Productivity of Norwegian Institutions of Higher Education 2004 – 2013*

by

Dag Fjeld Edvardsen,

Catenda AS

Finn R. Førsund♣,

Department of Economics. University of Oslo and

the Frisch Centre

Sverre A. C. Kittelsen,

The Frisch Centre

Abstract: Studies of productivity of institutions of higher institutions is of interest for two main reasons; education is an important factor of productivity growth for the economy, and in countries where higher education is funded by the public sector the effectiveness of spending the resources accountability is of key interest. A bootstrapped Malmquist productivity index is used to calculate productivity development for Norwegian institutions of higher education over the 10 year period 2004-2013. The confidence intervals got from the bootstrapping allows part of the uncertainty of point estimates to be revealed. The main result is that the majority of institutions have had a positive productivity growth over the total period. However, when comparing with growth in labour the impact on productivity vary quite a lot. Looking for optimal scale of institutions gives no clear conclusion as to any advantage of increasing the size except for quite small institutions.

Keywords: Institutions of higher education; Farrell efficiency measures; Malmquist productivity index; Bootstrapping

---

* The paper is based on a project for the Norwegian Ministry of Knowledge and Education (KD) at the Frisch Centre reported in Edvardsen et al (2014).

♣ Corresponding author

1. Introduction

The institutions in the sector of higher education are in many countries not-for-profit institutions. This is the case for Norway where institutions having the lion's share of students are state institutions providing educational services free of charge. Also many of the private institutions do not charge fees, and get support from the state. The fact that services are not sold on markets to prices reflecting marginal costs immediately points to the difficulty of assessing if the resources consumed in such activities are used efficiently. There is no automatic check of social revenues against costs in the accounts, only budget against expenditure.

One purpose of conducting a productivity study of the sector of higher education is to get information about the results for the considerable resources consumed out of public funds. One way of creating accountability is to conduct studies of productivity. The development of productivity will indicate if ongoing refocussing of objectives and improving efficiency may yield productivity gains.

A natural starting point for economic studies of the higher education sector is to use a production function approach; that is, identifying resources that are transformed into various service outputs that must be quantified. This is unavoidable if any insights into productivity development are to be obtained. This will be the approach of the present study. As tools for estimation we will use non-parametric techniques developed over the last decennia to analyse efficiency and productivity. Most of the studies of higher education focus on efficiency for units within institutions of higher education (see e.g. Worthington (2001) for a review). There are also some studies of productivity at the level of institutions that will be the theme of the present study (Førsund and Kalhagen 1999; Flegg et al 2004; Johnes 2005; Worthington et al 2008).

The plan of the paper is to present the methods in Section 2, and introduce the data in Section 3. Then the productivity analyses follow in Section 4 and Section 5 concludes.

2. Methods

Caves et al. (1982) introduced the bilateral Malmquist productivity index developed for discrete time based on the ratio of distance functions measures for two units (e.g. the same unit measured for two different time periods) relative to the same frontier production function.

A strength of the Malmquist productivity index is the possibility of calculating the productivity development of each unit in the data set. However, in many empirical applications of the index this possibility is under-utilised, focussing more on giving an aggregate picture over time or across units, or both (Färe et al, 2008). In this study efforts will be made to present results for individual units in ways more satisfactory in order to appreciate the results. However, overall impressions will also be given, but based on a unit-based approach of utilising a constructed average unit. How to set up LP problems to estimate the distance functions is so well known in the literature that we do not find it necessary to do it here, see e.g. Cooper et al (2000).

When applying the Malmquist productivity index attention should be paid to desirable properties (Färe et al, 2008). In the literature this is more often than not glossed over. We will therefore explain in more detail the choice of our specification. Productivity as measured by the Malmquist index may be influenced by changes in the scale of the operation, but two units that have the same ratio of outputs to inputs should be viewed as equally productive, regardless of the scale of production (Grifell-Tatjé and Lovell, 1995). Doubling all inputs and outputs keeping input and output mixes constant should not change productivity. Therefore the benchmark envelopment of data if we want to measure total factor productivity (TFP) is one that is homogenous of degree 1 in the input-output vector, and thus the linear-homogenous set that fits closest to the technology. The homogenous envelopment can be used to define the concept of technically optimal scale (Frisch 1965), termed TOPS in Førsund and Hjalmarsson (2004a, b). This is the scale where the elasticity of scale is 1, and is illustrated in Figure 1 at the point P_v^{tops} for a variable-returns-to-scale frontier (VRS). From classical production theory we know that the productivity is maximal at optimal scale where returns to scale is one, thus this is a natural reference for productivity changes over time. Observations of the same unit for the two periods *u* and *v* are indicated by P_u and P_v. The two corresponding VRS frontiers are drawn showing an outward shift indicating technological progress. The TOPS point for period *v* is labelled P_v^{tops}. Just as the productivity should be unchanged if the input-output vector is proportionally scaled, a measure of productivity should double if outputs are doubled and inputs are kept constant, and increase by half if inputs double, but outputs are constant. The desirable homogeneity properties of a TFP index is therefore to be homogenous of degree 1 in outputs in the second period and of degree (-1) in inputs of the second period, and homogenous of degree (-1) in outputs of the first period and homogenous of degree 1 in inputs of the first period. Using CRS to envelope the data is thus one way of

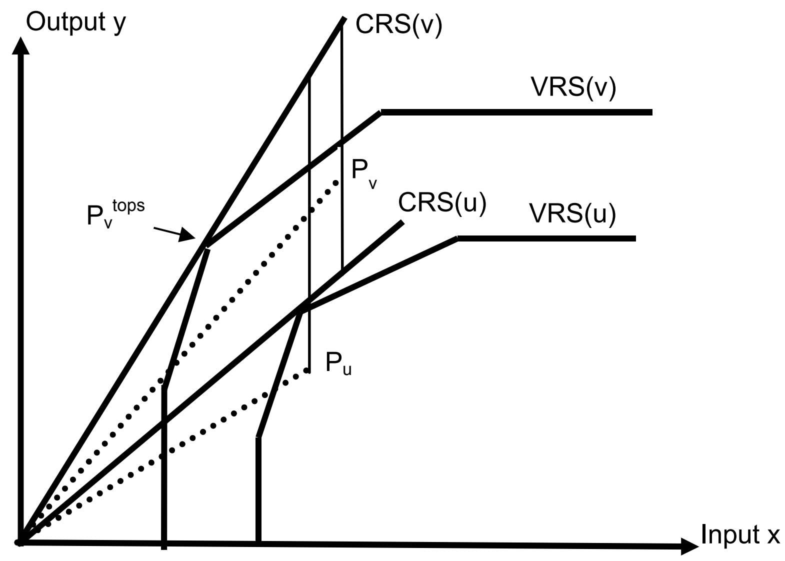


Figure 1. The Malmquist productivity index.
Productivity change for a unit measured relative to the benchmark CRS technology with maximal productivity.

obtaining all the required homogeneity properties of a Malmquist productivity index.

Another property of a productivity index that is important is the *circularity* of the index (Berg et al 1992) (see Gini (1931) for an interesting exposition). The implied transitivity of the index means that the productivity change between two non-adjacent periods can be found by multiplying all the pairwise productivity changes of adjacent periods between the two periods in question. We will transtivitise the Malmquist index by using a single reference frontier enveloping the pooled data. In Tulkens and van den Eeckaut (1995) this type of frontier was termed the *intertemporal frontier*. (In Pastor and Lovell (2005), missing out on this reference, it is called the global frontier.)

Using the same CRS reference frontier for all units means that we have made sure that efficiency for all units and time periods refer to the same frontier. Specifying CRS only is not sufficient to ensure that a specific data point occurring at different time periods get the same efficiency evaluation because both input- and output isoquants may differ in shape over time if the technology is allowed to change over time as in Färe et al. (2008). Using a linear homogeneous envelopment implies that the orientation of the distance function does not matter. The estimator of the Malmquist index for a unit *i*, using the Farrell efficiency indices that correspond to the distance functions, for the two periods relative to the same frontier is

$$\hat{M}_i^s(u,v) = \frac{\hat{E}^s(x_{iv}, y_{iv} \mid \hat{\hat{S}}^s)}{\hat{E}^s(x_{iu}, y_{iu} \mid \hat{\hat{S}}^s)} \quad , \quad i = 1,..,J,\ \ u,v = 1,..,T,\ u < v \tag{1}$$

where superscript *s* symbolises that all data are used as the technology reference set. The Malmquist productivity estimator is conditional on the estimator $\hat{\hat{S}}^s$ of a linear homogeneous envelopment set. The efficiency scores $\hat{E}^s$ are calculated for period *u* and *v* respectively for each unit *i*. The efficiency measures in (1) are the Farrell technical productivity measures (Førsund and Hjalmarsson 1979; Førsund et al 2006) and the productivity change is the change in the productivities of the observations relative to the benchmark maximal productivity.

The Malmquist index can be multiplicatively decomposed into an efficiency catching-up term *MC*, and a term capturing the shift of the frontier, *MF* if we assume that the period frontier changes over time (as the CRS frontiers in Figure 1) and keeping circularity:

$$\hat{M}_{ivu}^s = \frac{\hat{E}_v^s(x_{iv}, y_{iv})}{\hat{E}_u^s(x_{iu}, y_{iu})}\Bigg|\hat{S}^s = \frac{\hat{E}_{iv}^v}{\hat{E}_{iu}^u} \times \frac{\hat{E}_{iv}^s / \hat{E}_{iv}^v}{\hat{E}_{iu}^s / \hat{E}_{iu}^u} = \widehat{MC} \times \widehat{MF}, \quad i = 1,..,J,\ u,v = 1,..,T, u < v \tag{2}$$

Here superscripts *v* and *u* indicates that the technologies for the periods are different, while *s* stands for the benchmark technology based on the pooled dataset. The MC-measure shows how a unit is catching-up with the frontier. The *MF* measure of technology shift is calculated as a ‘double’ relative measure where both period efficiency measures are relative to the benchmark measure. There are two ways productivity can change over time; change in efficiency and shift in technology (Nishimizhu and Page 1982). Note that the decomposition does not mean that there is a causation; we cannot distinguish between productivity change due to increase in efficiency and due to shift in technology using the components in (2), as often believed in the literature (Johnes 2008; Worthington et al 2008). The MF-measure represents the relative gap between technologies and is thus the potential maximal contribution to productivity change, while the MC-measure is not the efficiency contribution to productivity change, but illustrates the actual catching-up that is also influenced by the technology shift.

Bootstrapping

We are following the bootstrap procedure outlined in Simar and Wilson (1998; 1999; 2000). Testing the period frontier function (Banker 1993) VRS turned out to be preferred. Choosing the Farrell output-oriented efficiency variable, distributed on (0,1], our resampling (Efron 1979) creating pseudo replicate data sets, is done on the basis of the calculation of output-oriented efficiency scores E_2 relative to the VRS frontier for each time period

$$y_{imt}^{ps} = \frac{y_{imt}}{\hat{E}_{2it}^{s}} E_{2t}^{KDE}, i = 1,..,J\ ,\ m = 1,...,M\ ,\ t = 1,...,T \tag{3}$$

where E_{2t}^{KDE} is a draw of the kernel density distribution estimated for the efficiency score. This distribution is used to smooth the empirical distribution of the original efficiency scores, using reflection (Silverman 1986), in order to avoid the accumulation of efficiency score values of 1.

A new DEA frontier is then estimated on these pseudo observations (x_i, y_i^{ps}). We make 2000 such draws and establish 2000 new DEA frontiers. Now, going back to each run for a pair of periods, the Malmquist productivity index, given by (1), is calculated using the linear homogeneous technology created for the pooled set of all pseudo observations as the benchmark.

Assuming estimators to be consistent, Simar and Wilson (1999) show that the bias can be estimated based on the relationship

$$(\tilde{M}^s(u,v) - \hat{M}^s(u,v)) \mid \hat{S}^s \sim (\hat{M}^s(u,v) - M^s(u,v)) \mid S^s,\ \ u,v = 1,..,T, u \neq v \tag{4}$$

Here M^s is the true unknown productivity, $\hat{M}^s$ is the original DEA estimate, $\tilde{M}^s$ is the bootstrapped estimate and S^s and $\hat{S}^s$ are the theoretical production possibility set and its DEA estimate, respectively.

However, it is pointed out in Simar and Wilson (2000) that the bias correction may create additional noise in the sense that the mean square error of the bias-corrected score may be greater than the mean square error of the uncorrected estimator. This turned out to be the case here. Therefore the point estimates of our Malmquist indices are based on the 'first round' of estimating the index. However, the confidence intervals are based on the estimates of the biases (Simar and Wilson, 1999).

2. Data

There are six variables in all used in the analysis; two inputs and four outputs set out in Table 1. The data are taken from the central register of data for institutions of higher education in

Table 1. Input- and output variables in the study

Inputs per year	**Outputs per year**
Faculty employees in man-years	Study points for courses of a lower degree (cost weighted)
Administration and other employees in man-years (excluding cleaning)	Study points for courses of a higher degree (cost weighted)
	Publishing points
	Ph.D.'s per year

Norway (DBH). Due to the degrees of freedom enforcing a parsimonious model we have restricted the variables to the key ones, excluding capital like buildings and equipment, and other expenses on the input side. As to quality variables for inputs there are quality of faculty measured by position and experience, and average grade at start of studies, and for outputs average grade level for degrees (Bachelor and Master), job success, citation index and impact of research, and external research funding. However, we have not included such quality variables, partly due to the fact that this information is not so readily available.

As to other employees than faculty cleaning is excluded because the institutions have different practices of outsourcing this activity or doing it in-house. A problem on the output side regarding study points is that the analysis had to be done at an aggregate level for each institution. But different types of studies require different resources of faculty and laboratory costs. We compensate for this by weighting the study points by cost weights based on yearly contributions per student from the state. A problem with Ph.D.'s as outputs is that there are several years (on average four) of use of resources on Ph.D. students before they obtain the degree.

The total number of units in the DBH database is 75. However, some units have missing data and extreme outliers are excluded based on the super-efficiency score (Andersen and Petersen, 1993) and the importance of the unit as a referent unit (Torgersen et al, 1996). We end up using 63 units in 2004 and 59 in 2013. There remain 44 units constituting a panel for the total period. The estimation of the reference CRs frontier is based on about 500 observations.

The development of our variables for the study period is set out in Figure 2 on index form with the values in 2004 as the base. The two outputs publishing points and Ph.D.'s have had

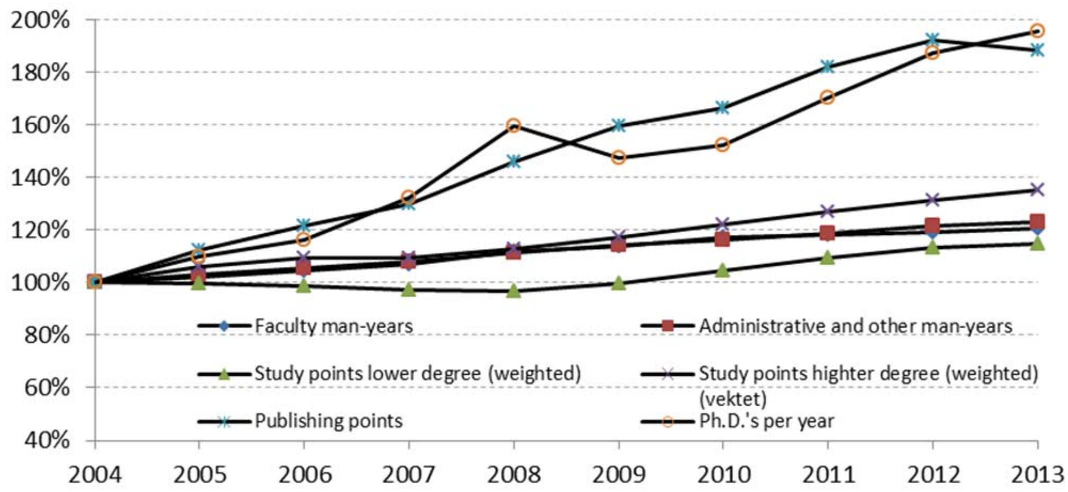


Figure 2. Development of the variables for the periods 2004 to 2013 relative to 2004

the most rapid growth with 88 and 96 % respectively. Of the two other outputs, weighted study points, the lower points have been growing most slowly with 15 % while the higher point have increased with 35 % . The two inputs have developed rather parallel with faculty increasing 21 % and administration and other 23 %. Partial reasoning indicates that there has been an aggregate productivity growth for the total period.

4. The productivity development

Aggregate development

We will use two variants of a bottom - up approach. One approach, linked to Farrell's way of measuring how the mean performance of a sector is compared with the frontier, is to form an average unit by averaging inputs and outputs and then enter this unit as a micro unit in the calculations (Førsund and Hjalmarsson 1979). Another more conventional approach is to take some mean, here a simple arithmetic one, of the individual results. Both approaches are illustrated in Figure 3. The difference in aggregate growth is moderate except for the growth from 2007 to 2008 with a positive jump in the productivity growth measured by the average unit and a negative growth for the average of productivities growth measure, and a similar development in the last period. This difference may be due to small units having a weaker productivity development than larger units. Inspecting the confidence intervals it is only for the same two periods that there is a significant difference between the two measures showing a higher productivity change by the average unit measure.

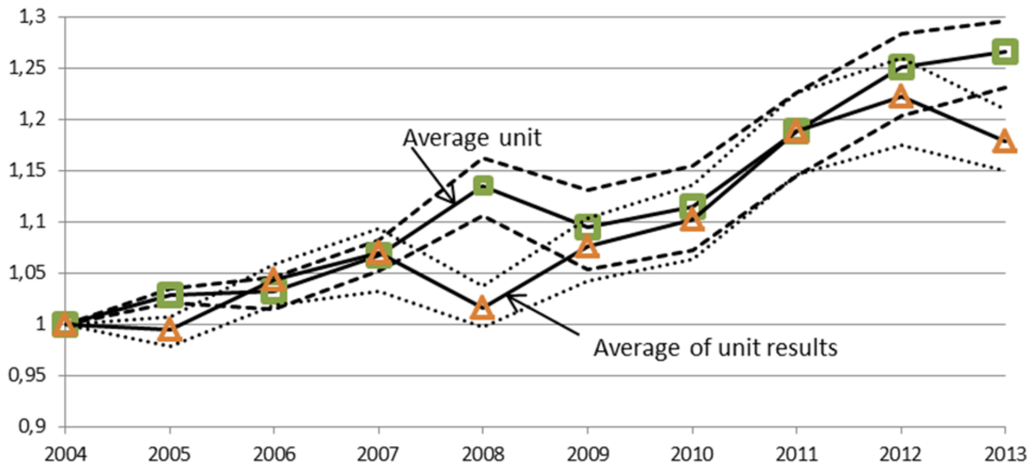


Figure 3. Aggregate productivity change (solid lines) for the periods 2004 to 2013 relative to productivity in 2004 measured by the average unit and average of individual productivities with 95 % confidence intervals (broken lines).

We have decomposed the productivity change measure into catching-up (MC) and frontier shift (MF) according to Eq. (2). The development is shown in Figure 4. We see

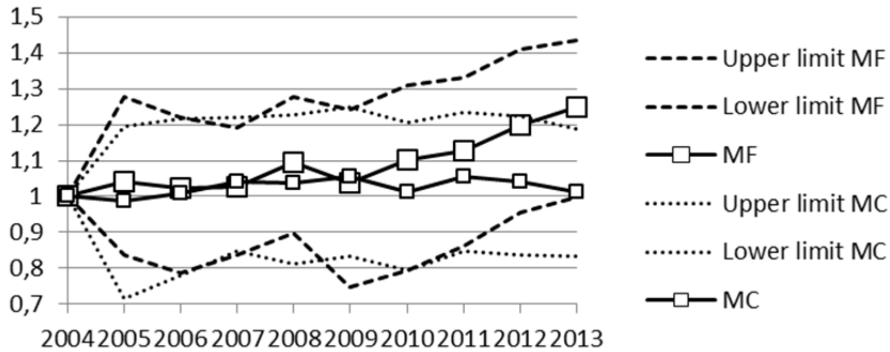


Figure 4. The decomposition of the Malmquist productivity index for the average unit into catching-up MC and frontier shift MF

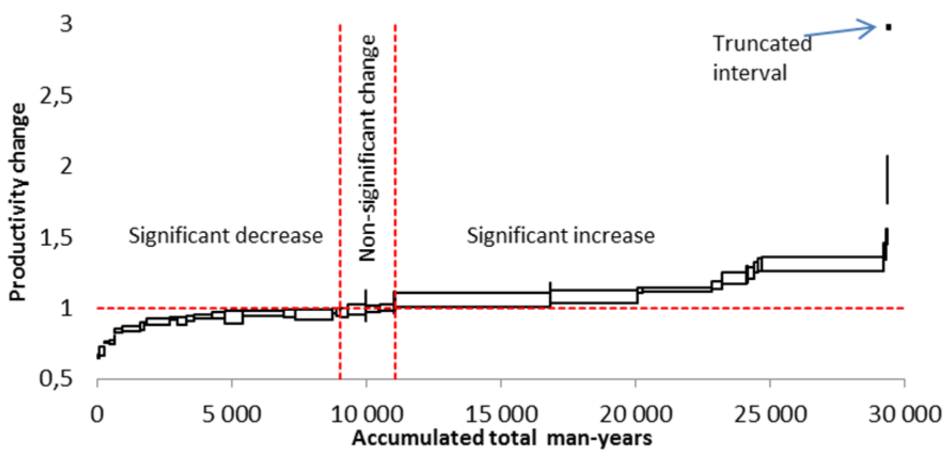
that MC and MF moves more or less parallel until 2009, but for the rest of the periods the MF measure grows markedly while the MC measure stagnates and even goes down. Thus, the point estimates indicate almost no change in the sector's ability to keep up with the own period best productivity, but that the productivity on the frontier has improved over time. However, we see from the confidence intervals that the differences are not significant (as also experienced in Edvardsen et al, 2006), but almost so for the last period.

Productivity development of individual units

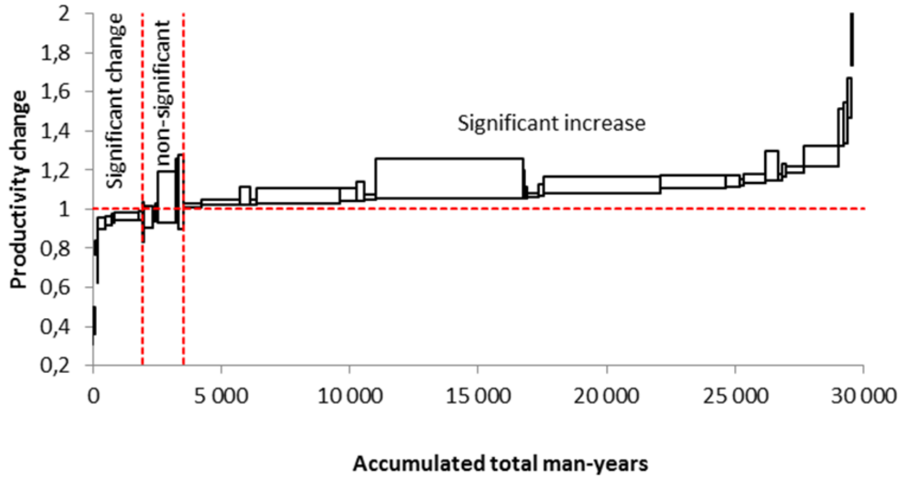
Due to bootstrapping it is now possible to assess the extent of uncertainty of the point estimates of productivity numbers represented by the bias of observing a limited sample.

(Even if all existing units are observed we may still speak of a sampling bias because the units could have had different levels of input- and output variables.) A new type of diagram is developed that allows the individual productivity results to be displayed together with the extent of uncertainty in the form of confidence intervals. The results are arranged in a way that directly facilitates a visual test of a unit's productivity performance at the same time as the information about size distribution is maintained.

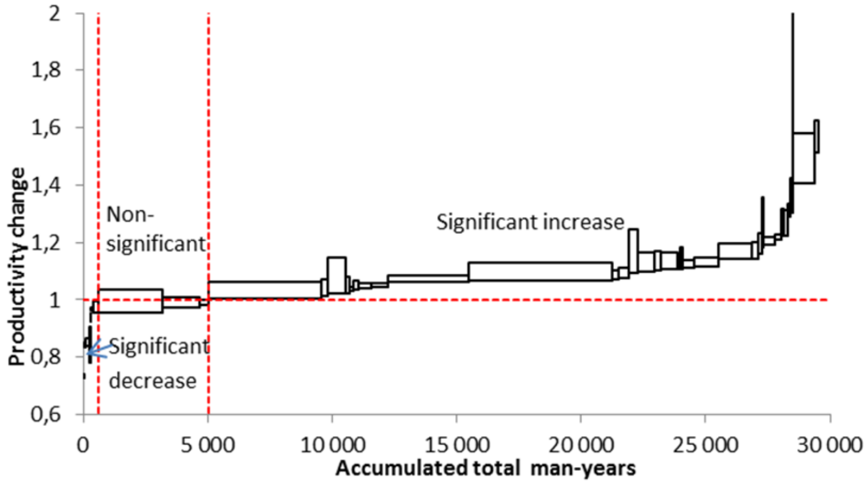
In Figure 5 four panels of productivity-change distribution are set out for three year periods, and the total period 2004 – 2013. Each unit is represented by a box. The width of a box is based on the relative share of total man-years as an average for all years for ease of identifying the units over the periods. The height of the box shows the width of the 95 % confidence interval. A unit may be in three states; exhibiting significant productivity decline, non-significant change, or significant growth. The position of a box for a unit relative to the crucial value of 1 signifies negative, positive productivity change, or no change. By sorting the units, starting from the left with units with significant decrease in productivity, then units with insignificant productivity change, and lastly units with significant increase, we get an immediate picture of the productivity change situation. The share of labour by units in each group can also be seen. The groups are delimited by the two broken vertical lines. In the first group the units are sorted according to ascending values of the upper limit of the confidence interval, thus securing that all units in the group have negative estimates of productivity change and the upper limits of confidence intervals below the value of 1. The second group is found by sorting both according to the upper and the lower limit of the confidence intervals



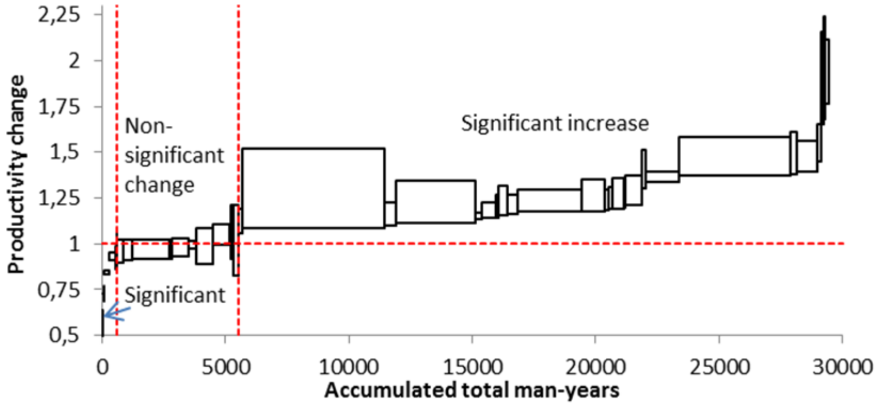
Panel (a) 2004-2007



Panel (b) 2007-2010



Panel (c) 2010-2013



Panel (d) 2004-2013

Figure 5. Productivity change for units sorted according to confidence status.
Width of boxes for confidence intervals proportional to average total man-years.

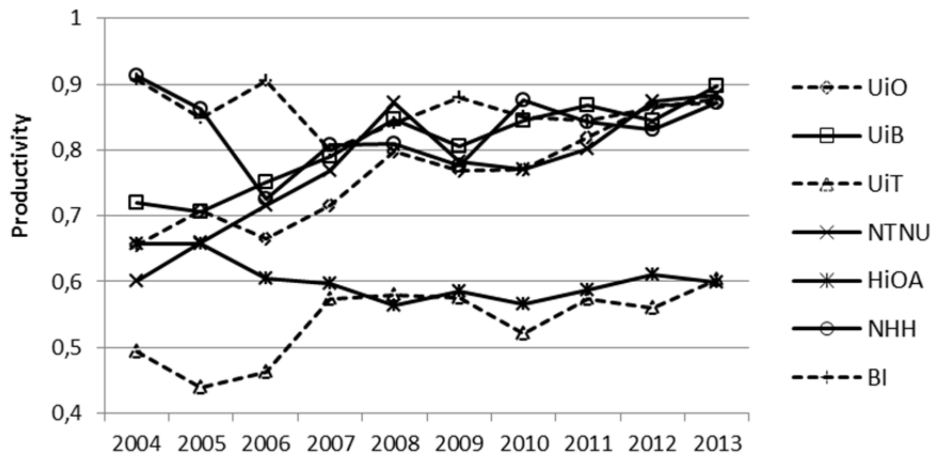
identifying the units securing that all units in the group have estimates of productivity change not significantly different from 1. The units are sorted according to ascending values of the median productivity change. In the third group the units are sorted according to ascending values of the lower limit of the confidence interval, thus securing that all units in the group have estimates of productivity change and the lower limits of confidence intervals above the value of 1, signalling significant productivity growth.

We will comment on some general features of the productivity distribution diagrams and leave details to the reader. The productivity numbers are on the whole relatively sharply determined; the confidence intervals are rather narrow. Large units tend in general to have wider confidence intervals than small and medium-sized units. The difference is not so pronounced in the first two panels, but is clearly seen in Panel (c) for 2010-2013 and especially Panel (d) for the total period. But a few quite small units have rather wide confidence intervals for the first two panels. The number of units with significant productivity decline is quite small for the panels except for the first period in Panel (a). A main result is that the share of man-years with significant productivity growth is considerably larger than for the other two groups, varying from 62 % for Panel (a) to 83 % and 88 % for the next two panels, and to 81 %, corresponding to 29 of the 44 units, for Panel (d) for the whole period.

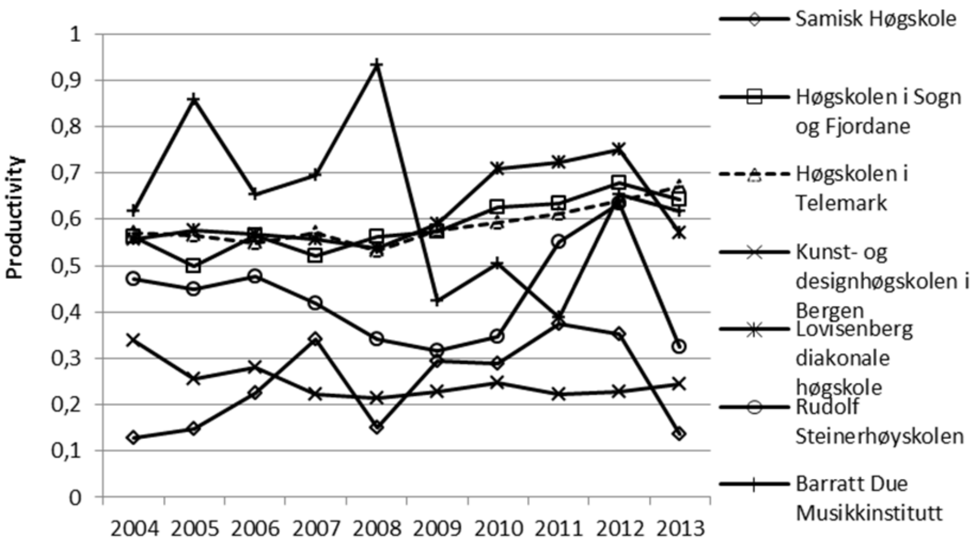
Productivity over time for selected units

We will select some large and small units to follow more closely over time. The two panels of Figure 6 show the level productivity developments year by year for a selection of large and a selection of small units. The four largest units are represented by the universities of Oslo (UiO), Bergen (UiB), the technical university (NTNU) and the university of Tromsø (UiT).The two largest business schools are represented (NHH and BI) and the largest college (HiOA).

The Malmquist productivity index is the ratio of consecutive values of the value of the level of productivity. This means that if productivity has gone down from e.g. 2004 to 2005, as is the case for the two business schools, this means the productivity change is negative and the Malmquist index for 2005 is less than 1. In fact, in Panel (a) we see that all units except two has productivity decline from 2004 to 2005. After that the productivity development of the units differ somewhat. UiT with the lowest productivity level of all in 2004, then from 2005 increase the productivity level until 2008, and then go up and down finishing in the last period



Panel (a). Large units



Panel (b). Small units

Figure 6. Development of level of productivity for selected large and small units

with its highest level of productivity. This means that over the period as a whole this university will come out with a positive productivity growth that is also significant. The productivity of the largest college HiOA is falling from 2004 to 2008 to the level of UiT and then evening out ending up with a non-significant negative change for the whole period.

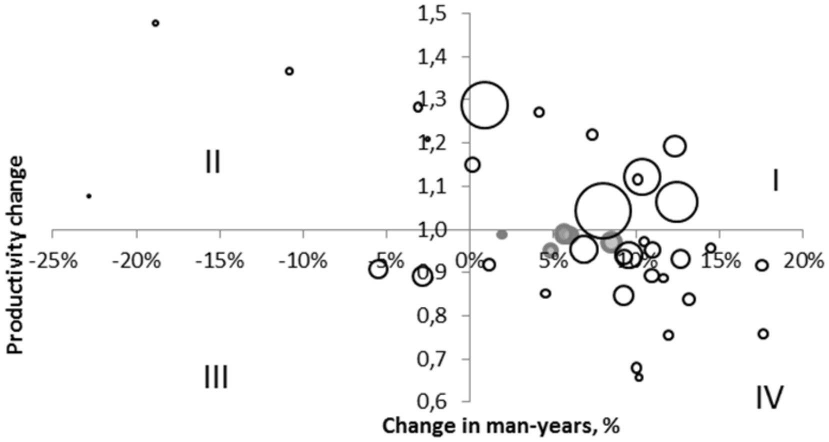
A striking trend of the development of the other units is that there is some turbulence in productivity up to 2008, but then the developments become more like and all units end up with about the same level of productivity close to 90 % implying a positive productivity growth for the universities, but an insignificant change for the two business schools because these start out with levels, considerably higher level of productivity than the universities, but also slightly higher than the end levels.The main purpose of showing the small units in Panel (b) is to illustrate the rather erratic performance regarding their productivity levels. This results in a similar erratic behavior of their productivity change, as observed in the cross-section panels of Figure 5. The most stable positive developments are shown by the two general colleges, while the colleges catering for special interests like arts, the Sami population, music, religious-based nursing and agriculture and village development, have the erratic developments. This can be attributed to the small scale of the institutions and the consequences of otherwise small absolute changes in man-years and study points.

Productivity and resource use

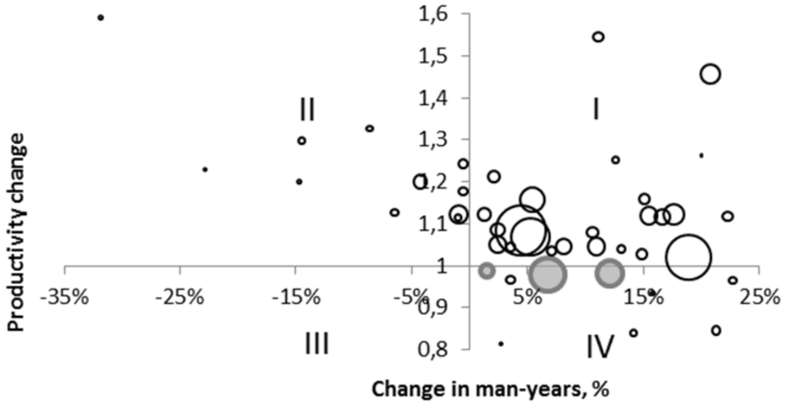
A recurrent policy question is the return on the resources allocated to higher education. Showing the change in total labour used and together with productivity change provides some answers (Førsund et al 2006). In Figure 7 productivity change for the same periods as for Fig. 5 is shown together with the relative change in total man-years. The area of a circle is proportional to the average level of man-years as also used as the size variable in Figure 5. The open circles are the units with significant productivity change (either negative or positive), while the circles with grey fill are units with insignificant change. The midpoints of the circle correspond to the medium of the productivity changes within the confidence intervals. The horizontal axis measures change in man-years. The vertical axis measuring productivity change is placed at zero change of labour use. To the left of the origin labour has decreased while to the right labour have increased.

The horizontal line at the value 1 delimitates the units with productivity decrease and increase, respectively, and the vertical axis from zero change in labour form four quadrants numbered I to IV. In Quadrant I units have had both productivity growth and increase in man-years. Such units may be said to have experienced *efficient labour expansion*. The units in Quadrant II have also had productivity growth, but experienced labour reductions. This may be termed *efficient labour saving*. In quadrant III productivity decrease is combined with labour decrease. This is *inefficient labour saving*. Units in Quadrant IV have the worst of both worlds with decreasing productivity and increasing labour. This is *inefficient labour expansion*.

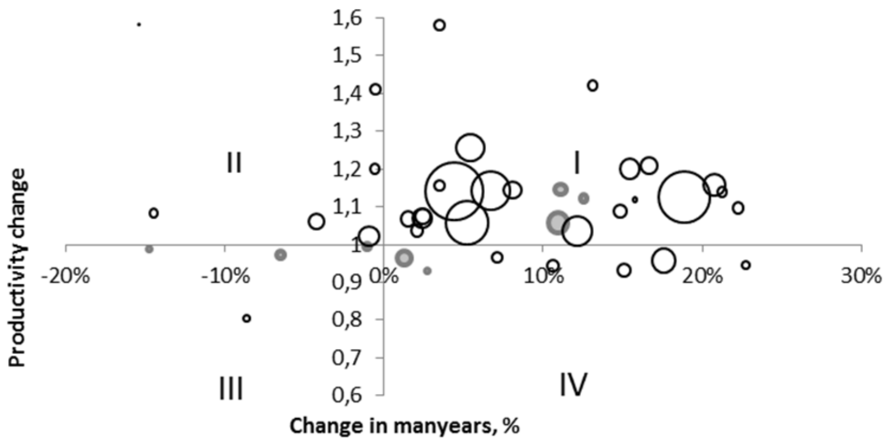
Due to a steady increase in labour for almost all units there are not many units in Quadrants II and III so Quadrants I and IV are the informative ones. (A few units with extreme changes



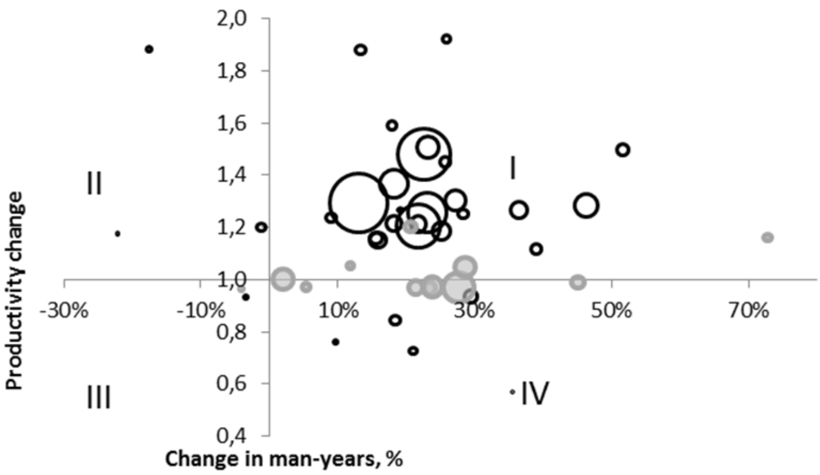
Panel (a) 2004-2007



Panel (b) 2007-2010



Panel (c) 2010-2013



Panel (d) 2004-2013

Figure 7. Change in productivity and man-years
The circles are proportional to size measures by average man-years 2004-2013.
Open circles represent units with significant change in productivity,
Filled circles represent units with non-significant change in productivity.

have been removed in order to keep the diagrams visually interpretable.) A general feature for all periods is that the large units from Figure 6 are in Quadrant I with efficient expansion of labour.

The total period in Panel (d) show quite a variety in the labour increase without a clear positive correlation with productivity change. The increase in labour range from 13 % for UiO, resulting in productivity growth of 29 %, to 23 % for NTNU resulting in the highest productivity growth of the large universities of 48 %. Note that the unit having the highest growth in labour of 73 % has an insignificant productivity change. However, this is the special purpose Sami College seen in Panel (b) in Figure 5 starting up with the lowest productivity in 2004 just above 10 % and ending not much higher in 2013 after up and down development of productivity. The two business schools (shown in Panel (a) of Figure 5) both have insignificant productivity growth, but while the private school BI has had a 2 % growth in labour the public school NHH has had 24 %.

Scale

The question of optimal scale is always interesting from a policy point of view, and especially when the public sector is providing almost all financing. However, optimal scale cannot be found by calculating productivity change using the Malmquist productivity index because CRS has to be specified. We must have VRS to be able to estimate an optimal scale, see

Figure 1 and TOPS points on the VRS period frontiers. Scale efficiency is measured as the ratio between the productivity of an observation corrected for technical inefficiency and maximal productivity on the VRS frontier (Førsund and Hjalmarsson 1979). Calculation of input-oriented scale efficiency yielded an average of 90 % showing units on the average being close to optimal scale when technical inefficiency is eliminated.
The productivity measures in the Malmquist index (1) give us information how close observed productivities are to the CRS benchmark productivity. Panel (a) in Figure 6 shows that the large units have levels around 90 % the last year, while Panel (b) shows that the small units have considerably lower productivity down to 10 % and fluctuating a lot. This indicates that the small units are too small, but we cannot say whether the large units are also too small or are too large.

To get another angle on the optimal scale issue we have calculated a VRS frontier for the pooled dataset (Johnes 2014) and then made a two-dimensional figure of the frontier by cutting the six-dimensional frontier with a plane through the origin and the average unit. By bootstrapping the confidence intervals are also calculated. Figure 8 show the results. A CRS

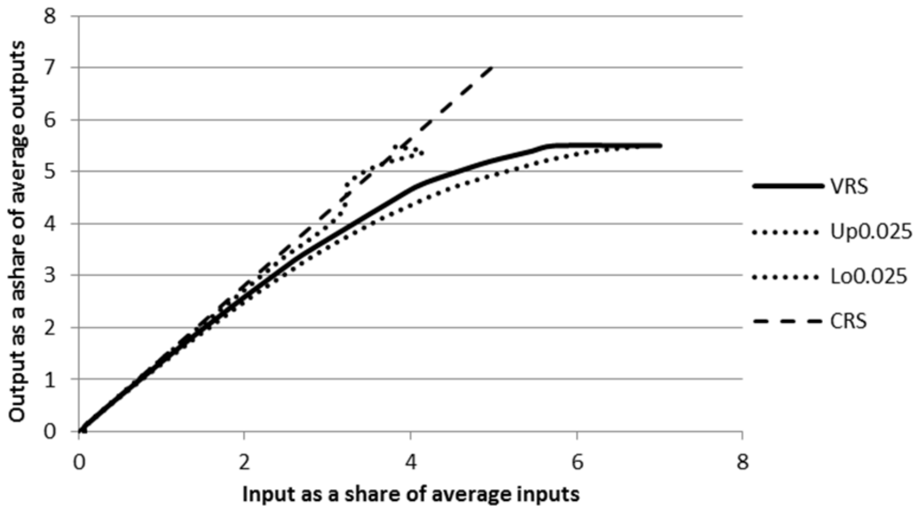


Figure 8. Cut of the frontier (solid curve) through the average unit with confidence intervals (broken curves).

frontier is also provided as a benchmark. Optimal scale is estimated to be about ¼ of the average unit, but this result is rather uncertain and the uncertainty increase with size. The confidence intervals show that we cannot exclude that the CRS and VRS frontier coincide over wide intervals so we cannot exclude that the optimal scale is five times the average unit. We may have a wide interval with optimal scales.

We have to recognize that optimal scale depends on the input mix and in our case especially the output mix (see Førsund and Hjalmarsson 2004a). The shape of the frontier in Figure 8 varies with the mix of variables. Investigating the impact of institutions being intensive in education or research indicate that education-intensive institutions tend to have smaller optimal scale than research-intensive units.

5. Conclusions

Studies of productivity of institutions of higher institutions is of interest for two main reasons; education is an important factor for productivity growth for the economy, and in countries where higher education is funded by the public sector the effectiveness of spending the resources is of key interest as to accountability. This study of Norwegian higher education institutions uses available primary data collected yearly. There is a choice of which variables to use and how many. The number of variables is limited by the number of observations. It turned out to be difficult to get variables covering interesting quality aspects of education, research and resources employed, including the quality of students, so we are left with variables easier to quantify like faculty and other employees for resources, and study points, publication points and Ph.D.'s for education and research, respectively. In order to make study points comparable for institutions having quite different focus of their education the study points are grouped into points for courses taken as part of basic studies (Bachelor) and points for courses within more advanced courses (Master), and then the study points are weighted with the size of contributions to types of courses from the Ministry of education.

As a tool for estimating productivity change for a 10-year period 2004-2013 a Malmquist productivity index is used. This index is based on extended Farrell efficiency measures and calculated employing a non-parametric benchmark using the DEA model. In order to get information about uncertainty a bootstrapping procedure is used for covering uncertainty created by sampling bias.

The results are presented by figures giving both information of productivity development of the units, but also giving visual information about uncertainty represented by confidence intervals. In the diagrams for sub-periods the units are classified in three categories representing significant decline, non-significant change, and significant growth, respectively. For the total period the majority of units have a significant growth. However, there is a considerable gap in size between units, and the small units show erratic results and cannot be taken at face value. Combining the productivity results with the change in labour it turns out that institutions achieve quite varying productivity results for varying resource allocations, but the majority of units can be classified as having efficient resource expansion, that is both productivity growth and increase in labour. All large units belong to this group. But there are also a number of medium-sized units in the group of non-significant productivity change, and several of these have had a considerable increase in resources.

Regarding the question of optimal scale the analysis cannot give clear answers. There are indications of falling returns to scale for the largest education-intensive institutions, but also large intervals with constant returns to scale. However, a hypothesis of increasing returns to scale is not supported by our results. Keeping the output levels and focusing on reducing technical inefficiency may be the best policy option.

There have been some mergers during the period covered but not enough to find any significant difference before or after, but given the yearly production of primary data this question should be studied later (Johnes 2014). Quality variables have not been used in the study. This a priority task for further research. Some types of quality variables are mentioned, but may be more relevant ones should be developed. Further research should also be done on scale, and scope (Bonaccorsi et al 2014).

References

Andersen P and Petersen NC (1993). A procedure for ranking efficient units in Data Envelopment Analysis. *Management Science* **39**: 1261-1264.

Banker RD (1993). Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science* **39**: 1265-1273.

Berg SA, Førsund FR and Jansen ES (1992. Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89. *The Scandinavian Journal of Economics* **94**: Supplement S211-S228.

Bonaccorsi A, Daraio C and Simar L (2014). *Efficiency and economies of scale and scope in European universities. A directional distance approach*. Technical Report 8, 2014, Sapienza Università di Roma, http://www.dis.uniroma1.it/~bibdis/ RePEc/aeg/ report/2014-08.pdf

Caves DW, Christensen LR and Diewert E (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica* **50**(6): 1393-1414.

Cooper WW, Seiford LM and Tone K (2000). *Data Envelopment Analysis: a comprehensive text with models, applications,references and DEA-solver software*. Kluwer Academic Publishers: Dordrecht/Boston/London.

Edvardsen DF, Førsund FR og Kittelsen SAC (2014). *Produktivitetsanalyse av universitets- og høgskolesektoren* [*Productivity analysis of the university and college sector*]. Rapport 3/2014, Ragnar Frisch Centre for Economic Research: Oslo.

Edvardsen DF, Førsund FR, Hansen W, Kittelsen SAC and Neurauter T (2006). Productivity and regulatory reform of Norwegian electricity distribution utilities. In: Coelli T and Lawrence D (eds). *Performance measurement and regulation of network utilities*. Edward Elgar Publishing, pp. 97-131.

Efron B (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**: 1-6.

Farrell MJ (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society* Series A, **120** (III): 253-281.

Flegg AT, Allen DO, Field K and Thurlow TW (2004). Measuring the efficiency of British universities: a multi-period Data Envelopment Analysis. *Education Economics* **12**(3): 231-249.

Frisch R (1965). *Theory of production*. D. Reidel Publishing Company: Dordrecht.

Førsund FR and Hjalmarsson L (1979). Generalised Farrell measures of efficiency: an application to milk processing in Swedish dairy plants. *Economic Journal* **89**: 294-315.

Førsund FR and Hjalmarsson L (2004a). Are all scales optimal in DEA? Theory and empirical evidence. *Journal of Productivity Analysis* **21**(1): 25-48.

Førsund FR and Hjalmarsson L (2004b). Calculating scale elasticity in DEA models. *Journal of the Operational Research Society* **55**: 1023-1038.

Førsund FR and Kalhagen KO (1999). Efficiency and productivity of Norwegian colleges. In: Westermann G (ed.). *Data envelopment analysis in the service sector*. Deutscher Universitäts-Verlag: Wiesbaden, pp. 269-308.

Førsund FR, Kittelsen SAC, Lindseth F and Edvardsen DF (2006). The tax man cometh - but is he efficient? *National Institute Economic Review* **197**(July): 106-119.

Färe R, Grosskopf S and Margaritis D (2008). Efficiency and productivity: Malmquist and more. In: Fried HO, Lovell CAK and Schmidt SS (eds). *The measurement of productive efficiency and productivity growth*, Chapter 5. Oxford University Press: New York, pp. 522-622.

Gini C (1931). On the circular test of index numbers. *Metron* **9**(2): 3-24.

Grifell-Tatjé E and Lovell CAK (1995). A note on the Malmquist productivity index. *Economics Letters* **47**: 169-175.

Johnes J (2008). Efficiency and productivity change in the English higher education sector from 1996/97 to 2004/5. *The Manchester School* **76**(6): 653-674.

Johnes J (2014). Efficiency and mergers in English higher education 1996/97 to 2008/9: parametric and non-parametric estimation of the multi-input multi-output distance function. *The Manchester School* **82**(4): 465-487.

Nishimizu M and Page JM (1982). Total factor productivity growth, technological progress and technical efficiency change: dimensions of productivity change in Yugoslavia 1965-78. *Economic Journal* **92**: 920-936.

Pastor JT and Lovell CAK (2005): A global Malmquist productivity index. *Economics Letters* **88**: 266-271.

Silverman BW (1986). *Density estimation for statistics and data analysis*. Chapman and Hall.

Simar L and Wilson PW (1998). Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* **44**: 49-61.

Simar L and Wilson PW (1999). Estimating and bootstrapping Malmquist indices. *European Journal of Operations Research* **115**(3): 459-471.

Simar L and Wilson PW (2000). Statistical inference in nonparametric frontier models: the state of the art. *Journal of Productivity Analysis* **13**: 49-78.

Torgersen AM, Førsund FR and Kittelsen SAC (1996). Slack adjusted efficiency measures and ranking of efficient units. *Journal of Productivity Analysis* **7**: 379–339.

Tulkens H and van den Eeckaut P (1995). Non-parametric efficiency, progress, and regress measures for panel data: methodological aspects. *European Journal of Operational Research* **80**: 474-499.

Worthington AC (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics* **9**(3): 245-268.

Worthington AC and Lee BL (2008). Efficiency, technology and productivity change in Australian universities, 1998 – 2003. *Economics of Education Review* **27**: 285-298.