ADAPTATION OF MEDITERRANEAN CROPS TO WATER PRESSURE IN THE EBRO BASIN: A WATER EFFICIENCY INDEX

AUTHORS

Zaira Fernández-Haddad and Sonia Quiroga

Department of Statistics, Economic Structure and International Organization,

Facultad de Ciencias Económicas y Empresariales

Universidad de Alcala

Plaza de la Victoria, 2. Alcala de Henares, 28802, Madrid, Spain

Phone: +34 918855198

Email: zairaferh@yahoo.com.mx

Paper prepared for presentation at the EAAE 2011 Congress

Change and Uncertainty

Challenges for Agriculture,

Food and Natural Resources

August 30 to September 2, 2011

ETH Zurich, Zurich, Switzerland

ADAPTATION OF MEDITERRANEAN CROPS TO WATER PRESSURE IN THE EBRO BASIN: A WATER EFFICIENCY INDEX

Z. Fernández-Haddad[1] and S. Quiroga[1]

[1]{Department of Statistics, Economic Structure and International Organization, Universidad de Alcala, Spain}

In this paper, we assess the output-oriented technical efficiency of agricultural production functions in order to compare, over time, economic and environmental production processes in the different regions of the Spanish Ebro basin, in a climate change context. The measurement of technical efficiency in agriculture can provide useful information about the competitiveness of farms and their potential to increase its productivity moreover can help in the crops adaptation to water pressure by improving the management of scarce resources. Here, we generate an agricultural water efficiency index to evaluate the adaptation of some Mediterranean crops to the water pressures in this area. We estimate frontier production functions and technical efficiency measures, using panel data models. This will allow us to observe changes in production due to individual specific effects and those that are time specific. To characterize our model, we use historical data, about crop yields, water requirements and climate as well as socio-economic and geographical aspects of the most representative crops in the provinces of the Ebro basin during 1976-2007. Then we generate a ranking of the most efficient crops across geographical areas, given their water use and other inputs, to evaluate policy scenarios with adjustments in water supply.

Keywords: water efficiency index in agriculture; Ebro basin; climate change adaptation

1 Introduction

Agriculture is the main user of water and other environmental and natural resources and therefore it plays an important in global ecosystem sustainability. According to the OECD, agriculture accounts about 70% of total available water which is mainly used for irrigation. Given that, a small change in agricultural water use, can have important economic and hydrological impacts. In this context, agricultural research and public policy have been prioritized the adaptation of crop yields to water pressures.

There is a lot of literature to study this problem in diverse areas of the knowledge. Econometric stochastic frontier analysis (SFA) of technical efficiency is an adaptable technique to help in the analysis of problem. The term of the stochastic frontier production function was proposed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), and it has been highly used in the econometric modeling of production and the estimation of technical efficiency of economic agents. The frontier have two random components, the first one is associated with the presence of technical inefficiency and the other being a traditional random error (Battese and Coelli, 1992). Nowadays, literature in this tool is large and it is growing quickly.

It is well known that the measurement of technical efficiency in agriculture can provide useful information about the competitiveness of farms and their potential to increase its productivity (Hallam and Machado, 1996). Moreover this technique can help in the crops adaptation to water pressure by improving the management of scarce resources. This methodology could be used in two directions input-output or output-input estimates the maximum possible production given a set of inputs (Alvarez-Pinilla, 2001; Battese and Coelli, 1992). In other words, deterministic and stochastic frontier functions estimate maxima or minima of a dependent variable given explanatory variables, usually to estimate production or cost functions.

Using this tool, we propose a water efficiency index to measure the degree of use of this input in the crop production in each province and later to rank the most efficient crops by province.This paper focuses on the Ebro basin in Spain, where agriculture can reach up to 90% or more of water consumption. The Ebro Basin is located in the Northeast of the Iberian Peninsula with a total area of 85,362 km2. This watershed is the largest in Spain, accounting for 17.3% of the total national area. It is made up of 347 major rivers, including the Ebro River, which drains the basin. It rises in the Cantabrian Mountains and ends in the Mediterranean and has a total length of 910 km and 12,000 km of main river network (CHEBRO, 2009). The climate in this basin is primarily Continental Mediterranean, with hot, dry summers, cold, wet winters and short, unstable autumns and springs. In the middle of the basin, the climate is semi-arid and in the northwest corner it is oceanic. Consequently, there is a wide heterogeneity in temperature.

The article is organized as follows: The second section provides general and detailed information on the methodology (variables, panel data, stochastic frontier functions and water index descriptions). The third section describes the main results of the estimates crop-water production functions for 8 main crops in the basin and water efficiency index. Finally, the fourth section presents the conclusions of the paper.

2 Methodology

Econometric stochastic frontier analysis has a lot of applications in many areas of the knowledge. In this study, it is applied to selected important crops in Ebro basin. We consider the most significant crops according their importance in the total agricultural area in the Ebro basin. Alfalfa, wheat, grapevine, olive, potato, maize and barley account for almost 60% of the total agricultural area in this region (MARM, 2007). However, we also include rice, which does not represent a large amount of the total cultivated area in the overall basin, but it is the main crop in the Ebro delta area and it is an intensively irrigated crop. It is important to keep in mind that alfalfa, maize, potato and rice are mainly irrigated while wheat, barley, grapevine and olive are primarily rainfed crops.

Currently, it is well known that natural resources, i.e. water, are very important to economic growth and environmental sustainability. An extended production function, well known as the Solow-Stiglitz model (Solow, 1974; Stiglitz, 1979), includes natural resources (R); it has proven useful to estimate the water requirements at different locations for selected crops. Also there is a lot of literature that these kinds of functions are useful to evaluate the effects of extreme contingencies among other biophysics and socioeconomic variables. The general form of this function is the next:

$$Y = L^{\alpha_1} K^{\alpha_2} R^{\alpha_3} \qquad \text{with } \alpha_1 + \alpha_2 + \alpha_3 = 1 \text{ y } \alpha_i > 0 \qquad (1)$$

Where L is labor; K is capital; R is the natural resources and $\alpha_1, \alpha_2, \alpha_3$ are the parameters which represents the elasticity of substitution among factors. We use Cobb-Douglas specification, as it allows a simple estimation and the coefficients obtained have a very intuitive interpretation in terms of elasticities. This function is not unique and varies among crops and zones and each approach to estimate production functions presents criticisms; however everyone has its strengths and limitations. The chapter is divided in 3 sections: (1) variable description and source of data, (2) Estimation of technical efficiency with panel data, and (3) The generation of the water efficiency index:

2.1. Variable description and sources of data

The variables, used here, are from regional, national and international sources of historical data for each crop in the 18 provinces of the Ebro river basin from 1976 to 2002. The dependent variable is the natural logarithm of the crop yield in a site i in the year t ($\ln Y_t$). Crop yield (Y) is defined as the ratio between production (in tones) and agricultural total area (in hectares). The source of data of these variables is the Statistical Division of the Spanish Ministry of Agriculture (MARM - Ministerio de Medio Ambiente y Medio Rural y Marino). As inputs, we used 4 categories of explanatory variables: management, water, climatic and socioeconomic variables.

Management variables were created to consider the effect of technology indicators, in this case we have incorporated the natural logarithm of irrigated area ($Irrig\_area_{it}$) and a linear combination of the diverse types of fertilizers and machinery like tractors and combines ($Comp1\_Tech_{it}$) (See Quiroga and Iglesias 2009; Iglesias and Quiroga 2007). $Irrig\_area_{it}$ is defined as the ratio between irrigated area and total crop land, by crop type. Data were obtained from the Spanish Ministry of Environment (MARM). In the other hand, fertilizers and machinery variables came from FAOSTAT of Food and Agriculture Organization of the United Nations (FAO). However, all these variables are highly correlated (see Table 1) and lead to problems of multicollinearity in the regression analysis. We used principal components analysis to solve this problem and we generated a new variable called $Comp1\_Tech_{it}$. The idea of using principal components in regression is not new. We can find literature since 1950s like Kendall, 1957; Hotelling, 1957; Jeffers, 1967. This technique consists in combine a large number of variables into a smaller number of related variables, retaining as much information as possible of the original variables (Blattberg, R., et al., 2008). Assuming an (n x k) matrix of X of n observations on k variables with $\sum$ variance-covariance matrix, the objective of principal components analysis involves an orthogonal transformation of a set of variables (k_1, k_2, …, k_n) into a set of components denoted by P, where P is (n x p) and $p \leq k$. These components are uncorrelated with each other, even though the original variables are quite highly correlated. It is good to say that after the analysis we will obtain the same number of components as original variables, and the total variance of the variables is preserved exactly in the total variance of the new components. The first principal component (p1) accounts for the highest proportion of the total variance, the second principal component (p2) reports for the largest share of the remaining variance, and so on (Blattberg, R., et al., 2008; Brook, G., et al., 1986; Jolliffe, I.; Jolliffe, I., 1982). In this case we only considered, in the regressions, components with an eigenvalue greater than 1; this value was selected given that a component with an eigenvalue less than 1 reports for less of the total variance than any single original variable. According to Table 2 only the first component has an eigenvalue greater than 1 (Var(Comp1) = 4.24939), which explains 85% of the variability of data. Complementarily, Figure 1 (screeplot) confirms the previous analysis. However, Jolliffe (1982) shows that there is a misconception about the principal components with small eigenvalues in a regression, and demonstrated that these components can be as important as those with large variance. The alternative approach to determine the number of components to use, is by using AIC, the Akaike's information criterion. This verifies the importance of the first component.

Table 1. Correlation matrix for technological variables

		Machinery		Fertilizers		
		Tractors	Combines	Nitrogen	Phosphate	Potash
Machinery	Tractors	1				
	Combines	0.9462	1			
Fertilizers	Nitrogen	0.7045	0.7758	1		
	Phosphate	0.6888	0.7232	0.8405	1	
	Potash	0.8958	0.8897	0.7587	0.89	1

Table 2. Principal components analysis: total variance explained

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	4.25	3.8	0.85	0.85
Comp2	0.45	0.22	0.09	0.94
Comp3	0.23	0.18	0.05	0.99
Comp4	0.05	0.02	0.01	1
Comp5	0.02		0	1

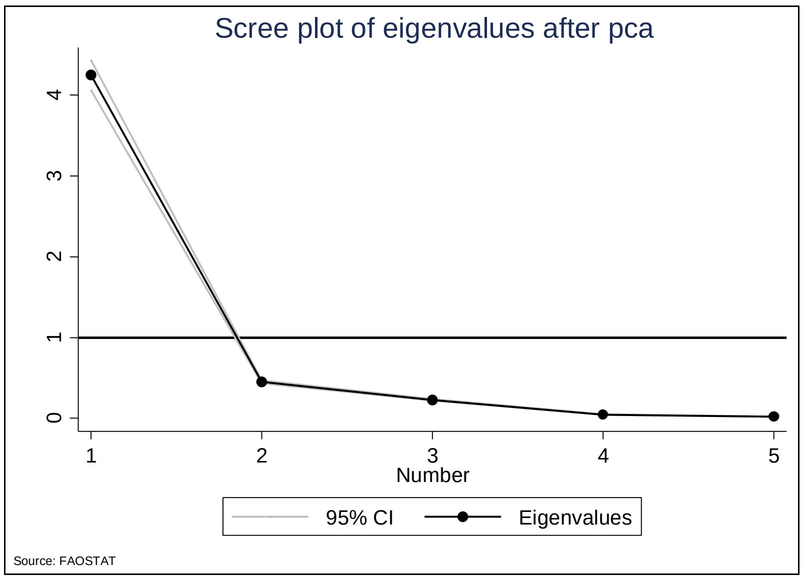

Figure 1 Significance of components using principal components analysis for technology variables

In the category of water variables, we consider precipitation (*Prec*$_{it}$) and water for irrigation (*Irrig*$_{it}$). *Prec*$_{it}$ is the total precipitation in mm in the *ith* site in a year *t*. It was taken from the Spanish Meteorological Agency (AEMET). To build a proxy variable for irrigation (*Irrig*$_{it}$), we used data on net crop water requirements from the Ebro basin management authority (CHEBRO, 2004). It is a good approximation given that currently there are no explicit restrictions on the irrigated area in the Ebro basin. We assume that water requirements of crops are being met.

Climate variables were taken from the Spanish Meteorological Agency (AEMET). In this case we used mean temperatures in degree Celsius (°C) per year in a *i* site (T_Mean$_{it}$). Finally, to characterize socioeconomic variables, we included the total employment of agricultural sector at a site *i* in year *t* in thousands of people (Labor$_{it}$). It was taken from the Labour Force Survey (LFS) of the Spanish National Institute of Statistics (INE). In Table 3, we summarize these variables.

Table 3. Summarized description of the variables

Type of variable	Name	Unit	Source of Data
Managment	$Comp1\_Tech_t$	Standarized units	Own elaboration from FAO data
	$Irrig\_area_t$	Per unit of crop land	MARM
Water	$Irrig_{it}$	m / month	CHEBRO
	$Prec_{it}$	mm / month	AEMET
Climate	T_Mean_{it}	° Celsius	AEMET
Socioeconomic	$Labor_{it}$	Thousands of people	INE

2.2. Estimating technical efficiency with panel data

Stochastic frontiers are an important step in estimating the technical efficiency with cross-sectional data, since they are able to incorporate stochastic modelling elements associated with any production process and also allow to decompose the random disturbance affecting production a symmetric component which includes factors beyond the control of the manager and an asymmetric component which includes systematic biases with respect to the frontier (inefficiency). However, the stochastic frontier has some limitations as: (1) the estimation of technical efficiency for each individual is not consistent, (2) the estimation of the model requires assumptions about the distribution of both components (symmetric and asymmetric) in the disturbance random, and these assumptions are often arbitrary, and (3) there is the assumption that technical efficiency is independent of the inputs, which has no microeconomic sense (Arias-Sampedro, 2001).

Given these problems, Schmidt and Sickles (1984) proposed the use of panel data as an option to solve some problems in the estimation of individual technical efficiency indexes with cross-sectional data. Also, Greene (1999) mention 2 reasons for the proliferation of studies using panel data: (1) panel data provide a rich environment for the development of estimation techniques and theoretical results; (2) from a practical standpoint, these data allow the estimation problems that cannot be studied in the context of time series or cross section, in example, the case for the unbiased estimation of technical efficiency.

The starting point for this analysis is a production model that can be represented by the following equation:

$$Y_{it} = \alpha + X'_{it}\beta + \varepsilon_{it} \qquad (2)$$

$$\varepsilon_{it} = v_{it} - u_i \qquad (3)$$

Where Y_{it} is the amount of output obtained, X_{it} is a vector of inputs, α and β are the parameters of the model is a random disturbance v_{it} independent and identically distributed with zero mean and constant variance σ_v^2, which represents the random factors beyond the control of producer, and u_i is a random disturbance with constant variance σ_u^2, which represents the individual inefficiency remains constant over time. The random variables v_{it} and u_i are considered independent.

According to Greene (2005), conventional panel data estimators (fixed and random effects estimators) suppose that technical inefficiency is time invariant and also they force any time invariant cross unit heterogeneity into the same term that is being employed to capture the inefficiency, given that, this inefficiency measures may have a heterogeneity problem. The Battese-Coelli maximum likelihood estimator (Battese and Coelli, 1988) extend the approach that Jondrow et al. (1982) did for cross section to panel data. They consider the maximum likelihood estimator (BC) of (2) and (3) which involves specification of the distributions of *v* and *u*. This formulation is frequently used in recent researches.

2.3. *Water efficiency index*

To aim this, we use a Cobb-Douglas production function as this, which include 3 input-variables to charaterize water use:

$$\ln Y_t = \alpha + \beta_1 \ln Labor_{it} + \beta_2 \mathrm{Comp1\_Tech}_{it} + \beta_3 \ln Irrig\_area_{it} + +\beta_4 \ln Irrig_t + \beta_5 \ln \mathrm{Pr}ec_{it} + \beta_6 T\_Mean_{it} + \varepsilon_{it}$$

To construct the water efficiency index, is better that the input-output are expressed in logarithms, then the technical efficiency of the *ith* farm is given by:

Using random effects:

$$TE_t^{GLS} = \exp(-u_i) = \exp(\alpha_i - \alpha)$$

One disadvantage is that GLS is biased if X is not independent of u.

Using Battese and Coelli estimator:

$$TE_i^{BC} = \exp(-u_i)$$

this measures the ratio of the *ith* province's production to what it would be if $u_i = 0$. Battese and Coelli's predictor for u_i conditional on estimates of ε_{it} is consistent as $T \rightarrow \infty$.

3 Results

Here we present the summarized results. According to Hallam and Machado (1996), the problem of multicollinearity - which generates a combination of insignificance and unexpected signs in the estimation of the coefficients although there is a good overall explanatory power - is not an important concern in the measurement of efficiency. Efficiency measure does not depends of the individual influence of the explanatory variables, it depends of the joint because it is based on the estimates of residuals

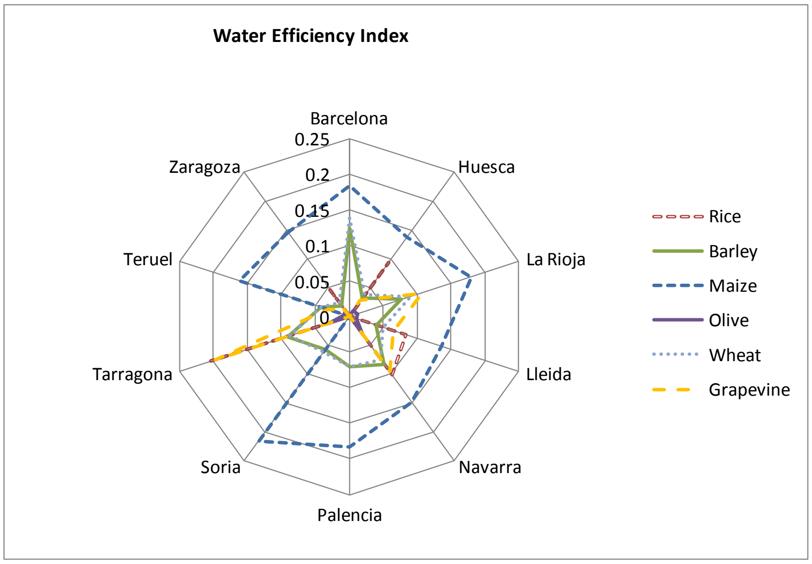

Figure 2 Technical efficiency using random effects estimator

Battesse and Coelli estimator the maximum efficiency is almost 0.94 for Alfalfa while in the GLS estimator, the same crop raise the maximum efficiency (in this case 1) for almost all the provinces of the study. Analyzing the results in both estimators, we can observe that there very similar and alfalfa and potato are the most efficient crops in almost all areas.

Table 4 Technical efficiency using Battesse and Coelly estimator

Crop	**Water index**
Alfalfa	0.89310581
Rice	0.05591807
Barley	0.04529622
Maize	0.14791077
Olive	0.00880517
Potato	0.43135603
Wheat	0.04984862
Grapevine	0.03801933
Period: 1976-2002 Provinces: 10	

4 Conclusions

Although here we do not use all the panel data estimators, we did the necessary test to the use of every one. It is well known the advantages of these estimators in the measurement of the technical efficiency. Analyzing the results in both estimators, we can observe that there very similar. In Battesse and Coelli estimator the maximum efficiency is almost 0.94 for Alfalfa while in the GLS estimator, the same crop raise the maximum efficiency (in this case 1) for almost all the provinces of the study. Then the choice of estimator does appear to influence the estimate of average technical efficiency and it proves some past studies. Finally, we can observe that given water inputs potato and alfalfa tend to be more efficient.

Acknowledgements

This research has been supported by the European Commission CIRCE project and the ARCO project of the Spanish Ministry of Environment, Rural, and Marine Affairs (MARM). We also acknowledge the data support of Agencia Estatal de Meteorología (AEMET).

5 References

AEMET (2010): Time series of temperature and precipitation, Years: 1961-2009. Spanish Meteorological Agency

Blattberg. R. (2008) International series in quantitative marketing: Database Marketing. Ed. Springer. Cap. 11 pp. 297-298

Brook, G.; Chin-hong Sun; and Carver, R. E., (1986). Predicting well productivity using principal components analysis. Professional Geographer. 38(4): 324-3

CHEBRO (2004): Revisi´on de las Necesidades H´ıdricas Netas de los Cultivos en la Cuenca del Ebro, 1961–2002, Confederación Hidrogr´afica del Ebro, Zaragoza, Spain.

FAOSTAT: Resources (Machinery), Years: 1961–2007, Food and Agriculture Organization of the United Nations (FAO), 2009.

Iglesias A, Quiroga S (2007). Measuring the risk of climate variability to cereal production at five sites in Spain. Climate Research, 34, 45 – 57.

INE (2009): Statistical Yearbook of Spain and Labour Force Survey (LFS), Years: 1976–2008, National Institute of Statistics, Madrid, Spain, http://www.INE.es.

Jeffers, J. N. R. (1967). Two case studies in the application of principal component analysis. Applied Statistics, 16: 225- 236.

Jolliffe, I., (1982). A Note on the Use of Principal Components in Regression. Applied Statistics. 31 (3): 300-303

Jolliffe, I., (1986). Principal Component Analysis. Springer-Verlag New York Inc.

Kendall, M. (1957). A Course in Multivariate Analysis. Ed. Griffin. London

MARM (2007): Anuarios de Estad´ıstica Agroalimentaria, Years: 1976–2007, Spanish Ministry of Environment and Rural and Marine, Statistical Division, Madrid.

Quiroga, S. and Iglesias, A.(2009): A comparison of the climate risks of cereal, citrus, grapevine and olive production in Spain, Agr. Syst., 101, 91–100.