Year 1 Pre-Analysis Plan: Impact of Alternate Wetting and Drying on Farm Incomes and Water Savings in Bangladesh

Ujjayant Chakravorty, Manzoor H. Dar, Kyle Emerick

July 21, 2017

Overview

This document describes the analysis to be carried out for the first year of data collected on the project “Impact of Alternate Wetting and Drying on Farm Incomes and Water Savings in Bangladesh”. We start by describing the experimental design and main parameters of interest. The document then describes the general specification that will be employed for estimation of the main average treatment effects. Each of the following sections then describes how each of the key impact parameters will be estimated. Finally, the analysis plan closes with a list of blank regression tables that will be populated after data are analyzed. This analysis plan has been written before any of the follow up data have been analyzed.

The treatment being studied is a water-management technique for irrigated rice called Alternate Wetting and Drying (AWD). AWD involves inserting a perforated PVC pipe into the soil to allow the farmer to observe soil moisture below the surface. The AWD guidelines suggest that the farmer let the field dry until the water level reaches 15 cm below the surface — which has a visible marking within the pipe. Once this water level is reached, the farmer should re-irrigate the field up to a level that depends on the current status of the crop. The process of alternatively wetting and drying the field should be practiced up to the time that the crop starts to flower or reproduce. The farmer should keep sufficient water in the field during flowering because the crop water requirements are much higher during flowering relative to the previous vegetative stage of growth. The farmer also drains the field approximately one to two weeks before harvest, regardless of their chosen method of irrigation. AWD is meant to reduced total irrigation withdraw relative to a system where the field is never allowed to dry, i.e. “continuous flooding”. In addition, agronomic trials on experiment stations have found that AWD reduces the methane emissions from rice relative to continuous flooding.

The study is a randomized control trial being carried out in 400 villages spread across 12 upazilas in 3 divisions of Bangladesh: Rajshahi, Rangpur, and Mymensingh. The first year of the study, which this document pertains to, is the boro (dry) season of 2017. Prior to this season, we identified 10 plots cultivated by different farmers that were adjacent to the tube well in each of these villages. This sample of 4,000 farmers serves as our sample for the estimation of impacts. We first randomly allocated treatment at the village level, where the randomization was stratified by upazila. The treatment was administered prior to the planting of boro rice, which occurred from January-March 2017, depending on the area. The treatment involved three components:

1. Delivery of an AWD pipe to each of the 10 farmers.

2. A short-training carried out by a local NGO on how to use the AWD pipe.

3. Assistance by the NGO with installation. The installation was required to be completed on the study plot identified close to the tube well, i.e. the plot identified in the beginning of the experiment.

Farmers in the control group were not provided with any of the above items 1-3.

The nature of water pricing varies across our sample. This is an important feature of the experiment because we expect the water-pricing regime might be important for the effectiveness of AWD. Water pricing varies mostly — but not entirely — across divisions. Rajshahi has metered and government-run deep tube wells (DTW) where farmers use prepaid cards to pay for each hour of irrigation. Our sample area of Rangpur also has government DTW, but those tube wells are operated based on seasonal contracts where farmers pay per unit of land area, not per unit of water. Mymensingh has a shallower groundwater table and farmers in this region irrigate rice with private shallow tube wells (STW). In this case the pricing regime depends on the contract between the owner and users of the tube well. In many cases the water is paid for on a seasonal basis and the owner of the tube well pays for the diesel fuel or electricity to lift the water. In some other cases there is a component of volumetric pricing because the farmer pays for the fuel or electricity, or because the tube well owner is part of the sample. We will use the baseline data to separate the sample into farmers that pay some form of volumetric prices and farmers that pay entirely area-based charges.

This phase of the experiment focuses on these important impact parameters:

1. Impact on water usage
2. Impact on methane emission
3. Impact on rice output, other inputs, and overall profitability per acre

General Specification

Our main specification will be

$$y_{ivj} = \alpha_j + \beta Treat_{vj} + \varepsilon_{ivj}, \tag{1}$$

where y_{ivj} is the outcome for farmer i located in village v and upazila j, $Treat_{vj}$ is an indicator for the 200 random treatment villages, and ε_{ivj} is the random error term. Standard errors will be clustered at the village level in all regressions. The term α_j is the upazila (randomization strata) fixed effects. Our main parameter of interest will be β which gives the intention to treat (ITT) effect of AWD on the various outcomes of interest. The sample will consist of the 4000 study plots, 2000 of which had AWD installed and 2,000 of which were identified at the start of the experiment, but were part of the control group and thus did not have AWD installed. We have baseline data for all 4,000 plots.

While our main specification will be that in (1), we will also include appendix tables where control variables are introduced. More specifically, the specification will be

$$y_{ivj} = \alpha_j + \beta Treat_{vj} + \delta X_{ivj} + \varepsilon_{ivj}, \tag{2}$$

where everything remains the same as in (1), except for the inclusion of the vector of controls X_{ivj}. This vector will include age of the farmer, years of education, household size, livestock ownership, landholdings, television ownership, refrigerator ownership, tube well ownership, baseline knowledge of AWD, indicator if the study plot was rented or sharecropped, area of the study plot, indicator for volumetric water pricing, number of crops being grown on the study plot, baseline number of irrigations in the dry season, baseline water cost per acre in the dry season, and baseline revenue per acre in the dry season.

1 Effects on water usage

In this section we discuss how we will estimate effects on water usage using direct field observations from enumerators. We discuss the estimation using self-reported data in Section 3. Each of the 4,000 fields was visited on two randomly chosen days: one during the first half of the growing season and the other during the second. This means we will have 8,000 observations where an enumerator visited the study plot on an unannounced and random day in order to observe the amount of water in the field.[1] Using these data, the main estimate

[1] The days were randomized at the village level so that the enumerator observed all 10 plots on the same day.

will be akin to that in (1), except for we will have two observations per farmer. There will be two dependent variables of interest: the amount of water (cm) in the field on that day, and and indicator variable for whether the field is dry (no water).

In addition to the average effects across the entire sample, it is important to estimate the effect of AWD separately for farmers with and without volumetric pricing. We will introduce this interaction term in the regression

$$y_{ivjd} = \alpha_j + \beta_1 Treat_{vj} + \beta_2 Volume_{ivj} + \beta_3 Treat_{vj} * Volume_{ivj} + \varepsilon_{ivjd}. \tag{3}$$

In this specification $Volume_{ivj}$ is an indicator for farmers with volumetric pricing and we have added a d subscript to note the day of the water-level observation.

In addition to the variation in water pricing, we would expect the difference between treatment and control plots to depend on the time of the growing season. AWD is not supposed to be practiced during flowering and farmers always drain their field during the end of the growing season. We would therefore expect treatment effects to be largest in the pre-flowering period of the growing season. Yet, we do not know the exact time of flowering, and this varies by crop variety and weather conditions. An approximate time for flowering is around 60-80 days after transplanting. We will estimate (3) separately for the period from 0-70 days after transplanting and the period beyond 70 days after transplanting. Since 70 days is an approximation, we will include appendix tables with both 60 and 80 day thresholds.

A simpler way to get at heterogeneity according to time of the growing season is to use the fact that we have randomized the day of the measurements. This allows us to estimate the non-parametric relationship between the two measures of water usage (levels in CM and an indicator for dry fields) separately for treatment and control groups. This approach has the advantage of not requiring us to impose a specific threshold. Rather, if the treatment effect at all varies by the time of the season, we will observe this by estimating non-parametric Fan regressions of the outcome on days after transplanting. These regressions are estimated separately for the treatment and control plots.

Finally, we will evaluate the change in the entire distribution of water levels using quantile regressions. We don't expect effects of AWD throughout the whole distribution. If AWD farmers irrigate their fields to the same level (when they choose to irrigate) then the distributions will slowly converge at the upper deciles. We will estimate a separate quantile regression for each decile, and report results in a graph where the decile is on the horizontal axis and the quantile regression coefficient (and 95% confidence interval) is on the vertical axis. This analysis will be done both for the entire sample, and separately for farmers with and without volumetric pricing.

The regression tables to be estimated are included in Tables 1 to 7.

2 Effects on methane emissions

We selected 104 random villages for measurements of methane fluxes. Within each village, one farmer was further randomly selected for measurement of methane. Trained individuals have taken 10 methane readings for 24 of these farmers (2 farmers per upazila). Three readings have been taken for each of the remaining 80 farmers. All readings were taken on random days. In combination we will have 480 observations to estimate the effect of AWD on methane emissions. The specification will be

$$y_{ivjd} = \alpha_j + \beta Treat_{vj} + \varepsilon_{ivjd}, \tag{4}$$

where in this case y is the methane flux (measured in mg per square meter per hour). We expect some skewness in the distribution of the methane readings given the difficulty of taking gas samples in the field. Our main estimates will trim the top 1.5% of readings. We will also report results on cumulative season emissions where we multiple the hourly flux by the number of hours in a season and the size of the plot (also trimming the top 1.5% of outliers). Standard errors will again be clustered at the village level since we have multiple observations per village.

The methane analysis will also test the same heterogeneity with regard to volumetric pricing. That is, we will introduce the interaction term into Equation 4 just as we did with Equation 3. We will also estimate the same non-parametric fan regressions for methane fluxes as we described for water levels in Section 1.

Adding certain control variables is likely to reduce residual variance and decrease standard errors for methane fluxes. While our main estimates will use only the experimental variation, we will also include an appendix table where we include certain controls in the interest of improving precision. These are:

1. Days after transplanting (DAT), included as a cubic function
2. Temperature
3. Precipitation on the day and day before the reading
4. Use of organic manure
5. Baseline yield
6. Baseline number of irrigations

In the robustness analysis we will also 1) exclude observations where no methane was detected 2) replace the level of the methane flux with the log as the dependent variable 3) check sensitivity to using robust regressions (which place less weight on outlier observations).

All of the regression tables to be estimated are included in Tables 8 to 10.

3 Effects on output, costs, input use, and profits from first-year followup

The follow up survey will be carried out with all 4,000 farmers and it will track production and input use on two plots: the study plot and another randomly selected plot where we collected baseline outcome data. Our main estimates will focus on the study plot and will be generated from a specification like (1). The outcome variables for input use will be:

1. Number of fertilizer applications
2. KG urea per acre
3. KG TSP per acre
4. KG potash per acre
5. KG of other fertilizer per acre
6. Pesticide expenditure per acre
7. Herbicide expenditure per acre
8. Hired labor expenditure for planting
9. Hired labor expenditure for weeding
10. Hired labor expenditure for harvesting
11. Imputed family labor cost for planting (opportunity cost imputed with village average wage)
12. Imputed family labor cost for weeding (opportunity cost imputed with village average wage)
13. Imputed family labor cost for harvesting (opportunity cost imputed with village average wage)

The follow-up survey will also contain self-reported measures of irrigation usage,

14. Number of times irrigation water was applied
15. Number of times field was drained
16. Cost per acre of water

The measures of revenue and profits will be

17. Yield (kilograms of output per acre)

18. Revenue per acre

19. Profit per acre

20. Log yield (kilograms of output per acre)

21. Log revenue per acre

22. Log profit per acre

23. Profit per acre, trim top and bottom 1.5%

24. Profit per acre, robust regression

25. Log profit per acre, trim top and bottom 1.5%

26. Log profit per acre, robust regression

Unlike the other analysis, this analysis includes multiple outcomes that measure different things. We will adjust the p-values using the Family-wise Error Rate in Anderson (2008 Journal of the American Statistical Association p.1486). The families of outcomes will be grouped as follows: chemical inputs (1-7 above), labor inputs (8-13), irrigation (14-16), and profitability (17-19).

Similar to the above analysis, these regressions will be run both with and without an interaction between treatment and the indicator for volumetric pricing. In addition, the analysis will also be repeated for the randomly chosen non-study plot to identify if there are any spillover benefits to other plots of the farmer.

The regression tables to be estimated are included in Tables 11 to 16.

4 Other estimates

4.1 Remote sensing

Our enumerators took the GPS locations for each of the 4,000 study plots during the baseline survey. We will match these locations to remote sensing data from the SENTINEL satellite of the European Space Agency in addition to Landsat using Google Earth Engine. There are two objectives of this exercise.

First, we want to obtain a measure of the "greenness" of each of these plots (throughout time) to measure whether treatment plots appear any more water stressed as a result of AWD. These data will directly complement the yield analysis described above in Section 3. In this case we will have multiple images for each plot, depending on cloudiness. We will take all cloud-free mages in between planting and harvesting to estimate the basic specification

$$y_{ivjd} = \alpha_j + \beta Treat_{vj} + \varepsilon_{ivjd}, \tag{5}$$

where y is the log of either Normalized Difference Vegtation Index (NDVI), Enhanced Vegetation Index (EVI), or Normalized Difference Moisture Index (NDMI) all of which are vegetation indices which are plausibly correlated with plant health. We will also estimate equation 5 more flexibly by estimating the non-parametric relationship between y and days after planting, separately for treatment and control groups. As a final and more parametric approach, we will bin the data into 10-day bins and estimate separate values of β for each 10 day bin, from transplanting to harvesting.

Second, we will test for differences in soil moisture between treatment and control plots, to the extent this is measurable with the imagery. The analysis will proceed identically to the analysis for greenness, except for that we will use either the Normalized Difference Water Index or the soil wetness index.

4.2 Deep driver heterogeneity

Some farmers in our sample are the "deep drivers" for the village tube well. The driver is the person responsible for collecting payments and planning the allocation of water. These individuals have a greater degree of control over water allocation and planning, as has been observed in qualitative field work prior to the study. While Rajshahi and Rangpur have deep drivers for the government-run tube wells, water allocation in Mymensingh is controlled by tube well owners. A plausible hypothesis is that these individuals will respond better to AWD because they have a greater degree of control over when and how long to irrigate.

We will estimate heterogeneous effects where an indicator for drivers or tube well owners is interacted with treatment. This analysis will be carried out for Sections 1, 3, and 4.1 above.

Tables

Table 1: Main effects on water usage

	(1) Level in CM	(2) Dry Field (0/1)	(3) Level in CM	(4) Dry Field (0/1)
AWD Treatment				
AWD Treatment * Volumetric Pricing				
Volumetric Pricing				
Strata Fixed Effects	Yes	Yes	Yes	Yes
Mean in Control				
p-Value: Treat+Treat*Volumetric				
Number of Observations	8000	8000	8000	8000
R squared				

Columns 3 and 4 will include the interaction term and the level effect of volumetric pricing.

Table 2: Effects on water usage, by growth stage of the crop

	0-60 Days After Planting		60+ Days After Planting	
	(1)	(2)	(3)	(4)
	Level	Dry	Level	Dry
AWD Treatment				
Strata Fixed Effects	Yes	Yes	Yes	Yes
Mean in Control				
Number of Observations				
R squared				

Table 3: Effects on water usage, by growth stage of the crop

	0-70 Days After Planting		70+ Days After Planting	
	(1)	(2)	(3)	(4)
	Level	Dry	Level	Dry
AWD Treatment				
Strata Fixed Effects	Yes	Yes	Yes	Yes
Mean in Control				
Number of Observations				
R squared				

Table 4: Effects on water usage, by growth stage of the crop

	0-80 Days After Planting		80+ Days After Planting	
	(1)	(2)	(3)	(4)
	Level	Dry	Level	Dry
AWD Treatment				
Strata Fixed Effects	Yes	Yes	Yes	Yes
Mean in Control				
Number of Observations				
R squared				

Table 5: Effects on water usage, by growth stage of the crop

	0-60 Days After Planting		60+ Days After Planting	
	(1)	(2)	(3)	(4)
	Level	Dry	Level	Dry
AWD Treatment				
AWD Treatment * Volumetric Pricing				
Volumetric Pricing				
Strata Fixed Effects	Yes	Yes	Yes	Yes
Mean in Control				
p-Value: Treat+Treat*Volumetric				
Number of Observations				
R squared				

All columns will include the interaction term and the level effect of volumetric pricing.

Table 6: Effects on water usage, by growth stage of the crop

	0-70 Days After Planting		70+ Days After Planting	
	(1)	(2)	(3)	(4)
	Level	Dry	Level	Dry
AWD Treatment				
AWD Treatment * Volumetric Pricing				
Volumetric Pricing				
Strata Fixed Effects	Yes	Yes	Yes	Yes
Mean in Control				
p-Value: Treat+Treat*Volumetric				
Number of Observations				
R squared				

All columns will include the interaction term and the level effect of volumetric pricing.

Table 7: Effects on water usage, by growth stage of the crop

	0-80 Days After Planting		80+ Days After Planting	
	(1)	(2)	(3)	(4)
	Level	Dry	Level	Dry
AWD Treatment				
AWD Treatment * Volumetric Pricing				
Volumetric Pricing				
Strata Fixed Effects	Yes	Yes	Yes	Yes
Mean in Control				
p-Value: Treat+Treat*Volumetric				
Number of Observations				
R squared				

All columns will include the interaction term and the level effect of volumetric pricing.

Table 8: Main effects on methane emissions

	Hourly Flux (mg/m^2/hr)		Season total (mg)	
	(1)	(2)	(3)	(4)
AWD Treatment				
AWD Treatment * Volumetric Pricing				
Volumetric Pricing				
Strata Fixed Effects	Yes	Yes	Yes	Yes
Mean in Control				
p-Value: Treat+Treat*Volumetric				
Number of Observations	480	480	480	480
R squared				

Columns 2 and 4 will include the interaction term and the level effect of volumetric pricing.

Table 9: Robustness of methane estimates

	(1) Flux	(2) Flux, Robust	(3) Exclude 0's	(4) Log Flux	(5) Flux	(6) Flux, Robust	(7) Exclude 0's	(8) Log Flux
AWD Treatment								
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Mean in Control								
Number of Observations	480	480	480	480	480	480	480	480
R squared								

Table 10: Robustness of methane estimates

	(1) Flux	(2) Flux, Robust	(3) Exclude 0's	(4) Log Flux	(5) Flux	(6) Flux, Robust	(7) Exclude 0's	(8) Log Flux
AWD Treatment								
AWD Treatment * Volumetric Pricing								
Volumetric Pricing								
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Mean in Control								
Number of Observations	480	480	480	480	480	480	480	480
R squared								

All columns will include the interaction term and the level effect of volumetric pricing.

Table 11: Effects on input usage

	Fertilizer					Chemical Expenditures		Hired Labor			Family Labor		
	(1) N apps	(2) Urea	(3) TSP	(4) Potash	(5) Other	(6) Pesticide	(7) Herbicide	(8) Plant	(9) Weed	(10) Harvest	(11) Plant	(12) Weed	(12) Harvest
AWD Treatment													
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control													
Number of Observations	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000
R squared													

Table 12: Effects on self-reported water use

	(1) Number Irrigations	(2) Times Drained	(3) Cost per Acre
AWD Treatment			
Strata Fixed Effects	Yes	Yes	Yes
Mean in Control			
Number of Observations	4000	4000	4000
R squared			

Table 13: Effects on revenues and profits

				Log:		
	(1) Yield	(2) Revenue Per Acre	(3) Profit Per Acre	(4) Yield	(5) Revenue Per Acre	(6) Profit Per Acre
AWD Treatment						
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control						
Number of Observations	4000	4000	4000	4000	4000	4000
R squared						

Table 14: Effects on input usage

	Fertilizer					Chemical Expenditures		Hired Labor			Family Labor		
	(1) N apps	(2) Urea	(3) TSP	(4) Potash	(5) Other	(6) Pesticide	(7) Herbicide	(8) Plant	(9) Weed	(10) Harvest	(11) Plant	(12) Weed	(12) Harvest
AWD Treatment													
AWD Treatment * Volumetric Pricing													
Volumetric Pricing													
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control													
p-Value: Treat+Treat*Volumetric													
Number of Observations	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000
R squared													

All columns will include the interaction term and the level effect of volumetric pricing.

Table 15: Effects on self-reported water use

	(1) Number Irrigations	(2) Times Drained	(3) Cost per Acre
AWD Treatment			
AWD Treatment * Volumetric Pricing			
Volumetric Pricing			
Strata Fixed Effects	Yes	Yes	Yes
Mean in Control			
p-Value: Treat+Treat*Volumetric			
Number of Observations	4000	4000	4000
R squared			

All columns will include the interaction term and the level effect of volumetric pricing.

Table 16: Effects on revenues and profits

				Log:		
	(1) Yield	(2) Revenue Per Acre	(3) Profit Per Acre	(4) Yield	(5) Revenue Per Acre	(6) Profit Per Acre
AWD Treatment						
AWD Treatment * Volumetric Pricing						
Volumetric Pricing						
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control						
p-Value: Treat+Treat*Volumetric						
Number of Observations	4000	4000	4000	4000	4000	4000
R squared						

All columns will include the interaction term and the level effect of volumetric pricing.