Productivity development of Norwegian institutions of higher education 2004-2013

By

Edvardsen DF, Førsund FR, Kittelsen SaC.

The definitive publisher-authenticated and formatted version:

Edvardsen DF, Førsund FR, Kittelsen SaC., 2017, Productivity development of Norwegian institutions of higher education 2004-2013, Journal of the Operational Reseach Society, 1-17, DOI: 10.1057/s41274-017-0183-x.

is available at:

https://doi.org/10.1057/s41274-017-0183-x

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

Frisch Centre

Sverre A. C. Kittelsen,

Frisch Centre

This is a post-peer-review, pre-copyedit version of an article published in *Journal of the Operational Research Society*. The definitive publisher-authenticated version

> **Edvardsen DF, Førsund FR, Kittelsen SaC. 2017. Productivity development of norwegian institutions of higher education 2004–2013. Journal of the Operational Research Society: 1-17. DOI: 10.1057/s41274-017-0183-x**

is available online at: http://rdcu.be/pixg

---

* 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

Abstract: 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 varies a lot.

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

JEL classification: C18, C43, C61, D24, H52, I21

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, as remarked in Parteka and Wolszczak-Derlacz (2013) (see e.g. Worthington, 2001; Johnes, 2004; De Witte and López-Torres, 2015; the last paper providing a recent comprehensive review).

1.1.Literature review

Productivity change will be studied at the level of institutions of higher education. A lower level like a department is also interesting; especially for internal policy purposes, but for external policy purposes the institutional level is often warranted. In addition this level is the one in our available database. Papers employing Malmquist productivity change index used at the more disaggregated level of departments, studying either education or research separately, or having less than three time periods are excluded. In Table 1 we have entered some characteristics of nine papers fulfilling these criteria, focusing on choice of variables and overall results. It should be emphasized that each paper contains more analyses than the focus of the table. The selection of variables reflects data possibilities as well as limitation on the number of variables due to limited samples, but the selections give a good insight into the possibilities of specifications. Carrington et al (2005) provide a very interesting discussion of type of variables to include, and the papers in Table 1 all give practical choices. Several papers mention quality variables, and Carrington et al (2005) perform a second stage analysis where efficiency or productivity scores are regressed on such variables. However, in the present paper we will focus on the productivity analysis only, because second stage analysis is a demanding research topic in itself and has to be left for further research.

Labour is a dominating input in service production like higher education. A common specification is to distinguish between three categories with different functions; academic staff, administrative staff and technical personnel (Kempkes and Pohl, 2010; Margaritis and Smart, 2011). A recurring question is whether students are inputs or outputs (Worthington and Lee, 2008). Students represent the “raw material”, but it is not an input in the standard way inputs are defined in a service activity; students are present and something happens with their human capital, so it seems more logical to specify the increase in human capital, i.e. the transformation from an unpolished diamond to a polished one, as an output. A more traditional input is capital, disaggregated into buildings (m^2) and equipment. Non-labour operating expenditure can represent capital (Worthington and Lee, 2008; Margaritis and Smart, 2011).

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. Table 1 reveals that the latter type of output is not included in any of the studies. The output of teaching is the addition of human capital. It can be measured by the type of degrees that are awarded, from lower grade Bachelor degrees, to Master degrees and finally Ph.Ds. (Flegg et al, 2004; Johnes, 2008; Worthington and Lee, 2008). Obviously there is a question of the quality of the degrees. A basic part of research output is published research papers. A quality dimension is also important here.

Table 1. Research papers applying non-parametric Malmquist productivity index to HEIs

Authors/ Country	Period	No. of obs.	Inputs	Outputs	Method	Yearly % change Productivity growth (M) Catching-up (MC) Frontier shift (MF)		
						M	MC	MF
Flegg et al (2004) British HEIs	1980/81-1992/93	45	Staff Undergraduates Postgraduates Aggregate expend.	Income research, consultancies Undergraduate degrees Postgraduate degrees	Geometric mean of adjacent periods	3.6	0.7	2.8
Carrington et al (2005) Australian HEI	1996-2000	35	Operating costs	Weighted student load Weighted publications	Geometric mean of adjacent periods	1.8	-0.7	2.1
Johnes (2008) English HEIs	1996/97-2004/05		Academic staff, Admin. and central services expenditure First degree & other undergraduates Postgraduates	First-degree & other qualifications awarded Higher degree qualifications incl. Doctorates Income grants & contracts	Fixed base period frontier (first period) against each year, yearly geometric mean	1.1	-4.6	6.0
Worthington & Lee (2008) Australian HEIs	1998-2003	35	Academic staff, Non-academic staff Non-labour expendi-ture, Undergrad. students, Post-graduate students	Undergraduate completions Postgraduate completions Ph.D. completions Grants, Publication points	Geometric mean of adjacent periods	3.3	0.0	3.3
Kempkes & Pohl (2010) German HEIs	1998-2003	72	Technical personnel Research personnel Current expenditure	Graduates External research grants	Geometric mean of adjacent periods	1.4	2.5	-1.1
Edvardsen et al (2010) Norwegian HEIs	2004-2008	38	Total man-years	Study-points lower degree Study points higher degree Publishing points	Intertemporal benchmark envelopment on pooled data Circularity, Bootstrapping No decomposition	3.3		

Margaritis & Smart (2011) Australian (AU), New Zealand (NZ) HEIs	1997-2005	AU36 ZN8	Academic staff General staff Non-labour operating expenditure Students	Undergrad. qualifications Postgrad. qualifications Indexed articles	Geometric mean of adjacent periods		NZ	
						1.1	1.1	0.0
							AU	
						2.8	0.9	1.9
Parteka & Wolszczak-Derlacs (2013) HEIs seven European countries	2001-2005	266	Students Academic staff Revenue	No. of Graduates No. of Publications	Geometric mean of adjacent periods. Bootstrapping	4.1	3.2	1.2
Fernández-Santos & Martínez-Campillo (2015) Spanish HEIs	2002/03-2008/09	39	Academic staff Registered students Total revenue	Graduate students’ qualifications Research publications R & D revenue	Geometric mean of adjacent periods Bootstrapping	2.8	-0.8	3.7

A problem with degrees like Ph.Ds. and research publications is that the use of resources for producing these outputs may not be in the same period as the outputs are registered, but in earlier periods (Flegg et al, 2004; Carrington et al, 2005). External research grants are used in some studies as a proxy for research outputs (Johnes, 2008; Worthington and Lee, 2008; Kempkes and Pohl, 2010; Fernández-Santos and Martínez-Campillo, 2015). However, the rationale for doing this is questioned in some of the papers. Grants are used to acquire inputs, and the connection to research publication and quality may be unclear and also subject to time lags. External grants play a rather minor role on average for Norwegian HEIs and this variable will not be used as a proxy for research in this study.

Measures of quality are a common theme in the reviewed papers. 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. As to quality of faculty it may be measured by position and experience, and quality of students can be measured by the grade of students at start of studies. Quality of education could be measured by grades achieved for courses and degrees (Bachelor and Master), time to get the first job after finishing, 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 research funding. A problem on the output side regarding study points is that the analysis has 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 (Carrington et al, 2005; Edvardsen et al, 2010) 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 reduces the number of observations and did not influence the results that much. Therefore we have chosen not to use lags.

All the papers except one in Table 1 have decomposed the Malmquist productivity measure into catching up and frontier shift as presented in the last three columns. However, the interpretation of this decomposition is not as straightforward as expressed in the papers, see the next section. Most of the papers have also done a further decomposition of the catching-up measure into what is called "pure" (this choice of term is not explained) efficiency and scale efficiency, following Färe et al (1994a,b) combining constant returns to scale and variable returns to scale of the frontiers. However, it is not clear that this last decomposition gives a real insight into productivity development with the special mixing between the two scale properties. (Flegg et al, 2004, decompose the frontier shift into three terms; a “pure” technical efficiency, scale efficiency and congestion efficiency, the last term being difficult to give any economic meaning due to the lack of an uneconomic part of an efficient frontier, cf. Farrell, 1957, pp. 255-256).

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. Methods used for estimating efficiency scores, and bootstrapping are presented in Appendices 1 and 2, and data for all units for mean, minimum and maximum for the variables over the total period, and Malmquist index results for the period 2004-20013 with decomposition and confidence intervals, are set out in Appendix 3.

2. Methods

The bilateral Malmquist productivity growth index was developed for discrete time based on the ratio of distance functions for two units relative to the same frontier production function (Caves et al, 1982). The distance functions correspond to Farrell measures of efficiency 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 fully utilise 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.

The properties we will give the Malmquist productivity change index are not the standard ones used in the literature (Färe et al, 2008). Our approach is set out in Førsund et al (2015). There an envelopment of data specified as constant returns to scale (CRS) is used as the benchmark, in order to satisfy the homogeneity property of a productivity change index

(Grifell-Tatjé and Lovell, 1995). Furthermore, in order to satisfy circularity (Gini, 1931; Berg et al, 1992), a fixed technology is used as the benchmark. An *intertemporal frontier* (Tulkens and van den Eeckaut, 1995) is specified, i.e. all observations are pooled and used to estimate the benchmark envelope. The common use of taking geometric means of two adjacent years is not compatible with circularity (Førsund et al, 2015).

An illustration of our approach is provided in Figure 1. A variable returns to scale technology

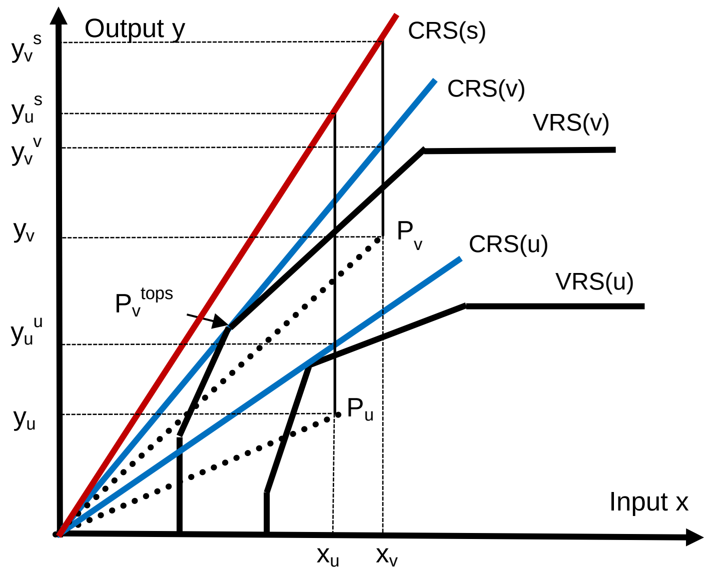


Figure 1. The Malmquist productivity change index.
Productivity change for a unit from period u to period v measured relative to the benchmark CRS(s) envelopment of the maximal productivity of the pooled dataset.

(VRS) is assumed for the contemporaneous technology shown by the frontiers VRS(v) and VRS(u). The productivity is maximal at optimal scale where the returns to scale is 1, termed a point of *technically optimal scale* (Frisch, 1965), illustrated by the point P_v^{tops} for VRS frontier for period *v*. Such points are then natural references 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 shift of the two VRS contemporaneous frontiers shows technological progress. The contemporaneous CRS benchmarks (blue) rays are tangents to the TOPS points.

The benchmark envelope is illustrated by the (red) ray CRS(s). There is a problem with changing reference frontiers over time as is a common practice (cf. Table 1). In the case of a fixed CRS reference envelopment for all units from all periods this means that technical productivity for all units and time periods refer to the same benchmark and not to different frontiers as in the adjacent frontier approach.

The estimator of the Malmquist index for a unit *i=1…n,* using the Farrell efficiency indices that correspond to the distance functions, for the two periods relative to the same frontier is (Førsund et al, 2015):

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

where superscript *s* symbolises that all data are used as the benchmark reference set. There is no orientation of the distance functions because when specifying a CRS benchmark envelopment input- and output orientations have identical scores. The Malmquist productivity estimator is conditional on the efficient border of the linear homogeneous envelopment set. The efficiency measures $\hat{E}_{iv}^s$ and $\hat{E}_{iu}^s$ in (1) are the Farrell technical productivity measures (the measure is 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* (Färe et al, 1992). 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

$$\hat{M}_i^s(u,v) = \frac{\hat{E}_{iv}^s}{\hat{E}_{iu}^s} = \frac{\hat{E}_{iv}^v}{\hat{E}_{iu}^u} \times \frac{\hat{E}_{iv}^s / \hat{E}_{iv}^v}{\hat{E}_{iu}^s / \hat{E}_{iu}^u} = \widehat{MC} \times \widehat{MF}, \quad i=1,..,J,\; u,v=1,..,T, u<v \qquad (2)$$

The 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 it has been assumed that the “true” period technology is VRS. As mentioned in Section 1 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, since mixing CRS and VRS assumptions is problematic (Kuosmanen and Sipiläinen, 2009, p. 140), 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 papers reviewed in Table 1, except Edvardsen et al, 2010, 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).

2.1. Bootstrapping

We are using the homogeneous bootstrap procedure outlined in Simar and Wilson, 1998; 1999; 2000). (For weaknesses with this bootstrap assumption see Olesen and Petersen, 2016.) Following Førsund et al (2015) 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. As in Førsund et al (2015) the Farrell output-oriented efficiency variable, distributed on (0,1], is chosen for the resampling (Efron, 1979). Pseudo replicate data sets (y_{imt}^{ps}) are created on the basis of the calculation of output-oriented efficiency scores for each output *m=1,…,M*, relative to the VRS frontier for each time period:

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

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.

Using these pseudo observations (x_i, y_i^{ps}) a new DEA frontier is then estimated. 2000 such draws was done and 2000 new DEA frontiers were established for each period. Going back to each run for a pair of periods, the Malmquist productivity index, given by (1), is calculated using the CRS 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. The mean square error of these bias-corrected scores may be greater than the mean square error of the uncorrected estimator (Simar and Wilson, 2000). 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. How to calculate the confidence intervals is shown in Simar and Wilson, 1999. The procedure is set out in Appendix 2.

3. Data and choice of model

When studying productivity the key to success is, first of all, to base the study on theoretically satisfactory definitions of inputs and outputs, and then to operationalise these definitions without compromising too much. The variables selected for our study are set out in Table 2. There are six variables used in our analysis; two inputs and four outputs. 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,

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 financial 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.

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 general rule in Norway is not to enrol more students than capacity allows).

However, we have not included quality variables discussed in Section 1.1, 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.

We see that no measures for the “third mission” are listed, and neither are quality variables. These variables are notoriously difficult to get measures for. 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.

It is difficult to assess the effect on productivity due to missing variables. However, we can try to conjecture cases where missing variables have typically different impacts on units. If higher quality means having to use more resources and quality variables are not specified, then units with a higher quality than the other units will tend to have lower efficiency scores. In the case of higher input qualities the impact on efficiency scores is the opposite. Enrolling better students without using extra resources to do this would increase efficiency, but if resources must be spent to attract such students the opposite may be the case. The same is the case if better quality faculty is attracted with or without using extra resources or with or without offering higher wages. However, productivity change is measured by the ratio of efficiency scores, so to speculate on the impact on productivity development is not so straightforward.

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 Fig. 2 on index form with the values in 2004 as the base. (See Appendix 3 for the individual average data for 2004 to 2013.) The two outputs publishing points and Ph.Ds. have had the most rapid growth with 88%

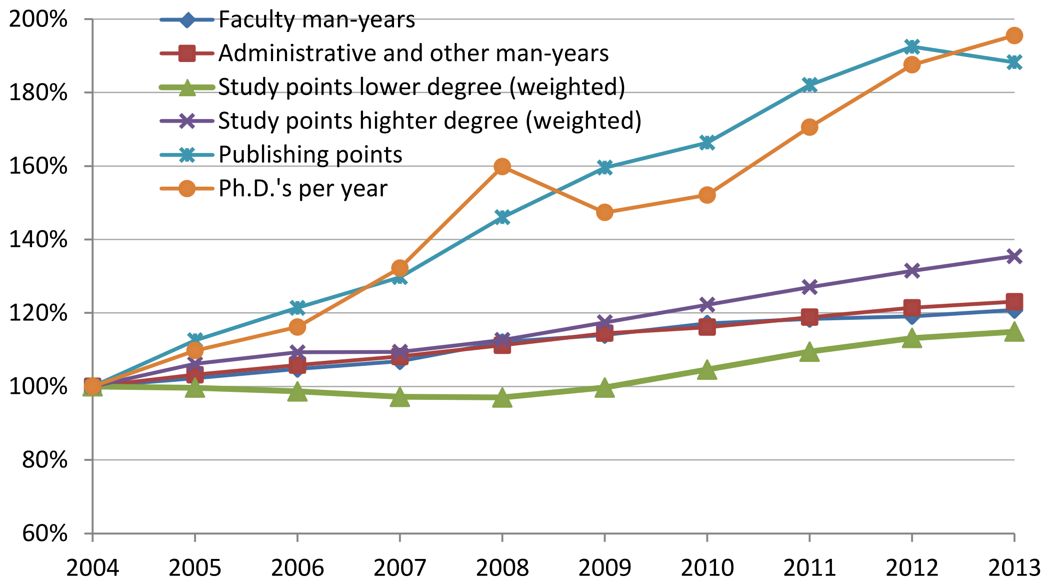


Figure 2. Development of the variables for the periods 2004 to 2013 relative to 2004
(See Table 2 for definitions of study points and publishing points)

and 96 % respectively. Of the two other outputs, weighted study points, the lower points have been growing most slowly with 15 % while the higher points have increased with 35 % . The two inputs have developed in 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

4.1. 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 growth from 2007 to 2008 with a positive jump in the productivity growth measured by the

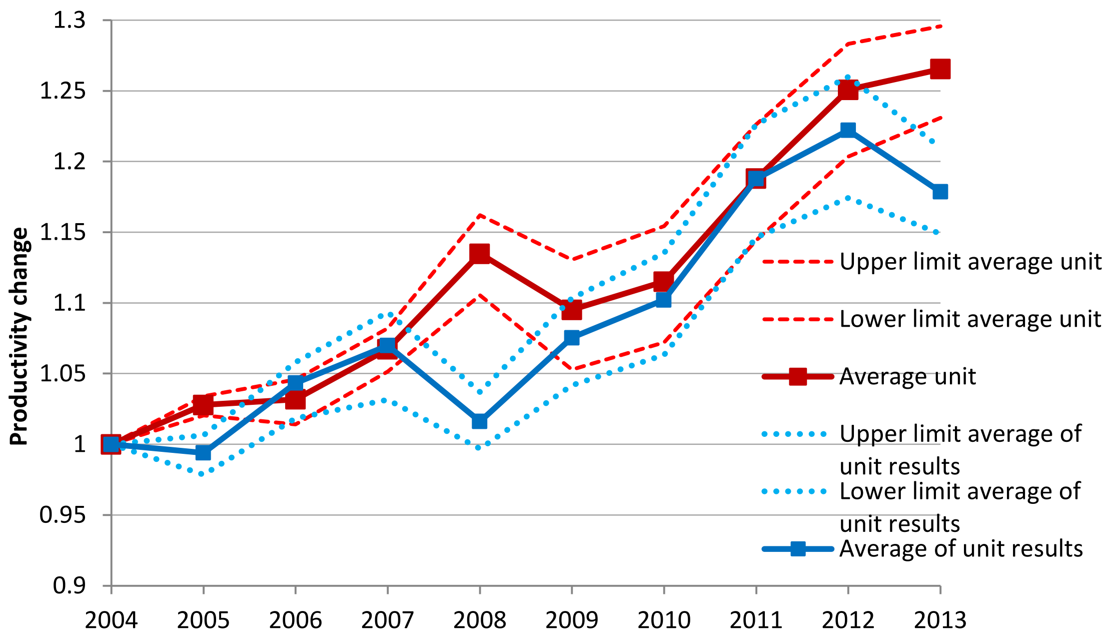


Figure 3. Aggregate productivity change (solid lines) for the periods 2004 to 2013 relative to productivity in 2004 measured by the average unit, and average of individual productivities with 95 % confidence intervals (broken lines).

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.

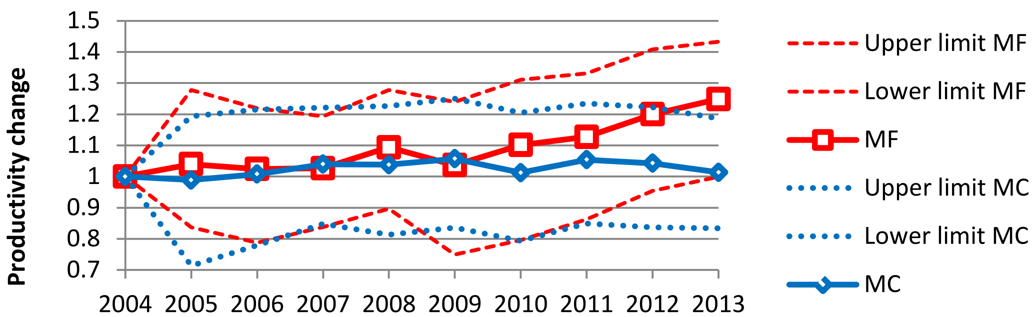


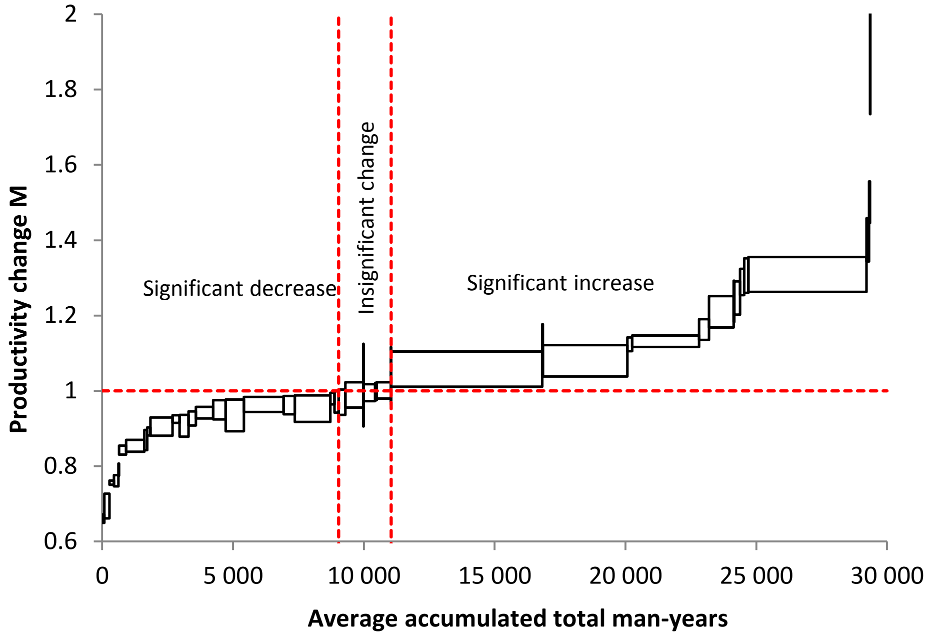
Figure 4. The decomposition of the Malmquist productivity index for the average unit into catching-up MC and frontier shift MF for periods 2004-2013 relative to 2004

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.

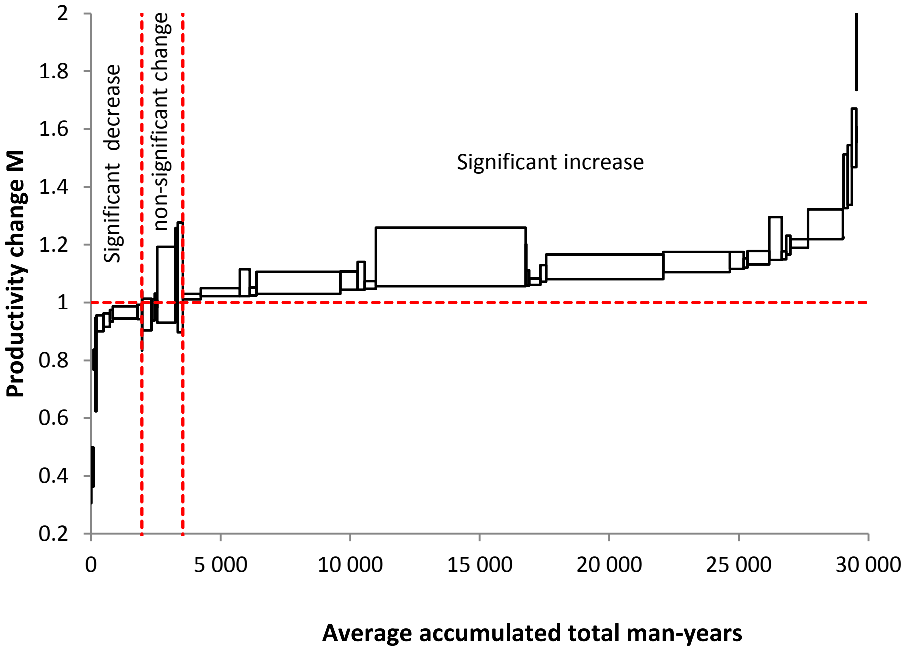
4.2. Productivity development of individual units

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.

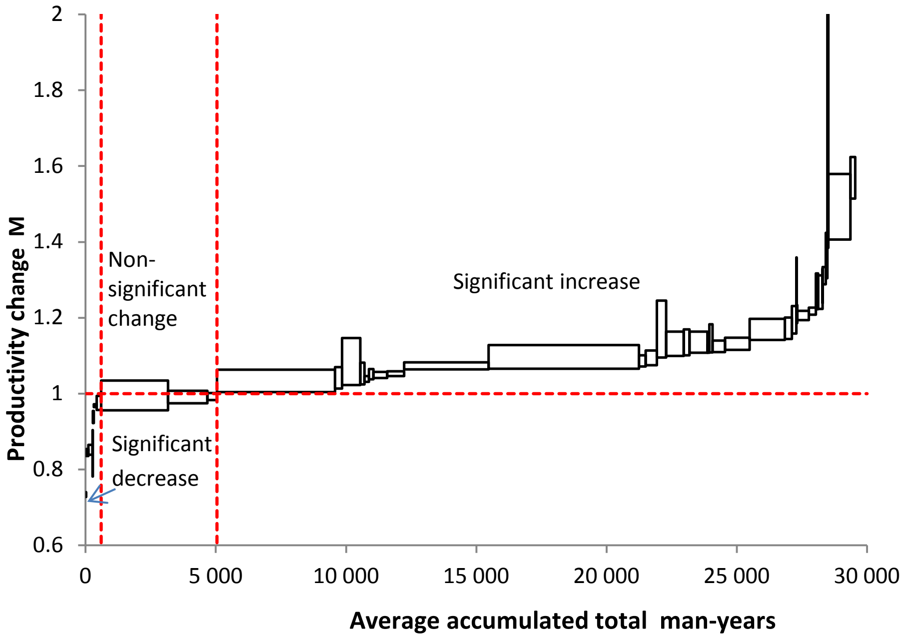
In Figure 5 four panels of productivity-change distribution for all the individual units are set out for three year periods, and the total period 2004 – 2013. (Due to perverse influence of the layout and readability of the diagrams a few units are not shown as indicated in the panel texts.) Each unit is represented by a box. The width of a box is the total man-years as an average for all years for ease of identifying the units over the periods.



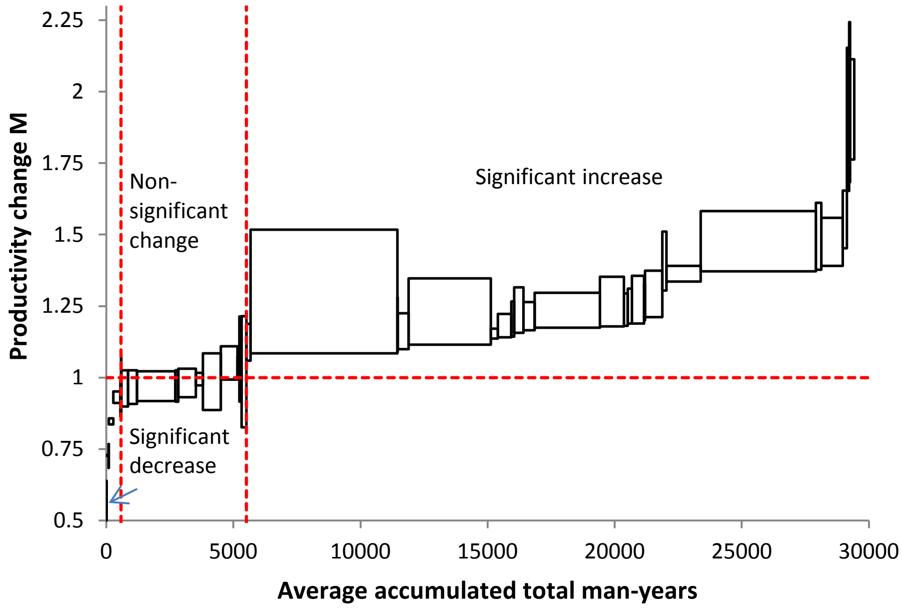
Panel (a) 2004-2007
(Two units with lowest and highest M, respectively, are not shown)



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



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



Panel (d) 2004-2013

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

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 or 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 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 last 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 hosts 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.

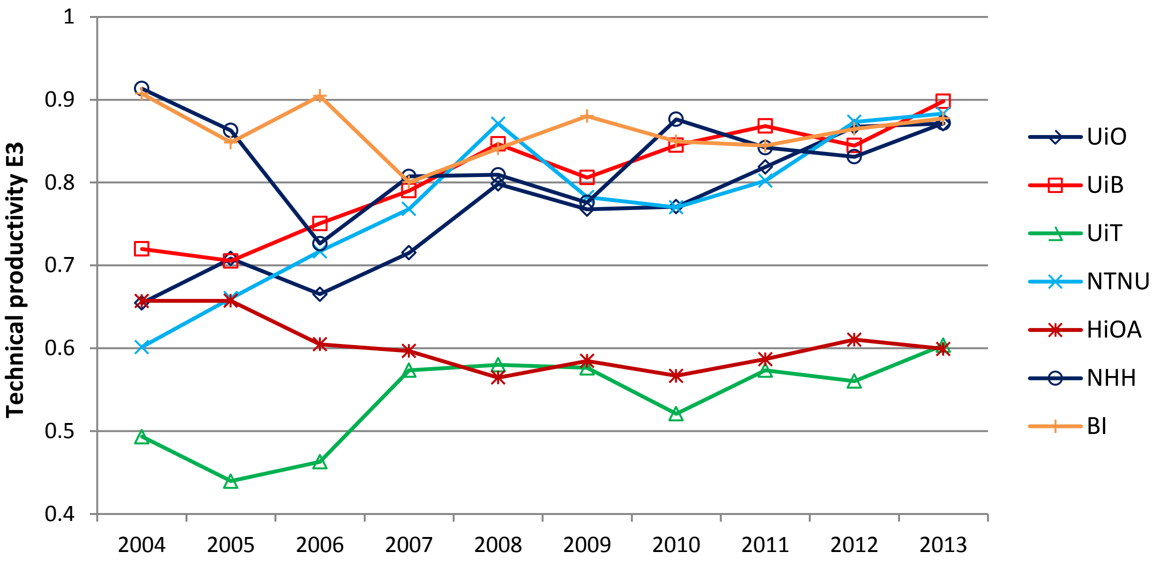
4.3. Decomposition of the productivity change

The decomposition results for the aggregate unit are representative of 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. 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 *caveat* about putting too much into attribution of the components as mentioned in Section 2.

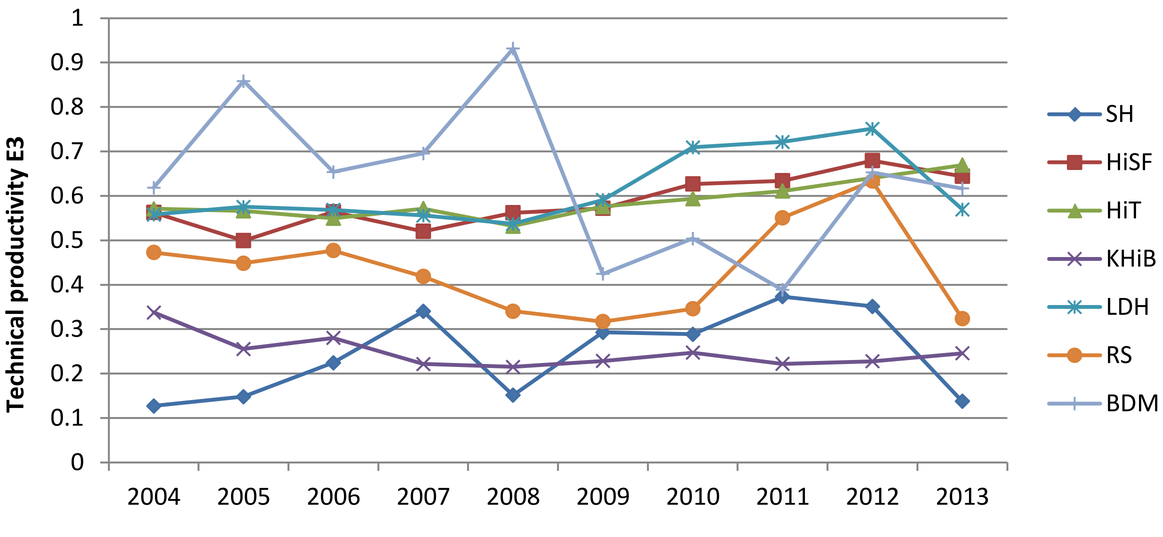
4.4. Productivity over time for sub-samples of selected large and small units

We will select some large and small units to follow more closely over time. The two panels of Figure 6 show the level of 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 has the lowest productivity level of all in 2004, then an increase in productivity level from 2005 until 2008, and then mixed productivity performance until the last period when it has its highest level of productivity. This means that over the period as a whole this university comes out with a positive productivity growth that is also significant. The productivity of the largest college HiOA falls from 2004 to 2008 to the level of UiT and then evens out ending up with a non-significant negative change for the whole period. A striking trend in the development of the other units is that there is some turbulence in productivity up to 2008, but then the developments become more alike 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 because these start out with high



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

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 behaviour 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 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.

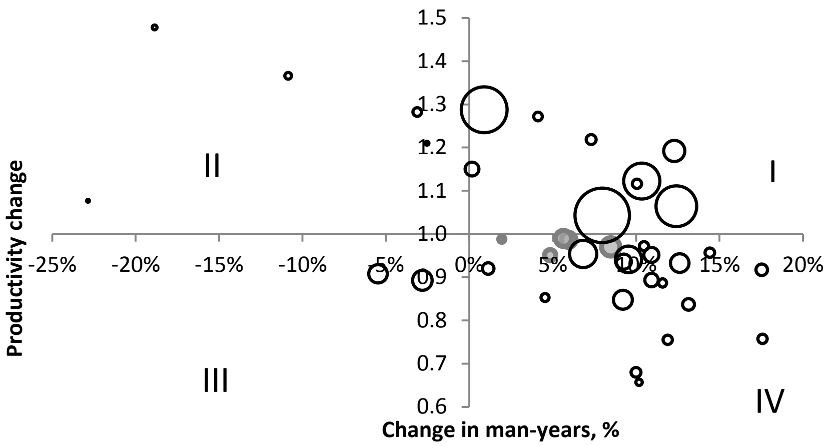
Panel (a) in Figure 6 shows that the large units have productivity levels all converging to around 90% in the last year, while Panel (b) shows that the small units have considerably lower productivity down to 10% and fluctuating a lot. This indicates that the small units are too small, being so far from the 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.

4.5. Change in productivity and resource use

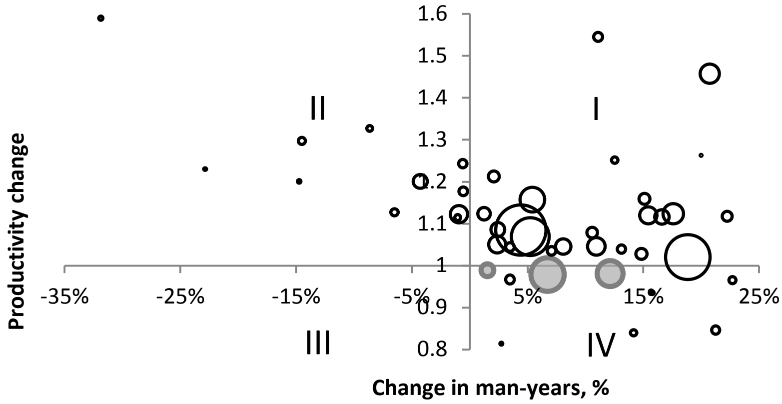
A recurrent policy question is the return on the resources allocated to higher education. Showing the change in total labour used together with productivity change provides some answers (Førsund et al, 2006). In Figure 7 productivity changes for the same periods as for Fig.5 are shown together with the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

The horizontal line at the value 1 delimits 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*. (See also Førsund and Kalhagen, 1999, where units in the quadrants II, III are termed having *positive* and *negative adjustment capability*, respectively).

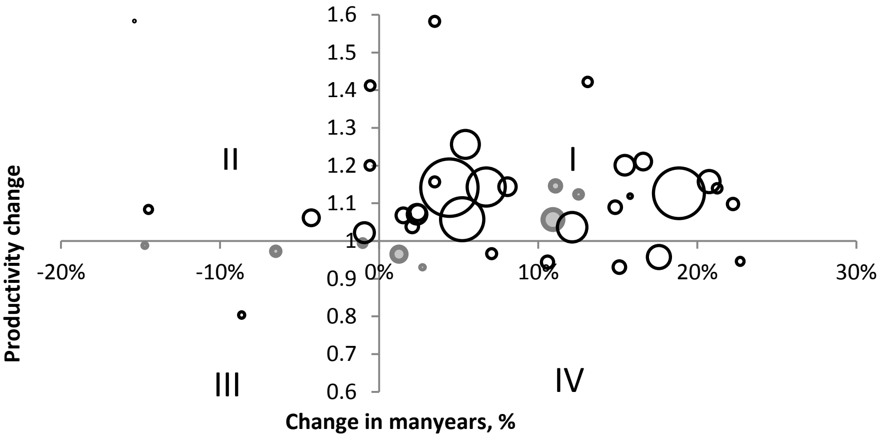
Due to a steady increase in labour for almost all units there are not many units in Quadrants II and III so Quadrants I and IV are the informative ones. (A few units with extreme changes have been removed in order to keep the diagrams visually interpretable.) A general feature for all periods is that the large units from Figure 6 are in Quadrant I with efficient expansion



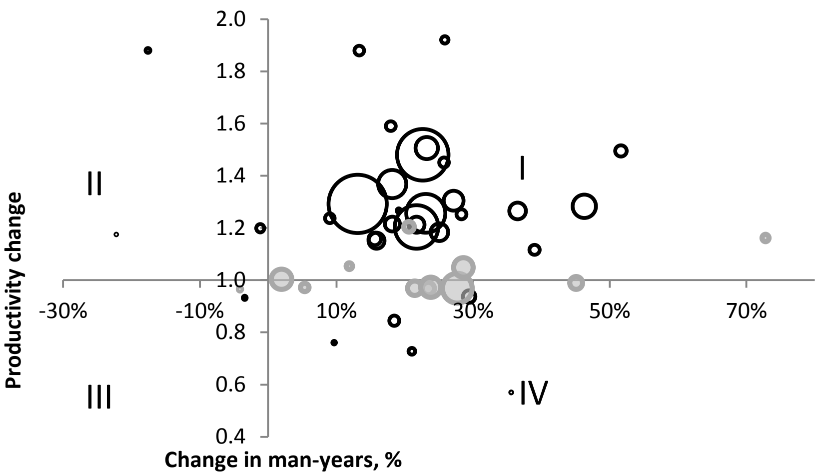
Panel (a) 2004-2007
(Three extreme units in QI and one in QIV are not shown)



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.

of labour. The total period in Panel (d) shows quite a variety in the labour increase without a clear positive correlation with productivity change. The increase in labour ranges 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 erratic 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 in the context of 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 more easily quantifiable such as faculty and other employees for resources, and study points, publication points and Ph.Ds. 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 1, 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 and Edvardsen DF (2015). Productivity of tax offices in Norway. *Journal of Productivity Analysis* **43**(3): 269–279.

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 (2004). Efficiency measurement. In: Johnes G and Johnes J (eds.). *The international handbook on the economics of education*. Edward Elgar: Cheltenham.

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.

Kuosmanen T and Sipiläinen T (2009). Exact decomposition of the Fisher ideal total factor productivity index. *Journal of Productivity Analysis* **31**(1): 137–150.

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.

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

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

The observation *it* is one of the *n* units in the panel 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

It is the pooled reference frontier that is bootstrapped and not the pair of period observations that are compared to this frontier. For each of the 2000 bootstrap iterations and for each period *t =1...T*, a new DEA frontier is estimated using the *i=1…n* pseudo observations (x_i, y_i^{ps}) we get from (3). Now, within each bootstrap iteration, the linear homogeneous technology created for the pooled set of all *Tn* pseudo observations is established as the benchmark. Still within each bootstrap iteration, Farrell efficiency indices are calculated using (A1) for each *actual* observation using the benchmark. Going back to each pair of periods, the Malmquist productivity index, given by (1), is calculated. The compared observations from period *t* and *t+1* are always the same actual observations of unit *i* in those periods. Since the pooled reference frontier estimate is based on a large number *Tn* of observations, the confidence intervals are expected to be narrower than they would have been for an annual frontier estimate with only *n* observations.

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:

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

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 centred 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)