

Efficiency analysis of Finnish crop farms by Stochastic Nonparametric Envelopment of Data (StoNED)

Timo Kuosmanen^{1,2} and Natalia Kuosmanen¹

1) *Economic Research Unit, MTT Agrifood Research Finland, Luutnantintie 13, 00410 Helsinki.*

2) *School of Economics, Aalto University, P.O. Box 1210, 00101 Helsinki, Finland.*

Abstract

A new two-stage approach to modeling the nutrient emissions of crop farms is developed. In the first stage, we estimate the classic production function using farm-level data. We apply the stochastic nonparametric envelopment of data (StoNED) method that combines the axiomatic, nonparametric modeling of the production technology with a stochastic, probabilistic treatment of inefficiency and noise. Given the estimated input saving and output expansion potential for each farm in the sample, we estimate the impacts of the decreased fertilizer use and the increased crop yield on the nutrient balance of farms. This provides an aggregate level estimate on how much nutrient emissions could be abated through improving the operational efficiency of farms without any new technical innovations or down-scaling of the activity.

Keywords

environmental performance, frontier estimation, productive efficiency analysis

Introduction

Productivity and efficiency analysis is frequently applied in the field of agricultural economics to estimate the total factor productivity growth at the sector and farm levels, to decompose the productivity changes to components of technological progress, operational efficiency and scale efficiency, to identify efficiency improvement targets and best-practice benchmarks, and to estimate shadow prices for undesirable outputs such as nutrient leaching to water systems (see e.g., Coelli 1995; Oude Lansink et al. 2002; Gorton and Davidova 2004). All different applications of productive efficiency analysis are based on the estimation of the production frontiers in one way or another.

The two main approaches to estimating the best-practice frontiers are the nonparametric *data envelopment analysis* (DEA: Farrell, 1957; Charnes et al., 1978) and the parametric *stochastic frontier analysis* (SFA: Aigner et al., 1977; Meeusen and van den Broeck, 1977). The main appeal of DEA lies in its nonparametric treatment of the frontier, which does not assume a particular functional form but relies on the general regularity properties such as monotonicity, convexity, and homogeneity. However, DEA attributes all deviations from the frontier to inefficiency, and completely ignores any stochastic noise in the data. The key advantage of SFA is its stochastic treatment of residuals, decomposed into a non-negative inefficiency term and a disturbance term that accounts for measurement errors and other random noise. However, SFA builds on the parametric regression techniques, which require an *ex ante* specification of the functional form.

Bridging the gap between SFA and DEA was recognized as an important research question already in the early 1990s. The emerging literature on semi/nonparametric stochastic frontier estimation has thus far mainly departed from the SFA side, replacing the parametric frontier function by a nonparametric specification that can be estimated by kernel regression or local maximum likelihood (ML) techniques. Fan et al. (1996) and Kneip and Simar (1996) were among the first to apply kernel regression to frontier estimation in the cross-sectional and panel data contexts, respectively. Fan et al. (1996) proposed a two-step method where the shape of the frontier is first estimated by kernel regression, and the conditional expected inefficiency is subsequently estimated based on the residuals, imposing the same distributional assumptions as in standard SFA. Kneip and Simar (1996) similarly use kernel regression for estimating the frontier, but they make use of panel data to avoid the distributional assumptions.

Banker and Maindiratta (1992) were the first to consider ML estimation of the stochastic frontier model subject to nonparametric shape constraints regarding the frontier. Their monotonicity and concavity constraints are analogous to DEA, but solving the resulting complex ML problem has proved extremely difficult, if not impossible in practical applications. No reported applications of the Banker and Maindiratta's constrained ML method are known in the literature.

In a series of recent papers, Kuosmanen (2006), Kuosmanen and Kortelainen (2007), Kuosmanen (2008), Kuosmanen and Johnson (2009), and Johnson and Kuosmanen (2009) develop a new encompassing framework referred to as *stochastic nonparametric envelopment of data* (StoNED). Specifically, they introduce a stochastic SFA-style composite error term to a nonparametric, DEA-style piecewise linear frontier. Similar to DEA, the StoNED approach does not assume any particular functional form for the frontier; it identifies the function that best fits the data from the family of continuous, monotonic increasing, concave functions that can be non-differentiable. Kuosmanen and Kortelainen (2007) noted that both SFA and DEA can be constrained special cases of the more general StoNED model.

An earlier study by the authors (Kuosmanen and Kuosmanen 2009) presented the first application of the StoNED method to the farm-level data from agricultural production, focusing on the environmental performance of a sample of Finnish dairy farms. The present paper extends the scope of the previous study in two important respects. First, we shift attention on crop farms specialized in wheat cultivation. Second, we elaborate the modeling of nutrient leaching. Some problems with the conventional approach of modeling nutrient surplus as an input or undesirable output are discussed. Further, an alternative two-stage method that avoids these problems is developed.

The rest of the paper is organized as follows. In the next section we introduce the theoretical model of production involving undesirable outputs. We then present the StoNED method that can be applied for estimating the theoretical model. We then describe our empirical data and present the estimation results. A brief summary and discussion of future research topics concludes the paper.

StoNED model

The vector of inputs (including labor, capital, energy, and intermediate inputs) is denoted by \mathbf{x} and the output is denoted by y . The production technology is represented by the *production function* $y = f(\mathbf{x})$. We assume that function f belongs to the class of continuous, monotonic increasing and globally concave functions. In contrast to the SFA literature, no specific functional form for f is assumed a priori; our specification of the production function proceeds along the nonparametric lines of the DEA literature.

The observed output y_i of firm i may differ from $f(\mathbf{x}_i)$ due to inefficiency and noise. We follow the SFA literature and introduce a composite error term $\varepsilon_i = v_i - u_i$, which consists of the inefficiency term $u_i > 0$ and the idiosyncratic error term v_i , formally,

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

Terms u_i and v_i ($i = 1, \dots, n$) are assumed to be statistically independent of each other as well as of inputs \mathbf{x}_i . Furthermore, we follow the standard SFA practice and assume $u_i \sim \left. N(0, \sigma_u^2) \right|$ and $v_i \sim N(0, \sigma_v^2)$.

In model (1), the deterministic part (i.e., production function f) is defined analogous to DEA, while the stochastic part (i.e., composite error term ε_i) is defined similar to SFA. As a result, model (1) encompasses the classic SFA and DEA models as its constrained special cases. Specifically, if f is restricted to some specific functional form (instead of the class F_2), model (1) boils down to the SFA model by Aigner et al. (1977). On the other hand, if we impose the restriction $\sigma_v^2 = 0$ and relax the distributional assumption concerning the inefficiency term, we obtain the single-output DEA model with an additive output-inefficiency, first considered by Afriat (1972). In this sense, both SFA and DEA can be seen as constrained special cases of model (1).

It is easy to write a theoretical model like (1); the main challenge is its estimation. In this paper we propose a new strategy to estimating model (1), referred to as *stochastic nonparametric envelopment of data* (StoNED). Our objective is to estimate the deterministic part of the model in a fully nonparametric fashion imposing a minimal set of assumptions, in the spirit of DEA. Similar to DEA, we estimate the shape of the frontier by exploiting the regularity properties from the microeconomic theory (i.e., continuity, monotonicity, and concavity of f), free of any distributional assumptions or assumptions about the functional form of f or its smoothness. However, in the cross-sectional setting it is impossible to distinguish between inefficiency and noise without imposing some distributional assumptions. Having estimated the shape of function f , we make use of the standard distributional assumptions adopted from the SFA literature to estimate the expected location of the frontier f , and the firm-specific conditional expected values for the inefficiency term. In summary, the StoNED method consists of two-steps:

Step 1: Estimate the shape of function f by Convex Nonparametric Least Squares (CNLS) regression

$$\begin{aligned} & \min_{\mathbf{a}, \mathbf{\beta}, \varepsilon} \sum_{i=1}^n \varepsilon_i^2 \\ & y_i = \alpha_i + \mathbf{\beta}'_i \mathbf{x}_i + \varepsilon_i \\ & \alpha_i + \mathbf{\beta}'_i \mathbf{x}_i \leq \alpha_h + \mathbf{\beta}'_h \mathbf{x}_i \quad \forall h, i = 1, \dots, n \\ & \mathbf{\beta}_i \geq \mathbf{0} \quad \forall i = 1, \dots, n \end{aligned}$$

Step 2: Using residuals ε from the CNLS regression, estimate the variance parameters σ_u^2, σ_v^2 by using the method of moments or pseudolikelihood techniques, and compute the conditional expected values of inefficiency. Subsequently, the expected value μ and the parameters of the inefficiency and noise distributions can be estimated based on the regression residuals by the method of moments or pseudolikelihood techniques (see Kuosmanen and Kortelainen 2007 for details).

Modeling of nutrient emissions

In the literature of productive efficiency analysis, the conventional approach to modeling nutrient emissions is to treat the nitrogen surplus as an input variable or as an undesirable output (e.g., Reinhard et al. 2002). However, the conventional approach is problematic both in the DEA and SFA-type modeling. Firstly, DEA requires that the inputs and outputs are ratio-scale variables. However, the nitrogen surplus is only an interval-scale variable: it is measured as a difference of the inflow and outflow of nitrogen. A related issue is that the nitrogen surplus may be less than or equal to zero. Secondly, from the perspective of the traditional production model assumed in the SFA literature, the nitrogen surplus is not a factor of production: it is possible to produce output even if the nitrogen surplus is negative or zero. Further, increase in the nitrogen surplus does not necessarily have an increasing effect on the output. Due to these problems, in this study we consider an alternative approach to modeling the nutrient problem.

In crop farms, high nitrogen surplus is generally due to the excessive use of fertilizers. Fertilizer is a ratio-scale input variable whereas the nitrogen surplus is not. Therefore, we resort to a two-stage approach where we first estimate the classic production function that includes the fertilizer use as one of the inputs. In the second stage, we estimate the potential impact of efficiency improvement on the nitrogen balance of the farm. Note that efficiency improvement can influence the nutrient surplus in two ways. Firstly, decreasing the input use, including the fertilizer input, will decrease the inflow of nitrogen. Secondly, increasing the output (crop yield) will increase the outflow of nutrient. Both have a favorable effect on the nutrient balance. In the application to be considered next, we examine both the input decrease and output expansion scenarios.

Data

We next apply the StoNED method described in the previous section to the empirical data of Finnish crop farms specialized in wheat. The data set is extracted from the Farm Accountancy Data Network (FADN) database. The following variables were considered: output variable is yield of wheat (100kg/ha), input variables are labor (hr/ha), farm capital (€/ha), total energy cost (€/ha) and fertilizers (kgN/ha), and the environmental variable of interest is the nitrogen surplus (kg/ha). An overview of the key characteristics of the data is presented in Table 1 in the form of the descriptive statistics: mean, standard deviation, minimum and maximum values.

Table 1: Descriptive statistics for the sample of wheat farms; year 2004, sample size n=141

Variable	Mean	St. Dev.	Min	Max
Yield of wheat, 100kg/ha	36.6	11.3	7.2	71.6
Labor, hr/ha	29.0	17.4	5.8	117.8
Farm capital, €/ha	2853.5	1302.1	977.1	7918.9
Energy, €/ha	89.9	41.3	0.0	279.9
Fertilizer, kgN/ha	135.8	62.6	0.0	385.3
Nitrogen surplus kgN/ha	26.9	12.3	-0.1	77.1

Main results

We first applied the StoNED method to estimate efficiency in input use at the farm level. Recall that the StoNED method allows us to distinguish whether deviations from the frontier that are due to inefficiency or random noise. The estimated standard deviation of inefficiency was 11.6, whereas the estimated standard deviation of noise was 7.82. Both standard deviations were found to be statistically significant at the one percent significance level.

The expected inefficiency loss was estimated as 9.25 €/ha. This is approximately 20% of the average wheat revenue per hectare. Thus, expressed in relative terms, the average efficiency at the sample level was approximately 80%.

Next, we estimated the potential impact on the nitrogen surplus that would be achievable through improvement of technical efficiency. Specifically, suppose that all farms in the sample could

catch up to the level of performance shown by the best farms in the sample. This would not require any new technical innovation, just that all farms could operate according to the current best practice. This is of course a very ambitious target as such, and unlikely to be met in practice. Nevertheless, it is interesting to examine what would be the impact of the maximum possible efficiency improvement on the nutrient balance. If the efficiency improvement does not suffice to bring the nutrient surplus to a sustainable level, the further decrease must come either through technical progress or down-scaling production. In this respect, utilizing the efficiency improvement potential to the maximum extent might be the easiest and the most cost-efficient way to decrease nutrient leaching.

Efficiency improvement could be implemented by producing the current level of output by less input, or using the current input to produce a larger amount of output. In the case of crop farms, both decrease of fertilizer input and the greater yield of output have a favorable effect on the nutrient balance. Thus, both input and output orientations were considered.

In the output orientation, the impact comes through larger uptake of nitrogen in the harvested crop keeping the land use and fertilizers use at the present level. This would result only 1.9 kg/ha decrease in nitrogen surplus. In the input orientation, we keep the present output level constant and decrease the input use including land and fertilizer, which would result in 5.4 kg/ha decrease in nitrogen surplus. While any kind of efficiency improvement has a desirable effect, the direction of input saving or output expansion matters as well. A more favorable effect is achieved through input-oriented efficiency improvement.

In conclusions, there is some scope for decreasing nitrogen surplus through efficiency improvement, but it is rather limited. Our results show that the input orientation allows for larger improvement potential. Even though it is not at all clear what would be the sustainable level of nitrogen surplus, at least, it is possible to estimate how much emissions could potentially be decreased if all farms operated as efficiently as their most productive peers.

Conclusions

We have explored a new approach to modeling and estimating the potential impact of efficiency improvement in reducing the nutrient emissions. The novel features of the study include the use of the StoNED method to estimating the frontier production function, and the subsequent estimation of the nutrient balance based on the decrease of fertilized input or increase of crop yield achievable through efficiency improvement. The attractive feature of the StoNED approach is that it enables us to combine the nonparametric, axiomatic approach to modeling the production technology with the stochastic modeling of inefficiency and noise. Further, drawing distinction between the productive input (fertilizer) and the associated environmental impact (nutrient surplus) avoids the theoretical and technical problems of the conventional modeling of nutrient surplus as an input or output variable.

Although we recognize that some environmental problems may depend on multiple inputs and/or outputs in a more complex way, we believe the two-stage approach considered in this paper has potential applications in other areas as well. For example, the green-house gas (GHG) emissions arising in the manufacturing sector are mainly due to one specific input: the energy use. Thus, a similar approach could be applied to estimating the potential reduction of GHG emissions through improving productive efficiency in the manufacturing industries.

References

- Afriat, S.** 1972. Efficiency estimation of production functions, *International Economic Review* 13, 568-598.
- Aigner, D.J. Lovell, C.A.K. and Schmidt, P.** 1977. Formulation and estimation of stochastic frontier models, *Journal of Econometrics* 6, 21-37.
- Banker, R.D. and Maindiratta, A.** 1992. Maximum likelihood estimation of monotone and concave production frontiers, *Journal of Productivity Analysis* 3, 401-415.
- Charnes, A. Cooper, W.W. and Rhodes, E.** 1978. Measuring the inefficiency of decision making units, *European Journal of Operational Research* 2(6), 429-444.

- Coelli, T.J.** 1995. Recent developments in frontier modelling and efficiency measurement, *Australian Journal of Agricultural Economics* 39(3), 219-245.
- Fan, Y., Li, Q. and Weersink, A.** 1996. Semiparametric estimation of stochastic production frontier models, *Journal of Business and Economic Statistics* 14(4), 460-468.
- Farrell, M.J.** 1957. The measurement of productive efficiency, *Journal of the Royal Statistical Society, Series A* 120, 253–281.
- Gorton, M., Davidova, A.** 2004. Farm productivity and efficiency in the CEE applicant countries: a synthesis of results, *Agricultural Economics* 30, 1-16.
- Johnson, A.L. and Kuosmanen T.** 2009. How operational conditions and practices affect productive performance? Efficient semi-parametric one-stage estimators, SSRN working paper (2 Nov 2009), available at <http://ssrn.com/abstract=1485733>.
- Kneip, A. and Simar, L.** 1996. A general framework for frontier estimation with panel data, *Journal of Productivity Analysis* 7, 187-212.
- Kuosmanen, T.** 2006. Stochastic nonparametric envelopment of data: Combining virtues of SFA and DEA in a unified framework, MTT Discussion Paper 3/2006.
- Kuosmanen, T.** 2008. Representation theorem for convex nonparametric least squares. *Econometrics Journal* 11: 308-325.
- Kuosmanen, T. and Johnson, A.L.** 2009. Data envelopment analysis as nonparametric least-squares regression. *Operations Research*, in press.
- Kuosmanen, T. and Kortelainen, M.** 2007. Stochastic nonparametric envelopment of data: Cross-sectional frontier estimation subject to shape constraints. University of Joensuu, Economics Discussion Paper #46, Joensuu, Finland.
- Kuosmanen, T. and Kuosmanen, N.** 2009. Role of benchmark technology in sustainable value analysis: an application to Finnish dairy farms. In Press.
- Meeusen, W., and van den Broeck, J.** 1977. Efficiency estimation from cobb-douglas production function with composed error, *International Economic Review* 8, 435-444.
- Oude Lansink, A., Pietola, K., and Bäckman, A.** 2002. Efficiency and productivity of conventional and organic farms in Finland 1994 – 1997, *European Review of Agriculture Economics*, 29(1), 51-65.
- Reinhard, S., Lovell, C.A.K., and Thijssen, G.** 2002. Analysis of environmental efficiency variation, *American Journal of Agricultural Economics* 84(4), 1054-1065.