Dieter Gstach

Data Envelopment Analysis in a Stochastic Setting: The right answer from the wrong model?

Paper

Data Envelopment Analysis in a Stochastic Setting: The right answer from the wrong model ?

Dieter GSTACH

Working Paper No. 29
August 1994

Comments and critique are welcome !

Authors adress:

Dieter Gstach
Department of Economics
Vienna University for Economics and Business Administration
Augasse 2-6, A-1090 VIENNA

phone: +43 1 31336-4965 fax: +43 1 31336-755,
e-mail: gstach@wu-wien.ac.at

Abstract: Data envelopment analysis (DEA) is compared to stochastic production function estimation (SPFE) in a noisy setting. The statistic of interest is the average efficiency estimator. Monte-Carlo simulations show that the mean squared error of the DEA-estimator even for considerable noise remains below the MSE of the SPFE analogue. A bootstrapping approach is designed to get some first-step statistical underpinning of this DEA average efficiency estimator. The coverage of the bootstrapping approximation to the distribution of this estimator is shown to be fairly good.

Introduction

Production processes are never under full control of the decision makers: They are noisy to some extent. Strictly speaking: Under such circumstances DEA is the wrong tool for efficiency estimation. In some sense, the well known techniques of stochastic production function estimation, would be more appropriate here. But the question for practical purposes is: How bad does DEA really perform with noisy production processes ? If it would turn out, that DEA from an overall perspective can in fact compete with SPFE, we were forced to ask a second question: How can statistical properties of DEA efficiency estimates be evaluated without relying on parametric theory ?

I will try to give answers to the first question by comparing the small sample performance of DEA and SPFE in terms of their ability to give accurate estimates of average efficiency. This is done by Monte-Carlo simulations and gives an impression of limits and scope of DEA as a valid alternative to SPFE in noisy environments. And as a first attempt to answer the second question I designed a bootstrapping procedure specifically to give statistical meaning to DEA by establishing empirical distribution functions for the relevant estimator. The statistical theory, on which I rely, is covered for example by Efron & Tibshirani's *Introduction to the Bootstrap* and by Härdle's *Applied Nonparametric Regression.*

Model and estimation procedures

All following simulations will be based upon a most simple constant-returns-to-scale production function $y = x$, with one input and one output. The stochastic model will consist of a normally distributed mean-zero noise term u and a half-normal efficiency term v, giving a data generating mechanism $y_i = x_i e^{u_i - v_i}$. In what follows I will refer to the sum of noise and efficiency term u-v as the error. To calculate the DEA frontier I use a variable-returns-to-scale, output-oriented, first-stage DEA model. To calculate the competing SPFE frontier I will rely on the correctly specified production function in log terms and the correctly parametrized distributions. This puts SPFE in very favorable position compared to DEA, which relies basically only on a convexity assumption.

Some explanations: The restriction to first-stage DEA comes from the univariate nature of the error term in SPFE and thus is a requirement for comparing the two approaches.[1] The focus

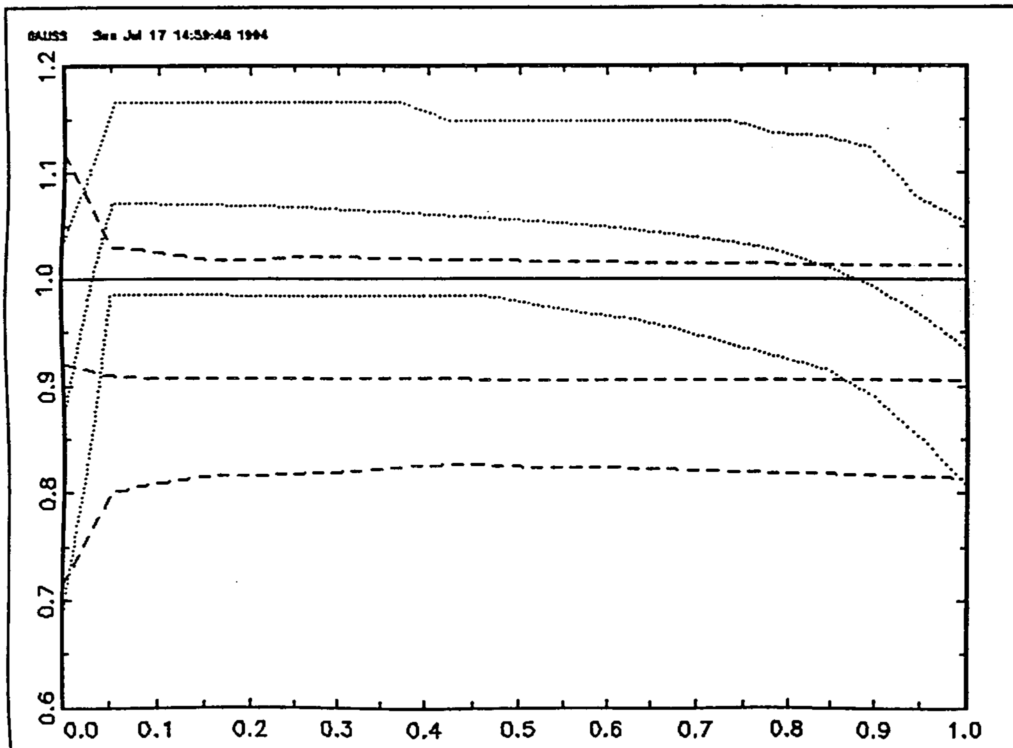
[1] Adding stochastics to second-stage DEA, where you optimize over slack variables and thus leave the proportional expansion framework, would of course require multivariate error terms.

on the one input-case has expositional purpose but also helps to justify, why the problem of programming slacks is neglected. A serious restriction is to have only one output. It seems as if canonical correlation analysis [2] could help at this point but as of yet, it is to early for serious claims.

So strictly speaking, the application of DEA in this setting is unjustified. But I want to show, that although being the wrong model it can be considered a viable alternative to SPFE to some extent.

DEA versus SPFE: Simulation results

Comparison of DEA and SPFE can't be based on closed form solutions for basic statistics, because they are unavailable. Instead one must invoke simulations to yield the sample performance of the relevant measures. Before going into the details and to illustrate the basic



Picture 1: Estimation of a frontier from 20 observations with 30% noise. DEA (dotted lines) vs. SPFE (dashed lines). 5%, 50% and 95% quantils are given. The model is $y_i = x_i e^{u_i - v_i}$ with $u \to N(0, \sigma_u = 0.1)$, $v \to |N(0, \sigma_v = 0.2)|$, both i.i.d and x equispaced in (0,1]. The picture is based on 1000 random samples of the above model. All figures are scaled by the respective true frontier values.

characteristics of my results take a look at picture 1.

It shows the average DEA estimated frontier with 5% and 95% quantiles (the dotted lines) and the average ML estimated frontier (the dashed lines), both scaled by the true frontier, which

[2] See the interesting approach of Friedman and Stern (1994).

therefore occurs as the solid horizontal line at one. 1000 random datasets from the above model were simulated to get these lines. Each of them consists of 20 observations with x equispaced in the interval [0.0001,1]. The variances of the noise and the efficiency term are chosen such, that noise accounts for 20% of the total variance and average efficiency is about 0.85. To get the SPF-estimates I used the iterative ML procedure implemented in a modul of the programming package GAUSS. Corrected ordinary least squares (COLS) parameter estimates served as starting values for the ML iterations.
As far as the location of the estimated frontiers is concerned, no method seems to be superior. But one can identify a characteristic deviation pattern of the DEA frontier from the true one: It underestimates the technological possibilities in the lowest and the highest regions of input space, while it overestimates them in the middle range. The reason for this behaviour is loss of comparability towards the ends of the scale. SPFE does not share this deficiency, because, after all it estimates an average tendency. As a consequence I will not deal with firm specific efficiency evaluations from DEA in noisy environments, but look only at average efficiency estimates.

Table 1 shows the small sample performance of DEA and SPFE , evaluated by the MSE and its two components variance and bias. Additionally it gives an idea of the sensitivity of these results to different stochastic specifications and sample sizes as shown.
Table 1 allows to draw some basic conclusions. The first one: Even with noise the DEA estimator does quite well on average, because it's variance is an order of magnitude lower than the variance of the SPF estimator. This of course comes at a cost, which is a significant bias as compared to the SPF estimator. The second one: Increasing the sample size deteriorates DEA's performance eventually, when noise is kept constant. This has to do with the inherent order statistic problem in DEA: The more observations you get, the higher the highest observation will be on average. And DEA is built upon these highest observations, no matter whether these result from high efficiency or from positive noise to the production process. Therefore in a stochastic environment it may be not be a good strategy to look for as many as possible observations to construct DEA frontiers. Of course more observations mean more information and this is valid for DEA too. But different techniques are required to exploit this information than simply construct convex envelopes for the data.
The third conclusion I draw from table 1 is, that reliance on asymptotic features of ML estimators like consistency is of little help in small samples like the ones used above. And these samples are not as small as it might seem at first sight: With 200 observations and 5 input variables the figures for comparable stochastic frameworks look quite similar. The failure of SPFE in terms of the MSE is due to high variance, which has two reasons: First in more than 30% of all cases the Waldman criterion, the third moment of the OLS-residuals, indicates misspecification, while in fact the model is perfectly specified. This failure leads to the conclusion of no inefficiency at all. The second reason is, that in more than 20% of all

	Obs	DEA estimator MSE	VAR	BIAS²	SPF estimator MSE	VAR	BIAS²	#{MLE fails} due to: $E[e^3]>0$	$\sigma_u^2<0$	no convergence
	20	**23**	19	3	**90**	81	9	222	334	148
25% noise	30	**37**	19	18	**83**	82	1	150	334	101
	40	**47**	16	31	**79**	78	1	105	413	68
	20	**35**	25	10	**104**	98	6	256	328	123
30% noise	30	**60**	22	38	**103**	102	1	202	312	79
	40	**84**	23	61	**105**	105	0	145	344	69
	20	**58**	31	27	**126**	118	8	314	254	119
40% noise	30	**107**	32	75	**129**	127	2	268	275	83
	40	**141**	30	111	**121**	121	0	207	259	57

Table 1: Statistic: $E\left[y_i/\hat{f}(x_i)\right]$ All figures times 0.0001. The variance of the efficiency term was kept constant at 0.023, which leads to a true average efficiency of about 0.85. When MLE failed the corrected ordinary least squares results were used as SPF-estimator. The figures for the 9 cases are based on 1000 repetitions each.

cases the estimated variance of the error term is negative, resulting in substituting the shifted COLS estimate for the ML estimate, with no noise at all. Together this accounts for the high MSE of the SPFE estimator.

So far DEA does not look bad but the limits of applicability seem clear too: With more than 30% noise, matters get tough. Because the main concern is efficiency estimation, the ultimate question is, whether the bias from DEA can be expected to be insignificant or not. DEA gives no answer to that question. But having a certain parametric alternative in mind, one could get approximate answers by doing simulations like the one above. But I will not push that idea any further here.

Instead I will proceed on the assumption, that there is not too much noise in the data. Table 1 indicates, that one can hope to get reasonable results from the DEA average efficiency estimator, in the sense of negligible bias compared to variance. The second question then is: How can confidence intervals for the average efficiency estimator be constructed ?

Statistical Interpretation of DEA results

According to bootstrapping theory one can simulate the distribution of an estimator by resampling techniques and thus derive statistics for an estimator where closed form solutions for these statistics are unavailable. DEA's average efficiency estimator is a good candidate in this respect. The naive bootstrapping procedure would be the following:

Estimate a DEA-frontier and construct the average efficiency estimator as usual as the mean of the ratios of observed outputs to frontier outputs. Then draw a random sample from the observations with replacement and construct a new DEA-frontier and a corresponding average efficiency estimator based on this new sample. Repeat this last step many times to get a vector of average efficiency estimators. Finally take the values corresponding to the desired quantiles from that vector.

Unfortunately this plug-in-principle does not work here, because it neglects some DEA specific features. The first is, that in fact two quite different things are estimated here at once: A frontier and the center of a distribution. If the frontier were known but nothing else, the above procedure would yield an unbiased confidence interval estimator with the smallest possible variance of any estimator based on as little information as this.

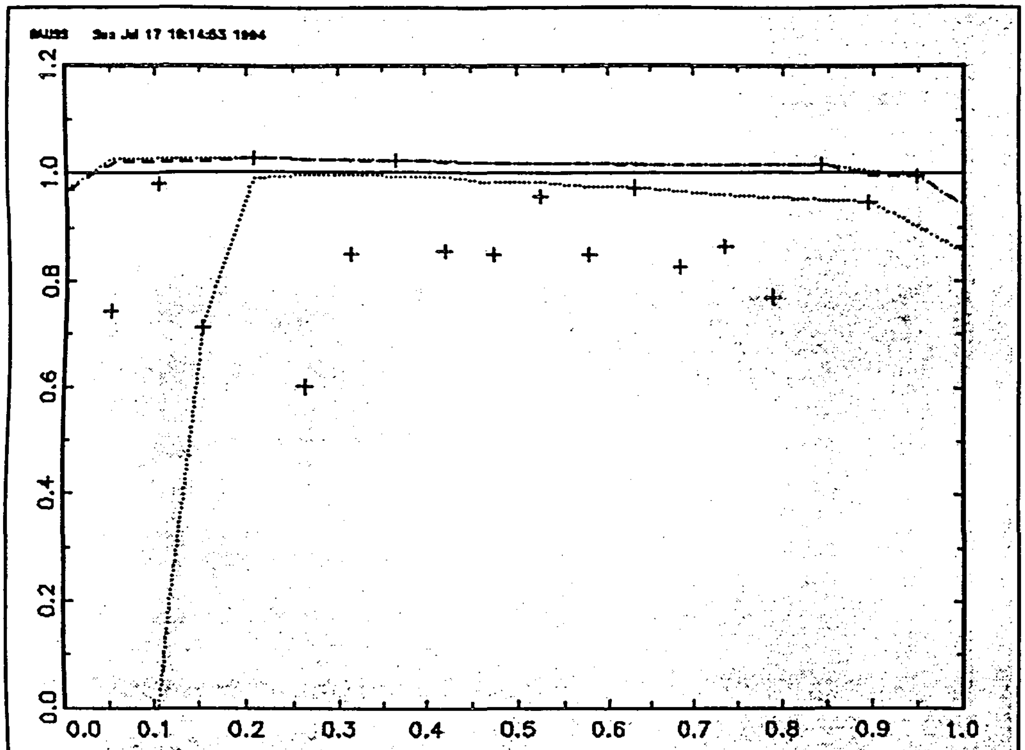
If on the other hand the average efficiency were known together with the efficiency term being i.i.d., a modified bootstrapping algorithm developed for regression analysis could be applied to estimate the frontier.[3]

The problem with trying to estimate both things at once is easy to explain by analogy: Suppose you would have a sample of positive values representing the population of something with a fixed upper bound. Suppose further the task were to estimate the average deviation of a value from this upper bound. Now consider two equally sized samples from this population and without loss of generality assume, the maximum value in sample 1 were greater than the maximum value in sample 2. Knowing that the true maximum for both samples must be the same you can conclude that in a probabilistic sense the mean deviation of sample 1 from the upper bound must be smaller than the mean deviation of sample 2. And this expected difference should enter the variance measure for the average deviation statistic. But in a procedure, estimating average and maximum at the same time some of that difference gets lost, because a center statistic and a maximum order statistic are always to some extent positively correlated.

I therefore conclude, that a confidence interval for the average efficiency estimator from the above procedure is biased upwards, because a DEA frontier is sort of a maximum order statistic. So to get more reliable confidence intervals I decoupled the two estimation problems by a nested bootstrapping algorithm. In one step I do resampling based on a fixed frontier to trace out the variation of the confidence interval related to the unknown center of the distribution. In another step I estimate the frontier based on a fixed average efficiency line, thus capturing the variation in the statistic due to variability in the frontier estimate. However, in this second step occurs another problem: It is the failure of the bootstrapping procedure in estimating a maximum. Let me illustrate that by Picture 2.

It shows three quantiles of an estimated frontier for the model from above, based on 1000 repetitions of the resampling scheme. The sample values are shown as crosses. The picture is translated into the X vs. efficiency space. The solid line represents the frontier itself and so is horizontal, the dashed line is the median of the frontier estimated by connecting the DEA frontier points and the dotted lines show the 5% resp. the 95% quantil of this frontier. Obviously the 5% quantil and the median collaps here, which corresponds to a mass point in the distribution of the frontier estimator. Because I assume the underlying distribution to be continous, this should not occur. It means, that from resampling the original observations, one does not get a reliable picture of the variability of the frontier estimate.

[3] In this instance of course there would be no point in estimating a confidence interval for the average efficiency.



Picture 2: The 5%, 50% and 95% quantils for the DEA frontier from a naive bootstrapping approach to detect the variability of DEA frontier estimator. The 20 observations are shown as crosses. The solid line is the true frontier. All figures are divided by the respective frontier value.

So in the second step I do not resample from the original observations. Rather I construct a non-parametric regression line, which by assumption of i.i.d. error terms gives a consistent estimate of the frontier shifted downward by a factor corresponding to the average efficiency. Then I calculate the deviations from this average efficiency line and construct a Gauss-Kernel density on the basis of these deviations. And this will finally be used for resampling. Notice that the theoretical justification for such a density estimate requires an unbounded support of the underlying random variable. And this in turn means, I actually need noise in the error term of my production function, because the efficiency term is bounded.

This algorithm yields a consistent estimate of the distribution of the average efficiency estimate from DEA, given that DEA's average efficiency estimate is unbiased in the first place.[4] Table 2 gives an impression of the small sample performance of this algorithm. It shows that a simple teststatistic, in this case the centered and standardized average efficiency estimator, performs very well compared to the true distribution of this teststatistic under the null hypotheses of an average efficiency of 0.85. The coverage probabilities say, that on average the bootstrapping distributions of the teststatistic based on 20 random observations cover x-percent of the true lower 5% and upper 5% ranges. [5]

---

[4] The latter requirement can be considered fulfilled in a neighborhood of certain sample sizes

[5] The true values for the quantils were themselves constructed by Monte Carlo Simulations, repeating DEA for 1000 random samples.

	5% Quantil	95% Quantil	Coverage Probabilities lower 5%	upper 95%
20% noise	-1.754	1.477	0.041 (0.004)	0.058 (0.004)
30% noise	-1.766	1.469	0.039 (0.004)	0.061 (0.004)
40% noise	-1.862	1.447	0.049 (0.003)	0.060 (0.004)

Table 2: Coverage probabilities of the estimated distribution of the centered and standardized statistic E[eff] derived from comparison with the true distribution under the Null-Hypothesis of E[eff]=0.85. In brackets the standarddeviations of these quantities. All figures are based on 100 random samples.

Conclusions

The case for DEA in noisy environments is there, as table 1 shows. For noise to error ratios of up to 30% and small samples DEA definitely outperforms SFPE as measured by the MSE. With increasing sample size and more trust as concerns a reasonable parametrization SPFE of course gains ground. But to my experience typical applications will lie in the range of up to 30 observations (if corrected for dimensionality or, statistically speaking: degrees of freedom) and the parametrization problem is rarely ever a trivial matter. So DEA must be considered a serious competitor to SPFE in real world applications.

Once DEA is established as a principal alternative, the problem of statistical interpretation arises. I gave some explanations on how to construct a bootstrapping procedure which allows to approximate the distribution function for DEA's average efficiency estimator. The problem thereby is to decouple the process of average estimation (used to construct the bootstrapping samples) from the process of maximum estimation (used to construct the frontier). Statistical results in terms of coverage probabilities as given in Table 2 show the power of this approach. Notice however, that I neglected the inherent bias problem . The possible usefulness of bootstrapping in DEA therefore ultimately depends on a comparison of bias against variance, which can only be made on a case by case basis.

Selected Literature

EFRON, Bradley and Robert J. TIBSHIRANI: *An Introduction to the Bootstrap.* Chapman & Hall, 1993

HÄRDLE, W: *Applied Nonparametric Regression.* Cambridge University Press, 1990

FRIED, O.F., C.A. Knox LOVELL and Shelton SCHMIDT (editors): *The Measurement of Productive Efficiency.* Oxford University Press, 1993

FRIEDMAN, Lea and Zilla SINUANY-STERN: *Scaling Units via the Canonical Correlation Analysis and the DEA.* Forthcoming in Management Science, 1994

Bisher sind in dieser Reihe erschienen:

Eigl R., Experimentielle Methoden in der Mikroökonomik, No. 1, Mai 1991.

Dockner E., Long N.V., International Pollution Control: Cooperative versus Non-Cooperative Strategies, No. 2, September 1991.

Andraea C.A., Eigl R., Der öffentliche Sektor aus ordnungspolitischer Sicht, No. 3, Oktober 1991.

Dockner E., A Dynamic Theory of Conjectural Variations, No. 4, Oktober 1991.

Feichtinger G., Dockner E., Cyclical Consumption Pattern and Rational Addictions, No. 5, Oktober 1991.

Marterbauer M., Die Rolle der Fiskalpolitik im Schwedischen Wohlfahrtsstaat, No. 6, Dezember 1991.

Pichler E., Cost-Sharing of General and Specific Training with Depreciation of Human Capital, No. 7, Dezember 1991.

Pichler E., Union Wage Bargaining and Status, No. 8, Dezember 1991.

Pichler E., Costs of Negotiations and the Structure of Bargaining - a Note, No. 9, Dezember 1991.

Nowotny E., The Austrian Social Partnership and Democracy, No. 10, Dezember 1991.

Pichler E., Walther H., The Economics of Sabbath, No. 11, April 1992.

Klatzer E., Unger B., Will Internationalization Lead to a Convergence of National Economic Policies?,No. 12, June 1992.

Bellak C., Towards a Flexible Concept of Competitiveness, No. 13, May 1992.

Koren St., Stiassny A., The Temporal Causality between Government Taxes and Spending, No. 14, August 1992.

Altzinger W., Ost-West-Migration ohne Steuerungsmöglichkeiten?, No. 15, September 1992.

Bellack Ch., Outsiders' Response to Europe 1992, Case of Austria, No. 16, December 1992.

Guger A., Marterbauer M., Europäische Währungsunion und Konsequenzen für die Kollektivvertragspolitik, No. 17, January 1993.

Unger B., van Waarden F., Characteristics, Governance, Performance and Future Perspectives, No. 18, January 1993.

Scharmer F., The Validity Issue in Applied General Equilibrium Tax Models, No. 19, May 1993.

Ragacs Ch., Minimum Wages in Austria: Estimation of Employment Functions, No. 20, June 1993.

Ragacs Ch., Employment, Productivity, Output and Minimum Wages in Austria: A Time Series Analysis, No. 21, September 1993.

Stiassny A., TVP - Ein Programm zur Schätzung von Modellen mit zeitvariierenden Parametern, No. 22, December 1993.

Gstach D., Scale Efficiency: Where Data Envelopment Analysis Outperforms Stochastic Production Function Estimation, No. 23, December 1993.

Gstach D., Comparing Structural Efficiency of Unbalanced Subsamples: A Resampeling Adaptation of Data Envelopment Analysis, No. 24, December 1993.

Klausinger H., Die Klassische Ökonomie und die Keynesianische Alternative. Revision ein Mythos?, No. 25, December 1993.

Grandner T., Gewerkschaften in einem Cournot-Duopol. Sequentielle versus simultane Lohnverhandlungen, No. 26, April 1994.

Stiasssny A., A Note on Frequency Domain Properties of Estimated VARs, No. 27, June 1994.

Koren St., Stiassny A., Tax and Spend or Spend and Tax ? An International Study, No. 28, August 1994.