The Economic Benefits of Irrigation Districts under Prior Appropriation Doctrine: An Econometric Analysis of Agricultural Land-allocation Decisions*

(First Draft)

Xinde "James" Ji† and Kelly M. Cobourn ‡

January 18, 2017

*This draft: January 18, 2017. Please do not cite or circulate without permission from the authors.

†Graduate Research Assistant, Virginia Tech
‡Assistant Professor, Virginia Tech

1 Introduction

In most Western states in the US, water use is governed by the prior appropriation doctrine, which allocates water based on the principle of "first in time, first in right". Water rights that are established later in time, or junior rights, will be curtailed during a water shortage to ensure sufficient water to satisfy senior rights. This means that junior water rights holders face greater risk in water availability. Burness and Quirk (1979) showed theoretically that in the absence of water markets, prior appropriation doctrine induces heterogeneous risk among water users, which leads to an inefficient allocation of water. Recent studies have empirically tested this hypothesis that risk heterogeneity in water availability drive wedges in the behavior of otherwise similar farmers, and have found generally supportive results. When facing water constraints, farmers usually respond by adjusting along the extensive margin (Hornbeck and Keskin, 2014). As a result, farmers with junior water rights tend to plant more land to drought-tolerant but low-profit crops than otherwise similar farmers holding senior water rights, which results in 5-10% loss in land rent Xu et al. (2014b); Cobourn et al. (2016); Brent (2016).

Given that the prior appropriation doctrine is not likely to be altered in the near future, are there ways to potentially mitigate this risk in water availability? Previous studies have looked at two potential channels of risk mitigation: access to additional water sources (i.e., a water portfolio), and establishing water markets. For the water access channel, Hornbeck and Keskin (2014) showed that access to groundwater supply expand farmers' extensive margin in the long run, and as a result farmers are able to switch to water intensive crops, which would ultimately lead to an increase in farmland value. Mukherjee and Schwabe (2015) demonstrated that in addition to groundwater, access to supplementary water from water districts also increases farmland value. For the water market channel, Burness and Quirk (1979) showed that in theory a competitive water transfer market is sufficient to achieve the efficient allocation of water, and the empirical efficiency gains from these markets are huge (e.g., Calatrava and Garrido, 2005; Ghosh et al., 2014). However, both channels are constrained by the current geographical and institutional environment. Access to groundwater is subject to the geographical distribution of underground aquafers, and large-scale water transfer markets are still very thin in the western U.S (Brewer et al. 2007; Bretsen and Hill 2006). In this paper, we present another channel that mitigates the risk of water availability - irrigation districts.

Irrigation districts are semi-governmental farmer cooperatives that allocate irrigation water acquired under prior appropriation water rights to their members. They differ from private farmers in three ways. First, they typically hold a broader portfolio of water rights that spans different seniorities, which reduces water supply volatility. Secondly, water is proportionally allocated inside most irrigation districts. Proportional allocation of water spreads the risk from individuals to all district members analogous to insurance pools, which reduces the chance of critical curtailment for every individual member. Lastly, water transfers between district members are less costly, and less likely to be opposed by a third party. The intra-district water transfer mechanism, which is effectively a small-scale water market, allows water allocation within the districts to reflect the marginal value of water use, and provides an efficiency gain over strict application of prior appropriation doctrine at the level of the individual water user. Because of the reasons above, members of irrigation districts will bear less risk of curtailment, and thus be able to expand their production towards more water-intensive crops.

The goal of this study is to empirically test whether there exist systematic differences in land-allocation decisions between farmers residing in and out of irrigation districts, in additional to their water rights ownership. Our empirical analysis focuses on the East Snake River Plain (ESRP) in Idaho, a major agricultural production region that relies heavily on irrigation. We take advantage of a series of rich, geo-referenced datasets, including a unique geospatial database of water rights for the state and the high-resolution Cropland Data Layer (CDL) to identify cropping choices made by each farm entity in the region. A panel dataset is constructed spanning 2007-2014, which includes water rights, crop choice, and control variables such as soil characteristics, temperature and precipitation. We use a fractional multinomial logit model to explain the land allocation decisions by farmers, which help us capture the influences of irrigation districts on land allocation.

This study contributes to the existing literature in two ways. First and foremost, we are able to disentangle the economic benefits provided by irrigation district from the effect of water rights seniority. This bridges the gap in the existing literature, which have provided estimates separately on the seniority effect and the irrigation district effect. Specifically, existing studies on the seniority effect used variations in seniorities either at the individual farmers' level (e.g., Cobourn et al., 2016; Xu et al., 2014b), which ignores the economic benefits of irrigation districts; or at the irrigation district level (Brent, 2016), which underestimates the seniority effect as a whole. Study on the economic value of districts (Mukherjee and Schwabe, 2015) did not consider the effect of seniority in determining water availability. This leads to an overestimation of the influence by irrigation district under the prior appropriation doctrine.

Secondly, we use a direct approach to model farmers' crop-specific land allocation decisions, which allows us to disentangle the land allocation dynamics as a function of different natural and institutional characteristics. Previous studies examine these differences indirectly through either hedonic approaches (e.g., in Brent, 2016; Schlenker et al., 2007), or at an aggregate level (e.g. Hornbeck and Keskin, 2014; Deschenes and Greenstone, 2007).

Our empirical results show that farmers in irrigation districts plant land to a more profitable set of crops than otherwise similar farmers outside districts. On average, irrigation districts grow more sugarbeets and potatoes, which are more drought-sensitive, higher-value crops. As a result of these differences in planting decisions, members of irrigation districts earn on average $16.20 per acre, or 6.0% more per year than those outside of irrigation districts. This is comparable to about 0.15 AF/acre in water delivery found in Buck et al. (2014), two-standard-deviation increase in water rights seniority found in Xu et al. (2014b), Brent (2016) and Cobourn et al. (2016), and larger than the benefit of having access to water districts in Mukherjee and Schwabe (2015). Our result indicates that there can be huge potential efficiency gains by deviating from strict application of prior appropriation doctrine on individuals.

The rest of the paper is organized as follows. Section 2 discuss the background on irrigation districts, Section 3 motivates the empirical strategy, Section 4 describes the data, and Section 5 provides the empirical results.

2 Background on Irrigation Districts

Most irrigation districts were originally established in the early 1900s, and the main purpose at that time were to exploit the natural monopoly of irrigation water supply. Irrigation districts facilitated the construction of water infrastructures such as pipes and canals, which exhibited high fixed infrastructure cost and increasing return to scales of water supply(Michelsen et al., 1999). They greatly reduced bargaining and transaction cost between irrigators who share the infrastructure at that time, and was regarded as an institutional innovation that speeded up the process of settling and development in the US West (Rosen and Sexton, 1993).

Irrigation districts still hold an important position in the current water appropriation arena. About a quarter of the irrigated areas of the U.S. West rely on irrigation districts; this number can be as much as one-half in some states like California (Smith, 1989). Like other special districts in the United States, irrigation districts are defined by fixed geographical boundaries. Any farmer who resides in an irrigation district is considered to be a member of the district, and is entitled to district water supply. Irrigation districts still need to hold prior appropriation water rights through each state's water appropriation system as individual farmers do. However, comparing to individual farmers, water rights held by irrigation districts are usually larger in quantity, more diverse in the spans of seniority, and are used to serve a collection of district members.

When irrigation districts face curtailment, water request will be awarded proportionally to their members according to their requests. Given that irrigation districts typically hold large amounts of water spanning across diverse water portfolios, the probability is very small for an irrigation district to be critically or completely curtailed. As a result of this risk alleviation, a farmer inside an irrigation district is able to plant a more water-intensive crop mix if the farmer has a concave production function (Burness and Quirk, 1979; Cobourn et al., 2016) or is risk-averse (Calatrava and Garrido, 2005; Li et al., 2016), which leads to economic gains.

Some irrigation districts also establish informal water transfer mechanisms between its members, typically held at the district office. These transfers are much less likely to be objected since water transfer between district members ensures that any non-consumptive water stays in the hydrological system. The potential efficiency gain from agricultural water transfers are huge, even if the market operates only locally (e.g., Ghosh et al., 2014; Calatrava and Garrido, 2005). However, outside irrigation districts, the presence of these markets are very limited under the current political and legal environment (see, e.g., Brewer et al. 2007; Bretsen and Hill 2006). Some states including Idaho have set up water banking programs to facilitate water transfers. But those banks typically have limited scope, and often place extra limitations on the sources and the geographical location that a transfer can take place (Clifford et al., 2004; Ghosh et al., 2014). Even if water transfer between two parties can be completed, the negotiation process are usually very costly, and often faces third-party objections.[1] In contrast, members of irrigation districts have easy access to a water-market with relatively small transaction cost. This small-scale water transfer market can thus achieve efficiency gains for district members.

Those two traits of irrigation districts corresponds to the two potential solutions that Burness and Quirk (1979) suggested in their seminal paper. First, if farmers' production functions are homogeneous, then equal sharing of all available water is Pareto optimal. Secondly, the establishment of a competitive water transfer market achieves Pareto optimal even if water is initially allocated under the prior appropriation doctrine, and farmers' production functions are heterogeneous. However, for individual farmers, both of these two solutions are currently constrained by the legal and political environment. The economic advantage of irrigation districts over individual farmers comes exactly from their ability to implement both solutions within districts, where farmers outside irrigation districts fail to achieve. By deviating from strict applications of prior appropriation doctrine at the individual level, as well as by offering risk pooling, irrigation districts are able to reduce the heterogeneous risk in water availability.

As a result of the risk alleviation, we hypothesize that farmers inside irrigation districts will plant more high-value, water intensive crops in their land allocation then otherwise similar farmers outside of irrigation districts. By doing so, we intentionally leave out farmers' adjustments in irrigation intensity from our model. In the context of agricultural production further West, both irrigation districts status and the prior appropriation hierarchy remains constant at least for the past 50 years. Thus, any variations in agricultural production cause by institutional differences in water availability will be reflected in the long-run rather than short-run response. In theory, when facing water availability constraints farmers can choose to adapt along the intensive margin or along the extensive margin. However, by decreasing irrigation intensity, farmers maintain the

[1] Private communication with IDWR.

current cropping pattern but become more susceptible to droughts and curtailments, which is ultimately a short-term coping method. In contrast, by changing their land allocation patterns towards drought-tolerant crops or plant less land as a whole, farmers become less sensitive to droughts and curtailments by trading off maximum profitability (Hornbeck and Keskin, 2014). Thus, intensive margin adjustments reflects the short-run responses, and should have little influence to farmers' long-run reaction to institutional differences. Thus, following Hornbeck and Keskin (2014),Xu et al. (2014b) and Cobourn et al. (2016), we focus on modelling the relationship between agricultural land allocation differences and water institutions.

Our analysis on the role of irrigation district is done under the assumption that water is allocated only between agricultural users. It is worth noting that a large set of economic literature (e.g. Bretsen and Hill, 2006; Libecap, 2010, 2011) looked at irrigation districts from the perspective of water allocation for multiple use, and especially, the conflict of interest between irrigation districts and urban water users. Those studies typically look at urbanized states like California and Arizona, where the opportunity cost of water is defined by the willingness to pay for stable water supply by the growing urban population, which is much higher than the marginal value of water used for irrigation purposes. The inability for irrigation districts to sell water to domestic users (Rosen and Sexton, 1993; Bretsen and Hill, 2008) or to adopt water-saving measures (Griffin, 2006; Michelsen et al., 1999) will then induce economic losses to the urban water user. The premise in those studies are drastically different from our study, and so does the policy implications. When agricultural interests dominate the local economy, as in Idaho, Montana or West Oregon, the opportunity cost of water used by one farmer is just water used by another farmer. As a result, the collective nature of irrigation districts help them bypass strict application of prior appropriation, and achieve economic benefits.

3 Empirical Strategy

We aim to draw inference from a multi-crop production model which predicts the crop allocation dynamics of farmers based on their water availability, soil characteristics, and price information. Specifically, we are interested in how the prior appropriation water rights places a constraint on farmers' land allocation decisions, and how residing inside an irrigation districts may alleviate that constraint.

We start by writing a generic multi-crop decision function. The fraction of land that farmer i allocate to crop j, y_{ij}, depends on that farmer's expected water availability $\mathbf{W}_i$, soil and climate characteristics $\mathbf{Z}_i$, and the price of input factors and output commodities for crop j as well as other alternative crops, $\mathbf{P}$. We write the problem as:

$$y_{ij} = f(\mathbf{W}_i, \mathbf{Z}_i, \mathbf{P}) \tag{1}$$

The water availability for farmer i, $\mathbf{W}_i$, can be disentangled into several factors: water acquired from the appropriation system W_i^a, water supply from precipitation W_i^p, supplemental groundwater sources W_i^s, whether she is a member of an irrigation district W_i^{ID}, and other factors that affect water availability W_i^o such as extreme heat conditions. We write that as:

$$W_i = f(W_i^a, W_i^s, W_i^p, W_i^{ID}, W_i^o) \tag{2}$$

And specifically, supply from the appropriation system can be further disentangled into three effects: supply effect, seniority effect, and portfolio effect. The supply effect α measures the total available surface water in the appropriation system. More water available in the system means more chance of getting water for everyone. Seniority effect μ measures the relative seniority of a right in the appropriation system. Senior rights are less likely to get curtailed than junior rights.

Portfolio effect σ measures the diversity of water source held by a farm entity. The more diverse the source is, the less likely a farm is going to be critically curtailed. The diversity here includes larger dispersion in surface water rights σ^s (Cobourn et al., 2016), and possibly holding additional groundwater sources σ^g (Mukherjee and Schwabe, 2015). We write it as:

$$W = f\Big(W^a\big(\alpha, \mu, \sigma(\sigma^s, \sigma^g)\big), W^p, W_i^{ID}, W^o\Big) \tag{3}$$

The supply effect α is unobserved in our model, and will be instead controlled by the year fixed effect. Priority effect μ is proxied by the average of water right quantile. Surface water portfolio effect σ^s is portrayed by the standard deviation of the water right quantile.[2] Groundwater sources σ^g is defined as a dummy variable of whether a farm holds any additional groundwater rights.

There are several ways to achieve a reduced-form model from the above characterization. Under the water-intensive versus drought-tolerant crop dichotomy, meaningful economic predictions can be reached with estimable reduce-form equations. These models usually assume that farmers choose to grow two type of crops, the higher-profit but water intensive crop, and the lower-profit but drought tolerant crop. Under this framework, Hornbeck and Keskin (2014) showed that in the long-run, groundwater depletion leads to a shrinkage of extensive margin by irrigating fewer lands and planting more drought-tolerant crops. Xu et al. (2014a) showed the converse argument that an increase in water availability will lead to an increase in land allocated to the water-intensive crops. Cobourn et al. (2016) further strengthened this argument by directly modeling the structure of water rights under the prior appropriation doctrine. They showed that both an increase in water rights seniority and an increase in the diversification of water rights motivate farmers to grow water-intensive crops. Although the water-intensive versus drought-tolerant type of model can not be used alone to produce an estimable multi-crop allocation model, it provides an important

[2]Specifically, define the standard deviation of a single water right to be zero, which can be viewed as a portfolio with no diversification at all.

theoretical construct that farmers adjusting to changing water availability by moving along their extensive margins.

The economic literature has also laid out multi-crop allocation models. For example, Moore and Negri (1992) proposes that under common multi-crop maximization settings, a linearized reduce-form model of multiple crop allocation can be achieved by assuming that the restricted profit function has a normalized quadratic functional form. However, their model does not lead to meaningful economic predictions as to what are the marginal effects of water supply on crop allocation. Another possible way of constructing a multi-crop allocation model is to adopt the conditional logit framework by McFadden (1974). Assume that the underlying profit of farmer i growing crop j on a unit of land can be linearly explained by a set of explanatory variables plus a random error term, i.e.,

$$\Pi_j = \mathbf{X}_i \boldsymbol{\beta}_j + \varepsilon_{ij} \tag{4}$$

It can be shown that if the random error term ε_{ij} follows an i.i.d. type I extreme value distribution, then the probability of farmer i choosing crop j, y_{ij}, has the form:

$$y_{ij} = \frac{exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{k=1}^{J} \mathbf{X}_i \boldsymbol{\beta}_k} \tag{5}$$

If we interpret y_{ij} as the share of crop j in land allocation (rather than the probability of choosing alternative j in a traditional conditional logit framework), then Equation 5 gives rise to the fractional multinomial logit model, which is the focus of this paper.

Fractional models assume that the dependent variables are generated by assigning shares to a series of choices. For each individual, the dependent variables are the shares for each choice categories. Those shares sum up to one, and are bounded by zero and one. There are three major advantages of adopting a fractional multinomial logit framework over traditional univariate linear models. First and foremost, it allows us to capture the dynamics of crop allocation changes with regard to different natural and institutional endowments. We are able to identify systematic differences in the cropping patterns exhibited between farmers inside and outside irrigation districts, with or without additional groundwater supply, etc. Univariate linear models, on the other hand, works like a black box in the sense that the model does not identify or explain any land use difference. Secondly, if the true data generating mechanism (DGM) is fractional, then a traditional linear estimator is mis-specified, which may provide inconsistent estimates as well as poor fits. The linear model will be especially problematic if the dependent variables takes near-boundary values (0 or 1) with non-trivial probabilities. Fractional logit models, on the other hand, correctly specifies the underlying DGM, and should be the preferred model to estimate data with share structures. Lastly, fractional multinomial logit model is able to capture the heterogeneity in partial effects, where in linear model partial effects are assumed to be homogeneous. In our application, it is particular interesting to capture the heterogeneous treatment effects of irrigation districts between farmers who are currently exposed to different level of risks in water supply.

The fractional multinomial logit model is a multivariate extension to the bivariate fractional logit model laid out by Papke and Wooldridge (1996). In that paper, Papke and Wooldridge proposed a quasi-maximum likelihood (QMLE) estimator for the fractional logit problem, along with the correction to achieve consistent standard errors following Gourieroux et al. (1984b,a). Papke and Wooldridge (2008) extended the univariate fractional logit model to a panel setting, and proposed to use either weighted non-linear least squares, QMLE, or a control function approach to estimate the panel fractional logit problem. In the economic literature,[3] Mullahy (2015) extended Papke and Wooldridge (1996, 2008)'s model to a multivariate setting, and demonstrated that QMLE is still consistent when there exists multiple share categories. Mullahy (2015) also showed that the Papke and Wooldridge (1996) standard error correction provides consistent stan-

[3]See, e.g., Sivakumar and Bhat (2002) for the development of fractional multinomial logit in the transportation literature.

dard error estimates for the fractional multinomial logit model. Empirically, fractional multinomial logit model has been widely used in agricultural land allocation modelling (e.g., see Kala et al. 2012; Fiszbein 2015; Cobourn et al. 2016).

The land allocation of farmer *i* for crop *j*, y_{ij}, is measured as the fraction of land allocated to crop *j* in all allocable lands. The right hand side is a linear combination of all explanatory variables with a multinomial logit transformation, with the form:

$$G(x_i\beta_j) = \frac{exp(x_i\beta_j)}{\sum_{k=1}^{J} exp(x_i\beta_k)} \tag{6}$$

of which $z_j = X\beta_j$ represents the right hand side combination for choice j with $j \in \{1, ..., J\}$. The entire model can be consistently estimated with a quasi-maximum likelihood estimator proposed by Papke and Wooldridge (1996), via maximizing the Bernoulli log-likelihood function with the form

$$\sum_{i=1}^{N} ln(L_i) = \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} ln(G(x_i\beta_j)) \tag{7}$$

of which L_i is the likelihood for observation $i \in \{1, ..., N\}$, y_{ij} is the fraction of choice $j \in \{1, ..., J\}$ made by observation i, x_i is a vector of of explanatory variables for observation i, and β_j is the choice-specific vector of coefficients of choice j.

The parameter estimates from the quasi-MLE problem represent the logit-transformed odds ratio for that specific choice against the baseline choice. In order to draw marginal or discrete conclusions, one need to obtain the average partial effects (APE). The marginal effect of the multinomial models has the form:

$$ME_{jk} = \frac{\partial p_j}{x_k} = p_j(\beta_{kj} - \bar{\beta}_i) \tag{8}$$

where p_j is an 1*N vector of predicted probabilities for choice j, and $\bar{\beta}_i = \sum_{m=1}^{J} \beta_{km} p_m$ is the probability weighted average of β_k. The discrete effect for a zero-one dummy variable has the form:

$$DE_{jk} = Pr(y = j|\mathbf{x}_{x_k=1}) - Pr(y = j|\mathbf{x}_{x_k=0}) \tag{9}$$

which is the change of predicted probability in the choice y_j when the dummy variable x_k switch from zero to one. Both the discrete and the marginal effects differ among different levels of right hand side variables, and thus different individuals. This means that the researcher has to aggregate the partial effects for different individuals in order to obtain APE. There are two ways to induce the average partial effect from the individual-heterogeneous partial effects. The partial effects at the mean (PEM) calculates the partial effects by setting all covariates at their sample mean, and use the partial effect at that point to represent the APE. The partial effects on average (PEA) calculates partial effects for every observation, and take the average of that as the APE. In this paper we will use the PEM method when calculating APE, although it should be pointed out that there is no agreement as to which one is preferred (Greene, 2008).

Our baseline fractional multinomial logit model has the following setup:

$$G^{-1}(y_{ijt}) = \alpha_j + \beta_j IrrDist_i + \gamma_j \mu_i + \delta_{1j}\sigma_i^g + \delta_{2j}\sigma_{it}^s + \zeta_j \mathbf{X}_{it} + \theta T_{it} + \varepsilon_{ijt} \tag{10}$$

where $G^{-1}(.)$ is the inverse of the multinomial logit function with $G(.)$ defined in equation 6; y_{ijt} is the fraction of crop j planted by farm i in year t; $IrrDist_i$ is a indicator variable of whether observation i is an irrigation district; μ_i is the mean of water rights quantiles; σ_i^g is a dummy variable indicating whether the farm owns groundwater rights in addition to surface water rights; σ_{it}^s is a measurement of surface water right portfolios; $\mathbf{X}$ is a matrix of control variables, including soil, weather, and price expectations. α is the intercept; and ε is the error term. Time dummies T_{it} are added to control for time-related heterogeneity, e.g., unobserved surface water supply effect.[4] Table 1 provides the description of variables included in our model, and Table 2 provides the

[4]Basin dummy can be added too, but adding them will cause model non-convergence.

summary statistics for them.

We propose to fit a pooled fractional multinomial logit model on a wide panel dataset with T=8. Although we possess a dataset with a panel structure, we do not intend to explicitly control for individual-specific heterogeneities using panel data methods for several reasons. First of all, panel data method for fractional multinomial logit model is not well-developed. Papke and Wooldridge (2008) proposed a control function method to estimate panel fractional probit models. However, there are two obstacles in applying their method in this study. First, the likelihood function for a multimonial probit model can only be meaningfully constructed on binary variables, but not on share variables. This prohibits the extension of panel fractional probit model to become multivariate. Secondly, Papke and Wooldridge's estimator requires within transformation of data, which would eliminate all time-invariant variables in the process. Since most of the variables in our model are time invariant, running a fixed-effect like model is impossible.

Secondly, using pooled cross-section model is sufficient to to explain the long-run impact of irrigation institutions. Our purpose is to draw inference mainly from between rather than within variations in land allocation. Since the institutional factors of water availability, prior appropriation seniority and irrigation districts, are fixed over our model period, any year to year land allocation differences for one specific farmers will not be a result of institutional factors. Rather, within variations mainly reflect crop rotation patterns and the expectation on weather conditions. Thus, using land allocation variations from a cross-sectional or short panel dataset are sufficient to identify the causal impact from water institutions if other determinants of water availability are properly controlled for.

Furthermore, irrigation institutions are exogenous in the sense that there does not exist any reverse causality that land allocation can affect irrigation institutions. Also, we are not worried about endogeneity emerging from sample selection or spatial sorting, which most hedonic models suffer from (Klaiber and Smith, 2013). Agricultural land sales in Idaho are usually accompanied by the selling of the water rights associated with that land parcel. Thus, even the ownership of a specific farm may change, the associated water rights or the claims to irrigation district water will not.

To minimize the impact of omitted variable bias, we control for as many factors as we can, guided by the structural agricultural production function. We have included factors that reflects water availability, soil conditions, and weather conditions that impact agricultural production and yield. Admittedly, there are factors that we are not able to control for because of data limitation, but any differences in endogeously determined inputs in the production such as water storage capacty, irrigation technology or fertilizer should be seen as part of the long-run impact of irrigation districts on farmers' production process. For example, we do not control for irrigation technology in our model, which affects irrigation efficiency and water requirements. For this to bias our result, irrigation districts have to use systematically inefficient technology than individual farmers. Furthermore, this systematically inefficient technology should not be a result of the reduced risk exposure. We have no reason to believe either of them are true. That is to say, our result is rather robust against potential omitted variable problems.

4 Data

Our empirical analysis focuses on the East Snake River Plain (ESRP) in Southeast Idaho (see Figure 1). The ESRP is a major agricultural production region in the intermountain West, and the region's economy is highly dependent on the agricultural sector. The main water source of the region is Snake River and its tributaries, and its water flows depend highly on winter precipitation and snow melts. Agricultural production in the region heavily relies on irrigation water: 74.7%

of farmlands are irrigated (NASS, 2012) in the region, which consumes 85.6% of all waters in the ESRP (Kenny et al., 2009). About 60% of the irrigated croplands are serviced by surface water and including surface water storage facilities. The other 40% are serviced by groundwater pumping (NASS, 2014).

[Insert Figure 1 here.]

One of the most important reasons that we focus on the ESRP is that Idaho maintains a spatially explicit water rights database provided by the Idaho Department of Water Resources (IDWR).[5] From this database we are able to identify the spatial boundaries of water rights entities, which we assume to be the boundary of farms. Additionally, we acquire the water right titles, source of water, and priority dates associated with the farm. This allows us to conduct analysis at the individual water rights level. And, more importantly, knowing the spatial boundaries of the farm entities allows us to match it with other spatially referenced dataset such as land use, soil and weather.

We are able to identify 6429 unique water rights residing in ESRP. Among them, 1679 farms hold at least one surface water rights, and 15 of them are irrigation districts. Figure 2 shows the geographical locations of irrigation district in the ESRP. We exclude those who hold only groundwater titles. Groundwater users do not face curtailment risk from the appropriation system, which make their water supply much more reliable than surface water users. This means that groundwater users are likely to behave systematically different from surface water users. Since our goal is to find valid counterfactual for farms residing inside surface water irrigation districts, groundwater users should not be included in our sample. We also exclude all observations that have areas less than 1 acre. Although the cutoff point is rather arbitrary, water rights with areas smaller than a certain threshold cannot be possibly identified as farmlands for major crops.

[5] Available from https://research.idwr.idaho.gov

For irrigation districts, we are only able to identify the water rights boundaries for each district, which usually coincide with their administrative boundaries. However, we are not able to distinguish property boundaries for individual farms inside irrigation districts. This means that land allocation decisions for irrigation districts are the measured at the aggregate level as the weighted mean of all farms residing in that district. This is a limitation of our dataset, and we provide robustness checks to show that this caveat does not undermine our main result. We also need to assume that every parcel of farmland inside the boundary of an irrigation district use district water, unless the parcel has access to other water sources. This assumption is widely used in previous literature when intra-district water delivery data is not available (Schlenker et al., 2007; Buck et al., 2014).

[Insert Table 1 here.]

[Insert Table 2 here.]

To better capture the marginal effect of priority effect on water availability, we perform a standardized rank transformation on water rights priority dates. Specifically, all surface water rights are ranked by their priority dates from the earliest to the latest, and are standardized to a zero-one range, of which we call the "quantile" of a right. The first water right in the system will be assigned a quantile of 0, the median water right of 0.5, and the last right of 1.

[Insert Table 2 here.]

The reason for this transformation is that the distribution of priority dates are not uniform in time. As shown in Figure 3, most surface water rights are filed during the progressive era, and fewer rights are filed after 1930. This means that a one-year seniority in the progressive era will represent a much larger increase in the rank of priority than a one-year seniority in the 1990s. So if we put priority dates directly into a linearized regression model, the marginal effect of a one-year seniority change will be heterogeneous across time. A rank transformation, on the contrary, guarantees that the quantile of each right is uniformly distributed in the appropriation system, and the marginal effect on a one-percentage quantile change becomes more homogeneous.

[Insert Table 3 here.]

We also acknowledge that the quantile of a water rights, through expressed in a percentage form, does not equal to the actual probability of curtailment. The link between water rights quantile and the actual probability of getting curtailed is non-linear, and to capture that relationship sophisticated hydrological-statistical models are required. The rank transformation thus serves as a second-best alternative, an improvement from using priority dates, or date range dummies (as in Xu et al. (2014b); Brent (2016)).

We obtain land allocation data from National Agricultural Statistics Service (NASS)'s Cropland Data Layer (CDL). CDL is a crop-specific land cover dataset for the continental US based on satellite imagery and calibrated classification algorithms (National Agricultural Statistics Service, 2014). It provides a moderate resolution imagery that classifies agricultural land use types. The dataset is available for year 2005, and from year 2007 onwards. For each farm entity, we are able to identify the percentage of land allocated to six major crops of the region: alfalfa, barley, corn, potato, sugarbeet and wheat, as well as land idlement.

Soil data is obtained from the SSURGO database, a soil database developed by USDA-NRCS. The SSURGO dataset contains a crop-specific yield estimate for each soil type, and from which we construct an average irrigated crop yield map for wheat and corn. This allows us to capture the possibility that a parcel of land is especially suitable for certain crops but not for others, which may explain some of the empirical cropping choices. We also include common soil quality indicators in our model, such as irrigated and non-irrigated soil capacity class, percent of clay, percent of slopes, and the k-factor.

Weather data is obtained from the PRISM climate dataset developed by Oregon State University, which provide small-scale climate maps and estimates. We believe that three weather indicators are important in determining crop productivity and water availability: growing degree days(GDD), extreme weather conditions, and growing season cumulative precipitation. We are not arguing that our list of climate variables are exhaustive. Rather, we believe that this three variables are representative to the expectation of short term crop growth and water demand.

Growing degree days is a non-linear transformation of temperature, which assumes that plant growth is linear only between moderate temperature ranges from 8°Cto 32°C(Ritchie and NeSmith, 1991). The use of growing degree days is common in estimating agroeconomic models (e.g. Schlenker et al., 2007; Deschenes and Greenstone, 2007), and is suggested by literature as a preferred method than using monthly average temperature (Schlenker et al., 2007). Extreme heat conditions[6] are detrimental to crop growth, and will significantly reduce crop yield (Burke and Emerick, 2016) Extreme heat conditions also contribute to increasing rates of plant evapotransporation, which cause increase water demands for crops as a result. Growing season cumulative precipitation[7] measures the supplemental water supply provided by precipitation process, which offsets demands for irrigation water.

[Insert Table 1 here.]

[Insert Table 2 here.]

[6]Here defined as daily maximum temperature exceeds 35°C

[7]Defined as cumulative precipitation between June.1 and Sep.30

5 Results

We begin by our "full model", which includes surface water and groundwater portfolio indicators, sets of control variables, and year dummy. Estimation results are shown in Table 3, and the average partial effects (APE) are shown in Table 4. Papke and Wooldridge (1996)'s robust standard errors are reported in Table 3, and are used to generate standard errors for the APEs via Krinsky-Robb simulations.

[Insert Table 3 here.]

[Insert Table 4 here.]

We also calculated the average partial effects on profits (APEP), shown in the first column of Table 5. APEP is analogous to the concept of the traditional parameter estimates and standard errors in a linear model. We do this by aggregating crop shares with respect to their profits per acre, as well as their respective standard errors, i.e.,

$$E(APEP_k) = \sum_{j=1}^{J} APE_{j,k} * profit_j$$

and

$$V(APEP_k) = \sum_{j=1}^{J} V(APE_{j,k}) * profit_j^2$$

where $APEP_k$ is the average partial effect on profits for explanatory variable k, and $APE_{j,k}$ is the average partial effect of crop shares for crop j, explanatory variable k. Here we assume that the standard error of each crop-specific APE is independent of each other, and thus the variance of APEP is the sum of the variances of all crop-specific APEs times the square of their respective profits.

[Insert Table 5 here.]

The average crop-specific profit statistics is calculated by subtracting the average cost from the average revenue between 2005 and 2013, which is shown in Figure 4. Revenue for a crop in a given year is calculated by multiplying the price received with the average yield of that crop in the state of Idaho, which is provided by USDA-NASS (National Agricultural Statistics Service, 2012). Cropping cost statistics are compiled from different sources, including USDA economic research services and the University of Idaho's crop costs and returns series.

[Insert Figure 4 here.]

Our model shows that irrigation districts allocate significantly more land to potato, sugarbeet, and wheat; less land to alfalfa and corn. This confirms our hypothesis that irrigation districts generally plant more water-intensive crops and less drought-tolerant crops, with the exception of wheat and corn. A farmer residing in an irrigation district will have an edge of a \$16.20 per acre, or 6.0% in profits, comparing to an otherwise equal farmer residing outside of an irrigation district. Holding additional groundwater rights significantly help farmers. Comparing with surface water only farmers, farmers who hold both groundwater and surface water rights on average allocate more crop to corn, potato, sugarbeet, and fallows more, less land to alfalfa and barley. These systematic differences leads to a \$31.23 per acre, or 11.5% advantage if a farmer or an irrigation district holds groundwater rights.

Both seniority variables in our model, the mean and the dispersion of water rights quantiles, have insignificant APEs in the model. This result differs from other studies, such as Xu et al. (2014b); Cobourn et al. (2016). However, this result is probably a reflection of the low statistical power that our model exhibits, and should not be interpreted as a nullification of the Burness and Quirk (1979) hypothesis. Other factors that have significant impacts on farm profits including expected growing degree days, expected precipitation, soil yield capacities for wheat and corn, and the soil k-factor. The sign of all these variables are as expected: farmers are more profitable when they expect warmer weather and more precipitation, as well as if they hold lands with more productive soils.

Heterogeneity in Partial Effects

One of the advantages of using fractional multinomial logit models is that it can capture the heterogeneities in partial effects among different observations. As Papke and Wooldridge (2008) pointed out, the difference between linear and non-linear models is not important with regard to the estimation results of average partial effects, but is important in determining whether and to what extent are the partial effects differ at different percentiles of the distribution of the variable in interest. In light of that, we calculate partial effects for the irrigation district dummy at different quantiles of water rights seniorities. Figure 5 shows the discrete effect of irrigation districts along the distribution of water rights seniority quantiles, holding all other variables at their average values. Result shows that the largest benefit by residing in an irrigation district happens for the most junior water right holders, at about 17.94(6.08)$/acre, while the lowest benefit happens for the most senior water rights holders, at 14.37(5.95)$/acre. This indicates that the treatment effect of irrigation districts are modestly larger for farmers holding more junior rights than those holding senior rights.

[Insert Table 5 here.]

Both the risk-sharing and the water transfer mechanism provided by irrigation districts provide theoretical support to the increasing return from irrigation districts over farmers holding junior rights. A senior right holder may find herself reluctant to join an irrigation district since her water right is already secure enough against supply volatility that additional risk-sharing mechanism provided by irrigation district may not provide much gain for her. On the contrary, a junior right holder may find herself in need of hedging against water volatility, and irrigation districts provide precisely that mechanism. Thus we would expect that irrigation districts provide more benefit to junior than senior water right holders. Similar logic also applies to the intra-district water transfer mechanism. A senior right holder has the option to grow a less water-intensive crop mix, and sell the additional water rights for monetary compensation. A junior right holder, on the other hand, has the option to grow a more water-intensive crop mix by buying additional water rights. This leads to the observation that junior rights holders can grow a more profitable crop mix than their endowments allowed to, and thus can have larger observed profits by joining an irrigation district.

APE in Linear vs. Logit Models

As discussed earlier, panel fractional multionomial logit model is not available for the purpose of this study. To check for whether individual heterogeneities may potentially bias our estimates, we propose to estimate two linear models, the pooled ordinary least square (OLS) and the panel random effect (RE) models . In doing so, we assume that all regressors are exogenous from the random unobserved individual effects as well as the idiosyncratic error term. This assumption cannot be formally tested using a Hausman-type test against the fixed effect (FE) model since our main variable of interest are time-invariant. However, this assumption can be justified on two grounds. First, the variables used in our model are all exogenously determined in the sense that any cropping choice made by the farmer will not alter their water rights, irrigation district status, soil quality, weather patterns, or the price for each crop. Secondly, we try to control for as many factors in farmers' production processes as the data permits.

Column 2 and 3 of Table 5 present the results from OLS and RE estimates. Although a Hausman test rejects the hypothesis that OLS and RE are equivalent, the point estimates for the two models are very close, especially in our main variable of interest. Other than *QMeanSurf*, which is statistically insignificant, the difference in point estimates between OLS and RE for water rights and irrigation district variables are less than 2%. This suggests that the inference of our main variables of interest should not be affected by ignoring individual-specific heterogeneities.

Furthermore, point estimates for the two variables that are significant in fractional multinomial logit, *IrrDist* and *GrndSurf*, are very close to that in the two linear models. The effect of residing in a irrigation district is 16.2 \$/acre in fractional multinomial logit, 15.62 \$/acre in OLS, and 15.68\$/acre in RE. The effect of holding additional groundwater rights is 31.2\$/acre in fractional multinomial logit, 34.91 and 33.94 in OLS and RE. The observation here echoes Papke and Wooldridge (2008), which in their case the fractional probit APEs are very close to that in linear models. This gives additional assurance that our point estimates on the effect of irrigation district is robust against different functional form specifications.

Aggregation of Irrigation Districts

Also, we need to address the problem that farms inside an irrigation district are measured at the aggregate level, whereas farms outside the districts are measured at the individual level. This means that the measured land allocation made by irrigation districts are essentially a weighted mean of the individual farmers residing inside that district. This is acceptable as long as land allocation with respect to farm size is homogeneous, i.e.,

$$E(\mathbf{y}|\mathbf{X}, A) = E(\mathbf{y}|\mathbf{X})$$

where A is the size of the farm, $\mathbf{y}$ is the land allocation vector, and $\mathbf{X}$ is all explanatory variables other than farm size. In this case, the expectation of the aggregated land allocation is the same as if they were the mean of each farm, i.e.,

$$E_D(\mathbf{y}|\mathbf{X}, A_D) = \frac{\sum_{d=1}^{D} A_d E_d(\mathbf{y}|\mathbf{X}, A_d)}{\sum_{d=1}^{D} A_d} = E(\mathbf{y}|\mathbf{X})$$

where D is the aggregated measurement of land allocation, and $d \in [1, D]$ is the farms being aggregated.

We empirically test the area-homogeneity assumption by running an augmented regression model and see how the area of the farm may influence farmer's crop allocation. To do so, we use a subsample that only contains farmers residing outside of an irrigation district, and runs two fractional multinomial logit models: one with the area of the farm, and the other with the natural log of farm area.

[Insert Table 6 here.]

The APE and APEP on the area variables are shown in Table 6. The model with *log(Area)* shows that an increase in log area decreases land allocation in alfalfa, and increases in all crop types other than barley. When aggregating these land allocation changes out, the marginal profit change due to log(Area) is statistically not significant at the 5% level. The model with *Area* depicts a similar picture, with a negative APE on alfalfa, and positive APE on all types other than barley and fallow. The margianl profit change is significant at the 5% level, indicating that controlling for all other factors such as water rights, soil and weather, an increase of one acre in farm acreage leads to an increase in profit of about \$0.008 per acre. To put that in perspective, if irrigation districts had no premium, than the average observed farm size inside an irrigation districts would have to be 2090 acres larger than the average size of an individual farm. There are only 3% of the farms in our sample that meets this cutoff. If we assume that the empirical distribution of farm size is similar between farms inside and outside irrigation districts, than it is highly unlikely that the the farm area is the main factor that drives the premium from residing in an irrigation district.

Furthermore, the aggregation process of irrigation districts will result in an over-representation of dryland crops because of the nature of our definition of farms. For individual farmers, we observe the spatial boundary of their water rights, and treat those as farm boundaries. It is likely that the water rights boundaries are smaller than the actual farm boundaries. Lands that are owned by that farm, but not covered by a water source, will practice dryland agriculture through all years (private communication with IDWR). These portions of lands will then be excluded from our sample. This is not the case with irrigation districts. Water rights boundaries of irrigation districts usually line up with their administrative boundaries. Thus, all lands inside an irrigation district will be aggregated, including those that practice dryland agriculture throughout, and potentially be excluded in the case of individual farmers. Thus, the spatial boundaries of irrigation districts will over-represent dryland agriculture, which will result in an underestimation of their premiums.

References

Brent, D. W. (2016). The value of heterogeneous property rights and the costs of water volatility. *American Journal of Agricultural Economics*.

Bretsen, S. N. and Hill, P. J. (2006). Irrigation institutions in the american west. *UCLA Journal of Environmental Law and Policy*, 25:283.

Bretsen, S. N. and Hill, P. J. (2008). Water markets as a tragedy of the anticommons. *William and Mary Environmental Law and Policy Review*, 33:723.

Brewer, J., Glennon, R., Ker, A., and Libecap, G. D. (2007). Water markets in the west: prices, trading, and contractual forms. *NBER Working Paper No. 13002*.

Buck, S., Auffhammer, M., and Sunding, D. (2014). Land markets and the value of water: hedonic analysis using repeat sales of farmland. *American Journal of Agricultural Economics*, page aau013.

Burke, M. and Emerick, K. (2016). Adaptation to climate change: Evidence from us agriculture. *American Economic Journal: Economic Policy*, 8(3):106–40.

Burness, H. S. and Quirk, J. P. (1979). Appropriative water rights and the efficient allocation of resources. *The American Economic Review*, pages 25–37.

Calatrava, J. and Garrido, A. (2005). Modelling water markets under uncertain water supply. *European review of agricultural economics*, 32(2):119–142.

Clifford, P., Landry, C. J., and Larsen-Hayden, A. (2004). *Analysis of water banks in the western states*. Washington State Department of Ecology.

Cobourn, K. M., Ji, X., Mooney, S., and Crescenti, N. F. (2016). Surface water rights and land allocation decisions: A fractional multinomial logit analysis using field scale remote sensing data. *Unpublished Manuscript*.

Deschenes, O. and Greenstone, M. (2007). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *The American Economic Review*, 97(1):354–385.

Fiszbein, M. (2015). Agricultural diversity, structural change and long-run development: Evidence from us counties. *Unpublished Manuscript*.

Ghosh, S., Cobourn, K. M., and Elbakidze, L. (2014). Water banking, conjunctive administration, and drought: The interaction of water markets and prior appropriation in southeastern idaho. *Water Resources Research*, 50(8):6927–6949.

Gourieroux, C., Monfort, A., and Trognon, A. (1984a). Pseudo maximum likelihood methods: Applications to poisson models. *Econometrica*, 52(3):701–720.

Gourieroux, C., Monfort, A., and Trognon, A. (1984b). Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700.

Greene, W. (2008). *Econometric Analysis*. Pearson / Prentice Hall, 6th edition edition.

Griffin, R. C. (2006). Achieving water use efficiency in irrigation districts. *Journal of water resources planning and management*, 132(6):434–442.

Hornbeck, R. and Keskin, P. (2014). The historically evolving impact of the ogallala aquifer: Agricultural adaptation to groundwater and drought. *American Economic Journal: Applied Economics*, 6(1):190–219.

Kala, N., Kurukulasuriya, P., and Mendelsohn, R. (2012). The impact of climate change on agro-ecological zones: evidence from africa. *Environment and Development Economics*, 17(06):663–687.

Kenny, J. F., Barber, N. L., Hutson, S. S., Linsey, K. S., Lovelace, J. K., and Maupin, M. A. (2009). *Estimated use of water in the United States in 2005*. US Geological Survey Reston, VA.

Klaiber, H. A. and Smith, V. K. (2013). Quasi experiments, hedonic models, and estimating trade-offs for local amenities. *Land Economics*, 89(3):413–431.

Li, M., Xu, W., and Rosegrant, M. W. (2016). Irrigation, risk aversion, and water rights under water supply uncertainty. *Agricultural and Applied Economics Association*, (235753).

Libecap, G. D. (2010). Water rights and markets in the us semi arid west: Efficiency and equity issues.

Libecap, G. D. (2011). Institutional path dependence in climate adaptation: Coman's Òsome unsettled problems of irrigationÓ. *The American Economic Review*, 101(1):64–80.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in econometrics*, pages 105–143.

Michelsen, A. M., Taylor, R. G., Huffaker, R. G., and McGuckin, J. T. (1999). Emerging agricultural water conservation price incentives. *Journal of Agricultural and Resource Economics*, pages 222–238.

Moore, M. R. and Negri, D. H. (1992). A multicrop production model of irrigated agriculture, applied to water allocation policy of the bureau of reclamation. *Journal of Agricultural and Resource Economics*, pages 29–43.

Mukherjee, M. and Schwabe, K. (2015). Irrigated agricultural adaptation to water and climate variability: The economic value of a water portfolio. *American Journal of Agricultural Economics*, 97(3):809–832.

Mullahy, J. (2015). Multivariate fractional regression estimation of econometric share models. *Journal of Econometric Methods*, 4(1):71–100.

National Agricultural Statistics Service (2007-2014). USDA National Agricultural Statistics Service Cropland Data Layer. Published crop-specific data layer. Available at $https://nassgeodata.gmu.edu/CropScape/$.

National Agricultural Statistics Service (2012). Census of agriculture, full report. Available at $https://www.agcensus.usda.gov/Publications/2012$.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632.

Papke, L. E. and Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1):121–133.

Ritchie, J. and NeSmith, D. (1991). Temperature and crop development. *Modeling plant and soil systems*, pages 5–29.

Rosen, M. D. and Sexton, R. J. (1993). Irrigation districts and water markets: an application of cooperative decision-making theory. *Land Economics*, pages 39–53.

Schlenker, W., Hanemann, W. M., and Fisher, A. C. (2007). Water Availability, Degree Days, and the Potential Impact of Climate Change on Irrigated Agriculture in California. *Climatic Change*, 81(1):19–38.

Sivakumar, A. and Bhat, C. (2002). Fractional split-distribution model for statewide commodity-flow analysis. *Transportation Research Record: Journal of the Transportation Research Board*, (1790):80–88.

Smith, R. T. (1989). Water transfers, irrigation districts, and the compensation problem. *Journal of Policy Analysis and Management*, 8(3):446–465.

Xu, W., Lowe, S., and Zhang, S. (2014a). An analysis of irrigated agricultural outcomes under the prior appropriation doctrine: hypotheses and applications. *Applied Economics*, 46(22):2639–2652.

Xu, W., Lowe, S. E., and Adams, R. M. (2014b). Climate change, water rights, and water supply: The case of irrigated agriculture in Idaho. *Water Resources Research*, 50(12):9675–9695.

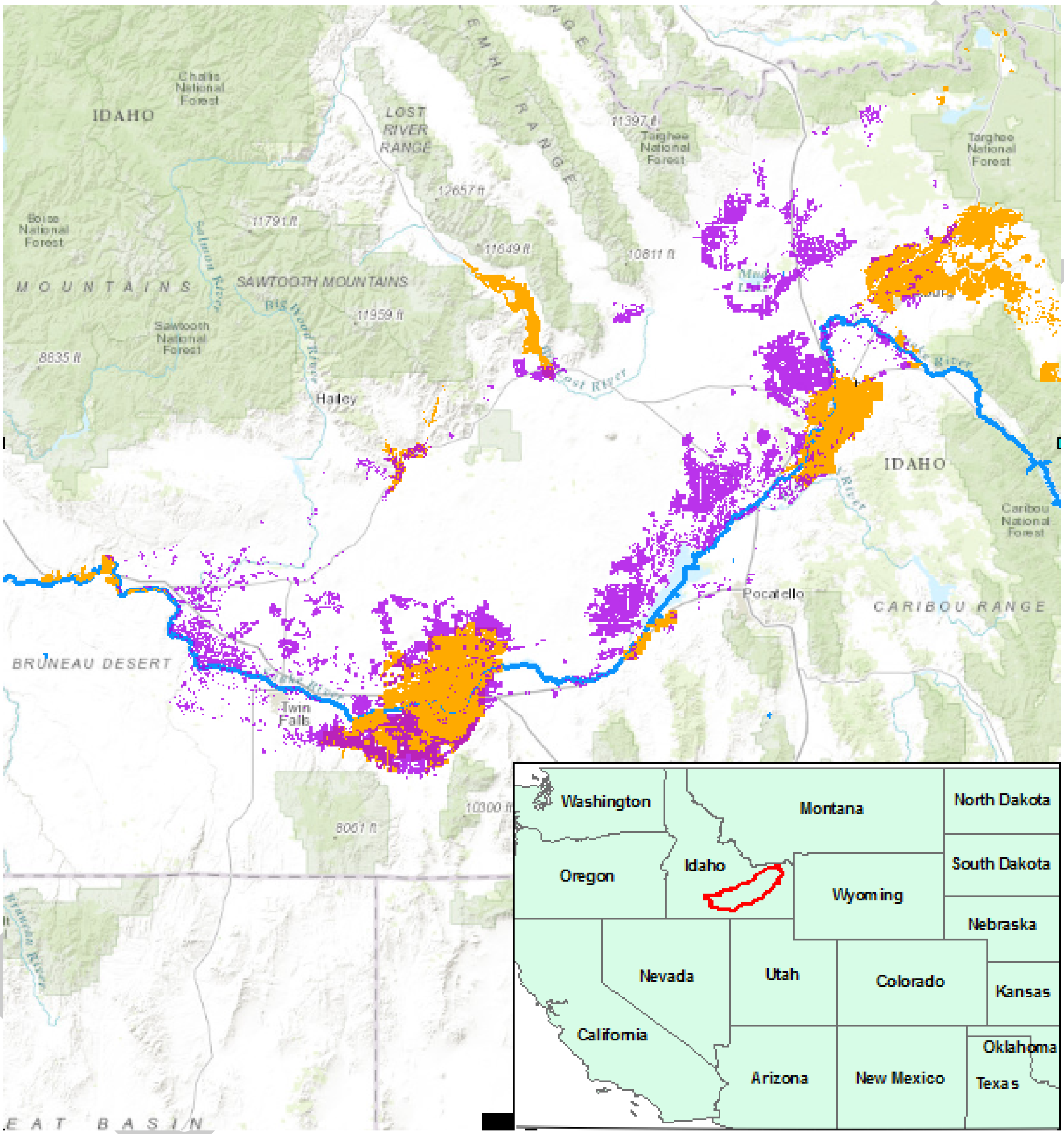

Figure 1: Map of the Eastern Snake River Plain. The blue line denotes the main stem of the Snake River. Purple areas are individual water rights, orange areas are irrigation district lands. Lower-right panel denotes the relative location of the ESRP (Red line denotes the watershed boundary of ESRP.)

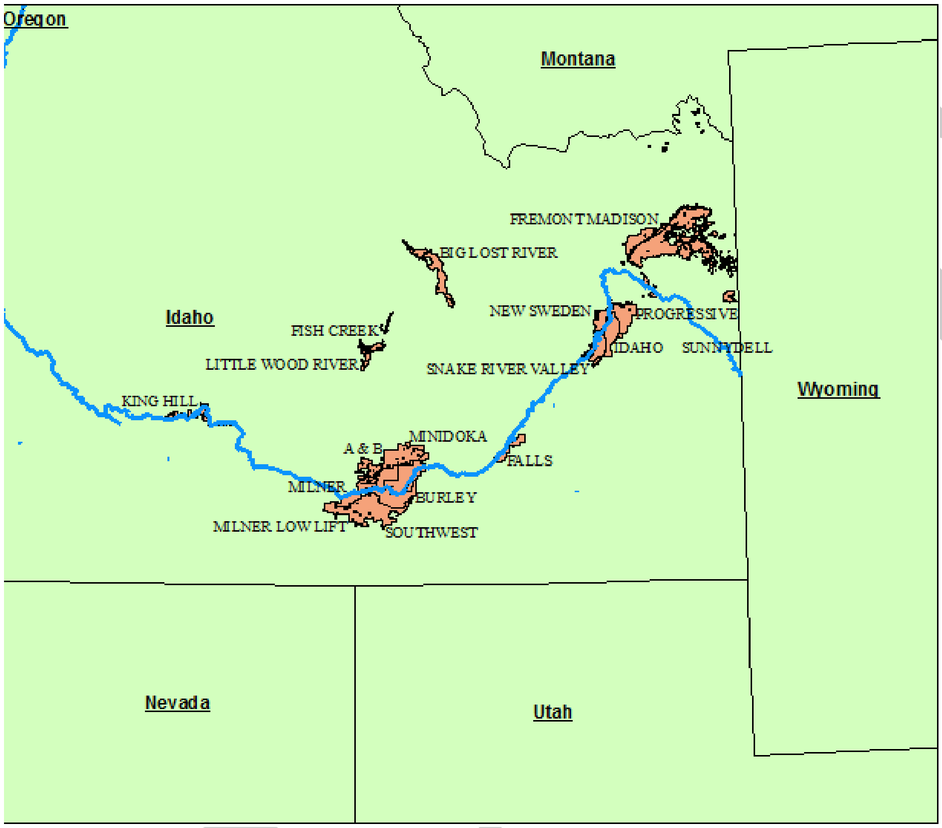

Figure 2: Irrigation Districts in the East Snake River Plain. Blue line is the Snake River. Captions are the names of the respective irrigation districts.

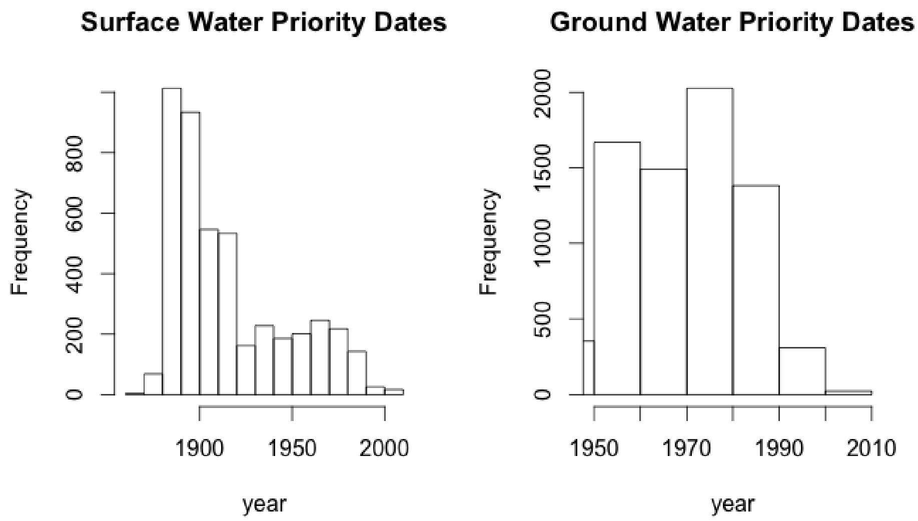

Figure 3: Water Rights Distribution Across Time

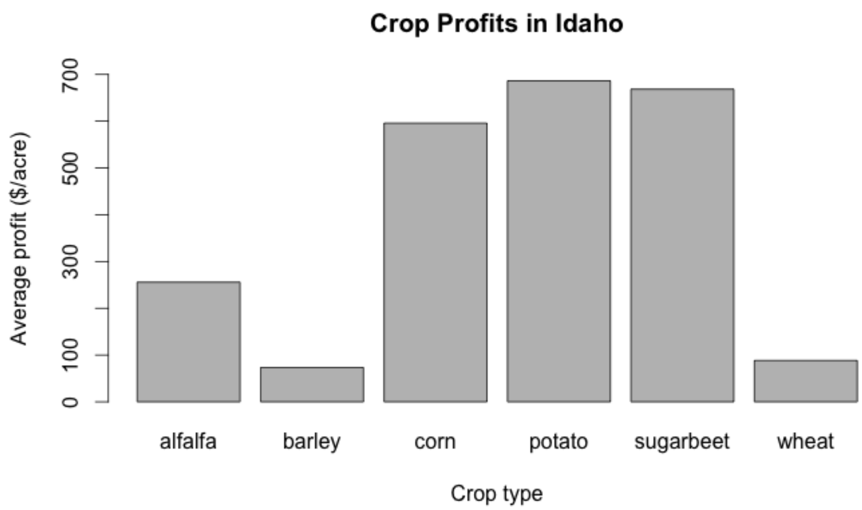

Figure 4: Average crop profits in Idaho, 2005-2013

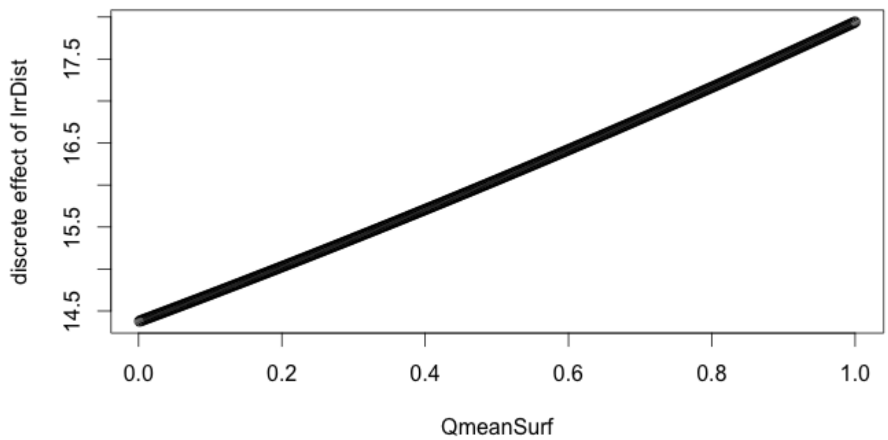

Figure 5: Partial effects of *IrrDist* at different levels of water rights quantile. The x-axis shows the distribution of water rights quantile, with 0 being the most senior, and 1 being the most junior water right. The y-axis shows the monetary value ($/acre) of the discrete effect of residing in an irrigation district.

Table 1: Description of Variable Names and Sources

Variable Name	Variable Description	Unit	Source
Area	area of individual farms	Acres	
DistArea	area of irrigation districts	Acres	
IrrDist	irrigation district dummy		IDWR
GrndSurf	groundwater dummy		
QmeanSurf	mean of water rights seniority quantile		
QsdSurf	standard deviation of water rights seniority quantile		
corn	fraction of corn planted		
wheat	fraction of wheat planted		
barley	fraction of barley planted		
alfalfa	fraction of alfalfa planted		USDA CDL
sugarbeet	fraction of sugarbeet planted		
potato	fraction of potato planted		
fallow	fraction of land fallowed		
exml3	average number of extreme heat days in last 3 years	days	
gddl3	average number of growing degree days in last 3 years	degree days	PRISM
precl3	average total summer precipitation in last 3 years	mm*100	
icclass	irrigated soil capacity class		
nicclass	non-irrigated soil capacity class		
slope	average slope of land		
ydwheat	average yield factor for wheat	bu/hectare	SSURGO
ydcorn	average yield factor for corn	bu/hectare	
claypc	percentage of clay in soil		
kfactor	soil k-factor		
pbarley	average normalized price for barley in the last year		
pcorn	average normalized price for cornin the last year		
pwheat	average normalized price for wheat in the last year		USDA NASS
psugarbeet	average normalized price for sugarbeet in the last year		
ppotato	average normalized price for potato in the last year		

Table 2: Summary Statistics of Variables. Number of observation N=7792.

Variable Name	Mean	Median	Min	Max	Std Dev
Area	206.701	94.689	2.340	6530.334	447.33
Dist_Area	37227.763	29642.514	3802.455	98166.152	28673.725
IrrDist	0.015	0	0	1	0.123
GrndSurf	0.103	0	0	1	0.304
QmeanSurf	0.54	0.546	0.001	1	0.28
QsdSurf	0.041	0	0	0.458	0.084
corn	0.144	0	0	1	0.286
wheat	0.119	0.005	0	1	0.232
barley	0.146	0.007	0	1	0.268
alfalfa	0.47	0.429	0	1	0.393
sugarbeet	0.019	0	0	1	0.105
potato	0.053	0	0	1	0.166
fallow	0.048	0	0	1	0.148
exml3	12.031	4.708	0	94.923	16.107
gddl3	1497.274	1482.925	860.119	2004.202	256.3
precl3	52.273	49.368	15.762	148.713	24.314
icclass	3.393	3.109	2	6	0.684
nicclass	5.557	6	3	6	0.938
slope	2.748	2.025	1	15.818	2.155
ydwheat	78.213	80	30	120	21.299
ydcorn	65.862	60	40	149.876	26.994
claypc	12.07	11.667	1.5	42.254	7.509
kfactor	0.271	0.254	0.02	0.57	0.13
pbarley	3.183	3.244	2.548	4.616	0.62
pcorn	3.271	3.469	2.149	4.15	0.616
pwheat	3.923	4.132	3.188	4.728	0.573
psugarbeet	30.306	29.593	20.725	45.118	8.208
ppotato	4.532	4.016	3.557	6.378	0.993

Table 3: Fractional multinomial logit parameter estimates

Variables	barley	corn	potato	sugarbeet	wheat	fallow
IrrDist	0.0499	-0.107	0.53***	1.63***	0.443***	0.167
	(0.0828)	(0.0869)	(0.0779)	(0.128)	(0.0591)	(0.136)
QmeanSurf	-0.099	-0.139	0.00372	0.271	0.143	0.325*
	(0.0964)	(0.112)	(0.146)	(0.266)	(0.1)	(0.154)
QsdSurf	0.341	-0.116	-0.049	-1.41	0.23	0.527
	(0.256)	(0.362)	(0.41)	(0.824)	(0.264)	(0.494)
GrndSurf	-0.247***	0.626***	0.356**	0.705***	-0.015	0.492***
	(0.0733)	(0.0854)	(0.111)	(0.163)	(0.0753)	(0.115)
pcorn	-0.964**	0.769	2.24***	-0.942	-0.148	-6.82***
	(0.373)	(0.449)	(0.542)	(0.991)	(0.395)	(0.784)
pbarley	0.0817	-0.347***	-0.302**	-0.213	-0.208***	-0.575***
	(0.0661)	(0.0719)	(0.102)	(0.155)	(0.0614)	(0.108)
pwheat	0.263	-0.113	-0.826*	0.325	0.896**	4.82***
	(0.266)	(0.297)	(0.374)	(0.672)	(0.279)	(0.597)
psugarbeet	0.0863**	-0.0204	-0.236***	0.113	0.00119	0.634***
	(0.0326)	(0.0392)	(0.0471)	(0.0859)	(0.0342)	(0.0685)
ppotato	-0.353	0.00678	1.14***	-0.386	-0.205	-3.78***
	(0.191)	(0.221)	(0.268)	(0.494)	(0.202)	(0.43)
exml3	-0.0248***	-0.013***	0.00759	-0.0544***	-0.00764*	0.00345
	(0.00491)	(0.00354)	(0.0048)	(0.00978)	(0.00353)	(0.0043)
gddl3	-0.00407***	0.00422***	0.000943	0.00215*	0.000528	-0.00474***
	(0.000308)	(0.000446)	(0.000534)	(0.000961)	(0.000315)	(0.00053)
precl3	-0.0218***	0.00259	0.0216***	-0.0211*	0.0159***	-0.0593***
	(0.00232)	(0.00375)	(0.00404)	(0.00877)	(0.0027)	(0.00443)
icclass	-0.268***	0.0305	-0.344***	-0.876***	-0.42***	-0.112
	(0.0554)	(0.0579)	(0.088)	(0.222)	(0.0644)	(0.0695)
nicclass	0.19***	0.593***	0.385***	1.25***	0.496***	0.766***
	(0.0338)	(0.0762)	(0.0695)	(0.164)	(0.0429)	(0.0678)
slope	0.0376	0.0371	0.0969**	0.217***	0.115***	0.154***
	(0.0202)	(0.0193)	(0.0295)	(0.0599)	(0.0229)	(0.0194)
ydwheat	0.000569	0.0187***	-0.0139***	-0.0121	-0.00646***	-0.0187***
	(0.00175)	(0.00313)	(0.00297)	(0.00638)	(0.00195)	(0.00248)
ydcorn	-0.00276	0.00633***	-0.00497	0.00645	-0.00197	-0.0144***
	(0.00178)	(0.00157)	(0.00276)	(0.00472)	(0.00205)	(0.00248)
claypc	-0.0452***	-0.0412***	-0.0689***	-0.02	-0.0225***	-0.0457***
	(0.00437)	(0.0078)	(0.00927)	(0.0141)	(0.00578)	(0.00588)
kfactor	3.78***	2.19***	5.82***	5.33***	3***	1.22***
	(0.219)	(0.371)	(0.382)	(0.651)	(0.269)	(0.301)
Number of Obs: 7792						
Log pseudo-likelihood: -10525.69						

Note: Papke and Wooldridge (1996)'s robust standard error reported in parenthesis. Alfalfa is the baseline choice and thus omitted. Year dummy and constant are suppressed from the table. A triple asterisk indicates $p < 0.001$; a double asterisk indicates $p < 0.01$; a single asterisk indicates $p < 0.05$.

Table 4: Average partial effects of fractional multinomial logit estimates

Variable	alfalfa	barley	corn	potato	sugarbeet	wheat	fallow
IrrDist	-0.0792***	-0.0126	-0.0161**	0.0233***	0.0406***	0.0433***	0.000669
	(0.0149)	(0.00962)	(0.0058)	(0.00379)	(0.00164)	(0.00692)	(0.00532)
GrndSurf	-0.0438***	-0.0379***	0.0483***	0.0156*	0.01***	-0.0117	0.0193**
	(0.0127)	(0.0111)	(0.00778)	(0.00771)	(0.00289)	(0.0112)	(0.00614)
QmeanSurf	-0.00672	-0.0146	-0.0107	-0.000413	0.0032	0.0168	0.0124
	(0.0169)	(0.0143)	(0.0104)	(0.00961)	(0.004)	(0.0134)	(0.00683)
QsdSurf	-0.0382	0.0361	-0.013	-0.0059	-0.0182	0.0208	0.0182
	(0.0528)	(0.0302)	(0.0248)	(0.0205)	(0.0111)	(0.0298)	(0.0192)
pcorn	0.147***	-0.0929***	0.0726***	0.127***	-0.00842***	0.0144***	-0.26
	(2.64e-09)	(1.71e-08)	(5.02e-11)	(5.68e-13)	(7.19e-07)	(1.07e-09)	(0.486)
pbarley	0.046***	0.0215***	-0.0188***	-0.0112***	-0.00163***	-0.0162***	-0.0196***
	(0.00157)	(0.000503)	(0.00121)	(0.00105)	(0.000291)	(0.00121)	(0.00191)
pwheat	-0.167***	-0.00423***	-0.0287***	-0.0568***	0.000378	0.077***	0.179
	(2.12e-08)	(9.64e-07)	(6.51e-07)	(1.36e-07)	(0.000875)	(1.51e-05)	(0.447)
psugarbeet	-0.014***	0.00813***	-0.00318***	-0.0132***	0.00109	-0.003***	0.0242
	(2.08e-09)	(5.05e-07)	(5.6e-08)	(2.73e-10)	(0.000742)	(3.75e-08)	(0.0567)
ppotato	0.0957***	-0.0242***	0.0124***	0.0666***	-0.00268***	-0.00469***	-0.143
	(1.29e-07)	(1.31e-07)	(1.69e-08)	(1.09e-10)	(2.9e-07)	(6.84e-08)	(0.233)
exml3	0.00302***	-0.00257***	-0.000544	0.000655**	-0.000605*	-0.000296	0.000348
	(0.000566)	(0.000744)	(0.000297)	(0.000244)	(0.000262)	(0.000473)	(0.000181)
gddl3	0.000161***	-0.000499***	0.000318***	6.23e-05***	3e-05***	0.000104***	-0.000177
	(1.39e-06)	(8.98e-05)	(3.11e-10)	(4.35e-08)	(5.26e-09)	(1.09e-07)	(0.000115)
precl3	0.00123***	-0.00259***	0.000336***	0.00121***	-0.000233***	0.00231***	-0.00226***
	(0.000153)	(9.61e-05)	(2.18e-05)	(6.07e-06)	(2.97e-05)	(1.42e-05)	(0.000313)
icclass	0.0678***	-0.0195*	0.0106***	-0.0114	-0.00933	-0.0384**	3e-04
	(0.0097)	(0.00887)	(0.00277)	(0.00908)	(0.0334)	(0.0126)	(0.00311)
nicclass	-0.111***	-0.00073	0.028	0.00958	0.013***	0.0385*	0.0226*
	(0.000215)	(0.00453)	(0.0327)	(0.03)	(0.00149)	(0.0167)	(0.0115)
slope	-0.0204***	0.000209	7.84e-05	0.00309	0.00224**	0.0101**	0.00468***
	(0.00273)	(0.00353)	(0.00208)	(0.00214)	(0.000857)	(0.00344)	(0.000974)
ydwheat	0.000583***	0.000211***	0.00139***	-0.000652***	-0.000137***	-0.000696***	-0.000701***
	(8.38e-05)	(1.52e-05)	(3.25e-06)	(3.27e-05)	(1.92e-05)	(2.84e-05)	(2.93e-05)
ydcorn	0.000518**	-0.000244***	0.000511***	-0.000206***	9.09e-05***	-0.000135***	-0.000536***
	(0.000191)	(2.69e-05)	(6.19e-06)	(1.75e-05)	(5.54e-06)	(2.93e-05)	(2.08e-05)
claypc	0.00981***	-0.00367***	-0.00169***	-0.00262***	-3.36e-05	-0.000671	-0.00113***
	(0.000936)	(0.000352)	(0.000343)	(0.000387)	(9.32e-05)	(0.000376)	(0.00015)
kfactor	-0.82***	0.307***	0.0524**	0.222***	0.0479***	0.199***	-0.00883
	(0.0521)	(0.0148)	(0.0192)	(0.00583)	(0.0027)	(0.0209)	(0.0112)

Note: Robust standard error reported in parenthesis, calculated via the Krinsky-Robb method. Discrete effects are reported for binary variables *IrrDist* and *GrndSurf*. Marginal effects are reported for all other variables. Year dummies and constant are suppressed from reporting. A triple asterisk indicates $p < 0.001$; a double asterisk indicates $p < 0.01$; a single asterisk indicates $p < 0.05$.

Table 5: Model estimates of average partial effect on profits ($/acre)

	Models		
	(1)	(2)	(3)
Variables	FML	OLS	RE
IrrDist	16.2(5.98)**	15.62***(4.373)	15.68*(9.082)
GrndSurf	31.2(8.06)***	34.91***(5.439)	33.94***(10.47)
QmeanSurf	-5.8(10.4)	-2.095(6.373)	-5.492(10.52)
QsdSurf	-29.2(25.5)	-47.15***(17.43)	-48.79(30.76)
pcorn	157(0.000444)***	81.46***(12.09)	80.87***(10.13)
pbarley	-8.07(1.12)***	-5.971*(3.147)	-5.466*(2.805)
pwheat	-92.1(0.786)***	-67.59***(9.737)	-71.13***(8.184)
psugarbeet	-13.5(0.52)***	-6.719***(1.270)	-6.278***(1.050)
ppotato	73.6(0.000349)***	42.44***(8.114)	44.13***(6.625)
exml3	0.278(0.34)	0.167(0.216)	0.629**(0.261)
gddl3	0.266(0.00666)***	0.235***(0.0207)	0.159***(0.0292)
precl3	1.2(0.0459)***	0.922***(0.148)	0.251*(0.151)
icclass	4.78(23)	7.819**(3.337)	8.935(5.943)
nicclass	6.92(28.9)	-3.344(2.245)	-1.401(3.727)
slope	-0.646(2.18)	-0.796(1.233)	-0.890(1.989)
ydwheat	0.394(0.0344)***	0.787***(0.136)	0.920***(0.220)
ydcorn	0.327(0.05)***	0.630***(0.114)	0.415**(0.162)
claypc	-0.641(0.421)	-1.345***(0.332)	-1.766***(0.536)
kfactor	46.1(18.3)*	76.25***(18.23)	84.61***(28.60)

Note: Column 1 shows the average partial effect on profits derived from the fractional multinomial logit (FML) model. Column 2 and 3 show linear estimates of farm profits using pooled ordinary least square (OLS) and panel random effect (RE) models. Robust standard errors reported in parenthesis. A triple asterisk indicates $p < 0.001$; a double asterisk indicates $p < 0.01$; a single asterisk indicates $p < 0.05$.

Table 6: Fractional multinomial logit model estimates on farm area variables

APE: Crop Type(%)	(1) logArea	(2) Area
alfalfa	-0.0522(0.00267)***	-8.1e-05(9.91e-06)***
barley	0.00162(0.00526)	6.74e-06(5.02e-06)
corn	0.0137(0.00254)***	1.73e-05(2.53e-06)***
potato	0.00909(0.00232)***	1.58e-05(1.36e-06)***
sugarbeet	0.00266(0.000881)**	5.04e-06(5.95e-07)***
wheat	0.0192(0.00419)***	3.93e-05(3.45e-06)***
fallow	0.00601(0.00229)**	-3.14e-06(2.78e-06)
APEP: profit ($/acre)		
profit	4.604(2.464)	0.00772(0.00315)*

Note: Model (1) includes the natural log of farm area (in acres) as an explanatory variable, and model (2) includes the level of farm area. All other control variables except the irrigation district dummy are included. Robust standard errors reported in parenthesis. A triple asterisk indicates $p < 0.001$; a double asterisk indicates $p < 0.01$; a single asterisk indicates $p < 0.05$.