

Analyzing Italian citrus sector by semi-nonparametric frontier efficiency models

Francesco Vidoli · Giancarlo Ferrara

Received: 22 July 2013 / Accepted: 9 September 2014 / Published online: 25 October 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract In this paper, we carried out an empirical productive analysis on agricultural Italian farms. In this research area, we propose a new approach of stochastic frontier analysis adopting a generalized additive model framework also compared with Stochastic semi-Nonparametric Envelopment of Z variables Data. By using the Italian National Institute of Agricultural Economics micro-data, we were able to map out the overall level of efficiency thereby focusing also on the evaluation of the differences observed due to presence of contextual variables. We obtained overall measures for the citrus sector that suggests an evaluation framework that can uphold policies to encourage and support farms.

Keywords Stochastic frontier · Generalized additive model · StoNEZD · Agriculture · Efficiency · Sustainable value

JEL Classification Q12 · C14 · Q57

1 Introduction

During the 1990s, policies introducing sustainability and efficiency principles in agriculture and natural resources were reconsidered. The need to support market orientation and to adapt it to the emerging demands of society in order to improve the quality

F. Vidoli (✉)

Department of Political Science, University of Roma Tre, Via G. Chiabrera, 199 Rome, Italy
e-mail: fvidoli@gmail.com

G. Ferrara

Department of Statistical and Economical Sciences, University of Palermo,
Viale delle Scienze Ed. 13, Palermo, Italy
e-mail: giancarlo.ferrara@gmail.com

of production and limit the impact of agricultural production on the environment has been the main driver for the EU Common Agricultural Policy (CAP), one of the oldest policies in the European Union. The initial objectives of the CAP (to increase productivity, ensure a fair standard of living for the agricultural community, stabilize markets, secure availability of supplies and provide consumers with food at reasonable prices) have remained unchanged over the years, even though the importance given to the different objectives has changed drastically. In response to this legal and policy framework demand, measuring firms' contributions to sustainability and efficiency has attracted increasing attention in recent years and many different practical approaches have been suggested (e.g. Tyteca 1998). Furthermore, in different field of research, as highlighted for example by D'Amico and Fernandez (2012), there is an increasing need to provide key policy evidence in a context in which optimization is critical both for social and financial purposes.

One of the most promising techniques in agricultural economics—linking sustainability and efficiency—is the Sustainable Value (SV) approach as introduced by Figge and Hahn (2005) which has stimulated, in recent years, a lively theoretical debate (Kuosmanen and Kuosmanen 2009a; Van Passel et al. 2009; Kuosmanen and Kuosmanen 2009b; Figge and Hahn 2009); in this paper, we refer to the Kuosmanen and Kuosmanen (2009b)'s version, also called the Generalized Sustainable Value (GSV). GSV is a systematic economic approach to measure the Sustainable Value creation of firms¹, and it measures the overall efficient use of a set of economic, environmental and social resources. A firm is said to create Sustainable Value whenever it uses its resources more efficiently than another firm would have. In principle, reallocating resources from firms that create negative Sustainable Value to ones that create positive one can increase the overall economic welfare while keeping all capital stock in the economy at a constant level. Thus, firms creating Sustainable Value would be able to compensate for any rebound effects that might occur. GSV methodology measures the use of environmental resources in a new way, as reported in ADVANCE (2008): "Rather than looking at how costly, painful or burdensome the use of an environmental resource is, it compares the value that can be created, at constant input, by different economic actors". From this point of view, companies create value whenever they use a certain resource more efficiently than others. In this context, the GSV approach may be considered the first value-based approach to the measurement and management of sustainability, and it appears appropriate for the agricultural sector where productive, ecological and social aspects are very closely related. The economic difference and statistical analogies between GSV analysis and production efficiency estimation methods are described in Kuosmanen and Kuosmanen (2009b) where it highlighted the possibility to use empirical frontier econometric methods for the quantification of the firms' economic sustainability. Hereinafter, given the correspondence, from a mathematical point of view, between GSV and technical efficiency definitions, in our paper, we refer it as *technical efficiency*.

Over the last decades, several studies have been carried out on the farms production efficiency. One of the first attempts to apply the efficiency theory and the practical

¹ In this paper, we refer to a farm (specifically "firm" in the introduction) even if the methodology remains valid for any public or private economic unit.

application on the agricultural economy was made by Coelli (1995) recommending to use the stochastic frontier method because measurement error, missing variables and weather play a significant role in this field.

In this context, even within models containing the random error part, it remains unsolved the correct choice of the production function; Hjalmarsson et al. (1996) suggest that a conventional Cobb–Douglas production function, in some situations, is not ever adequate. Moreover, forcing the production function to belong to a parametric family of functions (Translog, Cobb–Douglas) can be too restrictive, even inappropriate (Giannakas et al. 2003).

The purpose of this paper is both methodological and applied since a new approach of Stochastic Frontier Analysis (SFA) by using a Generalized additive model (GAM) framework is proposed with the aim of overcoming previous cited weakness about the specification of the functional form and it is compared with an empirical application of the Stochastic semi-Nonparametric Envelopment of Z variables Data (StoNEZD) (Johnson and Kuosmanen 2011), respectively. Thanks to the Italian National Institute of Agricultural Economics (INEA)'s micro-data, we were able to map out the level of efficiency of citrus sector for the whole territory. Citrus production is one of the most important Italian agricultural sector and it is the most studied sector in the field of production efficiency.

The remainder of the paper is structured as follows: first, we present an overview on the different methodologies considered in efficiency analysis, providing a brief introduction to the proposed GAM-SFA approach and the key concepts of existing StoNED models, secondly, description of data, model specifications and the application to the citrus sector are illustrated, while conclusions are reported in the last section.

2 Estimating efficiency

Productive efficiency methods can be classified on the basis of whether an average-practice or best-practice technology is estimated. The best-practice technologies can be further classified as non-stochastic and stochastic technologies, depending on the presence of a stochastic noise term. Furthermore, the methods can be further classified as being parametric and nonparametric in their orientation.

Parametric methods assume a specific functional form of the production function, which is usually linear in its parameters. Nonparametric methods do not assume a particular functional form, but estimate the benchmark technology based on a minimal set of axioms.

2.1 An overview

The determination of a technical efficiency measure is based on the knowledge of the so-called production function; it makes possible the relation of the production process of individual units to the *efficient* border of the production possibilities. The measure of the distance of each unit from the border is the most immediate way to assess its efficiency (Farrell 1957).

However, the production function as the efficient frontier is not generally known, but it has only a set of information on each production unit. It is therefore essential to develop techniques to estimate the production frontier.

In econometric literature, specification and estimation of production frontier functions are usually carried out by two different approaches:

- Stochastic Frontier Analysis (SFA) (Aigner et al. 1977; Meeusen and van den Broeck 1977)—Deterministic Frontier Analysis (DFA) (Aigner and Chu 1968);
- Data Envelopment Analysis (DEA) (Farrell 1957; Charnes et al. 1978)—Free Disposal Hull (FDH) (Deprins et al. 1984; Grosskopf 1996).

The two approaches listed above are classified in the literature, as parametric (DFA, SFA) and nonparametric (DEA, FDH) methods, respectively. In a parametric analysis, it is essential to specify *a priori* an explicit functional form of the boundary of the production set, while the nonparametric analysis is characterized by the ability to determine the relative efficiency of such units through linear programming, without specifying any functional form for the production function. In other words, unlike parametric techniques, the DEA-FDH approach allows for the determination of the relative efficiency of decision-making units in the absence of a similar detailed description of the production process. If the latter seems to make this approach particularly flexible and generalizable, the main drawback of DEA (or FDH) is its deterministic nature. When using this procedure, it is not possible to recognize whether the difference in efficiency, namely the distance between observed and maximum possible output, is due to technical inefficiency or effects of disturbance of an accidental type (Greene 2008). Therefore, it is not possible to determine whether, for example, inefficiency is due to an adverse condition of the contextual variables and therefore independent of the actions of the entrepreneur (i.e. a season of no rain will not help farm managers get good results) or it can be expressed as the determinant of other factors (such as the quality of personnel management within the enterprise). At the same time, a model with the parametric frontier production deterministic type of DFA, which allows for an explicit production function solely on the basis of inputs used, would explain only in part the inefficiency of the firm, as several inefficiency factors and measurement errors would be equally incorporated in the same error term.

The parametric model with stochastic production frontier (SFA) in addition to providing, as well in the classical DFA, useful information on the productive asset, exceeds the limits associated with the above model, resulting in an entire analysis of the sources of inefficiency that are not directly attributable to the production function or disturbances of an accidental type, and therefore are not directly attributable to corporate *policy*. The most important drawback associated with the SFA approach is the lack of flexibility associated with the specification of the production function. To overcome this problem, we consider a nonparametric specification, not yet considered in SFA models, to relax the choice of particular production frontier based on a Generalize Additive Model (GAM) framework. Finally, since over the last decade, many authors have tried to remove differences between the two competitors DEA-SFA by relaxing some assumptions or proposing semi-parametric or semi-nonparametric methods (Du et al. 2013), we will consider in our analysis the Stochastic semi-Nonparametric Envelopment of Data (StoNED) (Kuosmanen and Kortelainen 2012).

Due to the absence of the “best” approach for the analysis of efficiency and with the aim of introducing new insights for analysis and research, this study will be conducted by comparing different models based on stochastic (GAM-SFA, standard SFA), non-parametric (DEA) frontiers and the StoNED procedure. We conclude this section with a brief introduction to the proposed GAM-SFA approach and also provide the key concepts of StoNED framework.

2.2 GAM-SFA specification

Parametric stochastic frontier models, introduced by [Aigner et al. \(1977\)](#) and [Meeusen and van den Broeck \(1977\)](#), specify output in terms of a response function and a composite error term. The composite error term consists of a two-sided error representing random effects and a one-sided term representing technical inefficiency. Since their introduction, several related papers have been published in the relevant literature; see [Greene \(2008\)](#) and [Kumbhakar and Lovell \(2000\)](#) for overviews of developments in this area.

The stochastic frontier model can be written, in general terms, as:

$$y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) + v_i - u_i, \quad i = 1, \dots, n, \quad (1)$$

where $Y_i \in \mathbb{R}_+$ is the single output of unit i , $\mathbf{X}_i \in \mathbb{R}_+^p$ is the vector of inputs, $f(\cdot)$ defines a production (frontier) relationship between inputs \mathbf{X} and the single output Y , v_i is a symmetric two-sided error representing random effects and $u_i > 0$ is one-sided error term which represents technical inefficiency. If we recall that GSV may be defined as the residual between the observed output and the corresponding production frontier, it is evident that stochastic frontier model approach conforms to generalized SV formulation ([Kuosmanen and Kuosmanen 2009b](#)).

In applications, the two-sided error term is usually² assumed to be normally distributed: $v \sim N(0, \sigma_v^2)$. We assume u is distributed half-normally on the non-negative part of the real number line: $u \sim N^+(0, \sigma_u^2)$. In following common practice, we assume that v and u are each identically independently distributed (iid). In spite of the ease this model presents in terms of computation and interpretation of the results, there is an important drawback: a lack of flexibility. Indeed, in some situations, forcing $f(\cdot)$ to belong to a parametric family of functions (Translog, Cobb-Douglas) can be too restrictive, even inappropriate, and this may lead to a serious modelling bias and therefore misleading conclusions about the link between \mathbf{X} and Y ([Giannakas et al. 2003](#)).

To overcome drawbacks due to the specification of a particular production function, we propose a Generalized Additive Model (GAM) framework for the estimation of stochastic production frontier models.

A GAM fits a response variable Y using a sum of smooth functions of the explanatory variables, X_j for $j = 1, \dots, p$. In a regression context with Normal response, the model is:

² Various distributions can be assumed for the one-sided error term; e.g. half-normal, truncated normal, gamma, exponential, etc.

$$\mu = E(Y|X = \mathbf{x}) = \alpha + \sum_{j=1}^p f_j(X_j), \quad (2)$$

where the $f_j(\cdot)$'s are smooth functions (Hastie and Tibshirani 1990) standardized so that $E[f_j(X_j)] = 0$. GAMs can provide useful approximations to the regression surface, but relaxing the linear (polynomial) structure of the additive effects.

This additional flexibility alleviates the need to impose a perfect linear relationship between each explanatory variable and the response, yet it explains the variability of the response using an additive function of the inputs as in the corresponding SFA model. Generalized additive models are more flexible, and the advantage is that the best transformations are determined simultaneously. One useful feature of additive models is that nonparametric estimators of the unknown functions f_j have one-dimensional convergence rates (Stone 1986), which makes them much more accurate than estimating the p -dimensional function, and are able to avoid the “curse of dimensionality”.

In a cross-sectional setting, relatively few papers have dealt with differences from conventional production frontiers. Model (1) becomes:

$$y_i = \psi(\mathbf{x}_i) + v_i - u_i, \quad i = 1, \dots, n, \quad (3)$$

where the unknown function $\psi(\cdot)$ is modelled via GAMs (2) to relax the linear assumption between inputs and output (represented on log scale), but ensuring the additivity of the input factors.

For the estimation of the stochastic frontier model (3), we consider the following two step procedure as proposed by Fan et al. (1996):

- estimating the conditional expectation $E(Y|X = \mathbf{x})$ (i.e. the “mean” frontier),
- estimating error term parameters (σ_v, σ_u) by Fan et al. (1996) method,

where in the first step, we introduce a GAM specification for the estimation of $E(Y|X = \mathbf{x})$. In this sense, we call our proposal a GAM-SFA model, defining in this way a class of models capable of applying any class of nonparametric estimator of the inputs to the Fan et al. (1996) approach. This kind of models maintains, on the scale given by the smooth terms f_j 's, the same hypothesis of the corresponding classical SFA model (Aigner et al. 1977) in terms of additivity of the inputs, separability assumptions and independence between u and v conditionally on X .

More specifically, we consider a penalized regression splines approach as reported in Wood (2003) where proper penalties are introduced to guarantee the smoothness of the fitted production frontier. In particular, the f_j 's smooth functions are represented using thin plate regression splines with smoothing parameters selected by Generalized Cross Validation (GCV) criterion: $n * D / (n - DoF)^2$, where D is the deviance, n the number of data and DoF the effective degrees of freedom of the model. One key advantage of the approach is that it avoids the knot placement problems of conventional regression spline modelling (see Wood 2006 for further details).

After obtaining the “mean” frontier $E(Y|\widehat{X} = \mathbf{x})$, the estimation of the production function $\psi(\cdot)$ will be achieved by shifting the estimation of the conditional expectation in an amount equal to the average estimate of the expected value of the term of inefficiency (Fan et al. 1996).

Having obtained the pseudo maximum-likelihood estimates of the parameter σ_v and σ_u , the next step is to estimate the technical efficiency of each unit. [Jondrow et al. \(1982\)](#) provided a solution to the problem that involves deriving the conditional distribution of the component u respect to the compound error $\varepsilon = v - u$.

In this paper, we will consider the estimation of model (3) with unknown $\psi(\cdot)$ modelled using a penalized regression splines with penalty ([Wood 2003, 2006](#)) as a GAM.³ This class of models includes the linear model as a special case, where $f_j(x_j) = \beta_j x_j$ ([Aigner et al. 1977](#)) but it is clearly more general, because the f_j 's can be very arbitrary nonlinear functions. In addition, it is still possible to further customize the specification of the production frontier considering a generalization of Eq. (2) in such a way:

$$f(x_1, x_2, \dots, x_p, z) = \alpha + \sum_{i=1}^p f_j(x_j) + \sum_{i=1}^p \sum_{k < p} f_{kj}(x_k, x_j) + \beta z + \dots, \quad (4)$$

by introducing effects due to interactions among covariates or linear terms. Some possible estimations of model (4) have already been proposed in the literature including algorithms for backfitting, and more recently, [Sperlich et al. \(2002\)](#) used marginal integration for estimating additive models with interactions and the relative testing procedure. However, it should also be considered that introducing higher-order interactions would lead some interpretation and curse of dimensionality drawbacks.

2.3 Stochastic semi-Nonparametric Envelopment of Data

Stochastic semi-Nonparametric Envelopment of Data (StoNED, [Kuosmanen and Kortelainen 2012](#)) is an estimation method that, like other nonparametric methods, does not require any prior functional form assumption about $f(\cdot)$ but it assumes monotonicity and concavity of the production function.

The StoNED method differs from DEA and Corrected Concave Nonparametric Least Squares (C²NLS) in that it decomposes the deviations of y_i from $f(x_i)$ into two sources: the inefficiency term u_i and the stochastic noise term v_i , which is similar to SFA, trying to combine the deterministic part of DEA with the stochastic part of SFA.

The main advantage of StoNED over the nonparametric DEA is the better robustness to outliers, data errors and other stochastic noise in the data. While in a DEA approach, the benchmark technology is spanned by a relatively small number of efficient farms, in StoNED all observations influence the benchmark and also coefficients β_i can be “interpreted as the subgradient vector $\nabla f(x_i)$, and thus it represents the vector of marginal products of inputs at point x_i ” ([Kuosmanen and Kortelainen 2012](#)). As for the GAM-SFA specification, there is not an *ad hoc* assumptions regarding the functional form of the benchmark technology.

³ For the model estimation we used the R Environment (R: A language and environment for statistical computing [2012](#)) exploiting the functionality of the [mgcv \(2012\)](#) package.

Estimation of the StoNED model is conducted in two stages. In the first stage, the conditional expected value of $\ln(y)$ is estimated by CNLS (Convex Nonparametric Least Squares) regression [here in equivalent finite-dimensional quadratic programming expressed by a multiplicative model⁴ as suggested by [Kuosmanen and Kortelainen \(2012\)](#)]:

$$\min_{\alpha, \beta, \phi} \left\{ \sum_{i=1}^n (\ln y_i - \ln \widehat{\phi}_i)^2 \left| \begin{array}{l} \widehat{\phi}_i = \alpha_i + \beta'_i \mathbf{x}_i, \forall i = 1, \dots, n \\ \alpha_i + \beta'_i \mathbf{x}_i \leq \alpha_h + \beta'_h \mathbf{x}_i, \forall h, i = 1, \dots, n \\ \beta_i \geq \mathbf{0}, \forall i = 1, \dots, n \end{array} \right. \right\}. \quad (5)$$

Given the CNLS composite residuals⁵ $(\ln y_i - \ln \widehat{\phi}_i)$ from Eq. (5), we subsequently filter out the noise. This requires some distributional assumptions, e.g. the standard SFA assumptions $v \sim N(0, \sigma_v^2)$ and $u \sim N^+(0, \sigma_u^2)$. Parameters (σ_u, σ_v) can be estimated by the method of moments or by maximum pseudo-likelihood techniques and the conditional expectation of u is then computed using the [Jondrow et al. \(1982\)](#)'s formula.

Recently, [Johnson and Kuosmanen \(2011\)](#) introduced Stochastic semi-Nonparametric Development of Z variables Data (StoNEZD) model with the purpose of including contextual variables. The conventional approach tests the significance of the contextual variable by using a two-stage method where the efficiency estimates are regressed on the contextual variable, representing the operational conditions, using, for example, a censored Tobit regression.

The main problem, highlighted also by [Wang and Schmidt \(2002\)](#), is that “the second stage estimator is biased and inconsistent when the inputs are correlated with the contextual variables”, since the two-stage approaches ignore the correlations between inputs and contextual variables (see also [Simar and Wilson 2007](#)).

[Johnson and Kuosmanen \(2011\)](#) showed that contextual variable z can be directly incorporated in the objective function here in a linear form, developing the CNLS formulation as in Eq. (6):

$$\min_{\alpha, \beta, \delta, \phi} \left\{ \sum_{i=1}^n (\ln y_i - \ln \widehat{\phi}_i - z'_i \delta)^2 \left| \begin{array}{l} \widehat{\phi}_i = \alpha_i + \beta'_i \mathbf{x}_i, \forall i = 1, \dots, n \\ \alpha_i + \beta'_i \mathbf{x}_i \leq \alpha_h + \beta'_h \mathbf{x}_i, \forall h, i = 1, \dots, n \\ \beta_i \geq \mathbf{0}, \forall i = 1, \dots, n \end{array} \right. \right\}, \quad (6)$$

where δ represents the average effect of contextual variables z_i on performances and $z'_i \delta - u_i$ can be seen as the overall efficiency of farm i , where the term $z'_i \delta$ represents technical inefficiency that is explained by the contextual variables, and the component u_i represents the proportion of inefficiency that remains unexplained.

⁴ “The input-output data must be kept in the original units in order to use the Afriat inequalities for imposing concavity. Although the objective function involves logarithms of model variables” ([Kuosmanen and Kortelainen 2012](#)).

⁵ Note that log-transformation concerns Step 1 and makes no difference in the estimation of Step 2.

3 Data, models and application to citrus sector

The ADVANCE Guide to Sustainable Value Calculations (ADVANCE 2008) suggests calculating efficiency (i.e. GSV) in different steps: preparing for the assessment (choosing farms, benchmarks and defining resources to be included and output), collecting and harmonizing data and, the last one, calculating the scores. This section is therefore divided into three sections, one for each step mentioned above.

3.1 Data source

The Council Regulation No. 79/65/EEC of the 15 June, 1965 set up a network to collect accounting data on the incomes and business operations of agricultural holdings in the European Economic Community. The Farm Accountancy Data Network (FADN) is an annual survey carried out by the Member States of the European Union, and it represents an instrument to evaluate the income of agricultural holdings and the impacts of the Common Agricultural Policy. Based on national surveys, FADN is the only source of micro-economic data that is harmonized (in other words, the bookkeeping principles are the same in all countries).

In Italy, the FADN survey is carried out every year by the National Institute of Agricultural Economics (INEA), that is the liaison agency between the EU and Member States. The European Commission provides guidelines to define the instructions and recommendations to outline FADN's selection plan. It must ensure the representativeness of the returning holdings as a whole, and it defines the number of farms to be selected by region, type of farming (OTE) and classes of economic size, and it also specifies the rules applied for selecting the holdings. The Italian FADN sample is selected using the stratified random sampling technique and only includes commercial farms or those with an economic size of more than 4 Economic Size Units (ESU).⁶

The simple random sample allows for the extension, in terms of inference, of the results to the universe of farms as a whole that is formed by the subset of the EU universe; in this work, we used the Italian FADN survey database concerning the year 2007.

According to the EU FADN methodology, territorial location, economic size and type of farming were used as stratification variables: territorial location corresponds to Italian administrative regions; economic size is expressed in terms of ESU, calculated on the basis of the farm's Standard Gross Margin and divided in several classes, ranging from 4 to 8 ESU to more than 250 ESU; type of farming corresponds to the General OTE.⁷

⁶ EEC Regulation No. 1859/82 establishes the minimum threshold of economic size for inclusion in the FADN field of observation. The economic size of a farm is defined by the total standard gross margin expressed in ESU, where 1 ESU corresponds to 1,200 Euro.

⁷ Farms, in the European agricultural policy, are classified on the basis of type of farming (OTE). OTE provides information on the degree of specialization and production and it is determined on the basis of the percentage of the economic dimension (in terms of the standard gross margin) of one or more productive activities on the economic dimension of the total. The Community typological scheme provides 58 different combinations of production which are grouped into three successive levels of detail: General OTE, Main OTE and Particular OTE.

The different conditions in irrigation, energy consumption, work organization and profitability led us to focus our analysis on the citrus sector (Particular OTE 3220). We chose this sector for several reasons: citrus production is one of the leading Italian agricultural sector and it is the most studied sector in the field of production efficiency, given its strong spatial characteristics and economic importance in the territories (e.g. [Cisilino and Madau 2007](#); [Madau 2010](#)). Furthermore, the citrus industry is heavily subsidized by European policies and therefore it was interesting to see, with respect to other sectors not included in this paper, whether these policies distort the estimation of efficiency. Finally, it was chosen because of the specialization in specific regions. The sample covers 229 production units all located in Southern Italy.

3.2 Selection of input/output variables

Following criteria used in earlier studies (e.g. [Hani et al. 2003](#); [Reig-Martinez et al. 2011](#)), concerning the analysis of sustainable production in agriculture, we built an analysis framework, as reported in [Table 1](#), covering economic-financial and social-environmental sustainability aspects, trying to build a bridge between the economic aspect of the GSV and the field of production efficiency.

The response/output variable here used for efficiency models is the gross production (*PL*). Although it would have been preferable to choose a measure of physical output in order to avoid the influence of agricultural prices, it was not possible due to the lack of data; however, this critical issue is partially mitigated in the analysis in that we had considered a homogeneous production sector. We considered variables *LAB* (labour measured in terms of total number of hours worked per year), *CAP_L* (land farm capital) and *CAP_O* (operating farm capital) as inputs, as usually done in the analysis

Table 1 Variables involved in the estimation by dimension of sustainability

Dimension	Subdimension	Indicator/variable	Role in the analysis	Abbreviation
Economic	Gross production	Gross production	Output	PL
	Capital	Land farm capital	Input	CAP_L
		Operating farm capital	Input	CAP_O
Social	Employment	Agricultural employment	Input	LAB
		Index of utilization of family labour	Contextual	IUF
	Generational change	Risk of abandoning familiar agricultural activity	Contextual	RISK
		Farms elevation	Contextual	ELEV
Environmental	Energy	Energy and Water consumption	Input	E_W
	Soil	Fertilizers and pesticides	Input	PEST
		Organic production	Contextual	ORGANIC

of the production function. Moreover, we included as inputs two variables that, from an economic point of view, characterize environmental sustainability: Energy and water consumption (E_W) and Fertilizers and pesticides consumption (PEST).

We have divided the analysis into two steps: first, we used a conventional specification of efficiency models specifying the output PL in terms of inputs X (LAB, CAP_L, CAP_O, E_W , PEST) for the GAM-SFA, classical SFA, DEA and StoNED models; in the second stage, we considered the economic, environmental and social indicators (i.e. contextual variables) as z -variables by using StoNEZD model.

Many authors have proposed taking into account the effects of some contextual variables on efficiency with the aim of restricting the influence of multiple and heterogeneous dimensions. Agricultural production, in particular, as well as the individual farms productivity, can be strongly influenced by a multitude of random factors such as climatic and local factors related to the specific characteristics of the region or technical specifications. Since spatial external factors (natural or artificial) are often difficult to identify and to evaluate separately, the contextual variable selection was based on the literature cited above and on the availability and reliability of statistical data.

Contextual dimensions can be captured both by variables related to social sustainability as index of utilization of family labour (IUF)⁸ and risk of abandoning agricultural activity (RISK) both by the organic certification (ORGANIC)⁹ and the physical farm's characteristics like elevation (ALTITUDE).¹⁰ In particular, IUF and ORGANIC may be considered as endogenous variables, but only in the long term, since, in the agricultural sector, the production structure and culture techniques are usually difficult to change and suffer from a high degree of inertia.

The RISK composite indicator was created following the [Reig-Martinez et al. \(2011\)](#) approach using the [De Muro et al. \(2010\)](#) method called MPI. In particular, we chose as elementary indicators (i) the capability of continuing agricultural activity by the farmer's age, (ii) the profitability of the farm (in terms of ROI) and (iii) the presence of familiar co-working. The [De Muro et al. \(2010\)](#) composite indicator sets 100 as the mean indicator; values greater than 100, therefore, indicate that there is a high risk of the farm dropping out, while values less than 100 indicate a farm that is profitable and has a good chance of surviving. Table 2 lists the correlations¹¹ among contextual variables and explanatory variables (i.e. inputs).

The variable of seasonal workers was not included in the analysis, although it is present in the Italian FADN database; this variable is highly underestimated due to the large amount of undeclared workers. Other very interesting dimensions used in

⁸ Index of utilization of family labour (IUF) is the ratio between the hours worked by family members and the total working hours that they would have worked had they been full time employees (i.e. 2,200 h per employee).

⁹ As dummy contextual variables.

¹⁰ For the Italian demographic characteristics, we have considered this variable as a variable that explains a social context, for the increasing abandonment of the hills or mountainous land, but we are aware that it may capture other economic effects.

¹¹ ***= P value < 0.001 , **= $0.001 \leq P$ value < 0.05 , *= $0.05 \leq P$ value < 0.10 .

Table 2 Correlation among contextual variables and input variables

Input variable	RISK	IUF	ALTITUDE	ORGANIC
LAB	0.30046***	0.42520***	-0.14064**	0.19912**
CAP_L	0.03239	0.28925***	-0.05586	0.23965**
CAP_O	0.19843**	0.28629**	-0.11568*	0.11915*
E_W	0.01591	0.03137	-0.06796	-0.05712
PEST	0.16265**	-0.15362**	-0.23289**	-0.09225

Table 3 Kendall's τ between efficiency measures

	DEA	StoNED	GAM-SFA	SFA
DEA	1	0.878	0.655	0.650
StoNED		1	0.852	0.864
GAM-SFA			1	0.968
SFA				1

some empirical papers, such as variables concerning waste and residue or working conditions, are not included because of the lack of data.

Since nonparametric models, like GAM-SFA and StoNED-StoNEZD, do not allow for testing of the significance of each variable, for the selection of output/inputs and contextual variables we have used two sets of criteria. More specifically, we have considered both economic indications contained in scientific papers (Hani et al. 2003; Reig-Martinez et al. 2011) and both OECD's recommendations that in recent years (e.g. OECD 2001) tried to identify the main determinants of agricultural competitiveness.

Some determinants to be considered for differences in performance were: farm size, factor intensity indicators, farm specialization, consumer demand, natural environment or the presence of a specialized agri-food industry (see Latruffe 2010), as determinants of differences in performance. At the same time, the variable selection was carried out by using the estimation of the corresponding parametric SFA model, even if it represents only a descriptive criterion.

3.3 Application to the citrus sector

Following the scheme proposed in Table 1 and in Eqs. (3) and (5), we estimated production efficiency in the citrus sector without taking into account the contextual variables. The close relationship between the StoNED model and models based on stochastic frontiers is obtained from the Kendall rank correlation coefficient,¹² reported in Table 3, between efficiency measures. The GAM-SFA specification seems to show the presence of a higher average level of efficiency, as outlined by the corresponding lower average opportunity cost reported in Table 6. This result may be explained by the linearity hypothesis associated to the classical SFA specification and by the monotonicity and concavity for the StoNED.

¹² All correlations are 0.05 significant.

Table 4 StoNED versus StoNEZD's efficiency score

	StoNED	StoNEZD $z = \text{RISK}$	StoNEZD $z = \text{IUF}$	StoNEZD $z = \text{ALTITUDE}$	StoNEZD $z = \text{ORGANIC}$
σ_u	0.43849	0.40599	0.46676	0.37714	0.45083
σ_v	0.30435	0.31237	0.48054	0.32914	0.29720
δ	—	0.01831	0.00034	-0.00043	-0.02158

In the second part of our analysis, we evaluated separately (one by one) the determinants of efficiencies, using the StoNEZD model and following the schema proposed in Table 1. As previously described, we analysed four potential contextual variables: risk of abandoning familiar agricultural activity (RISK), index of utilization of family labour (IUF), organic certification (ORGANIC) and farm's elevation (ALTITUDE); we estimated one StoNEZD model separately for each contextual variable.

Table 4 reports the $\hat{\sigma}_v$ and $\hat{\sigma}_u$ estimates for the different StoNEZD models and the effect associated with the corresponding contextual variable represented by δ .

We found, in terms of $\hat{\sigma}_v$ and $\hat{\sigma}_u$, consistent results among the different specifications (i.e. different contextual variables) of the StoNEZD, and, on the other side, different signs of the δ coefficients: positive δ , such as for risk of abandoning familiar agricultural activity (RISK) or the utilization of family labour (IUF), means that high levels of the corresponding z variable lead farms, under equal conditions, to obtain lower efficiencies with respect to the basic model. At the same time, using organic production efficiency methods (ORGANIC), to which corresponds a negative δ , increases (on average) the efficiency of farms. On the other hand, altitude does not seem to provide meaningful indications.

One of the StoNEZD drawbacks is the lack of a proper diagnostic about the correct specification of the model, especially with respect to the contextual variable z that is introduced linearly to check for the presence of an "average" effect. To overcome these issues, especially to evaluate the linearity hypothesis, we propose a similar criterion originally proposed by [Daraio and Simar \(2007\)](#) for nonparametric methods, not just to obtain evidence of the z influence (*favourable* or *unfavourable*),¹³ but rather to have a qualitative diagnostic for the linear specification of z into Eq. (6). Following [Daraio and Simar \(2007\)](#), we examine a graphical representation that makes it possible to verify whether the impact of z is uniformly linear over the whole domain or whether there are discontinuity points and non-linearity; in such cases, a simple linear specification would not correct the z effect on the dependent variable properly.

Specifying the ratio between different efficiency estimations (ϑ) by the following equation:

$$Q_z = \frac{\vartheta_n(x, y|z)}{\vartheta_n(x, y)} \quad (7)$$

we can subsequently plot Q_z versus z , to check for the linearity assumption.

¹³ This would not make sense since δ is estimated.

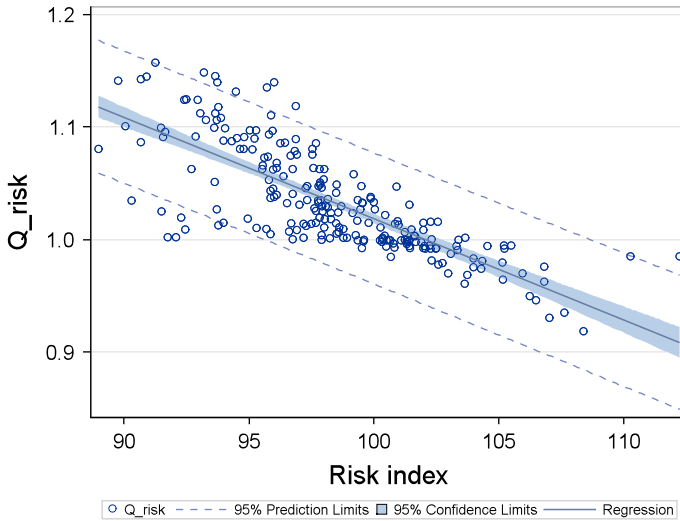


Fig. 1 Q_z versus z , contextual variable $z = \text{Risk}$

Since the variable RISK seems to show, among the z variables, the main contribution to explain the inefficiency not captured by StoNED, we evaluated the linearity assumption about RISK following the criterion given above and the result is given in Fig. 1: points are almost located within the confidence interval of a linear regressive model showing an evidence about linearity assumption.

Finally, we can now interpret the econometric results in terms of GSV. The difference between fitted values (production function or frontier) and observed values of output for inefficient units may be economically interpreted as a “whole system cost opportunity” reallocating resources in a relative sense: by reallocating resources, a higher economic well-being could be achieved without increasing the total resource use of the economy. The sum of the differences between inefficient farm’s production and the relative benchmark is a measure of the opportunity cost for the entire productive system. The estimated opportunity cost for the 229 citrus farms—based on the formulas given in Table 5—is shown in Table 6; inefficiencies are measured in terms of distance of PL from the average-practice estimated function (*mean*) and from the estimated production frontier (*frontier*), respectively.

A good modelling practice requires the evaluation of model reliability, possibly assessing the robustness associated with the modelling process and with the outcome of the model itself. In other terms, especially in applied analysis, following the idea “It is better to be vaguely right than exactly wrong” (Carveth 1901), the “last thing to do” is the study of how the uncertainty in the output has been due to different estimation models. The results provided by different methods are similar and robust with respect to the mean opportunity cost in terms of the distance from the *average-practice* and from the *frontier* production function, expect for the GAM-SFA which corresponds a lower average cost as shown in Table 6. The GAM-SFA approach seems to show greater flexibility due to its poor assumptions with respect to other frontier specifications and in addition it can provide proper diagnostics for the model evaluation.

Table 5 Opportunity cost in terms of average-practice function \hat{y} (left side) and production frontier \hat{y} (right side)

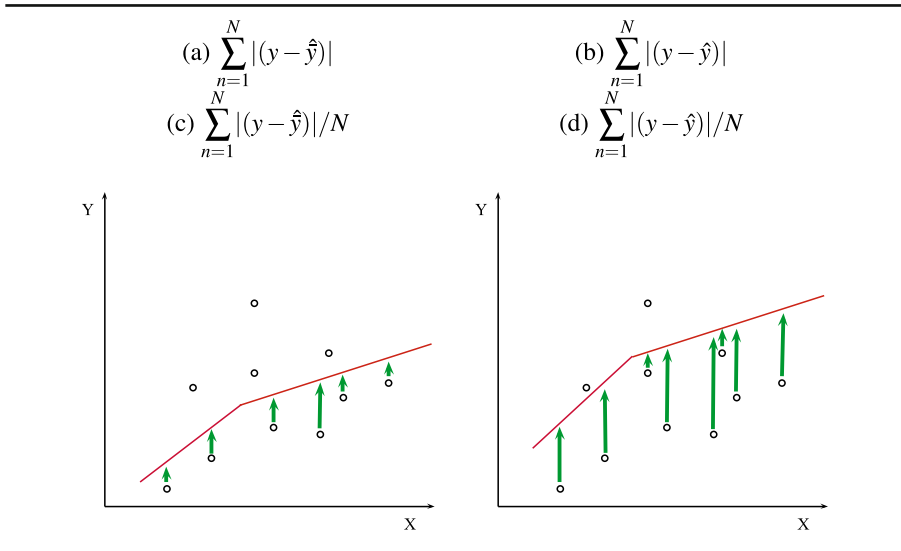


Table 6 Opportunity costs (€) calculated according to different estimation methods (source: INEA 2007)

Measure		StoNED	GAM-SFA	SFA
Total opportunity cost (frontier)	(b)	4,415,639	3,092,041	4,395,467
Mean opportunity cost (frontier)	(d)	23,740	21,623	25,704
Total opportunity cost (average-practice)	(a)	1,193,420	1,710,093	1,378,947
Mean opportunity cost (average-practice)	(c)	10,656	15,134	11,991

Moreover, regardless of the methodology, Table 6 points out the presence of a higher per-unit opportunity cost corroborating the recommendation about the need of optimizing social and financial resources.

4 Concluding remarks

In this paper, we have estimated efficiency levels in Italian agriculture with the purpose of introducing a new approach of Stochastic Frontier Analysis (SFA) by using a Generalized additive model (GAM) framework. The analysis, while starting from the literature on the productive efficiency estimation in the agricultural sector, extended its focus to include less tested and integrated aspects.

The Figge and Hahn (2005) approach, bridging a gap with classical efficiency estimation methods, turns out to be very modern in that it takes into account a multitude of aspects concerning production, in particular environmental and social aspects emphatically underlined in the EU Common Agricultural Policy.

The aim of this paper was to provide a contribution both to theoretical models in the field of semi-nonparametric estimation of efficiency and to knowledge related to territorial measures of efficiency in the agricultural sector.

From a theoretical point of view, we have proposed a model capable to overcome drawbacks due to the specification of a particular production function showing a greater flexibility due to its poor assumptions with respect to other frontier specifications; by this way, we have defined a family of models that permits of applying any class of nonparametric estimator of the inputs to the [Fan et al. \(1996\)](#) approach, ensuring the additivity of the inputs, separability assumptions and independence between noise and efficiency hypothesis.

Even if the StoNED-StoNEZD models provided a good economic interpretation of the coefficients as vector of marginal products of inputs for each unit, they need more efficient computational algorithm to solve the CNLS formulation and to be applied to practical problems on a wide scale; new algorithms have been recently proposed to solve StoNED computational issues ([Lee et al. 2011](#)), but the GAM-SFA model has still turned out to be easier to calculate.

In addition, since the effect of contextual variables on the model specification has been one of the issues covered in this paper, we propose a qualitative criterion to test the linearity hypothesis in the StoNEZD model.

From an applied point of view, this paper led to the outlining of an initial framework to investigate efficiency extending to production aspects not usually analysed and it allowed for the specification of the most appropriate measures and indicators for economic, environmental and social dynamics. The estimation, although based on different and contrasting assumptions, showed common traits and a good convergence of results, highlighting for the citrus sector a very consistent opportunity cost; the very extensive and local detailed INEA database should allow us to extend these estimates to other areas.

Other aspects to improve and explore further include the need to expand the database by cross-matching with data on undeclared work, working and social conditions and more accurate measurements on physical output and price levels in order to review the logic of information systems at the base of national surveys. Territorially disaggregated data would allow to include, in the proposed framework, supply chain variables and, more generally, to test the presence of Marshallian external economies given by the presence of a group of interdependent farms. Thus, information schema would not merely perform administrative functions, it would also provide information support for policies that will be more focused, more coherent and more quickly updated.

References

- ADVANCE (2008) Sustainable value of european industry: a value-based analysis of environmental performance of european manufacturing companies. Tech. rep, ADVANCE-project
- Aigner DJ, Chu SF (1968) On estimating the industry production function. *Am Econ Rev* 58:826–839
- Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. *J Econom* 6:21–37
- Carveth R (1901) *Logic, deductive and inductive*. Grant Richards, London
- Charnes A, Cooper WW, Rhodes E (1978) Measuring the inefficiency of decision making units. *EJOR* 2:429–444

- Cisilino F, Madau FA (2007) Organic and conventional farming: a comparison analysis through the Italian FADN. MPRA Paper 21786, University Library of Munich, Germany
- Coelli T (1995) Recent developments in frontier modelling and efficiency measurement. *Aust J Agr Econ* 39(3):219–245
- D'Amico F, Fernandez JL (2012) Measuring inefficiency in long-term care commissioning: evidence from english local authorities. *Appl Econ Perspect Policy* 34(2):275–299
- Daraio C, Simar L (2007) *Advanced robust and nonparametric methods in efficiency analysis*. Springer, Berlin
- De Muro P, Mazziotta M, Pareto A (2010) Composite indices of development and poverty: an application to mdgs. *Soc Indic Res* 104(1):1–18
- Deprins D, Simar L, Tulkens H (1984) Measuring labor inefficiency in post offices. In: Marchand M, Pestieau P, Tulkens H (eds) *The performance of public enterprises: concepts and measurements*. North-Holland, Amsterdam
- Du P, Parmeter C, Racine JS (2013) Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Stat Sinica* 23:1343–1372
- Fan Y, Li Q, Weersink A (1996) Semiparametric estimation of stochastic production frontier models. *J Bus Econ Stat* 14:460–468
- Farrell MJ (1957) The measurement of productive efficiency. *J R Stat Soc A Sta* 120(3):253–281
- Figge F, Hahn T (2005) The cost of sustainability capital and the creation of sustainable value by companies. *J Ind Ecol* 9:47–58
- Figge F, Hahn T (2009) Not measuring sustainable value at all: a response to Kuosmanen and Kuosmanen. *Ecol Econ* 69(2):244–249
- Giannakas K, Tran KC, Tzouvelekas V (2003) On the choice of functional form in stochastic frontier modeling. *Empir Econ* 28:75–100
- Greene W (2008) *The econometric approach to efficiency analysis*. In: Fried HO, Lovell CAK, Schmidt SS (eds) *The measurement of productive efficiency and productivity change*. Oxford University Press, Oxford
- Grosskopf S (1996) Statistical inference and nonparametric efficiency: a selective survey. *J Prod Anal* 7:161–176
- Hani F, Braga F, Stampfli A, Keller T, Fischer M, Porsche H (2003) Rise, a tool for holistic sustainability assessment at the farm level. *Int Food Agribus Manag Rev* 6(4):78–90
- Hastie T, Tibshirani R (1990) *Generalized additive models*. Chapman & Hall, London
- Hjalmarsson L, Kumbhakar S, Hesmati A (1996) Dea, dfa and sfa: a comparison. *J Prod Anal* 7:303–327
- Johnson A, Kuosmanen T (2011) One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent stoned method. *J Prod Anal* 36(2):219–230
- Jondrow J, Knox Lovell CA, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *J Econom* 19:233–238
- Kumbhakar SC, Lovell CAK (2000) *Stochastic frontier analysis*. Cambridge University Press, Cambridge
- Kuosmanen T, Kortelainen M (2012) Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *J Prod Anal* 38(1):1–18
- Kuosmanen T, Kuosmanen N (2009a) Role of benchmark technology in sustainable value analysis: an application to Finnish dairy farms. *Agric Food Sci* 18:302–316
- Kuosmanen T, Kuosmanen N (2009b) How not to measure sustainable value (and how one might). *Ecol Econ* 69(2):235–243
- Latruffe L (2010) *Competitiveness, productivity and efficiency in the agricultural and agri-food sectors*. Technical report, OECD Food, Agriculture and Fisheries Working Papers, No. 30, OECD Publishing
- Lee CY, Johnson L, Moreno-Centeno E, Kuosmanen T (2011) A more efficient algorithm for convex nonparametric least squares. Technical report, Texas A&M working paper
- Madau FA (2010) *Parametric estimation of technical and scale efficiencies in italian citrus farming*. MPRA Paper 26818, University Library of Munich, Germany
- Meeusen W, van den Broeck J (1977) Efficiency estimation from cobb–douglas production functions with composed error. *Int Econ Rev* 18:435–444
- OECD (2001) *Fostering productivity and competitiveness in agriculture*. OECD Publishing
- R Development Core Team (2012) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0

- Reig-Martinez E, Gomez-Limon JA, Picazo-Tadeo AJ (2011) Ranking farms with a composite indicator of sustainability. *Agr Econ* 42(5):561–575
- Simar L, Wilson PW (2007) Estimation and inference in two-stage, semi-parametric models of production processes. *J Econom* 136(1):31–64
- Sperlich S, Tjostheim D, Yang L (2002) Nonparametric estimation and testing of interaction in additive models. *Econom Theor* 18:197–251
- Stone CJ (1986) The dimensionality reduction principle for generalized additive models. *Ann Stat* 14: 590–606
- Tyteca D (1998) Sustainability indicators at the firm level: pollution and resource efficiency as a necessary condition toward sustainability. *J Ind Ecol* 2(4):61–77
- Van Passel S, Huylenbroeck GV, Lauwers L, Mathijs E (2009) Sustainable value assessment of farms using frontier efficiency benchmarks. *J Environ Manag* 90(10):3057–3069
- Wang HJ, Schmidt P (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *J Prod Anal* 18:129–144
- Wood SN (2003) Thin plate regression splines. *J R Stat Soc B Met* 65(1):95–114
- Wood SN (2006) An introduction to generalized additive models with R. CRC Press, Boca Raton
- Wood SN (R Development Core Team) (2012) *mgcv: mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation*, R package version 1.7-21