model based scenarios of the IPCC also indicate intensification of rainfall in most parts of India except for central and north east India (May 2004). More recently, Krishnamurthy (2012) studied frequency and intensity of rainfall extremes in India using non parametric trend analysis and finds a rising trend in rainfall extremes in the coastal and eastern regions respectively. Both temperature and rainfall are correlated across space and time, with global average surface temperature increases leading to heavy rainfall events through its effect on the moisture content of the atmosphere (Goswami 2006).

Rising trends in annual mean temperature coupled with an increased heat wave frequency since the middle of the 20th century have been observed in many parts of Asia. Results of the Coupled Model Inter Comparison Project Phase 5 (CMIP5) simulations for the Representative Concentration Pathway (RCP) scenarios predict a warming of 2°C in the mid 21st century, with warming exceeding 3°C for South and South East Asia to more than 6°C for high latitude regions (IPCC 2014). Regarding temperature trends for India, mean annual temperature shows a significant warming trend of 0.5°C per 100 years during the period 1901-2007 (Kothawale et al. 2010). More importantly, accelerated warming has been observed in the last forty years (1971-2007), due to intense warming in the recent decade (1998-2007). Increases in mean temperature have been accompanied by a rise in both maximum and minimum temperatures -- by 0.71 and 0.27°C, respectively, per hundred years during the period 1901-2007. At a regional level, homogeneous regions of East coast, West coast and the peninsula show an increasing trend in the frequency of hot days but Northern India does not. On the contrary, all regions show a decreasing trend in the frequency of cold days (GoI 2010). Dash et al. (2007) examine regional and seasonal trends in rainfall and temperature for seven zones, namely, north west, western Himalaya, north Central, north east, interior peninsula, east coast and west coast from 1901-2003. They find a significant increase in maximum temperature across all regions of India, with the highest being 1.2°C for the West Coast and lowest being 0.5°C in the Interior Peninsula. Further, increase in minimum temperature is found to be the most for the Interior Peninsula, with the least being 0.2°C for

---

1Similar results are reported in Pattanaik (2007), who found decreasing trend in monsoon rainfall over northwest and central India for the period 1941-2002.
the West Coast. Post monsoon (October to December) and monsoon (June to September) months experience significant warming.

Future projections of climate reveal that for India as a whole, there will be an increase in average surface temperature by 2 - 4°C, changes in the distribution of rainfall (inter-temporal and spatial) during both monsoon and non monsoon months, decrease in the number of rainy days by more than 15 days, an increase in the intensity of rainfall by 1 to 4 mm/day along with higher frequency and intensity of cyclonic storms (Ranuzzi and Srivastava 2012). Climate projections for the 2030’s derived from the Regional Climate Model PRECIS point to an increase in all India summer monsoon rainfall by 3 to 7 percent along with higher annual mean surface air temperature by 1.7 to 2 degree Celsius from the 1970s. Medium run projections indicate a warmer and wetter climate for India, with significant regional variation (GoI 2010). Frequency of occurrence of hot days and hot nights shows a widespread increasing trend, whereas that of cold days and cold nights shows decreasing trend during 1970-2005 (Bhutiyani et al. 2007).

2.2 Growing pattern and climatic requirements of rice and millets

Rice is cultivated throughout the country. Its production is highly sensitive to natural calamities. The Green Revolution has led to a significant improvement in rice productivity in Asia through a combination of new high-yielding varieties with increased input use, such as stable water supply from new irrigation systems, fertilizer, and biocide use (Hossain and Fischer, 1995). HYV seeds were adopted in many of the rainfed ecosystems of India as well. Expansion of tubewell irrigation was responsible for higher production in eastern Uttar Pradesh and West Bengal (Barah 2005).

Rice is a water intensive crop and is grown under rain fed conditions and/or irrigation. Cumulative rainfall required for growth is 1200-1300 mm. Drought stress is the largest constraint to rice production affecting 10 million hectares of upland rice and over 13 million hectares of rainfed lowland rice in Asia alone (Pandey et al. 2007). The 2002 drought, affected 55 percent of the land area and around 300 million people, lead to a 20 percent decline in rice production from the inter annual baseline trend (Pandey et al. 2007). Further, the likelihood of a drought in consequent years affects farmer’s investment decisions, which affects future agricultural productivity.

Depending on the depth of standing water in the fields, rainfed rice area has been further classified into upland and lowland rice area. Upland rice area accounts for 14 percent of the all India rice area and includes the states of Assam, Bihar, (east) Madhya Pradesh, Orissa, (east) Uttar Pradesh, West Bengal and hilly region of the North East. However, productivity of upland rice is rather low, with an average yield of 0.90 tonnes per hectare, which is less than the all India rice productivity of 1.90 tonnes per hectare. Upland rice production is characterised by no standing water in the fields after rain. Lowland rice area in India is around 14 million hectares, and accounts for 32 percent of the total rice area. Low land rice production is however, characterised poor soil quality and experiences drought/flood conditions resulting in higher variability in production. Low land rice area is further divided into shallow water, semi deep water and deep water rice area, where area with depth of (standing) water in the fields of less than 50 cm is considered shallow, with 50-100 cm semi deep and exceeding 100 cm is considered as deep water. However, the crop is also grown under irrigation in the states of Punjab, Haryana, Uttar Pradesh, Jammu and Kashmir, Andhra
Pradesh, Tamil Nadu, Sikkim, Karnataka, Himachal Pradesh and Gujarat (GoI 2014).

In India, rice is grown in three seasons, namely, autumn (pre-\textit{kharif}), winter (\textit{kharif}) and summer (\textit{rabi}), where these seasons have been named according to the season of harvest. Winter or \textit{kharif} rice (sown during June-July and harvested in November-December) is the main growing season and accounts for 84 percent of total rice production currently. Medium and long duration varieties of rice are grown during this season (GoI 2014). This is followed by summer rice (sown during November-February and harvested in March-June) at 9 percent and autumn rice (sown during May-August and harvested in September-October) which accounts for 7 percent of rice production. Early maturing varieties are grown during summer season, where as short duration varieties are mainly grown during autumn (GoI 2014). The sowing and harvesting period however, varies across states, depending on rainfall patterns, temperature, land quality, and availability of other agricultural inputs. However, it is grown throughout the year in southern and eastern states such as Andhra Pradesh, Karnataka, Kerala, Orissa and Assam (ICAR 2008). Further, the crop is grown mainly under irrigation in the states of Punjab, Haryana, Uttar Pradesh, Jammu and Kashmir, Andhra Pradesh, Tamil Nadu, Sikkim, Karnataka, Himachal Pradesh and Gujarat (GoI, 2014). Punjab and Haryana account for close to 11 percent of the all India rice production, but have emerged as major producers in recent years, owing to extensive use of HYV seeds coupled with use of irrigation water to meet crop input requirements.

June-September comprises the vegetative and reproductive phase of rice growth where as October-November is the ripening phase (Kavi Kumar et al. 2014). Among the three growth phases, namely, vegetative, reproductive and ripening phase, water stress during the reproductive stage affects rice production the most, with effects as severe as premature abortion of the seed in addition to depressing grain formation (Boonjung and Fukai 1996; Saini 1997; Saini and Westgate 2000). Rice is a thermal sensitive crop, with temperatures suitable for growth varying by growth phase of the crop. During June-September day temperatures upto 35°C and night temperatures upto 25°C are suitable for rice growth (Wassman et al. 2009). High temperature affects cellular and developmental processes leading to reduced fertility and grain quality (Barnabas et al. 2008). Decreased grain weight, reduced grain filling, higher percentage of white chalky rice and milky white rice are common effects of high temperature exposure during the ripening phase (Osada et al. 1973; Yoshida et al. 1981). Higher temperature causes serious reductions in grain size and amylase content (Yamakawa et al. 2007; Zhu et al. 2005). Studies based on General Circulation Models and SRES (Special Report on Emission scenarios) scenarios show that higher temperatures reduce rice yields significantly (IPCC 2014).

In addition to weather inputs, rice production is input intensive and requires organic and inorganic fertilizers for growth. Upland rice cultivation requires sufficient organic manure which increases water holding capacity of the soil. However, the amount of fertilizer used depends on soil fertility and existing climatic conditions (GoI, 2014).

Millets and sorghum are a group of grasses, which produce small, seeded cereal crops, grown in arid and semi arid regions of India (Pray and Nagarajan 2009). Eight types of millets are grown in India, namely, great millet (sorghum), spiked millet (pearl millet), finger millet, foxtail millet, little millet, kodo millet, proso millet and barnyard millet respectively (MNI, 

---

2This discussion is based on Singh (2009).
Millets considered in this study, sorghum and pearl millet, account for around 5 percent (each) of the total cropped area (ICAR 2006, Pray and Nagarajan 2009). Despite a significant reduction in area sown under millets since the Green Revolution, an increase in yields has been observed since the mid 1970s (Pray and Nagarajan 2009). Sorghum and pearl millet, are mostly grown in states with relatively higher temperature and lower rainfall (see Appendix 1 and district-level maps in Kurosaki and Wada (2015)).

Millets grow under varied climatic condition and have limited input requirements. They are grown under limited irrigation, with millets and sorghum accounting for less than 9 percent of the total irrigated area in India\(^3\) (MoA 2006; Pray and Nagarajan 2009). They are usually grown in dryland areas with soils less than 15 cm deep, using farmyard manures and household produced biofertilisers (MNI, 2009). Short duration varieties are grown in India with a duration of 65 days (Swaminaidu et al. 2015). They are mainly grown by small and marginal farmers (i.e. those with farm size less than 5 hectares) (Swaminaidu et al. 2015).

Among millets, pearl millet is most widely grown followed by sorghum. Because of its tolerance to difficult growing conditions such as drought, low soil fertility and high temperature, it can be grown in areas where other cereal crops, such as maize or wheat would not survive (Basavaraj et al. 2010). Water needed for growth is 350-400 mm (MNI, 2009). It is grown in India as a single season (kharif) crop. Pearl millet production is concentrated in the states of Rajasthan and Gujarat.

Sorghum is grown in two seasons, namely winter (kharif) season as a rain fed crop and summer (rabi) season under residual soil moisture i.e. limited irrigation conditions. Water needed for production ranges from 400-500 mm (MNI, 2009) Kharif sorghum accounts for 48 percent of total sorghum production, whereas the corresponding figure for rabi sorghum is 52 percent (GoI, 2014). Sorghum cultivated area has declined by 41.81 per cent from 2008-09 to 2014-15. However, despite significant reduction in area, all states have experienced a significant increase in sorghum productivity where the extent of increase is highest for Madhya Pradesh (14.95 q/ha) (GoI 2015). Major crop producers include Maharashtra, Andhra Pradesh, Karnataka and Tamil Nadu.

3. Framework and Relevant Literature

Our methodology is based on estimating an agricultural production function with exogenous climate variables, namely, precipitation and temperature. Our analysis is at the district level using a panel dataset for physical yield (output divided by area under crop) for rice, sorghum and pearl millet.

Several studies have analysed the economic impact of climate related variables on crop yields for India. Lahiri and Roy (1985) studied acreage and yield response to price. They postulated a gamma distribution for the effect of rainfall on yield (right skewed and bounded at zero), i.e., less rainfall (droughts) is worse than excess rainfall (floods). They also argue that with the advent of Green Revolution in mid 1960s, water requirement of the crops has increased

---

\(^3\)“Millet cultivation can save farmers fighting drought” (3 January 2016, Times of India)
making Indian agriculture more rainfall dependent. Further, despite significant irrigation expansion, bulk of the water requirement of the crops is fulfilled through natural downpour.

Kanwar (2006) extends this analysis to several food grains. He looks at supply response using a state level panel data set and finds that rainfall is a crucial input determining supply response (it is a supply shifter). Auffhammer et al. (2012) extend Auffhammer et al. (2006) find significant adverse effect of rainfall extremes on mean rice yields in India. They allow for asymmetric effect of rainfall deficiency and rainfall excess, by creating separate variables for both. Deficit rainfall is captured by a drought indicator, where years with south west monsoon rainfall at least 15 percent below the state average are considered to be drought years, rainfall is considered extreme if it equals or exceeds the state’s 95th percentile threshold. However, an indicator variable does not capture the extent of rainfall deficiency, which is particularly important for rice, since it is a water intensive crop. Another problem with state or national level studies is the need to collapse rainfall and other weather data over large geographies to a single value which ignores inherent heterogeneity across regions. Further, this may result in measurement error which may bias coefficients of weather variables downward (Auffhammer et al. 2012).

While none of the studies above control for the effect of climate variables there are several international supply response studies that do so (see Miao, Khanna and Huang (2016) for a discussion and critique which itself finds significant negative impacts of climate change on county level production of corn and soyabean in the United States).

District level panel datasets for India have been used in several studies like Dinar et al. (1998), Kumar and Parikh (2001), Sanghi and Mendelsohn (2008), Kumar (2009), Guiteras (2009) and recently by Krishnamurthy (2012) and Gupta et al. (2014). The first four use the Ricardian approach which estimates the impact of climate variables on net agricultural revenues per unit area. Coefficient estimates obtained from these studies are however, not reliable as they suffer from omitted variable bias, which the panel data analysis conducted in this paper accounts for. Cross sectional regressions of this type ignore significant information in the data by averaging both, the dependent and independent variables. Deschenes and Greenstone (2007) use a variant of the Ricardian approach and estimate the impact of variation in climate variables on US annual agricultural profits and find that climate change will increase annual profits by 4 percent. Among the more recent studies Guiteras (2009) examines the impact of temperature and rainfall on combined yield of six major crops in India which account for 75 percent of total revenue. The precipitation variable has been defined both as total monthly rainfall (for growing seasons months i.e. June – September) as well as total growing season rainfall. He finds that climate change could reduce yields by 4.5 percent to 9 percent in the medium run (2010-2039) and by as much as 25 percent in the long run (2070-2099) in the absence of long run adaptation. The main drawback of this study as highlighted by Krishnamurthy (2012) is of combining crops which differ significantly from each other (in terms of input requirements, growing season etc) to arrive at a monetary measure with ill defined prices. Fishman (2012) use district level data for India from 1970-2004 to estimate climate change impacts, *inter alia* on rice and wheat yields. Rainfall

---

4 “In other words, rainfall is the single most important factor determining supply response even today. Despite decades of massive irrigation schemes, the food crops continue to be rainfall-dependent.” (Basavaraj et al. 2010, p.80).

5 The six crops are rice, wheat, jowar, bajra, maize and sugarcane.
distribution during the monsoon months is captured by the frequency of rainy days (i.e. days with precipitation more than 0.1 mm, May 2004), intensity of rainfall, duration of the longest dry spell (Tebaldi et al. 2006) and shape parameter of the gamma distribution, which is a measure of skewness of the rainfall distribution and finds that expansion of irrigation can help mitigate adverse effects of higher rainfall variability. Kumar (2014) estimate the effect of climate variables on average rice yields in India, and find higher maximum temperatures to be reducing rice yields. Using a state level panel data set for India, Kumar et al (2014) finds significant adverse effects of maximum temperature, minimum temperature and changing rainfall pattern on rice yield in India.

However, few studies have estimated climate change impacts on yield variance in developing economies. Cabas et al. (2010) examine the effect of climate and non climate factors on mean and variance of corn, soybean and winter wheat yield in Canada from 1981-2006. Variation in climate is captured by the coefficient of variation (CV) for temperature and rainfall, where CV captures the effects of extreme events. However, their analysis is based on a limited number of counties. Higher temperature and precipitation variability is found to reduce average crop yields and increase crop yield variability.

Mc Carl et al. (2008) on the basis of their study for corn, soybean and sorghum yields for the United States conclude that precipitation intensity and extent of rainfall deficiency are important determinants of crop yields. Their analysis is based on a state level panel data set for US states from 1960-2007. Apart from annual precipitation, precipitation intensity and Palmer Drought Severity Index has been used to capture rainfall intensity and severity of a wet or dry spell respectively. They find that precipitation intensity and droughts are important predictors of crop yields than the total amount of precipitation. Temperature variability is captured by the standard deviation of temperature, which is found to reduce yields of all crops. Increased precipitation intensity increases sorghum yield variability.

Chen et al. (2004) use Maximum Likelihood Estimation (MLE) technique for the same crops for United States from 1973-1997 where in higher rainfall is found to increase sorghum yield variability. On the other hand, higher temperatures decrease sorghum yield variability. Their analysis, is however based on a limited set of observations (1200-1400) for each of the crops, Further, the MLE technique employed in their paper is suitable for small samples. They do not account for climate variability, and use total annual precipitation and average temperature as regressors.

Kumar et al (2015) estimate the impact of climate and non climate variables on mean yield and yield variability of sugarcane yields in India (as in our study) . Analogous to Cabas et al (2010) coefficient of variation was used to capture inter annual seasonal variation in maximum temperature, minimum temperature and rainfall, in addition to other non climate inputs. Higher temperature adversely affects crop yields, with both higher maximum temperature during summer and minimum temperature during winter reducing average yields. Higher variability in minimum temperature reduces average yields. Higher maximum temperature during winter and minimum temperature during summer increases yield variability. Higher rainfall during winter season increases yield variability. However, most of the climate variables do not explain yield variability significantly.

Boubacar (2012) examine the impact of climatic fluctuations on mean yields and yield variability of maize, millet and sorghum using data for 8 counties of Sahel from 1970-2000. Climate extremes are captured using Standardized Precipitation Index (SPI), which indicates
the severity of wet and dry spells, and degree days. However, increased degree days are found to be reducing sorghum and millet yield variability. In addition to degree days, higher SPI also increases yield variability. Increased precipitation intensity reduces maize yields; higher precipitation intensity and degree days reduce sorghum yields, on an average.

There have been limited number of studies estimating the stochastic production function using MLE technique. Isik and Devadoss (2007) use district level panel data on yields of four crops in Idaho, namely, wheat, barley, potato and sugar beet from 1939-2001, and find climate change to be reducing yield variance and covariance, apart from a modest effect on mean yields. Kelbore (2012) estimate the effects of seasonal (namely, belg and kiremt) rainfall on crop yield and variances of three cereal crops, namely, teff, maize and wheat from 1978-2009. They perform a regional analysis of the impacts and find that higher seasonal rainfall increases average yields and reduces yield variability. Kim and Pang (2009) also use MLE to estimate the relationship between climate variables and rice yield variability for eight regions of Korea from 1977-2008. Higher temperature and precipitation are found to be increasing rice yields and yield variability.

One such study for India, which uses the three step Feasible Generalized Least Squares Procedure (as in our study) to estimate the Stochastic Production Function, is by Barnwal and Kotani (2010), where the effect of temperature and precipitation on mean and variance of seasonal rice yields is studied for Andhra Pradesh for the time period 1969-2002. Standard deviation of climate variables (over months) are included as separate regressors to capture intra seasonal variance of the climate variables. Climate change impacts are allowed to vary across agro climatic zones by including interaction terms of climate variables with a dummy for that agro climatic zone as separate regressors. Higher intra seasonal variance of rainfall and temperature is found to reduce mean yields and increase yield variability across agro climatic zones. A similar study by Poudel and Kotani (2013), find a one to one relation between climate variability and yield variability of rice and wheat across agro climatic zones in Nepal. They also capture heterogeneity in climate impacts across altitudes by separate dummies for low, middle and high altitude regions. Rice is found to be more sensitive of the two crops.

Hasanthika et al (2013) use a district level panel of 6 major rice growing districts in Sri Lanka from 1980-2010, and find climate variables to augment production risk. Edeh et al (2011) use time series data on rice yield and climate variables for Ebonyi in Nigeria from 2000-06 and find that rainfall variations are a major cause of uncertainty in crop production in Africa. Higher growing season temperature increases rice yield variability; higher intensity reduces variability, where as higher relative humidity makes yields more variable.

Using a panel data set of 26 provinces in China from 1985-2007, Holst et al. (2011) investigate the impact of several climatic and non climatic inputs on mean aggregate output (which in turn comprises quantities of different varieties of rice, wheat, corn, sorghum, millet, tubers and beans) and output variability. Deviation in precipitation from the long period average increase risk for the southern provinces, in support of the fact that rainfall is a crucial input for rice production and deviations in rainfall affect the crop adversely. Sarker (2014) use both, three step FGLS procedure and MLE to analyse the impacts of rainfall and temperature on three major rice varieties, namely, Aus, Aman and Boro in Bangladesh. The impact of climate variables on average yields and variance vary in a crop specific fashion, and results are specific to the functional form used. Mean minimum temperature is found to be risk decreasing for Boro rice, where as it is risk decreasing for Aus and Aman rice, rainfall
is risk increasing for Aman rice where as it is risk reducing for Aus and Boro. For large sample sizes, FGLS is appropriate (Judge et al. 1988) and it corrects for heteroscedasticity and auto correlation, both of which are usually observed in panel data sets of a longer duration better than MLE.

However, one of the major limitations of these studies is that they do not account for agricultural inputs used in agricultural production. Cabas (2010) include input change as a regressor, which measures the aggregate inputs used. However, there is no mention of the exact set of inputs included in this variable. Another state level study by Auffhammer et al (2006) accounts for the use of irrigation, fertiliser and high yielding variety seeds. However, district level studies do not account for these inputs, which affect production in addition to weather. Our study looks at variability in climate while controlling for the major inputs used by the farmers, namely, irrigated water, HYV seeds and fertilizers.

4. Data and Methodology

4.1 Data

4.1.1 Agricultural Data

Data on agricultural variables for 311 districts of India from 1966-2011, and has been obtained from the ICRISAT (International Crops Research Institute for the Semi Arid Tropics) VDSA (Village Dynamics in South Asia) Apportioned Meso database. This is a district level database. Districts in this database are according to 1966 base, data on districts formed after 1966 is given ‘back’ to the parent districts i.e. apportioned, based on percentage area of parent district transferred to the new district. Out of the 311 districts, 249 districts were formed after 1966, for which data was apportioned. Hence, the final database comprises of data for districts according to 1966 boundaries.

The variables of interest in this database include area and production of rice, sorghum and pearl millet (measured in hectares and tons respectively), district-wise consumption of fertilizers (tons of nitrogen, phosphate and potash fertilizers used), total gross cropped area in each district (measured in hectares, and accounting for multiple cropping) area under HYV seeds for each crop (measured in hectares, again accounting for multiple cropping). Missing values of the variables were imputed by multiple imputation. Details of the imputation procedure are in Appendix 4.

In our study the dependent variable is yield (tons of output per hectare). Owing to non availability of crop specific data on fertilizers used, annual aggregate fertilizer consumption is weighted by proportion of gross cropped area devoted to each crop. The resulting variable is further divided by crop area (tons of fertilizers used per hectare). Crop specific irrigated area and area under HYV seeds (tons) are divided by area under the crop and thus measure the proportion of crop area irrigated and under HYV seeds, respectively.
4.1.2 Climate data

Information on temperature and rainfall was available from the Indian Meteorological Department (IMD). Our study makes use of gridded daily rainfall and temperature data sets recently developed by the IMD (Srivastava et al. 2009; Rajeevan 2005). Rainfall data of 0.25° x 0.25° latitude / longitude resolution from 1951-2013 and temperature data of 1° x 1° latitude / longitude resolution from 1901-2013 were used.

Numerous spatial interpolation techniques are used for obtaining district level estimates from gridded data. A weighted average of the data for grid points within 100 km from the district centroid is taken, using inverse square root of the distance as weight (Guiteras 2009, Krishnamurthy 2012, Barrios et al., 2010). However, reliability of estimation depends crucially on location of the district centroids, which could be closer to the edges or outside the district boundary for districts with irregular shapes, and thus are not a true representative of the district.

Alternatively, an area weighted average of grid level observations is taken (Fishman 2012; Jones 2012; Barrios et al., 2010), which is done for the present study, where grid points within the district boundary were weighted by proportion of the grid cell within the district boundary. For this purpose, actual district boundaries at the 0.25° x 0.25° resolution were obtained. A lower spatial resolution allows for greater accuracy in estimation. Further, out of 644 districts in India, average district area exceeds 17,000 km² for ten districts only (Census 2011), the above resolution seems appropriate. The procedure also corrects for intersecting grids at each district boundary and hence is an improvement over other methods commonly used.

The climate variables used for our study from this dataset are rainfall and temperature. We define the former as total annual rainfall and temperature as the annual average of daily average temperatures.

4.2 Methodology

The number of districts selected for each of the crops is 162 for rice, 119 for sorghum and 91 for pearl millet which account for around 95 percent of output (see list and maps in Appendix 1).

As previously mentioned, districts included in the ICRISAT database are those that existed as of 1966. However, climate dataset has been created taking into account district boundaries as of 2011, which are very different from those of 1966. Districts that comprise the panel sample have been selected on the basis of districts that existed in the ICRISAT database, and climate variables for these districts have been approximated from the district to which the largest area of the parent district was allocated⁶ (provided that it is more than 50 percent of total area of the parent district) (see Kumar and Somanathan 2009).

Past studies have focused on estimating climate change impacts on mean agricultural outcomes only (Guiteras 2009, Fishman 2012, Som et al. 2014, Gupta et al. 2014).

---

Guiteras (2009) uses a 40 year district level panel data set for India to estimate climate change impacts on crop yields of major crops, namely rice, wheat, jowar, bajra, maize and sugar, and find significant adverse impact of extreme heat on US crop yields. Increased precipitation increases crop yields, but does not mitigate the negative effect of temperature. Fishman (2012) use district level data for 580 districts of India from 1970-2004 to analyse climate change impacts on major food grains grown in India, focussing mainly on rice and wheat, and find significant negative effects for crop yield. They further find that irrigation expansion can mitigate some of these adverse impacts. Gupta et al. (2014) conduct similar analyses for rice, sorghum and pearl millet from 1966-2009 and find that higher temperatures reduce rice and pearl millet yields significantly. Increased precipitation is found to be beneficial for all crops. It also controls for other agricultural inputs used such as irrigation use and fertilisers. Som et al (2014) find higher temperatures reducing wheat yields with a 1°C increase in maximum temperature found to be reducing yields by 4 percent. Pattanayak and Kumar (2014) find significant adverse impact of both daytime and night time temperature on rice yields in India from 1969-2007. Further, higher daytime temperature affects the crop more than night time temperature. They also find that crop yield would have been much higher if pre 1960 climatic conditions would have prevailed.

The effects of increasing climate variability on Indian agriculture were first observed during the 1930s in the Central Great Plains (Worster, 1979; Rosenberg, 1980). Further Anderson and Hazell (1989) found that agricultural sector is more vulnerable to climatic fluctuations with the advent of the Green revolution in India in the 1960s. Several studies indicate a change in the mean rainfall and its higher order moments (Gordon et al. 1992, Whetton et al. 1993, Mearns et al. 1995). Katz and Brown (1992) have shown that changes in climate variability affect climatic extremes to a greater extent than mean changes in climate. One of the initial studies on crop yield variability is by Mearns (1992), who studied the effect of inter annual climate (i.e. temperature and precipitation) variability on simulated wheat yields by altering the monthly year to year variances of the climate variables. However, changes in inter annual variability were isolated from changes in daily variability. Crop yields were found to be significantly affected by changes in climate variability. Using Richardson’s stochastic weather generator, time series of the weather variables was generated and cumulative distribution functions for wheat yields were generated by altering climate variability. Increasing variability in daily climate was found to increase crop yield variability and reduce mean yields. In addition to studies based on biophysical simulation models, crop yield variability has been estimated as a shift in the distribution of crop yields due to changes in temperature and precipitation (Kaylen and Koroma 1991). There is growing evidence that climate change will change the mean and variance of crop yields (Mc Carl et al. 2008).

There are a handful of studies estimating climate change impacts on yield variability, with one such study for India (Barnwal and Kotani 2010) where in increased climate variability is found to augment rice yield variability. Results from studies for other countries also show similar results (Cabas et al. 2010, Mc Carl et al. 2008). Implicit in such an approach is the idea of climate change leading to a mean shift in agricultural outcomes, with no changes of the underlying relationship between agricultural outcomes and climate. This is problematic for several reasons; for instance, in much of the scientific literature on climate change, focus is on changes in variability (especially in the hydrologic cycle, which determines both long and short run availability of water supply, a critical ingredient in agriculture) as a result of an altered climate; while such changes are incorporated in a “mean effect” framework, they are
not restricted to it (Krishnamurthy 2012).

The basic specification of the stochastic production function is:

\[ y = f(X, \beta) + \mu = f(X, \beta) + h(X, \alpha)e \]  

(4)

where \( y \) is measure of output, \( X \) is the input vector, \( f(.) \) is the production function relating \( X \) to output with \( \beta \) being the vector of estimable parameters, \( h(X, \alpha) \) is the risk (variance) function, such that \( h^2 \) is the yield variance; \( e \) is random shock distributed with mean zero and unitary variance, \( \alpha \) is the vector of estimable parameters associated with the risk function (where \( \alpha > 0 \) implies that yield variance increases as \( X \) increases, and vice versa).

Just and Pope (1978) also show that the above specification satisfies the postulates regardless of specification of \( f(.) \).

Most empirical studies have used the method of Feasible Generalized Least Squares (FGLS). Alternatively, Maximum Likelihood Estimation (MLE) can be used. However, FGLS estimation is employed in most empirical studies, although MLE is more efficient and unbiased than FGLS for small samples (Saha et al. 1997). Given the large sample size here, FGLS was used, as described in Judge et al. (1988), to estimate a form of fixed effects panel model. The exact procedure is mentioned below (Just and Pope 1978, Cabas et al. 2010). First stage entails regressing \( y \) on \( f(X, \beta) \) which gives the least squares residuals, \( \hat{\mu} \) which (\( \hat{\mu} = y - f(X, \hat{\beta}) \)) is a consistent estimator of \( \mu \). The second stage uses least square residuals from the first stage to estimate marginal effects of explanatory variables on the variance of production (\( \alpha \)). In the second stage, \( \ln \hat{\mu}^2 \) is regressed on its asymptotic expectation \( h(X, \alpha) \) with \( h(.) \) assumed to be an exponential function (i.e. \( h = e^{X_i'\alpha} \)). The third and final stage uses predicted error terms from the second stage as weights for generating FGLS estimates for the mean yield equation. The resulting estimator of \( \beta \) in this final step is consistent and asymptotically efficient under a broad range of conditions and the whole procedure corrects for the heteroscedastic disturbance term (Just and Pope 1978).

5. Estimation

All variables were tested for non stationarity using panel unit root tests and were found to be stationary. We also performed Pesaran, Friedman and Frees tests and found cross sectional dependence. Finally, we tested for fixed versus random effects using the Hausman test and fixed effect model was found to be appropriate. One expects input usage to be correlated with district specific time invariant attributes, such as land quality, farming techniques and sowing patterns, which makes the fixed effects model suitable for a district level panel study like ours. In light of these results panel corrected standard error (PCSE) estimates were obtained which correct for cross sectional dependence, heteroscedasticity and autocorrelation. The parameters are estimated using a Prais Winsten (or OLS) regression. Equations have been estimated with district and year fixed effects.

7 Im Pesaran Shin and Fisher type tests (Augmented Dickey Fuller and Phillips Perron) were performed.
Two regressions were run for each crop, explaining mean yield and yield variability. Mean yield depends on climate and non-climate inputs, namely, fertiliser, irrigation and use of high yielding variety seeds, whereas yield variability depends on the transformed climate variables (called anomalies details of which are in Appendix 2). Further, to capture the effect of farming decisions at the extensive and intensive margin, gross cropped area under the crop has been included as a regressor. In Appendix 3 we also report results of alternate specifications where both mean yields and yield variability depend only on levels of climate variables or only on climate anomalies. In the appendix we also present results when mean yields and yield variability depend both on levels of climate variables and on climate anomalies. Our results, however, show mean yields are best explained by levels of rainfall and temperature whereas variability in yields is more a function of variability in climate (anomalies). We surmise therefore it is variability in climate that makes agriculture more risky.

To capture the asymmetry in yield response to climate extremes (Sivakumar 1987, Gadgil and Kumar 2006) for each of the climate variables, anomaly variables for both extremes have been included as regressors in the yield variability equation.

Regression equations estimated for the three crops are:

\[
\text{Mean Yield}_{it} = \beta_1 + \alpha_i + \delta_t + \beta_2 \text{Rainfall}_{it} + \beta_3 \text{Temperature}_{it} + \beta_4 \text{Area}_{it} + \\
\beta_5 \text{Fertiliser}_{it} + \beta_6 \text{Irrigation}_{it} + \beta_7 \text{HYV}_{it} + \nu_{it}
\]

\[
\text{Yield Variance}_{it} = \beta_1 + \alpha_i + \delta_t + \beta_2 \text{Drought Anomaly}_{it} + \beta_3 \text{Flood Anomaly}_{it} + \\
\beta_4 \text{Low Temp Anomaly}_{it} + \beta_5 \text{High Temp Anomaly}_{it} + \epsilon_{it}
\]

where \(i\) refers to the district and \(t\) refers to the year; \(\alpha_i\) denotes district level fixed effects; \(\delta_t\) denotes year fixed effects; \(\text{Area}_{it}\) denotes gross cropped area under the crop; \(\text{Irrigation}_{it}\) is the proportion of gross cropped area (under that crop) which is irrigated; \(\text{Fertiliser}_{it}\) is the total amount of fertilisers (nitrogen, phosphate and potash) used; \(\text{HYV}_{it}\) is the proportion of gross cropped area (under that crop) under HYV seeds; \(\text{Rainfall}_{it}\) is the annual rainfall; \(\text{Temperature}_{it}\) is the average temperature; \(\text{Drought Anomaly}_{it}, \text{Flood Anomaly}_{it}, \text{Low Temp Anomaly}_{it}\) and \(\text{High Temp Anomaly}_{it}\) are the climate anomaly variables capturing rainfall and temperature extremes respectively; \(\nu_{it}\) and \(\epsilon_{it}\) are stochastic error terms where \(\epsilon_{it} \sim N(0,1)\).

In our specifications we do not include inputs such as irrigation and fertilizer and HYV in the variance regression. Irrigation is likely to reduce production risk (Foudi and Erdlenbruch 2011, Espinoza 2012) though some argue otherwise (Guttormsen and Roll 2013). Fertilizer use typically increases production risk even as it increases expected output (Just and Pope 1979, Rosegrant and Roumasset 1985, Roumasset et al.1987, Ramaswami 1992, Di Falco, Chavas, and Smale 2007, Abdullah and Pandey 2004). Farmers in developing countries are not trained to use fertilisers. Fertiliser use may even poison crops (Feder 1980). Farnsworth and Moffitt, 1981 argue that fertiliser use helps crops survive even in adverse weather conditions. Use of high yielding variety seeds also increases production risk (Bakhshoodeh and Shajari 2006), since such seeds need more water than traditional seeds (Dalrymple 1979) making production sensitive to fluctuations in rainfall. However, since the two are correlated with each other and also with HYV use their interactive effect is unclear and we leave this for further research.
6. Results

Results using anomalies are reported in Table 1 below. Results for standardized anomalies are listed in Appendix 3. Coefficients of district and year fixed effects have been suppressed. For each crop, separate regressions were performed for mean yield and yield variance. Coefficients for mean yield are those obtained in third stage of the three step FGLS procedure, whereas second stage coefficients are the ones reported for yield variance. The explanatory variables for mean yield regression are: proportion of gross cropped area irrigated, proportion of gross cropped area under HYV seeds, crop specific fertilizers used, annual rainfall and annual average temperature, whereas climate extremes, captured using climate anomalies explain crop yield variability significantly. A positive (negative) coefficient in the mean yield regression can be interpreted as a marginal increase in input increasing (decreasing) crop yields, on an average. A positive (negative) coefficient in yield variance regression can be interpreted as higher deviation of rainfall and temperature from the long period average increasing (decreasing) crop yield variability.

6.1 Rice

Annual rainfall has been obtained by summation of the daily rainfall values, whereas for temperature average of the daily temperature values is taken.

An acreage expansion of 1000 hectares increases rice yields marginally, by less than 1 kg/ha, which is possibly because of diminishing returns from the marginal land devoted to rice production, owing to it being relatively inferior in quality. Coefficients of the irrigation, fertilizer and HYV variable are positive and highly significant, even at 1 percent level of significance. An extra tonne of fertiliser applied (per unit gross cropped area) increases yield, on an average by 4087 kg/ha. Similarly, irrigation expansion and use of HYV seeds increases yields, with a unit increase in rice irrigated area (as a proportion of gross cropped area) yielding an additional 434 kg/ha, whereas the corresponding estimate for HYV use is 138 kg/ha. Hence, rice is highly intensive in these inputs. Further, since rice is a water-intensive crop, higher rainfall is found to be increasing rice yields, however the coefficient is small, but highly significant with an additional 100 millimetres of rainfall incrementing yields by 8 kg/ha. Higher temperatures adversely affect crop yields with a 1ºC increase reducing yields by 80 kg/ha. Hence, temperature increase reduces average rice yield significantly.

For the yield variability regression, coefficients of both, drought and flood anomaly variables are positive and highly significant, hence rainfall variability, in particular rainfall extremes increase rice yield variability significantly. This is what one would expect, since rice is water intensive. The variance regression is log linear, with coefficients representing semi elasticity of crop yield variability. Annual rainfall of 100 mm below normal (or long period average) increases rice yield variability by 9 percent. Both rainfall deficit and rainfall excess increase crop yield variability, however insufficient rainfall increases variability to a greater extent. An annual rainfall of 100 mm more than the long period average increases yield variability by 5 percent. Coefficients of the temperature anomaly variables are opposite in sign, with coefficient of the high temperature anomaly being positive. An annual average temperature of 1ºC below normal reduces yield variability by 14 percent, whereas annual average temperatures of 1ºC above normal increases yield variability by 3 percent. Hence high temperature low rainfall conditions ameliorate rice yield variability, which is expected since
rice is a water intensive and thermal sensitive crop.

Table 1. Fixed effects results using climate anomalies

<table>
<thead>
<tr>
<th>Mean Yield</th>
<th>Rice</th>
<th>Sorghum</th>
<th>P Millet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>8.26E-05*** (1.20E-05)</td>
<td>-E-05 (1.41E-05)</td>
<td>3.27E-05 (2.08E-05)</td>
</tr>
<tr>
<td>Meantmp</td>
<td>-0.080*** (0.013)</td>
<td>-0.062*** (0.012)</td>
<td>-0.052*** (0.014)</td>
</tr>
<tr>
<td>Area</td>
<td>6.39E-04*** (9.10E-05)</td>
<td>2.80E-04*** (7.25E-05)</td>
<td>7.41E-04*** (9.09E-05)</td>
</tr>
<tr>
<td>Fertiliser</td>
<td>4.087*** (0.118)</td>
<td>0.737*** (0.105)</td>
<td>1.193*** (0.147)</td>
</tr>
<tr>
<td>Irrigation</td>
<td>0.434*** (0.022)</td>
<td>0.293*** (0.057)</td>
<td>0.560*** (0.044)</td>
</tr>
<tr>
<td>HYV</td>
<td>0.138*** (0.017)</td>
<td>0.104*** (0.013)</td>
<td>0.030* (0.016)</td>
</tr>
<tr>
<td>R²</td>
<td>0.87</td>
<td>0.70</td>
<td>0.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yield Variance</th>
<th>Rice</th>
<th>Sorghum</th>
<th>P Millet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drought Anomaly</td>
<td>9.39E-04*** (2.13E-04)</td>
<td>1.15E-03*** (3.25E-04)</td>
<td>4.85E-04 (4.03E-04)</td>
</tr>
<tr>
<td>Flood Anomaly</td>
<td>5.43E-04*** (1.76E-04)</td>
<td>3.92E-04* (2.34E-04)</td>
<td>7.23E-04*** (2.55E-04)</td>
</tr>
<tr>
<td>Low Temp Anomaly</td>
<td>-0.137 (0.221)</td>
<td>0.335 (0.250)</td>
<td>0.239 (0.268)</td>
</tr>
<tr>
<td>High Temp Anomaly</td>
<td>0.033 (0.211)</td>
<td>0.346 (0.228)</td>
<td>0.782*** (0.260)</td>
</tr>
<tr>
<td>R²</td>
<td>0.16</td>
<td>0.16</td>
<td>0.13</td>
</tr>
</tbody>
</table>

6.2 Sorghum

Acreage expansion increases mean yields, however, the coefficient is small. Increasing area under sorghum production by 1000 hectares increases average yields by less than 1 kg/ha. Coefficients of the irrigation, fertilizer and HYV variable are positive and highly significant, even at 1 percent level of significance. An extra tonne of fertiliser applied (per unit gross cropped area) increases yield, on an average by 737 kg/ha. Similarly, irrigation expansion and use of HYV seeds increases yields, with a unit increase in sorghum irrigated area (as a proportion of gross cropped area) yielding an additional 293 kg/ha, whereas the corresponding estimate for HYV use is 104 kg/ha. Coefficient of rainfall is small and insignificant. Higher rainfall reduces sorghum yields, with an additional 100 mm of rainfall
reducing yields by 1 kg/ha. However, coefficient of temperature is negative and highly significant, with a 1ºC temperature increase reducing sorghum yields, on an average by 62 kg/ha.

For the yield variability regression, coefficients of both, drought and flood anomaly variables are positive and significant. An annual rainfall of 100 mm below normal increases yield variability by around 12 percent. Rainfall exceeding the normal affects variability to a lesser extent, with a 100 mm deviation increasing yield variability by 4 percent. Coefficients of the temperature anomaly variables are positive but insignificant. Annual average temperatures of 1ºC below normal increases yield variability by 34 percent whereas temperatures of a similar magnitude above normal increases yield variability by 35 percent.

6.3 Pearl Millet

Increasing area under pearl millet production increases yields marginally with an additional 1000 hectares increasing yields by less than 1 kg/ha. This is possibly since the marginal land used in production is inferior in quality. Coefficients of the irrigation, fertilizer and HYV variable are positive and significant. An extra tonne of fertiliser applied (per unit gross cropped area) increases yield, on an average by 1193 kg/ha. Similarly, irrigation expansion and use of HYV seeds increases yields, with a unit increase in pearl millet irrigated area (as a proportion of gross cropped area) yielding an additional 560 kg/ha, whereas the corresponding estimate for HYV use is 30 kg/ha. Further, higher rainfall is found to be increasing pearl millet yields, with an additional 100 millimetres of rainfall incrementing yields by 3 kg/ha. Higher temperatures adversely affect crop yields with a 1ºC temperature increase reducing yields by 52 kg/ha.

For the yield variability regression, climate extremes increase pearl millet yield variability, however, coefficients of only the flood anomaly and high temperature anomaly variables are significant. An annual rainfall of 100 mm below normal increases yield variability by 5 percent. Rainfall of a similar magnitude above the long period average increases yield variability by 7 percent. A 1ºC positive deviation of annual average temperature from the long period average increases yield variability by 78 percent, which can have significant adverse consequences for the agriculture sector and Indian economy in general. On the other hand, temperatures below normal reduce yield variability, by 24 percent.

7. Concluding Remarks

Using a 46 year district level panel dataset for three major foodgrains in India we find that increased climate variability, climate extremes in particular, exacerbate risk. Indian agriculture is heavily monsoon dependent, with rainfall extremes increasing crop yield variability. Rice, being water intensive is most sensitive with both insufficient and excess rainfall increasing yield variability. Sorghum yields exhibit similar response. Pearl millet yields on the other hand, become more variable with rainfall and / or temperature exceeding the long period average. In addition to climate inputs, non climate inputs, namely, irrigation, fertilizer and HYV seeds are found to be increasing agricultural yields, on an average. Higher temperature adversely effects crop yields.

The analysis presented in this study has important policy implications. Production of rice, maize, and wheat has declined in many parts of Asia in the past few decades, due to
increasing water stress, arising partly from increasing temperatures, increasing frequency of El Nino events and reduction in the number of rainy days (Aggarwal et al. 2000, Fischer et al. 2002, Tao et al. 2004). In turn, this will increase vulnerability of poor rural farmers, especially in the arid and semiarid tropics in addition to implications for food security (Bates et al. 2008).

Given that agriculture is a major source of livelihood for farmers in India, increasing yield variability will have serious consequences for our economy. Hence, it is imperative to undertake suitable policies to mitigate climate change impacts on this sector to the extent possible.

As the econometric results could be subject to omitted variable bias, we leave it for further research to employ a richer set of mean and variability shifters than employed in this paper.
### Appendix 1. Data

#### Table A1. Summary Statistics

<table>
<thead>
<tr>
<th>Variable/Crop</th>
<th>Unit</th>
<th>Rice (162)#</th>
<th>Pearl Millet (Bajra) (119)#</th>
<th>Sorghum (Jowar) (91)#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Min</td>
</tr>
<tr>
<td>Rainfall</td>
<td>mm</td>
<td>1270</td>
<td>707</td>
<td>117</td>
</tr>
<tr>
<td>Temperature</td>
<td>° C</td>
<td>25.66</td>
<td>1.46</td>
<td>21.18</td>
</tr>
<tr>
<td>Area</td>
<td>000 hectares</td>
<td>223.51</td>
<td>166.69</td>
<td>0.00</td>
</tr>
<tr>
<td>Production</td>
<td>000 tons</td>
<td>372.89</td>
<td>333.96</td>
<td>0.00</td>
</tr>
<tr>
<td>Fertiliser Use</td>
<td>tons/hectare</td>
<td>0.08</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>Irrigation</td>
<td>proportion</td>
<td>0.55</td>
<td>0.38</td>
<td>0.00</td>
</tr>
<tr>
<td>HYV</td>
<td>proportion</td>
<td>0.46</td>
<td>0.35</td>
<td>0.00</td>
</tr>
<tr>
<td>Yield (Production/area)</td>
<td>tons/hectare</td>
<td>1.78</td>
<td>0.92</td>
<td>0.00</td>
</tr>
</tbody>
</table>

#Brackets denote number of districts.
Figure A1. Crop Area (‘000 hectare)

Figure A2. Crop Production (‘000 tons)
Figure A3. Crop Yield (tons/hectare)

Figure A4. Irrigated area as a proportion of gross cropped area
Figure A5. HYV area as a proportion of gross cropped area

Figure A6. Fertilizer Use (tons/hectare)
Table A2. Districts in the study

**Rice (162 districts)**

- **Andhra Pradesh**: Adilabad, Anantapur, Chittoor, Cuddapah, East Godavari, Guntur, Karimnagar, Khammam, Krishna, Kurnool, Mahabubnagar, Medak, Nalgonda, Nellore, Nizamabad, Srikakulam, Visakhapatnam, Warangal, West Godavari
- **Assam**: Cachar, Darrang, Dibrugarh, Goalpara, Kamrup, KarbiAnglong, Lakhimpur, Nowgong, Sibsagar
- **Bihar**: Bhagalpur, Champaran, Darbhanga, Gaya, Hazaribagh, Monghyr, Muzaffarpur, Patna, Purnea, Ranchi, Saharsa, SanthalParagana, Saran, Shahabad, Singhbhum
- **Gujarat**: Ahmedabad, Bulsar, Kaira, Surat
- **Haryana**: Ambala, Hissar, Jind, Karnal, Rohtak
- **Karnataka**: Bellary, Chickmagalur, Chitradurga, DakshinaKannara, Dharwad, Hassan, Mandya, Mysore, Raichur, Shimoga
- **Kerala**: Alapuzha, Eranakulam, Palakkad, Thrisur
- **Madhya Pradesh**: Balaghat, Bastar, Bilaspur, Durg, Raigarh, Raipur, Seoni, Shahdol, Surguja
- **Maharashtra**: Bhandara, Chandrapur, Kolhapur, Raigad, Ratnagiri, Thane
- **Orissa**: Balasore, Bolangir, Cuttack, Dhenkanal, Ganjam, Kalahandi, Keonjhar, Koraput, Mayurbhanj, Phulbani, Puri, Sambalpur, Sundergarh
- **Punjab**: Amritsar, Bhatinda, Ferozpur, Gurdaspur, Hoshiarpur, Jalandhar, Kapurthala, Ludhiana, Patiala, Roopnagar, Sangrur
- **Tamil Nadu**: Chengalpattu, Coimbatore, Madurai, North Arcot, Ramanthapuram, Salem, South Arcot, Thanjavur, Tiruchirapalli, Tirunelveli
- **Uttar Pradesh**: Allahabad, Azamgarh, Bhaich, Ballia, Barabanki, Bareilly, Basti, Bijnor, Deoria, Etawah, Faizabad, Fatehpur, Ghazipur, Gonda, Gorakhpur, Hardoi, Jaunpur, Kanpur, Kheri, Mainpuri, Mirzapur, Moradabad, Nainital, Pilibhit, Pratapgarh, Rae Bareily, Rampur, Saharanpur, Shahjahanpur, Sitapur, Sultanpur, Unnao, Varanasi
- **West Bengal**: Bankura, Birbhum, Burdwan, Hooghly, Howrah, Jalpaiguri, Malda, Midnapore, Murshidabad, Nadia, Purulia, Twenty Four Paragana, West Dinajpur
Sorghum (119 districts)

**Andhra Pradesh**: Adilabad, Anantapur, Cuddapah, Guntur, Hyderabad, Karimnagar, Khammam, Krishna, Kurnool, Mahabubnagar, Medak, Nalgonda, Nellore, Nizamabad, Warangal

**Gujarat**: Ahmedabad, Amreli, Banaskantha, Baroda, Broach, Junagadh, Mehsana, Rajkot, Surat, Surendranagar

**Haryana**: Rohtak

**Karnataka**: Belgaum, Bellary, Bijapur, Chickmagalur, Chitradurga, Dharwad, Gulbarga, Hassan, Mysore, Raichur, Shimoga, Tumkur

**Madhya Pradesh**: Betul, Bhind, Chhatarpur, Chhindwara, Damoh, Dewas, Dhar, Guna, Gwalior, Hoshangabad, Indore, Jabalpur, Jhabua, Khajuraho, Mandsaur, Morena, Narsinghpur, Rajgarh, Ratlam, Rewa, Sagar, Sehore, Shajapur, Shivpuri, Sidhi, Tikamgarh, Ujjain, Vidisha

**Maharashtra**: Ahmednagar, Akola, Amravati, Aurangabad, Beed, Buldhana, Chandrapur, Dhulia, Jalgaon, Kolhapur, Nagpur, Nanded, Nasik, Osmanabad, Parbhani, Pune, Sangli, Satara, Solapur, Wardha, Yeotmal

**Rajasthan**: Ajmer, Alwar, Bharatpur, Bhilwara, Bundi, Chittorgarh, Jaipur, Jhalawar, Kota, Nagaur, Pali, Sawai Madhopur, Tonk

**Tamil Nadu**: Coimbatore, Madurai, North Arcot, Ramanthapuram, Salem, South Arcot, Tiruchirapalli, Tirunelveli

**Uttar Pradesh**: Allahabad, Banda, Fatehpur, Hamirpur, Hardoi, Jalaun, Jhansi, Kanpur, Rae Bareily, Unnao

Pearl Millet (91 districts)

**Andhra Pradesh**: Anantapur, Chittoor, Cuddapah, Guntur, Kurnool, Mahabubnagar, Nalgonda, Nellore, Visakhapatnam

**Gujarat**: Ahmedabad, Amreli, Banaskantha, Baroda, Junagadh, Kaira, Kutch, Mehsana, Panchmahals, Rajkot, Sabarkantha, Surendranagar

**Haryana**: Gurgaon, Hisar, Jind, Karnal, Mahendragarh, Rohtak

**Karnataka**: Belgaum, Bellary, Bijapur, Gulbarga, Raichur

**Madhya Pradesh**: Bhind, Morena

**Maharashtra**: Ahmednagar, Aurangabad, Beed, Dhulia, Jalgaon, Nasik, Osmanabad, Pune, Sangli, Satara, Solapur

**Punjab**: Bhatinda, Ferozpur, Sangrur

**Rajasthan**: Ajmer, Alwar, Barmer, Bharatpur, Bikaner, Churu, Ganganagar, Jaipur, Jaisalmer, Jalore, Jhunjhunu, Jodhpur, Nagaur, Pali, Sawai Madhopur, Sikar, Tonk

**Tamil Nadu**: South Arcot, North Arcot, Salem, Coimbatore, Tiruchirapalli, Madurai, Ramanthapuram, Tirunelveli

**Uttar Pradesh**: Agra, Aligarh, Allahabad, Bareilly, Buduan, Bulandshahr, Etah, Etawah, Farrukhabad, Ghazipur, Jalaun, Kanpur, Mainpuri, Mathura, Mirzapur, Moradabad, Pratapgarh, Varanasi
Figure A8. Sorghum
Figure A9. Pearl Millet
Appendix 2. Variable transformation to capture climate extremes

Climate anomaly refers to deviation of a climate variable $x_{it}$ (e.g., annual rainfall), from its long period average (LPA). The anomalies were standardized as well. For India, asymmetric response of crop yields to rainfall and temperature extremes is well known. The impact of rainfall deficit is negative and large, whereas that of rainfall surplus is favourable but small (Kumar 2006). To incorporate this, four anomaly variables were defined capturing climate extremes.

If $x_{it}$ is the annual rainfall in district $i$ in year $t$, $\bar{x}_i$ and $\sigma_i$ its mean and standard deviation then rainfall anomaly ($RA_{it}$) is:

$$RA_{it} = x_{it} - \bar{x}_i$$

and standardized rainfall anomaly ($SRA_{it}$) as,

$$SRA_{it} = \frac{x_{it} - \bar{x}_i}{\sigma_i}$$

Analogously, temperature anomaly ($TA_{it}$) is defined as,

$$TA_{it} = x_{it} - \bar{x}_i$$

and standardized temperature anomaly ($STA_{it}$) as,

$$STA_{it} = \frac{x_{it} - \bar{x}_i}{\sigma_i}$$

Deviations of annual rainfall from the Long Period Average are normal if $x_{it} \in [\bar{x}_i \pm 0.04 \times \bar{x}_i]$.

Anomalies capturing rainfall extremes have been defined as

$$Drought\ Aomaly_{it} = RA_{it} \text{ if annual rainfall } \leq 0.96 \times \bar{x}_i$$

$$= 0, \text{ otherwise}$$

$$Flood\ Aomaly_{it} = RA_{it} \text{ if annual rainfall } \leq 1.04 \times \bar{x}_i$$

$$= 0, \text{ otherwise}$$

Deviations of annual average temperature from the Long Period Average exceeding 0.10 degree Celsius represent extreme temperature conditions. Anomaly variables capturing temperature extremes have been defined as

$$Low\ Temp\ Aomaly_{it} = TA_{it} \text{ if } TA_{it} \leq -0.10$$

$$= 0, \text{ otherwise}$$

$$High\ Temp\ Aomaly_{it} = TA_{it} \text{ if } TA_{it} \geq 0.10$$

$$= 0, \text{ otherwise}$$
Figure A10. Rainfall Anomaly (mm)

(a) Rice

(b) Sorghum
Figure A11. Temperature Anomaly (°C)

(c) Pearl Millet

(a) Rice
(b) Sorghum

(c) Pearl Millet
## Appendix 3. Supplementary Results

Table A3. Baseline results using standardised climate anomalies

<table>
<thead>
<tr>
<th>Mean Yield</th>
<th>Rice</th>
<th>Sorghum</th>
<th>P Millet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>7.96E-05***</td>
<td>-1.45E-05</td>
<td>1.44E-05</td>
</tr>
<tr>
<td></td>
<td>(1.17E-05)</td>
<td>(1.31E-05)</td>
<td>(1.84E-05)</td>
</tr>
<tr>
<td>Temperature</td>
<td>-0.079***</td>
<td>-0.062***</td>
<td>-0.056***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.012)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Area</td>
<td>0.001***</td>
<td>2.78E-04***</td>
<td>0.001***</td>
</tr>
<tr>
<td></td>
<td>(9.09E-05)</td>
<td>(7.24E-05)</td>
<td>(9.05E-05)</td>
</tr>
<tr>
<td>Fertiliser</td>
<td>4.092***</td>
<td>0.742***</td>
<td>1.193***</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.105)</td>
<td>(0.146)</td>
</tr>
<tr>
<td>Irrigation</td>
<td>0.434***</td>
<td>0.292***</td>
<td>0.557***</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.058)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>HYV</td>
<td>0.139***</td>
<td>0.104***</td>
<td>0.030*</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.013)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.82</td>
<td>0.61</td>
<td>0.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yield Variance</th>
<th>Rice</th>
<th>Sorghum</th>
<th>P Millet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std Drought Anomaly</td>
<td>0.141***</td>
<td>0.116***</td>
<td>0.080*</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.037)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>Std Flood Anomaly</td>
<td>0.093***</td>
<td>0.044</td>
<td>0.091*</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.041)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Std Low Temp Anomaly</td>
<td>-0.060</td>
<td>0.129</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td>(0.114)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>Std High Temp Anomaly</td>
<td>0.003</td>
<td>0.119</td>
<td>0.337***</td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.105)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.16</td>
<td>0.16</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Table A4. Levels of climate variables in both mean and variance

<table>
<thead>
<tr>
<th>Mean Yield</th>
<th>Rice</th>
<th>Sorghum</th>
<th>P Millet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>7.68E-05*** (1.09E-05)</td>
<td>-2.04E-05* (1.21E-05)</td>
<td>2.14E-05 (1.93E-05)</td>
</tr>
<tr>
<td>Temperature</td>
<td>-0.081*** (0.013)</td>
<td>-0.070*** (0.012)</td>
<td>-0.060*** (0.014)</td>
</tr>
<tr>
<td>Area</td>
<td>0.001*** (9.15E-05)</td>
<td>2.97E-04*** (7.21E-05)</td>
<td>0.001*** (8.93E-05)</td>
</tr>
<tr>
<td>Fertiliser</td>
<td>4.079*** (0.119)</td>
<td>0.756*** (0.105)</td>
<td>1.247*** (0.147)</td>
</tr>
<tr>
<td>Irrigation</td>
<td>0.441*** (0.022)</td>
<td>0.287*** (0.058)</td>
<td>0.557*** (0.044)</td>
</tr>
<tr>
<td>HYV</td>
<td>0.135*** (0.017)</td>
<td>0.103*** (0.013)</td>
<td>0.029* (0.016)</td>
</tr>
<tr>
<td>R²</td>
<td>0.81</td>
<td>0.60</td>
<td>0.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yield Variance</th>
<th>Rice</th>
<th>Sorghum</th>
<th>P Millet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>-9.49E-05 (1.11E-04)</td>
<td>-1.56E-04 (1.65E-04)</td>
<td>2.88E-04 (1.94E-04)</td>
</tr>
<tr>
<td>Temperature</td>
<td>0.125 (0.139)</td>
<td>0.091 (0.156)</td>
<td>0.329** (0.170)</td>
</tr>
<tr>
<td>R²</td>
<td>0.15</td>
<td>0.16</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Table A5. Climate Anomalies in both mean and variance

<table>
<thead>
<tr>
<th>Mean Yield</th>
<th>Rice</th>
<th>Sorghum</th>
<th>PMillet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drought Anomaly</td>
<td>-3.04E-04*** (2.45E-05)</td>
<td>-3.86E-04*** (2.77E-05)</td>
<td>-4.75E-04*** (3.92E-05)</td>
</tr>
<tr>
<td>Flood Anomaly</td>
<td>-6.84E-05*** (1.80E-05)</td>
<td>-2.21E-04*** (1.94E-05)</td>
<td>-1.92E-04*** (2.60E-05)</td>
</tr>
<tr>
<td>Low Temp Anomaly</td>
<td>0.040** (0.019)</td>
<td>-0.005 (0.018)</td>
<td>-0.045** (0.021)</td>
</tr>
<tr>
<td>High Temp Anomaly</td>
<td>-0.113*** (0.021)</td>
<td>-0.101*** (0.018)</td>
<td>-0.082*** (0.023)</td>
</tr>
<tr>
<td>Area</td>
<td>0.001*** (9.04E-05)</td>
<td>2.71E-04*** (7.08E-05)</td>
<td>0.001*** (9.36E-05)</td>
</tr>
<tr>
<td>Fertiliser</td>
<td>4.018*** (0.117)</td>
<td>0.705*** (0.099)</td>
<td>1.251*** (0.148)</td>
</tr>
<tr>
<td>Irrigation</td>
<td>0.420*** (0.022)</td>
<td>0.311*** (0.057)</td>
<td>0.519*** (0.043)</td>
</tr>
<tr>
<td>HYV</td>
<td>0.160*** (0.017)</td>
<td>0.118*** (0.013)</td>
<td>0.034** (0.016)</td>
</tr>
<tr>
<td>R²</td>
<td>0.87</td>
<td>0.70</td>
<td>0.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yield Variance</th>
<th>Rice</th>
<th>Sorghum</th>
<th>PMillet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drought Anomaly</td>
<td>0.001*** (1.95E-04)</td>
<td>4.17E-04 (3.12E-04)</td>
<td>4.69E-04 (4.16E-04)</td>
</tr>
<tr>
<td>Flood Anomaly</td>
<td>3.48E-04** (1.67E-04)</td>
<td>1.71E-04 (2.14E-04)</td>
<td>2.08E-04 (2.86E-04)</td>
</tr>
<tr>
<td>Low Temp Anomaly</td>
<td>0.051 (0.220)</td>
<td>0.177 (0.250)</td>
<td>0.513* (0.278)</td>
</tr>
<tr>
<td>High Temp Anomaly</td>
<td>0.065 (0.208)</td>
<td>0.098 (0.225)</td>
<td>0.443* (0.268)</td>
</tr>
<tr>
<td>R²</td>
<td>0.16</td>
<td>0.15</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Appendix 4. Handling Missing Data

Missing Data: Often in empirical research one finds observations for some of the variables in the data set to be missing. Depending on the proportion of missing data, number of observations for analysis and further estimation are reduced significantly, questioning reliability of the coefficients thus obtained. The problem is grave for longitudinal or panel studies, with observations on several variables for each unit of study. If the proportion of missing data is very high, observed sample of observations may not be a true representative of the population, hence resulting in biased estimates.

Data missingness could be due to errors at any stage of the data collection process, non response on part of the respondents, data recording errors and due to unanticipated causes, such as loss of data files. Graham et al. (2003) describe that missing data is caused by a combination of three kinds of processes, namely, random processes, processes which can be measured, and processes which cannot be measured. There are three missing data mechanisms, namely, missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). Data for a variable X is said to be missing completely at random if the probability of missing data of X is independent of the values of X and other measured variables. In such a scenario, the observed data are a simple random sample of the hypothetically complete data set (Enders, p. 8). Data for a variable X is missing at random if the probability of missing data is related to other measured variables apart from X. On the contrary, data is missing not at random (MNAR) when the probability of missing data on X is dependent on X, even after controlling for the effect of other variables. However, MCAR, by assuming missing data to be unrelated to the observed data, imposes stringent restrictions on the data which may or may not be satisfied for any given dataset.

In our data, the underlying cause of missingness is not fully known. Data for all variables is a compilation of several reports published by ministries under the Government of India and state governments, and other institutions involved in data collection and dissemination. Hence, missing data is due to data compilation from various sources.

One of the earlier approaches used for handling missing data is to discard the observations with missing data. These approaches discard all observations with missing data (known as list wise deletion or complete case analysis); or obtain estimates for pairs of observations where data for the dependent and all independent variables is available (known as pair wise deletion or available case analysis). However, both techniques reduce the sample size significantly. A major problem with both approaches is the assumption that missing data mechanism underlying the data is MCAR yielding distorted parameter estimates if this assumption is violated. Several techniques are used to fill in missing values in the data. List wise and pair wise deletion are among the popular methods used for filling in the missing values. However, they assume data to be missing completely at random, yielding distorted parameter estimates when data is not MCAR. Single imputation methods were further developed as an improvement over the traditional techniques based on deletion of the cases with incomplete data. These methods include arithmetic mean imputation, hot deck imputation and regression imputation, each of which replace the missing value with a single value. Arithmetic mean imputation involves replacement of the missing values with mean of the observations for which data is available. Despite reduction in variability of the data, the method attenuates covariance (and hence correlation) between variables. The method produces distorted parameter estimates, even when the data is missing completely at random. Rather, results
from simulation studies point to a significant bias in the imputed estimates (Brown, 1994; Enders and Bandalos 2001; Gleason and Staelin 1975; Kim and Curry 1977; Kromrey and Hines 1994; Olinsky, Chen and Harlow 2003; Raymond and Roberts 1987; Timm 1970; Wothke 2000) concluding that mean imputation is possibly the worst imputation technique and should be avoided (Enders, p. 43). Hot Deck imputation replaces missing observations with a random draw from a subset of observations with similar values of the other observed variables. However, the method results in biased correlation coefficients and regression estimates (Brown 1994; Schafer and Graham 2002) apart from underestimating the standard errors. Regression imputation (or conditional mean imputation) replaces missing observations with the predicted values from a regression equation of the missing variable on the set of regressors with complete data. However, imputed values lie exactly on the estimated regression line overestimating correlations between the variables. Stochastic regression imputation preserves the inherent variability in the data because of the missing observations by adding a normally distributed error term to the predicted values. Resulting parameter estimates are unbiased, even when the data is missing at random or missing completely at random. Each of the simple imputation techniques reduces variability in the imputed variables, increasing the risk of type I error. Another technique commonly used in longitudinal studies in medical sciences, is to replace missing observations with the value preceding them, hence termed as ‘last observation carried forward’. However, it is shown to produce distorted parameter estimates and is a poor strategy for dealing with missing data (Cook et al. 2004; Liu & Gould 2002; Mallinckrodt et al. 2001; Molenberghs et al. 2004; Shao and Zhong 2004).

Till the 1990s, researchers continued to rely on traditional missing data techniques, because of computational difficulties in using multiple imputation and maximum likelihood methods. Rubin (1996, 1987) found that these “ad hoc” methods of handling missing data yield statistically invalid answers for scientific estimands and proposed the method of multiple imputation. Data is assumed to be missing at random and normally distributed. There are no statistical tests to verify if the missing data mechanism is MAR. However, studies have found that the MAR assumption holds for most of the cases with rare incidences of serious violations (Graham et al. 1997 p. 354; Schafer & Graham 2002, p. 152). Resulting parameter estimates thus obtained have lesser bias and greater power. Rather, researchers consider both multiple imputation and maximum likelihood estimation as the current “state of the art” (Schafer and Graham, 2002). Both are asymptotically equivalent and produce similar results.

Multiple imputation uses stochastic regression imputation to impute the missing values. The entire process consists of three phases, namely, imputation phase, analysis phase and pooling phase respectively. However, only the first phase is relevant for our study. In the imputation phase, several copies of the original data set are created, depending on the number of imputations, where missing observations are replaced with the imputed counterparts. Hence, each missing value can be replaced with any of the plausible values obtained from imputation, which differs from the single imputation technique, where each missing observation can be replaced with a single point estimate. In the analysis phase, regression estimations are performed for each the imputed data sets. The resulting coefficient estimates and standard errors obtained are combined in the pooling phase to yield one set of coefficients and their standard errors.

As expected, data on crop production is not available for all observations with missing data.
on crop area, the only exception being data on rice area and production for Santhal Paragana district in 2004, where rice production is reported to be 536000 tonnes, where as data on rice area is not available. However, no such relationship can be found between missing data patterns in the variables for total irrigated area and crop area, total HYV area and crop area respectively.

Missing values were imputed using predictive mean matching. It is preferred over the fully parametric linear regression, since it preserves distribution of the data. The method requires regressing the incomplete variable on the missing variable. However, since all the variables were incomplete, each of the missing variables were regressed on the intercept. Further, regressions were performed separately for each district, to account for district level heterogeneity. Choice of the number of imputations depends on the fraction of missing data.

Simulation studies suggest that number of imputations affects power (Graham, Olchowski and Gilreath 2007) and efficiency of the coefficient estimates (Rubin, 1987, 1996; Schafer, 1997; Schafer and Olsen 1998). 10 imputations were considered suitable for the present study.

Variables: After imputing missing values in the data, the dependent and independent variables were created. Crop yield was obtained using data on crop production and crop area, (where crop yield is crop production per unit area). Estimates of crop area and production account for multiple cropping indicative of crop sown throughout the year. Crop area and production are reported in ‘000 hectares and ‘000 tonnes respectively, hence the dependent variable is measured in tonnes per hectare.

Fertiliser variable is the aggregate consumption of nitrogen, phosphate and potash used for production, measured in tonnes. However, district level aggregate fertiliser application is known. Hence, crop specific fertiliser consumption was obtained by pro rating total consumption by proportion of area under the crop (i.e. district level crop area as a proportion of gross cropped area), which is expressed per hectare. Similarly, data on crop irrigated area and area under high yielding variety seeds was expressed per hectare. However, crop irrigated area was found to be exceeding area under the crop for some of the observations. Similarly, crop wise area under high yielding variety seeds was found to be exceeding area under the crop. These observations are a result of data compilation from several sources. The number of such observations (percentage in brackets) for both variables is tabulated below:

<table>
<thead>
<tr>
<th>Crop / Variable</th>
<th>Irrigated Area (proportion)</th>
<th>HYV Area (proportion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice</td>
<td>350 (5)</td>
<td>210 (3)</td>
</tr>
<tr>
<td>Sorghum</td>
<td>8 (~0)</td>
<td>385 (7)</td>
</tr>
<tr>
<td>Pearl Millet</td>
<td>16 (~0)</td>
<td>335 (8)</td>
</tr>
</tbody>
</table>

(percentage reported in brackets has been rounded off; ~0 denotes value less than 0.5)

However, for these observations, values for both irrigation and hyv variables were found to
be close to one, and hence were treated to be equal to one.

**Choice of districts**: The number of districts selected for rice, sorghum and pearl millet are 162, 119 and 91 respectively, which have been selected on the basis of average crop production during 1966-2011.

Data on crop production for 311 districts of India was used to compute average production during the 46 years of study. For rice, districts with average production exceeding 100000 tonnes, whereas for sorghum and pearl millet, those with average production of at least 10000 tonnes were selected. In general, chosen districts account for around 95 percent of the all India crop production. Crop wise, selected districts for rice, sorghum and pearl millet account for 93.92 (≈ 94), 95.42 (≈ 95) and 93.99 (≈ 94) percent respectively of the all India production.

Rice is produced almost all over the country with Himachal Pradesh and Rajasthan being the only exception, possibly due to unfavourable and extreme climatic conditions observed in these states. On the other hand, production of millets is concentrated, with major sorghum and pearl millet producing districts spread across 9 and 10 states respectively.
References


Auffhammer, Maximilian, V. Ramanathan, and Jeffrey R. Vincent. 2006. “Integrated model shows that atmospheric brown clouds and greenhouse gases have reduced rice harvests in India.” *Proceedings of the National Academy of Sciences (PNAS)*, 103(52): 19668-19672(December 26).


Food and Agriculture Organization (FAO). 2013. A regional rice strategy for sustainable food security in Asia and the Pacific.


