Pre-Analysis Plan

This document provides a preliminary outline of econometric models that will be used to estimate the average treatment effect of the provision of social comparison information on groundwater use decisions.

1 Blocking Variables

We utilized several blocking variables to guide the assignment of treatment and control among wells and their owners in the study area. These blocks were determined by using groundwater use data from previous years to determine which factors best predict water use. This process resulted in a decision to block along:

1. $GWMD$ / $County$: A vector of groundwater management district (GWMD) or county dummy variables which constitute the comparison group in the experiment. $GWMD$ is the comparison group for Colorado well owners while $County$ is the comparison group for Kansas mailers.

2. $Q2_i^{owner}$: A dummy variable which = 1 if the i^{th} well's owner's mean pre-experiment water use, across all their wells, is greater than the median pre-experiment well-level water use at the well-level for their GWMD/county. The pre-experiment time period for the Colorado trial is 2018 and the pre-experiment time period for the Kansas trial is 2019.

2 Primary Econometric Models

Given the differences in data and experiments in Colorado and Kansas. We propose separately estimating econometric models for each state.

2.1 Primary Econometric Model: Colorado

Let $w_{i,t}$ represent the water use (acre feet) by the i^{th} well in trail year t ($t = 2019, 2020,\ or\ 2021$). The primary econometric model takes the following form

$$log(w_{i,t}) = \alpha * GWMD + \beta_1 * Q2_i^{owner} + \beta_2 * w_{i,t-1} + \beta_3 * Treated_i + \varepsilon_i \quad (1)$$

where α is a vector estimated coefficients related to GWMD dummy variables, β_1 is a coefficient capturing the effect of the dummy variable $Q2_i^{owner}$, β_2 measures the effect of pumping in time $t-1$ ($w_{i,t}$) on pumping in time t, $Treated_i$ is a dummy variable that indicates whether the owner of the i^{th} well received comparison information, β_3 is the treatment effect of interest, and ε_i is an idiosyncratic error term. Standard errors are clustered at the owner level.

2.2 Primary Econometric Model: Kansas

Let $w_{i,t}$ represent the water use (acre inches per acre irrigated) by the i^{th} well in trail year t ($t = 2020\ or\ 2021$). The primary econometric model takes the following form

$$log(w_{i,t}) = \alpha * County + \beta_1 * Q2_i^{owner} + \beta_2 * w_{i,t-1} + \beta_3 * Treated_i + \varepsilon_i \quad (2)$$

where α is a vector estimated coefficients related to County dummy variables, β_1 is a coefficient capturing the effect of the dummy variable $Q2_i^{owner}$, β_2 measures the effect of pumping in time $t-1$ ($w_{i,t}$) on pumping in time t, $Treated_i$ is a dummy variable that indicates whether the owner of the i^{th} well received comparison information, β_3 is the treatment effect of interest, and ε_i is an idiosyncratic error term. Standard errors are clustered at the owner level.

3 Secondary Econometric Models

This Section present three secondary econometric models, differentiating across states, aiming to uncover the mechanisms explaining agricultural producer's response to the social comparison information. The first model posits that high average use by a well owner across all their wells impacts response to the social comparison information at the well-level. The second model suggests that high use at the well-level influences response. The third model recognizes the panel nature of the pumping data and introduces both well and time fixed effects.

3.1 Model 1: Colorado

As in the primary model, let $w_{i,t}$ represent the water use (acre feet) by the i^{th} well in trail year t ($t = 2019, 2020, \ or \ 2021$). The model takes the following form

$$\begin{aligned} log(w_{i,t}) =& \alpha * GWMD + \beta_1 * Q2_i^{owner} + \beta_2 * w_{i,t-1} + \beta_3 * Treated_i + \beta_4 * (Treated_i \ X \ Q2_i^{owner}) \\ &+ \beta_5 * Well \ Capacity_i + \beta_6 * \# \ of \ Wells_i + \varepsilon_i \end{aligned} \tag{3}$$

where β_4 measures the treatment effect conditional on the i^{th} well's owner's mean water use in $t-1$, across all their wells, being greater than their GWMD's median pre-experiment water use at the well-level. β_5 measures the impact of well capacity, which constitutes a flow constraint imposed by local aquifer conditions which limits the volume of water that can be pumped within a given unit of time (e.g. $\frac{gal.}{min}$). β_6 measures the impact of the total number of wells owned by the owner of the i^{th} well.

3.2 Model 1: Kansas

As in the primary model, let $w_{i,t}$ represent the water use (acre inches per acre irrigated) by the i^{th} well in trail year t ($t = 2020 \ or \ 2021$). The model takes the following form

$$\begin{aligned} log(w_{i,t}) =& \alpha * County + \beta_1 * Q2_i^{owner} + \beta_2 * w_{i,t-1} + \beta_3 * Treated_i + \beta_4 * (Treated_i \ X \ Q2_i^{owner}) \\ &+ \beta_5 * Well \ Capacity_i + \beta_6 * \# \ of \ Wells_i + \varepsilon_i \end{aligned} \tag{4}$$

where β_4 measures the treatment effect conditional on the i^{th} well's owner's mean water use in $t-1$, across all their wells, being greater than their County's median pre-experiment water use at the well-level. β_5 measures the impact of well capacity, which constitutes a flow constraint imposed by local aquifer conditions which limits the volume of water that can be pumped within a given unit of time (e.g. $\frac{gal.}{min}$). β_6 measures the impact of the total number of wells owned by the owner of the i^{th} well.

3.3 Model 2: Colorado

As in the primary model, let $w_{i,t}$ represent the water use (acre feet) by the i^{th} well in trail year t ($t = 2019, 2020, \ or \ 2021$). The model takes the following form

$$\begin{aligned} log(w_{i,t}) =& \alpha * GWMD + \beta_1 * Q2_i^{owner} + +\beta_2 * w_{i,t-1} + \beta_3 * Treated_i + \beta_4 * (Treated_i \ X \ Q2_i^{owner}) \\ &+ \beta_5 * Well \ Capacity_{i,t} + \beta_6 * \# \ of \ Wells_i + \beta_7 * \gamma_i + \beta_8 * (\gamma_i \ X \ Treated_i) + \varepsilon_i \end{aligned} \tag{5}$$

where γ_i is a dummy variable which $= 1$ when the i^{th} well's pre-experiment water use was greater than their comparison group's median well-level water use and β_7 is a coefficient measuring the dummy variable's effect. β_8 measures the effect of higher than comparison group median pre-experiment water use conditional on inclusion in the treatment group.

3.4 Model 2: Kansas

As in the primary model, let $w_{i,t}$ represent the water use (acre inches per acre irrigated) by the i^{th} well in trail year t ($t = 2020$ *or* 2021). The model takes the following form

$$\begin{aligned} log(w_{i,t}) =& \alpha * County + \beta_1 * Q2_i^{owner} + +\beta_2 * w_{i,t-1} + \beta_3 * Treated_i + \beta_4 * (Treated_i \ X \ Q2_i^{owner}) \\ &+ \beta_5 * Well \ Capacity_{i,t} + \beta_6 * \# \ of \ Wells_i + \beta_7 * \gamma_i + \beta_8 * (\gamma_i \ X \ Treated_i) + \varepsilon_i \end{aligned} \tag{6}$$

where γ_i is a dummy variable which = 1 when the i^{th} well's pre-experiment water use was greater than their comparison group's median well-level water use and β_7 is a coefficient measuring the dummy variable's effect. β_8 measures the effect of higher than comparison group median pre-experiment water use conditional on inclusion in the treatment group.

3.5 Model 3: Colorado

Let $w_{i,t}$ represent the water use (acre feet) by the i^{th} well in year t ($t = 2011 \ldots 2021$). The panel econometric model takes the following form

$$log(w_{i,t}) = \theta_i + \delta_t + \alpha * GWMD + \beta_1 * Q2_i^{owner} + \beta_3 * Treated_{i,t} + \varepsilon_{i,t} \tag{7}$$

where θ_i is a well-level fixed effect, δ_t is a year fixed effect, α is a vector estimated coefficients related to GWMD dummy variables, β_1 is a coefficient capturing the effect of the dummy variable $Q2_i^{owner}$, $Treated_{i,t}$ is a dummy variable that indicates whether the owner of the i^{th} well received comparison information in time t, β_3 is the treatment effect of interest, and ε_i is an idiosyncratic error term. Standard errors are clustered at the well level.

3.6 Model 3: Kansas

Let $w_{i,t}$ represent the water use (acre inches per acre irrigation) by the i^{th} well in year t ($t = 2000 \ldots 2021$). The panel econometric model takes the following form

$$log(w_{i,t}) = \theta_i + \delta_t + \alpha * County + \beta_1 * Q2_i^{owner} + \beta_3 * Treated_{i,t} + \varepsilon_{i,t} \tag{8}$$

where θ_i is a well-level fixed effect, δ_t is a year fixed effect, α is a vector estimated coefficients related to County dummy variables, β_1 is a coefficient capturing the effect of the dummy variable $Q2_i^{owner}$, $Treated_{i,t}$ is a dummy variable that indicates whether the owner of the i^{th} well received comparison information in time t, β_3 is the treatment effect of interest, and ε_i is an idiosyncratic error term. Standard errors are clustered at the well level.

4 Data Inclusion Rules

Outliers: We excluded all wells that use less than 5% of their GWMD's 5^{th} percentile from the experiment. As such, these low consumption wells will not be included in our analysis of the experimental data.

Attrition: Any wells whose owner was involved in the social comparison mailer but does not report pumping in 2019 while reporting in 2018 will be dropped from the analysis.

Noncompliance: All wells associated with a mailer that was returned due to incorrect mailing address provided by the State of Colorado or Kansas will be dropped from the analysis. An exception to this rule is if a well owner's wells used ranked above the 5th percentile in the initial mailing and then less than the 5th percentile in subsequent mailings.

Duplicate Addresses: Despite the best efforts of the research team to identify and associate duplicate addresses within both Colorado and Kansas data, several duplicate addresses exist in both state's trails. A common example involves abbreviations e.g. Aaron's Farm LLC., address 123 County Road 4, Yuma, CO, owns wells 1 and 2 while Aaron's Other Farm LLC., address 123 C.R. 4, Yuma, CO, owns wells 3 and 4. Since C.R. is an abbreviation for County Road, both LLCs have the same address and should be considered the same entity. Duplicate addresses affect our analysis in several ways:

1. Over treatment i.e. sending multiple mailers (treatments) to the same entity, if both duplicate address were assigned to treatment
2. Over control i.e. duplicate addresses (well owners) were both assigned to the control group
3. Contamination of control, if one duplicate address was assigned to control the other to treatment groups

To address the issues related to duplicate addresses we plan to drop all the well owners that were assigned to both treatment and control. As a secondary check, we will also drop all over treatment and over control groups.