CEIS TOR VERGATA Research Papers

Università di Roma Tor Vergata

ISSN 2610-931X

CEIS Tor Vergata

RESEARCH PAPER SERIES

Vol. 17, Issue 5, No. 462 – January 2021

Imposing Monotonicity in Stochastic Frontier Models: An Iterative Nonlinear Least Squares Procedure

Federico Belotti and Giancarlo Ferrara

Imposing monotonicity in stochastic frontier models: an iterative nonlinear least squares procedure

Federico Belotti[a] and Giancarlo Ferrara[b,c]

ABSTRACT
Despite its importance, the monotonicity condition is typically overlooked in stochastic frontier analysis. This article illustrates a straightforward and useful method for the estimation of semiparametric stochastic frontier models imposing such constraint and incorporating exogenous inefficiency effects exploiting the scaling property. An iterative estimation algorithm based on nonlinear least squares is developed and the behavior of the proposed procedure is investigated through a set of Monte Carlo experiments comparing its finite sample properties with those of available alternatives. The simulation results highlight very good performance of the new algorithm which outperforms the competitors in small samples and in presence of small signal-to-noise ratios. Our results also show that the fraction of observations for which monotonicity naturally holds is generally quite small if this condition is not imposed. An application based on FADN agricultural data illustrates the usefulness of the proposed algorithm.

KEYWORDS
Efficiency, stochastic frontiers, monotonicity, nonlinear least squares, heteroskedasticity, simulation

1. Introduction

The Stochastic Frontier (SF) model, introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977), represents a very popular tool for efficiency analysis.[1] In its basic formulation, it is based on a parametric representation of production technology along with a compounded error term in which one component, often assumed to be normally distributed, deals with the randomness of production while the other, following a specific one-sided distribution, captures the inefficiency of productive units. Several methodological improvements of this basic formulation have been proposed in the last three decades, most of which deal with relaxing the restrictive assumptions about the frontier functional form and the inefficiency term.[2]

As for the flexible specification of production technology in a cross-sectional setting, Fan et al. (1996) introduced a two-step pseudo-likelihood procedure in which the production technology is estimated via kernel regression, while Kumbhakar et al.

Corresponding author: Giancarlo Ferrara. Email: giancarlo.ferrara@gmail.com. This paper was previously circulated under the title "Flexible estimation of heteroskedastic stochastic frontier models via two-step iterative nonlinear least squares". The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of SOSE S.p.A.. [a]Department of Economics and Finance, University of Rome Tor Vergata, Italy. [b]SOSE - Soluzioni per il Sistema Economico SpA, Rome, Italy; [c] Department of Economics, Business and Statistics, University of Palermo, Italy.

[1]Production frontiers represent the maximum amount of output that can be obtained from a given level of inputs, while a cost frontier characterizes the minimum expenditure required to produce a bundle of outputs given the prices of the inputs used in its production.

[2]See Parmeter and Kumbhakar (2014) for a recent survey.

(2007) proposed an alternative nonparametric approach based on the local maximum likelihood method which has been shown to be robust to arbitrary heteroskedasticity. However, both these approaches suffer from two drawbacks: *i*) they do not directly incorporate inefficiency determinants in the model, and *ii*) they do not ensure that the frontier function is monotonic, i.e. it monotonically increases in all production inputs. The first is a serious flaw since both frontier parameters and inefficiency estimates may be adversely affected when inefficiency determinants are neglected. As for the second, a non-monotone frontier may upset the final ranking of productive units, thus inhibiting a reasonable interpretation of the unit-specific efficiency scores (Henningsen and Henning, 2009). Nevertheless, the monotonicity condition is typically overlooked in stochastic frontier analysis.

In this paper, we propose an iterative algorithm in which the first step takes care of estimating the production function through any semi- or nonparametric monotonic regression approach, while the second exploits the nonlinear least squares (NLS) method for estimating the parameters of the compounded error term, including the effect of inefficiency determinants. These two steps are then iterated until an objective function is minimized. When the inefficiency determinants and the inputs are correlated, this iteration allows for reducing the bias characterizing currently available procedures (Wang and Schmidt, 2002).[3]

Albeit the proposed framework could exploit any flexible regression approach for the first step, we consider a Generalized Additive Model (GAM) specification in which monotonicity constraints are imposed by adopting the penalized B-splines approach, known as P-splines (Eilers and Marx, 1996; Ferrara and Vidoli, 2017).[4] Our approach is alternative to that of Parmeter et al. (2017) in which the production technology is parametrically specified and all the distributional assumptions on the error term are relaxed; see also Tran and Tsionas (2009) for an alternative nonparametric specification of the function describing the mean of technical inefficiency.

It is worth emphasizing that the standard "one-step" NLS (Kumbhakar and Lovell, 2000) or maximum likelihood (ML) approaches do not allow for a semiparametric/nonparametric monotonic specification of the production technology. Indeed, imposing the monotonicity constraint while allowing for inefficiency effects generally requires multiple steps or iterative schemes. To the best of our knowledge the Henningsen and Henning (2009) procedure is the only three-step approach for estimating a monotonic and parametric production function in presence of contextual variables.[5]

We carry out a Monte Carlo simulation analysis comparing the performance of our algorithm with those of one-step ML, one-step NLS and Henningsen and Henning (2009) three-step ML-based procedure assuming a translog specification. To our knowledge this is the first study providing evidence on the finite sample performances of one-step NLS and Henningsen and Henning (2009) procedure. Our simulation results show that, for small samples and in presence of small signal-to-noise ratios, the proposed algorithm outperforms standard one-step NLS and ML approaches; moreover, we find that not enforcing the monotonicity condition may lead to a non negligible

[3]Notice that, as pointed out by Wang and Schmidt (2002), even if the inefficiency determinants and the inputs are uncorrelated, when the former are omitted from the model the estimated inefficiencies are underdispersed causing the second-step estimate of the inefficiency effects to be downward biased.

[4]For a detailed description of the P-spline approach to ensure monotonicity see Bollaerts et al. (2006) and Muggeo and Ferrara (2008). For nonparametric kernel regression subject to monotonicity constraints see Hall and Huang (2001).

[5]Constrained ML and Bayesian MCMC approaches have been proposed but they are rarely used because rather complex and not readily available to practitioners, see among others O'Donnell and Coelli (2005) and Terrell (1996) for early approaches on monotone cost functions.

bias, especially in the estimation of the inefficiency effects. Given that the fraction of observations for which the monotonicity condition naturally holds is generally small, more attention should be devoted to enforcing *ex ante* or checking *ex post* this condition (Sauer et al., 2019). We apply the proposed iterative algorithm to deal with model specification issues using FADN agricultural data.

The plan of the paper is as follows. Section 2 provides the details of the proposed estimation algorithm. To investigate its finite sample properties, Section 3 provides the results of a Monte Carlo analysis, followed by an empirical application in Section 4; finally, Section 5 concludes.

2. Methods

In a cross-sectional setting, the conventional stochastic frontier model of Aigner et al. (1977) and Meeusen and van den Broeck (1977) can be written as

$$y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + v_i - u_i, \qquad i = 1, \ldots, n, \tag{1}$$

where y_i is the observed single output of unit i, $\boldsymbol{x}_i$ is the corresponding vector of inputs, $f(\cdot)$ denotes a production (frontier) relationship between the output and the inputs, $\boldsymbol{\beta}$ is the unknown vector used to parametrize $f(\cdot)$, v_i is a symmetric idiosyncratic error and $u_i > 0$ is a one-sided error term representing technical inefficiency[6]. This model is commonly estimated by ML methods assuming that v_i and u_i are i.i.d. with $v_i \sim N(0, \sigma_v^2)$ and $u_i \sim N^+(0, \sigma_u^2)$.

The presence of inefficiency determinants, such as firm size or managerial skills, is handled by relaxing the constant variance assumption on the the inefficiency term. In particular, σ_u^2 is specified as a function of some explanatory variables potentially associated with technical inefficiency; different specifications have been considered in literature, see for instance Kumbhakar et al. (1991), Reifschneider and Stevenson (1991) and Huang and Liu (1994).
Here, we assume that the inefficiency term depends on a set of exogenous covariates *via* the so called "scale" transformation (Wang and Schmidt, 2002) of some underlying process η_i, such as

$$u_i = \exp(\boldsymbol{z}_i \boldsymbol{\gamma}) \eta_i, \tag{2}$$

where $\boldsymbol{z}_i$ is a vector of contextual variables, $\boldsymbol{\gamma}$ the vector of unknown coefficients and η_i an i.i.d. random variable that does not depend on $\boldsymbol{z}_i$ with $\eta_i \geq 0$, $E(\eta_i) = 1$ and finite variance (Kumbhakar and Lovell, 2000). A single element of $\boldsymbol{\gamma}$ is positive (negative) when the corresponding element of $\boldsymbol{z}$ has a negative (positive) effect on technical efficiency. This formulation has the computational advantage of automatically constraining σ_{ui} to be positive and to directly nest the homoskedastic case. Moreover, changes in $\boldsymbol{z}_i$ change the scale but not the shape of the distribution of u_i (Alvarez et al., 2006). From a computational point of view, the main property of this formulation is that the parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ can be estimated by NLS without relying on distributional

[6] A cost frontier model can be obtained by changing the sign of the inefficiency component.

assumptions. Indeed, by substituting equation (2) into (1) leads to

$$\begin{aligned} y_i &= f(\boldsymbol{x}_i;\boldsymbol{\beta}) + v_i - \exp(\boldsymbol{z}_i\boldsymbol{\gamma})\eta_i, \\ &= f(\boldsymbol{x}_i;\boldsymbol{\beta}) - \exp(\boldsymbol{z}_i\boldsymbol{\gamma}) + \varepsilon_i, \end{aligned} \tag{3}$$

where $\varepsilon_i = y_i - E(y_i) = v_i - \exp(\boldsymbol{z}_i\boldsymbol{\gamma})(\eta_i - 1)$ and the ε_i's are independent but not identically distributed and $E(\varepsilon_i) = 0$ Kumbhakar and Lovell (see 2000, for more details).[7]

The parameters in (3) are then estimated by minimizing the following criterion

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = min \sum_{i=1}^{n} \left[(y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta}) + \exp(\boldsymbol{z}_i\boldsymbol{\gamma})\right]^2, \tag{4}$$

where $f(\boldsymbol{x}_i;\boldsymbol{\beta})$ is usually assumed to be a known parametric function (e.g. *translog* or *Cobb-Douglas*). Our aim with this paper is to allow for any frontier specification, i.e. $f(\boldsymbol{x}_i;\boldsymbol{\beta}) = f(\boldsymbol{x}_i)$, under monotonicity and incorporating inefficiency effects. To this aim, we propose to minimize (4) by exploiting the following algorithm:

1) set an initial guess for $f(\boldsymbol{x}_i)$;
2) given the current value of $\hat{f}(\boldsymbol{x}_i)$, estimate $\boldsymbol{\gamma}$ by NLS;
3) given the current value of $\hat{\boldsymbol{\gamma}}$, fit a regression model to obtain an updated estimate of $\hat{f}(\boldsymbol{x}_i)$;
4) repeat steps 2) and 3) until some convergence criterion is met.

The key advantage of this iterative nonlinear least squares (INLS) procedure is that it permits a more flexible specification of the production function $f(\boldsymbol{x}_i)$. Indeed, any semi- or nonparametric regression approach that satisfy monotonicity can be used in step 3), even when the monotonicity constraints require multiple steps or iterative schemes to be imposed. Notice that, in step 2), assuming $f(\boldsymbol{x}_i)$ as known corresponds to consider the "working" partial residual given by $y - \hat{f}(\boldsymbol{x}_i)$ as response variable. Further, while the NLS is less efficient than the ML estimator, it allows to construct such an iterative scheme, thus reducing the bias of $\boldsymbol{\gamma}$ in presence of correlation between $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$. The above algorithm can be actually applied to any NLS problem when a constrained semi/nonparametric specification for $f(\cdot)$ is required.

By restricting $f(\cdot)$ in equation (4) to be fully parametric can be too restrictive, even inappropriate, and may lead to biased estimates and misleading conclusions about the link between inputs and output. Giannakas et al. (2003) pointed out that technical efficiency measures are very sensitive to the choice of functional specification and that the identification of the most appropriate functional form among a set of parametric alternatives might not always be feasible. A more flexible nonparametric alternative that is referred to as the Generalized Additive Model takes the form

$$E[Y|x] = \psi(\boldsymbol{x}) = s_1(x_1) + \ldots + s_p(X_p), \tag{5}$$

[7] Model (3) can also be viewed as a partially nonlinear model, see for instance Hardle et al. (2000) and Carroll et al. (1997) for partial and generalized partial linear models, respectively, and Henderson and Parmeter (2015) for a full review. As in Tran and Tsionas (2009) and Parmeter et al. (2017), a drawback of model (3) is that the model intercept cannot be separately identified from the conditional mean of the inefficiency component. Nevertheless, Zhou et al. (2020) pointed out that this is not a relevant empirical issue for three reasons: *i*) practitioners are typically interested to the input elasticities, which do not require the identification of the intercept; *ii*) we can still compare productive units efficiency even if we cannot directly interpret the inefficiency level of each unit; *iii*) we are still able to estimate the marginal effects of $\boldsymbol{z}$ on $E(u|\boldsymbol{z})$.

where each $s_j(\cdot)$ is a smooth function of the x_j variable estimated from the data (Hastie and Tibshirani, 1990), standardized so that $E(s_j(x_j)) = 0$.
By allowing nonlinear dependence between output and inputs, the GAM specification may improve the overall fit compared to a fully parametric specification. The model remains fairly straightforward to interpret: the partial response function describes how the prediction of the outcome changes with observables. Further, the gradients of the nonparametric model can be interpreted as partial output elasticities and their sum as elasticity of scale (Henningsen and Kumbhakar, 2009).

If we also assume that $\eta_i \sim N^+(0, \sigma^2)$, it follows that $u_i \sim N^+(0, \sigma^2_{ui})$ with $\sigma_{ui} = \sqrt{(\pi/2)}\exp(\boldsymbol{z}_i\boldsymbol{\gamma})$ and the variance of the noise component v given by

$$\sigma^2_v = n^{-1}\sum_{i=1}^{n}\Big(\varepsilon^2_i - \exp(2\boldsymbol{z}_i\boldsymbol{\gamma})(\pi/2 - 1)\Big). \tag{6}$$

By denoting $\lambda_i = \sigma_{ui}/\sigma_v$ and $\sigma^2_i = \sigma^2_{ui} + \sigma^2_v$, estimates of unit specific inefficiency scores can be obtained through the Jondrow et al. (1982) estimator

$$E[u_i|\varepsilon_i] = \frac{\sigma_i\lambda^2_i}{1+\lambda^2_i}\Big[\frac{\phi(\omega_i)}{1-\Phi(\omega_i)} - \omega_i\Big], \omega_i = \lambda_i\frac{y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})}{\sigma_i}, \tag{7}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and cumulative distribution functions, respectively. Efficiency measure can then be obtained from $\exp(-E[u_i|\varepsilon_i])$.

The Monte Carlo analysis and the empirical application presented in the following sections have been carried out using the `R` Environment R Core Team (2020).[8]

3. Monte Carlo evidence

In this Section we study the behaviour of the proposed INLS algorithm and compare its finite sample performance with those of one-step ML, one-step NLS and the Henningsen and Henning (2009) procedure. Among these competitors, only the latter allows to impose monotonicity constraints.[9]

We investigate the effect of different sample sizes ($n = 100, 500, 1000$), different signal-to-noise ratios ($\bar{\lambda} = 1, 2$) and different degree of correlation between exogenous determinants of inefficiency and the inputs ($\phi = 0.3, 0.9$). We consider two data generating processes (DGPs). The first (DGP_1) is represented by the following heteroskedastic SF model in which the frontier function is linear

$$\begin{aligned} y_i &= 3 + 2x_i + v_i - u_i, \quad i = 1, \ldots, n, &(8)\\ v_i &\sim \mathcal{N}(0, \sigma^2_v), &(9)\\ \eta_i &= \mathcal{N}^+(0, \pi/2), &(10)\\ u_i &\sim \eta_i\exp(\gamma z_i), &(11)\end{aligned}$$

where we set $\gamma = 0.5$ and different (average) signal-to-noise ratios ($\bar{\lambda} = \frac{\bar{\sigma}_u}{\sigma_v} = 1, 2$) are obtained by setting $\sigma^2_v = (0.396, 1.587)$. Different degrees of correlation between

[8] For a recent review of SF analysis using R see Ferrara (2020).

[9] Henningsen and Henning (2009) procedure is based on the following three steps: i) the estimation of the unconstrained SF model; ii) the estimation of the constrained parameters by minimum distance; iii) estimation of inefficiency effects and scores.

the input x_i and the inefficiency factor z_i are obtained by exploiting the Cholesky factorization of the correlation matrix $\boldsymbol{C} = \boldsymbol{\phi I_2} = \boldsymbol{L'L}$ with ϕ representing the correlation coefficient between x_i and z_i and $\boldsymbol{I}_2$ a 2×2 identity matrix. First, two i.i.d. random variables have been drawn from the standard normal distribution, $\boldsymbol{s}_i = (s_{i1}, s_{i2})$, $i = 1, \ldots, n$; then the correlated variables are obtained as $\tilde{\boldsymbol{S}} = \boldsymbol{SL}$. Finally, $x_i = \Phi(\tilde{s}_{i1}) + 1$ while $z_i = \tilde{s}_{i2}$.

As for the second DGP (DGP_2), we consider the following heteroskedastic SF model in which the frontier function is highly nonlinear

$$y_i = 1 + 0.7\sqrt{x_{i1}} + x_{i2} + \frac{\sin(2\pi x_{i2})}{(2\pi)} + \frac{1}{1 + \exp(5 - 10x_{i3})} + v_i - u_i, \quad (12)$$

$$v_i \sim \mathcal{N}(0, \sigma_v^2), \quad (13)$$

$$\eta_i = \mathcal{N}^+(0, \pi/2), \quad (14)$$

$$u_i \sim \eta_i \exp(z_i\gamma) \quad i = 1, \ldots, n. \quad (15)$$

As for the first DGP, different (average) signal-to-noise ratios $\bar{\lambda} = 1, 2$ are obtained by setting σ_v^2, different degrees of correlation between the input x_{i1} and the inefficiency factor z_i are obtained by exploiting the Cholesky factorization, while $x_{i2} \sim \mathcal{U}(3)$ and $x_{i3} \sim \mathcal{U}(4)$ are uncorrelated with z_i. For this second DGP, a translog functional form is assumed when the estimation is performed via one-step ML, one-step NLS and the three-step procedure of Henningsen and Henning (2009, ML-MONO). It is worth noting that the monotonicity condition is not naturally fulfilled by the translog technology when estimation is performed via one-step ML or NLS.

Simulation results are summarized by reporting the average bias and Mean Squared Error (MSE) of the estimates over replications, together with the linear ($r_{u,\hat{u}}$) and the Spearman rank ($s_{u,\hat{u}}$) correlation coefficients between the (true) simulated inefficiencies and the estimated ones. The estimated inefficiency's bias and MSE are also reported and are computed for each replication over the n observations, and then these quantities are averaged over replications, e.g., MSE($\hat{u}_i$)$= R^{-1}\sum_{r=1}^{R}(n)^{-1}\sum_{i=1}^{n}(\mathbb{E}(u_i|\hat{\varepsilon}_i) - u_i^0)^2$, where $\mathrm{E}(u_i|\hat{\varepsilon}_i)$ is the JLMS estimate and u_i^0 is the simulated (true) inefficiency for unit i. For the second DGP, we also report, for the one-step ML and NLS approaches, the average proportion of units $P(m)$ for which the monotonicity condition is fulfilled; by indicating the elasticity of y with respect to x_{ik} ($k = 1, 2, 3$) with $e_{y,x_{ik}}$, $P(m) = R^{-1}\sum_{r=1}^{R}(n)^{-1}\sum_{i=1}^{n} 1\left[e_{y,x_{i1}} \geq 0 \;\&\; e_{y,x_{i2}} \geq 0 \;\&\; e_{y,x_{i3}} \geq 0\right]$, where $1[\cdot]$ is the indicator function.[10] Finally, we use for each experiment $R = 1000$ replications.

Simulation results are reported in Tables from 1 to 2 for DGP_1 and from 3 to 4 for DGP_2, for the two levels of correlation $\phi = 0.3, 0.9$, respectively.

Overall, the proposed algorithm shows very good performances: an increase in the cross-sectional dimension produces significant reductions in both bias and MSE of both the parameters of interest, γ and σ_v. Even for large value of ϕ, both parameters are accurately estimated by the proposed algorithm regardless of the signal-to-noise ratio. Interestingly, we find that, in presence of very small samples and small signal-to-noise ratios ($\bar{\lambda} = 1$), the INLS algorithm always outperforms the one-step competitors (NLS and ML) and is in line with three-step procedure (ML-MONO).

As far as DGP_1 is concerned, the bias of γ is almost the same compared to NLS and ML, even though, as expected, ML provides a more efficient estimation. As for σ_v, NLS and INLS show very similar performances. In particular, the bias appears to

[10] The monotonicity condition is automatically fulfilled for INLS thanks to the adoption of the P-splines method.

be higher than ML, even thought it is numerically negligible when $n > 100$, ranging from 2 to 3.5 percent when $\bar{\lambda} = 1$ and from 1.5 to 2 percent when $\bar{\lambda} = 2$. The speed of convergence to the true values of the INLS algorithm appears to be lower when $\phi = 0.9$. This last result may be due to the unfairness of DGP_1 for INLS. Interestingly, when the true production function is nonlinear (DGP_2), INLS appears to be superior in the estimation of both γ and σ_v compared to the standard NLS regardless of $\bar{\lambda}$ and ϕ. In this case, ML and the proposed algorithm show comparable performances in the estimation of γ even with $\phi = 0.9$ while, as before, ML provides better results for the σ_v parameter; we find a similar evidence when comparing our algorithm with ML-MONO. It is worth emphasizing that, overall, the monotonicity condition is not fulfilled by the translog technology when the latter is not constrained as in ML-MONO. The fraction of observations for which this condition holds is generally quite small, ranging from 25-30 percent when $n = 100$ to 40-45 percent when $n = 1000$. We also find that this fraction is slightly higher when estimation is performed by ML. Furthermore, our simulations suggest that ignoring the monotonicity requirement may lead to a non negligible bias, especially in small sample and for the estimation of the inefficiency effects. This evidence confirms the Henningsen and Henning (2009) "empirical" argument that "... non-monotone production frontier inhibits not only a reasonable interpretation of the individual (relative) efficiency estimates, but also the analysis of factors that might affect technical (in)efficiency".

As expected we find that the inefficiencies are inconsistently estimated regardless of the approach. Nevertheless, we find that the proposed algorithm generally outperforms its competitors, especially in small samples. Consistently with previous literature, the performance of the JLMS estimator are driven by the signal-to-noise ratio, showing smaller bias when $\bar{\lambda} = 2$.

Table 1. Simulation results for INLS, NLS and ML when $\phi = 0.3$ (DGP_1).

$\bar{\lambda} = 1$

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0118	0.0223	-0.0131	0.0224	-0.0194	0.0155
σ_v	-0.0981	0.0499	-0.0833	0.0457	-0.0225	0.0155
$\hat{y}$	-0.0036	0.0627	-0.0030	0.0514	0.0025	0.0456
E(u|ε)	-0.1294	0.5191	-0.1317	0.5183	-0.1453	0.5056
$r_{u\hat{u}}$	0.7805		0.7814		0.7822	
$s_{u\hat{u}}$	0.6197		0.6203		0.6205	
β_0			-0.0161	0.7327	0.0076	0.6241
β_1			0.0128	0.3360	-0.0067	0.2811

$\bar{\lambda} = 2$

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0360	0.0147	-0.0362	0.0147	-0.0189	0.0102
σ_v	-0.0683	0.0422	-0.0528	0.0386	-0.0149	0.0081
$\hat{y}$	0.0090	0.0299	0.0091	0.0238	0.0013	0.0176
E(u|ε)	-0.0652	0.2533	-0.0686	0.2508	-0.0774	0.2326
$r_{u\hat{u}}$	0.8951		0.8964		0.8991	
$s_{u\hat{u}}$	0.7710		0.7727		0.7755	
β_0			0.0400	0.3606	0.0229	0.2297
β_1			-0.0329	0.1641	-0.0165	0.1030

(a) $n = 100$

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0001	0.0049	0.0001	0.0049	-0.0033	0.0029
σ_v	-0.0211	0.0069	-0.0188	0.0068	-0.0055	0.0028
$\hat{y}$	-0.0051	0.0126	-0.0052	0.0100	-0.0023	0.0089
E(u|ε)	-0.1422	0.4884	-0.1423	0.4881	-0.1457	0.4866
$r_{u\hat{u}}$	0.8017		0.8018		0.8016	
$s_{u\hat{u}}$	0.6266		0.6267		0.6264	
β_0			-0.0018	0.1435	0.0100	0.1262
β_1			0.0046	0.0640	-0.0051	0.0551

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0012	0.0039	-0.0011	0.0039	-0.0028	0.0019
σ_v	-0.0368	0.0157	-0.0342	0.0154	-0.0043	0.0014
$\hat{y}$	-0.0035	0.0063	-0.0036	0.0048	-0.0020	0.0034
E(u|ε)	-0.0674	0.2282	-0.0679	0.2277	-0.0780	0.2223
$r_{u\hat{u}}$	0.9110		0.9113		0.9121	
$s_{u\hat{u}}$	0.7823		0.7826		0.7834	
β_0			-0.0032	0.0711	0.0065	0.0464
β_1			0.0045	0.0325	-0.0030	0.0204

(b) $n = 500$

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	0.0004	0.0026	0.0005	0.0026	-0.0014	0.0014
σ_v	-0.0148	0.0038	-0.0138	0.0038	-0.0047	0.0016
$\hat{y}$	-0.0017	0.0068	-0.0018	0.0055	-0.0004	0.0048
E(u|ε)	-0.1458	0.4894	-0.1459	0.4892	-0.1478	0.4882
$r_{u\hat{u}}$	0.8032		0.8033		0.8032	
$s_{u\hat{u}}$	0.6267		0.6267		0.6265	
β_0			0.0046	0.0777	0.0075	0.0691
β_1			-0.0019	0.0349	-0.0047	0.0300

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	0.0003	0.0021	0.0004	0.0021	-0.0006	0.0009
σ_v	-0.0241	0.0080	-0.0230	0.0080	-0.0037	0.0008
$\hat{y}$	-0.0013	0.0036	-0.0013	0.0028	-0.0007	0.0019
E(u|ε)	-0.0733	0.2251	-0.0735	0.2249	-0.0795	0.2223
$r_{u\hat{u}}$	0.9128		0.9129		0.9133	
$s_{u\hat{u}}$	0.7835		0.7837		0.7838	
β_0			0.0041	0.0400	0.0026	0.0269
β_1			-0.0018	0.0184	-0.0012	0.0116

(c) $n = 1000$

Table 2. Simulation results for INLS, NLS and ML when $\phi = 0.9$ (DGP_1).

$\bar{\lambda} = 1$

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0433	0.0478	-0.0590	0.0641	-0.0462	0.0290
σ_v	-0.0998	0.0604	-0.0866	0.0528	-0.0217	0.0156
$\hat{y}$	-0.0011	0.0988	-0.0018	0.1019	0.0085	0.0652
$E(u\|\varepsilon)$	-0.1281	0.5786	-0.1307	0.5894	-0.1519	0.5307
$r_{u\hat{u}}$	0.7613		0.7427		0.7729	
$s_{u\hat{u}}$	0.6060		0.5912		0.6141	
β_0			0.2102	1.8980	0.1761	1.0173
β_1			-0.1388	0.9083	-0.1236	0.4917

$\bar{\lambda} = 2$

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0908	0.0354	-0.0853	0.0360	-0.0302	0.0145
σ_v	-0.0236	0.0390	-0.0248	0.0395	-0.0163	0.0084
$\hat{y}$	0.0247	0.0508	0.0213	0.0464	0.0019	0.0249
$E(u\|\varepsilon)$	-0.0928	0.2912	-0.0893	0.2875	-0.0800	0.2410
$r_{u\hat{u}}$	0.8889		0.8880		0.8972	
$s_{u\hat{u}}$	0.7629		0.7630		0.7727	
β_0			0.3500	0.8353	0.0996	0.3590
β_1			-0.2476	0.4087	-0.0680	0.1749

(a) $n = 100$

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0206	0.0128	-0.0087	0.0104	-0.0060	0.0045
σ_v	-0.0186	0.0104	-0.0223	0.0097	-0.0058	0.0028
$\hat{y}$	0.0017	0.0255	-0.0035	0.0197	-0.0016	0.0122
$E(u\|\varepsilon)$	-0.1472	0.5132	-0.1422	0.5039	-0.1463	0.4910
$r_{u\hat{u}}$	0.8013		0.8017		0.8021	
$s_{u\hat{u}}$	0.6256		0.6258		0.6258	
β_0			0.0292	0.3344	0.0259	0.1952
β_1			-0.0172	0.1651	-0.0163	0.0915

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0225	0.0091	-0.0128	0.0076	-0.0033	0.0026
σ_v	-0.0233	0.0229	-0.0308	0.0216	-0.0047	0.0015
$\hat{y}$	0.0050	0.0146	0.0006	0.0109	-0.0019	0.0046
$E(u\|\varepsilon)$	-0.0766	0.2469	-0.0711	0.2409	-0.0783	0.2237
$r_{u\hat{u}}$	0.9101		0.9105		0.9118	
$s_{u\hat{u}}$	0.7822		0.7827		0.7827	
β_0			0.0522	0.1849	0.0111	0.0678
β_1			-0.0352	0.0945	-0.0062	0.0325

(b) $n = 500$

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0116	0.0070	-0.0018	0.0054	-0.0030	0.0020
σ_v	-0.0129	0.0064	-0.0165	0.0055	-0.0043	0.0016
$\hat{y}$	0.0019	0.0162	-0.0026	0.0116	-0.0002	0.0062
$E(u\|\varepsilon)$	-0.1486	0.5051	-0.1445	0.4973	-0.1482	0.4906
$r_{u\hat{u}}$	0.8036		0.8038		0.8038	
$s_{u\hat{u}}$	0.6267		0.6265		0.6262	
β_0			0.0045	0.1913	0.0177	0.0977
β_1			-0.0013	0.0970	-0.0117	0.0458

	INLS		NLS		ML	
	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0119	0.0050	-0.0039	0.0040	-0.0015	0.0012
σ_v	-0.0197	0.0151	-0.0261	0.0129	-0.0038	0.0008
$\hat{y}$	0.0034	0.0095	-0.0005	0.0066	-0.0006	0.0023
$E(u\|\varepsilon)$	-0.0773	0.2377	-0.0727	0.2323	-0.0799	0.2228
$r_{u\hat{u}}$	0.9123		0.9127		0.9133	
$s_{u\hat{u}}$	0.7834		0.7837		0.7835	
β_0			0.0175	0.1092	0.0084	0.0337
β_1			-0.0113	0.0568	-0.0052	0.0159

(c) $n = 1000$

4. Empirical application

To deal with the model specification issue, researchers usually do specification searches by comparing different models. With this in mind, we exploit an empirical application to show the usefulness of the proposed algorithm for model selection purposes.

We consider the estimation of model (1) where $f(\boldsymbol{x}_i; \boldsymbol{\beta})$ is specified by: *i*) nonparametric kernel without monotonicity constraints [NPK], *ii*) GAM with monotonic P-splines [PSM], *iii*) parametric translog with monotonicity constraint [PTM]. For *i*) and *ii*) we exploit the proposed INLS algorithm while *iii*) is estimated *via* the Henningsen and Henning (2009) procedure.
Albeit we can virtually exploit any existing nonparametric regression technique as the first step of our INLS procedure, we adopt here the nonparametric kernel approach without constraints to evaluate if monotonicity is "naturally" satisfied at all data points.[11] In all the cases, we assume a heteroskedastic inefficiency, i.e. $u_i \sim N^+(0, \sigma^2_{ui})$, exploiting the scaling property as in equation (3). Finally, inputs and output are specified in logs allowing us to interpret the first derivatives of the conditional mean with respect to the inputs as output elasticities and to test for monotonicity across models.

The Farm Accountancy Data Network (FADN) is an annual survey carried out by

[11] We use a local-linear kernel regressor with an Epanechnikov kernel for the (continuous) regressors as described in Hayfield and Racine (2008) with bandwidth specification using the method of Racine and Li (2004).

Table 3. Simulation results for INLS, NLS, ML and ML-MONO when $\phi = 0.3$ (DGP_2).

$\bar{\lambda} = 1$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0035	0.0202	-0.0147	0.0223	-0.0262	0.0164	-0.0143	0.0145
σ_v	-0.1155	0.0525	-0.1687	0.0699	-0.0990	0.0255	-0.0531	0.0178
$\hat{y}$	-0.0153	0.0968	-0.0105	0.2671	-0.0028	0.2460	-0.0087	0.1105
E(u|ε)	-0.1233	0.5147	-0.1234	0.5424	-0.1407	0.5254	-0.1383	0.5043
$r_{u\hat{u}}$	0.7806		0.7672		0.7699		0.7805	
$s_{u\hat{u}}$	0.6162		0.6038		0.6061		0.6157	
$P(m)$			0.2560		0.2569			

$\bar{\lambda} = 2$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0425	0.0143	-0.0503	0.0158	-0.0271	0.0111	-0.0190	0.0102
σ_v	-0.0515	0.0323	-0.1059	0.0452	-0.0667	0.0129	-0.0311	0.0083
$\hat{y}$	0.0063	0.0526	0.0098	0.1416	-0.0026	0.1052	-0.0056	0.0543
E(u|ε)	-0.0671	0.2564	-0.0592	0.2927	-0.0669	0.2611	-0.0692	0.2415
$r_{u\hat{u}}$	0.8914		0.8742		0.8834		0.8932	
$s_{u\hat{u}}$	0.7651		0.7409		0.7523		0.7662	
$P(m)$			0.2809		0.2973			

(a) $n = 100$

$\bar{\lambda} = 1$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0026	0.0043	-0.0047	0.0044	-0.0056	0.0027	-0.0029	0.0026
σ_v	-0.0233	0.0081	-0.0301	0.0084	-0.0169	0.0031	-0.0100	0.0029
$\hat{y}$	-0.0044	0.0277	-0.0032	0.0610	-0.0020	0.0556	-0.0036	0.0338
E(u|ε)	-0.1422	0.4926	-0.1431	0.4972	-0.1454	0.4932	-0.1443	0.4904
$r_{u\hat{u}}$	0.7991		0.7972		0.7973		0.7985	
$s_{u\hat{u}}$	0.6245		0.6229		0.6229		0.6237	
$P(m)$			0.3429		0.3481			

$\bar{\lambda} = 2$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0039	0.0033	-0.0049	0.0034	-0.0039	0.0017	-0.0025	0.0017
σ_v	-0.0354	0.0158	-0.0399	0.0164	-0.0093	0.0016	-0.0041	0.0016
$\hat{y}$	-0.0023	0.0185	-0.0018	0.0370	-0.0021	0.0280	-0.0029	0.0204
E(u|ε)	-0.0681	0.2335	-0.0676	0.2401	-0.0772	0.2310	-0.0772	0.2285
$r_{u\hat{u}}$	0.9086		0.9058		0.9079		0.9090	
$s_{u\hat{u}}$	0.7781		0.7736		0.7765		0.7782	
$P(m)$			0.3794		0.4005			

(b) $n = 500$

$\bar{\lambda} = 1$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0016	0.0025	-0.0024	0.0025	-0.0013	0.0015	0.0003	0.0014
σ_v	-0.0097	0.0034	-0.0113	0.0034	-0.0060	0.0016	-0.0029	0.0016
$\hat{y}$	-0.0008	0.0186	-0.0004	0.0364	-0.0004	0.0333	-0.0014	0.0230
E(u|ε)	-0.1458	0.4895	-0.1463	0.4920	-0.1467	0.4897	-0.1459	0.4882
$r_{u\hat{u}}$	0.8022		0.8012		0.8013		0.8018	
$s_{u\hat{u}}$	0.6260		0.6250		0.6250		0.6254	
$P(m)$			0.3769		0.3887			

$\bar{\lambda} = 2$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0016	0.0020	-0.0020	0.0020	-0.0007	0.0010	0.0000	0.0010
σ_v	-0.0171	0.0070	-0.0151	0.0068	0.0008	0.0008	0.0030	0.0008
$\hat{y}$	-0.0008	0.0127	-0.0005	0.0242	-0.0005	0.0194	-0.0010	0.0158
E(u|ε)	-0.0746	0.2283	-0.0751	0.2324	-0.0799	0.2284	-0.0797	0.2273
$r_{u\hat{u}}$	0.9112		0.9095		0.9103		0.9108	
$s_{u\hat{u}}$	0.7811		0.7781		0.7793		0.7801	
$P(m)$			0.4093		0.4394			

(c) $n = 1000$

Table 4. Simulation results for INLS, NLS, ML and ML-MONO when $\phi = 0.9$ (DGP_2).

$\bar{\lambda} = 1$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	0.0082	0.0267	-0.0626	0.0674	-0.0591	0.0317	-0.0044	0.0175
σ_v	-0.1276	0.0619	-0.1814	0.0908	-0.0940	0.0252	-0.0509	0.0180
$\hat{y}$	-0.0243	0.1123	-0.0098	0.3430	0.0065	0.2809	-0.0160	0.1154
E(u|ε)	-0.1063	0.5381	-0.1169	0.6517	-0.1504	0.5657	-0.1300	0.5084
$r_{u\hat{u}}$	0.7784		0.7233		0.7547		0.7794	
$s_{u\hat{u}}$	0.6157		0.5725		0.5962		0.6145	
$P(m)$			0.2386		0.2381			

$\bar{\lambda} = 2$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0596	0.0182	-0.1276	0.0466	-0.0403	0.0170	-0.0083	0.0118
σ_v	-0.0308	0.0368	-0.0508	0.0421	-0.0681	0.0137	-0.0323	0.0091
$\hat{y}$	0.0141	0.0524	0.0348	0.1886	0.0012	0.1192	-0.0121	0.0564
E(u|ε)	-0.0822	0.2717	-0.1021	0.3649	-0.0760	0.2792	-0.0674	0.2455
$r_{u\hat{u}}$	0.8914		0.8577		0.8778		0.8927	
$s_{u\hat{u}}$	0.7670		0.7263		0.7492		0.7667	
$P(m)$			0.2237		0.2794			

(a) $n = 100$

$\bar{\lambda} = 1$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0008	0.0073	-0.0174	0.0126	-0.0127	0.0045	0.0000	0.0035
σ_v	-0.0333	0.0119	-0.0371	0.0142	-0.0159	0.0031	-0.0103	0.0029
$\hat{y}$	-0.0078	0.0342	-0.0018	0.0797	0.0005	0.0608	-0.0064	0.0358
E(u|ε)	-0.1359	0.5003	-0.1427	0.5234	-0.1482	0.5009	-0.1414	0.4919
$r_{u\hat{u}}$	0.7981		0.7950		0.7963		0.7976	
$s_{u\hat{u}}$	0.6247		0.6226		0.6232		0.6235	
$P(m)$			0.3169		0.3322			

$\bar{\lambda} = 2$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0134	0.0052	-0.0252	0.0084	-0.0089	0.0023	-0.0028	0.0020
σ_v	-0.0314	0.0184	-0.0320	0.0234	-0.0078	0.0016	-0.0027	0.0016
$\hat{y}$	0.0001	0.0217	0.0049	0.0479	-0.0011	0.0304	-0.0045	0.0214
E(u|ε)	-0.0701	0.2407	-0.0746	0.2598	-0.0791	0.2338	-0.0762	0.2297
$r_{u\hat{u}}$	0.9074		0.9034		0.9065		0.9078	
$s_{u\hat{u}}$	0.7787		0.7731		0.7760		0.7775	
$P(m)$			0.3355		0.3899			

(b) $n = 500$

$\bar{\lambda} = 1$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	0.0035	0.0049	-0.0039	0.0071	-0.0035	0.0023	0.0034	0.0020
σ_v	-0.0193	0.0061	-0.0194	0.0067	-0.0057	0.0017	-0.0032	0.0017
$\hat{y}$	-0.0053	0.0243	-0.0025	0.0477	0.0002	0.0363	-0.0036	0.0243
E(u|ε)	-0.1395	0.4955	-0.1430	0.5066	-0.1474	0.4939	-0.1436	0.4893
$r_{u\hat{u}}$	0.8025		0.8011		0.8016		0.8022	
$s_{u\hat{u}}$	0.6262		0.6252		0.6252		0.6253	
$P(m)$			0.3484		0.3728			

$\bar{\lambda} = 2$

	INLS		NLS		ML		ML-MONO	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
γ	-0.0019	0.0039	-0.0066	0.0053	-0.0027	0.0014	0.0004	0.0013
σ_v	-0.0279	0.0137	-0.0263	0.0164	0.0010	0.0008	0.0033	0.0008
$\hat{y}$	-0.0020	0.0164	-0.0004	0.0318	-0.0000	0.0207	-0.0019	0.0164
E(u|ε)	-0.0694	0.2355	-0.0710	0.2461	-0.0804	0.2298	-0.0788	0.2279
$r_{u\hat{u}}$	0.9109		0.9086		0.9103		0.9108	
$s_{u\hat{u}}$	0.7813		0.7778		0.7790		0.7797	
$P(m)$			0.3704		0.4289			

(c) $n = 1000$

the Member States of the European Union and can be used 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.[12]
More specifically, we consider a dataset for the *Fieldcrops* production sector relative to 109 main FADN European regions already analyzed by Ferrara and Vidoli (2017): the output variable is represented by the total amount of crops and crop products (SE135 code) while Labour (SE010 code), Machinery (SE455 code) and Energy (SE345 code) are the inputs. Since efficiency can be strongly affected by a wide range of climatic factors, the average macro-regional temperature and average high temperature are considered as contextual variables to control for observable territorial heterogeneity (Battese and Coelli, 1995).[13]

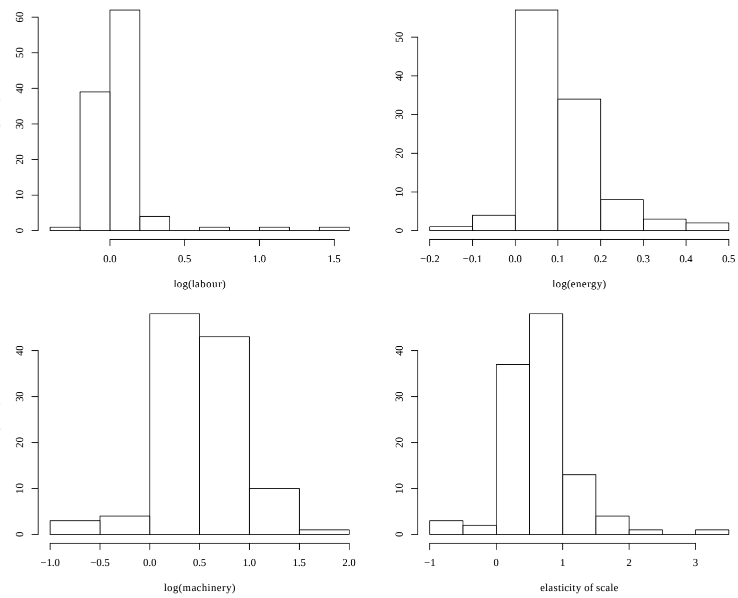

Figure 1. Output elasticities estimated under kernel specification (FADN data).

The output elasticities with respect to each input (fixing all other input variables at their median values), and the relative elasticity of scale, are reported in Figure 1 for NPK, in Figure 2 for PSM and in Figure 3 for PTM, respectively. We find that the output elesticities obtained via the NPK specification do often violate the monotonicity condition ($\approx 45\%$ of the units) highlighting a potential model misspecification.

Overall we find evidence of variability among units in terms of output elasticities, especially for machinery. This implies the need of a flexible model (even parametric) to handle heterogeneous technologies. As expected, Figure 2 and 3 show that both the P-splines and translog specifications, under the monotonicity constraints, do not violate regularity conditions.

In Figure 4 we show the estimated partial relationship between each input and the output when the production technology is estimated by GAMs with monotonic P-splines. It is worth emphasizing that this graphical tool can be used to investigate

[12]FADN elementary data aggregated by year, region and typology are available at: http://ec.europa.eu/agriculture/rica/.
[13]For more information about the FADN data and the several specific issues relative to the measurement of inputs and outputs please, see Ferrara and Vidoli (2017).

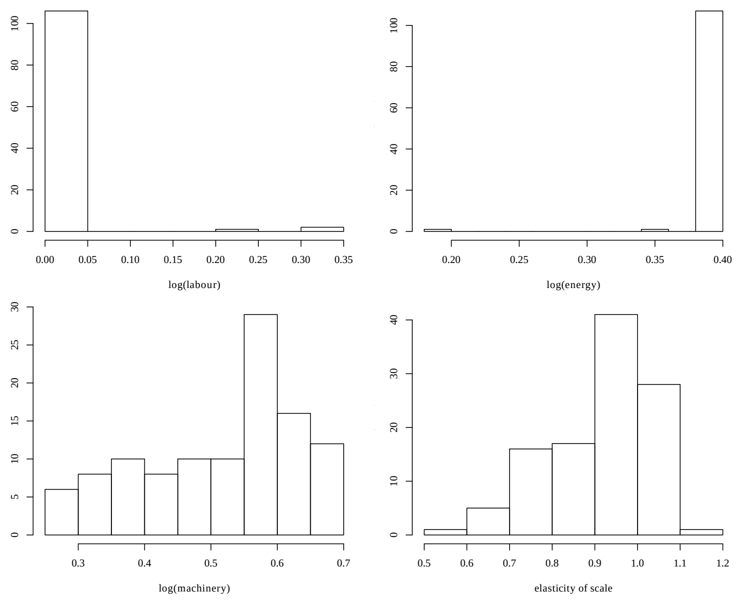

Figure 2. Output elasticities estimated using GAMs with monotone P-splines (FADN data).

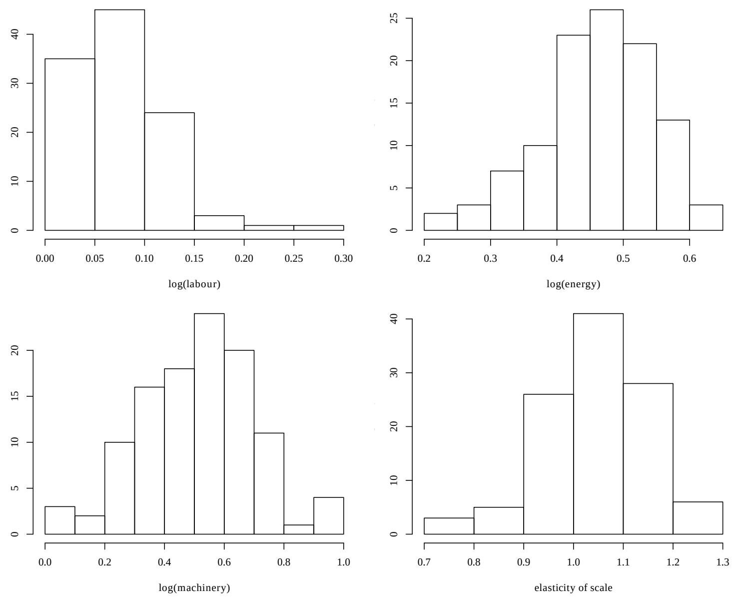

Figure 3. Output elasticities estimated using Translog with monotonicity constraint (FADN data).

the underline relationship between the output and each input in the model, i.e. the form of production technology. Indeed, we find that a nonlinear relationship is required with respect to labour and machinery while a simpler linear specification seems to be appropriate and noteworthy for energy.

Finally, we report in Table 5 estimated coefficients related to the average macro-regional temperature (AT) and the average high temperature (AHT) for the three different models. All coefficients are statistically significant for all model specifications but the sign and magnitude of the estimated effects is different. Indeed, the coeffi-

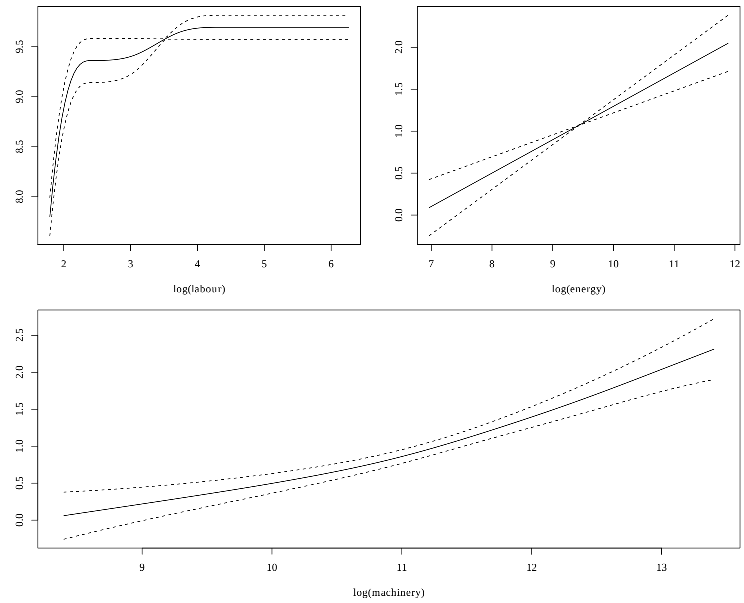

Figure 4. Estimated production technology by GAMs with monotonic P-splines (FADN data).

Table 5. Contextual variables estimates (FADN data).

	PTM		NPK		PSM	
	Est	SE	Est	SE	Est	SE
AT	-152.116*	65.488	1.044***	0.227	1.085***	0.314
AHT	-213.809*	92.255	-0.855***	0.188	-0.900***	0.260

Note: Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1.

cient associated to AT for the parametric translog monotonic specification is negative in contrast with previous studies (e.g., Ferrara and Vidoli, 2017) suggesting a positive correlation between efficiency estimates and temperature. Moreover, the translog specification point towards increasing returns to scale in contrast with NPK and PSM.

The end result of this exercise is that, even thought the flexibility of the translog functional form is comparable to nonparametric/semiparametric smoothers, inference about both inefficiency determinants and scale elasticity can be (very) different, potentially leading to misleading conclusions. This evidence emphasizes the usefulness of the INLS algorithm for evaluating the appropriateness of the functional form specification. Finally, Figure 5 shows the relationship between firm size and the elasticity of scale based on the PSM specification. Smaller firms, in general, would gain from increasing their size. Of special note is that this plot is also revealing possible leverage points, as indicated by the two units which correspond to the lowest output level but share the highest value of elasticity of scale.

5. Concluding remarks

We have introduced an iterative algorithm based on nonlinear least squares allowing for a great deal of flexibility in the specification of a production frontier satisfying

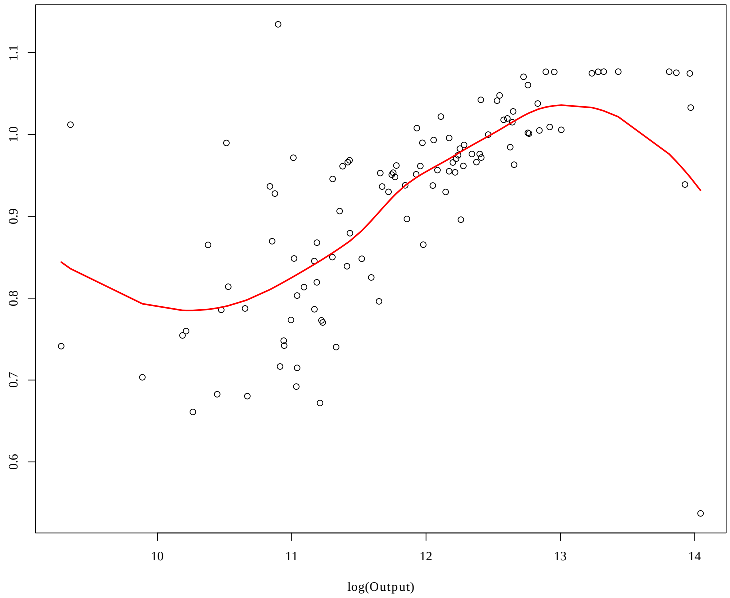

Figure 5. Firm size vs elasticities of scale estimated by GAMs with monotonicity constraints (FADN data).

monotonicity and incorporating exogenous determinants of inefficiency. Although there are alternative approaches in the literature, they are rarely applied mainly due to their computational complexity; the proposed framework can be applied using any statistical software that allows the estimation of semi- or nonparametric regression and standard nonlinear least squares.

We investigate the behavior of the proposed procedure through a set of Monte Carlo experiments comparing its finite sample properties with those of available alternatives. To our knowledge this is the first study providing evidence on the finite sample performances of one-step NLS and the Henningsen and Henning (2009) three-step procedure. The new algorithm provides very good performance, outperforming the competitors in small samples and in presence of small signal-to-noise ratios. An application on agricultural data illustrate the usefulness of the proposed algorithm, even when it is used as a "goodness-of-specification" tool.

Most empirical studies adopt a flexible - even parametric - specification of the production technology but do not enforce *ex ante* the monotonicity condition or check *ex post* if the estimated frontier fulfils this requirement. This is crucial since a non-monotonic frontier may produce biased inefficiency effects and upset the final ranking of productive units, thus inhibiting a reasonable interpretation of the unit-specific efficiency scores. Since the true frontier is typically unknown, a semi- or nonparametric specification may be useful to test for the appropriateness of the assumed frontier specification. Moreover, with the semiparametric approach, unit-specific measures of the output elasticises can be easily obtained. With regard to potential extensions of our algorithm, one important area for further work concerns panel data analysis.

6. Disclosure statement

No potential conflict of interest was reported by the authors.

References

Aigner, D., Lovell, C.A.K., & Schmidt, P. (1997). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6*, 21-37.

Alvarez, A., Amsler, C., Orea. L., & Schmidt, P. (2006) Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis, 25*, 20-212.

Bollaerts, K., Eilers, P.H.C., & van Mechelen, I. (2006). Simple and multiple P-splines regression with shape constraints. *British Journal of Mathematical and Statistical Psychology, 59*, 451-469.

Battese, G.E., & Coelli, T. (1995). A Model for technical inefficiency effects in a stochastic frontier Production function for panel Data. *Empirical Economics, 20*, 325-332.

Carroll, R.J., Fan, J., Gijebels, I., & Wand, M.P. (1997). Generalized partially linear single-index models. *Journal of American Statistical Association, 92*, 477-489.

Eilers, P.H.C., & Marx, B.D. (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science, 11*, 89-121.

Fan, Y., Li, Q., & Weersink, A. (1996). Semiparametric Estimation of Stochastic Production Frontier Models. *Journal of Business and Economic Statistics, 14*, 460-468.

Ferrara, G. (2020). Stochastic frontier models using R. In: *Handbook of Statistics - Financial, Macro and Micro Econometrics Using R*, Vinod and Rao, eds., Elsevier.

Ferrara, G., & Vidoli, F. (2017). Semiparametric stochastic frontier models: A generalized additive model approach. *European Journal of Operational Research, 258*, 761-777.

Giannakas, K., Tran, K.C., & Tzouvelekas, V. (2003). On the Choice of Functional Form in Stochastic Frontier Modeling. *Empirical Economics, 28*, 75-100.

Hall, P., & Huang, L. (2001). Nonparametric Kernel Regression Subject to Monotonicity Constraints. *Annals of Statistics, 29*, 624-647.

Hayfield, T., & Racine, J.S. (2008). Nonparametric Econometrics: The `np` Package. *Journal of Statistical Software, 27*, 1-32.

Hardle, W., Liang, H., & Gao, J. (2000). *Partially Linear Models*. Physica-Verlag.

Hastie, T.J., & Tibshirani, R.J. (1990). *Generalized additive models*. London: Chapman & Hall.

Henderson, D., & Parmeter, C.F. (2015). *Applied Nonparametric Econometrics*. Cambridge: Cambridge University Press.

Henningsen, H., & Henning, C. (2009). Imposing Regional Monotonicity on Translog Stochastic Production Frontiers with a Simple Three-Step Procedure. *Journal of Productivity Analysis, 32*, 217-229.

Henningsen, A., & Kumbhakar, S.C. (2009). Semiparametric stochastic frontier analysis: an application to polish farms during transition. European Workshop on Efficiency and Productivity Analysis. EWEPA, Pisa, Italy.

Huang, C.J., & Liu, J-T. (1994). Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis, 5*, 171-180.

Jondrow, J., Lovell, C.A.K., Materov, I.S., & Schmidt, P. (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Models. *Journal of Econometrics, 119*, 233-238.

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

Kumbhakar, S.C., Ghosh, S., & McGuckin, J.T. (1991). A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in United-States

Dairy Farms. *Journal of Business and Economic Statistics, 9*, 279-286.

Kumbhakar, S.C., Park, B., Simar, L., & Tsionas, E. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics, 137*, 1-27.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from cobb−douglas production functions with composed error. *International Economic Review, 18*, 435-444.

Muggeo, V., & Ferrara, G. (2008). Fitting generalized linear models with unspecified link function: a p-spline approach. *Computational Statistics and Data Analysis, 52*, 2529-2537.

O'Donnell, C.J., & Coelli, T.J. (2005). A Bayesian approach to imposing curvature on distance functions. *Journal of Econometrics, 126*, 493-523

Parmeter, C.F., & Kumbhakar, S.C. (2014). Efficiency Analysis: A Primer on Recent Advances. *Foundations and Trends(R) in Econometrics, 7*, 191-385.

Parmeter, C.F., & Wang, H., & Kumbhakar, S.C. (2017). Nonparametric estimation of the determinants of inefficiency. *Journal of Productivity Analysis, 47*, 205-221.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2020.

Racine, J.S., & Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics, 119*, 99-130.

Reifschneider, D., & Stevenson, R. (1991). Systematic departures from the frontier A framework for the analysis of firm inefficiency. *International Economic Review, 32*, 715-723.

Sauer, J., Frohberg, K., & Hockmann, H. (2019). Stochastic efficiency measurement: the curse of theoretical consistency. *Journal of Applied Economics, 9*, 139-165.

Terrell, D. (1996). Incorporating Monotonicity and Concavity Conditions in Flexible Functional Forms. *Journal of Applied Econometrics, 11*, 179-94.

Tran, K.C., & Tsionas, E.G. (2009). Estimation of nonparametric inefficiency effects stochastic frontier models with an application to British manufacturing. *Economic Modelling, 26*, 904-909.

Wang, H.J., & Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis, 18*, 129-144.

Zhou, J., Parmeter, C.F., & Kumbhakar, S.C. (2020). Nonparametric estimation of the determinants of inefficiency in the presence of firm heterogeneity. *European Journal of Operational Research, 286*, 1142-1152.

Imposing monotonicity in stochastic frontier models: an iterative nonlinear least squares procedure

Federico Belotti[a] and Giancarlo Ferrara[b,c]

[a]Department of Economics and Finance, University of Rome Tor Vergata, Italy; [b]SOSE - Soluzioni per il Sistema Economico SpA, Rome, Italy; [c] Department of Economics, Business and Statistics, University of Palermo, Italy

ARTICLE HISTORY
Compiled December 29, 2020

ABSTRACT
Despite its importance, the monotonicity condition is typically overlooked in stochastic frontier analysis. This article illustrates a straightforward and useful method for the estimation of semiparametric stochastic frontier models imposing such constraint and incorporating exogenous inefficiency effects exploiting the scaling property. An iterative estimation algorithm based on nonlinear least squares is developed and the behavior of the proposed procedure is investigated through a set of Monte Carlo experiments comparing its finite sample properties with those of available alternatives. The simulation results highlight very good performance of the new algorithm which outperforms the competitors in small samples and in presence of small signal-to-noise ratios. Our results also show that the fraction of observations for which monotonicity naturally holds is generally quite small if this condition is not imposed. An application based on FADN agricultural data illustrates the usefulness of the proposed algorithm.

KEYWORDS
Efficiency, stochastic frontiers, monotonicity, nonlinear least squares, heteroskedasticity, simulation

1. Introduction

The Stochastic Frontier (SF) model, introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977), represents a very popular tool for efficiency analysis.[1] In its basic formulation, it is based on a parametric representation of production technology along with a compounded error term in which one component, often assumed to be normally distributed, deals with the randomness of production while the other, following a specific one-sided distribution, captures the inefficiency of productive units. Several methodological improvements of this basic formulation have been proposed in the last three decades, most of which deal with relaxing the restrictive assumptions

CONTACT G. Ferrara. Email: federico.belotti@uniroma2.it, giancarlo.ferrara@gmail.com

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of SOSE SpA

[1]Production frontiers represent the maximum amount of output that can be obtained from a given level of inputs, while a cost frontier characterizes the minimum expenditure required to produce a bundle of outputs given the prices of the inputs used in its production.

RECENT PUBLICATIONS BY *CEIS Tor Vergata*

Ideology in times of crisis.
A principal component analysis of votes in the European Parliament, 2004-2019
Anatole Cheysson and Nicolò Fraccaroli
CEIS Research Paper, 461, June 2019

Competing Mechanisms and Folk Theorems: Two Examples
Andrea Attar, Eloisa Campioni, Thomas Mariotti and Gwenael Piaser
CEIS Research Paper, 460, June 2019

Estimation of stochastic frontier panel data models with spatial inefficiency
Federico Belotti, Giuseppe Ilardi and Andrea Piano Mortari
CEIS Research Paper, 459, May 2019

Hurricanes, Climate Change Policies and Electoral Accountability
Stefano Gagliarducci, M. Daniele Paserman and Eleonora Patacchini
CEIS Research Paper, 458, May 2019

Demography and Productivity in the Italian Manufacturing Industry: Yesterday and Today
Carlo Ciccarelli, Matteo Gomellini and Paolo Sestito
CEIS Research Paper, 457, May 2019

Real Effective Exchange Rates Determinants and Growth: Lessons from Italian Regions
Silvia Calò and Mariarosaria Comunale
CEIS Research Paper, 456, March 2019

Predictability, Real Time Estimation, and the Formulation of Unobserved Components Models
Tommaso Proietti
CEIS Research Paper, 455, March 2019

Sovereign Spread Volatility and Banking Sector
Vivek Sharma and Edgar Silgado-Gómez
CEIS Research Paper, 454, March 2019

The Societal Benefits of Money and Interest Bearing Debt
Luis Araujo and Leo Ferraris
CEIS Research Paper, 453, February 2019

A Horse Race in High Dimensional Space
Paolo Andreini and Donato Ceci
CEIS Research Paper, 452, February 2019

DISTRIBUTION

Our publications are available online at www.ceistorvergata.it

DISCLAIMER

COPYRIGHT

MEDIA INQUIRIES AND INFORMATION

For media inquiries, please contact Barbara Piazzi at +39 06 72595652/01 or by e-mail at piazzi@ceis.uniroma2.it. Our web site, www.ceistorvergata.it, contains more information about Center's events, publications, and staff.

DEVELOPMENT AND SUPPORT

For information about contributing to CEIS Tor Vergata, please contact at +39 06 72595601 or by e-mail at segr.ceis@economia.uniroma2.it