MEMORANDUM

No 04/2016

Productivity Development of Norwegian Institutions of Higher Education 2004 – 2013

Dag Fjeld Edvardsen, Finn R. Førsund and Sverre A. C. Kittelsen

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 03/16	Finn R. Førsund *Multi-equation modelling of Desirable and Undesirable Outputs Satisfying the Material Balance*
No 02/16	Ingrid Hjort *Potential Climate Risks in Financial Markets: Report from a workshop, January 20, 2016*
No 01/16	Ingrid Hjort *Potential Climate Risks in Financial Markets: A Literature Overview*
No 22/15	Geir B. Asheim and Frikk Nesje *Destructive intergenerational altruism*
No 21/15	Rolf Golombek, Alfonso A. Irarrazabal, Lin Ma *OPEC's market power: An Empirical Dominant Firm Model for the Oil Market*
No 20/15	Moritz A. Drupp, Mark C. Freeman, Ben Groom and Frikk Nesje *Discounting Disentangled*
No 19/15	Simen Markussen and Marte Strøm *The Effects of Motherhood*
No 18/15	Marit Linnea Gjelsvik, Ragnard Nymoen, and Victoria Sparrman *Have Inflation Targeting and EU Labour Immigration Changed the System of Wage Formation in Norway*
No 17/15	Geir B. Asheim and Ivar Ekeland *Resource Conservation across Generations in a Ramsey-Chichilnisky Model*
No 16/15	Nina Drange, Tarjei Havnes and Astrid M. J. Sandsør *Kindergarten for All: Long-run Effects of a Universal Intervention*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/

Productivity Development 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 growth of institutions of higher education is of interest for two main reasons; education is an important factor for productivity growth of the economy, and in countries where higher education is funded by the public sector accountability of resource use is of key interest. Educational services consist of teaching, research and the "third mission" of dissemination of knowledge to the society at large. A bootstrapped Malmquist productivity change index is used to calculate productivity development for Norwegian institutions of higher education over the 10 year period 2004-2013. The confidence intervals from bootstrapping allow part of the uncertainty of point estimates stemming from sample variation 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 input the impact on productivity vary a lot.

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

Higher education is important for economic growth and managing structural changes in economies. 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 growth study of the sector of higher education is to get information about the results for the considerable resources consumed out of public funds. Of the central government 11.5 % of the budget for 2016 goes to higher education. 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 productivity study will signal whether the pace of the sector's productivity development can contribute to growth in the economy.

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. This will be the approach of the present study. As tools for estimation we will use non-parametric techniques developed over the last decades to analyse efficiency and productivity. Most of the performance studies of higher education focus on efficiency for units within institutions of higher education using cross-section data (see e.g. Worthington (2001), and De Witte and López-Torres (2015) for a recent comprehensive review), as remarked in Parteka and Wolszczak-Derlacz (2013). Productivity change at the level of institutions of higher education will be the theme of the present study.

Literature review

We start with giving summaries of most of the published productivity-change papers, all using the non-parametric method of calculating Malmquist productivity change indexes for universities as units focusing on choice of variables and overall results. We exclude papers employing Malmquist productivity change index used at the more disaggregated level of departments, and specifying education or research only.

Flegg et al (2004) study efficiency and productivity change for 45 British universities over the academic years 1980/81 – 1992/93. The four input variables are number of staff, number of undergraduate students, number of postgraduate students, and aggregate expenditure. The three output variables are income from research and consultancy, number of undergraduate degrees awarded (adjusted for quality) and the number of postgraduate degrees awarded. The Malmquist productivity change index grew 51.5 % over the total period. Decomposition showed the upward shift in technology to be the major component of productivity change. The split on individual units is not shown.

Carrington et al (2005) study 35 Australian universities for the period 1996 -2000. There is an interesting general discussion of output and input measures and quality attributes. Operating costs is the only input used. However, it is not quite clear which output variables that are actually used of the ones presented in Table 1 (p.154). The nine output measures in the table are student load, science student load, non-science student role (all three full-time equivalent enrolment), student load, research higher degree student load, non-research higher degree student load (last three weighted full-time equivalent enrolment), completions (full-time equivalent enrolment), weighted number of publications, and research grants. Most of the paper is devoted to efficiency measurement and only a short description of productivity change is presented (without discussing relations or choice of model). The productivity growth on average was found to be 1.8 %, and decomposed into an average of 2.1 % frontier shift and an average efficiency decline of 0.7 %. Scale efficiency contribution of 0.4 % indicated that expansion contributed to increased productivity change.

Johnes (2008) uses five inputs and three outputs. The inputs are full-time academic staff, expenditure on administration and central services, expenditure on centralised academic services (library, computer and networks, museums, observatories), full-time equivalent first-degree and other undergraduates, and number (full-time equivalent) postgraduate students. The outputs are number of full-time equivalent first-degree and other undergraduates qualifications awarded, number (full-time equivalent) of higher degree qualifications and post-graduate qualifications awarded (doctorates), and income received in funding and research grants and contracts. A fixed base period, the first one, is used for determining the frontier, and the Malmquist index is calculated for all other periods in turn versus the base period. There is a productivity growth of 1 % for the total period. The sample is split into three sub-groups, universities pre 1992, post 1992, and universities that were previously mostly polytechnics including specialised institutions. A decomposition of the Malmquist index shows that frontier shift has contributed to positive growth, while catching-up is negative. The post-1992 university group and specialised group have significantly higher positive frontier shifts than the pre-1992 university group.

Worthington and Lee (2008) study 35 Australian universities for the period 1998-2003. The five inputs specified are academic staff, non-academic staff, non-labour expenditure, undergraduate student load, and postgraduate student load. The six categories of outputs are undergraduate completion, postgraduate completion, Ph.D. completions, national competitive grants, industry grants, and publication points. The choice of variables is discussed extensively, as are the productivity change results and the various decompositions. Annual productivity growth averaged 3.3 % and the contribution from efficiency change was zero underlining the importance of frontier shift. A test of robustness was done aggregating inputs and outputs to three each. Mean productivity growth increased, but frontier shift was still dominating efficiency change.

Kempkes G and Pohl C (2010) compute Malmquist indexes for 72 German universities for the period 1998-2003. The three inputs are number of technical personnel, number of research personnel and current expenditure and the two outputs are number of graduates and amount of external research grants. The Malmquist productivity change index is decomposed into change in efficiency or catching-up and technical change or frontier shift. The universities from former East Germany perform on average better than universities from West Germany. A result for the total sample is that change in efficiency represents a main determinant of TFP change. Technical change was rather low over the period, 72 % of the units had a negative contribution to TFP while 28 % had a negative contribution of catching-up.

Data for 36 Australian universities and the eight universities in New Zealand for the period 1997-2005 are used in Margaritis and Smart (2011). The four inputs are academic staff, general staff, non-labour operating expenditure, and number of students. The three output variables are undergraduate and postgraduate qualification completions, and number of indexed articles and reviews in Web of Science lagged one year (weighting entries from social sciences and humanities with a factor of 2). The Australian universities are split into three groups; eight large metropolitan universities, Old universities (founded during 1960s and 1970s) and New universities (founded during late 1980s and early 1990s after the Dawkin reform). The Malmquist productivity change index (following Färe et al 2008) is decomposed multiplicatively into efficiency (catching-up) and frontier change, and catching-up is split into technical efficiency change based on a contemporaneous variable returns to scale frontier and scale efficiency change. Aggregated yearly result for all Australian universities is 2.8 % with the subgroup New universities showing the strongest growth with 3.6 % and Old universities the weakest with 1.4 %. The New Zealand universities have a markedly lower yearly growth with 0.1 %. The yearly contribution from frontier shift is 1.9 % and from catching-up 0.9 %. The former effect dominates, and this is also the case for the subgroups varying from 2.9 % for metropolitan universities to 1.3 % for Old universities and the latter varying from 1.7 % for New universities to 0.3 % for the metropolitan group. The yearly technology contribution for New Zealand universities is negative with -0.1 %, and the efficiency component is 0.2 %. The decomposition of the efficiency term reveals that scale efficiency change is most important for universities in both countries.

All the previous studies report point estimates of productivity change. However, after the development of a bootstrapping procedure for the Malmquist productivity change index (Simar and Wilson 1999), taking care of sampling variation only, this seems to be a necessary tool to apply.

Parteka and Wolszczak-Derlacz (2013) study productivity change for 266 public universities from seven European countries over four periods 2001-2005. A novel feature for university studies was claimed to be the first use of bootstrapping to obtain confidence intervals for the point estimated of the Malmquist index and its decompositions. [However, bootstrapping was used in Edvardsen et al (2010) in a study of Norwegian public universities including regional University colleges for the period 2004-2008 using a single input; total employment, and three outputs; study points for lower degree (Bachelor), study points for higher degree (Master), and publishing points for research publications. The bias-corrected productivity growth over the total period was 13.7 % with a (95 %) confidence interval of 10.1 % - 17.4 %. The yearly individual confidence intervals were rather narrow. Because the report is in Norwegian we do not review it further.] Due to the problems of getting comparable data for different countries the number of variables is less than for single country studies. The three inputs are number of students, total academic staff and total revenue. The two outputs are number of graduates and number of publications (in addition to scientific articles proceedings papers, meeting abstracts, reviews, letter, and notes were also included). One of the tests of robustness was estimating country-specific frontiers. The average productivity change was 4.1 % with the decomposition 3.2 % for catching up and 1.2 % for frontier shift. Instead of presenting confidence intervals the number of units with significant productivity improvements is stated to be 56 % of all units, 32 % had significant positive catching up, while 17 % had significant positive frontier shift. The robustness checks roughly showed the same pattern, as did the single country analyses.

Fernández-Santos and Martínez-Campillo (2015) study a sample of 39 public universities of Spain over the academic years 2002/03 – 2008/09. The three input variables are academic staff, registered students and total revenue. Three output variables are specified; graduate students' qualifications, research publications and research and development revenues. The main result is a productivity growth from the first academic year to the last of 11.2 %. The decomposition of the growth yields a negative change of -3.2 % from efficiency change and a 14.7 % positive growth from technological change. Bootstrapping is applied but no details are offered, only total numbers of universities with significant growth of the Malmquist productivity change index (59 %) and its two components, 10 % and 64 % for the efficiency change and technological change, respectively.

The plan of the paper is to present the methods in Section 2, and to introduce the data in Section 3. Then the productivity analyses follow in Section 4 using some special illustration allowing a visual impression of developments. Section 5 concludes.

2. Methods

Caves et al. (1982) introduced the bilateral Malmquist productivity growth 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, based both on constructing an average unit and taking averages of the individual units. The specific linear programming problems used for estimation are set out in Appendix 1.

When applying the Malmquist productivity change index attention should be paid to desirable properties. In the literature this is more often than not glossed over. We therefore find it necessary to explain in more detail the choice of our specification. Productivity growth 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 growth (TFP) is one that is homogenous of degree 1 in the input-output vector, and thus supported by the linear-homogenous set that fits closest to the observations. The homogenous constant returns to scale (CRS) 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 (VRS) frontier for period *v*. From classical production theory we know that the productivity is maximal at optimal scale where the returns to scale is

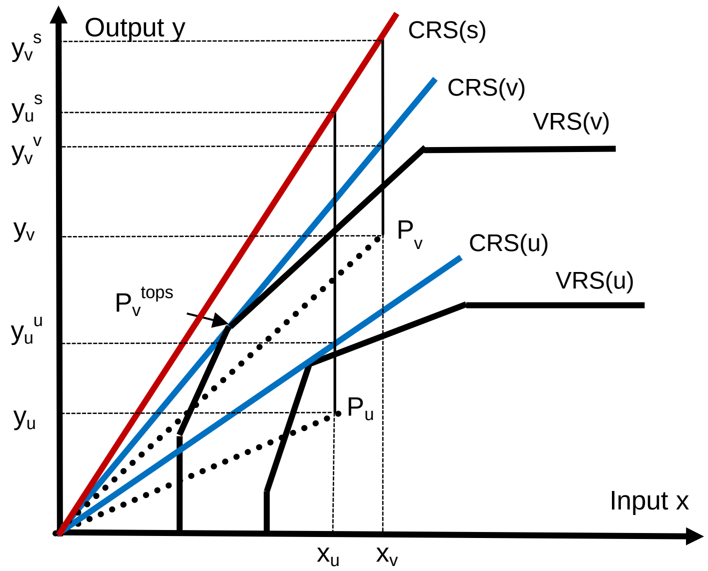


Figure 1. The Malmquist productivity change index.
Productivity change for a unit from period u to period v measured relative to the benchmark CRS(s) envelopment of the maximal productivity of the pooled dataset.

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 contemporaneous frontiers are drawn showing an outward shift indicating technological progress. The contemporaneous CRS benchmarks (blue) rays are tangents to the TOPS points. 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 obtaining all the required homogeneity properties of a Malmquist productivity index.

Another important property of a productivity index 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 make the Malmquist index transitive by using a single benchmark enveloping the pooled data, as illustrated by the upper (red) ray CRS(s) in Fig. 1. 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 technical productivity 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 that is traditionally done (see e.g. 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

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

where superscript *s* symbolises that all data are used as the benchmark reference set. The Malmquist productivity estimator is conditional on the estimator $\hat{S}^s$ 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 (the measure termed E_3 in 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 (Førsund, 2015). In Fig. 1 the Malmquist index (1) estimator $\hat{M}^s(u,v)$ is $(y_v / y_v^s)/(y_u / y_u^s)$. We should be able to see that observation P_v is relatively much closer to the benchmark than observation P_u, i.e. $\hat{M}^s(u,v) > 1$.

There are two ways productivity can change over time; change in efficiency and shift in technology (Nishimizhu and Page 1982). If contemporaneous frontiers are calculated the Malmquist index can be multiplicatively decomposed into an efficiency term, or catching-up term *MC*, and a term capturing the shift of the frontier, *MF*. In order to keep the proportionality property the contemporaneous benchmark must also be CRS, as illustrated in Fig. 1 with the (blue) CRS rays for periods *u* and *v*, respectively. Keeping the circularity of both components we have the decomposition

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

Here superscripts *v* and *u* indicates the contemporaneous benchmark envelopments, while *s* stands for the benchmark envelopment based on the pooled dataset. The *MC*- measure shows how a unit is catching-up with the frontier, and the MF measure shows the potential frontier shift. In the literature is has been assumed that the “true” period technology is VRS. The catching-up term has then been decomposed into a product of an efficiency term relative to each VRS frontier and a scale efficiency change using the definition of scale efficiency in Førsund and Hjalmarsson (1979) (see Färe et al 1994a,b). However, scale issues will not be pursued here.

In Fig. 1 the catching-up term can be calculated as $MC = (y_v / y_v^v)/(y_u / y_u^u)$. It should be possible to see that observation *v* is relative closer to its own period CRS benchmark than observation *u*, i.e. *MC* > 1. The *MF* measure of technology shift is calculated as a ‘double’ relative measure where both period benchmark efficiency measures are relative to the pooled benchmark measure; $MF = (y_v^v / y_v^s)/(y_u^u / y_u^s)$ in Fig.1. It should be easy to see that $MF > 1$.

However, note that the standard decomposition does not mean that there is a causation; we cannot unambiguously distinguish between productivity change due to increase in efficiency and due to shift in technology using the components in (2), as often appear to be believed in the literature (all eight papers reviewed in Section 1 adopt the standard definition of decomposition, however, Johnes (2008) and Worthington et al (2008) have some discussion). Following the assumption made in Nishimizu and Page (1982), introducing this decomposition for discrete time, 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 *per se*, but illustrates the actual relative catching-up to the frontier that is also influenced by the technology shift. There is no objective way to decompose efficiency effects and frontier shift effects without making specific assumptions, according to Nishimizu and Page (1982) (see Førsund (2015) for a detailed exposition).

Bootstrapping

We are following the homogeneous bootstrap procedure outlined in Simar and Wilson (1998; 1999; 2000). For weaknesses with the bootstrap assumptions see Olesen and Petersen, (2016). Testing the period frontier function form, CRS versus VRS, using the S_1 measure in Simar and Wilson (2002), the latter turned out to be accepted. 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 relative to the VRS frontier for each time period:

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

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 for each period. Now, going back to each run for a pair of periods, the Malmquist productivity index, given by (1), is calculated using the linear homogeneous benchmark envelopment created for the pooled set of all output pseudo observations in the benchmark set.

Assuming estimators to be consistent, Appendix 2 shows how the sampling bias can be estimated. 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. Simar and Wilson (1999) show how confidence intervals can be calculated in this case. The procedure is set out in Appendix 2.

3. Data

The traditional outputs of institutions of higher education are connected to teaching, research and the "third mission", i.e. dissemination of knowledge to - and various interactions with - the society at large. When studying productivity the key to success is, first of all, to base the study on theoretically satisfactorily definitions of inputs and outputs, and then to operationalise these definitions without compromising too much. Based on the literature reviewed in Section 1 some of the mostly used operationalisations are set out in Table 1. The number of students is used both as an output and as an input. Research grants are used as a variable catching research output. In Norway external grants play a minor role in research financing. We see that no measures for the "third mission" are listed, neither are quality variables. These variables are notoriously difficult to get measures for. Interactions with society could be measured by number of popular media appearances by faculty, participation in government committees and in writing white papers, and consultancies. Quality of education could be measured by grades achieved, time to get the first job after finishing, and expected life-time earnings.

Table 1. Input- and output variables used in productivity studies

Inputs	**Outputs**
Total employees	Total completions
Faculty employees	Undergraduate qualifications
Administration	awarded
Technical personnel	Graduate qualifications
	Awarded
Total expenditure	
Spending on labour	Study points for courses of a lower
Non-labour expenditure	degree (cost weighted)
Spending administration	
and central services	Study points for courses of a higher
Spending on academic services	degree (cost weighted)
(libraries, IT equipment and	
systems, museums,	Total student load
observatories, laboratory	Science student load
equipment)	Other student load
Buildings m^2	Number of publications
Total number of students	Publishing points
Undergraduates	
Graduates	Doctorates/Ph.Ds
	External research grants

There are six variables in all used our analysis; two inputs and four outputs set out in Table 2. The data are taken from the Database for Statistics on Higher Education (DBH), a state-run central register of data for institutions of higher education in Norway, covering a broad range of topics in the sector of higher education institutions including research. Due to the degrees of freedom enforcing a parsimonious model we have restricted the variables to the key ones. Capital, like equipment and buildings (measured by area m^2), or measured by expenses, had to be excluded because these variables are not reported for private institutions. However, capital is rather generic and should not discriminate much between institutions, provided that the capacity to produce educational services is not restricted by buildings (the rule in Norway is not to enroll more students than capacity allows). As to quality variables for inputs one could measure the quality of faculty by position and experience, and the grade of students at start of studies, and measure the quality of outputs as grade level for degrees (Bachelor and Master), job success, and expected life-time earnings. As to quality of research its impact measured by citations can be used, as well as prestige of the journal of the publication, and external

Table 2. Inputs and outputs used in the study

Inputs	Outputs
Faculty employees	Study points for courses of a lower degree (cost weighted)[a)]
Administration and other employees (excluding cleaners)	Study points for courses of a higher degree (cost weighted)[a)]
	Publishing points[b)]
	Doctorates/Ph.Ds

a) Study points are calculated as the norm of number of 60 course points per year weighted according to state contributions to seven different types of studies such as, medical studies, science studies, architecture, design and arts, humanities higher level, humanities lower level, nursing and teacher students, and students coming in and leaving, catching typical differences in cost of students.

b) There are three types of research publications and two levels of quality giving publishing points: journal article level 1 (1) and level 2 (3), book chapter level 1 (0.7) and level 2 (1), book level 1 (5) and level 2 (8). Publishing points ranging from 0.7 to 8 are given in parentheses. The points are weighted with the share of authors from the institution in question of total authors.

research funding. However, we have not included such quality variables, partly due to the fact that this information is not available in the data base. As to other employees than faculty in Table 2 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. Using a lag between resource use and completion of the Ph.Ds. of e.g. three years reduced the number of observations and did not influence the results that much. Therefore we have chosen not to use lags. We have also, following the literature, neglected the lags between use of resources on research and the actual research publication being published.

We have formally tested if the model can be reduced further by aggregating variables such as employees or study points, dropping Ph.Ds. and cost weighting of study points, but these changes were all rejected.

The total number of units appearing one or more years in the DBH database is 75, varying from 63 in 2004 to 59 in 2013. We did not have the opportunity to control data at the institution level (it would be prohibitively costly and time consuming), so the only option is to delete units with missing data. Then there is the question of extreme outliers influencing the benchmark envelope. One possibility is that there are errors in reporting, blowing up one or more outputs and/or shrinking one or more inputs. However, the downside of deleting extremely efficient units is that we may lose correct information. There are various approaches to detecting efficient outliers, from the first suggestion in Timmer (1971) of "peeling the onion" by removing one efficient unit at a time until a prescribed number (or share) of units is removed, a variation of this approach using super-efficiency scores (Andersen and Petersen, 1993) and eliminating units with higher values than a predetermined level (Banker and Chang, 2006), and using the importance of the extreme-efficient unit as a referent unit (Torgersen et al, 1996). We end up deleting 7 observations with super-efficiency scores above 1.25 and/or being the referent for inefficient units having more than 25 % of the saving potential for inputs. There remain 42 units that have observation for all years, thus constituting a balanced panel for the total period. The number of units appearing is 49. A few units have been merged during the period, and are aggregated artificially for all years when estimating productivity change. However, the original actual units are used for the premerger period in the benchmark set. The estimation of the benchmark CRS envelopment is based on about 500 observations. We do not need a balanced panel to calculate the benchmark envelope; in fact we would lose information if we used the balanced panel only

The development of our variables for the study period is set out in Figure 2 on index form

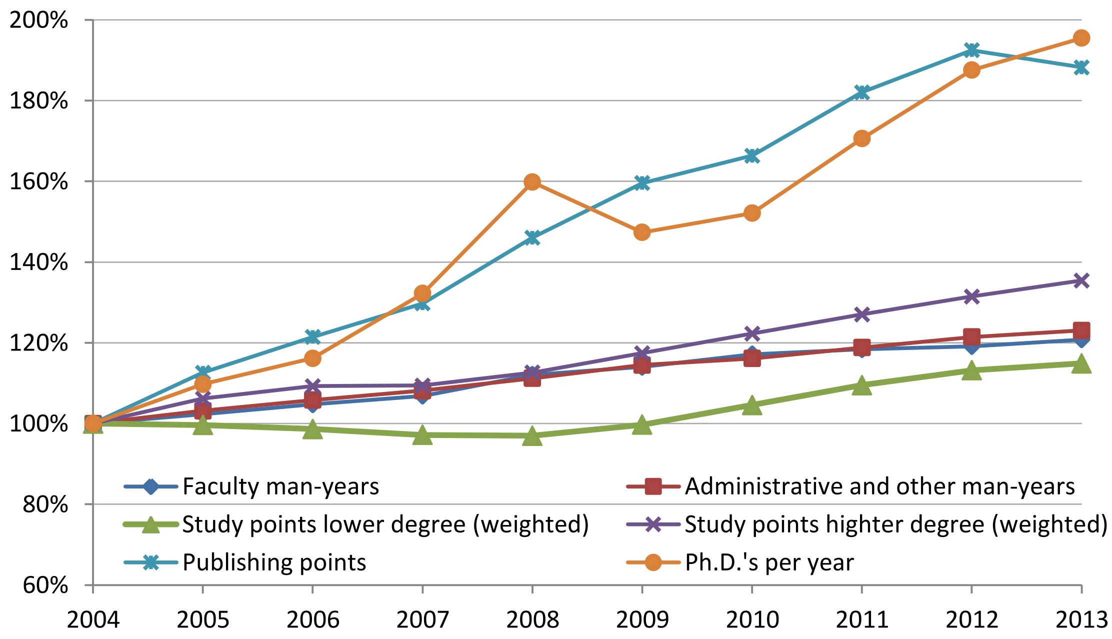


Figure 2. Development of the variables for the periods 2004 to 2013 relative to 2004
(See Table 2 for definitions of study points and publishing points)

with the values in 2004 as the base. (See Appendix 3 for the individual average data for 2004and 2013.) The two outputs publishing points and Ph.Ds have had 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 man-years 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, based on Farrell's way of measuring how the performance of a sector as a whole 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

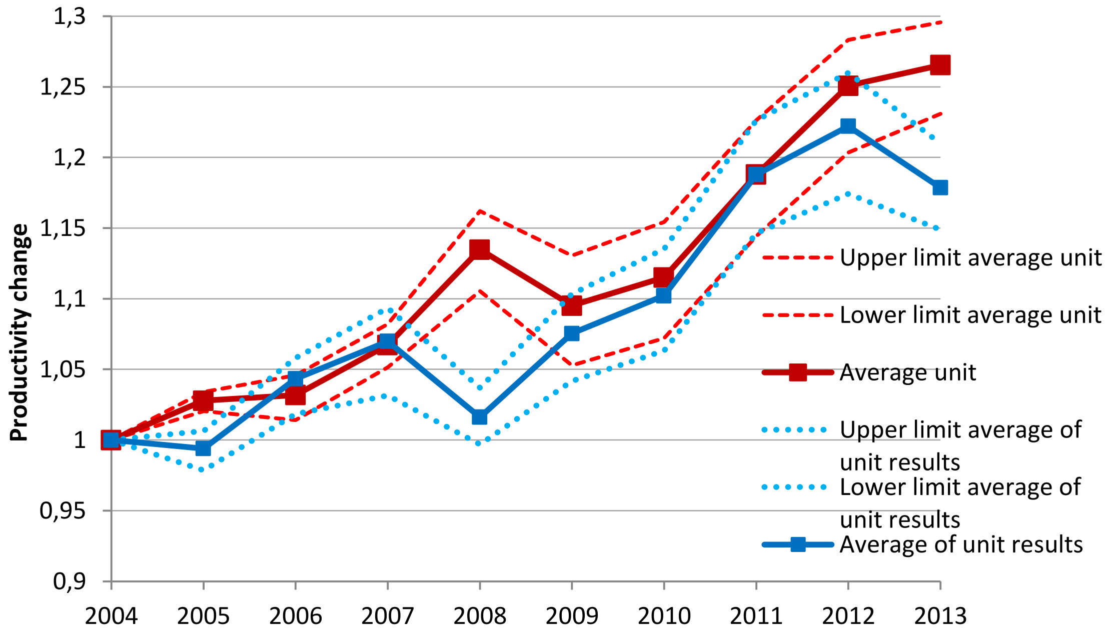


Figure 3. Aggregate productivity change (solid lines) for the periods 2004 to 2013 relative to productivity in 2004 measured by the average unit, and average of individual productivities with 95 % confidence intervals (broken lines).

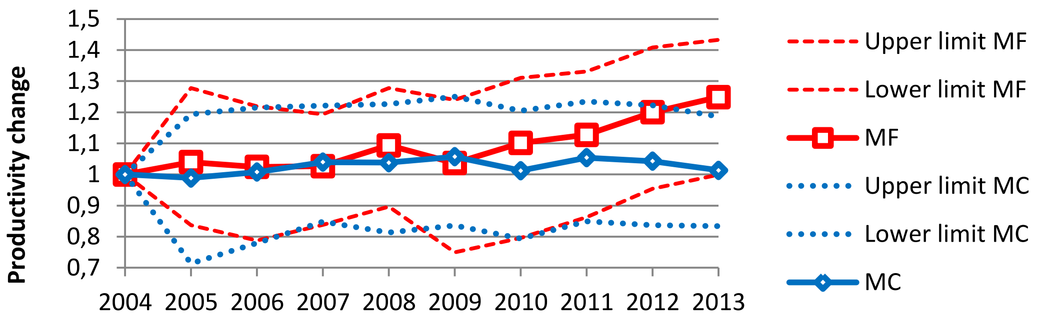


Figure 4. The decomposition of the Malmquist productivity index for the average unit into catching-up MC and frontier shift MF for periods 2004-2013 relative to 2004

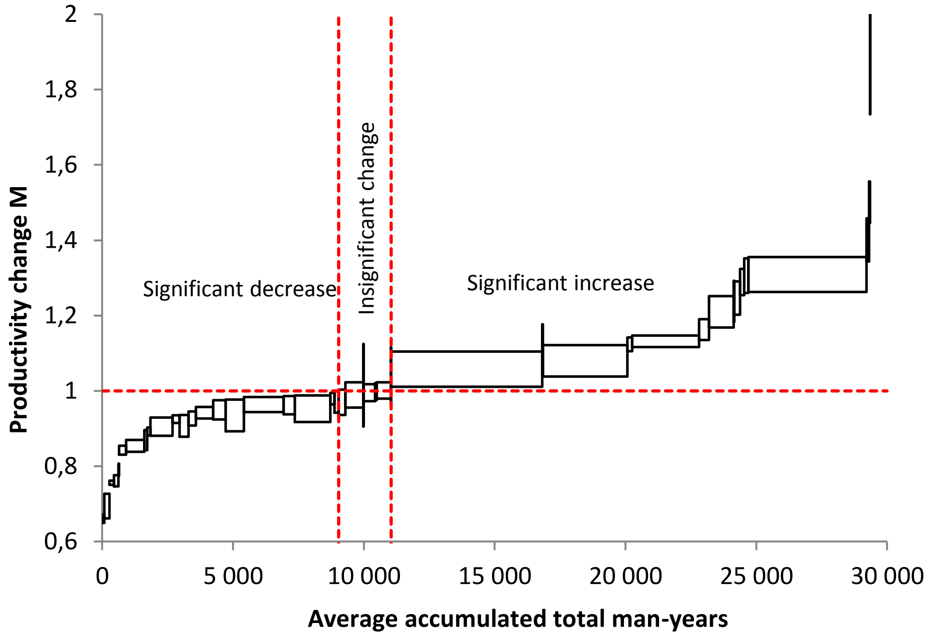
growth from 2007 to 2008 with a positive jump in the productivity growth measured by the average unit and a negative growth for the average of productivities growth measure, and a similar development in the last period. This difference may be due to small units having a weaker productivity development 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.

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

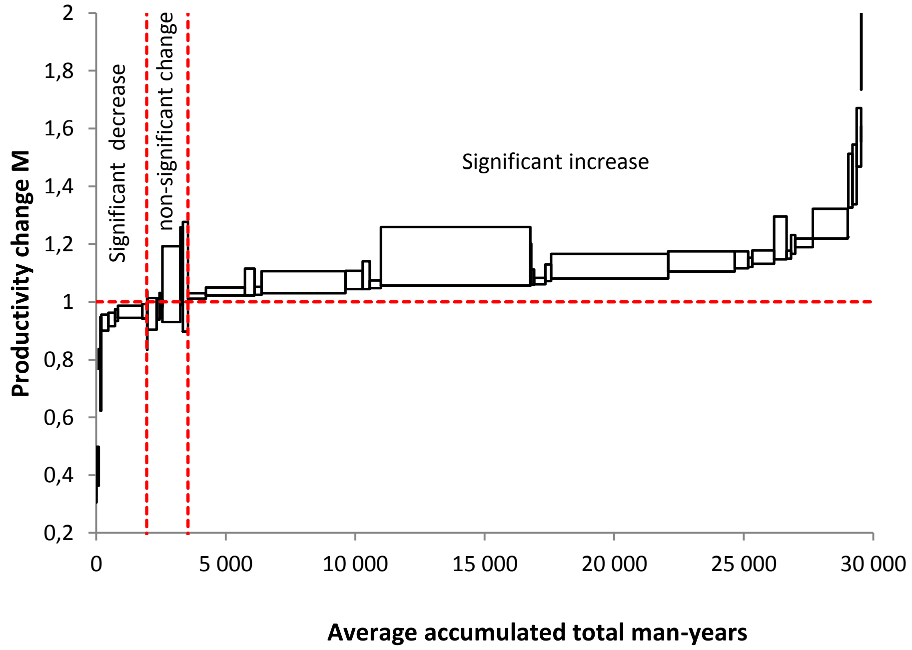
Productivity development of individual units

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. The individual productivity results, together with the extent of uncertainty in the form of confidence intervals, can be displayed as a sorted distribution in a special type of diagram. (The numerical results are set out in Appendix 3 for period 2004-2013.) 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 location of units according to size is revealed.

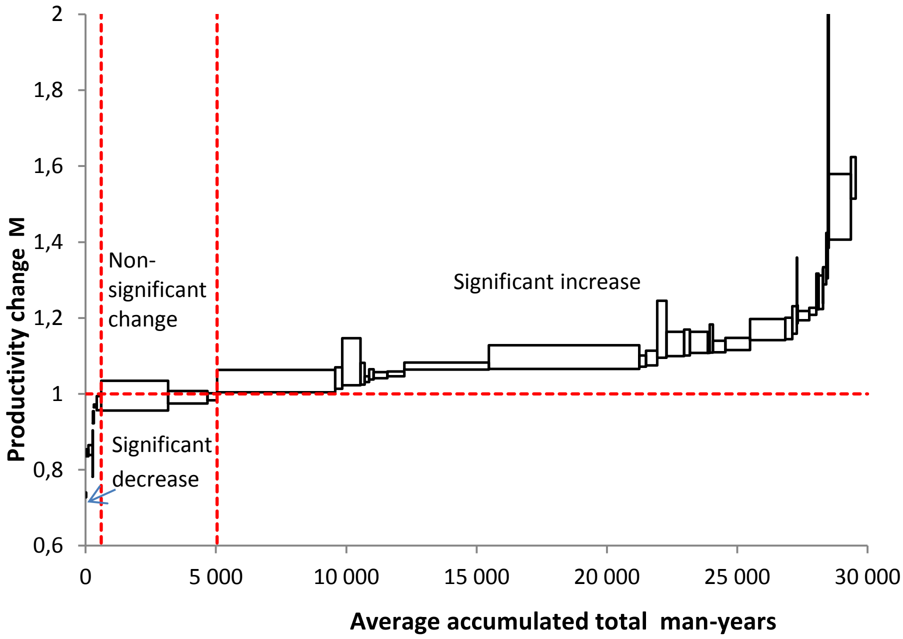
In Figure 5 four panels of productivity-change distribution for all the individual units are set out for three year periods, and the total period 2004 – 2013. (Due to perverse influence of the layout and readability of the diagrams a few units are not shown as indicated in the panel



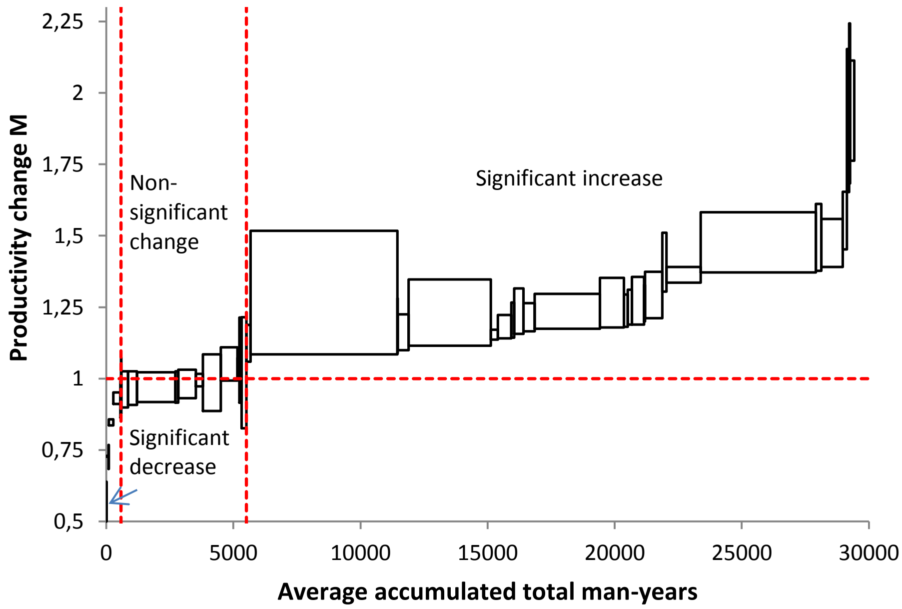
Panel (a) 2004-2007
(Two units with lowest and highest M, respectively, are not shown)



Panel (b) 2007-2010
(One unit with highest M not shown)



Panel (c) 2010-2013
(One unit with lowest M not shown)



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.

texts.) 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. As a measure of size 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 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.

The series of sub-period productivity change distributions allow us to see structural change regarding features such as the range of the range of distributions, shifts in the size of the three subgroups as to significance of productivity change, change of location of small and large units, and movement of units along the distributions.

The four largest units are easily identified in Panels (a)-(d) because the same size is used for all years. Some very small units have both the lowest and the highest productivity in 2004-2007 (lowest and highest not shown in the figure). The four largest units are all located in the subgroup of units having significant growth in Panel (a). Five units only are in the subgroup of insignificant change, while the highest number of small units is in the subgroup of significant decrease of productivity. Moving on to Panel (b) the number of units in the latter group has contracted considerably but still consisting of very small units only but for one. The position of the largest units has changed in the subgroup and the confidence interval for the largest unit has increased. Panel (c) shows that both the two first subgroups have continued to shrink, the significant decrease group now consists of very small units only, while the insignificant group rooms three units only, one of them being the 4th largest university. Panel (d) spanning the whole period reveals that the subgroup of very small units with significant decrease in productivity has all but vanished; the insignificant group consists of medium-sized units. All the four large units are in the group having significant productivity growth. Although the distribution of point estimates of productivity change has shifted upwards the confidence intervals have increased substantially for the large units due to variation in productivity change over the periods and a trend of upward movement.

Some common features are that the productivity numbers on the whole are relatively sharply determined; the confidence intervals are rather narrow. Large units tend in general to have wider confidence intervals than medium-sized units. Small units tend to have the widest confidence intervals. A few quite small units have rather wide confidence intervals for all the panels. A general structural feature is the shrinking of the group of units with significant productivity decrease and the increase in the number of units with productivity increase. 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 (b) and (c), and to 81 %, corresponding to 29 of the 44 units, for Panel (d) for the whole period.

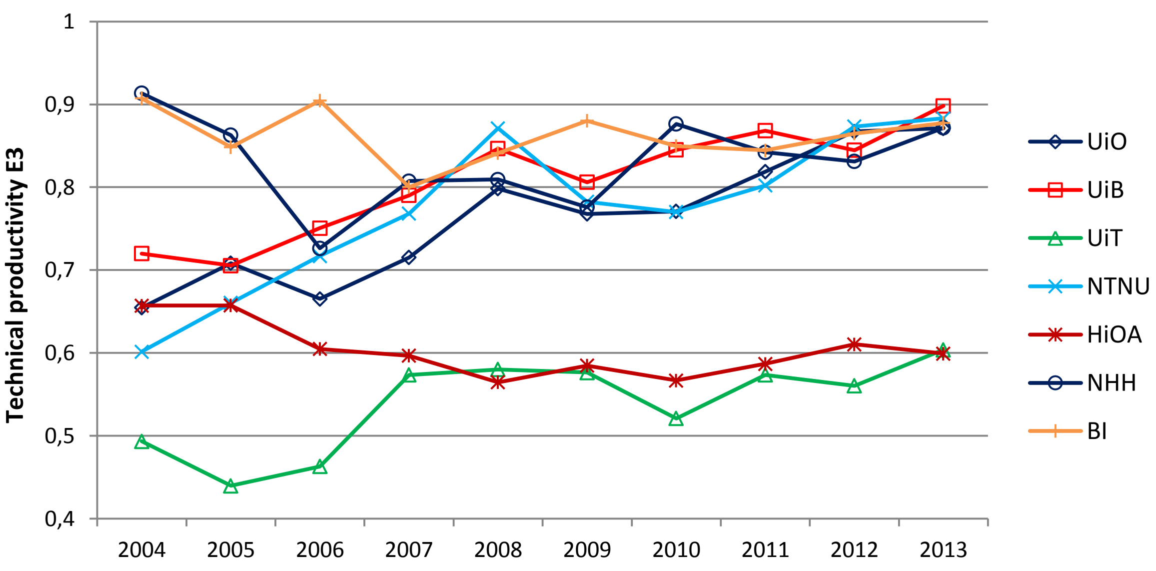
Decomposition of the productivity change

The decomposition results for the aggregate unit are representative for the results for the individual units. In Appendix 3 the results for the indexes calculated for 2013 relative to 2004 are set out. The significant results are set in bold. We have that while the Malmquist productivity change index has 14 % of the units with significant decrease, 27 % with insignificant growth and 59 % with significant increase (see Panel (d) of Fig. 5), 16 % of the units have a significantly catching-up index less than 1, i.e. a significant decline, 57 % have insignificant change, and 27 % significant positive contribution to the Malmquist index. The impact of frontier shift is slightly more positive; 7 % of the units show a significant decrease, 66 % an insignificant change and 27 % a positive impact. (However, remember the caution about putting too much into attribution of the components as mentioned in Section 2.)

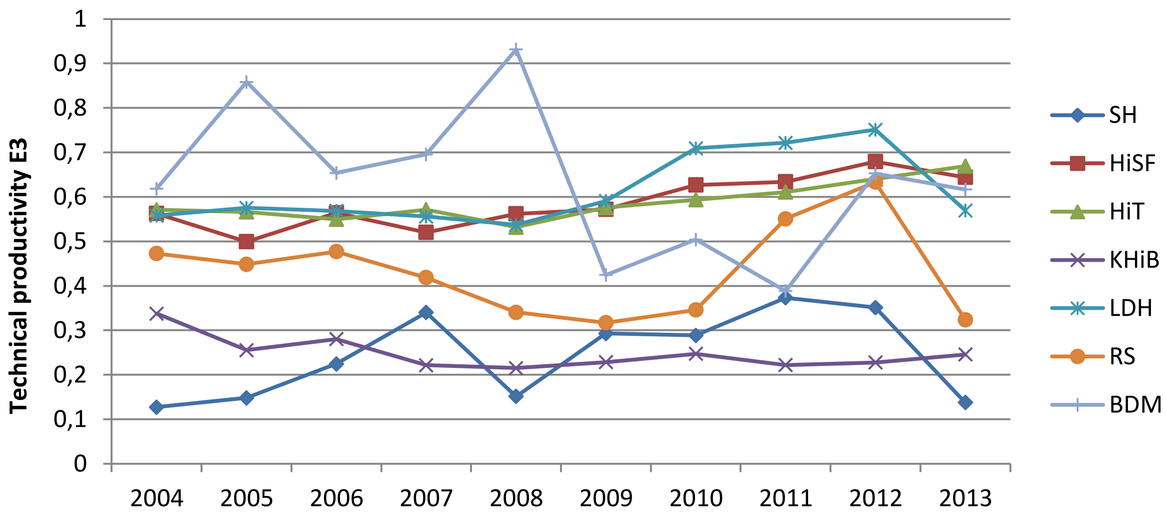
Productivity over time for sub-samples of selected large and small units

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 university college (HiOA).

The Malmquist productivity index is the ratio of consecutive values of the value of the level of productivity (technical productivity measure E_3 in Førsund and Hjalmarsson, 1979). This means that if productivity has gone down from e.g. 2004 to 2005, as is the case for the two business schools, 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 have 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 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



Panel (a). Sub-sample of large units



Panel (b). Sub-sample of small units

Figure 6. Development of level of productivity relative to benchmark (E_3) for selected large and small units

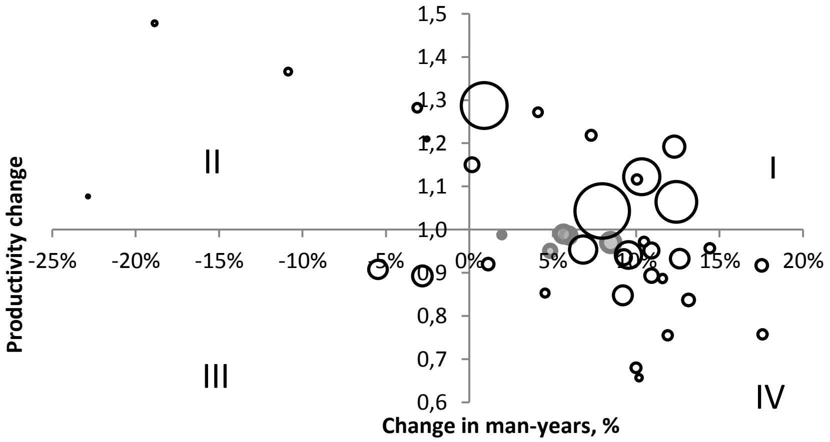
HiOA is falling from 2004 to 2008 to the level of UiT and then evening out ending up with a non-significant negative change for the whole period. A striking trend of the development of the other units is that there is some turbulence in productivity up to 2008, but then the developments become more like and all units end up with about the same level of productivity close to 90 % implying a positive productivity growth for the universities. For the two business schools, however, this is an insignificant change for because these start out with high levels of productivity, considerably higher levels 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.

Panel (a) in Figure 6 shows that the large units have productivity levels all converging to 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, being so far from optimal scale of the benchmark technology, but we cannot say whether the large units are also too small or are too large without conducting a further, more detailed, analysis.

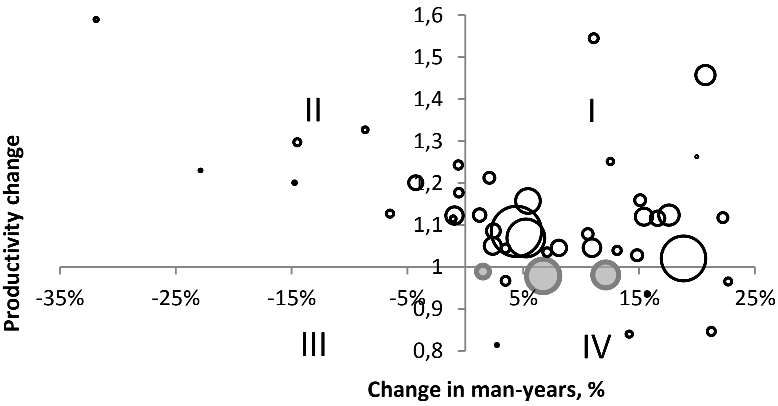
Productivity and resource use

A recurrent policy question is the return on the resources allocated to higher education. Showing the change in total labour used together with productivity change provides some answers (Førsund et al, 2006). In Figure 7 productivity changes for the same periods as for Fig.5 is shown together with the relative change in total man-years illustrating the heterogeneity. 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

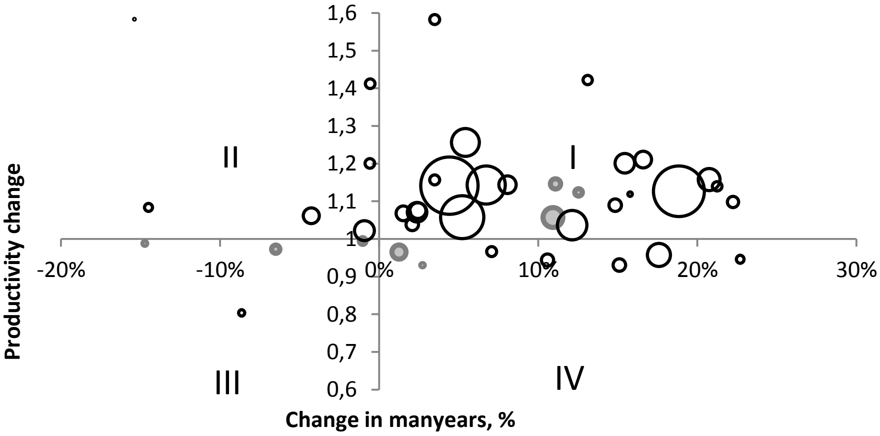
The horizontal line at the value 1 delimitates the units with productivity decrease and increase, respectively, and the vertical axis from zero change in labour form four quadrants numbered I to IV. In Quadrant I units have had both productivity growth and increase in man-years. Such units may be said to have experienced *efficient labour expansion*. The units



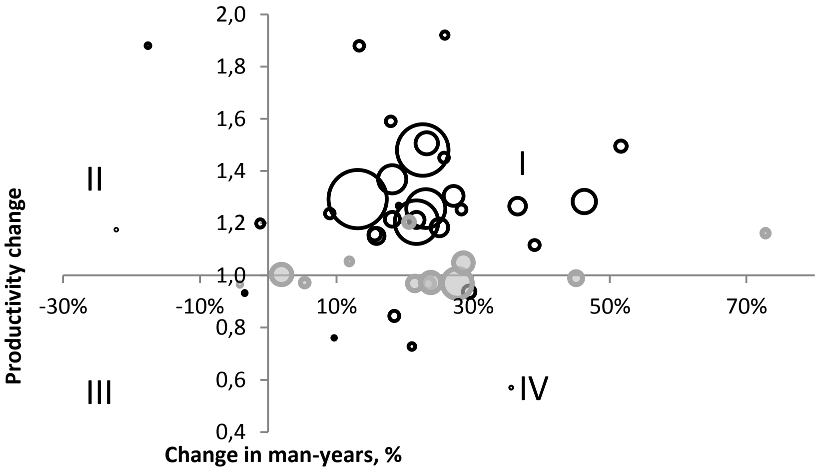
Panel (a) 2004-2007
(Three extreme units in QI and one in QIV are not shown)



Panel (b) 2007-2010
(One extreme unit in QI and one in QIV are not shown)



Panel (c) 2010-2013
(Two extreme units in QI and two in QIV are not shown)



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.

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*. (See also Førsund and Kalhagen (1999) where units in the quadrants II, III are termed having *positive* and *negative adjustment capability*, respectively).

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 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 %, and 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 unit SH seen in Panel (b) in Figure 6 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 BI and NHH (shown in Panel (a) of Figure 6) 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 %.

5. Conclusions

Studies of productivity of institutions of higher institutions are of interest for two main reasons; education is an important factor for productivity growth for the macro 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 by a public agency. 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 financial contributions to types of courses from the Ministry of Knowledge and 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.

There are several ways to extend the study of productivity change. Optimal scale of institution of higher learning is a "hot" topic in Norway and can be undertaken based on the notion of optimal scale that maximizes the productivity level. An interesting policy question is whether scale should be increased in order to improve productivity or efforts should be concentrated on reducing technical inefficiency.

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). Mergers are one obvious way of increasing size, but the question remains whether this will increase productivity.

Although the institutions of higher education studied have had the same type of variables there is heterogeneity that should be investigated forming subgroups. Some institutions are more specialized than others, and the effects of specialization or scope as to outputs is an interesting topic (Daraio et al, 2015). Some units are serving special interests, whether political or cultural, and should be investigated as a separate group. In Norway there has been a development of regional colleges founded to provide shorter more "practical" education than traditional universities, to become universities, so there we have two sub-groups for further investigation. Another classification is according to ownership being private or public.

Quality variables have not been used in the study. This a priority task for further research. Some types of quality variables are mentioned in Section 3, but these and may be more relevant ones need to be developed.

References

Andersen P and Petersen NC (1993). A procedure for ranking efficient units in Data Envelopment Analysis. *Management Science* **39**(10): 1261-1264.

Banker RD and Chang H (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Research* **175** (2): 1311–1320.

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.

Carrington R, Coelli and Rao DSP (2005). The performance of Australian universities: conceptual issues and preliminary results. *Economic Papers* **24**(2): 145–163.

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.

Coelli TJ, Rao DS, O'Donnell CJ. and Battese GE (2005). *An introduction to efficiency and Productivity Analysis (second editon).* Springer: New York.

Daraio C, Bonaccorsi A and Simar L (2015). Efficiency and economies of scale and specialization in European universities: A directional distance approach. *Journal of Informetrics* **9**(3), 430–448.

De Witte K and López-Torres L (2015) Efficiency in education. A review of literature and a way forward. *Journal of the Operational Research Society.* Pre-published 16 December 2015, DOI 10.1057/jors.2015.92

Edvardsen DF, Førsund FR og Kittelsen SAC (2010). *Effektivitets- og produktivitetsanalyser på StatRes-data* [*Efficiency- and productivity analysis based on StatRes data*]. Rapport 2/2010, Kapittel 4, 31-47. Ragnar Frisch Centre for Economic Research: Oslo.

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: Cheltenham, UK, pp. 97-131.

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

Färe R, Grosskopf S and Lovell CAK (1994a). *Production frontiers*. Cambridge University Press: Cambridge.

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*. Oxford University Press: New York, pp. 522-622.

Färe R, Grosskopf S, Lindgren B and Roos P (1992). Productivity changes in Swedish pharmacies 1980-1989: a non-parametric approach. *Journal of Productivity Analysis* **3**(1-2): 85-101.

Färe R, Grosskopf S, Norris M and Zhang Z (1994b). Productivity growth, technical progress and efficiency change in industrialized countries. *American Economic Review* **84**(1): 66-83.

Farrell MJ (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society* Series A, **120**(3): 253-281.

Fernández-Santos Y and Martínez-Campillo A (2015). Has the teaching and research productivity of Spanish public universities improved since the introduction of the LOU? Evidence from the bootstrap technique. *Revista de Educación* **367**(January-March): 90-114. DOI: 10.4438/1988-592X-RE-2015-367-284

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 (2015). Productivity interpretations of the Farrell efficiency measures and the Malmquist index and its decomposition. *Memorandum* 14/2015, Department of Economics, University of Oslo.

Førsund FR and Hjalmarsson L (1979). Generalised Farrell measures of efficiency: an application to milk processing in Swedish dairy plants. *Economic Journal* **89**(354): 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**(10): 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. (Also issued as Memorandum11/1999 from Department of Economics, University of Oslo).

Førsund FR, Kittelsen SAC, Lindseth F and Edvardsen DF (2006). The tax man cometh - but is he efficient? *National Institute Economic Review* **197**(1): 106-119.

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**(2): 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 2008/09: parametric and non-parametric estimation of the multi-input multi-output distance function. *The Manchester School* **82**(4): 465-487.

Kempkes G and Pohl C (2010). Efficiency of German universities–some evidence from nonparametric and parametric methods. *Applied Economics* **42**(16): 2063–2079.

Margaritis D and Smart W (2011). Productivity change in Australasian universities 1997-2005: a Malmquist analysis. Paper presented at the 52 Annual Conference of the New Zealand Association of Economics 29 June - 1 July 2011, Wellington, New Zealand.
http://nzae.org.nz/wp-content/uploads/2011/Session5/54_Smart.pdf

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**(368): 920-936.

Olesen OB and Petersen NC (2016) Stochastic Data Envelopment Analysis—a review. *European Journal of Operational Research* 251(1), 2–21

Parteka A and Wolszczak-Derlacz J (2013). Dynamics of productivity in higher education: cross-European evidence based on bootstrapped Malmquist indices. *Journal of Productivity Analysis* **40**(1): 67-82.

Pastor JT and Lovell CAK (2005): A global Malmquist productivity index. *Economics Letters* **88**(2): 266-271.

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

Simar L and Wilson PW (1998). Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* **44**(1): 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**(1): 49-78.

Simar L and Wilson PW (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research* **139**(1): 115–132

Timmer CP (1971). Using a probabilistic frontier production function to measure technical efficiency. *Journal of Political Economy* **79**(4): 776-794.

Torgersen AM, Førsund FR and Kittelsen SAC (1996). Slack adjusted efficiency measures and ranking of efficient units. *Journal of Productivity Analysis* **7**(4): 379–398.

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**(3): 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**(3): 285-298.

Appendix 1.

Efficiency scores

The calculation of Farrell technical efficiency scores for the units in the panel based on a CRS benchmark envelopment

$$
\begin{aligned}
& 1/E_{it} = Max_{\lambda,\theta}\,\theta \\
& s.t. \\
& Y\lambda \geq \theta y_{it}, i = 1,...,n, t = 1,...,T \qquad \text{(A1)} \\
& X\lambda \leq x_{it}, i = 1,...,n,\ t = 1,...,T \\
& \lambda \geq 0,\ \theta \geq 0
\end{aligned}
$$

The observation *it* is one of the *n* in the panel of units for time period *t*, E_{it} is the efficiency score for unit *i* in period *t*, θ is the output expansion factor, *Y* is the $m \times T \cdot n$ matrix of *m* outputs in the reference set, *Tn* is the number of units in the pooled data, λ is the $Tn \times 1$ vector of intensity weights defining the projection of unit *it* to the CRS benchmark envelopment, and *X* is the $k \times T \cdot n$ matrix of *k* inputs in the reference set.

The calculation of the Malmquist productivity change index from period *t* to period *t* + 1 then follows from inserting the scores obtained from solving (A1) for the unit and periods in question into (1) and (2).

Appendix 2

Bootstrapping

A new DEA frontier is estimated using the pseudo observations (x_i, y_i^{ps}) we get from (3). We make 2000 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 *Tn* 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

$$(\tilde{M}^s(u,v) - \hat{M}^s(u,v)) \mid \hat{S}^s \sim (\hat{M}^s(u,v) - M^s(u,v)) \mid S^s,\ \ u,v = 1,..,T, u \neq v \qquad \text{(A2)}$$

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 benchmark envelopment 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 (MSE) 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. Simar and Wilson (1999) suggested another way to calculate the confidence intervals. The confidence interval limits (dropping the two periods for convenience) may be defined by:

$$\Pr(-\hat{b}_{\alpha i} \leq \tilde{M}_i^s - \hat{M}_i^s \leq -\hat{a}_{\alpha i} \,\big|\, \hat{S}^s) = 1 - \alpha \qquad \text{(A3)}$$

The estimates for the limits are found from the distribution of $(\tilde{M}_{i,b}^s - \hat{M}_i^s)$ for b =1,..,B (B = 2000) by sorting in increasing order and finding the values for $\hat{a}_{\alpha i}$ (lower) and $\hat{b}_{\alpha i}$ (higher) matching the chosen degree of confidence. The estimated (1 - α) confidence interval for the true Malmquist index M^s is then

$$\hat{M}_i^s + \hat{a}_{\alpha i} \leq M^s \leq \hat{M}_i^s + \hat{b}_{\alpha i} \qquad \text{(A4)}$$

Since the mean square error (MSE) of the bias-corrected Malmquist index estimate is larger than the estimated MSE of the original deterministic estimate $\hat{M}_i^s$ the confidence interval is centered around the point estimate $\hat{M}_i^s$.

Appendix A3

Table A. 1 Data

Short name	Faculty man-years		Admin and other man-years		Study points lower degree (weighted)		Study points higher degree (weighted)		Publishing points		PhDs	
	Mean	Min - Max	Mean	Min - Max	Mean	Min - Max	Mean	Min - Max	Mean	Min - Max	Mean	Min - Max
HiH	81.1	(74.5-86.4)	40.4	(37.8-46.5)	1027.5	(934.0-1147.2)	19.9	(0.0-76.7)	17.5	(4.8-30.7)	0.0	(0.0-0.0)
HiN	107.6	(96.7-124.0)	56.3	(48.6-62.5)	789.3	(732.5-941.5)	203.9	(163.0-232.5)	36.7	(10.8-67.7)	0.0	(0.0-0.0)
HiNe	76.3	(66.4-84.7)	36.3	(31.3-39.4)	798.9	(678.3-895.2)	19.2	(0.0-36.3)	13.4	(4.9-22.8)	0.0	(0.0-0.0)
SH	39.8	(29.2-44.9)	46.4	(23.2-57.2)	118.4	(79.4-151.8)	3.4	(0.0-9.7)	20.1	(0.0-34.3)	0.0	(0.0-0.0)
HiNT	242.9	(228.0-276.2)	131.5	(116.6-149.7)	3461.7	(3091.8-3883.0)	180.8	(115.3-235.1)	40.3	(2.5-89.9)	0.0	(0.0-0.0)
HiST	432.8	(411.3-477.3)	258.6	(219.0-303.1)	6732.2	(5998.4-7620.2)	407.3	(179.9-686.8)	88.1	(29.6-151.5)	0.0	(0.0-0.0)
HiB	449.7	(391.4-511.7)	211.0	(186.9-231.8)	6564.8	(6059.5-7413.9)	144.7	(38.7-344.5)	81.4	(37.1-141.7)	0.0	(0.0-0.0)
HiM	108.2	(99.2-120.1)	49.2	(44.1-53.0)	1163.8	(981.1-1546.1)	261.6	(163.0-369.1)	43.8	(15.9-74.6)	2.9	(0.0-6.0)
HiSF	184.9	(168.9-208.3)	83.3	(80.6-88.7)	2536.4	(2051.1-3125.1)	83.6	(49.0-153.4)	37.8	(20.6-50.1)	0.0	(0.0-0.0)
HSH	170.5	(156.5-190.2)	84.7	(71.3-97.2)	2268.0	(2069.6-2364.3)	58.9	(30.3-84.6)	39.7	(8.9-65.3)	0.0	(0.0-0.0)
HiVo	179.4	(151.4-203.8)	93.0	(78.6-105.6)	2391.5	(2078.4-2747.7)	214.6	(132.6-294.8)	71.4	(45.1-113.0)	0.0	(0.0-0.0)
HiÅ	106.5	(92.9-124.7)	67.2	(55.2-84.0)	1566.2	(1318.6-1811.8)	46.1	(0.0-125.7)	16.7	(3.8-32.6)	0.0	(0.0-0.0)
HiT	332.0	(295.3-384.9)	197.8	(182.2-214.7)	4900.6	(4420.1-5766.1)	485.7	(368.1-639.8)	76.5	(37.5-149.9)	0.9	(0.0-5.0)
HiØ	270.8	(261.5-283.5)	163.5	(133.5-187.4)	3807.3	(3541.3-4361.4)	193.8	(111.7-326.7)	58.9	(23.4-111.1)	0.0	(0.0-0.0)
HiAk	917.5	(811.4-1024.3)	595.8	(508.1-678.5)	13797.2	(12855.6-14927.5)	977.9	(491.7-1514.6)	289.5	(103.1-414.8)	1.4	(0.0-5.0)
HBu	434.8	(372.7-473.0)	243.3	(203.0-278.0)	5677.7	(4809.0-6707.5)	528.2	(102.0-1063.3)	143.7	(39.0-260.0)	0.3	(0.0-2.0)
HiG	154.1	(125.6-189.4)	64.6	(53.9-82.7)	1777.5	(1333.5-2233.9)	186.9	(108.1-325.5)	51.2	(8.2-89.5)	0.8	(0.0-4.0)
HiHe	266.0	(234.8-301.7)	169.4	(142.5-187.1)	4419.1	(3856.4-5534.7)	217.2	(1.8-466.6)	73.5	(27.6-124.6)	0.0	(0.0-0.0)
HiL	166.1	(127.6-193.1)	105.9	(89.2-123.7)	3161.4	(2474.1-4235.9)	307.8	(99.5-475.7)	95.4	(57.7-132.1)	0.0	(0.0-0.0)
UiO	3240.6	(2976.2-3393.8)	2535.3	(2346.6-2645.6)	9538.3	(9016.4-10559.9)	17347.6	(16345.8-18244.5)	3464.1	(2839.5-4064.0)	392.5	(266.0-524.0)
UiB	1939.3	(1655.2-2079.5)	1297.0	(1183.7-1400.2)	6046.3	(5638.4-6518.8)	10979.1	(10009.4-11639.8)	1776.1	(1441.2-2048.2)	215.9	(157.0-265.0)
HiFm	1483.7	(1322.2-1589.8)	1078.9	(934.3-1187.2)	6023.7	(5403.5-6445.9)	5366.3	(4487.2-6873.4)	888.6	(496.1-1163.6)	95.2	(60.0-123.0)
NTNU	2786.4	(2447.3-3173.3)	1736.7	(1532.2-1990.9)	6114.1	(5814.7-6493.9)	16557.9	(13801.4-19400.8)	2325.6	(1247.5-3180.3)	282.3	(191.0-374.0)
HiS	592.4	(505.9-684.4)	350.8	(259.1-434.5)	5927.1	(5602.4-6178.0)	1697.5	(986.9-2725.6)	385.1	(122.5-558.4)	20.6	(3.0-34.0)

Continue Table A.1 Data

Short name	Faculty man-years		Admin and other man-years		Study points lower degree (weighted)		Study points higher degree (weighted)		Publishing points		PhDs	
	Mean	Min - Max	Mean	Min - Max	Mean	Min - Max	Mean	Min - Max	Mean	Min - Max	Mean	Min - Max
HiA	523.4	(467.3-570.6)	317.3	(271.3-366.5)	6453.2	(6095.4-7271.7)	1178.8	(792.7-1793.2)	317.0	(0.0-568.3)	7.2	(0.0-18.0)
HiBo	302.0	(262.8-330.0)	177.8	(138.2-228.6)	2981.7	(2732.5-3174.1)	718.0	(498.7-959.7)	123.0	(84.3-186.4)	7.4	(0.0-19.0)
NVH	718.3	(616.1-793.3)	634.0	(608.3-653.5)	1284.7	(1065.0-1614.6)	4198.4	(3344.8-5311.0)	524.8	(370.8-779.4)	75.5	(56.0-103.0)
NMH	124.3	(107.1-133.6)	45.4	(36.3-52.1)	1152.2	(1034.4-1351.9)	372.5	(289.7-500.3)	0.0	(0.0-0.0)	2.7	(0.0-6.0)
AHO	75.9	(61.3-90.4)	37.0	(29.5-44.7)	0.0	(0.0-0.0)	1455.2	(1161.4-1641.0)	0.0	(0.0-0.0)	5.1	(4.0-6.0)
NHH	223.7	(198.0-242.0)	127.0	(119.6-143.7)	1126.9	(363.4-1325.8)	1931.0	(1414.0-2462.4)	155.1	(130.4-188.3)	12.7	(8.0-19.0)
NIH	101.1	(79.1-115.2)	88.7	(81.2-98.2)	838.2	(629.0-1036.2)	229.6	(148.1-362.6)	99.5	(54.6-181.4)	8.1	(4.0-14.0)
KHiO	82.1	(73.1-90.0)	90.5	(79.4-100.5)	493.2	(457.8-565.3)	0.0	(0.0-0.0)	0.0	(0.0-0.0)	0.0	(0.0-0.0)
KHiB	41.1	(34.0-45.0)	36.4	(33.3-41.4)	300.3	(277.0-361.5)	0.0	(0.0-0.0)	0.0	(0.0-0.0)	0.0	(0.0-0.0)
DH	93.7	(70.9-105.2)	59.2	(43.9-67.3)	1498.1	(1300.3-1685.8)	225.2	(80.0-298.8)	38.3	(12.1-70.5)	0.0	(0.0-0.0)
LDH	50.6	(47.6-57.8)	20.1	(17.5-22.8)	719.0	(650.9-871.3)	5.4	(0.0-36.5)	8.2	(2.7-16.8)	0.0	(0.0-0.0)
HD	29.0	(19.2-41.0)	8.5	(4.0-15.6)	410.9	(349.9-535.7)	31.4	(0.0-78.0)	3.5	(0.0-7.5)	0.0	(0.0-0.0)
HB	17.1	(14.8-18.9)	10.3	(7.6-12.9)	290.6	(282.8-311.3)	0.0	(0.0-0.0)	2.1	(0.0-5.0)	0.0	(0.0-0.0)
HDH	21.8	(19.7-24.2)	10.1	(8.9-11.7)	331.4	(297.0-414.0)	0.0	(0.0-0.0)	6.3	(0.0-13.5)	0.0	(0.0-0.0)
MG	106.3	(91.8-117.9)	56.6	(51.5-60.7)	1370.9	(1220.0-1625.5)	138.2	(70.7-171.6)	39.0	(4.1-90.5)	0.0	(0.0-0.0)
DMMH	71.4	(60.2-87.6)	25.5	(20.8-32.7)	856.0	(733.6-987.3)	70.0	(25.1-128.9)	27.0	(1.4-62.4)	0.0	(0.0-0.0)
RS	14.7	(12.3-17.6)	5.6	(4.3-6.2)	132.0	(98.8-168.5)	26.1	(0.0-40.0)	1.4	(0.0-6.2)	0.0	(0.0-0.0)
EH	3.1	(2.0-4.1)	1.5	(0.6-3.2)	39.6	(18.7-60.9)	0.0	(0.0-0.0)	0.0	(0.0-0.0)	0.0	(0.0-0.0)
BDM	12.1	(8.2-17.9)	6.9	(4.6-9.7)	198.2	(157.4-230.4)	1.0	(0.0-9.6)	0.0	(0.0-0.0)	0.0	(0.0-0.0)
NDH	12.8	(10.5-17.0)	4.1	(1.0-7.0)	224.0	(91.3-281.7)	0.0	(0.0-0.0)	0.5	(0.0-4.5)	0.0	(0.0-0.0)
BI	319.4	(288.9-348.8)	379.7	(359.7-414.7)	7627.9	(6335.8-9625.2)	3205.4	(1939.8-3941.2)	181.7	(84.0-221.5)	8.5	(6.0-12.0)
HLB	5.2	(1.5-10.0)	2.1	(0.4-4.0)	29.1	(23.4-36.3)	0.0	(0.0-0.0)	0.1	(0.0-0.7)	0.0	(0.0-0.0)
HLT	5.3	(3.3-8.0)	2.8	(1.3-6.3)	58.4	(13.8-139.2)	0.0	(0.0-0.0)	1.6	(0.0-6.0)	0.0	(0.0-0.0)
NRH	27.0	(19.8-39.6)	23.2	(16.0-28.5)	1025.9	(403.5-1569.2)	9.6	(0.0-58.9)	16.8	(0.0-32.3)	0.0	(0.0-0.0)
NITH	24.1	(7.6-63.1)	16.0	(8.3-30.2)	570.5	(392.6-1081.6)	2.7	(0.0-15.5)	7.0	(0.0-14.7)	0.0	(0.0-0.0)

Table A.2 Decomposition of Malmquist index 2004-2013 with 95% confidence intervals

Unit	M	MC	MF
HiH	0.972 (0.908-1.017)	1.017 (0.940-1.103)	0.956 (0.837-1.029)
HiN	**1.451 (1.276-1.530)**	1.126 (0.908-1.295)	1.288 (0.990-1.475)
HiNe	**1.199 (1.141-1.265)**	**1.153 (1.033-1.232)**	1.040 (0.952-1.154)
SH	1.162 (0.796-1.225)	0.952 (0.655-1.034)	1.220 (0.757-1.385)
HiNT	**1.215 (1.155-1.314)**	**1.150 (1.064-1.249)**	1.057 (0.953-1.167)
HiST	0.973 (0.922-1.034)	**0.910 (0.836-0.980)**	1.069 (0.958-1.165)
HiB	1.049 (0.988-1.100)	1.073 (0.972-1.153)	0.977 (0.878-1.073)
HiM	**1.880 (1.790-2.094)**	1.302 (0.738-1.403)	**1.444 (1.330-1.906)**
HiSF	**1.158 (1.113-1.177)**	**1.228 (1.131-1.409)**	0.943 (0.770-1.005)
HSH	0.968 (0.889-1.043)	**0.895 (0.816-0.948)**	1.081 (0.977-1.207)
HiVo	**0.938 (0.901-0.951)**	**0.853 (0.788-0.962)**	1.100 (0.923-1.172)
HiÅ	**1.116 (1.059-1.196)**	1.012 (0.929-1.080)	**1.103 (1.011-1.218)**
HiT	**1.184 (1.118-1.229)**	**1.116 (1.011-1.227)**	1.061 (0.921-1.154)
HiØ	**1.152 (1.110-1.231)**	1.007 (0.921-1.064)	**1.144 (1.073-1.266)**
HiAk	0.970 (0.906-1.017)	**0.886 (0.811-0.958)**	1.094 (0.971-1.192)
HBu	**1.305 (1.186-1.367)**	**1.181 (1.056-1.267)**	1.104 (0.958-1.213)
HiG	**1.495 (1.380-1.632)**	**1.239 (1.031-1.345)**	**1.207 (1.075-1.417)**
HiHe	**1.213 (1.153-1.267)**	**1.154 (1.064-1.247)**	1.051 (0.938-1.133)
HiL	0.990 (0.978-1.016)	1.000 (0.921-1.308)	0.990 (0.564-1.069)
UiO	**1.292 (1.062-1.510)**	1.000 (0.687-1.336)	1.292 (0.688-1.621)
UiB	**1.204 (1.131-1.324)**	1.000 (0.785-1.325)	1.204 (0.693-1.444)
HiFm	**1.256 (1.167-1.295)**	0.991 (0.812-1.190)	1.268 (0.924-1.442)
NTNU	**1.480 (1.369-1.567)**	1.045 (0.806-1.279)	1.417 (0.992-1.677)
HiS	**1.283 (1.165-1.362)**	1.001 (0.851-1.089)	**1.282 (1.098-1.448)**
HiA	**1.506 (1.284-1.587)**	1.206 (0.974-1.388)	1.249 (0.895-1.391)
HiBo	**1.265 (1.181-1.366)**	0.963 (0.793-1.104)	**1.314 (1.078-1.527)**
NVH	**1.369 (1.324-1.388)**	0.977 (0.756-1.143)	**1.401 (1.107-1.654)**
NMH	**1.252 (1.193-1.323)**	0.928 (0.605-1.099)	**1.350 (1.043-1.709)**
NHH	0.969 (0.899-1.020)	1.000 (0.739-1.548)	0.969 (-0.227-1.152)
NIH	1.204 (0.561-1.216)	1.000 (0.335-1.272)	1.204 (0.135-1.399)
KHiO	**0.845 (0.827-0.856)**	**0.682 (0.492-0.734)**	**1.238 (1.120-1.494)**
KHiB	**0.728 (0.727-0.730)**	**0.609 (0.478-0.658)**	**1.196 (1.091-1.407)**
DH	**1.237 (1.159-1.288)**	1.116 (0.948-1.216)	1.108 (0.970-1.240)
LDH	1.054 (0.961-1.083)	**1.195 (1.096-1.304)**	**0.882 (0.749-0.940)**
HD	**1.880 (1.830-2.256)**	**1.456 (1.361-1.648)**	**1.291 (1.146-1.561)**
HB	**0.932 (0.856-0.991)**	1.012 (0.888-1.104)	0.921 (0.799-1.035)
HDH	**1.266 (1.152-1.328)**	**1.306 (1.113-1.415)**	0.970 (0.842-1.098)
MG	**1.590 (1.249-1.652)**	**1.436 (1.145-1.525)**	1.107 (0.835-1.267)
DMMH	**1.921 (1.523-2.164)**	1.323 (0.832-1.444)	**1.452 (1.076-1.819)**
RS	**0.761 (0.621-0.770)**	0.943 (0.822-1.039)	**0.807 (0.592-0.859)**
EH	**1.175 (1.070-1.308)**	**1.073 (1.010-1.223)**	1.095 (0.881-1.224)
BDM	0.965 (0.904-1.097)	**0.883 (0.833-0.994)**	1.093 (0.922-1.247)
BI	1.004 (0.804-1.087)	1.000 (0.283-1.486)	1.004 (0.072-1.379)
HLB	**0.570 (0.525-0.651)**	**0.689 (0.607-0.741)**	**0.828 (0.755-0.981)**
Average unit	**1.264 (1.226-1.302)**	1.019 (0.853-1.233)	1.241 (0.917-1.414)
Mean	**1.144 (1.037-1.219)**	1.003 (0.820-1.146)	1.138 (0.869-1.310)