Decomposing the productivity differences between hospitals in the Nordic countries

Sverre A.C. Kittelsen
Frisch Centre, Oslo.

Et al.

UNIVERSITY OF OSLO
HEALTH ECONOMICS
RESEARCH PROGRAMME

Working paper 2013: 4

HERO

Decomposing the productivity differences between hospitals in the Nordic countries

Sverre A.C. Kittelsen
Frisch Centre, Oslo.
Corresponding author, e-mail: sverre.kittelsen@frisch.uio.no

- Benny Adam Persson, Frisch Centre, Oslo, and Oslo University Hospital, Oslo
- Kjartan S. Anthun , SINTEF, Trondheim
- Fanny Goude, Karolinska Institutet, Stockholm
- Øyvind Hope, SINTEF, Trondheim
- Unto Häkkinen, National Institute for Health and Welfare – THL, Helsinki
- Birgitte Kalseth, SINTEF, Trondheim
- Jannie Kilsmark, Danish Institute of health research, Copenhagen
- Emma Medin, Karolinska Institutet, Stockholm
- Clas Rehnberg, Karolinska Institutet, Stockholm
- Hanna Rättö, National Institute for Health and Welfare – THL, Helsinki

Health Economics Research Programme at the University of Oslo
HERO 2013

Health Economics Research Programme at the University of Oslo
Financial support from The Research Council of Norway is acknowledged.
ISSN 1890-1735 (online), ISBN 978-82-7756-234-6

Decomposing the productivity differences between hospitals in the Nordic countries[*]

Sverre A.C. Kittelsen, Frisch Centre, Oslo. sverre.kittelsen@frisch.uio.no

Benny Adam Persson, Frisch Centre, Oslo, and Oslo University Hospital, Oslo

Kjartan S. Anthun , SINTEF, Trondheim

Fanny Goude, Karolinska Institutet, Stockholm

Øyvind Hope, SINTEF, Trondheim

Unto Häkkinen, National Institute for Health and Welfare – THL, Helsinki

Birgitte Kalseth, SINTEF, Trondheim

Jannie Kilsmark, Danish Institute of health research, Copenhagen

Emma Medin, Karolinska Institutet, Stockholm

Clas Rehnberg, Karolinska Institutet, Stockholm

Hanna Rättö, National Institute for Health and Welfare – THL, Helsinki

---

[*] We acknowledge the contribution of other participants in the Nordic Hospital Comparison Study Group (http://info.stakes.fi/nhcsg/EN/index.htm) in the collection of data and discussion of study design and results. During this study the NHCSG consisted of Mikko Peltola and Jan Christensen in addition to the authors listed. The data has been processed by Anthun, Kalseth and Hope, while Kittelsen and Persson have performed the DEA and SFA analysis respectively and drafted the manuscript. All authors have critically reviewed the manuscript and approved the final version. We thank the Norwegian board of health (www.hdir.no) and the Health Economics Research Programme at the University of Oslo (HERO – www.hero.uio.no), as well as the respective employers, for financial contributions. We finally thank the participants of the Conference in Memory of Professor Lennart Hjalmarsson in December 2012 in Gothenburg for helpful comments and suggestions.

Abstract

Previous studies indicate that Finnish hospitals have significantly higher productivity than in the other Nordic countries. We decompose the productivity levels into technical efficiency, scale efficiency and country specific possibility sets (technical frontiers). Data have been collected on operating costs and patient discharges in each DRG group for all hospitals in the Nordic countries. We find that there are small differences in scale and technical efficiency between countries, but large differences in production possibilities (frontier position). The results are robust to the choice of bootstrapped Data Envelopment Analysis (DEA) or Stochastic Frontier Analysis (SFA) as frontier estimation methodology.

JEL classification: C14, I12

Keywords: Productivity, Hospitals, Efficiency, DEA, SFA

1. *Introduction*

In previous studies (Kittelsen *et al.* 2008; Kittelsen *et al.* 2009; Linna *et al.* 2006; Linna *et al.* 2010) one has found persistent evidence that the somatic hospitals in Finland have a significantly higher average productivity level than hospitals in the other major Nordic countries (Sweden, Denmark and Norway)[2]. These results indicate that there could be significant gains from learning from the Finnish example, especially in the other Nordic countries, but potentially also in other similar countries. The policy implications could however be very different depending on the source of the productivity differences. This paper extends earlier work by, a) decomposing the productivity differences into those that stem from technical efficiency, scale efficiency and differences in the possibility set (the technology) between periods and countries, and b) exploring the statistical associations between the technical efficiency and various hospital-level indicators such as case-mix, outpatient share and status as a university or capital city hospital. Finally, c) we examine the robustness of the results to the choice of method.

International comparisons of productivity and efficiency of hospitals are few, primarily because of the difficulty of getting comparable data on output (Derveaux *et al.* 2004; Linna *et al.* 2006; Medin *et al.* 2013; Mobley and Magnussen 1998; Steinmann *et al.* 2004; Varabyova and Schreyögg 2013). Such analyses often find quite substantial differences in performance between countries. Differences may be due to the dissimilar hospital structures and financing schemes, but may also result from methodological problems.

[2] Although the Nordic countries also include Iceland, comparable data on Icelandic hospitals have not been available. In this article we will therefore use the term Nordic countries about the four largest countries.

Cross-national analyses are often based on data sets that only to a limited extent are comparable – in the sense that inputs and outputs are defined and measured differently across countries. It is, however, well known that the way we measure hospital performance may influence the empirical efficiency measures (Halsteinli *et al.* 2010; Magnussen 1996). In this article we will therefore use both the non-parametric data envelopment analysis (DEA) method and the stochastic frontier analysis (SFA) method, and provide evidence of the robustness of our results.

2. Methods

Efficiency and productivity

Efficiency and productivity are often used interchangeably. In our terminology productivity denotes the ratio of inputs and outputs, while efficiency is a relative measure comparing actual to optimal productivity. Since *productivity* is a ratio, it is by definition a concept that is homogenous of degree zero in inputs and outputs, i.e. a constant returns to scale (CRS) concept. This does not imply that the underlying technology is CRS. Indeed, the technology may well exhibit variable returns to scale (VRS), and equally efficient units may well have different productivity depending on their scale of operation, as well as other differences in their production possibility sets.

Most productivity indexes rely on prices to weigh several inputs and/or outputs, but building on Malmquist (1953), Caves *et al.* (1982) recognised that one can instead use properties of the production function, i.e. rates of transformation and substitution along the frontier of the production possibility set, for an implicit weighting of inputs and outputs.

This analysis departs from Farrell (1957) who defined (the input-oriented) *technical efficiency* as:

$$E_1^{T^{tc}} = \text{Min}\left\{\theta \middle| (\theta \mathbf{x}_i, \mathbf{y}_i) \in T^{tc}\right\} \tag{1}$$

Where $(\mathbf{x}_i, \mathbf{y}_i)$ is the input/output vector for an observation *i*, and T^{tc} is the technology or production possibility set for year *t* and country *c*. For an input/output-vector $(\mathbf{x}, \mathbf{y})$ to be part of the production possibility set, we need to be able to produce **y** using **x.** As shown in Färe & Lovell (1978), this is equivalent to the inverse of the Shephard (1979) input distance function.

If there are variable returns to scale, Farrell's measure of technical efficiency depends on the size of the observation, so that we can account for (dis)economies of scale. The measure of *technical productivity* can following Førsund and Hjalmarsson (1987) be defined by rescaling inputs and outputs[3]:

$$E_i^{\lambda T^{tc}} = \text{Min}_{\theta,\lambda}\left\{\theta \middle| (\theta \mathbf{x}_i, \mathbf{y}_i) \in \lambda T^{tc}\right\}, \tag{2}$$

where the the convex cone of the technology λT^{tc}, contains all input-output combinations that are a proportionate rescaling of a feasible point in the technology set T^{tc}. While this is formally identical to a "CRS technical efficiency" measure, our definition

---

[3] Førsund, F. R. and L. Hjalmarsson (1987), *Analyses of industrial structure: A putty-clay approach.*, Almqvist & Wiksell International, Stockholm. used the symbols e_1 for input technical efficiency, as did Farrell, M. J. (1957), The measurement of productive efficiency., *Journal of the Royal Statistical Society*, ***120***, 253-281., and e_3 for technical productivity which they call "overall scale efficiency".

here is instead that the reference surface is a homogenous envelopment of the underlying technology. This is the same assumption normally used in Malmquist indices of productivity change, see e.g. Grifell-Tatjé and Lovell (1995).

Furthermore, it is not necessary to assume that the technologies of different countries and time periods are identical in order to compare productivity, as long as one has a common reference set. It is common to use a specific (base) time period as a reference, as in Berg *et al.* (1992):

$$M_{ij}^{tc} = \frac{E_i^{\lambda T^{tc}}}{E_j^{\lambda T^{tc}}}, \tag{3}$$

which compares the productivity of two observations *I* and *j* using a fixed time period *t* as the reference, even if the observations *I* and *j* are from different time periods. A widespread alternative method is to construct geometric averages of indices based on consecutive time periods, as in Färe *et al.* (1994), which avoids the arbitrary choice of reference period *t*, but instead introduces a circularity problem. The approach followed here is instead to use information from all time periods for the country specific productivity reference:

$$T^c = \underset{t}{\mathrm{Env}}\left(T^{tc}\right) \tag{4}$$

where Env() is the convex envelopment of the time specific technologies. Furthermore, to compare the productivity across countries we will need the envelopment of all time and country specific technologies:

$$\overline{T} = \underset{c}{\mathrm{Env}}\left(T^c\right) \tag{5}$$

The reference sets (4) and (5) are not themselves technologies, only envelopment of technologies, as are the convex cones (rescaled sets) $\lambda T^{c}, \lambda \overline{T}$. Analogous to (2), it is then possible to define the productivity levels relative to the country specific references and the pooled references as $E_i^{\lambda T^C}$ and $E_i^{\lambda \overline{T}}$ respectively.

The country c specific Malmquist index of productivity change over time can then be defined as

$$M_{ij}^{c} = \frac{E_i^{\lambda T^c}}{E_j^{\lambda T^c}}, \qquad (5)$$

which normally is reported for two observation *i* and *j* of the same unit at two points in time. In this analysis we are primarily concerned with comparing observations from different units in different countries, and there is no natural pairing of *i* and *j.* Edvardsen and Forsund (2003) develop and report geometric means of Malmquist indices between a unit in one country and all units in another country. We will instead take a simpler approach and report the productivity and efficiency *levels* of each unit and their country means.

Decomposition

As discussed e.g. in Fried *et al.* (2008), the Malmquist index can be decomposed in various ways, where the original decomposition is into frontier shift and efficiency change. When working in productivity and efficiency levels, the starting point is instead the decomposition of technical productivity into technical efficiency and scale efficiency:

$$E_i^{\lambda T^{tc}} = E_i^{T^{tc}} \frac{E_i^{\lambda T^{tc}}}{E_i^{T^{tc}}} = (TP_i = TE_i * SE_i), \qquad (6)$$

where the parenthesis denotes the conventional way of writing this relationship. By including the possibility of comparing productivity across both time and countries, this decomposition naturally expands into:

$$E_i^{\lambda \bar{T}} = E_i^{T^{tc}} \frac{E_i^{\lambda T^{tc}}}{E_i^{T^{tc}}} \frac{E_i^{\lambda T^{c}}}{E_i^{\lambda T^{tc}}} \frac{E_i^{\lambda \bar{T}}}{E_i^{\lambda T^{c}}} = (TTP_i = TE_i * SE_i * PP_i * CP_i), \qquad (6)$$

where we have decomposed the now *total technical productivity (TTP)* into technical efficiency (TE), scale efficiency (SE), period productivity (PP) and country productivity (CP). Each of these is specific to the observation *i*[4].

---

[4] Note that dividing this decomposition for two observations of one unit at different points in time, and ignoring the country productivity, one gets the common Malmquist decomposition of technical efficiency change, scale efficiency change and frontier change.

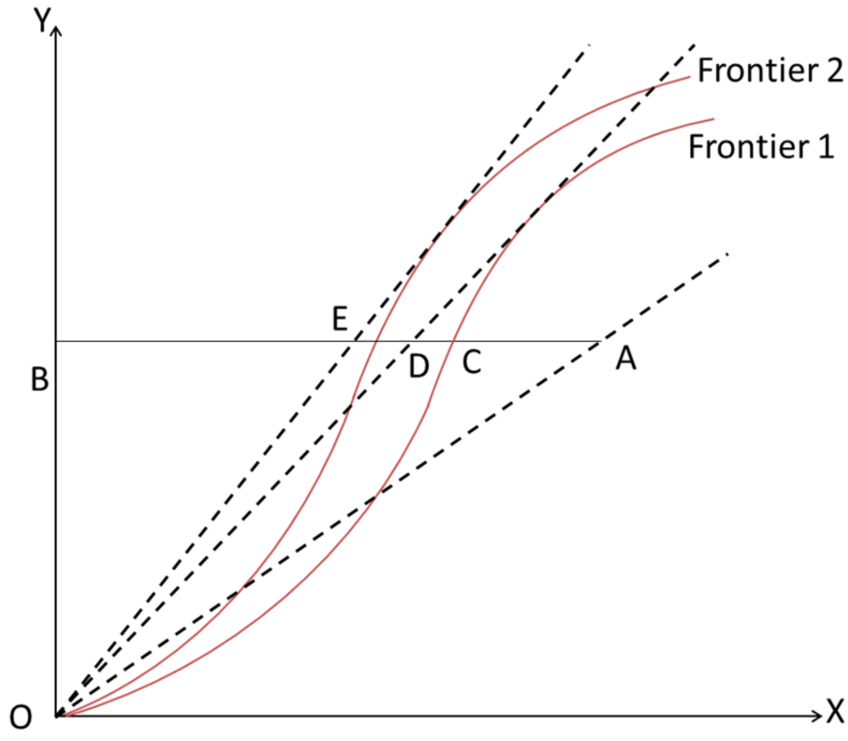

Figure 1. The components of hospital total technical productivity in input-output space. For observation A in country 1, TTP (Total technical productivity) =BE/BA, TE (Technical efficiency) = BC/BA, TP (Technical productivity) = BD/BA, Scale efficiency (SE) = BD/BC and CP (Country productivity) = BE/BD.

These concepts are illustrated in Figure 1, where we ignore the time dimension and concentrate on country differences. For an observation A in country 1 with a production possibility set bounded by the production function Frontier 1, we can define the technical efficiency by (1) above as the ratio BC/BA of necessary inputs to actual inputs for a given output. The productivity of A is the slope of the diagonal OA, but we can normalise this in (2) by comparing it to the maximal productivity given by the slope of the diagonal OD. The technical productivity of A is then the ratio BD/BA. Using the definition implicit in (6), scale efficiency is BC/BD. Assume that country 2 has a production possibility set bounded by Frontier 2, and that the maximal productivity of country 2 given by the slope OE is also the maximal for all countries, i.e bounding the convex cone of all possibility sets $\lambda \overline{T}$. This slope OE will serve as the reference for the total technical productivity in (8), which for observation A is given by BE/BA. The country productivity for observation A is then the ratio BE/BD.

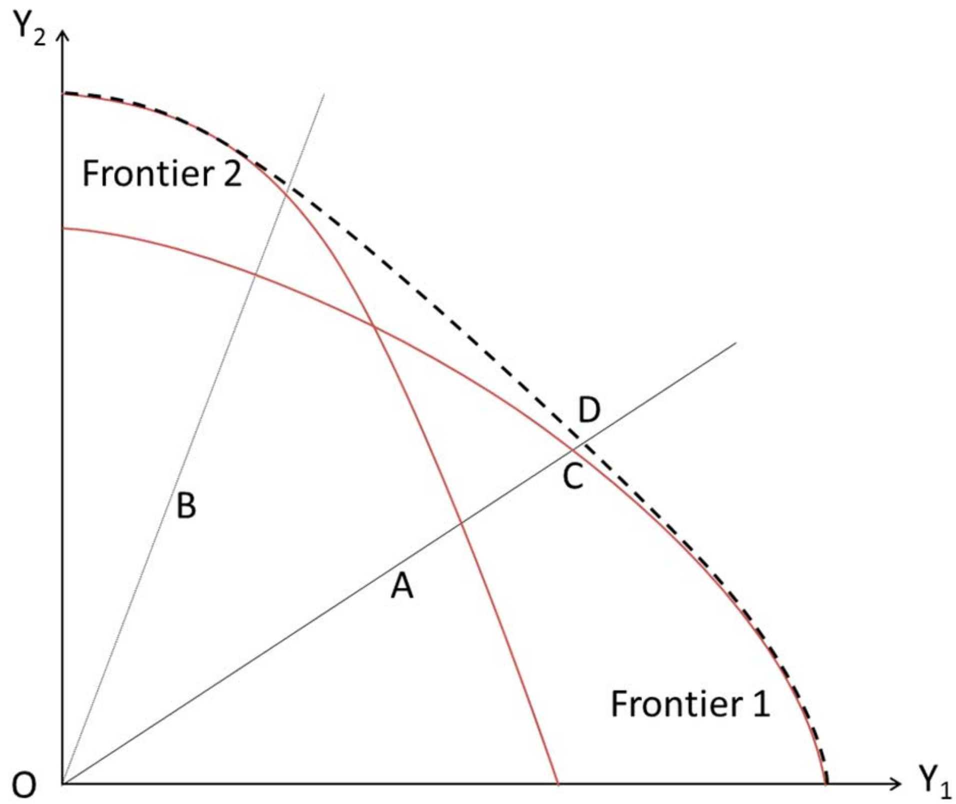

Figure 2. The components of hospital total technical productivity in output-output space. For observation A in country 1, TTP (Total technical productivity) =OA/OD, TP (Technical productivity) = OA/OC, and CP (Country productivity) = OC/OD.

With only one input and one output as in Figure 1, one country will define the reference and all observations in each country will have the same country productivity. With two outputs as in Figure 2, the convex cone of each country's frontier λT^{C} can be drawn as the curved lines for a given level of the single input. The convex cone of all the country frontiers $\lambda\bar{T}$ is represented by the dashed line which serves as the reference for total technical productivity defined in (8). If the country frontiers cross as in this example, the country productivities will depend on the output mix of the observation.

Cost efficiency and productivity

Finally note that since we have only one input in our data, cost minimization for a given input price is formally equivalent to input minimization. Thus cost efficiency, which is defined as the ratio of necessary costs to input costs, is also equivalent to technical efficiency. The decomposition of productivity and the Malmquist index is most often shown in terms of technical efficiency and technical productivity but could easily have been developed in terms of cost efficiency and cost productivity. Note that in the general multi-input case the numbers will differ in technical and cost productivity decompositions, but in our one-input case, the actual numbers will be identical. Thus, we may view the terms technical efficiency and cost efficiency as equivalent in discussing the results in this analysis.

Estimation method

The DEA and SFA methodologies build upon the same basic production theory basis. In both cases one estimates the production frontier (the boundry of the production possibility set or technology) or the dual formulation in the cost frontier, but the methods are quite different in their approach to estimating the frontiers and in the measures that are easily calculated and therefore commonly reported in the literature (Coelli *et al.* 2005; Fried *et al.* 2008). While the major strengths of DEA has been the lack of strong assumptions beyond those basic in theory (free disposal and convexity) and the fact that the frontier fits closely around the data, SFA has had a superior ability to handle the prescense of measurement error and to perform statistical inference. The latter shortcoming of DEA has been allieviated somewhat with the bootstrapping techniques introduced by Simar and Wilson (1998); (2000).

In our data there are good reasons to choose either method. While the prescense of measurement error is probably limited for those activities that are actually measured, there is a strong case for omitted variable (i.e. quality) bias that may be more severe in DEA. The DEA method can easily estimate the country specific frontiers without strong assumptions, thereby making country differences dependent on the input-output mix, while the SFA formulation generally introduces a constant difference between country frontiers. The prescense of country dummies in SFA implies however, that information from other countries are used to increase the precision of the estimates and therefore the power of the statistical tests.

In the DEA analysis the frontiers have been estimated using the homogenous bootstrapping algorithm from Simar and Wilson (1998), while the second stage analysis of the statistical association of technical efficiency and the environmental variables has been conducted using ordinary least square (OLS) regressions. The SFA analysis has used the simultanous estimation of the frontier component and the (in)efficiency component proposed in Battese and Coelli (1995)[5].

3. Data

Inputs are measured as operating costs, which for reasons of data availability are exclusive of capital costs. The Swedish data is further limited by the lack of cost information at the hospital level, nescessitating the use of the administrative county ("landsting") level as the unit of observation. In addition, it was not possible to get ethical permission for the use of data for 2007 in Sweden.

---

[5] The DEA bootstrap estimations have been done in FrischNonParametric, while second stage regressions and the SFA analysis has been done in STATA 12, StataCorp (2011), *Stata: Release 12. Statistical Software*, StataCorp LP, College Station, TX.

Since we do not have data on teaching and research output, the associated costs are also excluded. Costs are initially measured in nominal prices in each country's national currency, but to estimate productivity and efficiency one needs a comparable measure of "real costs" that is corrected for differences in input prices.

To harmonize the cost level between the four countries over time we have constructed wage indices for physicians, nurses and four other groups of hospital staff, as well as one for "other resources". This removes a major source of nominal cost and productivity differences between the countries, a difference that can not be influenced by the hospitals themselves, nor by the hospital sector as a whole. The wage indices are based on official wage date and include all personnel costs, i.e. pension costs and indirect labour taxes (Kittelsen *et al.* 2009). The index for "other resources" is the purchaser parity corrected GDP price index from OECD. The indices are weighted together with Norwegian cost shares in 2007. Thus we construct a Paasche-index using Norway in 2007 as reference point. Note that this represents an approximation, the index will only hold exactly if the relative use of inputs is constant over time and country.

Outputs are measured by using the Nordic version of the Diagnosis related groups (DRGs). We define three broad output categories; inpatient care, day care and outpatient visits. Within each category patients are weighted with the Norwegian cost weights from 2007[6]. Outpatient visits were not weighted. Considerable work has gone into reducing problems associated with differences in coding practice, including moving patients between DRGs, eliminating double counting etc. The problem of DRG-creep, where hospitals that face

---

[6] From a common initial starting point, the Danish DRG system has diverged significantly from the other Nordic systems after 2002. Danish DRG-weights were used for the specific Danish DRG groups, while the level was normalized using those DRG-groups that were common in the two systems.

strong incentives to upcode from simple to more severe DRGs based on the number of co-morbidities has been reduced by aggregating these groups. In the DEA analysis this had the effect of reducing the mean productivity level of Norwegian hospitals by 2 percentage points while the other countries were not affected, presumably because activity based financing is a more entrenched feature in Norway.

In addition to the single input and the three outputs, we have collected data for some characteristics that vary between hospitals within each country or over time, and that may be associated with efficiency. These include dummies for university hospital status which may capture any scope effects of teaching and research. This must be effects beyond the costs attributed to these activities which are already deducted from the cost variable. University hospitals may also have a more severe mix of patients within each DRG-group. The main case-mix effect should presumably already be captured by the DRG weighting scheeme. Capital city hospitals may have a different patient mix due to the socio-economic composition of the catchment area, as well as shorter travelling times and a greater potential for daypatient or outpatient treatment. The case-mix index (CMI) is calculated as the average DRG-weight per patient, and may again capture patient severity if the average severity within each DRG-group is correlated with the average severity as measured by the DRG-system itself. The length of stay (LOS) deviation variable is calculated as the DRG-weighted average LOS in each DRG for each hospital divided by the average LOS in each DRG across the whole sample (i.e. expected LOS). Again this could capture differences in severity within each DRG group, but may also indicate excessive, and therefore inefficient, LOS. Finally, the outpatient share is an indicator of diffences in treatment practices across hospitals. These variables are collectively termed "environmental variables", although they are not always strictly exogenous to the hospital.

Some variables are missing from the list. In earlier studies, the extent of activity based financing (ABF) has been an important explanatory variable, but in the period covered by our dataset there has been too little variation in ABF within each country. If a variable is perfectly highly correlated with the country then it is not possible to statistically separate the effect from other country specific fixed effects. This also holds for structural variables such as ownership structure, financing system etc. Travelling time to hospital can be an important cost driver but is not included here due to lack of data[7]. Finally, no indicators of the quality of treatment have been available for this analysis.

Table 1 shows the distribution of hospitals between countries and summary statistics for the varibles in the analyses. When interpreting the size of the Swedish observations, remember that these are not physical hospitals but the larger administrative "Landsting" units. To a lesser extent, the Norwegian observations of health enterprises can also encompass several physical hospitals.

[7] While we do not have data for travelling time in Denmark, we have calculated the average travelling time for the catchment area of emergency hospitals. A separate analysis reported in Kalseth, B., et al. (2011), Spesialisthelsetjenesten i Norden. Sykehusstruktur, styringsstruktur og lokal arbeidsorganisering som mulig forklaring på kostnadsforskjeller mellom landene,*Rapport* SINTEF Health Services Research. indicate that travelling time can explain some of the cost differences between the Norwegian regions, but not a significant amount of the differences between countries.

Table 1. Descriptive statistics. Observation means and standard deviations.

Variable		Finland	Sweden	Denmark	Norway	Total
Observation type		Hospital	Landsting/ County	Hospital	Health enterprise	
Number of observations		96	40	105	75	316
Period		2005-2007	2005-2006	2005-2007	2005-2007	
Variables in production frontier function (deterministic part)						
Real Costs in billion NOK [#]		1,112	4,812	1,516	1,864	1,893
	St.Dev.	*1,563*	*5,178*	*1,167*	*1,248*	*2,488*
Outpatient visits		150 128	368 134	178 620	129 609	182 321
	St.Dev.	*170 646*	*445 542*	*125 012*	*70 008*	*212 219*
DRG Inpatients		22 516	65 262	22 517	31 447	30 047
	St.Dev.	*27 834*	*68 200*	*17 647*	*18 414*	*34 440*
DRG Daypatients		3 119	18 000	2 651	4 044	5 067
	St.Dev.	*4 092*	*18 207*	*2 028*	*2 532*	*8 576*
Variables in SFA efficiency part or DEA second stage (environmental variables)						
University hospital dummy (share)		0,156	0,250	0,381	0,200	0,253
	St.Dev.	*0,365*	*0,439*	*0,488*	*0,403*	*0,436*
Capital city dummy (share)		0,031	0,050	0,257	0,160	0,139
	St.Dev.	*0,175*	*0,221*	*0,439*	*0,369*	*0,347*
Case Mix Index DRG patients		0,848	0,655	0,915	0,918	0,862
	St.Dev.	*0,089*	*0,096*	*0,166*	*0,083*	*0,146*
Length of stay deviation		0,968	1,118	1,017	0,859	0,977
	St.Dev.	*0,092*	*0,111*	*0,193*	*0,082*	*0,156*
Outpatient share		0,841	0,731	0,865	0,773	0,819
	St.Dev.	*0,028*	*0,049*	*0,044*	*0,026*	*0,061*

4. Results

DEA results

In the DEA analysis, the total technical productivity level is calculated with reference to a homogenous frontier estimated from the pooled set of observations for all countries and periods. Figure 3 show that the considerable productivity superiority of the Finnish hospitals found in previous studies is also present and highly significant in this dataset. The other Nordic countries are in some periods significantly different from each other, but in general have a similar productivity level.

Figure 3 also shows a slight time trend towards declining productivity. However, the DEA bootstrap tests did not reject a hypothesis of constant technology across time periods. This implies that we can ignore the time dimension and report the simpler three-way decomposition

$$E_i^{\lambda\bar{T}} = E_i^{T^c} \frac{E_i^{\lambda T^c}}{E_i^{T^c}} \frac{E_i^{\lambda\bar{T}}}{E_i^{\lambda T^c}} = (TTP_i = TE_i * SE_i * CP_i), \qquad (7)$$

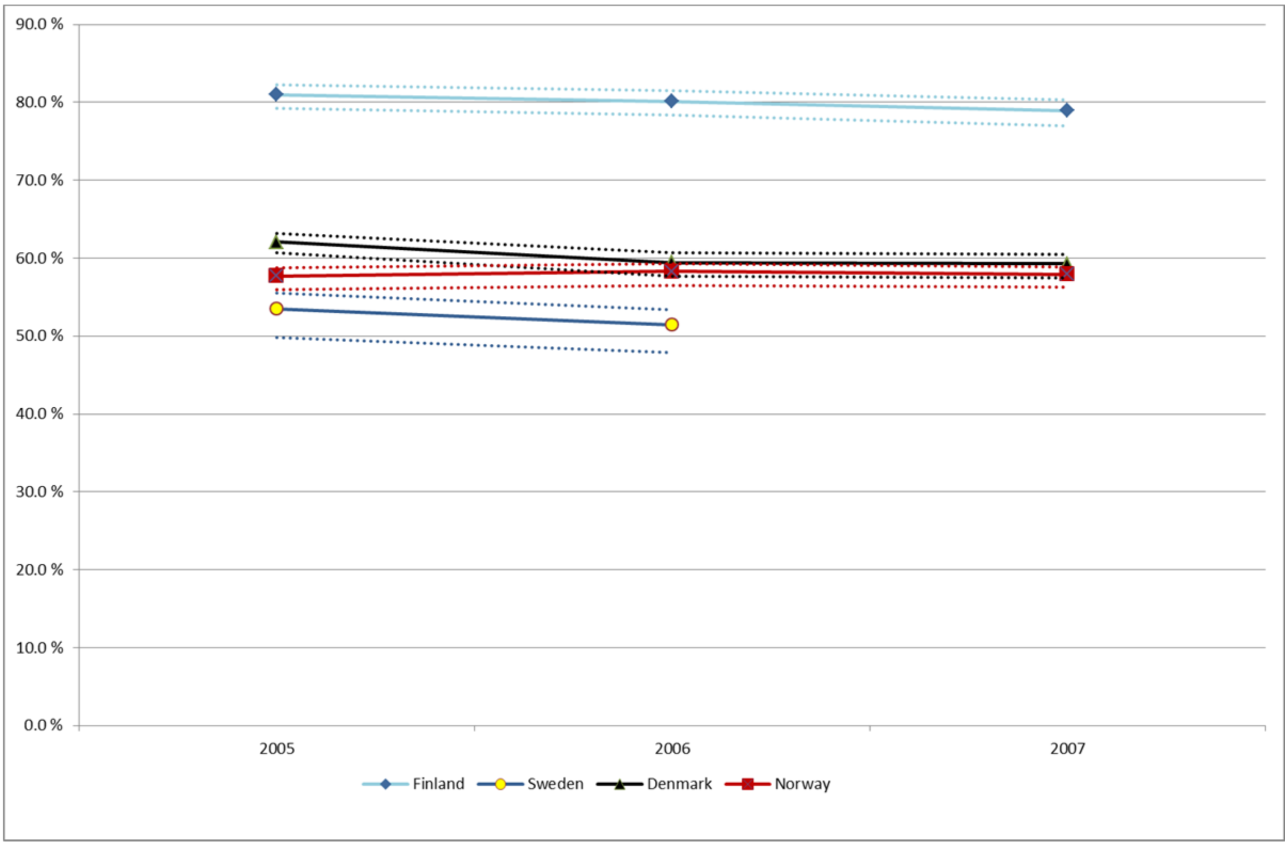

Figure 3. DEA productivity estimates by country and year with common reference frontier. Mean of observations and 95% confidence interval.

The productivity estimates for the individual observations are shown in figure 4. The hypothetical full productivity frontier is represented by productivity equal to 1.0, but since these numbers are bootstrapped estimates no observation is on the frontier. Clearly, the Finnish productivity level is consistently higher, with *all* Finnish observations doing better than most observations in Denmark and Norway and almost all in Sweden. Confidence intervals are quite narrow so this is a robust result. In all countries one can see that smaller units tend to be more productive, while comparisons between countries are confounded by the fact that the Swedish units are not hospitals but observations on the administrative "Landsting" level.

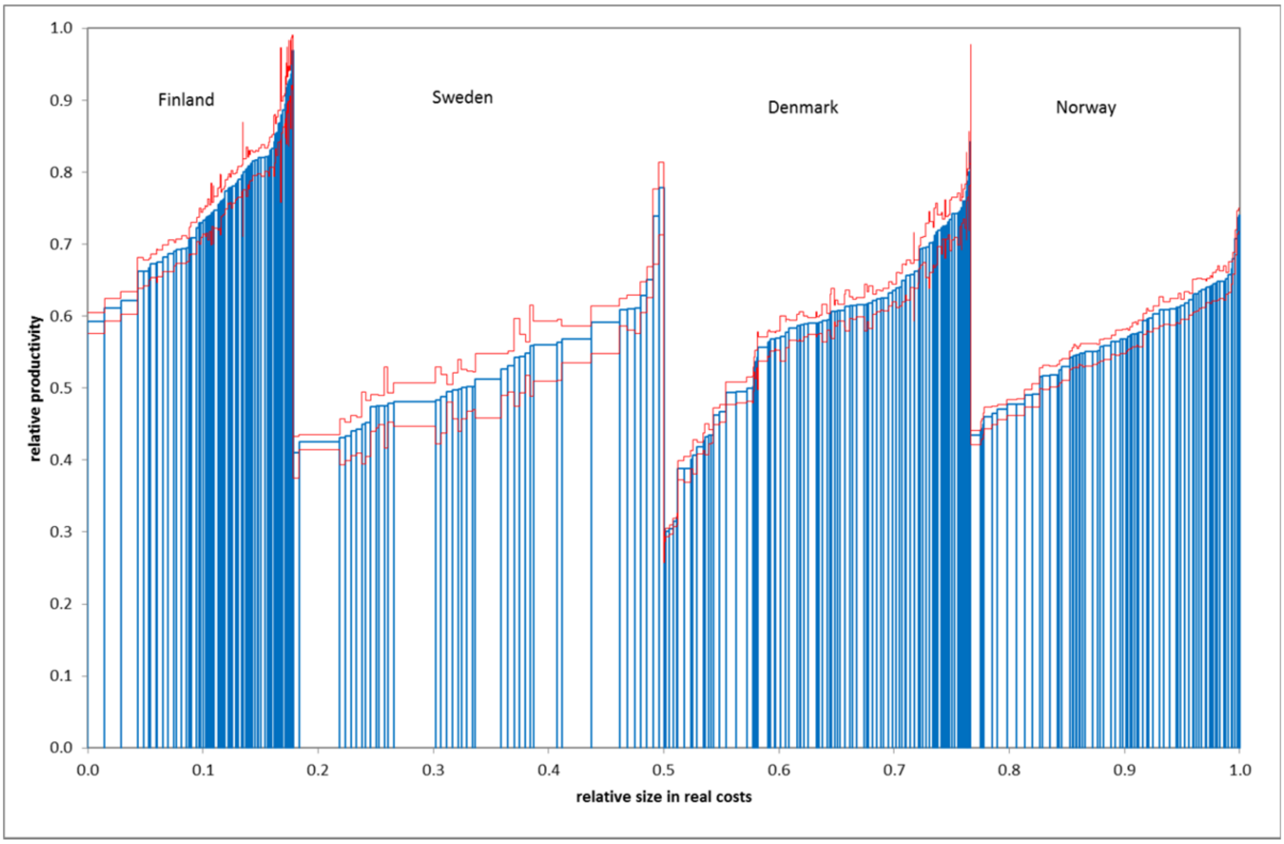

Figure 4. Hecksher-Salter diagram of DEA total technical productivity estimates with pooled common reference frontier. Height of each bar is productivity estimate for each observation with 95% confidence interval, and width is proportional to the observation size measured by real costs.

Table 2 reports the mean country productivity results and its decomposition. While Finland has an average productivity of around 80% measured relative to the pooled frontier, the decomposition reveals that this is wholly due to lack of scale efficiency and technical efficiency, which are at around 90% each. The country productivity mean is almost precisely 100%, which means that it is the Finnish hospitals that define the pooled reference frontier alone. For Sweden and Norway the picture is quite different; here the country productivity is the major component in the lack of total productivity. In fact, the cost efficiency and scale efficiency components are quite similar for Finland, Norway and Sweden. Denmark is in between, with significantly higher country productivity than Sweden and Norway, but still lagging far behind Finland. On the other hand, Denmark has clearly the lowest technical efficiency level of the Nordic countries.

Table 2. Mean productivity in each country as measured against the pooled reference frontier in DEA. Decomposition of total technical productivity into productivity of country specific frontier, scale efficiency and technical efficiency respectively. Although not part of this decomposition, the mean scale elasticity is given in the last line .

	Finland	Sweden	Denmark	Norway
Productivity with pooled reference frontier, TPP	79,1 %	52,6 %	57,7 %	56,6 %
	(77,0 - 81,0)	(49,8 - 54,2)	(55,4 - 59,6)	(53,0 - 58,6)
Decomposition of productivity				
Productivity of country specific frontier, CP	100,0 %	65,1 %	78,5 %	68,6 %
	(99,8 - 100,0)	(62,3 - 68,7)	(75,8 - 81,4)	(66,1 - 72,7)
Scale efficiency, SE	89,7 %	94,3 %	93,7 %	94,2 %
	(87,8 - 91,8)	(91,9 - 96,3)	(91,9 - 95,2)	(93,1 - 95,1)
Technical efficiency, TE	89,8 %	84,1 %	77,1 %	89,7 %
	(88,9 - 90,6)	(81,7 - 86,2)	(75,4 - 78,6)	(88,6 - 90,6)
Scale elasticity	0,935	1,137	0,940	0,941
	(0,917 – 0,956)	(1,000 – 1,255)	(0,911 – 0,982)	(0,884 – 0,982)

Geometric mean of productivity with 95% confidence interval for observations in each country.

Table 2 also reports the scale elasticities in the last line. Since the DEA numbers are based on separate frontier estimates for each country, the fact that the units are of a different nature represents no theoretical problem but must be reflected in the interpretation of the results. For Finland, Denmark and Norway, where the units are hospitals or low-level health enterprises, the scale elasticities below 1 indicate decreasing returns to scale on average, a result that is often found in estimates of hospital scale properties. Thus, optimal size is smaller than the median size. For Sweden, however, the scale elasticity is larger than one, although only just significantly. Thus, even though the units of observation are clearly larger in Sweden, the optimal size is even larger. The natural interpretation of this paradox is that while the optimal size of a hospital is quite small, the optimal size of an administrative region (or purchaser), such as the Swedish Landsting, is quite large. Of course, other national differences that are not captured by our variables may also explain this result.

Table 3. Simplified test tree in the SFA analysis

	Log-likelihood ratio	Critical value (degrees of freedom)	Result
Should country enter the frontier function?	287,952	7,05 (3)	Yes
Is Translog better than Cobb-Douglas?	42,892	11,91 (6)	Yes
Should year enter frontier function?	2,798	5,14 (2)	No
Should environmental variables enter efficiency term?	22,867	10,37 (5)	Yes
Should country enter efficiency term?	57,751	7,05 (3)	Yes
Should year enter efficiency term?	1,821	5,14 (2)	No

The Log-likelihood ratio indicator is distributed as χ^2 with degrees of freedom equal to the number of additional variables.

SFA results

The testing tree for the SFA model is shown in table 3. The formulation by Battese and Coelli (1995) implies that factors that determine the position of the frontier function in the deterministic part of the equation are estimated simultaneously as the variables in the "explanation" of the inefficiency term. Right hand side variables can potentially enter both components.

Clearly, the strongest result is that country dummies should enter the frontier term. This implies that there are highly significant fixed country effects that are not explained by any other variables, and that by the assumptions of the model specification the country dummy should primarily shift the frontier term. The functional form of the inefficiency term is not easily tested but the exponential distribution is the one that fits the data most closely. The functional form of the frontier function itself is, however, testable, and the simple Cobb-Douglas form is rejected in favour of the flexible Translog form. The time period dummies are also rejected in both terms, which means that the period can be ignored as in the DEA case.

The full estimation results for the preferred model is given in the appendix, but the normalized marginal effects are shown in table 4 together with the corresponding DEA results. The normalization is done so that a positive coefficient shows the percentage point increase in the productivity level (or decrease in costs) stemming from a one per cent increase in the explanatory variable. The frontier and efficiency terms are shown in separate columns. For the DEA results, the marginal effects are dependent on the input-output mix, and the numbers shown are for the average Norwegian observation.

Table 4. Marginal normalized effects on productivity in SFA and DEA, 95 per cent confidence intervals.

	SFA		DEA	
Parameter	Frontier (deterministic component)	Efficiency component	Frontier distance	Technical efficiency in second stage regression
Finland	0,300 *** (0,233 – 0,361)	0,049 (-0,085 – 0,183)	0,322 *** (0.295-0.370)	-0,029 (-0,083 – 0,025)
Sweden	0,071 (-0,020 - 0,154)	-0,024 (-0,085 – 0,037)	-0,021 (-0.068 – 0.061)	-0,004 (-0,072 – 0,064)
Denmark	0,208 *** (0,132 - 0,277)	-0,118 *** (-0,174 – (-0,062))	0,050 *** (0.010 – 0.094)	-0,160 *** (-0,246 – (-0,075))
Outpatient share		0,658 *** (0,259 – 1,057)		0,666 ** (0,014 – 1,319)
Length of stay deviation		-0,063 (-0,142 – 0,015)		-0,138 ** (-0,263 – (-0,013))
Case mix index		-0,048 (-0,160 – 0,064)		-0,064 (-0,204 – 0,075)
Capital city dummy		0,030 (-0,015 – 0,075)		0,040 (-0,021 – 1,101)
University hospital dummy		-0,010 (-0,049 – 0,029)		0,012 (-0,040 – 0,064)
Constant		-0,216 (-0,504 – 0,072)		0,533 ** (0,104 – 1,002)

Significant coefficients at 10, 5 and 1 per cent level respectively are marked with *, **, ***. Reference units are hospitals in Norway in 2007 that are not in the capital and not university hospitals. The reference unit in SFA has a technical efficiency estimate of 0.9176. In the DEA model the distance between the frontiers is measured at the average product mix of Norwegian hospitals

The results are generally very robust across method. The Finnish hospitals are strongly more productive than the other countries. The Swedish and Norwegian frontiers are not significantly different, while the Danish frontier is in between. In the efficiency term, the only significant country effect is that the Danish hospitals are less efficient. Of the environmental variables, the outpatient share has a significant positive effect on productivity while the LOS deviation has a weaker negative effect. The case-mix index and the dummies for university and capital city hospitals have no effect on costs so in these models there seems to be no sign that the central hospitals have a more costly case mix than is accounted for by the DRG system.

5. Conclusion

International comparisons can reveal more about the cost and productivity structure of a sector such as the somatic hospitals than a country specific study alone. In addition to an increase in the number of observations and therefore in the degrees of freedom, one gets more variation in explanatory variables and stronger possibilities for exploring causal mechanisms. This study has found evidence of a positive association between efficiency and outpatient share, a negative association with LOS, and no association with the case-mix index or university and capital city dummies. We have further found evidence of decreasing returns to scale at the hospital level, with a possibility of increasing returns to scale at the administrative or purchaser level. There is also evidence of cost/technical inefficiency, particularly in Denmark.

As so often, the strongest results are not what we can explain, but what we cannot explain. There is strong evidence, independent of method, that there are large country specific differences that are not correlated with any of our other variables. Finland is consistently more productive than the other Nordic countries. There are systematic differences between countries that do not vary between hospitals within each country. Without observations from more countries, or more variables that vary over time or across hospitals within each country, such mechanisms cannot be revealed by statistical methods. On the other hand, qualitative information can give some plausible explanations. Based on interviews of 8 hospitals in Nordic countries, the possible reasons for the Finnish good results are the good coordination between somatic hospitals and primary care including inpatient departments of health centres, a smaller number of personnel as well as better organization of work and team work between different personnel groups inside hospitals (Kalseth *et al.* 2011). However, these findings are still preliminary. An important policy question is whether the higher productivity in Finland is related to worsen quality.

What we can say is that the country productivity differences are consistent with possible differences in system characteristics that may vary systematically between countries. Such characteristics include the financing structure, ownership structure, regulation framework, quality differences, standards, education, professional interest groups, work culture, etc. Some of these characteristics, such as quality, may also vary between hospitals in each country and should be the subject of further research. Differences in estimated country productivity are also consistent with data definition differences, but the underlying analysis in Kalseth *et al.* (2011) does not support this. In summary, these country effects are essentially not caused by factors that can be changed by the individual hospitals acting in isolation to become more efficient, but rather factors that must be tackled by relevant organizations and authorities at the national level.

References

Battese, G. E. and T. J. Coelli (1995), A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, *Empirical Economics*, ***20***(2), 325-332.

Berg, S. A., et al. (1992), Malmquist indices of productivity growth during the deregulation of Norwegian banking 1980-1989, *Scandinavian Journal of Economics, Supplement*, ***94***, 211-288.

Caves, D. W., et al. (1982), The economic theory of index numbers and the measurement of input, output, and productivity, *Econometrica*, ***50***, 1393-1414.

Coelli, T. J., et al. (2005), *An introduction to efficiency and productivity analysis 2nd edition*, Springer Verlag.

Derveaux, B., et al. (2004), Comparing French and US hospital technologies: a directional input distance function approach, *Applied Economics*, ***36***, 1065-1081.

Edvardsen, D. F. and F. R. Forsund (2003), International benchmarking of electricity distribution utilities, *Resource and Energy Economics*, ***25***(4), 353-371.

Farrell, M. J. (1957), The measurement of productive efficiency., *Journal of the Royal Statistical Society*, ***120***, 253-281.

Fried, H. O., et al. (2008), *The measurement of productive efficiency and productivity growth*, Oxford University Press, Oxford.

Färe, R., et al. (1994), Productivity developments in Swedish hospitals; A Malmquist output index approach., in A. Charnes, et al. (eds.),*Data Envelopment Analysis: Theory, Methodology and Applications*, Kluwer Academic Publishers, Massachusets.

Førsund, F. R. and L. Hjalmarsson (1987), *Analyses of industrial structure: A putty-clay approach.*, Almqvist & Wiksell International, Stockholm.

Grifell-Tatjé, E. and C. A. K. Lovell (1995), A note on the Malmquist productivity index, *Economics Letters*, ***47***(2), 169-175.

Halsteinli, V., et al. (2010), Productivity growth in outpatient child and adolescent mental health services: The impact of case-mix adjustment, *Social Science & Medicine*, ***70***(3), 439-446.

Kalseth, B., et al. (2011), Spesialisthelsetjenesten i Norden. Sykehusstruktur, styringsstruktur og lokal arbeidsorganisering som mulig forklaring på kostnadsforskjeller mellom landene,*Rapport* SINTEF Health Services Research.

Kittelsen, S. A. C., et al. (2008), Hospital productivity and the Norwegian ownership reform - A Nordic comparative study,*STAKES discussion paper* STAKES.

Kittelsen, S. A. C., et al. (2009), En komparativ analyse av spesialisthelsetjenesten i Finland, Sverige, Danmark og Norge: Aktivitet, ressursbruk og produktivitet 2005-2007,*Rapport* SINTEF Health Services Research.

Linna, M., et al. (2006), Comparing hospital cost efficiency between Norway and Finland, *Health Policy*, ***77***(3), 268-278.

Linna, M., et al. (2010), Measuring cost efficiency in the Nordic Hospitals-a cross-sectional comparison of public hospitals in 2002, *Health Care Management Science*, ***13***(4), 346-357.

Magnussen, J. (1996), Efficiency measurement and the operationalization of hospital production, *Health Services Research*, ***31***, 21-37.

Malmquist, S. (1953), Index numbers and indifference surfaces, *Trabajos de estadistica*, ***4***, 209-224.

Medin, E., et al. (2013), International hospital productivity comparison: Experiences from the Nordic countries, *Health policy (Amsterdam, Netherlands)*.

Mobley, L. and J. Magnussen (1998), An International Comparison of Hospital Efficiency. Does Institutional Environment Matter?, *Applied Economics*, ***30***, 1089-1100.

Salvanes, K. G. and S. Tjotta (1998), A Note on the Importance of Testing for Regularities for Estimated Flexible Functional Forms, *Journal of Productivity Analysis. March*, ***9***(2), 133-143.

Simar, L. and P. W. Wilson (1998), Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models, *Management Science*, ***44***, 49-61.

Simar, L. and P. W. Wilson (2000), Statistical inference in nonparametric frontier models: The state of the art, *Journal of Productive Analysis*, ***13***(1), 49-78.

StataCorp (2011), *Stata: Release 12. Statistical Software*, StataCorp LP, College Station, TX.

Steinmann, L., et al. (2004), Measuring and comparing the (in)effciency of German and Swiss Hospitals, *European Journal of Health Economics*, ***5***.

Varabyova, Y. and J. Schreyögg (2013), International comparisons of the technical efficiency of the hospital sector: Panel data analysis of OECD countries using parametric and non-parametric approaches, *Health policy (Amsterdam, Netherlands)*.

Appendix Raw coefficients in SFA analysis

Table A.1. SFA-analysis with an exponentially distributed efficiency component. Dependent variable is total real costs in billion 2007 NOK. Reference units are hospitals in Norway in the year 2007, which is neither in the capital nor are university hospitals.

	Coefficient	Z-value	
Cost frontier (deterministic part)			
Constant	-10,349	-9,760	***
Ln Outpatients	0,410	0,930	
Ln DRG inpatients	0,093	0,260	
Ln DRG daypatients	0,500	2,090	**
(Ln Outpatients)* (Ln DRG inpatients)	-0,246	-2,030	**
(Ln Outpatients)* (Ln DRG daypatients)	0,030	0,510	
(Ln DRG inpatients)* (Ln DRG daypatients)	-0,240	-2,690	***
(1/2) (Ln Outpatients)^2	0,193	1,940	*
(1/2) (Ln DRG inpatients)^2	0,530	2,760	***
(1/2) (Ln DRG daypatients)^2	0,200	3,650	***
Finland	-0,356	-7,660	***
Sweden	-0,074	-1,540	
Denmark	-0,233	-4,970	***
Inefficiency part			
Constant	5,457	1.470	
Finland	-1,236	-0.720	
Sweden	0,610	0.770	
Denmark	2,984	4,120	***
Outpatient share	-16,634	-3,230	***
Length of stay deviation	1,602	1,580	
Case mix index	1,214	0,840	
Capital city dummy	-0,755	-1,290	
University hospital dummy	0,244	0,480	
Log likelihood		218,275	
Scale elastisity		0.928	
Gradient vector		2,31e-7	
Number of observations		316	
Number of regular observations		249	

Significant coefficients at 10, 5 and 1 per cent level respectively are marked with *, **, ***. In the inefficiency part, positive coefficients indicate reduced efficiency. Scale elasticity is calculated as in Coelli *et al.* (2005). Regularity conditions for the cost frontier part are as calculated in Salvanes and Tjotta (1998).