Marazzano Colon April 27, 2020
Professor: Robert Garlick Econ 442

Final Research Paper

Introduction:

Research Question:

In developing countries, do farmers adjust what plants they grow after weather shocks such as rainfall? If so, is the adjustment different if there is a low amount of rainfall, negative shock, versus if there is a great amount of rainfall, a positive shock? Furthermore, are there substitutions between crops depending on if there was a negative shock or positive shock?

Economic Relevance:

This is a relevant economic question for the very reasons why the answers to these question are not immediately obvious. The behavioral economic and the agricultural policy aspects of this question will be fleshed out in this section because they serve as useful models for understanding how farmers could react to rainfall shocks and, thus, help with the interpretability of the results.

First, it has been well documented that agriculture is a significant source of income for many household in the developing world (Banerjee 2007). For some households, agriculture is the primary source of income but the majority of households use agriculture to supplement their income and diets. A rainfall shock is thus indicative or an income shock as well. The investigation of the adjustments of what crops a farmer grows tells us about the farmers future outlook, current situation, and goals.

There is much to be gained by looking at this problem from a behavioral economic lens. Rainfall is a proxy for an income shocks and there are multiple ways a farmer could react to this. The decision depends on if, for the farmer, the income effect is positively or negatively associated with more risk. For example, if there was a lot of rainfall this year which caused a positive effect on income, a farmer could try planting riskier crops that have higher yield because he/she can now afford to take on more risk. Or the farmer could plant low yield crops that have lower risk because they are content with the higher than normal income they currently have.

Another behavioral economic phenomena this research paper evaluates is if there is negativity bias for any reactions that farmers do have in response to rainfall shocks. The negativity bias is the notion that things of a more negative nature (e.g. unpleasant thoughts, emotions, or social interactions; harmful/traumatic events) have a greater effect on one's psychological state and processes than neutral or positive things (Baumeiste 2001). The negativity bias would manifest in the data if the magnitude of change for negative shocks are greater than the magnitude of change for any similarly sized positive shock.

The behavioral economic questions are insightful because their answers tells us about how rainfall/income shocks impacts the individual. This information would allow policy makers to better help farmers impacted by rainfall shocks. In a future where climate change will increase the variability of weather, and thus increase amount of droughts, it is important to understand how farmers react to negative and positive rainfall shocks in the present day.

Another major question and concern is what other factors impact what crops a farmer decides to grow (e.g. agricultural subsidies, access to credit, soil and topographic suitability) or the long term change in crops. These concerns will be addressed throughout the paper but mostly in the interpretation and limitations section of this paper.

Data Preview:

The research questions evaluates the relationship between a change in crops and rainfall during the previous year. To elucidate this relationship, agricultural and rainfall over the same geographic area and a method to match them is necessary.

The agricultural data used in this research report comes from a series of yearly statistical reports published by the Indian state of Haryana. The statistical abstracts contains agricultural information on each of the districts that make up Haryana. The agricultural information of use to this paper was is amount of land committed to individual crops and the total amount of land committed to all crops in each districts. This information is used to measure the main outcome variable used in the paper. There is also information on the what percentage of various crops were of the high yield variety (HYV) in each district. An important thing to mention about this data is that nowhere in the statistical abstracts does it mention how the agriculture data is collected. This most definitely impacts any interpretation of the data and blindly opens the results of this paper to omitted variable bias and other unknown types of measurement error. This opaque, if not unclear, provenance of the data will be spoken about in the limitations sections of this research paper.

The rainfall data is made publicly available by the Department of Geography at the University of Delaware. The rainfall data contains all terrestrial rainfall dating back from 1900 to 2017 on a yearly basis. The rainfall data is spatially interpolate to a 0.5-degree by 0.5-degree latitude/longitude grid, where the grid nodes are centered on the 0.25 degree. It is important to know that rainfall amounts in Haryana is a random variable, in that rainfall one year does not predict rainfall another year (Shah 2017). The random nature of the rainfall is also reconfirmed in this paper.

The location of each Haryana district was mapped to one of the grid nodes manually using google maps. The resulting values were also cross referenced with a paper that mapped the Haryana districts to University of Delaware data (Shah 2017).

Empirical Strategy Preview:

The research question of this paper can be empirically evaluated by measuring how the portion of land committed to a crop changes conditional on a negative or positive rainfall shock in the previous year. Where a negative rainfall shock is considered to be a yearly rainfall value in the lower 20th percentile of all yearly rainfall data values and a positive rainfall shock is a value in the upper 80th percentile all yearly rainfall data. The regression that measures this relationship is this one:

$$Y_{ijt} = a_0 + a_1 U_{j(t-1)} + a_2 L_{j(t-1)} + \epsilon \quad (1)$$

Where variable Y is the change from the previous year of the percentage of total land committed to a plant. Variable U is 1 if yearly rainfall is in the upper 80th percentile in the previous year and 0 otherwise. Variable L is 1 if the yearly rainfall is in the lower 20th percentile in the previous year and 0 otherwise. Coefficients a_1 and a_2 is the expected percentage point change in the area committed to a crop conditional on a positive or negative shock respectively. This simple regression, and slight variations of it, are the main tools to investigate the relationship between the change in crops planted and rainfall shocks the previous year.

There is precedent in the literature for using rainfall and rainfall shocks as a predictive variable—through the use of the positive relationship between rainfall and income—to investigate how rainfall shocks impact food insecurity (Demeke 2011), childhood school attendance and human capital (Jacoby 1997, Shah 2017) .

It is also important to understand the relative water needs of crops grown in and what season each crop is typically grown in India to interpret the results. This data is surmised in the empirical strategy section of this paper. The most important components of this is that Bajra needs little water while Rice needs a lot of water. Rice and Bajra are both grown in the wet season. Maize, Wheat, and Barley need a lot, an average amount, and very little water respectively to grow and that they are all grown in the dry season. Crops that grow during the same season are should be primarily compared with each other.

Preview of Results:

This research paper finds that there is no evidence of a trend between crops planted and rainfall shocks and that there is also no substitution between crops (eg. Rice vs Bajra) and crop types (eg. drought resistant vs non-drough resistant) on a district level. Under closer analysis of the data, the lack thereof of any relationship is due to that, on a district-aggregated basis, farmers do not adjust crops on a yearly basis. The data used in this paper only provides agricultural data at a district level. Thus, the results cannot be used to make substantial claims about individual farmer behavior.

Review of Data:

Rainfall Data:

The rainfall data is made publicly available by the Department of Geography at the University of Delaware.[1] The rainfall data contains global terrestrial rainfall dating back from 1900 to 2017 on a yearly basis. This papers uses the rainfall data from 1975-2017. University of Delaware created this data set by aggregating weather station data from multiple sources. The rainfall data from India in particular comes from the National Center for Atmospheric research. Researchers from the University of Delaware then looked through the data for and removed duplication and unrealistic data points. This rainfall data was then interpolated to a .5-degree by 0.5-degree latitude/longitude grid, where the grid nodes are centered on the 0.25 degree. Also, statistical analysis of this data set determined that one year does not predict rainfall another year in the state of Haryana, thus random rainfall can be treated as a random variable. Also, there is not a significant amount of rainfall heterogeneity within the districts of Haryana (Shah 2017).

Haryana District Agricultural Data:

The agricultural data used in this research report comes from a series of yearly reports issued by the Department of Economic and Statistical Analysis, Haryana[2]. Each report having a separate name, the report for the 2016-2017 year would be called “Statistical Abstract of Haryana 2016-17.” These reports can be found on the government agencies’ website (link here) from 2006 to 2018, for a total of 12 reports.

The statistical abstracts from each year contain agricultural information on each of the districts that make up Haryana. Some districts changed name (Mewat changed to Nehal) and some were removed (Palwal) or added (Charkhi Dadri). In the end, this paper was able to use data from 20 districts because their data was able to be followed for all 12 years.

Each statistical abstract gave information on the same 20 individual crop variables (e.g. Rice, Maize, Bajra) and 7 variables that aggregated the crop variables by type (e.g. Total Cereals,

[1] http://climate.geog.udel.edu/~climate/html_pages/download.html#P2009
[2] http://esaharyana.gov.in/en-us/State-Statistical-Abstract-of-Haryana

Total Foodgrains) for a total of 27 crop variables of interests. Importantly, the area sown for each crop variable was given for every district and each yearly report provided the total area sown for each district. These datapoints allowed for the calculations of the proportion of total area sown for each crop variable for all 12 years, the primary outcome variable for this research paper.

The most prevalent crops by proportion of total area sown were Rice, Wheat, Maize, and Bajra in descending order, each with values hovering between 20%-40% depending on the year and district.

The yearly abstracts also have high yield variety (HYV) crop data for each district for Maize, Rice, Bajra, and Wheat. Amongst other information, the percentage of each of the four crops that were HYV is recorded for each district is in the abstracts.

The quality of data for each crop variable varied depending on the prevalence of the crop in Haryana and in each district. For example, data for linseed, one of the 20 individual crop variables, was unusable because it may have not been grown at all in some districts and if so in very small amounts. Thus, the number of a crops variable observations is strongly associated with the prevalence of the crop. This also means that Rice, Wheat, Maize, and Bajra have some of the highest number of observations. The HYV crop data was of lower quality in comparision to the 27 crop variables of interest. All of these variations in the quality of data mean that the sample sizes are quite different for the different crops, ranging from 240 for wheat to 32 for HYV maize. Beyond year and importance, it unknown why the data quality varies because it is unknown how the data was collected.

On this topic, the largest concern of the agricultural data is that nowhere in the statistical abstract does it explain how the data was generated. This limits the interpretation of the data because there is no way to know from whom or what type of farmers (or even how a farmer is defined) this data is collected from. However, throughout the project there were no glaring issues with the data that would hint at some systematic issue, as trends stayed relatively steady overtime. Further discussion of how the opaque provenance of this data limits and affects the interpretation of results is in the limitation section of this research paper.

Finally, the abstracts are only available in pdf format. To get the data in a form that could be used, an online pdf to excel converter was used[3]. Great care was taken to make sure the data was copied correctly. The empirical strategy actually cross referenced the data with itself in the creation of the what percentage of the total crops were one of the 27 crop variables of interest. Abnormalities in the data were corrected and noted. The one abnormality that couldn't be fixed was the Total_Cereals crop variable of interest, but results from this variable are not seen in this paper.

Location Data:

A necessity of this research paper is to map rainfall data from the University of Delaware to each of the 20 districts. To do this each districts needed to be matched a .5 by .5 latitudinal/longitudinal quadrant of the rainfall data. Luckily there was a research paper that already matched the Haryana districts to quadrants of the University of Delaware rainfall data (Shah 2017). For the interest of time, these matching were used as a guide and deviations were only made when it appeared that there was another quadrant that covered more of the district area. The amorphous boundaries of the Haryana districts resulted in that some districts are mapped to the same rainfall quadrant. This should be of little concern because 1. the multicollinearity of rainfall means that any rainfall values of nearby quadrants were going to be

[3] https://smallpdf.com/pdf-to-excel

similar and 2. there seemed to be no reason why a quadrant should be chosen other than the one that covered the most surface areas.

Empirical Strategy

Rainfall shocks as proxies for Income Shocks:

Throughout this paper an assumption has and will be made that positive and negative rainfall shocks serve as a proxy for positive and negative income shocks. There is considerable evidence for this being a reasonable assumption to make. First, it is a given that there is a relationship between rainfall and agricultural production. Second, according to the Department of Economics and Statistical Analysis of Haryana, two-thirds of the households in Haryana depends on agriculture for at least a portion of their income. It is not explicitly stated if the households sampled to create the Haryana agricultural data used in this paper depend on agriculture for their income, but it would be very unlikely if that was not so, especially since we know two-thirds of households in Haryana already depend on agriculture for a part of their income. These reasons lead this paper to assume that rainfall data can be used as a proxy for income shocks. The strength of this assumption is critiqued in the interpretation and limitation section of this paper.

Discussion on Crops and Rain sensitivity:

Wet Season			Dry Season		
Drought	Normal	Water Intensive	Drought	Normal	Water Intensive
Bajra	Gram	Rice	Barley	Wheat	Maize
	Moong		Rapeseed	Jowar	Sugarcane
	Massar			Mash	
				Potatoes	
				chillies	
				Tobacco	

Figure 0: Tabulation of what season the crop is planted and the relative water needs of the crop

This chart was made using irrigation requirement data for each of the 20 individual crop variables from apnikheti.com, a website that provides data on common agricultural crops for farmers in India. The purpose of this chart is to categorize crops that grow in the wet or dry season by their relative water consumption. It is meant to help guide the reader's interpretation of how a crop would react to positive or negative rainfall shocks. This chart should be taken lightly as it was not made using any serious empirical methods, only the relative water needs of the crops. This chart could be greatly improved with more time (e.g. Tobacco is omitted because data on it in India was not found). The most important crops are Bajra, considered the poor man's crop and is extremely drought resistant, Rice, a crop that cannot grow without a lot of water, Maize, also needs a lot of water, Wheat, and Barley, renowned for its drought tolerance.

Strategy to answer the main empirical question—Change in Crops caused by Rainfall Shocks:

The overall purpose of this paper is to elucidate the relationship between the change in what a farmer grows and a rainfall shock in the previous year. This question must be reformulated in a way so that it can be empirically evaluated: 'how does the percentage of the total land sown that an individual crop is sown change, conditional on a positive or negative rainfall shock in the previous year?" Where a positive rainfall shock is a yearly rainfall value in the lower 20th percentile of all yearly rainfall values. Similarly, an positive shock is a yearly rainfall value in the upper 80th percentile of all yearly rainfall values.

The following quantities and OLS models investigate the overall research question:

(1) $$Y_{ijt} = (\frac{C_{ijt}}{T_{jt}} - \frac{C_{ij(t-1)}}{T_{j(t-1)}}) * 100$$

(2) $$Y_{ijt} = a_0 + a_1 U_{j(t-1)} + a_2 L_{j(t-1)} + \epsilon$$

(3) $$Y_{ijt} = a_0 + a_1 U_{j(t-1)} + a_2 L_{j(t-1)} + \mu_j + \epsilon$$

Variable & Index List:

i ≔ **crop,** j ≔ **district** , t ≔ **year**

C ≔ area committed to crop “i”

T ≔ total area committed to all crops

U≔ as “1” **if** rainfall in upper **80th percentile**

L≔ as “1” **if** rainfall in lower **20th percentile**

μ ≔ district fixed effect

To see how these values investigate the research question, first look at equation (1) and the variable and index list. Variable Y_{ijt} is the change in the percentage of total land sowed that crop “i” is sown on in district “j” in year “t”. The variable Y, and similar quantities, will be referred to as YoY, short for “year over year change” or “year over year”, throughout the rest of this paper for the purpose of simplicity and it is the main outcome variable used in this paper. Note that not all subscripts may be listed or explicitly stated for the purpose of simplicity here on after for the simplicity also.

YoY change is calculated using variables C and T. Since C_{ijt} is the total area sown for crop “i” in district “j” and T_{jt} is the total area sown for all crops in district “j”. Then C divided by T in period “t” minus C divided by T in period “t-1” is the change in the percentage of the total land sown that an individual crop “i” is sown in.

Now evaluate equation (2). Equation (2) is an OLS model that investigates the relationship between a change in crops grown and rainfall shocks. To see this one could evaluate the coefficients and intercept of the model. The coefficient “a_1” is the expected percentage point change of the total amount of land sown where crop “i” is sown if, in the previous year, there was a yearly rainfall value in the upper 80th percentile of all yearly rainfall values. Similarly, coefficient “a_2” is the expected change in YoY if there was a negative rainfall shock in the previous year.

Equation (3) is the OLS model (2) but with district fixed effects added. The reason that there are no yearly fixed is because of the multicollinearity of variables U and L. For example, if there is a positive shock in one district then, due to the geographic proximity of the other districts, then most other districts (~80%) have positive shocks also. In this case the yearly fixed effect will "capture" the impact of a positive shock instead of coefficient "a_1".

Strategy for HYV as a measure of risk tolerance:

The yearly Haryana abstracts also contain information on the what percentage of Rice, Maize, Bajra, and Wheat are of high yield variety (HYV) in each district. This is very fortunate for the research project.

HYV crops require more water and more fertilizers. This results in that HYV crops are also riskier to grow than the non-HYV strains, but with the benefit of higher yields. This risk-reward tradeoff makes the percentages of HYV crops grown in a year a proxy for the measurement of risk tolerance when compared to the non-HYV strain.

This project exploits this in order to measure whether or not rainfall shocks in the previous year, as proxies for income shocks, induce farmers to take more or less, HYV crops, a proxy for risk, the next year. This effect can be measured with OLS models that differs from OLS model (2) and (3) where variable Y is defined as the year over year change of the percentage of crop "i" that is HYV. Let's consider equations (4) and (5) below modeled after equation (2) and (3):

$$(4) \qquad Y_{ijt} = a_0 + a_1 U_{j(t-1)} + a_2 L_{j(t-1)} + \epsilon$$

$$(5) \qquad Y_{ijt} = a_0 + a_1 U_{j(t-1)} + a_2 L_{j(t-1)} + \mu_j + \epsilon$$

Let Y, henceforth referred to as YoY also, be defined as the year over year change of the percentage of crop "i" that is HYV. Then, with the use of the variable and index list used earlier in this section, in OLS model (4) the coefficients "a_1" and "a_2" are the expected change in the percentage of crop "i" that is HYV if there is a positive or negative rainfall shock respectively. This analysis hold true in OLS model (5) with the added caveat of controlling for the district averages through the added district fixed effects.

OLS models (4) and (5) serve as a measure of how positive or negative income shocks impact risk tolerance because of the relationship between rainfall and household income described in the "*Rainfall shocks as proxies for Income Shocks*" subsection. For example, if "a_1" is positive and "a_2" is negative and both, then that would mean that if there is a positive relationship between rainfall shocks and risk tolerance.

There are no yearly fixed effects for the heterogeneity problem discussed in the previous subsection.

Relationship between Rainfall and YoY:

The research question, in part, seeks to discover whether or not rainfall shocks can induce farmers to change what crops they grow. Due to the heterogeneity of YoY changes where there was no rainfall shock the previous year, it was necessary to rule out any underlying relationship between a change in crops and the amount of rainfall the previous year. This was accomplished with the following model:

$$(6) \quad Y_{ijt} = c_0 + c_1 R_{j(t-1)} + \epsilon$$

Variable Y_{ijt} is the change in the percentage of total land sowed that crop "i" is sown on in district "j" in year "t". Variable R_{jt} is the total yearly rainfall (mm) in district "j" and "t" in district. Thus the OLS model coefficient "c_1" is the effect of an addition mm of rainfall has on the change in crop "i" the next year. If the coefficient "c_1" is sufficiently large and significant then there is evidence for a general relationship between rainfall the previous year and the change in what is planted the next year.

This regression was run for every crop variable of interest including HYV crops.

Rainfall Data Investigation Strategy:

The University of Delaware rainfall data was recompiled in order to match the July 1^{st} to June 30^{th} agricultural year used by the agricultural data in the Haryana statistical abstracts. For this reason whenever referencing yearly rainfall in this paper it actually refers to the agricultural year in Haryana.

The yearly 20^{th} and 80^{th} percentiles for each district were determined using 1975 to 2017 data, despite the data going back to 1900. The year 1975 is a slightly trivial number and it was chosen in part because that was a value used in other literature trying to calculate percentiles using the Delaware Univerisity rainfall data (Shah 2017). The reason why it is not completely trivial is because of climate change. In order to mitigate the impact of climate change on the results, 1975 was chosen out of convenience.

It has also been independently verified that that there is no serial correlations of rainfall in Haryana and that there also was not significant within-district variations in rainfall (Shah 2017). To verify that there was, at least, no serial correlation a simple regression was run for each crop:

$$(7) \quad R_{jt} = b_0 + b_1 R_{j(t-1)} + \epsilon$$

Where R_{jt} is the total yearly rainfall value (mm) in district "j" at time "t". If b is small and close to zero (which it is as b_1= -.03) then that reinforces the claim that rainfall one year does not predict rainfall another year.

Empirical Analysis of Long Run Change in crops:

The variable of interest YoY, is a measure of yearly change. The overall empirical analysis schema does not measure long term change. To compensate for this shortfall, the change over totalis be analyzed. Looking at equation (7), this was done by using histograms to observe the quantity C/T in year 2018 minus C/T in 2008 for each district. These values referred to as the “long_change” for each district. The most relavant of these findings are presented in the results section.

Admittedly, this is a rather crude way to evaluate the long run change in the proportion of total area committed to an individual crop. A better strategy would be to run a linear regress through each C/T value for each of the 13 available years. This was, sadly, not done in the interest of time.

General Remarks about Empirical Strategy:

Most of this section has gone over regressions models. However there are various scatter plots that were used to investigate the data and research question. These plots well be explained as they are presented in the results section of this research because of the simplicity of interpretation and for the conservation of space and time.

Results

Change in Crop Variables due to Rainfall Shocks:

Regression Results Table:

Table 1a & 1b
Tabulation of regression results from OLS Models (2) & (3):

	Dependent variable: YoY					
	Rice (Wet) (1)	Bajra (Wet) (2)	Maize (Dry) (3)	Wheat (Dry) (4)	Barley (Dry) (5)	Total_Foodgrains (6)
Upper	0.586*	-0.249	-0.192	-0.954*	-0.087	-0.333
	(0.340)	(0.285)	(0.203)	(0.536)	(0.065)	(0.854)
Lower	-0.228	-0.220	0.122	0.604	-0.102*	-0.129
	(0.325)	(0.270)	(0.201)	(0.509)	(0.060)	(0.811)
Intercept	0.348	0.032	-0.118	0.143	0.037	0.237
	(0.211)	(0.176)	(0.126)	(0.331)	(0.044)	(0.528)
Observations	229	226	105	240	136	240
R^2	0.023	0.005	0.019	0.031	0.024	0.001
Adjusted R^2	0.014	-0.004	0.000	0.022	0.009	-0.008
Residual Std. Error	2.080 (df=226)	1.726 (df=223)	0.858 (df=102)	3.347 (df=237)	0.295 (df=133)	5.331 (df=237)
F Statistic	2.652* (df=2; 226)	0.510 (df=2; 223)	1.005 (df=2; 102)	3.740** (df=2; 237)	1.632 (df=2; 133)	0.076 (df=2; 237)
Note:						*p<0.1; **p<0.05; ***p<0.01
						IMPORTANT : This regression has no fixed effects

W/O FE

	Dependent variable: YoY					
	Rice (Wet) (1)	Bajra (Wet) (2)	Maize (Dry) (3)	Wheat (Dry) (4)	Barley (Dry) (5)	Total_Foodgrains (6)
Upper	0.577	-0.186	-0.169	-0.971*	-0.084	-0.289
	(0.355)	(0.296)	(0.220)	(0.554)	(0.072)	(0.884)
Lower	-0.208	-0.217	0.114	0.699	-0.105	0.000
	(0.337)	(0.278)	(0.215)	(0.521)	(0.065)	(0.832)
Intercept	0.240	0.112	-0.079	0.360	-0.000	0.535
	(0.640)	(0.527)	(0.275)	(1.018)	(0.313)	(1.626)
Observations	229	226	105	240	136	240
R^2	0.055	0.055	0.142	0.077	0.040	0.044
Adjusted R^2	-0.041	-0.043	-0.038	-0.012	-0.117	-0.048
Residual Std. Error	2.138 (df=207)	1.758 (df=204)	0.874 (df=86)	3.404 (df=218)	0.313 (df=116)	5.438 (df=218)
F Statistic	0.574 (df=21; 207)	0.563 (df=21; 204)	0.788 (df=18; 86)	0.869 (df=21; 218)	0.256 (df=19; 116)	0.474 (df=21; 218)
Note:						*p<0.1; **p<0.05; ***p<0.01
						IMPORTANT : This regression has fixed effects

W/ FE

Table 1 contains the key result of regressions (2) and (3). Only six crop variables are displayed in order to conserve space and for the sake of interpretability. Crops that are grown during the wet season are labeled with "Wet" in parenthesis and are in a box. Similarly, crops that are grown during the dry season are in a box. Wet season and dry season crops should only be compared with each other. "Upper" and "Lower" correspond to the U and L variables in OLS models (2) and (3). These OLS models investigate the question of 'how does the percentage of the total land sown committed to an individual crop change conditional on a positive or negative rainfall shock in the previous year?"

The results from multiple regressions of (2) and (3) give no evidence for any relationship between the change in what crop a farmer grows and rainfall shocks the previous year. Thus, there is also no evidence of substitution between crops caused by rainfall shocks.

Discussion of Regression Result:

The above table gives no evidence for a relationship between how farmers adjust their crops and rainfall shocks during previous year for two main reasons: 1. The signs of the coefficients seem random and counterintuitive 2. The size of coefficients are very small and OLS coefficients are mostly statistically insignificant.

To see how the coefficients seem random and counter intuitive, first look at the coefficients for Rice and Bajra, both wet season crops. Rice is a crop that has a relatively high yield in comparison to the other crops grown in Haryana (and in general). It is also slightly riskier crop to grow because it is a rain intensive crop. These characteristics of Rice crop can be rationalized with a positive "Upper" coefficient could positive and a negative "Lower." One could say that at a district level, if there was a positive rainfall shock last year then farmers can afford to take on more risk and grow more rice the next year and that if there was a negative shock they would take on less risk the next year and grow less rice. However, there is no way rationalize that both the "Upper" and "Lower" coefficients are negative. Bajra is considered the poor man's crop because it is very drought resistant. There are no characteristics of the crop that would induce someone to plant less of it the next year if there was too little or very plentiful rainfall the previous year. If anything, the "Lower" coefficient should be positive for Bajra because of it's drought resistance. In short, evaluation of the signs of the other crop variables in the regression are indistinguishable from random, this includes all the 21 other crop variables.

Just because coefficients don't appear to be interpretable, doesn't mean there is not a relationship between YoY and rainfall shocks. However, the metaphorical final nail(s) in the coffin is that the coefficients themselves are very small, most are <1 pp , such p-values imply that the coefficients are not significantly distinct from zero, and the R^2 values are very small or negative, implying that the OLS models (2) and (3) explain little to none of the variation in YoY change.

For these reasons, the table 1 gives no evidence for a relationship between how farmers adjust their crops and rainfall shocks the previous year and no evidence for substitutions between crops rainfall caused by rain shocks the previous year.

Change in HYV crops due to Rainfall Shocks

Regression Results Table:

Table 2:

Tabulation of regression results from OLS Models (4) & (5):

	Dependent variable:YoY			
	HYV Rice	HYV Maize	HYV Bajra	HYV Wheat
	(1)	(2)	(3)	(4)
Upper	1.272	13.427	-0.743	1.710^{**}
	(2.139)	(13.288)	(3.032)	(0.738)
Lower	7.109^{***}	-3.940	-0.683	0.823
	(1.951)	(13.288)	(2.984)	(0.679)
Intercept	-1.031	0.529	-3.119^{*}	-0.632
	(1.194)	(8.312)	(1.757)	(0.402)
Observations	154	32	165	200
R^2	0.084	0.052	0.001	0.028
Adjusted R^2	0.071	-0.014	-0.012	0.018
Residual Std. Error	10.344 (df=151)	31.102 (df=29)	15.817 (df=162)	4.059 (df=197)
F Statistic	6.882^{***} (df=2; 151)	0.793 (df=2; 29)	0.042 (df=2; 162)	2.805^{*} (df=2; 197)
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
				IMPORTANT : This regression has no fixed effects

W/O FE

	Dependent variable:YoY			
	HYV Rice	HYV Maize	HYV Bajra	HYV Wheat
	(1)	(2)	(3)	(4)
Upper	1.265	13.342	-1.666	1.805^{**}
	(2.320)	(13.521)	(3.081)	(0.794)
Lower	7.301^{***}	-2.781	-0.826	0.857
	(2.054)	(13.521)	(3.070)	(0.718)
Intercept	-0.896	7.142	-2.959	-0.553
	(1.211)	(13.117)	(5.163)	(1.382)
Observations	154	32	165	200
R^2	0.111	0.175	0.110	0.034
Adjusted R^2	-0.007	-0.023	-0.013	-0.080
Residual Std. Error	10.773 (df=135)	31.239 (df=25)	15.828 (df=144)	4.256 (df=178)
F Statistic	0.940 (df=18; 135)	0.886 (df=6; 25)	0.892 (df=20; 144)	0.302 (df=21; 178)
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
				IMPORTANT : This regression has district fixed effects not shown here

W/ FE

Table 2 contains the result of OLS models (4) and (5) OLS models. OLS models (4) and (5) serve as a measure of how positive or negative income shocks impact risk tolerance because of the relationship between rainfall and household income described in the previous section. Similar to the previous subsection, the results of this data indicate no evidence for a relationship between rainfall shocks during the previous year and change in what percentage of crops that are HYV. Thus, there results provide no evidence for a relationship between income shocks and a change in risk tolerance. Also district fixed effects has little effect on the models. A comment of why this is the case is in the interpretation and limitation section of this paper.

Discussion of Regression Result:

The above table gives no evidence for a relationship between rainfall shocks the previous year and a change in what percentage of crops are grown the next year because of the lack of interpretability of the results and the statistically insignificant nature of most of the results. These are the same problems that plagued the regressions in the previous section.

The OLS model results are uninterpretable because of the sign of the "Upper" and "Lower" coefficients. For example, both the "Upper" and "Lower" coefficients for rice are positive. This means that if there is plentiful or very little rainfall one year, then the next year farmers, on a district level, plant more HYV rice. This is difficult to interpret because rice is a water intensive crop, so to imagine that farmers would plant more HYV rice, which is even more water intensive than non-HYV rice, is unlikely.

However, just because there appears to be no interpretable relationship, does not mean that there is no actual relationship. Perhaps, farmers "go big or go home," by planting more rice during the rainy season when they have too little rainfall or have lots of it the previous year. This is very unlikely because of the statistically insignificant p-values for most of the regression accompanied by the low R^2 values that imply that OLS models (4) and (5) simply do not explain much of any of the variation in the YoY change in HYV crops.

For these reasons the above tables provide no evidence for a relationship between rainfall shocks during the previous year and change in what percentage of crops that are HYV.

These results give no evidence for a relationship between income shocks and a change in risk tolerance. This is because the relationship for rainfall shock and change in the proportion of HYV crops planted are considered proxies for the relationship between income shocks and a change in risk tolerance, for which no relationship was found.

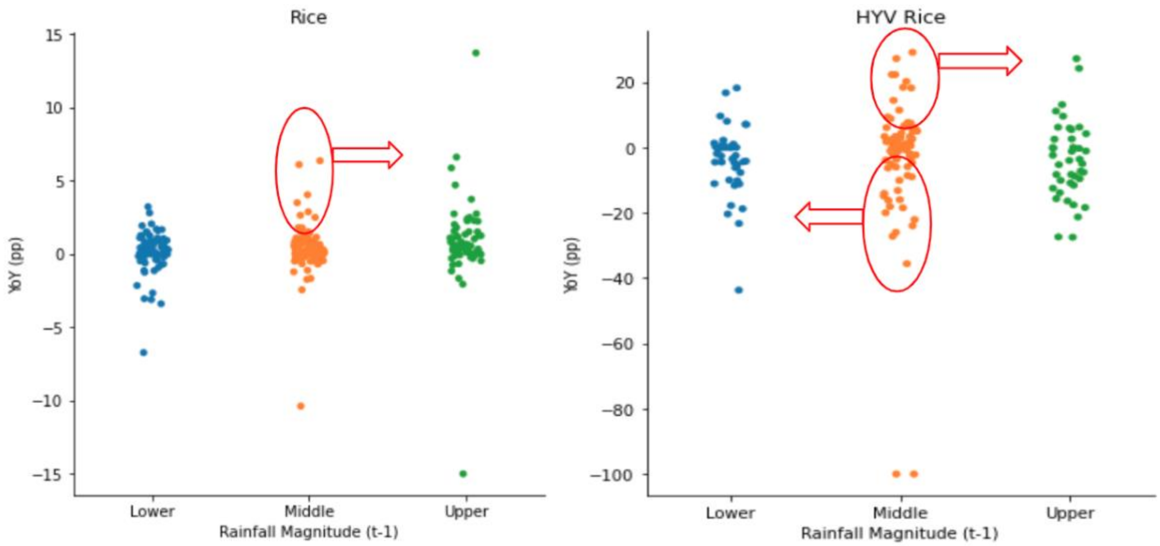
As in the previous subsection, district fixed effects did not affect these results. A comment of why this is the case is in the interpretation and limitation section of this paper.

HYV & Crop Variable Change due to Rainfall:

Visualization of Data and Regression Results:

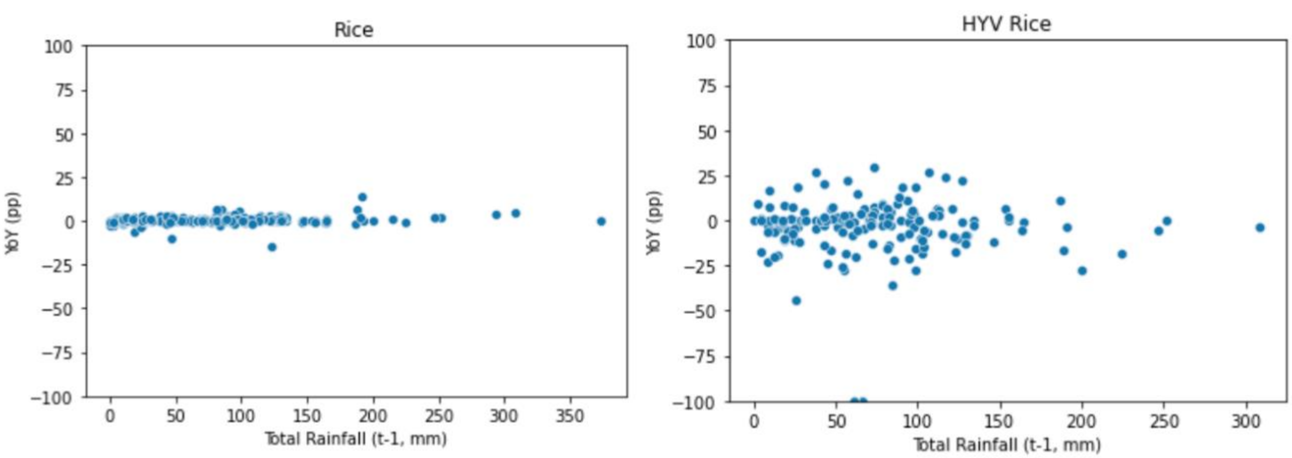
Figure 2a & 1b:

Category plots that show there is YoY variation not being captured by the "Lower" and "Upper" category variables. Everything that is not "Upper" or "Lower" is labeled as "Middle" in these charts. The encircled red values are examples of how the data could theoretically shift categories if the bounds of "Upper" (>80th percentile) and "Lower" (<20th percentile) are redefined.



Figures 1a & 2b:

Scatterplot of YoY and the Total Rainfall in the previous year.



Tables 3a & 3b:

Tabulation of regression results from OLS Models (6) for HYV crops in 3(a) and regular crops in 3(b):

	Dependent variable:YoY			
	HYV Rice	HYV Maize	HYV Bajra	HYV Wheat
	(1)	(2)	(3)	(4)
Total_Rainfall	-0.040*	0.081	-0.004	0.001
	(0.020)	(0.072)	(0.025)	(0.007)
Intercept	4.132**	2.432	-3.223	0.046
	(1.653)	(13.508)	(5.730)	(1.543)
Observations	154	32	165	200
R^2	0.052	0.171	0.109	0.005
Adjusted R^2	-0.066	0.011	-0.008	-0.106
Residual Std. Error	11.085 (df=136)	30.721 (df=26)	15.789 (df=145)	4.308 (df=179)
F Statistic	0.440 (df=17; 136)	1.070 (df=5; 26)	0.929 (df=19; 145)	0.045 (df=20; 179)
Note:				*p<0.1; **p<0.05; ***p<0.01
				IMPORTANT : This regression has district fixed effects not shown here

	Dependent variable:YoY					
	Rice (Wet)	Bajra (Wet)	Maize (Dry)	Wheat (Dry)	Barley (Dry)	Total_Foodgrains
	(1)	(2)	(3)	(4)	(5)	(6)
Total_Rainfall	0.007***	0.000	-0.001	-0.006	0.000	0.005
	(0.003)	(0.002)	(0.001)	(0.004)	(0.001)	(0.006)
Intercept	-0.408	-0.047	0.036	0.884	-0.028	-0.137
	(0.672)	(0.560)	(0.288)	(1.091)	(0.317)	(1.718)
Observations	229	226	105	240	136	240
R^2	0.070	0.052	0.137	0.053	0.021	0.046
Adjusted R^2	-0.020	-0.041	-0.032	-0.033	-0.129	-0.041
Residual Std. Error	2.116 (df=208)	1.757 (df=205)	0.871 (df=87)	3.441 (df=219)	0.314 (df=117)	5.419 (df=219)
F Statistic	0.779 (df=20; 208)	0.557 (df=20; 205)	0.810 (df=17; 87)	0.615 (df=20; 219)	0.142 (df=18; 117)	0.528 (df=20; 219)
Note:						*p<0.1; **p<0.05; ***p<0.01
						IMPORTANT : This regression has fixed effects

Discussion of Visualized Data and Regression Result:

Figures 1a and 1b show that the "Upper" and "Lower" categorical variables are not capturing some of the variation in YoY. The YoY values that were not associated with shocks, 60% of the data, can be seen in "Middle" category. Only the figures for rice are shown in this paper, but this phenomena is present in all YoY values for the 27 crop variable of interest and HYV crops. A reasonable question is would "Upper" and "Lower" variables capture some of this variable if the criteria for "Upper" and "Lower" rainfall shocks were redefined. This can be seen in the red encircled values in the data. Shifting the bounds of the categorical variables is, fundamentally, an attempt to see if there is a broader relationship with total rainfall yearly the previous year and YoY the next year.

Figures 2a and 2b and tables 3a and 3b are investigations into the question between YoY and total rainfall the previous year. Figures 2a and 2b plot the YoY change for rice and HYV change against rainfall the previous year. What is immediately noticeable is that YoY values appear to be mean centered around zero. This trend can be seen in all other YoY plots of the 27 crop variable of interest and HYV crops. This implies that farmers do not usually change what crops they grow on a yearly basis at all, let alone because of positive or negative shocks the previous year.

To verify this visual observation OLS model (6) was run, the results of which are in tables 3a and 3b. OLS model (6) measures the effect that the total rainfall in the previous has on the change in crops. The coefficients for "Total Rainfall" are all close to zero and not statistically significant, which indicates that there is no relationship between total rainfall and YoY change in HYV crops and the crop variables of interest.

Rainfall as Random Variable

Table 4

Tabulation of OLS equation (7) result:

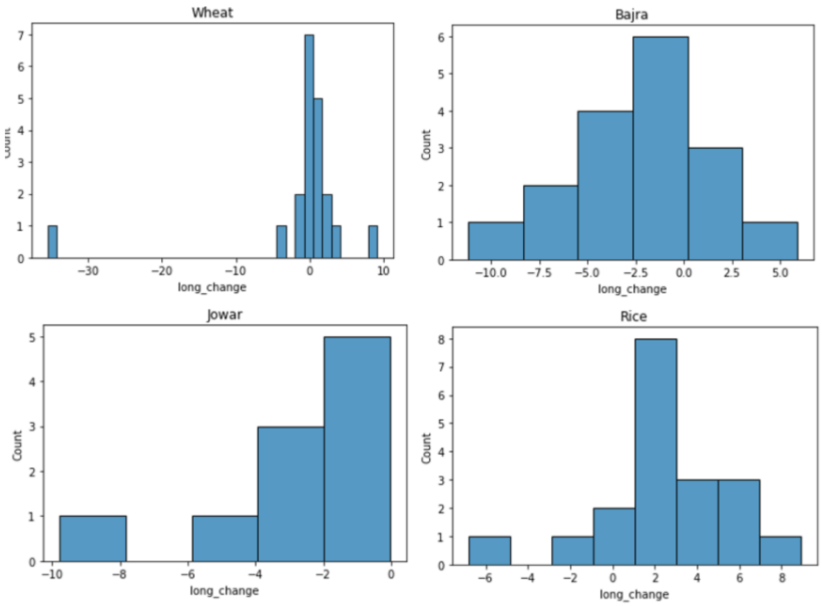
	Dependent variable:t1
	(1)
Intercept	100.145***
	(5.577)
t0	-0.363***
	(0.062)
Observations	240
R^2	0.127
Adjusted R^2	0.124
Residual Std. Error	54.893 (df=238)
F Statistic	34.778*** (df=1; 238)
Note:	*p<0.1; **p<0.05; ***p<0.01

These are the results of the OLS model (7). OLS model (7) was an investigation of if there was a serial relationship between rain one year and rain the next year. Coefficient of "t_0" is the effect of an additional 1 mm of rain one year on the rainfall the next year. This value, though statistically significant, is very small. This corroborates with prior investigation of a serial correlation of rainfall in the Haryana district rainfall data (Shah 2017).

Long Run Change of YoY:

Figure 3:

Histograms of the long_change values of various crops



There is evidence for a long run change in various crop variable of interest. In figure 3 we see that Wheat and Rice have increased in share over time, that Jowar has decreased in share over time, and that Bajra has mainly stayed the same over time.
This is not sufficient analysis to say strongly that there is long run change and the reasons for that change, but it is evidence that there is most likely long run change in the prevalence of crops in Haryana.

Interpretations and Limitations

Main Conclusion and Takeaways:

The main conclusion of this paper is that farmers do not adjust what crops they plant very much on a yearly basis, and if they do, there is no evidence in the results of this paper to suggest that it is because of rainfall shocks the previous year. If there is substitution between crops then there is no evidence in this paper that it is caused by rainfall shocks. Since rainfall shocks are known to be income shocks in the developing world (Shah 2017), this paper sees no evidence that farmers change what they grow due to rainfall induced income shocks. This all suggests that there are other mechanisms and factors that decide what crops are grown in Haryana, and India as a whole.

Limitations of Rainfall as a Predictor of the Change in Crops Sown:

The fact that rainfall is a random variable in Haryana (Shah 2017), and India as a whole, is very important. This means that in order to maximize one's yield, the farmer should have the same strategy every year. There is no reason to believe that farmers believe that rainfall last year would effect rainfall the next year. In fact, over thousands of years, Indian farmers have most likely learned to treat rainfall as a random variable which could explain why farmers do not change what they grow because of rainfall shocks.

Though, what is true is that rainfall shocks induce income shocks as discussed throughout this paper. The effect that income shocks have on what a farmer decides to plant, as discussed in the economic relevance section, is not obvious. Furthermore, an implicit assumption made throughout this paper is that income shocks one year can effect decisions the next year. This may not be the case if farmers do not have the ability or incentive to save (Banerjee 2007). This could weaken any intertemporal relationship between income shocks and agricultural decisions.

In the districts of Rewari and Mahendragarh, bajra and oilseeds make up a disproportionate percentage of crops sown. This quirk in the suggest that there could be centralized mechanism that effect agricultural decisions. One of these mechanisms could be agricultural subsidies provided by various levels of government to Indian farmers. Agricultural subsidies most definitely play an important role in what crops are planted (Kumar 2020). These subsidies could be in the form of outright cash transfers or subsidized credit conditional on growing specific crops. These subsidies most likely play a decisive role in what some farmers plant on a yearly to basis. Changes in subsidy policies, in all likelihood, are not induced by rainfall shocks. The effect of policy on agricultural decisions could be captured in some type of variable included in the OLS models used in this research paper. It is important to note that its inclusion into the models would not change the conclusions of this paper.

Access to irrigation fundamentally impacts what crops a farmer decides to grow. If a farmer can supply more water to his/her crops then the potential risk of planting higher yield crops such as rice and maize would decrease while increasing the average yield. Changes to access to irrigation on a district wide level (and the adoption of water saving technologies and techniques) would impact what crops are grown. This could explain the long run shift away from bajra to rice seen in the data. Since access to irrigation is a variable that would change slowly over time, the empirical techniques used in this research paper would struggle to capture this.

All of these aspects are an extensive but not exhaustive analysis of factors that could influence what crops a farmer will grow (e.g. topography, soil characteristics, centralized planning are not discussed). These are all forms of omitted variable bias because they can predict how a farmer will adjust what crops are grown on a yearly basis, even though their addition still would not impact the findings of this paper. The point of this subsection is to make it clear that the amount of rainfall during the previous year is probably a small aspect, amongst many, in the considerations farmers make when deciding what to grow. The results of this research paper give no evidence that it even a consideration at all.

District Fixed Effect:

Throughout this paper, district fixed effects do not impact the OLS results. This was primarily due to the facts the YoY has mean on zero. Though not inspected because of time, it is most likely that the district averages hovered around zero also.

Limitations of the Agricultura Data:

First and foremost of the problems with the Haryana statistical abstract data does not explain how agricultural data was collected. The opaque provenance of the Haryana agricultural data makes a general discussion about error and bias in the data difficult. Given that the YoY was consistently mean-zero, the prevalence of non-classical error is highly unlikely because non-classical error would most likely deviate YoY away from zero or manifest in systematic abnormalities in the data. The only systematic abnormality in the data was that the crop variable Total_Cereals. The total area of all the individual cereal crops did not sum up to the area value for Total_Cereals, however YoY change for this variable was still centered around zero. Therefore, there is no evidence for non-classical measurement error in the agricultural data even though it can't be ruled out with our current understanding of the data.

Another consequence of the opaqueness of the data is that it is unclear what is considered to be a farmer or from whom this agricultural data was collected. Particularly, if the data does not distinguish between or exclude farmers of a small holdings farmer or the farmer derives the majority of his/her income from agriculture. This could bias the estimates in unknown ways since the income shock caused by rainfall shock would be considerably smaller for a small holdings farmer than the individual who depends on it for most of their income, thus inducing different behavior.

The final limitation of the agricultural data is that it is only given on a district level. This prevents any claim that is stated in this paper from actually being strongly applied to individual farmers and households. To truly rule out the effect or rainfall shocks on households then a data set that follows individual farmers and what crops they sow over time would be necessary.

Bibliography

Banerjee, Abhijit, V., and Esther Duflo. 2007. "The Economic Lives of the Poor." *Journal of Economic Perspectives*, 21 (1): 141-168.

Baumeister, Roy F.; Finkenauer, Catrin; Vohs, Kathleen D. 2001. "Bad is stronger than good" (PDF). Review of General Psychology. 5 (4): 323–370. doi:10.1037/1089-2680.5.4.323. S2CID 13154992. Retrieved 2014-11-19

Demeke, Abera Birhanu, et al. “Using Panel Data to Estimate the Effect of Rainfall Shocks on Smallholders Food Security and Vulnerability in Rural Ethiopia.” *Climatic Change*, vol. 108, no. 1-2, 2011, pp. 185–206., doi:10.1007/s10584-010-9994-3.

Jacoby, Hanan G., and Emmanuel Skoufias. “Risk, Financial Markets, and Human Capital in a Developing Country.” *The Review of Economic Studies*, vol. 64, no. 3, 1997, pp. 311–335. *JSTOR*, www.jstor.org/stable/2971716. Accessed 21 Apr. 2021.

Kumar, Sanjeev. “Impact of Subsidies on Agriculture Sector in India.” Agri Mirror: Future India. Vol 1 Issue 2, Agri Mirror: Future IndiaIssue, June 2020, https://aiasa.org.in/wp-content/uploads/2020/09/32.pdf.

Shah, Manisha, and Bryce Millett Steinberg. “Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital.” *Journal of Political Economy*, vol. 125, no. 2, 2017, pp. 527–561., doi:10.1086/690828.