CERE

CERE Working Paper, 2012:11

# Aggregate Technical Efficiency and Water Use in U.S. Agriculture

Andrew L. Zaeske

Centre for Environmental and Resource Economics

SLU and Umeå University

S-901 83 Umeå, Sweden

# Aggregate Technical Efficiency and Water Use in U.S. Agriculture

Andrew L. Zaeske*
Centre for Resource and Environmental Economics
Umeå, Sweden

May 9, 2012

**Abstract**

In the United States freshwater withdrawals for agriculture account for 80% of all out of stream water withdrawals from 1985 to 2005. To assess what drives water use in agriculture, we use the two error stochastic frontier analysis model of Battese and Coelli (1995) to estimate a translog production frontier for agriculture at the state level. The inclusion of non-negative technical inefficiency effects allows us to account for additional characteristics in our assessment of production inefficiency.

The average marginal value of irrigation is $380, while we find that on average $1 of intermediate inputs provides $0.96 of final output. These results are driven by a small subset of states with large negative values, indicating persistent misallocation of resources. The inefficiency effects regression finds that government subsidies increase in value of output of 0.083 per real dollar of subsidies and that that shifts from larger acreage farms to lower acreage ones will generally be efficiency increasing.

This analysis highlights differences in water use and how they can have major implications for farm policy as a whole. Of particular note is the measured positive correlation between having a negative marginal product of intermediates and having a positive marginal product of irrigation, which suggests that shifts in inputs from intermediates to irrigation are a ripe target for efficiency gains in many states.

*Contact E-mail: andrew.zaeske@slu.se

# 1 Introduction

Water withdrawals in the United States are primarily driven by use for irrigation. Freshwater withdrawals for agriculture account for 80% of all out of stream water withdrawals from 1985 to 2005,[1] with agricultural withdrawals in the western U.S. alone account for over 55% of all such withdrawals over that same time frame. Thus understanding the driving forces behind water use in agriculture is key for a comprehensive understanding of water demand in the United States.

To isolate the effects of water use in agriculture, this paper uses stochastic frontier analysis estimate a translog production frontier for agriculture at the state level, which allows an analysis of the structure of agriculture without the need for large quantities of detailed farm level data. Specifically this paper adopts the structure of Battese and Coelli (1995), which uses two uncorrelated errors, a random production error and a non-negative technical inefficiency effect. The inefficiency effects utilize a variety of producer and state level characteristics to estimate production inefficiency simultaneously with the production frontier. This provides a measure of each production unit's distance from the frontier, an improvement over the standard assumption that all deviations from the frontier are due to inefficiency.

Our production frontier estimation uses four inputs, cropland, labor, intermediates and total water. Specification tests validate the use of the two error structure of Battese and Coelli (1995) over ordinary least squares (OLS) and provide strong evidence for the chosen inefficiency effects structure compared to various alternatives. The adjusted $R^2$ from OLS indicates that over 92% of the variance in the value of agricultural output is explained by the basic translog model, while the inefficiency effects are found to explain nearly 91% of the total variance in the stochastic frontier model.

[1] An out of stream withdrawal is considered to be any withdrawal where water is transferred from the souce before end use. This is in contrast to in-stream uses such as hydroelectric power, where the full amount withdrawn is immediately returned to the source.

The mean marginal effect of cropland is found to be negative, which indicates that it is used too much relative to the choices of an efficient producer. The other marginal effects are positive, but only that of intermediate inputs (1.4785) is statistically significant. Converting these marginal effects, which are output elasticities due to the use of a translog production function, into marginal products, allows us to better analyze intermediate and irrigation factor use in particular. The average marginal product of intermediates is slightly below one, so on average using an extra dollar of intermediate inputs will add less than a dollar of output. The full distribution across state-years is skew, with states that have negative marginal products tending to have larger magnitude effects than those that have positive marginal products. Marginal products are negative for irrigation for the same 26 states in each year, indicating persistent misallocation of water resources in agriculture. California specific values are compared to a small sample of prices, revealing that at least in that state water is providing far more value as an agricultural input than it costs to acquire and deliver. Measured positive correlation between having a negative marginal product of intermediates and having a positive marginal product of irrigation suggests that a shift in inputs from intermediates to irrigation provided a natural avenue for substantial gains in efficiency in many states.

The inefficiency effects regression allows for the determination of the effects of state and farm level qualities on efficiency and output. Government subsidies are estimated to have a small positive effect on producer efficiency, with an average increase in value of output of 0.083 per real dollar of subsidies. Having either the riparian or prior appropriations water rights regimes boosts output relative to a hybrid system of the two, although only the riparian effect is statistically significant. Farm size effects have relatively large output elasticities, ranging from basically zero to 0.0286, with the highest values being for farms less than 50 acres and farms between 50 and 500 acres. The model suggests that a shift from larger acreage farms to lower acreage ones will generally be efficiency increasing.

The individual crop coefficients are each relatively small in magnitude, causing at most a 0.1% change in output for a 1% change in acreage for an average state. The ranking of these values does not closely correspond to existing rankings of crops by water intensity, with relatively water intensive wheat and soybeans being some of the most efficiency increasing crops and water intensive barley, sorghum, rice and corn all being efficiency reducing. On the whole, the crop specific results are more suitably viewed as a measure of the match of crops to their growing conditions rather than a direct measure of water efficiency of growing the crop.

The primary aim of this paper is to provide a tractable aggregate level view of the factors and characteristics of agricultural production and their effects on efficiency and output. This paper adapts the SFA methods for use in estimating the frontier for each state as a means of approximating an aggregate production function rather than estimating a farm or firm level production function explicitly. There is precedent for using SFA for aggregate production analysis in such a way. Coelli et al. (2003) examines agriculture in Bangladesh over the period of 1961-1992, using regions as decision making units, while Koop et al. (1999) uses SFA methods to investigate factors behind GDP growth in OECD countries. In each case the use of such methods provided useful insight into the mechanisms of production for their respective targets.

Section two outlines the models and methods used in this paper. Section three summarizes the data and discusses the empirical model estimation. Section four will conclude.

# 2 Stochastic Frontier Analysis

Stochastic frontier analysis, independently developed by Aigner et al. (1977) and Meeusen and van den Broeck (1977), is a procedure for estimating a production frontier with two types of errors, a random error and a non-negative technical inefficiency effect. Although this method requires parametric assumptions, it is desirable because it allows for the use of standard hypothesis testing procedures and because it does not restrict producer observations to lie within the frontier estimated. The latter property is particularly desirable because it allows for the presence of measurement errors or other forms of statistical noise in the model, while with nonparametric approaches all deviations from the frontier are assumed to be due to inefficiency.

This paper adopts the particular stochastic frontier model of Battese and Coelli (1995), which jointly estimates the structure of the technical inefficiency effects using a set of observable characteristics. This model has the following form for panel data,

$$y_{it} = x'_{it}\beta + \nu_{it} - u_{it}. \tag{1}$$

The random errors, $\nu_{it}$, are assumed to be independent and identically distributed standard normals with variance $\sigma_v$, while the inefficiency effects, $u_{it}$, are assumed to be independently distributed as positive truncations of a normal distribution with mean $z'_{it}\alpha$ and standard deviation $\sigma_u$.

In order to assess the effects of observable characteristics on technical inefficiency, it might seem natural to compute a stochastic frontier under the assumption of independent and identically distributed $uit$s across producers and then run a second stage regression of the observable characteristics on the resulting inefficiency values. An iid assumption of this type is often made in the literature, with half-normal, truncated normal and gamma distributions all being used as distributions for the inefficiency effects[2] and such a two-stage procedure is widely used in such analyses. However, this procedure directly violates the assumption the that the inefficiency effects are identically distributed unless all of the second stage coefficients are simultaneously equal to zero (Coelli et al., 2005).

To avoid this contradiction, the procedure this paper uses simultaneously esti-

[2]Details on the structure and properties of each of these models can be found in Kumbhakar and Lovell (2000).

mates the production function and the mean of the inefficiency effects, $z_{it}'\alpha$. In order to better understand the structure of this estimation it is helpful view the technical efficiency effects as follows,

$$u_{it} = z_{it}'\alpha + \omega_{it}, \tag{2}$$

where $\omega_{it}$ is a positive truncation of a standard normal with standard deviation $\sigma_u$ such that $\omega_{it} > -z_{it}'\alpha$. Note that this is mathematically equivalent to the distributional assumption for $u_{it}$ made in equation (1) and that by assumption the $\omega_{it}$s must be independently distributed but is not necessarily identically distributed nor positive.

To more easily obtain the estimates for the model it is convenient to parameterize it in terms of the total variance ($\sigma^2 = \sigma_\nu^2 + \sigma_u^2$) and the share of the total variance due to the idiosyncratic inefficiency effects ($\gamma = \sigma_u^2/\sigma^2$). As long as the inefficiency effects are stochastic[3] then $\beta$, $\alpha$, $\sigma$ and $\gamma$ can be estimated simultaneously using the method of maximum likelihood.[4]

# 3 Data and Empirical Model

The empirical exercise of this paper uses a balanced panel of aggregated farm data at the state level. Looking at agriculture at the state level allows for a macro level of analysis that does not require detailed farm level data that is often hard to find for even a single state, let alone the entire continental U.S. Due of the nature of this setup, we use the terms producer and state(-year) interchangeably. We must take caution when it comes to policy recommendations from an estimation with aggregate variables, an issue which will be discussed further in Section 4.

Data on the value of agricultural output, inputs and a variety of producer characteristics comes from the U.S. Department of Agriculture's National Agricultural

[3] That is $\gamma \neq 0$. If $\gamma = 0$ then the correct specification is an ordinary least squares regression.
[4] The full likelihood function and first order conditions are presented in Battese and Coelli (1993).

Statistical Service (NASS) and the U.S. Census Bureau's CenStats database. Water withdrawal data is from the United States Geological Survey (USGS) National Water Use Information Database, rain data from the National Oceanic and Atmospheric Administration and evapotranspiration estimates from Carr et al. (1990). There are a total of 240 observations for the years 1985-2005 in five year intervals, including all 48 continental states. Table 1 presents a list of variables and their summary statistics. All variables are aggregated at the state level unless otherwise noted and all real values have a base year of 2005.

Our main analysis uses a translog functional form for (1), with the logarithms of cropland, labor, intermediate inputs (e.g. fertilizer, pesticides and equipment) and total available water as inputs to estimate the real value of agricultural output. Total available water is of particular note, because its value does not come directly from a single data source. It is calculated for each state (in millions of gallons) as,

$$W = C \cdot \tau \cdot RAIN - ET_R \cdot \frac{C}{AREA} + (1-\delta)IR \; = ET_0 + (1-\delta)IR, \qquad (3)$$

where $\tau$ is a universal unit conversion factor, $RAIN$ is yearly rainfall in inches, $C$ is acres of cropland utilized, $ET_R$ is total raw evapotranspiration, AREA is total acreage, $\delta$ is the rate of conveyance losses in irrigation and $IR$ is the amount of water withdrawals for irrigation. Thus by construction we have that $ET_0$ represents effective evapotranspiration. Counting only water withdrawn from existing sources for irrigation ignores that rainfall and irrigation are at least partial substitutes in agricultural production. The primary concept captured in this formulation is that if you expand cropland it will have a bigger immediate effect on total water use in agriculture in a rainy state versus a dry state, holding irrigation withdrawals constant.

For notational simplicity we refer to the sets of inputs, regions and additional inefficiency variables as $\mathcal{X}$, $R$, and $\mathcal{V}$, respectively. Our stochastic frontier model to

Table 1: Variable Description and Summary Statistics

| Variable | Mean | Median | Max | Min | SD |
|---|---|---|---|---|---|
| Value of Goods Produced (th $/year) | 1765300 | 1156260 | 23531914 | 27950 | 2552436 |
| *Inputs* | | | | | |
| Cropland (acres) | 9037922 | 5580779 | 38414769 | 23766 | 9178643 |
| Labor (workers) | 65973 | 60131 | 329942 | 1313 | 59299 |
| Intermediate Inputs (th $/yr) | 1833311 | 1222019 | 16071787 | 11460 | 2108384 |
| Effective Water (Mgal/yr) | 870839 | 168586 | 10874899 | 995 | 1639870 |
| *Additional Characteristics* | | | | | |
| California* (1/0) | 0.0208 | - | - | - | - |
| Northern Forests (1/0) | 0.0625 | - | - | - | - |
| Eastern Temperate Forests (1/0) | 0.5625 | - | - | - | - |
| Great Plains (1/0) | 0.1875 | - | - | - | - |
| North American Desert (1/0) | 0.1042 | - | - | - | - |
| Northwestern Forested Mountains (1/0) | 0.0625 | - | - | - | - |
| Hybrid Water Rights* (1/0) | 0.2292 | - | - | - | - |
| Prior Appropriations Water Rights (1/0) | 0.1667 | - | - | - | - |
| Riparian Water Rights (1/0) | 0.6042 | - | - | - | - |
| Federal Farm Subsidies (th $/yr) | 270747 | 125311 | 2286637 | 107 | 385289 |
| Surface Withdrawals (Mgal/yr) | 1734 | 98 | 20225 | 0 | 3673 |
| Flood Irrigation (acres) | 1669679 | 219026 | 12474480 | 0 | 2616618 |
| Family Farms | 37153 | 29370 | 216651 | 539 | 32924 |
| S1 (<50 Acres) | 13279 | 10427 | 86172 | 370 | 12298 |
| S2 (50-500 Acres) | 21845 | 12993 | 113548 | 269 | 21033 |
| S3 (500-1000 Acres) | 3732 | 2474 | 19568 | 6 | 4177 |
| S4 (1000-2000 Acres) | 2086 | 1094 | 12704 | 2 | 2567 |
| S5 (2000+ Acres) | 1517 | 625 | 10818 | 0 | 2186 |
| *Crops (acres planted)* | | | | | |
| Wheat | 13267 | 11575 | 70373 | 0 | 11950 |
| Rice | 61901 | 0 | 1631800 | 0 | 219023 |
| Corn | 100899 | 33691 | 930854 | 0 | 169330 |
| Barley | 93758 | 178 | 3500000 | 0 | 355274 |
| Sorghum | 147095 | 0 | 3573431 | 0 | 550472 |
| Cotton | 180735 | 0 | 6415000 | 0 | 713343 |
| Sugarcane | 12380 | 0 | 419000 | 0 | 64747 |
| Sugarbeets | 28192 | 0 | 488200 | 0 | 75862 |
| Tobacco | 11912 | 0 | 284195 | 0 | 39912 |
| Flaxseed | 8231 | 0 | 883400 | 0 | 66529 |
| Peanuts | 18877 | 0 | 750600 | 0 | 82273 |
| Soybeans | 8974 | 63 | 41740 | 0 | 11609 |
| Sunflowers | 43485 | 0 | 2190000 | 0 | 217306 |
| Beans | 61956 | 0 | 1827738 | 0 | 179585 |
| Hay | 756472 | 149734 | 4000000 | 0 | 1046576 |
| Potatoes | 21124 | 0 | 406000 | 0 | 60896 |
| Irrigation Loss Rate | 0.0808 | 0.0000 | 0.87 | 0.0000 | 0.1329 |

(*) Omitted categories in regressions.

be estimated is thus is defined as

$$log(y_{it}) = \beta_0 + \beta_t t + \sum_{r \in R} \beta_r D_{ri} + \sum_{x \in \mathcal{X}} \beta_x x_{it} + \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} \beta_{xy} x_{it} y_{it} + v_{it} - u_{it}, \qquad (4)$$

$$u_{it} = \alpha_0 + \alpha_t t + \sum_{v \in \mathcal{V}} \alpha_d z_{vit} + \omega_{it}, \qquad (5)$$

where $D_{ri}$ is a dummy variable for state $i$ being in region $r$, and $\nu_{it}$ and $\omega_{it}$ are distributed as in (1) and (2), respectively.

Regions were chosen by matching each state with its predominant Level I ecoregion (Commission for Environmental Cooperation, 1997), by majority if possible and by plurality if necessary.[5] California is taken as our base region, since it was the only state that constituted its own region under this criterion. This is intended to capture similarity of growing conditions and effects that could not be accounted for with data in a more meaningful way than choosing a more artificial division merely related to geographical proximity.

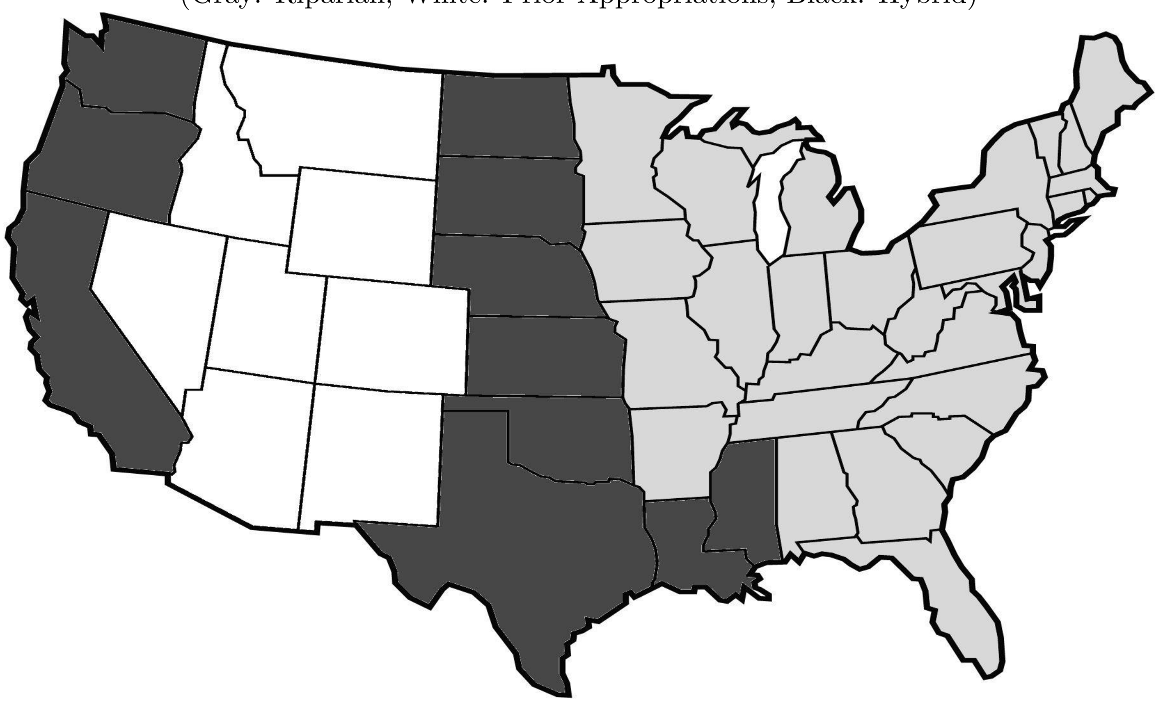
The set of additional characteristics, $\mathcal{V}$, includes the log of the real value of federal farm subsidies, dummies for water rights regimes, acreage for sixteen different crop categories, the number of family run farms, the amount of surface withdrawals for irrigation, acres flood irrigated and a farm size distribution, separated into five categories by acreage. A number of these factors are included to account for effects due to issues that affect irrigation type, availability and intensity. Crop choices will be analyzed through the lens of water intensity per acre, while surface withdrawals and acres flood irrigated provide rough measures of irrigation efficiency, as will be discussed further in the Results section.

Of particular interest are the dummies regarding the three primary water rights regimes in the United States, riparian, prior appropriations and hybrid, which simply merges elements of the other two. Riparian systems attach water rights to any property that abuts a body of water and are generally proportional to frontage on the

[5]Current map available: ftp://ftp.epa.gov/wed/ecoregions/cec_na/NA_LEVEL_I.pdf

source, while prior appropriations follow a doctrine generally referred to as "first in time, first in right", where age of claims take precedence over physical relationships between land and a water source. Importantly, under the prior appropriations doctrine, water rights are severable from land. The spatial distribution of water rights regimes are shown in Figure 1

Figure 1: Water Rights Regimes
(Gray: Riparian, White: Prior Appropriations, Black: Hybrid)



Maximum likelihood estimates for (4) and (5) are obtained using the *frontier* package for the R statistical computing environment (Coelli and Henningsen, 2011). This package uses the Fortran code of FRONTIER 4.1[6] as a base, and computes the maximum likelihood estimates using the ordinary least squares estimates (OLS) as an initial guess. Model specification tests and final results presented are all output from routines contained in this package.

The results of specification tests using likelihood ratio (LR) statistics are presented in Table 2. Each involves a standard LR test, except in the case of the null hypothesis

[6] `http://www.uq.edu.au/economics/cepa/frontier.htm`

Table 2: Specification Tests

| Null hypothesis | Log-likelihood | $\chi^2$ Statistic | Critical Value | Decision |
|---|---|---|---|---|
| Baseline Model | 24.8545 | | | |
| (1) $\gamma = 0$ | -81.649 | 213.01 | Mixed $\chi^2_{2.0.95} = 5.138$ | Reject $H_0$ |
| (2) $\delta_0 = \delta_d = 0$ | -114.885 | 279.48 | $\chi^2_{33..95} = 47.400$ | Reject $H_0$ |
| (3) No Crop Effects | -78.980 | 207.67 | $\chi^2_{16..95} = 26.296$ | Reject $H_0$ |
| (4) No Farm Size Effects | -9.8965 | 69.502 | $\chi^2_{5..95} = 11.070$ | Reject $H_0$ |
| (5) No Water Effects | 16.983 | 15.743 | $\chi^2_{4.0.95} = 9.488$ | Reject $H_0$ |

Mixed $\chi^2$ critical value from Kodde and Palm (1986).

of no inefficiency effects, when the generalized likelihood ratio asymptotically has a mixed chi-square distribution (Kodde and Palm, 1986). In (1) and (2) we test two restrictions of the structure of the technical inefficiency effects, that the share of variance explained by the inefficiency effects, $\gamma$, is equal to zero and that all of the coefficients of the second stage regression are equal to zero. The first null implies that our model is not an improvement over OLS, while the second implies that the technical inefficiency effects have a half-normal distribution (Aigner et al., 1977). In each case the null hypothesis is rejected, with strong evidence in favor of the chosen specification of the technical inefficiency effects.

Finally, we test three restrictions of the inefficiency effects. The joint significance of the crop effects is tested in (3) and this null is rejected, indicating that the set of crop variables provide valuable information as a whole, even though most are not individually significant. The sixteen crops included cover 18% of total cropland reported for the nation and many of them show up in various agricultural studies (e.g. Pimentel et al. (1997) and Postel (1998)) as water intensive crops. The farm size effects are tested in (4) and this null is also rejected, indicating that the farm size distribution is jointly signficiant. Finally, we test the joint significance of the water related variables, including both irrigation measures and water rights dummies, in (5) and this null is rejected as well, justifying the inclusion of this array of direct and indirect measures of irrigation productivity in the inefficiency effects. On the whole these results providing evidence that the chosen specification for the inefficiency effects is preferred over one with these variables omitted.

# 4 Results

Table 3 presents the results of the production function estimation. The adjusted $R^2$ from an OLS specification shows that these variables alone explain nearly 93% of the variation in agricultural production. Of the fourteen input coefficients, nine are statistically significant at the 5% level. The two measures of variance, $\sigma^2$ and $\gamma$, are each significant at the 1% level, supporting use of the technical inefficiency effects framework. The magnitude of $\sigma$ is equivalent to a 35% change in output, all else held equal, while $\gamma = 0.9091$ means that the inefficiency effects explain 90.91% of that variance. The time trend is significant and negative, indicating a slight deterioration of agricultural productivity over time on an aggregate level.[7]

The regional fixed effects can be interpreted as a weak measure of productivity in this model, with indicators for other ecoregions measuring differences from our base region, California, whose value (the constant) is significant at the 1% level. The indicators for the Great Plains, North American Desert are negative and signficiant, while the Northwestern Forested Mountains is significant and positive. Differences in these coefficiences can generally be attributed to any number of differences that may affect growing seasons or agricultural quality relative to California. The Great Plains and North American Desert regions cover the relatively dry plains (Kansas, Nebraska, North and South Dakota) and southwestern (Arizona, Utah, New Mexico) states, so likely factors driving these coefficient's negative values are issues of water availability, sun intensity and soil quality. The Northwestern Forested Mountains include Oregon, Washington and Idaho, so similarly, soil quality and water availability seem plausible factors explaining a positive indicator value.

With a translog production function the values of the input coefficients themselves

[7] A specification with regionally differing time trends was tested, and these additional coefficients were found to be jointly insignificant using a LR test.

Table 3: Stochastic Frontier Estimation For United States Agriculture

| | | |
|---|---|---|
| Intercept | 18.880 | 8.7037*** |
| Time | -0.110 | -4.3612*** |
| Northern Forests | -0.181 | -1.0511 |
| Eastern Temperate Forests | 0.146 | 0.8467 |
| Great Plains | -0.339 | -2.2113** |
| North American Desert | -0.596 | -2.4579** |
| Northwestern Forested Mountains | 0.441 | 2.3815** |
| log(CROP) | -3.815 | -7.2598*** |
| log(LABOR) | 1.174 | 2.0268** |
| log(INT) | 0.405 | 0.7957 |
| log(WATER) | 1.606 | 6.5067*** |
| $CROP^2$ | 0.611 | 11.3937*** |
| CROP*LABOR | -0.415 | -5.1747*** |
| CROP*WATER | -0.160 | -6.0445*** |
| CROP*INT | 0.051 | 1.1433 |
| $LABOR^2$ | 0.228 | 1.5816 |
| LABOR*INT | 0.253 | 4.4656*** |
| LABOR*WATER | -0.045 | -1.3723 |
| $INT^2$ | -0.241 | -3.0533*** |
| INT*WATER | 0.019 | 0.8272 |
| $WATER^2$ | 0.091 | 2.8250*** |
| $\sigma^2$ | 0.0921 | 7.0936*** |
| $\gamma$ | 0.9091 | 26.0558*** |
| Log-likelihood | | 24.8545 |
| Adjusted $R^2$ from OLS | | 0.9291 |
| Observations | | 240 |
| Years | | 5 |

Significance levels: $* = 10\%, ** = 5\%, *** = 1\%$

do not have an easily interpretable meaning, so to truly assess input effects we look at the marginal effect for each input. With a standard translog production function the marginal effect for input $x$, which is equivalent to that input's partial output elasticity, is calculated as

$$ME_x = \frac{\partial \log(y)}{\partial \log(x)} = \beta_x + \sum_{k \in \mathcal{X}} \beta_{xk} \log(k). \tag{6}$$

In our model this formula still holds for labor, intermediates and total available water. However, due to the dependence of total water available on cropland and irrigation withdrawals, the marginal effects of these two variables on output are slightly more complicated. Calculating directly from the production function yields the following formulas,

$$\frac{\partial \log(y)}{\partial \log(C)} = ME_C + \frac{ET_0\, C}{W}\, ME_W, \tag{7}$$

$$\frac{\partial \log(y)}{\partial \log(IR)} = \frac{(1-\delta)\, IR}{W}\, ME_W, \tag{8}$$

again with $ET_0$ representing effective evapotranspiration and $\delta$ being the irrigation loss rate. For $C$ an additional marginal effect comes from more land being subject to rain which increases effective evapotranspiration, while for $IR$ the marginal effect is mitigated by irrigation withdrawals only accounting for a fraction of total water available and being subject to losses due to transportation and evaporation.

The raw marginal effects vary widely across state-year observations, with there being few cases of sign changes for a state from one year to another, so to better assess agricultural production behavior, we focus our attention on the mean values. Table 4 provides the mean marginal effects along with bootstrapped confidence intervals for each.[8] We know from standard producer theory that in a perfectly competitive market an optimizing producer will use a factor until its marginal product is equal

[8]Computed using the percentile method, appropriate percentiles of the empirical distribution. For each marginal effect we are unable to reject the null that the data is distributed normally using the nonparametric Anderson-Darling test (Anderson and Darling, 1952), justifying the use of this method over a more complicated procedure.

Table 4: Marginal Effects

| Factor | Marginal Effect | 95% Confidence Interval | | Marginal Product |
|---|---|---|---|---|
| C | -2.5098 | -5.0023 | 0.0352 | -1.2766 |
| L | 1.2142 | -0.5776 | 2.9828 | 8.1217 |
| I | 1.4785 | 0.3351 | 2.6117 | 0.9643 |
| W | 0.3808 | -0.2368 | 0.9996 | 2.8432 |
| IR | 0.1344 | -0.3255 | 0.5846 | 0.3789 |

to its price. So for any input that is necessary to produce a positive level of output, a negative marginal effect automatically implies overuse of a factor relative to an efficient allocation, while a positive value is ambiguous without additional analysis.

The mean marginal effect of cropland is $-2.5098$, meaning that a 1% increase in cropland would on average lead to nearly a 2.5% decrease in output. It is only positive 60 state-years, comprising the same twelve states each of the five years of the sample, with no discernable geographic pattern to their distribution. The marginal effects of water and irrigation are 0.3808 and 0.1344, respectively corresponding to 0.38% and 0.13% changes in output for a 1% change in the underlying factor. The marginal effect of intermediate inputs is the only value which is statistically significantly, measured to be 1.4785. Converting these marginal effects, which are output elasticities, into true marginal products, the raw change in value of output for a one unit change in a factor, gives us implied factor prices, or shadow prices. Marginal products are particularly easy to interpret for intermediate inputs, since that factor is measured in terms of the dollar value of intermediates rather than as some raw quantity of inputs. The model finds that adding $1 of intermediate inputs yields about $0.96 of final agricultural output. Over the whole sample, 164 of the 240 state-year observations have marginal products greater than 1,so the negative values are on average larger than the positive one. Examining the average of positive and negative marginal product producers highlights this, with the positive average being $5.34 and the negative $-12.16, while the median marginal product is $2.25. This implies that there are large margins for more efficient use of intermediate inputs, particularly in the Eastern Forests region, where the majority of these negative marginal product states are located.[9]

The marginal product results for water and irrigation find that the average values of one million gallons per day for a year are respectively \$2,843 and \$378.86. To assess what this means for agricultural water use, we would need to compare irrigation prices to the marginal products of individual states. Thanks to some irrigation price data from Wichelns (2010), we can do exactly this for California, which has the highest marginal product of water in each year except 1985, when it is second behind Florida. Estimated water charges per 1,000 $m^3$ from the California State Water Project ranged from \$28 to \$466 in 2006, while water supplied from the U.S. Bureau of Reclamation to the Westlands Water District had a price of \$75.66 in 2008 (all converted into 2005 dollars).[10] Converting the marginal values above into the same units yields an implied value of water (i.e. shadow price) in California that increases from \$5,279 in 1985 to \$15,223 in 2005. So water use in California does not buck traditional economic theory, with the marginal value added of this crucial input on average exceeding the price paid for it. However, it is important to note that these values used for comparison are not market prices, per se, but prices determined by regional irrigation districts to cover costs, often after being bought from the Bureau of Reclamation at relatively low rates.

Altogether the same 24 states have a positive marginal product of water in each year,[11] and altogether they comprise nearly 60% of total agricultural value produced over the sample period. This means that, at least for present uses, the other 26 states should reduce their relative usage of water. This could be accomplished by

[9]The full list of states that have a negative marginal product in at least one year is Arkansas, Delaware, Illinois, Kentucky, Massachusetts, Missouri, New Hampshire, North Carolina, Oregon, South Dakota, Vermont and Wisconsin.

[10]Examples from 1992 yields prices between \$4.64 and \$224.67.

[11]These states include Arkansas, California, Delaware, Florida, Illinois, Indiana, Kentucky, Louisiana, Massachusetts, Missouri, Montana, New Hampshire, New Jersey, North Dakota, Oregon, Pennsylvania, South Dakota, Tennessee, Vermont, Virginia, Wisconsin and Wyoming.

improving water efficiency (e.g. growing less water intensive crops or using new irrigation methods) or altering the mix of inputs to increase production while keeping water use constant. ltogether they comprise nearly 60% of total agricultural value produced over the sample period. This means that, at least for present uses, the other 26 states should reduce their relative usage of water. This could be accomplished by improving water efficiency (e.g. growing less water intensive crops or using new irrigation methods) or altering the mix of inputs to increase production while keeping water use constant.

Another interesting interaction these results displays is that there is a high level of correlation ($\rho = 0.577$) between having a positive marginal product of water and having a negative marginal product of intermediates, suggesting that at some level intermediate inputs could be substituted for water in these states and provide substantial gains in efficiency while keeping output constant. The states that possess these two qualities, positive marginal product of water and negative marginal product of intermediate inputs, are Arkansas, Delaware, Illinois, Kentucky, Massachusetts, Missouri, New Hampshire, Oregon, South Dakota, Vermont and Wisconsin. Ideally, this could be achieved by shifting agricultural subsidies, due their ineffective policy effects.[12]

This choice of model provides an additional margin to examine inefficiency, the inefficiency effects regression. It should be noted that a positive sign indicates that the variable increases inefficiency and a negative sign indicates that it lowers inefficiency, with the reverse being true if we consider the effect on final output. Because our dependent variable is logged, the coefficient magnitudes conveniently measure the total percentage change in input given a single unit change in that variable, holding all else equal. For an inefficiency effect not taken in logs the average percentage change in output for a one unit increase will be $e^{-\beta_z} - 1$, while if the variable is

[12]On average \$0.083 of value created for each dollar of subsidies, a result which is outlined further in the inefficiency effects results.

logged the percentage change for a 1% increase this average will be equal to $-\beta_z$. For a state with the mean characteristics, the inefficiency effects lead to a 8% increase in the betvalue of output compared to the case with no inefficiency effects present. The full inefficiency regression results are given in Table 5.

The coefficient on agricultural subsidies is negative and significant, resulting in a small marginal product of 0.083. So while subsidies have a mild positive effect on efficiency, for every dollar of subsidies, this estimation finds that less than ten cents of output is created. The two water rights regime coefficients are negative, indicating that having a fixed water rights regime reduces inefficiency versus the hybrid system. However, only the riparian rights coefficient is statistically significant, with an output elasticity of 39.69%. It is difficult to attribute too much meaning to the water rights coefficients, as there is significant geographic correspondence between ecoregions and water rights regimes. An analysis with more disaggregated data may be able to separate these effects more clearly.

Five of the sixteen crop coefficients are significant, those for wheat, barley, sorghum, peanuts and potatoes. The coefficients for wheat and potatoes each decrease inefficiency, while the others increase it. Examining output elasticit

This highlights that these coefficients should be considered as a measure of the suitability of current chop choices, in essence how well each crop is currently matched with the growing conditions.[13] Viewed this way, the coefficients suggest that the most optimally grown crops are wheat, sugarcane and soybeans, while rice, corn, barley and sorghum are grown in the least ideal places. With his interpretation of the crop coefficients it is important to note that decreasing crops whose coefficients are negative and increasing those coefficients are positive cannot be taken as a quick fix to make a state a more efficient agricultural producer. Suitability of climate, soil

[13]Along with the productivity coefficients in the stochastic frontier, this is another way in which elements of soil quality or other dimensions of growing suitability find their way into the results of this estimation.

Table 5: Inefficiency Effects Regression

| | | |
|---|---|---|
| Intercept | 2.093 | 4.4860*** |
| Year | 0.026 | 0.6007 |
| Prior Appropriations Rights | -0.227 | -1.0095 |
| Riparian Rights | -0.334 | -1.7042* |
| log(Subsidies) | -0.091 | -1.9171* |
| Wheat | -1.27E-05 | 1.6614* |
| Rice | 1.84E-07 | 0.8209 |
| Corn | 3.17E-07 | 0.8560 |
| Barley | 4.29E-07 | 2.8614*** |
| Sorghum | 2.50E-07 | 1.9531* |
| Cotton | -2.13E-07 | -1.4690 |
| Sugarcane | -2.86E-05 | -1.2555 |
| Sugarbeets | -6.64E-07 | -0.6218 |
| Tobacco | 3.36E-07 | 0.2738 |
| Flaxseed | -4.14E-06 | -1.2829 |
| Peanuts | 1.08E-06 | 2.3775** |
| Soybeans | -6.45E-06 | -1.0004 |
| Sunflowers | 6.65E-08 | 0.0990 |
| Beans | -3.23E-07 | -1.0114 |
| Hay | -5.72E-08 | -0.9835 |
| Potatoes | -1.96E-06 | -1.6805* |
| Surface Withdrawals | 1.03E-04 | 3.0692*** |
| Flood Irrigation | -6.40E-08 | -1.5424 |
| Family Farms | 1.75E-04 | 3.6292*** |
| S1 (<50 Acres) | -2.21E-04 | -4.7291*** |
| S2 (50-500 Acres) | -1.03E-04 | -2.2047** |
| S3 (500-1000 Acres) | -3.67E-04 | -3.7491*** |
| S4 (1000-2000 Acres) | -3.29E-05 | -0.1841 |
| S5 (2000+ Acres) | -1.96E-04 | -2.5147** |

Significance levels: $* = 10\%, ** = 5\%, *** = 1\%$

and other growing factors is crucial.

Both the family farm and farm size coefficients measure the effect of farm scale on inefficiency. The family farm effect is significant and positive, with an output elasticity of $-0.069$. In terms of farm size effects, four of the five coefficients are statistically significant and all of the coefficients are positive, indicating that increasing the number of farms in a state should always cause at least a modest increase efficiency. Shifting the distribution towards farms below 50 acres (S1) and farms between 50 and 500 acres (S2) relatively increases efficiency, with output elasticities of 0.0286 and 0.022 respectively. The final inefficiency effects examined are the two direct measures of irrigation, the amount of surface withdrawals and the acreage of flood irrigation. Howell (2003) generally evaluates water use efficiency by type of irrigation system, and finds that surface methods are on average 65% efficient, versus 70-80% efficiency for sprinkler methods. However, our analysis finds a statistically insignificant result of a slightly positive output elasticity for flood irrigation. The surface irrigation coefficient is negative, as expected, and statistically significant, with an output elasticity of $-0.0018$. The values for these two output elasticities are nearly equal and opposite to each other, meaning that a 1% increase in surface withdrawals for flood irrigation should have nearly no effect on the value of output. One major drawback to the analysis of irrigation is that the USGS did not split irrigation data into more than two categories until 1995, leaving us unable to account for the rise of microirrigation methods, which Howell (2003) finds are on average 85-90% efficient and thus would be expected to play a larger role in the future of irrigation.

# 5 Conclusion

This paper presented a state-level estimation of agricultural production to analyze water use behavior in that sector in the United States. Using the two-error stochastic frontier model of Battese and Coelli (1995) with data from a variety of U.S. government agencies, we simultaneously estimated a translog production function and technical inefficiency effects, which allow us to measure raw technical inefficiency using a variety of farm variables and characteristics.

Our four input translog estimates yield on average a negative output elasticity for cropland, indicating that it is used more than an efficient producer would choose to. The average marginal product of intermediate inputs is slightly below 1, indicating that on average an additional dollar of intermediates adds less than a dollar of agricultural value. However, there is wide heterogeneity across state-years, with a few states with particularly low marginal products of intermediate inputs dragging down the average effect. The average values for water and irrigation are positive, however they are negative for 26 individual states, which account for 42% of the total value of output. Calfornia specific values are compared to a small sample of prices, revealing that at least in that state water is providing at least ten times as much value as an agricultural input than it costs to acquire and deliver. Overall the primary regression results suggest that major shifts are possible to improve water efficiency or alter the mix of inputs to increase production while keeping water use constant. In particular, measured positive correlation between having a negative marginal product of intermediates and a positive marginal product of irrigation suggests that a shift in inputs from intermediates to irrigation seems a promising avenue to provide substantial gains in efficiency in many states.

The inefficiency effects regression finds that government subsidies have a small positive effect on producer efficiency, which translates into a small marginal product per dollar of 0.083. Having the non-transferable riparian water rights regime provides a statistically significant 39.7% boost to output relative to being under a hybrid regime. Farm size effects have output elasticities ranging from basically zero to 0.0286, with the highest values being for farms less than 50 acres and farms between 50 and 500 acres. The model suggests that a shift towards smaller farms at the expense of larger ones should lead to sizeable efficiency gains. Irrigation type effects are found to be small, although the data available provided incomplete information about irrigation choices for most of the sample.rable riparian water rights regime provides a statistically significant 39.7% boost to output relative to being under a hybrid regime. Farm size effects have output elasticities ranging from basically zero to 0.0286, with the highest values being for farms less than 50 acres and farms between 50 and 500 acres. The model suggests that a shift towards smaller farms at the expense of larger ones should lead to sizeable efficiency gains. Irrigation type effects are found to be small, although the data available provided incomplete information about irrigation choices for most of the sample.

As would be expected individual crop coefficients are each relatively small in magnitude, implying at most a 0.1% change in output for a 1% change in a crop in an average state. The ranking of these values does not closely correspond to existing rankings of crops by water intensity, with relatively water intensive wheat and soybeans being some of the most efficiency increasing crops and water intensive barley, sorghum, rice and corn all being efficiency reducing. On the whole, the crop specific results should be viewed as a measure of the suitability of crop choices and not necessarily a direct measure of water efficiency of growing the crop.

This analysis implies that differences in water use have major implications for farm policy as a whole and provide interesting directions for future investigations. Additionally, better data, for example about irrigation choices, soil quality and other possible legal restrictions that may exist beyond the standard rights frameworks could provide valuable information to a stochastic frontier estimation such as this one. Observed relationships in the model, between water and intermediate inputs in particular, suggest concrete areas that policies could target to spur efficiency gains. Use of a wide panel of individual farm data would be ideal for providing more specific policy guidance and improving both the efficiency of water use and agricultural production in the United States.

# References


Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):pp. 193–212.

Battese, G. and Coelli, T. (1993). A model for technical inefficiency effects in a stochastic frontier production function for panel data. Working Papers in Econometrics and Applied Statistics 69, Department of Econometrics, University of New England, Armidale.

Battese, G. and Coelli, T. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20:325–332.

Carr, J. E., Chase, E. B., Paulson, R. W., and Moody, D. W. (1990). National water use summary 1987: hydrologic events and water supply and use. Water Supply Paper 2350, U.S. Geological Survey.

Coelli, T. and Henningsen, A. (2011). *frontier: Stochastic Frontier Analysis*. R package version 0.997-0.

Coelli, T., Rahman, S., and Thirtle, C. (2003). A stochastic frontier approach to total factor productivity measurement in bangladesh crop agriculture, 196192. *Journal of International Development*, 15(3):321–333.

Coelli, T. J., Rao, D. P., O'Donnell, C. J., and Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. Springer Science + Business Media, New York, 2nd edition.

Commission for Environmental Cooperation (1997). *Ecological regions of North America – Toward a Common Perspective*. CEC, Quebec, Canada.

Howell, T. (2003). *Irrigation Efficiency*, pages 467–472. Marcel-Dekker, Inc.

Kodde, D. A. and Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, 54(5):pp. 1243–1248.

Koop, G., Osiewalski, J., and Steel, M. F. (1999). The Components of Output Growth: A Stochastic Frontier Analysis. *Oxford Bulletin of Economics and Statistics*, 61(4):455–87.

Kumbhakar, S. and Lovell, C. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge.

Meeusen, W. and van den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, 18(2):435–44.

Pimentel, D., Houser, J., Preiss, E., White, O., Fang, H., Mesnick, L., Barsky, T., Tariche, S., Schreck, J., and Alpert, S. (1997). Water Resources: Agriculture, the Environment, and Society. *BioScience*, 47(2):pp. 97–106.

Postel, S. L. (1998). Water for food production: Will there be enough in 2025?. *BioScience*, 48(8):629 – 637.

Wichelns, D. (2010). Agricultural Water Pricing: United States. Background report, Organization for Economic Co-operation and Development.