ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 2/2015

A parametric frontier model for measuring eco-efficiency

Luis Orea, Alan Wall

Departamento de Economía

Universidad de Oviedo

Available online at: http://economia.uniovi.es/investigacion/papers

A parametric frontier model for measuring eco-efficiency

Luis Orea

Alan Wall

Oviedo Efficiency Group, School of Economics and Business, University of Oviedo[1]

Abstract

Eco-efficiency has been defined by the OECD as "the efficiency with which ecological resources are used to meet human needs" and can be considered a measure of environmental performance that takes into account both the environmental and economic objectives of firms. Frontier models are an ideal tool for measuring eco-efficiency. While the literature applying frontier models to the empirical measurement of eco-efficiency has been growing steadily in recent years, it has exclusively relied on non-parametric Data Envelopment Analysis (DEA) methods to measure eco-efficiency and its determinants. We propose a parametric Stochastic Frontier (SF) model to measure eco-efficiency, arguing that this has several potential advantages. We provide an empirical application using cross-sectional data from Spanish dairy farms which includes information on environmental and economic indicators as well as a series of potential socio-economic determinants of eco-efficiency.

[1] Correspondence to: Alan Wall, Oviedo Efficiency Group, Department of Economics, School of Economics and Business, University of Oviedo, 33006 Oviedo, Spain.
email: awall@uniovi.es.

1. Introduction

Concerns about the sustainability of economic activity has led to an increasing interest in the concept of eco-efficiency and the literature on this topic has been growing in recent years (Oude Lansink and Wall, 2014). The term eco-efficiency was originally coined by the World Business Council for Sustainable Development in their 1993 report (Schmidheiney, 1993) and is based on the concept of creating more goods and services using fewer resources. In turn, the OECD defines eco-efficiency as "the efficiency with which ecological resources are used to meet human needs" (OECD, 1998). Clearly, the concept of eco-efficiency takes into account both the environmental and economic objectives of firms.

When evaluating firm performance in the presence of adverse environmental impacts, production frontier models are a popular tool (Tyteca, 1996; Lauwers, 2009; Picazo-Tadeo et al., 2011; Pérez-Urdiales et al., 2015). The measurement of eco-efficiency in a frontier context, which Lauwers (2009) refers to as the "frontier operationalisation" of eco-efficiency, involves comparing economic results derived from the production of goods and services with aggregate measures of the environmental impacts or 'pressures' generated by the production process. To date, only the non-parametric Data Envelopment Analysis (DEA) method has been used in the literature. While DEA has many advantages, it has the drawback that it can be extremely sensitive to outliers and measurement errors in the data.

In the present work we propose a Stochastic Frontier Analysis (SFA) approach to measuring eco-efficiency, which has the advantage that it well-suited to dealing with measurement errors in the data. Using a stochastic frontier model to measure eco-efficiency involves the estimation of only a few parameters, so the model can be implemented even when the number of observations is relatively small. Moreover, the SFA approach permits an analysis of the potential substitutability between environmental pressures, and can easily be extended to incorporate determinants of eco-efficiency in a one-stage procedure.

We illustrate our simple proposal with an empirical application using a sample of 50 dairy farmers from the Spanish region of Asturias. This data set includes information from a questionnaire specifically carried out to permit the accurate measurement of eco-efficiency and provides information on farmers' socioeconomic characteristics and attitudes towards the environment, and has been used by Pérez-Urdiales et al. (2015) to measure eco-efficiency and identify its determinants using the DEA-based bootstrapped

truncated regression techniques of Simar and Wilson (2007). The results from that paper therefore provide a useful point of comparison for the results from our proposed stochastic frontier model.

The paper proceeds as follows. In Section 2 we discuss the concept of eco-efficiency and the DEA approach often used to estimate eco-efficiency scores. Section 3 introduces our stochastic frontier model, which can be viewed as a counterpart of the DEA eco-efficiency model. Section 4 describes the data we use. The results are presented and discussed in Section 5, and Section 6 concludes.

2. Background

To measure eco-efficiency using frontiers, Kuosmanen and Kortelainen (2005) defined eco-efficiency as a ratio between economic value added ('profit') and environmental damage and proposed a pressure-generating or pollution-generating technology set $T = \{(\pi,\ p) \in R^{(1+K)} | \pi \text{ can be generated by } p\}$. This technology set describes all the feasible combinations of profit, π, and environmental pressures, p. Environmental damage, $D(p)$, is measured by aggregating the *K* environmental pressures $(p_1, \dots, p_K)$ associated with the production activity.

Figure 1 provides an illustration for the simple case of two environmental pressures, p_1 and p_2. The set of eco-efficient combinations is represented by the eco-efficient frontier, which represents the minimum combinations of the two environmental pressures which can be used to produce a profit of π_0. Combinations of pressures below the frontier are unfeasible whereas combinations above it, such as point A, are eco-inefficient. The combination of pressures represented by point A is clearly eco-inefficient as the environmental pressures could be reduced equiproportionally to point E on the frontier without reducing profit.[2]

Eco-inefficiency can be measured using the radial distance from a point A to the efficient frontier. The eco-efficiency score is given by the ratio OE/OA which takes the value 1 for

[2] Instead of measuring the extent to which pressures can be reduced while maintaining profit, we could measure the extent to which the firm, given its present combination of pressures, could increase its profit. Thus, if the firm was using the combination of pressure represented by A efficiently, it would be operating on a new eco-efficient frontier passing through that point, and could achieve a higher profit corresponding to this new frontier. Answering either of these questions will uncover eco-inefficient behaviour but we will follow the existing literate by focusing on the capacity of firms to reduce environmental pressures while maintaining profit.

eco-efficient combinations of pressures and economic profit and values less than 1 for inefficient combinations such as A.

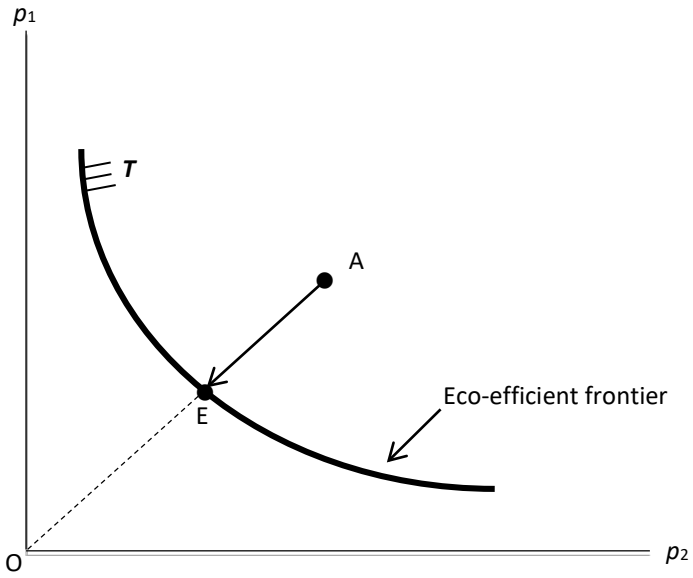

Figure 1. Eco-efficiency

Individual eco-efficiency scores for producer i can be found using the following expression:

$$EEF_i = \frac{\text{Economic value added}}{\text{Environmental pressure}} = \frac{\pi_i}{D_i(p)} \tag{1}$$

where $D_i(p)$ is a function that aggregates the environmental pressures into a single environmental pressure indicator. This can be done by taking a linear weighted average of the individual environmental pressures:

$$D_i(p) = w_1 p_{1i} + w_2 p_{2i} + \cdots + w_K p_{Ki} \tag{2}$$

where w_k is the weight assigned to environmental pressure p_k. Kuosmanen and Kortelainen (2005) and Picazo-Tadeo et al. (2012), among others, use DEA as a non-subjective weighting method.

The DEA eco-efficiency score of firm *i* can be computed from the following programming problem:[3]

$$\max_{w_{ki}} EEF_i = \frac{\pi_i}{\sum_{k=1}^{K} w_{ki} p_{ki}} \tag{3}$$

[3] This formulation involves a non-linear objective function and non-linear constraints, which is computationally difficult. This problem is often linearized by taking the inverse of the eco-efficiency ratio and solving the associated reciprocal problem (Kuosmanen and Kortelainen 2005; Picazo-Tadeo et al. 2011).

subject to the constraints

$$\frac{\pi_j}{\sum_{k=1}^{K} w_{ki} p_{kj}} \leq 1 \;\; j = 1, \ldots, N$$

$$w_{ki} \geq 0 \quad k = 1, \ldots, K$$

The above two constraints force weights be non-negative and eco-efficiency scores take values between zero and one, that is:

$$EEF_i = \frac{\pi_i}{\sum_{k=1}^{K} w_{ki} p_{ki}} \leq 1 \quad , \forall i = 1, \ldots, \mathrm{N} \tag{4}$$

The DEA eco-efficiency score which solves this problem for firm *i* indicates the maximum potential equiproportional reduction in all environmental pressures that could be achieved while maintaining economic profit constant, i.e., it corresponds to the ratio OE/OA for a firm operating at point A in Figure 1 and would take the vaue 1 for an eco-efficient firm.

3. The SFA eco-efficiency model

In this section we introduce our SFA counterpart of the above DEA eco-efficiency model. We first introduce a basic (i.e. homoskedastic) specification of the model in order to focus our attention on the main characteristics of the model and the differences between the SFA and DEA approaches. We then present a heteroskedastic specification of the model that allows us to identify determinants of firms' eco-efficiency in a simple one-stage procedure. Finally, we explain how we obtain the estimates of eco-efficiency for each farm.

3.1. Basic specification

In this section we introduce our SFA counterpart of the DEA eco-efficiency model. Our approach relies on the constraint in equation (4). If we assume that the environmental pressure weights, w_k, in (2) are parameters to be estimated, we can impose that these be positive by reparameterizing them as $w_k = e^{\beta_k}$. The natural logarithm of equation (4) can be written as:

$$\ln EFF_i = \ln\left(\frac{\pi_i}{\sum_{k=1}^{K} e^{\beta_k} \cdot p_{ki}}\right) \leq 0 \tag{5}$$

The above equation can be rewritten as:

$$\ln(\pi_i) = \ln\left(\sum_{k=1}^{K} e^{\beta_k} \cdot p_{ki}\right) - u_i \quad (6)$$

where $u_i = -\ln EFF_i \geq 0$ can now be viewed as a non-negative random term capturing firm *i*'s eco-inefficiency.

Equation (6) is a non-linear regression model with a nonpositive disturbance that can be estimated using several techniques (see Kumbhakar and Lovell, 2000, section 3.2.1.). Regardless of the technique, however, note that in (6) we are measuring farms' eco-efficiency relative to a deterministic environmental pressure frontier. This implies that all variation in profit not associated with variation in individual environmental pressures is entirely attributed to eco-inefficiency. In other words, this specification does not make allowance for the effect of random shocks, which might also contribute (positively or negatively) to variations in economic value.

As is customary in the SFA literature in production economics, in order to deal with this issue we extend the model in (6) by adding a symmetric random noise term, v_i, and a non-zero intercept θ:

$$\ln(\pi_i) = \theta + \ln\left(\sum_{k=1}^{K} e^{\beta_k} \cdot p_{ki}\right) + v_i - u_i \quad (7)$$

This model is more complex than a deterministic eco-efficiency frontier model but it is also more realistic as deviations from the frontier due not only to eco-inefficiency but also to uncontrollable or unobservable factors (i.e. random noise) are incorporated. We have also added a non-zero intercept in order to obtain unbiased parameter estimates in case the unobservable factors or measurement errors have a level effect on firms' profit.

The error term in (7) thereby comprises two independent parts. The first part, v_i, is a two-sided random noise term, often assumed to be normally distributed with zero mean and constant standard deviation, i.e. $\sigma_v = e^{\gamma}$. The second part, u_i, is a one-sided error term capturing underlying eco-inefficiency that can vary across firms and over time. Following Aigner *et al.* (1977) it is often assumed to follow a half-normal distribution, which is the truncation (at zero) of a normally-distributed random variable with mean zero. Moreover, these authors also assumed that the variance of the pre-truncated normal variable (hereafter σ_u) is *homoskedastic* and common to all farms, i.e. $\sigma_u = e^{\delta}$. The identification of both random terms in this model (ALS henceforth) relies on the asymmetric and one-sided distribution of u_i. If the inefficiency term could take both positive and negative values, it would not be distinguishable from the noise term, v_i.

It should be pointed out that under these distributional assumptions the density function of the composed error term $\varepsilon_i = v_i - u_i$ in (7) is the *same* as the well-known density function of a standard normal-half normal frontier model. Following Kumbhakar and Lovell, (2000, p. 77), the log likelihood function for a sample of N producers can then be written as:

$$lnL(\theta,\beta,\gamma,\delta) = \frac{-N}{2} ln[\sigma_v^2 + \sigma_u^2] + \sum_{i=1}^{N} ln\Phi\left[-\frac{\varepsilon_i(\theta,\beta)\cdot\sigma_u/\sigma_v}{(\sigma_v^2+\sigma_u^2)^{1/2}}\right] - \frac{1}{2(\sigma_v^2+\sigma_u^2)}\sum_{i=1}^{N}\varepsilon_i^2(\theta,\beta) \quad (8)$$

where $\beta = (\beta_1, \dots, \beta_K)$, and

$$\varepsilon_i(\theta,\beta) = \ln(\pi_i) - \theta - \ln\left(\sum_{k=1}^{K} e^{\beta_k}\cdot p_{ki}\right) \quad (9)$$

The likelihood function (8) can be maximized with respect to $(\theta,\beta,\gamma,\delta)$ to obtain consistent estimates of all parameters of our eco-efficiency model. The only difference between our SFA *eco-efficiency* model and a traditional SFA *production* model is the computation of the error term $\varepsilon_i(\theta,\beta)$. In a traditional SFA production model, this is a simple linear function of the parameters to be estimated and hence the model can be estimated using standard econometric software, such as Limdep or Stata. In contrast, $\varepsilon_i(\theta,\beta)$ in equation (9) is a non-linear function of the β parameters. Although the non-linear nature of equation (9) prevents using the standard commands in Limdep or Stata to estimate our SFA eco-efficiency model, it is relatively straightforward to write the codes to maximize (8) and obtain our parameter estimates.

The model in (7) can be also estimated using a two-step procedure. In the first stage, the intercept θ and the environmental pressure parameters β of equation (7) can be estimated using a non-linear least squares estimator. In the second step, the aforementioned distributional assumptions regarding the error terms are made to obtain consistent estimates of the parameters describing the variance of v_i and u_i (i.e., γ and δ) conditional on the estimated parameters from the first step. This two-step approach is advocated for various models in Kumbhakar and Lovell (2000). The main advantage of this two-step procedure is that no distributional assumptions are used in the first step. Standard distributional assumptions on v_i and u_i are used only in the second step. In addition, in the first step the error components are allowed to be freely correlated.

An important issue that should be taken into account when using a two-step procedure is that the expectation of the original error term in (7) is not zero because u_i is a non-

negative random term. This implies that the estimated value of the error term ε_i in equation (7) should be decomposed as follows:

$$\varepsilon_i = v_i - u_i + E(u_i) \tag{10}$$

If u_i follows a half-normal distribution, then $E(u_i) = \sqrt{2/\pi} \cdot \sigma_u$. Thus, the stochastic frontier model in the second stage is:

$$\varepsilon_i = v_i - u_i - \sqrt{2/\pi} \cdot e^{\delta} \tag{11}$$

Note that there are no new parameters to be estimated. The parameters γ and δ are estimated by maximizing the likelihood function associated to this (adjusted) error term.[4] As Kumbhakar et al. (2013) have recently pointed out, the stochastic frontier model based on (11) can accommodate heteroskedastic inefficiency and noise terms simply by making the variances of σ_u and σ_v functions of some exogenous variables (see, for instance, Wang, 2002; Álvarez et al., 2006). This issue is addressed later on.

Compared to the DEA eco-efficiency model, our SFA approach will attenuate the effect of outliers and measurement errors in the data on the eco-efficiency scores. Moreover, it is often stressed that the main advantage of DEA over the SFA approach is that it does not require an explicit specification of a functional form for the underlying technology. However, the "technology" here is a simple index that aggregates all environmental pressures into a unique value. Thus, we would expect that the parametric nature of our SFA approach is not as potentially problematic in an eco-efficiency analysis as it may be in a more general production frontier setting where theses techniques are used to uncover the underlying (and possibly quite complex) relationship between multiple inputs and outputs. Another often-cited advantage of the DEA approach is that it can be used when the number of observations is relatively small. We reiterate, however, that the

[4] An alternative second-stage approach based on the Method of Moments (MM) can also be used. This empirical strategy relies on the second and third moments of the error term ε_i in equation (7). This approach takes advantage of the fact that the second moment provides information about both σ_v and σ_u whereas the third moment only provides information about the asymmetric (one-sided) random inefficiency term. Olson *et al*. (1980) showed using simulation exercises that the choice of estimator (ML versus MM) depends on the relative value of the variances of both random terms and the sample size. When the sample size is large and the variance of the one-sided error component is small compared to the variance of the noise term, ML outperforms MM. The MM approach has, in addition, some practical problems. It is well known in the stochastic frontier literature, for example, that neglecting heteroskedasticity in either or both of the two random terms causes estimates to be biased. Kumbhakar and Lovell (2000) pointed out that only the ML approach can be used to address this problem. Another practical problem arises in homoscedastic specifications of the model when the implied σ_u becomes sufficiently large to cause σ_v<0, which violates the assumptions of the econometric theory.

"technology" of our SFA model is extremely simple, with few parameters to be estimated, so that the model can be implemented even when the number of observations is not large.

Finally, note that the estimated β parameters have an interesting interpretation in the parametric model. In the expression for eco-efficiency in (1), we note that eco-efficiency is constant and equal to 1 along the eco-efficiency frontier. Differentiating (1) with respect to an individual pressure p_k for farm *i* we obtain:

$$\frac{\partial D_i(p)}{\partial p_k} = \frac{\partial \pi_i}{\partial p_k}\frac{D_i(p)}{\pi_i} \tag{12}$$

For any two pressures p_j and p_k, therefore, we have:

$$\frac{\frac{\partial D_i(p)}{\partial p_j}}{\frac{\partial D_i(p)}{\partial p_k}} = \frac{\frac{\partial \pi_i}{\partial p_j}}{\frac{\partial \pi_i}{\partial p_k}} \tag{13}$$

From the expression for eco-efficiency in the reparameterized model in (5) it is clear that $\frac{\partial D_i(p)}{\partial p_k} = e^{\beta_k}$, so that in this particular case (13) becomes:

$$\frac{e^{\beta_j}}{e^{\beta_k}} = \frac{\partial \pi_i / \partial p_j}{\partial \pi_i / \partial p_k} \tag{14}$$

Once the β parameters have been estimated, e^{β_k} therefore represents the marginal contribution of pressure p_k to firm *i*'s profit, i.e., it is the monetary loss in profits if pressure p_k were reduced by one unit.

As expression (14) represents the *marginal rate of technical substitution of environmental pressures*, it provides valuable information on the possibilities for substitution between pressures. If this marginal rate of substitution took a value of 2, say, we could reduce pressure p_j by two units and increase p_k by one unit without changing economic profit. This also sheds light on the consequences for firms of legislation requiring reductions in individual pressures. Continuing with the previous example, it would be relatively less onerous for the firm to reduce pressure p_k rather than p_j as the fall in profit associated with a reduction in p_k would be only half that which would occur from a reduction in p_j.

3.2. Heteroskedastic specification

Aside from measuring firms' eco-efficiency, we also would like to analyse the *determinants* of eco-efficiency. The concern about the inclusion of contextual variables or z-variables has generated the development of several models either using parametric, non-parametric or semi-parametric techniques.[5] The inclusion of contextual variables in DEA has been carried out in one, two or even more stages. Ruggiero (1996) and other authors have highlighted that the one-stage model introduced in the seminal paper of Banker and Morey (1986) might lead to bias. To solve this problem, other models using several stages have been developed in the literature. Ray (1988) was the first who proposed a second stage where standard DEA efficiency scores were regressed on a set of contextual variables. This practice was widespread until Simar and Wilson (2007) demonstrated that this procedure is not consistent because the first-stage DEA efficiency estimates are serially correlated. These authors proposed a bootstrap procedure to solve this problem in two stages which has perhaps become the most-widely used method in DEA to identify inefficiency determinants.

As the inefficiency term in the ALS model has constant variance, our SFA model in (7) does not allow the study of the determinants of farms' performance.[6] To deal with this, we could also estimate a *heteroskedastic* frontier model that incorporates z-variables into the model as eco-efficiency determinants. The specification of u_i that we consider in this paper is the so-called RSCFG model (see Alvarez et al., 2006), where the z-variables are treated as determinants of the variance of the pre-truncated normal variable. In other words, in our frontier model we assume that

$$\sigma_{ui} = h(z_i) \cdot \sigma_u \tag{15}$$

where

$$h(z_i) = e^{\alpha' z_i} \tag{16}$$

is a deterministic function of eco-inefficiency covariates, $\alpha = (\alpha_1, \dots, \alpha_J)$, is a vector of parameters to be estimated, and $z_i = (z_{i1}, \dots, z_{iJ})$ is a set of *J* potential determinants of

5 For a more detailed review of this topic in SFA and DEA, see Johnson and Kuosmanen (2011, 2012).

6 It might also yield biased estimates of both frontier coefficients and farm-specific eco-inefficiency scores (see Caudill and Ford, 1993).

firms' eco-inefficiency.[7] This specification of σ_{ui} nests the *homoskedastic* model as (15) colapses into e^{δ} if we assume that $h(z_i) = 1$ or α=0.

The log likelihood function of this model is the same as equation (8), but now σ_{ui} is *heteroskedastic* and varies across farms. The resulting likelihood function should then be maximized with respect to $\theta, \beta, \gamma, \delta$ and α to obtain consistent estimates of all parameters of the model. As both frontier parameters and the coefficients of the eco-inefficiency determinants are simultaneously estimated in one stage, the inclusion of contextual variables in our SFA model is much simpler than in DEA.

3.3. Eco-efficiency scores

We next discuss how we can obtain the estimates of eco-efficiency for each firm once either the homoskedastic or heteroskedastic model has been estimated. In both specifications of the model, the composed error term is simply $\varepsilon_i = v_i - u_i$. Hence, we can follow Jondrow *et al.* (1982) and use the conditional distribution of u_i given the composed error term ε_i to estimate the asymmetric random term u_i. Both the mean and the mode of the conditional distribution can be used as a point estimate of u_i. However, the conditional expectation $E(u_i|\varepsilon_i)$ is by far the most commonly employed in the stochastic frontier analysis literature (see Kumbhakar and Lovell, 2000).

Given our distributional assumptions, the analytical form for $E(u_i|\varepsilon_i)$ can be written as follows:

$$E(u_i|\varepsilon_i) = \bar{\mu}_i + \bar{\sigma}_i \left[\frac{\phi(-\bar{\mu}_i/\bar{\sigma}_i)}{1-\Phi(-\bar{\mu}_i/\bar{\sigma}_i)}\right] \tag{17}$$

where

$$\sigma_i^2 = \sigma_v^2 + h(z_i)^2\sigma_u^2$$

$$\bar{\mu}_i = \frac{\varepsilon_i h(z_i)^2 \sigma_u^2}{\sigma_i^2}$$

[7] The so-called scaling property (Alvarez et al., 2006) is satisfied in this heteroskedastic version of our SFA model in the sense that the inefficiency term in (7) can be written as $u_i = h(z_i) \cdot u_i^*$, and $u_i^* \rightarrow N^+(0, e^{\delta})$ is a one-sided random variable that does not depend on any eco-efficiency determinant. The defining feature of models with the scaling property is that firms differ in their mean efficiencies, but not in the shape of the distribution of inefficiency. In this model u_i^* can be viewed as a measure of "basic" or "raw" inefficiency that does not depend on any observable determinant of firms' inefficiency.

$$\bar{\sigma}_i = \frac{h(z_i)\sigma_u\sigma_v}{\sigma_i}$$

To compute the conditional expectation (17) using the heteroskedastic model, we should replace the deterministic function $h(z_i)$ with our estimate of (16), while for the homoskedastic model we should assume that $h(z_i) = 1$.

4. Data

The data we use come from a survey which formed part of a research project whose objective was to analyse the environmental performance of dairy farmers in the Spanish region of Asturias. Agricultural activity has well-documented adverse effects on the environment and the increasing concerns among policymakers about environmental sustainability in the sector are reflected in the recent Common Agricultural Policy (CAP) reforms in Europe. Dairy farming, through the use of fertilizers and pesticides in the production of fodder as well as the emission of greenhouse gases has negative consequences for land, water, air, biodiversity and the landscape, so it is of interest to see whether there is scope for farmers to reduce environmental pressures without profits being reduced and identify any farmer characteristics that may influence their environmental performance.

A questionnaire was specifically designed to obtain information on individual pollutants, including nutrients balances and greenhouse gas emissions. These individual pollutants were then aggregated using standard conversion factors into a series of environmental pressures. Questions were included regarding farmers' attitudes towards aspects of environmental management as well as a series of socioeconomic characteristics. The data collected correspond to the year 2010.

A total of 59 farmers responded to the questionnaire and the environmental and socioeconomic data were combined with economic data for these farmers which is gathered annually through a Dairy Cattle Management Program run by the regional government. Given that there were missing values for some of the variables we wished to consider, the final sample comprised 50 farms.

These data were used by Pérez-Urdiales et al. (2015) to measure the farmers' eco-efficiency and relate it to attitudinal and socioeconomic factors. These authors used the two-stage DEA-based bootstrapped truncated regression technique proposed by Simar and Wilson (2007) to estimate eco-efficiency and its determinants, finding evidence of considerable eco-inefficiency. We will use the same variables as Pérez-Urdiales et al. (2015) to estimate eco-efficiency and its determinants using the SFA methods proposed in the previous section, which will permit us to see whether the SFA model yields similar results. We will use the results from Pérez-Urdiales et al. (2015) as a reference for comparison but it should be stressed that the dataset is far from ideal for using a SFA approach. In particular, the number of observations is relatively small and there are several determinants of eco-efficiency whose parameters have to be estimated.

The variables are described in detail in Pérez-Urdiales et al. (2015) but we will briefly discuss them here. For the numerator of the eco-efficiency index, economic value added (*Econvalue*) is the difference between revenues from milk production (including milk sales and the value of in-farm milk consumption) and direct costs. The environmental pressures comprise nutrients balances and greenhouse gas emissions. The nutrients balances measure the extent to which a farm is releasing nutrients into the environment, defined as the difference between the inflows and outflows of nutrients. The nutrients balances used are nitrogen (*SurplusN*), phosphorous (*SurplusP*) and potassium (*SurplusK*), all measured in total kilograms. The volume of greenhouse gas emissions captures the contribution of the farm to global warming and the dataset contains information on the emissions of carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O). Each of these greenhouse gases is converted into CO_2 equivalents, so that the variable used is (thousands of) kilos of carbon dioxide released into the atmosphere (*CO_2*).

The second set of variables are the potential determinants of eco-efficiency, which comprises socioeconomic characteristics and attitudes of farmers. The socioeconomic variables are the age of the farmer (*Age*); the number of hours of specific agricultural training that the farmer received during the year of the sample (*Training*); and a variable capturing the expected future prospects of the farm and which is defined as a dummy variable taking the value 1 if the farmer considered that the farm would continue to be in operation five years later, and 0 otherwise (*Prospects*). As explained in Pérez-Urdiales et al. (2015), eco-efficiency would be expected to be negatively related to age (i.e., older farmers should be less eco-efficient) and positively related to professional training and the expectation that the farm continue.

Three attitudinal variables were constructed from responses to a series of questions on farmers' beliefs regarding their management of nutrients and greenhouse gas emissions as well as their attitudes towards environmental regulation. Thus, on a five-point Likert scale respondents had to state whether they strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4) or strongly agree (5), with a series of statements regarding their habits and attitudes towards environmental management. The variables *HabitsCO_2* and *HabitsNutrients* are constructed as dummy variables that take the value 1 if respondents stated that they agreed or strongly agreed that management of grenhouse gases and nutrients was important, and 0 otherwise. The final variable measuring attitudes towards environmental regulation, defined as a dummy variable

taking the value 1 if respondants agreed or strongly agreed that environmenatl regulation should be made more restrictive and 0 otherwise (*Regulation*).

Some descriptive statistics of the variables used for measuring eco-efficiency and the determinants of estimated eco-efficiency are presented in Table 1.

Table 1. Descriptive statistics of variables

Variable	Description	Mean	S.Dev.
Econvalue	Value added (€)	77137	40423
Environmental pressures			
SurplusN	Nitrogen surplus (kg)	5966	4705
SurplusP	Phosphorous surplus (kg)	2770	2168
SurplusK	Potassium surplus (kg)	2096	1681
CO_2	Greenhouse gases ('000s kg)	427	142
Eco-efficiency determinants			
HabitsCO_2	Attitude towards greenhouse gas management	0.09	0.29
HabitsNutrients	Attitude towards nutrient management	0.77	0.43
Age	Age of head of household	45.98	7.97
Prospects	Continuity of farm	0.98	0.14
Regulation	Attitude towards regulation	0.58	0.50
Training	Hours of specific training in last year	45.14	63.10

5. Results

We focus initially on the results from the stochastic frontier models and then on the comparison of these with the DEA results.

Table 2 presents estimates from different specifications of the homoskedastic (ALS) and heteroskedastic (RSCFG) stochastic eco-efficiency frontier, with their corresponding eco-efficiency scores presented in Table 3. Columns (A) and (B) of Table 2 report estimates from the ALS model with all environmental pressures included and it can be seen that all the estimated coefficients on the pressures were highly significant. The parameter δ corresponding to $\ln \sigma_u$ was also highly significant, implying that the frontier specification is appropriate.

Table 2. Parameter estimates

		ALS-50A		**ALS-50B**		**ALS-50C**		**ALS-40C**		**RSCFG-40C**	
		(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)
Variable	*Parameter*	*Estimates*	*t-stat*	*Estimates*	*t-stat*	*Estimates*	*t-stat*	*Estimates*	*t-stat*	*Estimates*	*t-stat*
Intercept	θ	-0.15	-1.65	1.42	6.41	2.06	21.79	2.06	21.26	1.96	20.62
SurplusN	β_1	1.59	10.78	0.03	0.02						
SurplusP	β_2	-5.25	-37.11								
SurplusK	β_3	2.24	16.05	0.66	0.31	0.93	7.04	1.17	8.05	1.27	8.63
CO_2	β_4	5.30	42.89	3.73	16.53	3.12	25.13	2.89	20.06	2.91	20.41
ln σ_v	γ	-2.15	-15.19	-2.15	-4.63	-2.26	-3.68	-2.64	-4.67	-2.45	-15.52
ln σ_u	δ	-0.47	-3.41	-0.47	-3.47	-0.45	-2.91	-0.53	-3.92	-0.32	-2.00
HabitsCO_2	α_1									-1.06	-6.71
HabitsNutrients	α_2									-0.21	-1.33
Age	α_3									0.01	1.64
Prospects	α_4									-0.64	-4.06
Regulation	α_5									0.27	1.71
Training	α_6									0.00	-1.35
SurplusN	exp(β_1)	4.90	6.79	1.03	7.15						
SurplusP	exp(β_2)	0.01	7.07								
SurplusK	exp(β_3)	9.37	7.17	1.93	7.10	2.55	7.67	3.22	6.81	3.55	6.80
CO_2	exp(β_4)	200.91	8.09	41.85	8.31	22.74	8.56	17.94	6.99	18.43	7.00
Mean log-likelihood		-0.3999		-0.3999		-0.4039		-0.2867		-0.1744	
Observations		50		50		50		40		40	

As the pressure function parameters β_k enter the eco-efficiency specification exponentially rather than linearly (5), in the bottom part of Table 2 the exponents of the coefficients are presented. The t-statistics here correspond to the null that e^{β_k} is equal to zero for each of the *k* pressures, and this is rejected in all cases.

However, focusing on the magnitudes rather than the statistical significance, it can be seen that the marginal contribution of the phosphorous balance to profit is almost negligible. Also, the value of e^{β_k} for potassium is almost twice as large as that of nitrogen. Recalling our discussion of the interpretation of these parameters after equation (14) above, this implies that potassium contributes twice as much to economic value as nitrogen and would therefore be more costly for the farmer to reduce. Similarly, if farmers were required to reduce nitrogen, this could in principle be substituted by potassium: for a given reduction in kilos of nitrogen, farmers could increase their use of potassium by half this number of kilos and maintain the same profit. In this particular application, such substitution could be achieved through changes in the composition of feed, fertilizers, and a change in the composition of forage crops. Reducing phosphorous, on the other hand, would be virtually costless.

In light of the negligible contribution of phosphorous to economic value, we reestimate the ALS model eliminating the phosphorous balance from the pressure function, and the results are presented in columns (C) and (D). The parameters on the nutrients are not significantly differently from 0, implying that the e^{β_k} are not significantly different from 1. Note that the frontier specification is still appropriate and a comparison of the the efficiency scores from the two models in Table 3 shows that are practically identical.

We now turn to the heteroskedastic (RSCFG) specification of the stochastic frontier where we incorporate the determinants of eco-efficiency described in the previous section. Some of the farms had missing values for one or more of these determinants, and after eliminating these observations we were left with 40 farms with complete information. When estimating the model for these 40 observations with all nutrients balances included, it did not converge. We then eliminated the phosphorous balance as we had done in columns (C) and (D) for the homoskedastic (ALS) specification, but the model still did not converge. Following our earlier strategy of eliminating the nutrient balance with the lowest marginal contribution, from column (A) we see that the nitrogen balance has a far lower marginal contribution than the potassium balance. We therefore specified the model without the nitogen balance, keeping only the potassium balance and grenhouse gas emissions as pressures. With this specification the model converged successfully and the results are reported in columns (I) and (J).

Table 3. Eco-efficiency scores

Farm	DEA-50	DEA-40	ALS-50A	ALS-50B	ALS-50C	ALS-40C	RSCFG-40C
1	0.615	0.891	0.682	0.683	0.664	0.802	0.878
2	0.562	n.a	0.627	0.627	0.608	n.a	n.a
3	0.233	0.290	0.257	0.257	0.259	0.293	0.305
4	0.407	0.609	0.448	0.448	0.445	0.54	0.577
5	0.906	1.000	0.892	0.892	0.911	0.964	0.961
6	0.698	0.739	0.761	0.761	0.733	0.789	0.818
7	0.690	1.000	0.751	0.751	0.716	0.872	0.905
8	0.540	0.531	0.586	0.587	0.554	0.554	0.578
9	0.790	0.854	0.824	0.824	0.843	0.893	0.906
10	0.307	0.294	0.333	0.334	0.305	0.313	0.327
11	0.809	0.830	0.855	0.855	0.839	0.861	0.879
12	0.675	0.710	0.734	0.734	0.7	0.765	0.797
13	0.617	0.596	0.648	0.648	0.594	0.531	0.539
14	0.797	0.836	0.851	0.851	0.839	0.87	0.886
15	0.465	0.515	0.519	0.519	0.509	0.557	0.581
16	0.238	0.255	0.261	0.261	0.266	0.273	0.281
17	0.702	0.701	0.761	0.762	0.734	0.721	0.74
18	1.000	1.000	0.937	0.937	0.943	0.942	0.937
19	0.566	0.580	0.61	0.61	0.622	0.591	0.606
20	0.515	0.514	0.568	0.568	0.547	0.528	0.543
21	0.559	0.575	0.621	0.621	0.61	0.59	0.603
22	0.567	1.000	0.607	0.607	0.613	0.756	0.806
23	0.844	0.898	0.879	0.879	0.868	0.926	0.932
24	0.298	0.306	0.326	0.326	0.333	0.315	0.323
25	0.608	0.633	0.661	0.661	0.665	0.652	0.669
26	1.000	n.a	0.937	0.937	0.939	n.a	n.a
27	0.384	n.a	0.432	0.432	0.415	n.a	n.a
28	0.358	0.435	0.404	0.405	0.394	0.445	0.471
29	0.618	0.718	0.681	0.681	0.671	0.751	0.813
30	0.703	n.a	0.746	0.746	0.696	n.a	n.a
31	0.852	0.921	0.848	0.848	0.886	0.935	0.937
32	0.727	0.746	0.784	0.784	0.752	0.787	0.814
33	0.535	0.576	0.582	0.582	0.585	0.61	0.634
34	0.765	n.a	0.823	0.823	0.802	n.a	n.a
35	0.993	n.a	0.935	0.935	0.937	n.a	n.a
36	0.514	0.647	0.54	0.54	0.56	0.643	0.685
37	0.774	0.843	0.818	0.818	0.829	0.888	0.914
38	0.518	0.538	0.547	0.547	0.569	0.554	0.571
39	0.543	0.557	0.594	0.594	0.585	0.577	0.593
40	0.287	0.303	0.321	0.321	0.32	0.318	0.335
41	0.655	0.679	0.722	0.722	0.706	0.704	0.753
42	0.811	n.a	0.837	0.837	0.846	n.a	n.a
43	0.427	0.433	0.467	0.467	0.437	0.466	0.487
44	0.649	n.a	0.685	0.685	0.705	n.a	n.a
45	0.807	0.829	0.849	0.849	0.86	0.84	0.867
46	0.904	0.888	0.906	0.906	0.905	0.871	0.875
47	0.204	n.a	0.234	0.234	0.228	n.a	n.a
48	0.348	n.a	0.39	0.39	0.384	n.a	n.a
49	0.423	0.450	0.447	0.448	0.399	0.454	0.479
50	0.757	0.755	0.816	0.817	0.8	0.765	0.783
Mean	*0.611*	*0.662*	*0.647*	*0.647*	*0.639*	*0.663*	*0.685*

The homoskedastic specification with the potassium and greenhouse gases as the only pressures for both the complete sample of 50 observations (ALS-50C) and the reduced sample of 40 observations (ALS-40C) are reported in Columns (E)-(H), and a comparison of these estimates reveals that the coefficients on the pressures change very little across the three models.

To compare the eco-efficiencies estimated by DEA and SFA, a scatterplot of the DEA efficiency scores and the efficiency estimates from the SFA model is presented in Figure 2.[8] As can be seen, the eco-efficiencies are almost identical. Also plotted on Figure 2 is the regression line from the regression of the SFA estimates on the DEA scores, where the R^2 is 0.9805. The Spearman Rank Correlation Coefficient (Spearman's rho) was 0.996, showing that the models yielded virtually identical rankings of eco-efficiency levels. Even when the reduced sample of 40 observations is used, the eco-efficiencies are again very similar, with almost identical mean values and a Spearman Rank Correlation Coefficient of 0.959.

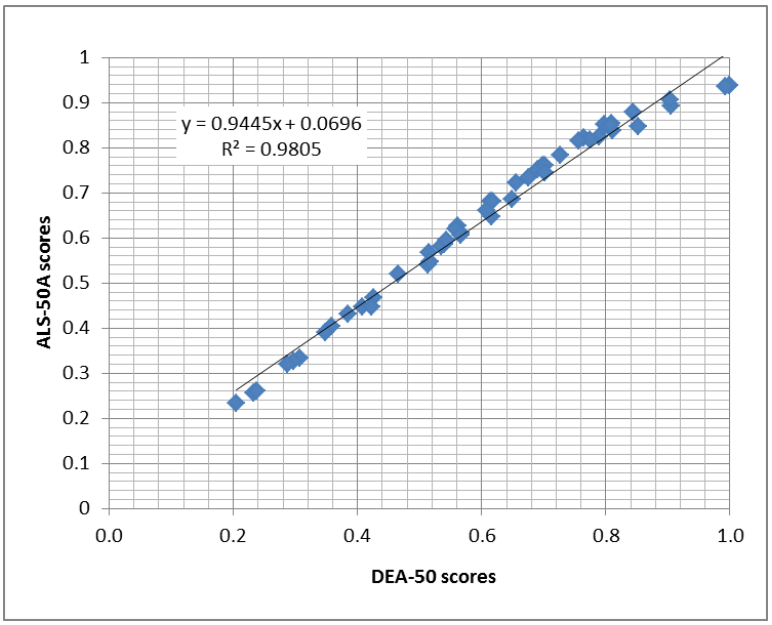

Figure 2. Comparison of Eco-efficiency scores

While the raw eco-efficiency scores between DEA and SFA are very similar, the questions remains as to whether the models yield similar results with regard to the determinants of eco-efficiency. The estimates of the efficiency determinants from the

[8] These DEA scores are based on a simple DEA calculation as opposed to the bootstrapped DEA scores reported in Pérez-Urdiales et al. (2015).

Table 4. Estimated coefficients and significance of eco-efficiency determinants

Variable	SFA*		DEA	
	Estimated parameter	Significant at 95% level?	Estimated parameter	Significant at 95% level?
HabitsCO_2	-1.060	Yes	-0.689	Yes
HabitsNutrients	-0.210	No	-0.231	Yes
Age	0.011	No	0.008	Yes
Prospects	-0.641	Yes	-2.144	Yes
Regulation	0.270	No	0.230	Yes
Training	-0.004	No	-0.002	Yes
Intercept	-0.317	Yes	0.161	Yes

* The SFA results are from Table 2. DEA results are obtained from Pérez-Urdiales et al. (2015, Table 3).

SFA model from Table 2 are presented in Table 4 alongside the parameter estimates reproduced from Pérez-Urdiales et al. (2015). While all the determinants in Pérez-Urdiales et al. (2015) were found to be significant at the 95% level, only two of the determinants – *HabitsCO_2* and *Prospects* – are significant at this level in the heteroskadastic SFA model (though two other variables – *Age* and *Regulation* - were significant at the 90% level). Notably, however, the SFA model yields exactly the same signs on the eco-efficiency determinants as the bootstrapped DEA-based truncated regression used by Pérez-Urdiales et al. (2015).

6. Conclusions

Measurement of eco-efficiency has been carried out exclusively using non-parametric DEA techniques in the literature to date. In the present work we have proposed using a (parametric) stochastic frontier analysis (SFA) approach. While such models are highly non-linear when estimating eco-inefficiency, in an empirical application we find that such an approach is feasible even when the sample size is relatively small and determinants of eco-inefficiency - which increases the number of parameters to be estimated - are incorporated. Using data from a sample of 50 Spanish dairy farms previously used by Pérez-Urdiales et al. (2015), we begin by estimating a stochastic frontier model without eco-efficiency determinants, and find that our model yields virtually identical eco-efficiency scores to those calculated by DEA. Estimating eco-efficiency without determinants involves relatively few parameters, so sample size should not be a major obstacle to using SFA. Our results corroborate this.

We then estimated a heteroskedastic SFA model which incorporated determinants of eco-inefficiency. We use the same determinants used by Pérez-Urdiales et al. (2015), who carried out their analysis applying bootstrapped truncated regression techniques. As extra parameters have to be estimated, the small sample size became more of an issue for the stochastic frontier model. Indeed, in order for the model to converge we had to use fewer environmental pressures in our application than Pérez-Urdiales et al. (2015). Encouragingly, however, we found the exact same signs on the determinants of eco-efficiency as those found by Pérez-Urdiales et al (2015). Thus, even with a small sample size and multiple determinants of eco-inefficiency, the stochastic frontier model yields similar conclusions to those obtained by truncated regression techniques based on DEA estimates of eco-efficiency.

Using stochastic frontier models for eco-efficiency measurement has some advantages over the bootstrapped truncated regression techniques that have been employed in the litrature to date. In particular, the stochastic frontier model can be carried out in one stage and the coefficients on the environmental pressures ('technology' parameters) have interesting interpretetations which shed light on the contribution of these pressures to firm profits. The estimated coefficients also uncover potentially useful information on the substitutability between environmental pressures. As such, we advocate the use of SFA for measuring eco-efficiency as a complement to or substitute for DEA-based approaches. When sample size is small and we wish to incorporate determinants of eco-efficiency, the DEA-based truncated regression techniques may permit more environmental pressures to be included in the analysis. However, with larger sample sizes, we would expect this advantage to disappear and the stochastic frontier models can provide extra valuable information for producers and policymakers, particularly with regard to substitutability between pressures.

References

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

Alvarez, A., Amsler, C., Orea, L., and Schmidt, P., (2006), "Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics", *Journal of Productivity Analysis*, 25(3), 201-212.

Banker, R.D., and Morey, R.C., (1986), "Efficiency analysis for exogenously fixed inputs and outputs", *Operations Research*, 34(4), 513-521.

Behr, A. and Tente, S., (2008), "Stochastic frontier analysis by means of maximum likelihood and the method of moments", *Discussion Paper Series 2: Banking and Financial Studies*, No 19/2008.

Caudill, S.B., and Ford, J.M., (1993), "Biases in frontier estimation due to heteroscedasticity", *Economics Letters*, 41(1), 17-20.

Johnson, A.L., and Kuosmanen, T., (2011), "One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-*n* consistent StoNEZD method", *Journal of Productivity Analysis*, 36(2), 219-230.

Johnson, A.L., and Kuosmanen, T., (2012), "One-stage and two-stage DEA estimation of the effects of contextual variables", *European Journal of Operational Research*, 220(2), 559-570.

Jondrow, J., Lovell, C.A.K., Materov, I.S., and Schmidt, P., (1982), "On the estimation of technical inefficiency in the stochastic frontier production function model", *Journal of Econometrics*, 19 (2-3), 233-238.

Kumbhakar, S.C., and Lovell, C.A.K., (2000), *Stochastic frontier analysis*, Cambridge University Press, Cambridge.

Kumbhakar, S.C., Asche, F., and Tveteras, R., (2013), "Estimation and decomposition of inefficiency when producers maximize return to the outlay: an application to Norwegian fishing trawlers", *Journal of Productivity Analysis*, 40(3), 307-321.

Kuosmanen, T., and Kortelainen, M., (2005), "Measuring eco-efficiency of production with data envelopment analysis", *Journal of Industrial Ecology*, 9(4), 59-72.

Lauwers, L., (2009), "Justifying the incorporation of the materials balance principle into frontier-based eco-efficiency models", *Ecological Economics*, 68(6), 1605-1614.

OECD, (1998), *Eco-Efficiency*, Organization for Economic Co-operation and Development, OECD, Paris.

Olson, J.A., Schmidt, P., and Waldman, D.M., (1980), "A Monte Carlo study of estimators of stochastic frontier production functions", *Journal of Econometrics*, 13(1), 67-82.

Oude Lansink, A., and Wall, A., (2014), "Frontier models for evaluating environmental efficiency: an overview", *Economics and Business Letters*, 3(1), 43-50.

Pérez-Urdiales, M., Oude Lansink, and Wall, A., (2015), "Eco-efficiency among dairy farmers: the importance of socio-economic characteristics and farmer attitudes", *Environmental and Resource Economics*, DOI: 10.1007/s10640-015-9885-1.

Picazo-Tadeo, A.J., Beltrán-Esteve, M., and Gómez-Limón, J.A., (2012), "Assessing eco-efficiency with directional distance functions", *European Journal of Operational Research*, 220(3), 798-809.

Picazo-Tadeo, A.J., Gómez-Limón, J.A., and Reig-Martínez, E., (2011), "Assessing farming eco-efficiency: a data envelopment analysis approach", *Journal of Environmental Management*, 92(4), 1154-1164.

Ray, S.C., (1988), "Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation", Socio-Economic Planning Sciences, 22(4), 167-176.

Ruggiero, J., (1996), "On the measurement of technical efficiency in the public sector", *European Journal of Operational Research*, 90(3), 553-565.

Schmidheiney, S., (1993), *Changing course: a global business perspective on development and the environment*, Technical report, MIT Press, Cambridge, MA.

Simar, L., and Wilson, P.W., (2007), "Estimation and inference in two-stage, semi-parametric models of production processes", *Journal of Econometrics*, 136(1), 31-64.

Tyteca, D., (1996), "On the measurement of the environmental performance of firms - a literature review and a productive efficiency perspective", *Journal of Environmental Management*, 46(3), 281-308.

Wang, H.-J., (2002), "Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model", *Journal of Productivity Analysis*, 18(3), 241-253.