AN EVALUATION OF ECONOMIC EFFICIENCY OF FINNISH REGIONS BY DEA AND TOBIT MODELS* **

Heikki A. Loikkanen
Department of Economics
FIN-00014 University of Helsinki, Finland
heikki.loikkanen@helsinki.fi

Ilkka Susiluoto
Department of Economics
FIN-00014 University of Helsinki, Finland
ilkka.susiluoto@hel.fi

Abstract: In the first phase of the study, private sector economic efficiency scores for 83 Finnish regions in 1988-1999 were estimated by the Data Envelopment Analysis (DEA) method. Inputs in the DEA analysis included capital stock, employment by education level, years of schooling and volume of local public sector activity. Outputs were regional value added and personal direct real income from employment. Efficiency differences between regions proved to be considerable and they were correlated with several regional factors. In the second phase the differences in the efficiency scores were explained by using Tobit and logistic regression models. In these cross-section models (1988, 1993, 1999 and average of 1988-1999) the explanatory variables included regional characteristics such as population size, distance from national centres, structure of regional economy (concentration), the existence of university, number of students, accessibility index, innovativity index and the number of patents.

* Paper prepared for the 42nd Congress of the European Regional Science Association, Dortmund, Germany, 27. – 31.8. 2002

** Preliminary version, not to be quoted without the permission of the authors.

1. Introduction

The purpose of this paper is to present some results concerning economic performance of Finnish regions. More specifically, we study inter-regional differences in private sector efficiency (or productivity). Our data consists of regional input and output variables and other regional characteristics concerning 83 NUTS 4-level regions in Finland during the period 1988-1999. We use a two stage modelling approach. In the first stage we apply non-parametric programming techniques by using Data Envelopment Analysis (DEA). By applying DEA for each year in our study period we get annual efficiency scores for regions. Fully efficient regions get a score of one and other ones below one. In the second stage we explain (in)efficiency differences between regions with econometric methods by applying logistic regression and Tobit models. In this stage the explanatory variables are different from the DEA stage, describing the environment of productive activity (indirect inputs or externality effects), rather than direct inputs.

Empirically the period 1988-99 is very exceptional in Finnish economic history. In the end of 1980s favourable international economic developments and financial deregulation lead to a boom, which was followed by a deep crisis. Unfavourable international developments, fall of exports to the former Soviet Union, domestic currency and bank crises, and pursued economic policies led to a cumulative decline of GDP of more than 10 per cent in 1991-93. Unemployment, which had been below 5 per cent in the end years of 1980s, sky rocketed to 17 per cent. From 1995 on economic growth has been exceptionally fast and the structure of the economy has changed, as IT industries have been the fastest growing sectors. Despite favourable developments unemployment has remained at high level. Also regionally recent growth has been less evenly distributed than earlier. Fewer urban areas than earlier are attracting new investment and gaining from net migration. Thus it is of interest to see whether efficiency differences obtained from DEA in early, middle and end part of the period 1988-99 differ. On the other hand, we estimate logistic regressions and Tobit models both to explain average scores during the whole period and also study the years 1988, 1993 and 1999 separately.

We want to shed light on these developments by studying regionally the relation between value added of private non-residential sector and/or taxable income, and input factors including capital stock, labour force, regional knowledge base and volume of public sector activity.

This paper is organized as follows. In section 2 we describe briefly the main features of the Data Envelopment Analysis method. In section 3, data sources, as well as input and output variables are introduced. In this connection we also present the models to be employed. In section 4 we present some empirical results concerning efficiency differences across Finnish regions. In section 5 we introduce the econometric models, namely the Tobit model and logistic (log odds) regression model, which are used to explain (in)efficiency differences. Results from econometric models are presented in section 6. Section 7 offers some conclusions.

2. Data Envelopment Analysis

The Data Envelopment Analysis (DEA) method of measuring (in)efficiency is fundamentally based on the work by Farrell (1957) which was further elaborated by Charnes et al. (1978) and Banker et al. (1984). This approach (see e.g. Färe et al.1985) has been widely used in empirical efficiency (or productivity) analysis especially in cases where the units (DMUs) use multiple inputs to produce multiple outputs, and there are problems in defining weights and/or specifying functional forms to be employed in analysis. As DEA does not require input or output prices in determining empirical efficiency frontiers based on best practise technology and related measures of inefficiency, it has become especially popular in the study of public sector.

In the last few years several regional applications of DEA have emerged. Charnes et al. (1989) studied the economic performance of 28 China's cities in 1983 and 1984. Chang et al. (1995) use DEA and the Malmquist productivity index approach to study the economic performance of 23 regions in Taiwan in 1983 and 1990. Tong applied DEA to investigate the changes in production efficiency of 29 Chinese provinces in two papers with somewhat different emphasis (Tong 1996, 1997). Bernard and Cantner (1997) calculate the efficiency of the 21 French provinces in 1978-1989. In a recent study, Maudos, Pastor and Serrano (2000) analyse the relationship between efficiency and production structure in Spain 1964-93. Regional aspects are present also in several

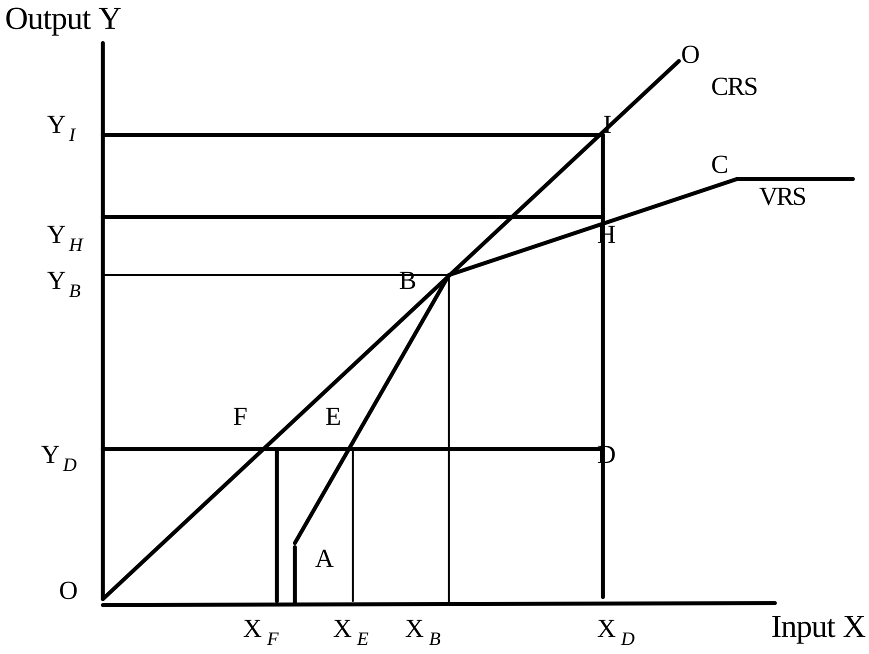
DEA studies, which concern agricultural productivity, see Weaver (1984), Mao and Koo (1997) or Millan and Aldaz (1998).

To keep this paper short, we shall not present mathematically the linear programming background for DEA. We will instead graphically describe a basic case of the method. Four decision making units are described in Figure 1 below; these are the points A, B, C and D. The DMUs use one input X to produce one output Y. Either constant returns to scale (CRS) or variable returns to scale (VRS) can be assumed for the production possibility frontier. In practical research several inputs and possibly more than one output are used, creating a multidimensional situation.

Under CRS, the most efficient unit is B, for which the tangent of the angle measured from the origin (output/input) is greatest (Y_B / X_B). Accordingly, the efficiency frontier under CRS is the line OO. Compared with B, points A, C and D are clearly inefficient. Point D for example uses more of the input (X_D) to produce less of the output (Y_D) than point B. In order to be efficient, only X_F should be used to produce Y_D, or alternatively Y_I should be produced with input use X_D. From this we get X_F/X_D as the relative efficiency of D in the input direction; in the output direction the efficiency score is Y_D/Y_I. Under CRS these two ratios are equal, or $(X_F/X_D) = (Y_D/Y_I)$.

Under VRS the efficiency frontier passes through the points A, B and C. Consequently the relative efficiency of D is X_E/X_D in the input direction and Y_D/Y_H in the output direction, these ratios being generally unequal. In VRS efficiency can be further decomposed into scale efficiency and technical efficiency. Scale efficiency relates the size of the DMU to optimal size; in the input direction it is given by the ratio (efficient input use under CRS)/(efficient input use under VRS), or X_F/X_E in figure 1. Similarly, scale efficiency in the output direction is Y_H/Y_I. This efficiency loss is due to the inoptimal size of the DMU. The rest of the inefficiency of D is technical inefficiency, measured by X_E/X_D in the input direction, or Y_D/Y_H in the output direction.

Figure 1. Efficiency of decision making units in DEA, basic case



Finally, the change in total factor productivity of each DMU can be calculated in DEA, using the so-called Malmquist index approach. This change can be further decomposed into the change in the relative position of the DMU with respect to the efficiency frontier (PPF), and to the movement of PPF itself. For this, see Cooper, Seiford and Tone (2000).

3. Data and models in the DEA estimation of regional efficiency scores

In the first phase of the study, the DEA method was used to estimate regional efficiency scores for the 83 regions for 1988-99. Real value added in the business sector was used as the main output variable. Public sector, non-profit organisations and the residential sector were excluded. Direct real income from private production was used as another output measure, consisting of wages, income from business, trade and profession, and agriculture. Pensions, income originating in the public sector and capital income were excluded. Differences in regional consumption price levels were taken into account. Consequently the figures describe real regional purchasing power of the income earned.

On the input side business sector real capital stock was used as a main variable together with the number of employed. Regional capital stock was separately constructed for the study; this is a crucial resource variable often missing in regional economic research. The number of employed was obtained from regional employment statistics, and it was divided into skilled labour with at least a lower college degree and unskilled labour. Sum of years of education of the population was used as an input measure of knowledge base of the region, supplementing the division of employment into skilled and unskilled labour. As a final input factor, regional value added of the public sector was used.

Outside the proper inputs, domestic economic accessibility of the regions was seen as a potential background factor for efficiency. Consequently an (inverse) accessibility measure was calculated, weighing for each region the road distances to all other regions by the value added of the destination regions. This distance factor was then used as one of the explaining factors in the Tobit and logistic regression models.

Five DEA models were constructed, ranging from a basic one output – two inputs system (value added / labour and capital) traditionally used in production function studies to a system with two outputs and four inputs (value added, real income / construction capital, machinery and equipment, skilled labour, unskilled labour), the latter demonstrating the possibilities of the DEA method, compared to ordinary parametric production function analysis. These models formed a succession in two directions, namely increasing the number of inputs and varying the outputs. The assumption of constant returns to scale was applied throughout in the estimation.

As a result we got efficiency scores, which ranged from 1 for efficient units forming the production possibility frontier (PPF) each year to values lower than 1 indicating the degree of inefficiency. Due to problems in data availability, only three of the five models were applied for 1998-99.

The five models gave somewhat varying efficiency scores for the individual regions, the correlation coefficients between the results of the model pairs averaging at +0.59. Finally, single efficiency numbers for each region and each year were obtained by taking averages over the five models. This completed the DEA estimation.

4. Results from the DEA estimation of regional efficiency

Some interesting results and regularities were obtained from the efficiency estimation. First of all, regional differences in efficiency proved to be considerable. During the whole period 1988-1999, the Helsinki Region (the capital region) topped the list with an average efficiency score 0.993, while the lowest ranking region had a score 0.671. Consequently, it can roughly be stated that the weakest area produces about 33 per cent less outputs than the strongest for the same inputs, or alternatively, if the weakest area had been as efficient as the strongest one, it would have produced its outputs with 33 per cent less resources.

Table 1: Correlation coefficients between average DEA-efficiencies and some regional variables

		Average DEA-efficiency			
		1988-90	1991-95	1996-99	1988-1999
Population	+0.40	+0.36	+0.34	+0.38	
Net migration rate		+0.26	+0.38	+0.51	+0.48
Average specialisation in production (Herfindahl index)		-0.05	+0.25	+0.37	+0.27
Domestic distance		-0.34	-0.51	-0.56	-0.52

The size of a region seemed to be connected with efficiency. Among the ten highest-ranking areas, three were among the country's biggest cities, i.e. Helsinki, Tampere (sixth in our efficiency ranking) and Oulu (tenth). Moreover, all the ten biggest regions by population ranked above the median in 1988-99. It seems that large size brings certain advantages of agglomeration that raise efficiency. Similarly, all university cities were more efficient than median, with two exceptions. However, universities are typically located in large cities, making a further analysis necessary.

Another factor that can be discerned among the top regions is specialisation, six of the top ten regions being specialised in paper and pulp industry. A different example of specialisation is given by Salo region in Southern Finland with a Nokia mobile phone factory. Salo ranked second for the whole period. In these specialised and often small regions we may talk about the effects of localisation advantages and/or economies of scale at plant or firm level.

The most inefficient regions were usually sparsely populated and peripheral areas of primary production. Among the ten most inefficient regions, only one had a population above median.

To some extent, the efficiency scores were also correlated with regional growth in the 1990's and regional recovery after the Finnish economic depression of the early 1990's. On the average, employment has developed more favourably in the more efficient regions (r = +0.68 between private sector employment growth rate 1996-99/1988-90 and average efficiency 1988-99). The result was not self-evident, because high DEA efficiency rates also imply economical use of labour. Efficient regions often also had a more positive net migration than inefficient ones.

As to the geographical perspective, the most efficient areas usually lie in the southern and western part of Finland, while the regions in the northern provinces of the country predominantly ranged among the weakest quarter, particularly after the depression of early 1990s. Possibly a peripheral location has even become a clearer obstacle to efficiency during the time period 1988-1999. It is difficult for peripheral regions to be efficient, unless they are strongly and favourably specialised. Yet the picture has also mosaic-like elements, and perhaps even increasingly so in the southern and central parts of the country.

According to the results, differences in efficiency between regions increased between 1988-1999. In 1988-90, the difference between the most and the least efficient regions was around 30 per cent, but in 1996-99 it was already about 40 per cent. The average score for a median region was 0.851 in 1988-90 but only 0.783 in 1996-99. During the depression of the early 1990's, the position of the weakest areas weakened even further, and for most regions the DEA scores fell during the twelve years from 1988 to 1999. In other words, the majority of regions were increasingly lagging behind the strongest ones at the end of he study period.

5. Explaining (in)efficiency by Tobit and logistic regression models

In this section we introduce two types of econometric models in order to explain (in)efficiency differences among regions, namely logistic (log odds) regression models and Tobit models. As a result of cross-sectional DEA analyses with annual data during 1988-99 we obtained efficiency scores for each region in each year. In each year fully efficient regions (at least one) got an efficiency score equal to one (100 %) and inefficient ones below one (vlow 100 %). Our aim is to explain efficiency differences at three points of time: in boom (1988), bust (1993) and a new boom (1999). For this purpose we first define inefficiency score I as

(1) Inefficiency score = 1 - efficiency score

and define the dependent variable (Y) in Tobit model in two alternative ways. First, defining Y to be equal to I, which is equal to censoring it to a lower limit zero (i.e. ll = 0). In this case only fully efficient regions have Y=0 value. Alternatively, we censor all regions with I <= 0.1 to Y value= 0, and otherwise Y = I. This means that we form a group of efficient regions by the criterion that efficiency score is at least 90 % (inefficiency at most 10 %). When presenting results in the text we apply the lower limit 0.1 (ll= 0.1) to censor inefficiency scores. At the end of next section, we compare these results in one case to those obtained by applying lower limit zero (ll= 0) in connection of the dependent variable in Tobit models.

In both cases the standard Tobit model (see e.g. Maddala 1983) can be defined as

$$(2)\qquad \begin{aligned} Y_i^* &= X_i \beta + \mu_i \\ Y_i &= Y_i^*, \text{ if } Y_i^* > 0 \\ Y_i &= 0 \text{, otherwise.} \end{aligned}$$

Above, X_i is a vector of explanatory variables, I_i refers to region and β is a vector of parameters to be estimated. Y_i^* is a latent variable which can be viewed as a threshold beyond which the explanatory variables must affect in order for Y_i to "jump" from 0 (here being efficient) to some positive value (being inefficient in various degrees). The Tobit model can be estimated by the maximum likelihood method by assuming normally distributed errors μ_i .

In addition to Tobit models applied to cross sections data in 1988, 1993 and 1999, we also consider the whole period 1988-99 and explain average inefficiency scores obtained as an average of annual DEA models. None of the regions is fully efficient every year, which implies that minimum value of inefficiency is not zero but above it. In this case we apply two types of models First, the above Tobit model where average inefficiency scores less than or equal to 0.1 are censored to zero. As a second alternative, we apply logistic (log odds) regression model defined as

(3) $$\ln(I_i/1-I_i) = X_i \beta + \mu_i .$$

In (3) the dependent variable is logarithm of the odds of being inefficient. This model can be applied as none of the I_i is zero or 1, rather all regional inefficiency scores are in the (0,1) interval. We estimate the parameter vector β by OLS.

As for the explanatory variables in Tobit and logistic regression models, we use the following variables. **Population** of the region is aimed to catch agglomeration effects. As a measure of **concentration** of private sector economic activity, we use regional Herfindahl index measure. Its high values indicate specialisation and low values are related to diversified structure. Domestic economic accessibility of the regions was measured inversely by **distance** variable. It was calculated weighing for each region the road distances to all other regions by the value added of the destination regions. **University** is a dummy variable if there is at least one in the region. Alternatively, the number of **students** in the region is used. As an alternative to distance we use a regional **accessibility index**. We also use a regional index of **innovativity** and the number of **patents**. The measures on accessibility, innovativity and patents were obtained from the basic data of Huovari, Kangasharju and Alanen (2001), for which the authors wish to express their thanks.

6. Results from Tobit and logistic regression models

In presenting results we first report models based on average inefficiency scores for the whole period 1988-1999. The dependent variables in the two model types (logistic regression and Tobit) are thus based on an average of 12 cross sectional DEA analyses, and the explanatory variables are averaged correspondingly. In Tobit models the lowest inefficiency scores are censored with a lower limit of 0.1. The number of censored regions is also given in connection of Tobit models.

According to the results of logistic regression models 1 and 2 in Table 2 regions with big population are more efficient as inefficiency decreases with population. This also true for Tobit model 1 whereas the respective coefficient in Tobit model 2 has an opposite sign and is insignificant (at 5 % level). Remote areas are more inefficient as inefficiency increases with distance. A summary index of accessibility of regions gives a similar result. The existence of

Table 2. Parameter estimates of logistic regression models and Tobit models (ll=0.1)* explaining inefficiency of regional economies for 1988-1999

	Regression model 1		Regression model 2		Tobit model 1		Tobit model 2	
	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Constant	-1.295	-7.39	0.711	2.48	0.177	7.54	0.447	13.08
Population (100 000)	-0.304	-6.76	-0.184	-4.00	-0.0429	-2.61	0.0371	0.38
Distance (100 km)	0.157	4.54			0.022	5.46		
Accessibility index			-0.017	-5.26			-0.0024	-6.35
University	-0.161	-0.95			-0.022	-0.96		
Number of students			-2.483	-1.50			-0.575	-2.80
Concentration of structure	-7.258	-6.08	-8.443	-7.41	-0.549	-3.46	-0.661	-4.37
N	83		83		83		83	
Left-censored observations					12		12	
Adj R^2	0.604		0.651					
Pseudo R^2					-0.319		-0.386	

* ll = lower limit of censoring the dependent variable.

a university (dummy) in the region gets an expected negative sign, which, however, is clearly insignificant. When the university dummy is replaced by the number of students (and accessibility index instead of distance) it gets negative coefficients in logistic regression model 2 and Tobit model

2, but only the latter is significant. The more concentrated (specialized) is the structure of the regional (private sector) economy, the higher efficiency.

When Tobit models with the same variables as model(s) 1 in table 2 are estimated with cross-section data for a boom year (1988), last year of depression (1993) and the 5th year of fast growth (1999) we get results reported in table 3. In this table the last column repeats the respective results from table 2 (see Tobit model 1) for the whole period 1988-99. Note that the number of censored (lower limit of inefficiency 0.1) regions is 17 in both 1988 and 1993 whereas it is 7 in 1999.

In table 3 the size of regional population always has a negative effect on inefficiency (i.e. increases efficiency), but the coefficient is significant only for 1988 and the whole period 1988-99, not for 1993 and 1999 (at 5 % level). Distance from other centres within the country always increases inefficiency and is significant, too. University dummy always gets an expected negative coefficient but is never even close to being significant. Concentration (specialisation) of regional economy decreases inefficiency and the coefficient is significant except in 1988.

Table 3. Parameter estimates of Tobit model 1 explaining inefficiency of regional economies for 1988, 1993 and 1999 and 1988-1999 (ll=0.1)*

	1988		1993		1999		1988-1999	
	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Constant	0.179	5.69	0.162	5.27	0.235	7.98	0.177	7.54
Population (100 000)	-0.785	-3.20	-0.373	-1.91	-0.274	-1.37	-0.429	-2.61
Distance (100 km)	0.014	2.58	0.024	5.15	0.024	4.43	0.022	5.46
University	-0.118	-0.38	-0.033	-1.22	-0.028	-0.89	-0.022	-0.96
Concentration of structure	-0.216	-1.04	-0.693	-3.00	-0.643	-3.55	-0.549	-3.46
N	83		83		83		83	
Left-censored observations	17		17		7		12	
Pseudo R^2	-0.259		-0.365		-0.297		-0.319	

* ll = lower limit of censoring the dependent variable.

When the distance variable and the university dummy in table 3 are replaced by accessibility index and the number of students (table 4), the results concerning concentration of economy remain much the same as before but the coefficients of population become clearly insignificant and in two cases also change sign to positive (1993 and 1988-99). Obviously the number of students and population are correlated, and the same is true with accessibility and population. Anyhow, now good accessibility always decreases inefficiency and the coefficient is very significant in all models. Inefficiency decreases with the number of students in all cases and the effect is significant except in 1999.

Table 4. Parameter estimates of Tobit model 2 explaining inefficiency of regional economies for 1988, 1993, 1999 and 1988-1999 (ll=0.1)*

	1988		1993		1999		1988-1999	
	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Constant	0.391	8.28	0.434	9.09	0.497	10.31	0.447	13.08
Population (100 000)	-0.125	-0.49	0.011	0.08	-0.071	-0.55	0.037	0.38
Accessibility index	-0.0020	-3.65	-0.022	-4.55	-0.0023	-4.18	-0.0024	-6.35
Number of students	-1.038	-2.54	-0.619	-2.37	-0.158	-0.64	-0.575	-2.80
Concentration of structure	-0.368	-1.88	-0.907	-3.17	-0.711	-3.73	-0.661	-4.37
N	83		83		83		83	
Left-censored observations	17		17		7		12	
Pseudo R^2	-0.352		-0.355		-0.286		-0.386	

* ll = lower limit of censoring the dependent variable.

Unfortunately some variables were not available for the whole period 1988-99 but only for the last few years. For instance information on innovativeness index and the number of patents by region was available for the last years. In model 3 of table 5 we include the number of patents during 1995-99 in a cross section Tobit model explaining inefficiency in 1999. In Model 4 innovativity index is included instead of patents.

According to the results (Table 5) both an increase in the number of patents (model 3) and innovativity index (model 4) decreases inefficiency (i.e. increase efficiency). Both coefficients are highly significant, too. In models 3 and 4 concentration of economic structure gets negative and significant coefficients as before. The distance variable in model 3 is positive and significant, whereas when it is replaced by accessibility index in model 4, the coefficient is insignificant (at 5 % level). Also the university dummy (model 3) and the number of students (model 4) get insignificant coefficients.

Table 5. Parameter estimates of Tobit models 3 and 4 including innovativity and patents in explaining inefficiency of regional economies for 1999 (ll=0.1)*

	Model 3		Model 4	
	Coefficient	t-ratio	Coefficient	t-ratio
Constant	0.267	8.97	0.572	10.88
Population (100 000)	-0.125	-1.05	-0.067	-0.61
Distance (100 km)	0.020	3.61		
Accessibility index			-0.0012	-1.83
University	-0.0079	-0.29		
Number of students			0.353	1.25
Concentration of structure	-0.630	-3.44	-0.586	-3.21
Patents	-0.554	-2.85		
Innovativity index			-0.0024	-3.09
N	83		83	
Left-censored observations	7		7	
Pseudo R^2	-0.350		-0.349	

* ll = lower limit of censoring the dependent variable.

In the above Tobit models we censored the inefficiency scores using a lower limit of 0.1 which means that all regions with inefficiency score in the range 0.00-0.1 were regarded as zeros (efficient) in Tobit estimations. This means that we have separated a group of top performers from other regions (inefficiency score above 0.1). This separation is somewhat artificial as efficiency scores are continuous variables. On the other hand, we are not using all the information on efficiency differences in estimating the Tobit models. In order to find out the effect of censoring the dependent variable, we estimated two Tobit models with 1999 data using both 0.1 and 0.0 as lower limit.

As a result of lowering the censoring limit (see table 6) the number of efficient (censored) regions decreases from seven to three. With the lower limit at 0.0, the models fit improves (Pseudo R^2 increases). The signs of coefficients of all variables remain the same in comparable models. The size of coefficients changes somewhat and their t-values increase somewhat. Most notable change is related to the population variable in model 1. Its negative coefficient becomes significant, when censoring limit is 0.0 instead of 0.1.

Table 6. Parameter estimates of Tobit models 1 and 2 explaining inefficiency of regional economies for 1999, with ll=0 and ll=0.1*

	Model 1				Model 2			
	ll=0.1		ll=0.0		ll=0.1		ll=0.0	
	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Constant	0.235	7.98	0.240	8.91	0.497	10.31	0.524	10.82
Population (100 000)	-0.274	-1.37	-0.141	-2.09	-0.071	-0.55	-0.025	-0.34
Distance (100 km)	0.024	4.43	0.026	4.58				
Accessibility index					-0.0023	-4.18	-0.0024	-4.24
University	-0.028	-0.89	-0.048	-1.75				
Number of students					-0.158	-0.64	-0.285	-1.18
Concentration of structure	-0.643	-3.55	-0.819	-4.92	-0.711	-3.73	-0.896	-5.05
N	83		83		83		83	
Left-censored observations	7		3		7		3	
Pseudo R^2	-0.297		-0.351		-0.286		-0.349	

* ll = lower limit of censoring the dependent variable.

7. Conclusions

In this paper efficiency differences of the private sector between 83 Finnish regions in 1988-1999 were investigated by using Data Envelopment Analysis (DEA) and Tobit as well as logistic regression models. Regional efficiency scores were first estimated with five different DEA models, ranging from a basic one output – two inputs case to a model with two outputs and four inputs. Outputs included regional value added and real personal income from employment, inputs covered capital stock, employment by education level, years of schooling and volume of local public sector activity. Regional efficiency scores were obtained as averages of the five models.

According to the DEA estimates regional differences in efficiency proved to be considerable, and efficiency scores correlated with several regional background factors. All the ten biggest regions by population ranked above median for 1988-99; also most university cities fared fairly well in the comparison. Several strongly specialised small regions rated near the top. As to accessibility, the most efficient areas were usually in the southern and western part of Finland, while the regions in the peripheral northern provinces of the country predominantly ranged among the weakest quarter. The efficiency scores were also positively correlated with the regions' employment growth and net migration rates. Differences in efficiency between regions have increased during the period 1988-1999.

In the second part of the study two types of econometric models were introduced in order to explain the (in)efficiency differences among regions, namely Tobit and logistic regression models. Both methods were used to explain the average efficiency scores of the whole period 1988-1999. As none of the regions was fully efficient every year, the minimum value of inefficiency was not zero but above it for the whole period 1988-1999. Secondly, efficiency differences at three points of time were explained: in boom (1988), bust (1993) and a new boom (1999); here only the Tobit model was applied. In the Tobit models inefficiency scores were censored to zero whenever they were less than or equal to 0.1.

The following explanatory variables were used in Tobit and logistic regression models. Population of the region was aimed to catch agglomeration effects. As a measure of concentration of economic activity, we used regional Herfindahl index measure. Domestic economic accessibility was measured inversely by a distance variable and alternatively by a regional accessibility index. University was a dummy explanatory variable if there was at least one in the region, with the number of students in the region as an alternative. Finally, a regional index of innovativity and the number of patents by region were used in some models.

For the whole period 1988-1999 regions with big population were significantly more efficient in three cases out of four. Remote areas were more inefficient as inefficiency increases with distance, a result confirmed by the summary index of regional accessibility. Also, the more concentrated (specialised) was the structure of the regional economy, the higher efficiency. The existence of a university and the number of students in the region got expected negative signs, but only one case was significant.

For the separate years 1988, 1993 and 1999 population lost its significance in all cases but one. Concentration of regional economy decreased inefficiency and the coefficient was significant except in 1988. Distance and accessibility gave expected signs and were significant. University dummy got expected negative coefficients but was never significant, while inefficiency decreased with the number of students in all cases and the effect is significant except in 1999. When the distance variable and the university dummy were replaced by accessibility index and the number of students for the separate years

1988, 1993 and 1999, the coefficients of population became insignificant and changed sign in two cases, perhaps due to correlation between the variables. According to the results both an increase in the number of patents and innovativity index decreased inefficiency for 1999, both coefficients being highly significant.

Lowering the censoring limit in the Tobit models for 1999 improved the fit, while the signs of coefficients of the variables remained same and their t-values increased somewhat.

References

Banker, R.D. - Charnes, A. - Cooper, W.W. (1984): Some models of estimating technical and scale inefficiencies in data envelopment analysis. Management Science, Vol. 9, No. 9, 1078-1092.

Bernard, J. - Cantner, U. (1997): French Regional Performance and Variety. A Non-Parametric Frontier Approach. Paper Presented at the 37th Congress of the European Regional Science Association, Rome.

Chang, P. - Hwang, S. - Cheng, W.(1995): Using Data Envelopment Analysis to Measure the Achievement and Change of Regional Development in Taiwan. Journal of Environmental Management, Vol. 43, 49-66.

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

Charnes, A. - Cooper, W. - Li, S.(1989): Using Data Envelopment Analysis to Evaluate Efficiency in the Economic Performance of Chinese Cities. Socio-Economic Planning Sciences, Vol. 23, No. 6, 325-344.

Farrell, M.J. (1957): Measurement of Productive Efficiency. Journal of the Royal Statistical Society, Series A, 253-281.

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

Huovari, J. - Kangasharju, A. - Alanen, A. (2001): Alueiden kilpailukyky. Pellervon taloudellisen tutkimuslaitoksen raportteja n:o 176, Helsinki (in Finnish, an abridged English version available).

Maddala, G.S. (1983): Limited-Dependent and Qualitative Variables in Econometrics. Econometric Society Monographs No. 3. Cambridge University Press.

Mao, W. - Koo, W. (1997): Productivity Growth, Technological Progress, and Efficiency Change in Chinese Agriculture after Rural Economic Reforms: A DEA Approach. China Economic Review, Vol. 8. No. 2, 157-174.

Maudos, J. – Pastor, J. – Serrano, L. (2000): Efficiency and Productive Specialisation: An Application to the Spanish Regions. Regional Studies vol. 34, 9, 829-842.

Millan, J. - Aldaz, N. (1998): Agricultural Productivity of the Spanish Regions: A Non-Parametric Malmquist Analysis. Applied Economics, Vol. 30, 875-884.

Susiluoto, I. - Loikkanen H. (2001 a): The Economic Efficiency of Finnish Regions 1988-1999, an Application of the DEA Method. Paper presented at the 41st Congress of the European Regional Science Association, Zagreb, Croatia, 29.8. - 1.9.2001

Susiluoto, I. - Loikkanen H. (2001 b): Seutukuntien taloudellinen tehokkuus. Helsingin kaupungin tietokeskuksen tutkimuksia 9/2001, Helsinki (in Finnish, English summary).

Tong, C. (1996): Industrial Production Efficiency and its Spatial Disparity among the TVEs of China: A DEA Analysis. Singapore Economic Review, Vol. 41, No. 1, 85-101.

Tong, C.(1997): China's Spatial Disparity within the Context of Industrial Production Efficiency: A Macro Study By the Data Envelopment Analysis (DEA) System. Asian Economic Journal, Vol. 11, No. 21, 207-217.

Weaver, R.(1984): Multiple Input, Multiple Output Production Choices and Technology in the U.S. Wheat Region. American Journal of Agricultural Economics, Vol 65, 45-56.