Dieter Gstach

Comparing Structural Efficiency of Unbalanced Subsamples: A Resampling Adaptation of Data Envelopment Analysis

Paper

Comparing Structural Efficiency of Unbalanced Subsamples: A Resampling Adaptation of Data Envelopment Analysis

Dieter Gstach

Working Paper No. 24
December 93

Comments and critique are welcome.

Adress of the author:

Dieter Gstach
Department of Economics,
Vienna University of Economics and Business Administration
Augasse 2-6, A-1090 VIENNA, Austria
Phone: 31336-4965 Fax: 31336-755
e-mail: gstach@isis.wu-wien.ac.at

Abstract *- At a general level the question is adressed, how to compare structural efficiencies for subgroups in the framework of data envelopment analysis. If a multiproduct technology is employed in estimating the production possibility frontier and if the subgroups are distinguished by products, then specific problems arise. It is illustrated by Monte-Carlo techniques that there may be considerable bias in relative structural efficiencies due to unbalanced numbers of observations for different subgroups. A resampling procedure is shown to outperform the simple estimators from data envelopment analysis in such circumstances. The method proposed is applied to austrian farming data to derive conclusions about the relationship between certain specialization patterns and efficiency.*

1. Introduction

Data envelopment analysis (DEA) has prooved to be a versatile tool for efficiency estimation and related matters of interest. This is especially true for comparing multiproduct firms (or decision making units = DMU's) as Charnes, Cooper and Rhodes have pointed out in there seminal paper from 1978. Traditional statistical techniques would in such circumstances have to rely on price information to aggregate for example several outputs measured in quantities to a single revenue figure. But this aggregation may fail, when either market prices are unavailable, as for public goods, or are unreliable in the sense of being accurate scarcity indicators. Unfortunately this and other advantages of DEA are accompanied by some serious drawbacks. Most important and often discussed is DEA's lack of statistical grounding. Another one, arising from problems related to sample size, will be discussed in this paper.

There exist lots of empirical applications dealing with one or another issue in efficiency estimation with DEA.[1] Many of these applications are used to draw conclusions about the relative performance of certain groups of DMU's. In what follows, I will refer to this topic as the problem of evaluationg *relative structural efficiency*.[2] By this term I mean the ratio of two efficiency indices for two subgroups to be compared. For ease of exposition I will restrict to the ratio of the *mean efficiencies*. Using other location parameters would not affect the qualitative conclusions.

Now comparisons of structural efficiencies are a straightforward matter, as long as some conditions are met. The first of these is, that all DMU's are used in the pooled sample for estimating the frontier. Separate estimations instead are likely to lead to biased structural efficiencies, when the numbers of observations are unequal. Generally speaking: The more observations a subgroup consists of, the lower the structural efficiency index for this subgroup will be. A characteristic example for ignoring this condition is found in Aly et.al. (1990). Other authors do not explicitly mention this condition but are obviously aware of it, like Tulkens (1993), who uses balanced samples or Leibenstein (1992), who restricts attention to comparisons of single DMU's.

[1] A helpfull, comprehensive bibliography is from Seiford (1990).

[2] This must not be confused with the traditional use of the term structural efficiency, which, as in Färe et.al. (1985) means congestion efficiency.

The second condition to keep in mind gets important when moulding multi-output technologies, which are typical for DEA applications. To illustrate, what might go wrong when ignoring it, assume, that in a first step of DEA a multi-product technology would be calculated along with the corresponding efficiency measures, based on the pooled sample. In a second step then, structural efficiencies among subgroups of the original sample were compared. The second condition for the reliability of such comparisons in this context is, that the distinction between the subgroups is not related to the specialization pattern of the firms. Because if firms were distinguished for example by the biggest revenue share of one output, this would lead to quite the same difficulties as pointed out above: Unbalanced numbers of firms in the specialized subsamples here is equivalent to estimate different production possibility frontiers for each subgroup defined in this way. Note that this problem would not arise, if one would compare structural efficiencies for example of certain production regions, if these in turn would not be distinguishable by specialization patterns.

But as the example of regional comparisons also makes clear: The problem of unbalanced observations is likely to persist, because specialization often is related to all kinds of production characteristics, like firm size, composition of capital and so on. This is a serious objection to DEA, which is claimed to be designed especially for efficiency comparisons in multi-product environments. On the other hand, distinctions of firms according to there main product seems a good perspective in general and it is therefore desirable to adopt DEA for this special purpose.

The following section of the paper will present the DEA-model used subsequently for estimation. Section 3 then outlines the consequences of unbalanced numbers of observations under specific distributional and specialization conditions. Monte Carlo simulations are used in section 4 to derive the empirical distribution of the estimator for relative structural efficiency for two subgroups, which are actually drawn from identical efficiency distributions. As a first remedy against the bias of the standard DEA estimator an estimator based on resampling is proposed, which has a less spread and less biased distribution. This result again is derived with simulation techniques. Section 5 of this paper deals with austrian agriculture and compares the results form the two approaches. Especially the relative structural efficiency of five subgroups, distinguished explicitly by specialization patterns is investigated. It will be seen, whether comparative results change, when the issue of unbalanced subsample sizes is taken into account.

2. The DEA-Model

For the construction of the multi-product technological frontiers to be estimated below, I use the standard DEA approach introduced by Charnes, Cooper and Rodes [1978]. This covers the constant-returns-to scale (CRS) case, while the variable-returns-to-scale (VRS) is neglected here. To give the precise formal statement of the program used, consider scalar valued inputs x_i, with i=1,2,...n and (1×6)-dimensional output vectors $y_{i,.}$ with typical elements $y_{r,i}$, all of which are supposed to be observed. Let λ be a (1×n)-dimensional vector of weights to be attached to the sample firms for construction of an efficient frontier firm $(\lambda x, \lambda y_1,...,\lambda y_6)$. The programming problem to calculate efficiency for firm k is then:

$$\begin{aligned} &\max_{\lambda,\theta} \quad \theta \\ \text{s.t.: } \lambda y_{.,r} &\geq \theta y_{r,k} \qquad r = 1,2,....,6 \\ \lambda x &\geq x_k \\ \lambda,\theta &\geq 0 \end{aligned}$$

The solution value θ^* gives the maximum possible output expansion for the firm k under consideration within the production possibility set for a given input level. The reciprocal of it gives the (output oriented) efficiency measure $e = 1/\theta^*$ for firm k. Varying the index k from 1 to n and thus solving n programming problems yields efficiency measures for each firm in the sample. The mean of these, $E[e_i]_{i=1...n}$ gives a picture of the overall performance of the pooled sample. But of interest here is especially the ratio

$$r = E[e_i]_{i=1...n_2} \,/\, E[e_j]_{j=1...n_1}$$

which is used to measure relative structural efficiency for two subsamples with sample sizes of n_1 resp. n_2. Of course $n_1 + n_2 \leq n$.

3. Structural efficiency and sample size

As a first step towards demonstrating the effect of unbalanced numbers of observations on relative structural efficiency consider the following simple case: Two groups of DMU's, $i = 1,2,...n_1$ and $j = 1,2,....n_2$, produce goods $y_{1,i}$ respectively $y_{2,j}$ with one input x_i resp. x_j .Both groups consist of completely specialized DMU's. This in fact is a degenerate case of no practical interest, but it illustrates the issue well. Now let $n_2 >> n_1$, so that there are many more observations of group-2 DMU's than of group-1 DMU's. But assume,that the efficiencies e_i resp. e_j of the sample DMU's were drawn independently from the same distribution f(e). So relative structural efficiency r of the two groups should be 1 on theoretical grounds.

Restricting attention to the CRS-case, the situation just described could be graphically illustrated in two-output space (with equalized amounts of input) with all observations lying on the axes y_1 or y_2. Using DEA to construct a CRS production possibility frontier would therefore yield a line connecting the respective maximum efficient firms from both samples.

So on average only the ratio of the mean to the maximum order statistics matter as far as structural efficiency is concerned. And this might favor smaller groups considerably as a glance at Figure 1 below makes clear. It shows, how structural efficiency for two simple and highly skewed efficiency distributions is linked to sample size.

Increasing the sample size obviously reduces structural efficiency on average. So a typical group-2 measure for structural efficiency will be considerably lower than a corresponding group-1 measure. And as will be seen shortly, this argument carries over to non-degenerate multi-product situations with even more power.

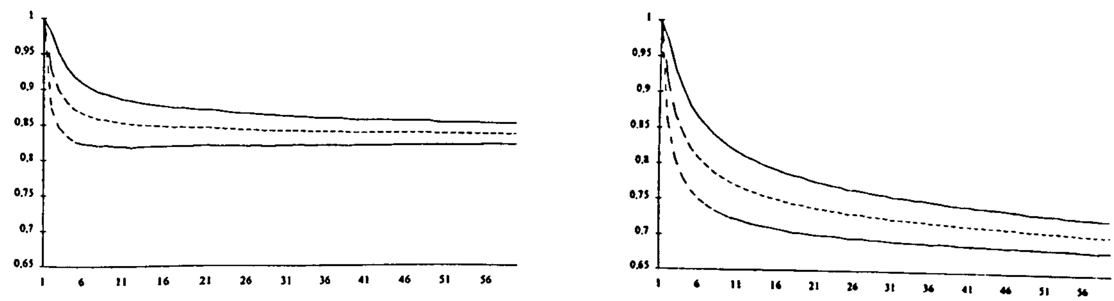

Figure 1: STRUCTURAL EFFICIENCY DEPENDING ON SAMPLE SIZE
Left figure: Efficiency e distributed as $f(e) = 8(e-0.5)$ for $e \in [0.5, 1]$, else $f(e) = 0$
Right figure: Efficiency e distributed as $g(e) = 8(1-e)$ for $e \in [0.5, 1]$, else $f(e) = 0$

4. A Monte-Carlo Simulation

Consider now DMU's $i = 1,2,...n$, producing two different products $(y_{1,i}, y_{2,i})$ in different proportions with a single input x_i , from which every firm uses an equal amount. This corresponds to the CRS-case again. And assume, that interest focusses on a comparison of structural efficiency of two subgroups, which are distinguished by their respective degree of specialization. Neglecting the scaling issue for the moment, I measure specialization by the share $\alpha_r = y_{r,i}/(y_{1,i}+y_{2,i})$, as if both output prices would equal one.

The main assumption in the following Monte-Carlo simulations is, that the efficiencies for all firms are independently drawn from an identical distribution f(e). In what follows, I considered only one special efficiency distribution., which is

$$\begin{aligned} f(e) &= f_N(e \mid \mu_e = 0.75, \sigma_e^2 = 0.13) / [F_N(1)-F_N(0)] \quad \text{for } e \in [0,1] \\ &= 0 \text{ else.} \end{aligned}$$

where $f_N(. \mid \mu, \sigma^2)$ is the shorthand for the normal distribution function with mean μ and variance σ^2 and $F_N(.)$ is the corresponding cumulative distribution function. Such a truncated normal distribution has often been encountered in DEA applications and also plays an important role in estimating stochastic production functions, which might rationalize this specific assumption.

Instead of resampling at random from the distribution f(e) I use order statistics $e_r^{(n)}$, $r = 1,2...n$. This leads to more homogenous results concerning the average performance of the estimators under investigation.

The next step towards simulation is the characterization of the underlying technology. This is assumed to be characterized with a typical input-isoquant

$$y_1^2 + y_2^2 = 1 \text{ for } y_1, y_2 \in [0,\pi/2].$$

So the first quarter of the unit circle constitutes the production possibility frontier in output-space. Technologies (i.e.: DMU's) are supposed to be located along this part of the unit circle according to the following distribution function

$$g(l) = f_N(l \mid \mu_l, \sigma_l^2)/[F_N(\pi/2)-F_N(0)] \quad \text{for } l \in [0,\pi/2]$$
$$= 0 \text{ else.}$$

Again I will make use of order statistics, this time $l_s^{(n)}$, $s = 1,2...n$, instead of random resampling from g(l). By definition, a technology l_i translates into (efficient) output figures by

$$y_{1,i} = \cos(l_i) \text{ and } y_{2,i} = \sin(l_i) \text{ for all } i=1,2,...n$$

The distinction of the two subgroups is based on the maximum revenue share of one product. So if $\alpha_{1,i} = y_{1,i} / (y_{1,i}+y_{2,i}) > 0.5$, firm i is classified as a group-1 firm, else as a group-2 firm. This distinction corresponds geometrically to a partitioning of the first quadrant in output space into two 45° cones. With $\pi/4 < \mu_l < \pi/2$ the technology distribution is skewed to the left, which will result in $n_2 > n_1$. The degree of asymmetry in sample size will depend on μ_l and σ_l^2. For the simulation below three different combinations of these are chosen, such that the ratios n_1:n_2 are 1:2, 1:4 and 1:10 respectively.

Each single loop of the simulation thus starts with attaching efficiencies from the set of order statistics $\{e_r^{(n)}, l_s^{(n)} | r=1,2...n\}$ to DMU's (i.e. to technologies) from the set of order statistics $\{l_s^{(n)} | s=1,2...n\}$ at random without replacement to yield datasets of the form

$$\{x_i = 1, y_{1,i} = \cos(l_s^{(n)}) \cdot e_r^{(n)}, y_{2,i} = \sin(l_s^{(n)}) \cdot e_r^{(n)}, i = 1,2,...n\}$$

In a second step of each loop DEA is applied to this dataset to calculate the firmspecific efficiencies in relation to a two-output best practice frontier based on the pooled sample.
The final step in each loop is to calculate the relative structural efficiency r. As n1 and n2, the subsample sizes approach infinity, this ratio should get approximately 1. But, as outlined in section 3, it is reasonable to expect that this ratio for finite samples is off its target of one on average and thus biased.

Figure 2 below, giving results based on 500 loops of the kind described above, show, that this in fact is true for finite samples. The extent of bias thereby hinges on two things: The first is the degree of asymmetry in sample size as expressed by the ratio n_1:n_2. Lowering this ratio for a given size $n = n_1+n_2$ of the pooled sample, increases this bias as measured by 1 - r. The second determination comes from the size n of the pooled sample. Increasing n while keeping the ratio n_1:n_2 constant, lowers this bias.

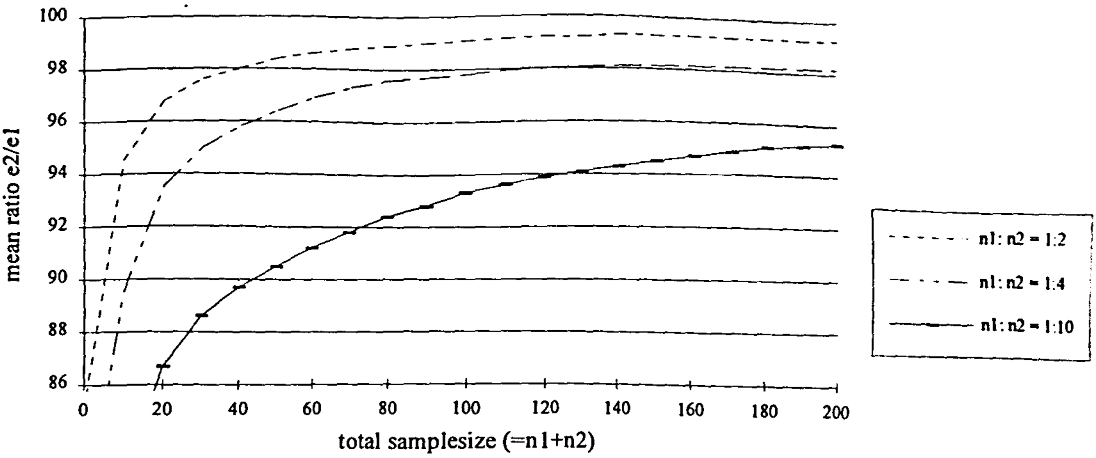

Figure 2: IMPACT OF SAMPLE SIZE ON SAMPLE MEANS FOR RELATIVE STRUCTURAL EFFICIENCY

5. A resampling approach

It is natural to ask, what could be gained, if one were to look for better balanced subsamples. This question can not be answered, without resort to some screening device to cut down the number of observations from the bigger subsample. But data screening in general is likely to introduce a bias of its own. A second caveat with screening is the possible waste of information as compared with a full sample DEA. The key to overcome these difficulties is restriction of attention to relative structural efficiencies. But in a way, this is quite in the spirit of DEA, because DEA gives only relative efficiencies anyhow.

From here to the idea to resample data within the respective subgroups it isn't far. At first, this immediately helps balancing the sample size, when defining $n_B = 2 \cdot \min\{n_1, n_2\}$. The inspiration for the index B comes from bootstrapping, although bootstrapping techniques typically resample from the estimated errors, see Hall (1992). Simplifying I will use the term bootstrapping also for the proposed resampling procedure. To eliminate the effect of sample size on the spread of the order statistics and in accordance with other bootstrapping procedures, I draw samples of size $n_B/2$ with replacement from both subgroups. So any bootstrapping sample consists of equal numbers of DMU's from both subgroups and captures *comparable* effects from the underlying true distributions. For the ratio of 1:10 of n_1:n_2 and various values of n such a bootstrapping procedure (with 300 repetitions each) was applied to 300 random combinations of technologies and efficiency as from above with subsequent DEA of the new sample. Then the ratios r from bootstrapping, labeled r_B along with the corresponding ratios r_D from full fledged DEA were averaged to yield the typical biases for both relative-structural-efficiency-estimators.

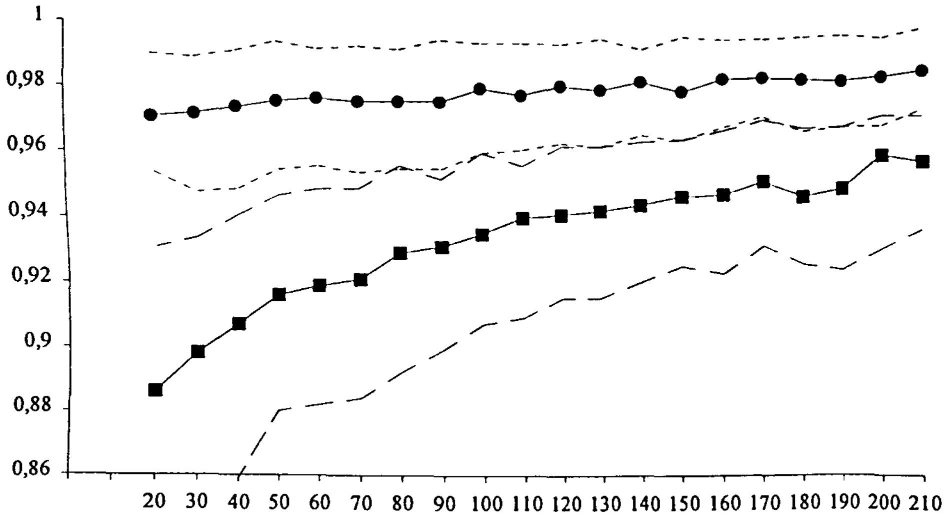

Figure 3: SAMPLE DISTRIBUTION OF TWO ESTIMATORS FOR RELATIVE STRUCTURAL EFFICIENCY

Upper line (circles) corresponds to the bootstrapping DEA estimator and the lower line (squares) corresponds to the standard DEA estimator. The dashed lines represent the 10% and 90% quantiles for the respective estimators.

The results are summarized in Figure 3 above, which also gives the 10% and 90% quantiles. As can be seen, the bootstrapping estimator outperformes the standard DEA estimator in the sense of less bias and more efficiency, especially of course in the lower range of total sample size n. Although this result rests on rather specific assumptions, it is not farfetched to assume, that it will carry over to situations with more than two outputs and that it will be of even more significance, when distributions are more skewed, as the results from section 3 suggest.

6. An Application

In the following I will show the results from an application of the proposed bootstrapping approach and compare it with standard DEA. I used data about austrian farms for the year 1990, the last available year. Screening the original data set left 333 observations for estimation. According to the focus of this paper I modelled a technology which deals accurately with outputs but in a crude fashion with inputs. The six output categories used are, in revenue terms (= millions of austrian schillings) respectively: corn, root crop, calf and cattle breeding, dairy products, pig breeding and a residual category other activities (like renting land, lodging and so on).

Standard DEA according to the program specified in section three yielded first results concerning the relative structural efficiencies of subgroups according to there main products. These results are collected in Figure 4 as the bright bars. The structural efficiency of calf and cattle breeding, which is the most efficient subgroup (according to both kinds of measurement) was used to standardize the other figures. Figure 4 exhibits most clearly the effect of sample size on structural efficiency. So one does not really wonder, why corn farms look better than dairy farms and pig breeding farms. But it is in the same vein that one is suspicious, because despite comparatively big and similar sample sizes, pig breeding farms seem to perform so much better on average than dairy farms. But as the arguments in the preceeding sections suggest, these results should not be taken for granted. Instead, one should apply the bootstrapping mechanism illustrated above and check the results against the ones from standard DEA.

The darker shaded bars in Figure 4 show these relative structural efficiencies from the bootstrapping procedure, again standardized by the figure for the category cattle breeding.

It can be seen immediately, that the negative result concerning the structural efficiency of dairy farms is confirmed, although the result is softened. Anyway: It is not the sample size, that makes the group of dairy farms to appear inefficient, but rather real effects. On the other hand the group of pig breeding farms gained a lot through bootstrapping. For this subgroup sample size did have a strong and misleading influence on the structural efficiency figure from standard DEA and this very possibility is, what makes bootstrapping worthwile.

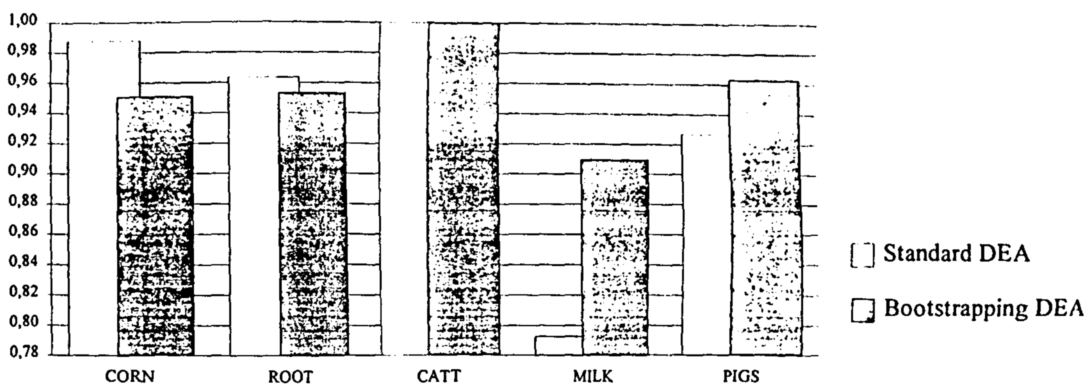

Figure 4: RELATIVE STRUCTURAL EFFICIENCIES COMPARED

7. Concluding remarks

At the heart of this paper was the question of how sample size may influence structural efficiency comparisons. It was demonstrated, that in the context of multi-product DEA the distinction of subgroups to be compared along the lines of specialization patterns gives relative (sub-) sample size a key role in the estimation of structural efficiencies. This impact is undesirable, because it impedes the applicability of DEA.

The paper showed the extent of expectable bias of standard DEA estimators for relative structural efficiency by Monte-Carlo simulations for several distributions of technologies. The more skewed this distribution is, which corresponds to highly asymmetric subsample sizes, and the less observations there are in the pooled sample, the more severe gets this bias.

In section 5 a bootstrapping approach was proposed to overcome this dependency on sample size. Monte-Carlo simulation to establish the distribution of the DEA estimator as compared to its bootstrapping counterpart for a typically observed efficiency distribution shows, that the latter outperforms the former especially in the lower range of total sample size.
The comparative application to austrian farming data suggest, that bootstrapping indeed may help to correct wrong impressions from simple DEA as concerns relative structural efficiency.
A problem of course remains: A deterministic approach like DEA can't be used to derive measures of statistical reliability, unless one introduces efficiency distributions explicitly, at one stage or another of the estimation procedure. Although the proposed bootstrapping approach itself is innocent in this respect (it requires no distributional assumptions or parametrizations whatsoever) the arguments in support of its superiority compared to standard DEA are not.
This should be kept in mind, when applying bootstrapping in the DEA context.

References

Aly, Y. Hassan, Richard Grabowski, Carl Pasurka and Nanda Rangan: *Technical, Scale and Allocative Efficiencies in US Banking: An Empirical Investigation.* The Review of Economics and Statistics, 211 - 218, 1990

Charnes, A., W.W. Cooper and E. Rhodes: *Measuring the Efficiency of Decision Making Units.* European Journal of Operational Research, 429 - 444, (2) 1978

Färe, R., S. Grosskopf and C.A.K. Lovell: *The Measurement of Efficiency of Production.* Kluwer-Nijhoff Publishing, Boston, 1985

Hall, Peter: *The Bootstrap and Edgeworth Expansion.* Springer Verlag, New York,1992

Leibenstein, H. and S. Maital: *Empirical Estimation and partitioning of X-Inefficiency: A Data Envelopment Approach.* American Economic Review, Papers and Proceedings, 428 - 433, (82/2) 1992

Seiford, L.M.: *A Bibliography of Data Envelopment Analysis (1978-1990).* University of Massachusetts at Amherst, Department of Industrial Engineering and Operations Research, 1990

Tulkens, H.: *On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit.* Journal of Productivity Analysis, 1993

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.