Dieter Gstach

Scale Efficiency: Where Data Envelopment Analysis Outperforms Stochastic Production Function Estimation

Paper

Scale Efficiency: Where Data Envelopment Analysis Outperforms Stochastic Production Function Estimation

Dieter Gstach

Working Paper No. 23
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 - Robustness of DEA scale efficiency scores is investigated in the context of non-radial efficiency measures. Most efficient scales are identified with DEA's reference firms instead of traditional clustering techniques. The systematic difference between single- and multi-output technologies as concerns most efficient scales is then examined by comparing applied DEA results. These provide evidence in favor of my proposition, that single-output techniques, like stochastic production function estimation, yield upwardly biased most efficient scales.

1. Introduction

Data Envelopment Analysis (DEA) is now a well established technique for efficiency measurement and related matters. It has proven useful in many applications, involving the evaluation of technical efficiency of multi-product firms or non-market producers, cases with qualitative rather than quantitative data and so on. In such instances DEA is often better suited than the competing technique of stochastic production function estimation (SPFE), which in turn is superior under different conditions. Anyway, both approaches are often specifically used to determine most efficient scales (MES) of production. So the natural question arises: Which one does the better job ? This is the core question of this paper.

I will therefore focus on methodological issues located at the borderline between the two competing techniques. Three consecutive arguments shall lead to the final proposition concerning the systematic difference between single- and multi-output approaches to MES.

The first issue is reliability of DEA's scale efficiency measures. This arises for two reasons. One is errors in the data, which DEA as a true frontier estimation must neglect. The second is the metric used to construct the efficiency measures, from which in turn the scale efficiency measure is derived. DEA requires the projection of a firm to the production possibility frontier. But this projection is not unique and the special projection employed can have a critical impact on the scale efficiency results. Reliability in this respect can be achieved by a simple adaptation of DEA, as will be argued. In applied work this issue, appearing as positive slacks in the programming solutions, is typically neglected.[1]

The second step of my argument, building on reliable scale efficiency measures, is to show a more convincing way to go from these measures to MES's. A traditional way would be to cluster a sample of firms into several groups depending on size (however measured) and than looking at there respective average scale efficiency scores. The average size for the group with the highest score would be considered the MES. But obviously such clustering of firms is arbitrary and the specific classification used can yield different results, as will be demonstrated. An alternative to clustering, which I propose, uses the scale efficient reference

[1] Some authors like Leibenstein & Maital [1992] point towards the possibly destructive impact of slacks on the results, while others, like Aly et.al. [1990] seem to ignore this issue completely.

firms from DEA to establish these MES's. It is shown, that this is an analytically and numerically straightforward procedure, not bound to any awkward clustering scheme.
The final point to raise concerns the question of whether MES-estimates from single-output-models exhibit any systematic difference to analogues from multi-output-models. Some evidence in this respect can be found in the literature (see for example Vandenbroek et.al. [1980]), but the issue has not yet been systematically investigated. Based on the two steps before, I will support further evidence for the proposition, that single-output techniques, like SPFE, yield upward biased estimates for MES. This is considered as yet another argument for using programming approaches like DEA, as opposed to statistical procedures like SPFE, for efficiency measurement.

2. Methodology

The framework to analyze these issues will be traditional convex hull programming [2, 3] based on a true multiproduct-technology.[4] Models with constant returns to scale (CRS) and with variable returns to scale (VRS) will be applied, the former yielding measures for total efficiency and the latter what I will call primary efficiency. It is necessary to calculate them both, because the measure of final interest, namely scale efficiency is derived from these. The orientation of projection towards the production possibility frontier will be radial with respect to the outputs. This means, that efficiency has to be interpreted as possibel proportional output increase given the amounts of inputs. As mentioned above, radiality is necessary to get scale efficiency measures at all, while the output-orientation could be replaced by input-orientation, although leading to different results! [5]
Evaluating the efficiency of firm i under CRS resp. VRS assumptions requires the following two programming problems to be solved:

[2] The standard references are Charnes et.al. [1978] and Banker et.al. [1984]. The major alternative programming approach is based on a free disposal hull (FDH) instead (Tulkens [1993]).
[3] I do not employ non-Archimedian constants though. Therefore the programming slacks must be checked too to identify efficient firms and some positivity constraints concerning the input and output variables must be invoked to guarantee existence of solutions to the programming problems.
[4] In fact, a multiproduct-technology is the one and only option when interest focusses on isolated technological efficiency rather than on economic efficiency, which captures technical as well as allocative effects. And such a focus is reasonable from a policy maker point of view, when pricing in input or output markets is non-competitive. Because in these circumstances there is no point in measuring allocative efficiency of private producers at all. Markets for agricultural products in many countries seem highly regulated, so the above considerations apply here for example. highly regulated, so the above considerations apply here for example highly regulated, so the above considerations apply here for example.

[5] The task of input minimization is by no means neglected with this decision. But I felt very uncomfortable with exogenous output and less so with exogenous inputs. See Schmidt [1986] for a good exposition on this issue.

CRS-Modell (1)

$$\max_{\theta,\lambda,s_O,s_I} \theta \quad s.t.: \quad \lambda' Y - s_O = \theta\, y_i \qquad \lambda' X + s_I = x_i \qquad \lambda,\theta,s_O,s_I \geq 0$$

VRS-Modell (2)

$$\max_{\phi,\gamma,r_O,r_I} \phi \quad s.t.: \quad \gamma © Y - r_O = \phi\, y_i \qquad \gamma © X + r_I = x_i \qquad \textstyle\sum_{j=1}^{n} \gamma_j = 1 \qquad \gamma,\phi,r_O,r_I \geq 0$$

where Y is an $n \times s$ matrix of the observed s outputs of n sample firms. y_i is the $1 \times s$ output vector of firm i. Correspondingly, X is the $n \times r$ matrix of the s inputs used by the n firms. x_i is the $1 \times r$ output vector of firm i. The $n \times 1$ vectors λ and γ give the weights to construct the reference firm from the sample firms in each modell. The only difference between the CRS and VRS-modell is, that the latter requires the production possibility frontier to be constructed from convex combinations of sample firms only. The scalars θ and ϕ are the expansion factors for projection to the frontier. The slack vectors of the output and input restrictions are s_O and s_I respectively r_I and r_O

The efficiency measures of firm i are than defined as:

Total radial efficiency (3)

$$e \equiv \frac{1}{\theta^*} = \frac{\|y_i\|}{\|\lambda^{*\prime} Y - s_O^*\|}$$

Primary radial efficiency (4)

$$e_P \equiv \frac{1}{\phi^*} = \frac{\|y_i\|}{\|\gamma^{*\prime} Y - r_O^*\|}$$

where * indicates a solution to (1) resp. (2). But if s_O^* is positive, $e = 1$ in fact does no longer mean 100% total efficiency, because radial expansion in this case does not exhaust all of the reference firms outputs. Analogous reasoning applies to the case of positive s_O^* with $e_p = 1$. To take account of these possibilities I will additionally use tougher, non-radial [6] measures, readily available from the solutions to (1) resp. (2), namely:

Total non-radial efficiency (5)

$$\tilde{e} \equiv \frac{\|y_i\|}{\|\lambda^{*\prime} Y\|}$$

Primary non-radial efficiency (6)

$$\tilde{e}_P = \frac{\|y_i\|}{\|\gamma^{*\prime} Y\|}$$

Scale efficiency e_S is defined only for the radial case as:

$$e_P \equiv \frac{e}{e_P} \tag{7}$$

A meaningful non-radial analogue to (7) does not exist and thats where the problems arise, this paper deals with in the following sections.

The complete efficiency evaluation of a sample of firms would require the above steps to be repeated for every firm, the firm index i thereby going from 1 to n with obvious notational changes in the programming problems (1) and (2).

[6] Non-radial measures must be used cautiously, because they normally depend on the units of measurement. See Russel [1985] for a rigorous treatment and Charnes et.al. [1993] for a comprehensive exposition.

3. The Data

To illustrate my arguments I will repeatedly use austrian farming data, which the Austrian ministry for agriculture collects each year from over 2000 bookkeeping farms. I used the latest available figures from 1990 and built two rather homogenous subsamples from these observations: The first consists of 131 crop farms and the second of 530 dairy farms. The criterion of distinction was, that revenues from crop farming respectively from dairy farming exceeded 50% of total net revenues.

Four inputs are considered: Labor, land, capital usage and current expenses. *Labor* is measured in numbers of full time workers during the year. *Land* is measured in acres used, including land hold under lease and excluding farmed out land and some other corrections. The *Capital* usage is simply identified as depreciation and given in nominal terms (millions of Austrian Schillings). The final input is current *Expenses* for fuel, fodder, insurance, veterinarian care and the like.

A multiproduct technology was moulded with five different output categories, all measured as revenues in millions of austrian schillings, accruing from the sources *crop farming, dairy farming, pig-breeding, forrestry* and a residual category *other*. A control model used the sum of these revenues as single output-category.

It is assumed, that output prices across regions and production types are equal, which is a harmless assumption in a heavily regulated market like the one for austrian agricultural products. So I interpret these outputs as adequately scaled versions of some imaginary quantities.

4. Scale efficiency and slacks

Theoretically, scale efficiency under radial projections is a simple measure derived from formula (7) as the ratio between total and primary efficiency.

Due to the incorporated slacks in (3) and (4) however, measure (7) has to be interpreted cautiously. The reason for this is illustrated in Figure 1 with a production possibility frontier for a technology with one input and two outputs.

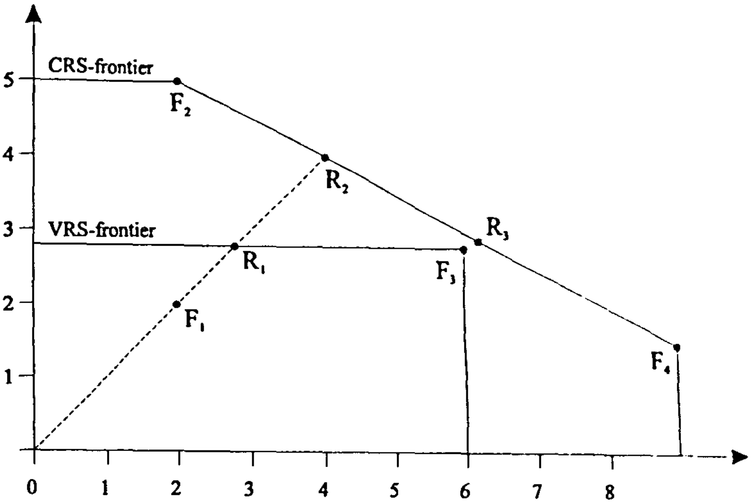

Figure 1: Input Isoquants

Consider firm F_1. According to radial measurement total efficiency of firm F_1 is $e=\overline{OF_1}/\overline{OR_2}=0.50$, primary efficiency is $e_P=\overline{OF_1}/\overline{OR_1}=0.72$ and scale efficiency therefore $e_S=0.69$. Measuring non-radial instead yields $\tilde{e}_P=\overline{OF_1}/\overline{OF_3}=0.42$, showing that primary efficiency in fact is much lower, while a reasonable, non-radial scale efficiency concept like $\overline{OF_3}/\overline{OR_3}$ is much higher, here about 0.98. This deviation makes a big difference when assessing the scope of possible efficiency improvement by resizing firms. One can not rely solely on non-radial measures though, because they are unit-variant, so it should become a standard in applied DEA, to report both sorts of metrics.

Table 1 will help to illustrate this problem:

	cropfarms		dairyfarms	
	all 131 obs.	40 random obs.	all 530 obs.	40 random obs.
e	0.867	0.942	0.790	0.919
e^*	0.857	0.881	0.781	0.896
e_p	0.892	0.965	0.821	0.953
e_p^*	0.886	0.902	0.813	0.931
e_S	0.971	0.977	0.963	0.964

Table 1: RADIAL VS. NON-RADIAL EFFICIENCY MEASURES

Using only 40 of the crop farm observations (drawn at random) led to the figures in column 3 of crop farms, based on 100 randomizations and taking the mean. These figures show clearly, that small samples (especially in relation to the number of outputs used, namely five) creates significantly different results between radial (e, e_p) and non-radial (e^*, e_p^*) efficiency measures. In the light of the above considerations therefore, the derived measure of 0.977 for the average scale efficiency cannot be given any serious meaning.

Looking at column 2 instead, reporting the corresponding results for the whole sample of 131 observations, shows no significant difference between the two measures (based on one sided Mann-Whitney tests for location difference). I therefore conclude, that the average scale efficience of 0.971 is reliable. The main reason for this positive result certainly is the higher number of observations (131), which gives a much richer description of the efficiency frontier than only 40 observations. Columns 4 and 5 of Table 1 give the corresponding figures for dairy farms. Although the results seem less compelling, Mann-Whitney-tests allow the same qualitative conclusions at significance levels of 0.1% or less.

Of course the borderline between meaningfull and irrelevant scale efficiency figures is only sketched by these considerations, but the need to look at non-radial measures should be obvious by now. Anyway: In what follows I will pursue my arguments on the presumption, that my scale efficiency figures for both samples are reliable in the sense indicated.

5. Clustering vs. using DEA's reference firms

To determine MES's in the framework of DEA one typically clusters the scale efficiency measures according to a size variable and then looks at the maximum average efficiency among these clusters. But this approach must remain arbitrary, because an optimal way of clustering does not exist. A look at Figure 2, showing the results for different modes of clustering of crop- resp. dairy farms scale efficiency figures, illustrates the issue well:

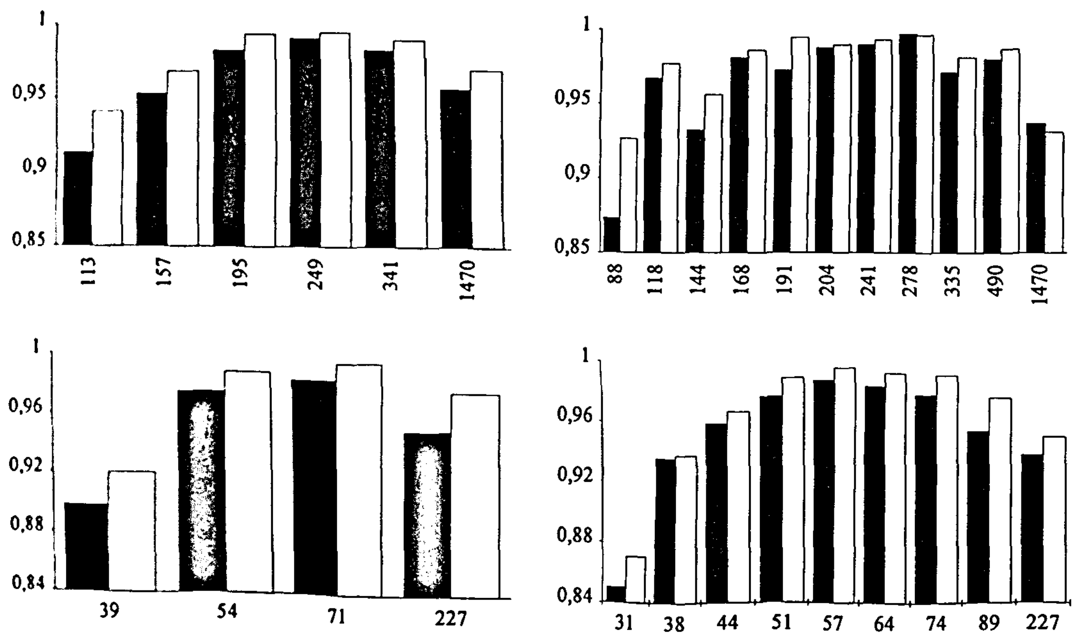

Figure 2: Mean and median scale efficiencies for crop-farms resp. dairy -farms
Legend: Top row = crop farms, bottom row = dairy farms, solid bars = means, empty bars = medians, X-axis gives farm size in acres as upper bounds for the respective cluster, Y-Axis shows the mean resp. median scale efficiency of clusters. In any one picture all clusters contain an equal number of observations.

With six clusters of crop farms, chosen that each comprise of an equal number (22) of farms, the scale efficient size, as measured in acres of land in use, seems to be roughly 220 acres.

Clustering the same sample with 11 size classes would yield a corresponding optimum size of 260 acres. For dairy farms the two figures for optimal size are approximately 63 versus 54 acres. Now this might not look like much of a difference. But in terms of policy recommendations, concerning restructuring agricultural production, this difference matters: If the MES for dairy farming in Austria is 63 acres we find, that 330 farms are too small while 200 are too big. With a MES of 54 acres the proportion is about 265:265. So the former MES clearly indicates a demand for increased farm size, while the latter does not. One must therefore conclude, that traditional clustering techniques serve their purpose rather badly.
But DEA offers another option to evaluate MES's, which does not resort to such ambiguous methodology. The following formula, in fact a simple derivative of DEA-calculations gives the scale efficient reference firm for firm i

$$x_i^* = \frac{\|\lambda^{*\prime} X\|}{\|\Sigma_{j=1}^{n} \lambda_j^*\|} \tag{4}$$

x_i* is a vector of input levels, calculated from the input levels of the CRS-reference firm standardized by the sum of weights. This is the optimum sized firm, built as convex combination of observed scale efficient firms. If one takes care of the homogeneity of the sample firms with respect to the production type, averaging over these reference firms yields a straightforward typical scale efficient firm, that is to say: MES's in all inputs. The reason this technique doesn't work for heterogenous samples is, because the average of MES's for differently specialized firms describes an unrealistic and hence meaningless 'everything-producer'.

6. MES's: Single- vs. Multi-Output Models

As outlined in section 3 I screened the original data to get two homogenous samples. So they qualify for applying the above technique for determining MES's, which leads to the results reported in table 2 for crop-farms resp. dairy-farms. The calculations are based on two different models: One with a single output (as sum of revenues) and one with five outputs. The two models give different MES's: The triple greater-than signs in the column labeled "1:5" indicate 0.1% significance in Mann-Whitney-Tests of the hypotheses, that the firm-specific MES's scales from the single output-model are greater than their counterparts from the five-output model. How can this be explained ?
One answer would be to say, that the differences reflect an inherent deficiency of DEA: The more variables a DEA model contains, the higher is measured efficiency of any kind. Therefore MES's figures should exhibit a tendency towards the sample means as the number of variables increases. This potential loss of information due to excessive use of variables

can't be overemphasized, but it does not explain for example the differences in the optimum land figures for the crop-farm sample or the labor figures in the dairy-farm sample. In these instances the single-output model gives MES's that are lower than the sample mean (although not significantly so) while the five-output model gives figures even lower and further away from than the sample means (with 5% resp. 0.1% significance). So there must be another reason for the observed differences.

		1 output (1)			5 outputs (5)				sample (S)	
		MEAN	STD	1:5	MEAN	STD	1:S	5:S	MEAN	STD
Crop-Farms 131 obs.	Land	54.53	12.10	>>>	46.39	17.92	~	<	61.71	48.45
	Labor	2.62	1.04	>>>	2.16	0.95	>>	~	2.27	1.03
	Capital	0.22	0.09	>>>	0.18	0.09	~	<	0.21	0.11
	Expenses	0.78	0.46	>>>	0.63	0.34	>	~	0.68	0.42
Dairy-Farms 530 obs.	Land	23.99	5.79	>>>	23.09	8.28	>>>	~	23.54	10.72
	Labor	2.11	0.43	>>>	1.91	0.45	~	<<<	2.12	0.60
	Capital	0.14	0.03	>>>	0.13	0.04	>>>	~	0.14	0.06
	Expenses	0.39	0.23	>>>	0.37	0.26	>>>	>	0.37	0.27

Table 2: MES's of Crop and Dairy Farms for two models (with sample means)

This other reason can be found in the fact, that single- and multi-output models use different prices for aggregating the different output-quantities to a single value figure. Single-output models use observed prices and their method of aggregation is simple summation. Multi-output models instead calculate shadow prices within the process of estimating the production possibility frontiers. And these shadow prices fulfill certain optimality conditions like best-rationalizing the quantitiy data in the sense of the profit-maximization hypotheses. Now it is easy to verify, that in a multi-output context a relative scale efficient firm under one price regime can be quite scale inefficient under another. And from there it is straightforward to realize, that higher prices (compared to shadow prices) in favor of products, which are better produced on a big-scale, must, from the single-output perspective lead to upward biased MES's. And that is exactly what we observe.

But which prices to use is not a matter of discretion, because efficiency measurement ultimately rests on a competitive economy point of view. So it is logical to require the prices used for aggregation to reflect this properly and this leads unambigously to shadow rather than observed prices. If the two were the same, one could expect the MES results to be equal. But, as is the case not only for Austrian agriculture, prices for products like wheat or milk are regulated and too high. Therefore the most specialized wheat or milk farmers, using much more land resp. labor than smaller mix producers, have too much weight in determining MES's. That's exactly, why the MES's from single-output models exhibit the sort of upward bias reported in table 3. I think, this also explains much of the differences between MES's calculated with DEA resp. SPFE, reported in other articles.

Conclusions

This paper considered in detail the problem of estimating scale efficiency and determining MES's. I argued, that the two main techniques in this field, DEA and SPFE, are not equally suited for this task.

A problem with DEA is, that its scale efficiency figures depend crucially on the robustness of the logical prior efficiency measures. I proposed comparing radial with non-radial measures to get control over this point.

A problem of both approaches concerns the way of deriving MES's from scale efficiency figures. I illustrated the weakness of clustering techniques in this respect and showed, that DEA offers an alternative determination of MES's, based on reference firms. This alternative is free from arbitrariness, has a straightforward economic interpretation and is easy to calculate from readily available DEA results. SPFE can't cope with DEA here.

The last and main problem investigated was the difference between MES's as identified by single-output models (SPFE in fact always uses single-output models!) and MES's from multi-output models. Using the methodology proposed in section 5, MES's for a single-output and a five-output model were calculated and led to the typical finding of related studies, that MES's from single output-models are higher. I explained this phenomenon by the different prices used to aggregate outputs in the two sorts of models. And I concluded, that whenever prices are non-competitive, one could no longer use single-output models and that MES's from stochastic production function estimation are therefore often upward biased.

References

Aly, H.Y., Richard Grabowski, Carl Pasurka and Nanda Rangan. *Technical, scale and allocative efficiencies in U.S. banking: An empirical investigation.* The Review of Economics and Statistics, 211 - 218, 1990

Banker, R.D. *Stochastic data envelopment analysis.* Working Paper, University of Texas at Austin. 1988

Banker, R.D., A. Charnes and W.W. Cooper. *Some models for estimating technical and scale inefficiencies in data envelopment analysis.* Management Science 30 (9), 1078 - 1092, 1984

Charnes, A., W.W. Cooper and E. Rhodes. *Measuring the efficiency of decision making units.* European Journal of Operational Research 2, 429 - 444, 1978

Charnes, A., W.W. Cooper, A.Y. Lewin and L.M. Seiford. *Basic DEA Models.* In A. Charnes et.al. (editors): Data Envelopment Analysis: Theory, Methodology and Applications. Quorum Books, forthcoming, 1993

Leibenstein, H. and S. Maital. *Empirical estimation and partitioning of X-Inefficiency: A data-envelopment approach.* American Economic Review, Papers and Proceedings 2 (82), 428 - 433, 1992

Russel, R.R. *Measures of technical efficiency.* Journal of Economic Theory 35, 109 - 126, 1985

Schmidt, P. *Frontier production functions.* Econometric Review 4, 289 - 328, 1986

Sengupta, J.K. *Data envelopment analysis for efficiency measurement in the stochastic case.* Computers and Operations Research 2 (14), 117 - 129, 1987

Tulkens, H. On FDH analysis: *Some methodological issues and applications to retail banking, courts and urban transit.* Journal of Productivity Analysis, forthcoming, 1993

Vandenbroek, J., F.R. Forsund, L. Hjalmarsson and W. Meeusen. *On the estimation of deterministic and stochastic production functions.* Journal of Econometrics 13, 117 - 138, 1980

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.