

Cost and efficiency in the Hospital Sector

A Comparative Study of Finland, Sweden, Denmark and Norway

Benny Adam Persson



Master Thesis
Department of Health Management and Health
Economics

UNIVERSITY OF OSLO

May 13, 2011

Are the differences in hospital costs between Finland, Norway, Denmark and Sweden, over time, due to differences in the production possibilities that are specific to country and period, or due to the distribution of efficiency behind the frontier?

© Benny Adam Persson

2011

Cost and efficiency in the Hospital Sector

Benny Adam Persson

<http://www.duo.uio.no>

Print: Reprosentralen, University of Oslo

Preface

It is expected that the reader has an understanding of general economic theory and common statistical knowledge, such as the ordinary least square method (OLS).

I would like to take this opportunity to thank my supervisor Sverre A.C. Kittelsen whose support, interest and guidance has been valuable in the work on this thesis. Thank you for being available and giving good advice along the way.

I would also like to thank my friend Søren T. Klitkou and my wife Eirin for their positive support. To everyone working at the Frisch center – I cannot think of a better place to write a Master's thesis – Thank you.

Benny Adam Persson

Oslo, May 2011

Abstract

Background: The Nordic countries' welfare state is under increased pressure. Health care expenditures are rising and expected to do so in the coming decades. It is therefore interesting to find out if the Norwegian specialist health care is cost efficient compared to its neighboring countries, or whether there are resources which could be allocated differently. The SINTEF report A12200 (Kittelsen et al., 2009a) was published in 2009. It found that Finland executes its specialist health care with higher productivity compared to Norway. This thesis builds upon the SINTEF report and extends it.

The objective: To find out if the differences in hospital costs between the countries, over time, are due to differences in the productivity of the best performing hospitals and periods, which defines the efficiency frontier, or due to the distribution of efficiency behind the frontier.

Method: The parametric method of Stochastic Frontier Analysis is chosen in order to decompose the country specific frontiers. The dataset originates from the SINTEF report (Kittelsen et al., 2009a).

Results: Compared to Norway; the frontiers of Finland and Denmark are significantly different from the Norwegian frontier. The cost penalty for providing specialist health care in Norway as compared to Finland and Denmark is estimated to 26.9 % and 9 %. There is found no statistical significant difference between Sweden and Norway. There are differences between country specific frontiers but not the efficiency behind each frontier. In addition, significant differences are found within Norway which implies regional level frontiers.

Conclusion: Norway could reduce its expenses by learning from the way Finland has organized their specialist health care. The results in this thesis makes it evident that the differences in hospital costs are due to differences in country specific productivity, and not the distribution of efficiency behind the frontier. It is recommended that future research attempts to develop a generic tool for assessing the best deterministic function as well as probability density function among several competing models. It would be interesting to examine allocative efficiency by shadow price models applied to this dataset. Quality outcomes ought to be incorporated in future efficiency evaluation.

Table of Contents

Preface.....	IV
Abstract	V
List of figures	IX
List of tables.....	X
List of Acronyms	XI
1 Introduction.....	1
2 Aim	5
3 Background.....	7
3.1 Specialist health care	8
3.2 Previous studies.....	9
4 Methodology.....	11
4.1 Techniques.....	11
4.1.1 Data Enveloped Analysis	13
4.1.2 Stochastic Frontier Analysis	15
4.2 The distribution of the residuals.....	18
4.3 Algebraic forms.....	20
4.4 Considering the functional form and residuals.....	21
4.4.1 The residuals	23
4.4.2 Production vs. costs and multiple outputs.....	24
4.5 Estimation.....	25
4.5.1 Hypothesis testing	26
4.5.2 Scale elasticity.....	28
5 The Data.....	29
5.1 Activity and costs	30
5.1.1 The variables	31
5.2 Selecting deflator.....	33
5.2.1 Other considerations	33
6 The model(s).....	35
6.1 Estimating the models	36
7 Results.....	39

8	Discussion.....	51
8.1	The deterministic component.....	54
8.2	The inefficiency.....	55
8.3	The elasticity of scale.....	57
8.4	Limitations.....	58
9	Conclusion.....	61
	References.....	65

List of figures

Figure 1	General governmental fiscal account, revenue and expenditure	1
Figure 2	OLS, DEA and SFA frontier in a single input / output orientation	12
Figure 3	The assumptions defining the DEA estimate of the feasible technology set	14
Figure 4	Alternative methods to examine hospital efficiency	17
Figure 5	The cost frontiers of Norway and Finland.....	53

Figure 5 by the author.

List of tables

Table 5.1 Descriptives	32
Table 7.1 Assessment of different models.....	39
Table 7.2 Assessment of <i>environ</i>	40
Table 7.3 Regularity and Parameter testing.....	41
Table 7.4 Sensitivity analysis	42
Table 7.5 The final models.	44
Table 7.6 Percentage deviation between the frontiers.....	48

All tables by the author.

List of Acronyms

C-D	Cobb Douglas function
CMI	Case mix index
CRS	Constant return to scales
DEA	Data Enveloped Analysis
DMU	Decision Making Unit
DRG	Diagnosis Related Groups
<i>Environ</i>	The variables: university hospital, capital city, LOS deviation, CMI and share of outpatients
F-test	F – statistic test
GDP	Gross domestic product
LM test	Lagrange multiplier test
LOS	Length of stay
LR-test	Likelihood ratio statistic test
i.id.	Independent and identically distributed probability density function
MDG	Main diagnostic group
ML	Maximum likelihood estimation
MLE	Maximum likelihood estimator
The Nordics	The countries: Denmark, Finland, Iceland, Norway and Sweden. The thesis excludes Iceland when referring to the Nordic(s).

OLS	Ordinary Least Squares regression
RLS	Restricted least squares
Scandinavia	The countries: Denmark, Norway and Sweden
SFA	Stochastic Frontier Analysis
SINTEF	SINTEF is an independent research group based in Trondheim.
VRS	Variable return to scales
W test	The Wald test

1 Introduction

The increasing health care expenditures in the Nordic countries have driven authorities to investigate methods for enhancing hospital efficiency. This has been done through several reforms in the past. One example is the prospective activity based financing schemes that reimburse the activity in the hospital, based on the diagnosis related group (DRG) system. The thought behind this kind of payment system is that it gives incentives for cost containment and at the same time increases efficiency and activity (Magnussen et al., 2009).

The health sector is a substantial part of the public sector in the Nordic countries. According to statistics, nominal general governmental expenditures in the health sector in Norway for the year 2000 was around 16 % of total costs, and increased to about 17 % by 2010 (ssb.no, 2010b). This is illustrated in figure 1.

General government expenditure, by sector as a percentage of total costs. 2000 and 2010

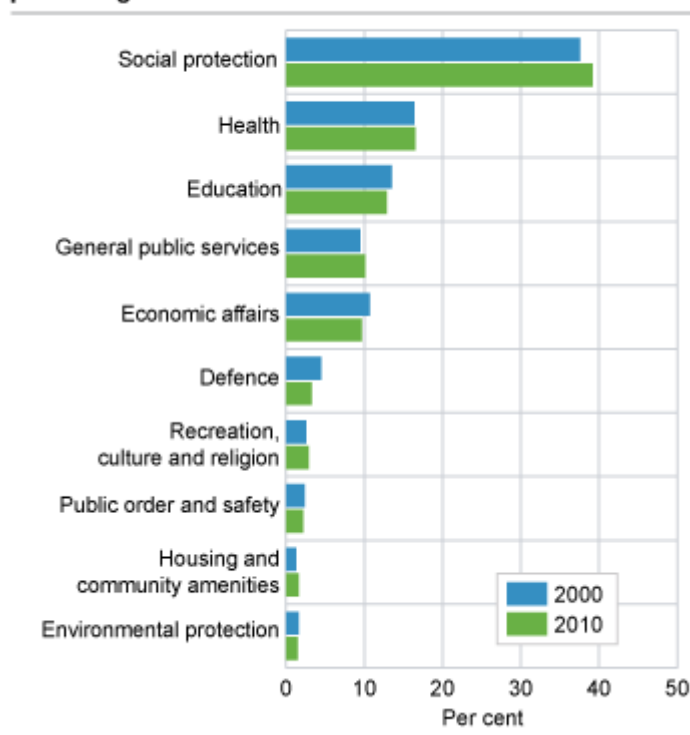


Figure 1. General governmental fiscal account, revenue and expenditure. Source: Statistics Norway.

With the high expenditure in the Norwegian health sector in mind, it may not come as a surprise that the levels of cost and efficiency, or inefficiency, are of interest. Furthermore, it is interesting to examine what the relative levels of efficiency between the Nordic countries are, and whether there is a best practice of allocating resources that we can learn from. The answers to these questions are not easy to provide, nor are they the main research questions in this thesis. However, the methodology used in this thesis may provide a good starting point.

The research question which is examined in this thesis is:

Are the differences in hospital costs between Finland, Norway, Denmark and Sweden, over time, due to differences in the production possibilities that are specific to country and period, or due to the distribution of efficiency behind the frontier?

In order to investigate the research question, there are two main methods to consider. The general literature (e.g. Jacobs et al. (2006), Coelli et al. (2005)) divides the techniques in two categories; the non-parametric Data Enveloped Analysis (DEA) method and the parametric Stochastic Frontier Analysis (SFA) method.

This thesis builds on previous studies applying a multi-output cost function in assessing cost efficiency and estimating inefficiency. A contribution of this thesis to the literature is to apply SFA in an environment where DEA is frequently used, and assess country specific frontiers.

In the literature it is common to make assumptions on issues such as whether efficiency varies over time or not (Coelli et al., 2005). By letting the data decide what type of assumptions are best suited for presenting the material, this thesis is not making any a priori specifications. Here it is not assumed how technology, time, country nor region influences costs in one way or the other; rather I will test for it.

Researchers often have a notion of explanatory variables that they assume can account for (in)efficiency. The dataset used in this thesis provides information on environmental variables that is expected to influence the (in)efficiency component directly. The expectancy is not taken for granted, instead I will test if the environmental variables are correctly assumed to influence the (in)efficiency or not.

This thesis is organized as follows: chapter 2 presents the aim and motivation for the thesis. Chapter 3 gives a background to the subject of the thesis and defines concepts such as a

Nordic model of health care; it attempts to define specialist health care and provides a brief literature review. Specific methodological issues of concern are explored in chapter 4. I include a brief introduction to the DEA technique in order to account for the methodology that is omitted, and provide a thorough understanding of SFA. Among several issues, special consideration is given to functional forms for assessment of the cost estimation as well as the distributions commonly imposed on the residuals.

Chapter 5 investigates the data and challenges. It gives an account of the variables and what is measured as output. Chapter 6 provides a discussion of the model(s) and how estimation is conducted, before presenting the results in chapter 7. A discussion of the results is provided in chapter 8 followed by a conclusion in chapter 9.

2 Aim

The objective of this thesis is to find out if the differences in hospital costs between the countries, over time, are due to differences in the productivity of the best performing hospitals and periods, which defines the efficiency frontier, or due to the distribution of efficiency behind the frontier.

A parametric method is chosen over a non-parametric method. The SFA method provide means of separating measurement error (or noise) from inefficiency in the residual as opposed to DEA which assumes that there is no measurement error (Jacobs et al., 2006). In order for separation to occur SFA assumes functional restrictions on the residual with respect to the noise and inefficiency component respectively (Kittelsen, 1997).

In addition to use DEA methodology, the SINTEF report A12200 “En komparativ analyse av spesialisthelsetjenesten i Finland, Sverige, Danmark og Norge: Aktivitet, ressursbruk og produktivitet 2005-2007” (Kittelsen et al., 2009a) tried to employ an SFA analysis, under the condition of truncated normal distribution of the inefficiency term¹. Unfortunately, convergence was not achieved (Kittelsen et al., 2009a, p. 120). ‘Convergence not achieved’ simply means that the maximization of the deterministic part in the model fails, i.e. the function does not exhibit a global concavity or it is not found (StataCorp, 2009). Therefore, another aim of this thesis is to try alternative assumptions of the probability density function of the inefficiency term as well as the deterministic part to see whether convergence can be achieved. The dataset used stems from the above mentioned SINTEF report.

A working hypothesis in this thesis is that the previously discovered global DEA frontier discussed in the SINTEF report is not entirely representative. In the report (Kittelsen et al., 2009a) Norway emerged as clearly less productive compared to Finland. If it can be shown that the global frontier under DEA is unrepresentative, it could imply that the decision making units (DMUs) within each country might be closer to their own frontier. This in turn, may suggest that Norwegian public hospitals are as efficient as they have prevailing conditions to be. One possibility, if this is verified, is that there are underlying structural differences that influence the efficiency. This debate however, will not be the focus of this thesis.

¹ Personal transcript provided by Sverre A.C. Kittelsen

Nevertheless, a good starting point for further discussion of the structural differences is Magnussen et al. (2009).

3 Background

One often talks about the Nordic model, which refers to the Nordic welfare state. As with the Nordic model, the Nordic health care systems' are reasonably similar. While the SINTEF report (Kittelsen et al., 2009a) defines a health care system in line with Roemer (1989, p. 77) as “*any health system can be characterized through analysis of five major components: (1) its production of resources, (2) organization of programs (including a residual private market), (3) sources of economic support, (4) modes of management, (5) patterns of providing services.*”, the notion of a Nordic health care system is less familiar.

According to Magnussen et al., (2009, p. 6) the “*Nordic model of healthcare*” has a goal of equity in health without compromising universalism. Furthermore, it is financed through tax based funding and has a large proportion of public governance. The proportion of private actors is relative small, and few people have private health care insurance (Magnussen et al., 2009).

Yet there are differences. These differences relate to how the systems are organized in terms of centralization / decentralization, the use of incentives to get the organizations to adhere national directives, how well integrated primary and secondary health care is, etc. (Kittelsen et al., 2009a).

The Nordic countries are also comparable with respect to demographic forecasts. In general we have a birth rate under 2, meaning below the rate to sustain the population, although this is anticipated to be offset by immigration. All countries have an increasing proportion of elderly needing health care within the next 20 years, and in addition the relative proportion of workers to retired workers is expected to be reduced further (ssb.no, 2010a, scb.se, 2010, dst.dk, 2001). In other words, the demographic forecasts imply increased pressure on the welfare states' sustainability.

These forecasts bring about questions regarding whether the health care sector is using its resources efficiently, and whether resources could be allocated differently in order to provide the population with even more / better health care services. As a response, a comparative study was conducted and presented in 2009. The SINTEF report concluded that Norway has a

higher resource usage and labor costs per capita of all the four countries and a lower productivity level, particularly when compared to Finland (Kittelsen et al., 2009a).

Using Norway as a base comparison when considering per capita resource spending in somatic care, Sweden used 94 % of the resources, Denmark used 92 % and Finland used only 65 % in 2007. However, the trend shows that all countries have had a larger real-growth in expenditures compared to Norway in the years 2005-2007. The Finnish productivity was 18 % above the level of the Norwegian productivity measured using the DEA method. The differences between Norway, Denmark and Sweden seems statistically non significant at the 5 % significance level (Kittelsen et al., 2009a).

3.1 Specialist health care

The SINTEF report made an attempt to provide a common definition of the specialist health care for all the four countries based on Norway's division of health care. It discovered that the Norwegian definition of the specialist health care originates mainly from the legal division between what is to be conducted at municipal and regional level². Roughly, this division implies that primary care is to be conducted at the municipal level, while the specialist care is to be provided at the regional level by regional health care authorities (Kittelsen et al., 2009a).

Thus, the Norwegian health care is divided between two pillars of public sector, as opposed to the Swedish and Finnish. While Sweden has placed the responsibility of the primary and specialist health care on the regional level, Finland has placed the responsibility on the municipal level. For Denmark, Kittelsen et al. (2009a) find that whereas the Danish law describes the local and regional authorities responsibilities, it does not specify what is to be understood as the specialist health care obligations. Due to the challenges this discrepancy introduce and the additional problem in the dataset that could occur when trying to separate the services provided in each country, the attempt to define one mutual understanding of the specialist health care is abandoned (Kittelsen et al., 2009a).

² For details on the law, consult: <http://www.lovddata.no/all/tl-19821119-066-001.html#1-2> chapter 1 for municipal responsibility, and <http://lovddata.no/all/tl-19990702-061-002.html> §2-1a. for the regional health authorities responsibility

In fact, due to the small, but yet significant differences between the countries, it was deemed most appropriate to restrict the analysis to somatic care only, as well as provide an exploratory analysis of the psychiatric health care (Kittelsen et al., 2009a). This thesis however will focus on the specialist health care.

After this introduction to the topic, the reader should now have a general understanding of the most vital results from the SINTEF report (Kittelsen et al., 2009a) that affects this thesis. For further details the interested reader is advised to consult the report itself.

- The SINTEF report (Kittelsen et al., 2009a) used the DEA methodology.
- A definition of a health care system is provided as well as a notion of a Nordic health care model.
- An attempt was made in the parametric tradition of SFA, but this fails. The analysis is thus restricted to public hospitals in the somatic health care.
- The Nordic countries are comparable in demographic forecasts, which informs of future challenges to the sustainability of the Nordic health care model.
- Norway has a higher resource usage per capita, highest labor cost and a lower productivity. Finland represents the opposite case, using 65 % of the resources that the Norwegian health care consumes per capita, but has 18 % higher productivity.

3.2 Previous studies

In a recent paper Medin et al. (2010) presents a bootstrapped DEA analysis of cost efficiency among university hospitals in the Nordic countries, excluding Iceland. Some of the conclusions are in line with previous findings which argue that the added responsibilities (teaching and research) university hospitals have potentially interferes with the routine work of patient care and thereby raises the costs. What is new in the article is the inclusion of variables explaining research activities and teaching. These have a major impact on the

models cost efficiency. Geographic location and patients with a high (>5) DRG weight are also shown to affect university hospital efficiency.

Linna et al. (2006) conducted a comparative study of hospital cost efficiency through DEA in Norway and Finland using data from 1999. They found a significant difference between and within the countries. The average cost efficiency of Norwegian hospitals was 17-25 % lower to that of Finnish hospitals per capita, and had a longer duration of patient stay, commonly expressed as Length of stay (LOS).

Linna et al. (2010) compared the average hospital cost efficiency in the Nordic countries in the year 2002. One of several results indicates that Swedish hospitals were less cost efficient than the Norwegian hospitals. The paper supported previous findings where the Finnish hospital sector was found to be the most cost efficient sector on average among the Nordic countries.

In a working paper, Kittelsen et al. (2009b) investigate the effect of the substantial Norwegian hospital reform of 2002. Norway, as opposed to the other Nordic countries, started to re-centralize its hospitals and changed ownership from regional to state ownership. The paper concentrates on the period before and after the shift of ownership (the years 1999-2004) to see whether the reform had the effects one hoped for regarding reduced waiting time, cost control and increased efficiency. Even though not yet published, some of the findings indicate that productivity has increased in the order of 4 % during the time period. However, it is unclear why this effect took place.

In an earlier version in Norwegian, Kittelsen et al. (2007) find that the Norwegian hospital reform of 2002 improved hospital productivity with 3-4 %. With some reservations, they also found that Norwegian productivity was above the Swedish, yet below the Finnish. For Denmark, data was only available for 2002. At that time Denmark was on the same level as Finland. Furthermore, one conclusion is that the increased efficiency effect is not due to any different technology changes in the other Nordic countries that divert from the technology changes in Norway.

4 Methodology

In standard economic theory, the productivity of a firm would be explained through the relationship between inputs and outputs. The relationship can be exemplified with two firms. If firm 1 uses two units of input to produce one unit of output, it would be considered as having higher productivity than firm 2 that uses three identical units of input yet produces only one identical unit of output. Thus, productivity expresses how much output, or production, one achieves with the available resources (Frank and Glass, 2006). This is how productivity is to be understood throughout this thesis.

Of perhaps greater interest is the efficiency. As opposed to productivity, efficiency is always expressed in relative terms. Efficiency expresses the productivity in relation to the best feasible input / output mix (Carlton and Perloff, 2005). Consequently, firm 1 above not only has higher productivity, but is also more efficient than firm 2. Given the prices of the input and levels of output are equal for both firms, firm 1 is also more cost efficient compared to firm 2, as it uses less monetary resources in production (Frank and Glass, 2006). Now that we have a notion of productivity and efficiency, let us turn our attention to the subject of this chapter.

4.1 Techniques

As noted earlier, a parametric method is chosen in this thesis rather than a non-parametric method. Some theorists (Kittelsen, 1997, Jacobs, 2001) divide the two techniques further into deterministic (best practice comparison) and stochastic (random error) categories. DEA would be deemed non-parametric but deterministic and SFA would be categorized as parametric stochastic. This thesis will however only differentiate between the non-parametric and parametric categories.

I will here briefly provide an explanation of the non-parametric DEA method before continuing with the parametric SFA method. Both are techniques for estimating inefficiency on aggregate level down to firm specific level. I will not consider index number methods,

since it is believed that some Decision Making Units (DMU) are less efficient than others, contradicting the underlying assumption of index number methods (Coelli et al., 2005). Neither will hybrid non-neutral stochastic models like those of Huand and Liu (1994) be discussed, as this is not the interest of this thesis.

Figure 2 provides a simple graphical presentation of the two techniques discussed, in comparison to the simpler ordinary least squares (OLS) method.

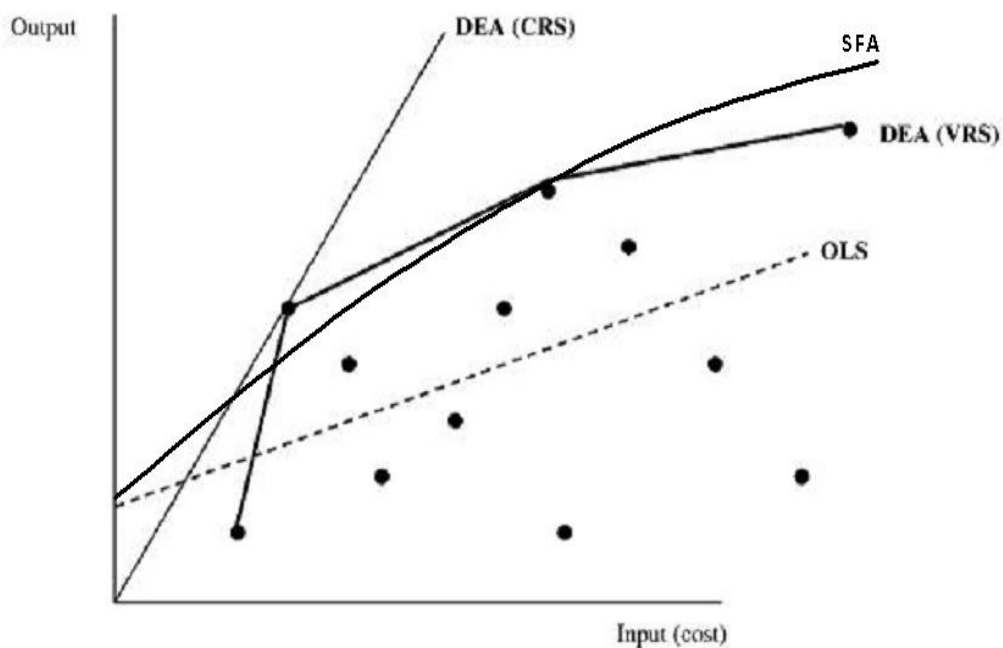


Figure 2. OLS, DEA and SFA frontier in a single input / output orientation. Source: Modified from Jacobs, R., (2001).

As can be seen, SFA does not align it self to the “best fit” as OLS. Neither does it “rest on” the best observed units as DEA (Jacobs et al., 2006)

4.1.1 Data Enveloped Analysis

According to Jacobs et al. (2006) DEA is now established as the prominent empirical methodology for efficiency evaluation within sectors where we cannot assume cost minimizing behavior, and especially so when evaluating the health care sector.

The idea of DEA is rather simple. The underlying key assumption is that there is no measurement error that results in noise. Therefore, DEA does not need to consider the form of the cost or production function.

In SFA as in DEA the dataset defines the frontier. Some authors (e.g. Jacobs et al., 2006) would say that in approaching efficiency measurement SFA is more theoretical while DEA is more empirical. Irrespective of Jacobs, the units in DEA that are most efficient constitute the frontier. All other units are behind the frontier. Hence, the frontier is enveloped around the most efficient units, while the efficiency of a unit that is not fully efficient is measured as its relative distance to the frontier. Of course, the DMUs that constitute the frontier has an efficiency equal to 1, while those that are behind the frontier take any value between 0 and 1. This flexible property makes DEA very attractive when assessing efficiency (Jacobs et al., 2006).

Kittelsen (1997) points out that while the DEA has no assumptions regarding the functional form of the frontier itself, there exists underlying assumptions regarding the feasible technology (or the production possibility set). From Kittelsen (1997), these are, a) that all observations constituting the data are feasible, b) free disposal of inputs and outputs, and c) “*convex envelopment of the data*” (Kittelsen, 1997, p. 11). Figure 3 below provides an illustration of the assumptions where each + indicates an observation. Together the three assumptions make d) the DEA estimation.

DEA method

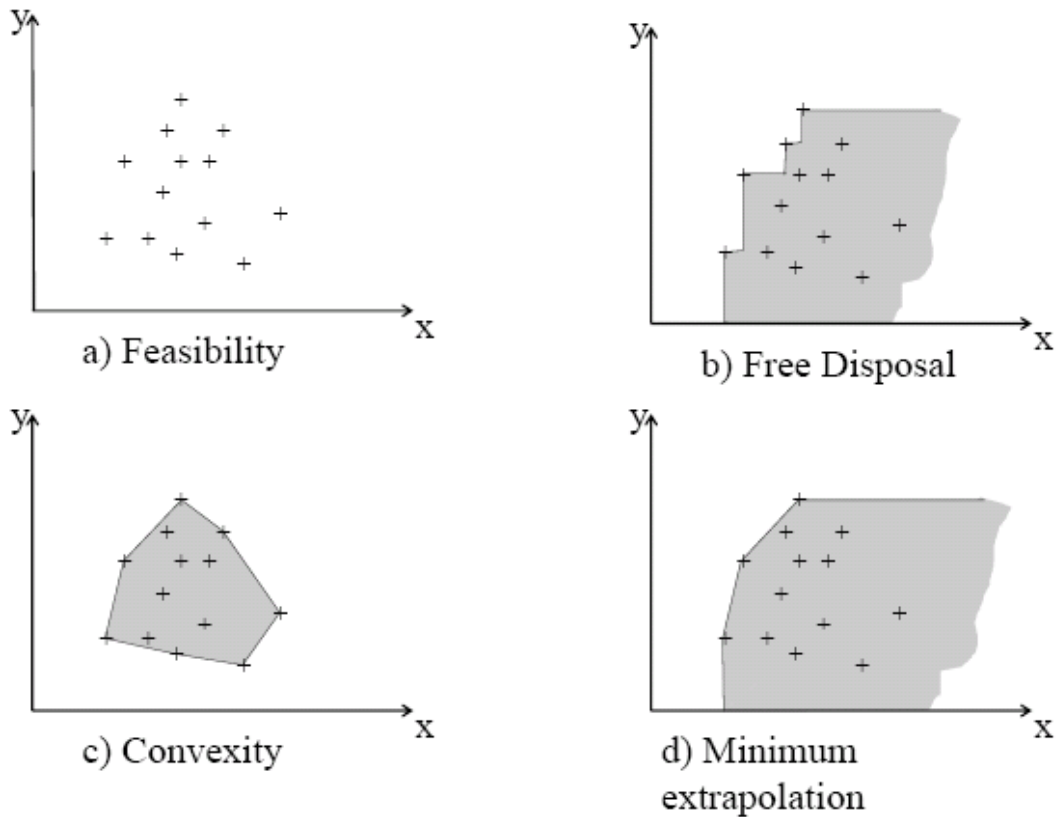


Figure 3. The assumptions defining the DEA estimate of the feasible technology set. Source: adapted from Kittelsen (1997, p. 11)

As with SFA, DEA can be calculated both under constant return to scale (CRS) or variable return to scale (VRS), as illustrated in figure 2 above. If it is assumed that the units are operating at an optimal level, CRS is appropriate, if not VRS is often applied. It is important to note, that if the data used is expressed in ratios, such as discharge rate or proportion of input x and the assumption is that the units are operating at non-optimal levels (VRS), the underlying implication becomes a CRS estimation (Jacobs et al., 2006). However, without going into details, there are ways to work around the issue of having ratio data but wanting to use VRS by applying the Banker, Charnes and Cooper estimation of DEA (as cited in Jacobs et al., (2006, p. 104)).

Finally, since DEA is a model comparing units against a best practice benchmark it can be thought of as yardstick model. Comparison against a benchmark makes the result sensitive to few observations in the dataset and to extreme outliers. Another potential impairment effecting efficiency estimates occurs if the units' conduct different activity but have similar measurement of output / input mix. To give an example, a hospital with emergency care would seem to have more outputs than a hospital that does not have emergency care. If there are no other peers to benchmark against (hospitals') with emergency care in the sample, the uniqueness of the unit that has emergency care would lead to full efficiency score, and thus bias the results (Jacobs et al., 2006).

4.1.2 Stochastic Frontier Analysis

While the DEA technique is appealing with its flexibility and empirical attraction, making it unnecessary to a priori determine the functional form, the SFA technique requires the consideration of functional form of the regression (mind the usual assumptions and problems associated with common regression techniques, e.g. degrees of freedom) and the efficiency term. In fact, the separation of the deterministic part from noise and efficiency is why Zuckerman et al. (1994) deems SFA superior to DEA. As opposed to DEA, SFA uses all information available in the dataset when deriving the efficiency scores of each unit (Jacobs et al., 2006).

The first basic SFA model was proposed separately by Aigner et al. (1977) building on the work of Farrell 1957 (as cited in Aigner et al. (1977)), as well as by Meeusen and van Den Broeck, (1977) criticizing the Afriat-Richmond production frontier (as cited in Meeusen and van Den Broeck (1977)). The basic production model proposed in Coelli et al. (2005) yields:

$$\ln y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i \quad (1)$$

and,

$$\varepsilon_i = v_i - u_i$$

whereby y_i is an output variable, \mathbf{x}_i' is a vector of inputs, β is a vector of unknown parameters, ε_i is the residual composed of v_i which is symmetric random error and u_i , which is of particular interest, represents estimation of the random inefficiency. Conveniently, the function is specified in log form, and therefore the coefficients are expressed in percentage or elasticity change (Jacobs et al., 2006), while the inefficiency term is expressed in deviated percentage from the frontier (Greene, 2003). As noted above, the functional restriction on the error term is usually a normal distribution, and for the inefficiency term it is usually half-normal, exponential, truncated or a gamma distribution (Coelli et al., 2005).

When considering a production function the following condition must hold;

$$y_i \leq \exp(\mathbf{x}_i' \beta + v_i), \text{ and } u_i \geq 0 \quad (\text{Aigner et al., 1977, p. 24}) \quad (2)$$

Implying that total output units (production), y_i , must be smaller than or equal to the total amount of inputs, $\exp(\mathbf{x}_i' \beta + v_i)$. Thereby, the deterministic frontier exhibits diminishing returns to scale in line with economic theory of production, or what Coelli et al., (2005) calls “*output values are bounded from above by the stochastic... variable $\exp(\mathbf{x}_i' \beta)$* ” and the random error varies around the deterministic frontier (Coelli et al., 2005, p. 243).

Figure 4 below illustrates the above features of (2), the Stochastic production frontier in an input / output orientation. Consider two firms, A and B. If $u_i = 0, i = A, B$ (no inefficiency) the production of firm A and B, q_i^* , would be equal to $x_i = \otimes$ in the figure, sometimes referred to as the frontier output (Coelli et al., 2005). To clarify, note that Coelli et al., (2005) uses q_i instead of y_i .

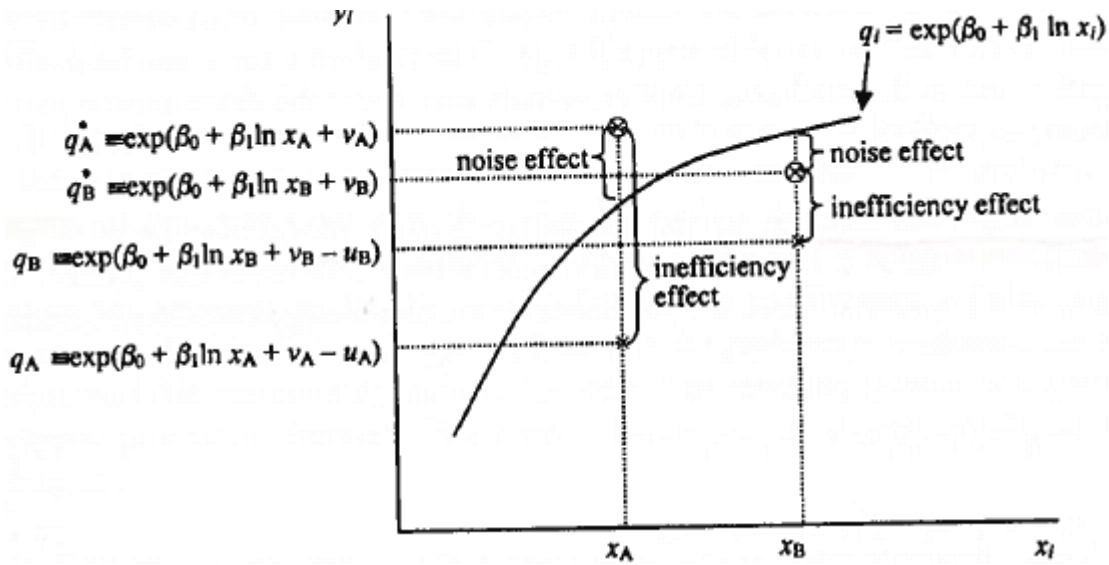


Figure 4. A Stochastic Production Frontier. Source: Coelli et al. (2005, p. 244)

If the noise component is positive, that is $v_i > 0$, the frontier output point would lie above the deterministic production function, and vice versa if the noise component is negative. In addition, if the sum of $(v_i - u_i) < 0$, the observed output of firm $i = A, B$ must lie below the deterministic production function given by $q_i = \exp(\beta_0 + \beta_1 \ln x_i)$. Thus, the frontier output tends to vary above and below the functional form of the deterministic compound, while the observed output tends to lie below the very same. In this way, one can derive an estimation of the inefficiency term $\exp(-u_i)$ by considering the ratio of observed output to the frontier output (Coelli et al., 2005). Considering (1) this observation gives us:

$$\text{Technical efficiency} = TE_i = \frac{y_i}{e^{(\beta_0 + \beta_1 \ln x_i + v_i)}} = e^{-u_i} \quad (3)$$

4.2 The distribution of the residuals

As mentioned above, the SFA can estimate the inefficiency term with the assumption of a distributional form. The distributional forms often used in the literature were half-normal, exponential, truncated normal and gamma distributed. I will return to the issue of how the assumptions affect the efficiency term.

Under the half normal distribution, the assumptions regarding the residual ε_i in (1) above can be expressed as (Coelli et al., 2005):

$$v_i \sim iid \text{ Normal } (0, \sigma_v^2) \quad (4)$$

$$u_i \sim iid \text{ Half - normal } (0, \sigma_u^2) \quad (5)$$

(4) tells us that the stochastic error term is independently and identical (i.id.) normal distributed with a mean of 0 and a variance of σ_v^2 , while assumption (5) imposes an independent and identical half-normal distribution with 0 mean and variance of σ_u^2 on the efficiency term u_i .

The Gamma, Exponential and Truncated normal distributional assumptions changes (5) to, (6), (7) and (8) respectively:

$$u_i \sim iid \text{ Gamma } (\lambda, m) \quad (6)$$

$$u_i \sim iid \text{ Exponential } (\lambda) \quad (7)$$

$$u_i \sim iid \text{ Truncated Normal } (\mu, \sigma_u^2) \quad (8)$$

The restriction (6) tells us that the gamma distribution is i.i.d distributed with mean and degrees of freedom equal to λ, m respectively (Greene, 2003). While (7) is the exponential distribution i.i.d. with mean and variance equal to λ , and is a special case of the gamma distribution (Johnson and Kotz, 1970). The Truncated normal distribution (8) is i.i.d. distributed with mean and variance, μ, σ_u^2 , being limited by a range of values (Greene, 2003), and is a general case of the half normal distribution (Jacobs et al., 2006).

A note is made regarding the variance of σ_u^2 in the residual, applying to the half normal distribution (5) only. In discussing the half normal distribution, Greene notes (in Schmidt et al., 2008) that it is a common mistake to take the value of σ_u^2 as a face value. By doing so, one overestimates σ_u^2 by a factor of 3. Hence, the true variance of the u_i component in a half normal distribution is: $\text{var}[u_i] = \sigma_u^2 [(\pi - 2) / \pi]$.

Since the initial SFA model in the SINTEF report (Kittelsen et al., 2009a) did not converge, it is of interest to investigate the (5) – (8) distributions. In discussing the Gamma distribution, Ritter and Simar (1997) points out that it is preferred to have several thousand observations. They also point to the risk of correlation between the stochastic element (noise) and the efficiency component when using the Gamma distribution. Hence, all distributions but the Gamma distribution will be conducted in this thesis.

Following the parsimonious criteria above the half normal distribution or the exponential distribution would be preferred. As a drawback, they are less flexible than the Gamma and truncated normal distribution. On the other hand, the first two are likely to be better in separating noise from inefficiency due to the mode at zero (Coelli et al., 2005) and, thereby, evade the correlation issue pointed out by Ritter and Simar (1997) above.

Thus, the different distributions affect the prediction of the efficiency since they affect the efficiency component, u_i . In discussing the impact of the different probability distributions, Coelli et al. (2005) finds 0.09 % change when comparing exponential distribution to half normal distribution, while, Greene concludes that the answer to whether “...*the distribution matter?*” is that it “... *does not have an analytical answer*” (Schmidt et al., 2008, p. 180). Nevertheless, in line with Coelli et al. (2005), a likelihood ratio test will be used when assessing which distribution of the inefficiency term is most appropriate.

4.3 Algebraic forms

The choice of which function to use in the deterministic part of the SFA is an important matter. For instance, choosing a Cobb Douglas (C-D) function gives other properties than a quadratic or translog function would yield. According to Coelli et al. (2005), there are four considerations to take into account when choosing functional form in estimating the production technology:

a) Flexibility, b) Linear in the parameters, c) Regularity, d) Parsimonious.

Let us consider each in turn.

a) Flexibility

it is common to differentiate between first order flexibility and second order flexibility. A first order flexible function gives estimates in a linear form, while second order flexibility gives a quadratic approximation. Generally a function of second order flexibility is preferred as it is more exact in estimating several points in space. However, the flexibility reduces the degrees of freedom as the number of parameters to be estimated increases.

b) Linear in the parameters

when the function exhibits linear parameters we can use the function with linear regression methods.

c) Regularity

the cost or production function must satisfy what is termed regularity properties of economics. This implies that the function must be a non-negative real value, show concavity or quasi concavity and exhibit monotonicity. Furthermore, output cannot exist without input, neither can output be negative. If input price increases, the total cost must increase given the same output. More units of input cannot result in a decrease of output units, and costs must increase if units of output increases (Caves and Christensen, 1980, Coelli et al., 2005).

d) *Parsimonious*

implies that the simplest functional form should be used. In other words, the simpler the function is the better.

4.4 Considering the functional form and residuals

A perennial consideration in this thesis is the technical aspects of the functional form of both the deterministic part, and the distribution of the residual(s). I will here consider each in turn. Let me start with rewriting (1) above into a generic form:

$$y_i = f(\mathbf{X}_i) + \varepsilon_i \quad (9)$$

Where as previously,

$$\varepsilon_i = v_i - u_i \quad (10)$$

Still, this thesis will not deal with inputs in relation to production as in (9). Instead, it considers a measure of outputs in relation to costs, thus I rewrite (9) to (11):

$$c_i = f(\mathbf{y}_i) + \varepsilon_i \quad (11)$$

c_i stands for costs. It can be the individual organizations' costs or the total industry costs.

$f(\mathbf{y}_i)$ indicates the deterministic functional part of outputs, whereas the ε_i is as before.

Following these considerations when estimating the properties of the production technology (and in effect the efficiency term), I choose to try different functions in order to see the impact

on the model. In a recent literature review and analytical paper, Rosko and Mutter (2008, p. 161) found little impact on the estimation of the mean inefficiency term when considering alternative cost functions (the deterministic part). The author of this thesis is skeptical to their conclusion. Nevertheless, Rosko and Mutter (2008) find reasons to prefer a translog cost function over a C-D function, which is in line with the flexibility requirement above. In this thesis the following cost functions from Coelli et al. (2005) are used when assessing the deterministic part in addition to a C-D function. I will return to the results later in the thesis.

Cobb–Douglas

$$C_i = \beta_0 \prod_{n=1}^N y_{in}^{\beta_n} e^{v_i - u_i} \quad (12)$$

Quadratic

$$C_i = \beta_0 + \sum_{n=1}^N \beta_n y_{in} + 0.5 \sum_{n=1}^N \sum_{m=1}^N \beta_{nm} y_{in} y_{im} + v_i - u_i \quad (13)$$

Translog

$$\ln C_i = \beta_0 + \sum_{n=1}^N \beta_n \ln y_{in} + 0.5 \sum_{n=1}^N \sum_{m=1}^N \beta_{nm} \ln y_{in} \ln y_{im} + v_i - u_i \quad (14)$$

In (12) and (13) above, C_i are observed costs for hospital i , the β_n and β_m are the unknown coefficients to be estimated. While β_0 is a constant, y_{in} and y_{im} are the observed output variables, v_i and u_i are the error term and the efficiency term respective. The only difference in (14) is that the function is logarithmic on both sides of the equation.

4.4.1 The residuals

Until now, we have assumed that the v_i in (4) is normal distributed, while the u_i is subject to alterations. STATA 11, offers possibilities to specify the u_i in more detail (StataCorp, 2009). In particular, specification of the variance in either / or both error term(s), that is (4), (5), (7), and (8). Specified in StataCorp (2009) as:

$$\sigma_i^2 = \exp(\mathbf{w}_i \delta), \quad i = v, u \quad (15)$$

where the vector \mathbf{w}_i consists of a constant, and a function of explanatory variables, while δ is an unknown vector of coefficients. Still, I will model u_i only, thus not altering the variance of (4).

Generalize Battese and Coelli (1995) the u_i can be expressed as:

$$u_i = z_i \delta + \mathbf{W}_i \quad (16)$$

where \mathbf{W}_i is a random variable, defined by the distribution, and the z_i is the function of the explanatory variables, and δ is as above. (15) and (16) actually express the same, albeit with different use of notation. Following (15); (5), (7) and (8) can be rewritten to (17), (18) and (19) (StataCorp, 2009);

$$u_i \sim \text{independently Half-normal } (0, \sigma_u^2), \text{ distributed} \quad (17)$$

$$u_i \sim \text{independently Exponential } (\sigma_u^2), \text{ distributed} \quad (18)$$

$$u_i \sim \text{independently Truncated Normal } (\mu, \sigma_u^2), \text{ distributed with truncation point } 0 \quad (19)$$

Note that (18) has mean and variance embodied in one parameter like Greene discusses in Schmidt et al., (2008).

4.4.2 Production vs. costs and multiple outputs

As often is the case with publicly provided goods, price information is usually not available on the outputs. This is also the case for the data in this thesis. However, since information about the costs generated from providing the public goods are available, efficiency analysis can be conducted without hindrance. For example by using (11) above.

In addition to the problem of public goods, Jacobs et al. (2006) points to the problem of multiple outputs. Because SFA is a parametric method, it is ill suited for estimation of a *production function* with multiple outputs (dependent variables) as opposed to DEA. However, setting up a *cost function* avoids the problem of multiple outputs, since the dependent variable becomes total costs. Other methods do exist for handling multiple outputs, such as distance functions (Coelli et al., 2005) but will not be considered in this thesis.

According to Jacobs et al., (2006) an underlying assumption when estimating cost functions is the need to assume cost minimizing behavior. If this assumption holds, the cost function is equivalent to that of the production function. Jacobs et al. (2006) problematize the cost minimization assumption in the health care sector. As the Nordic hospitals are publicly funded and operate in a suboptimal market, the usual market mechanisms are not applicable. In discussing economic behavior Coelli et al. (2005) point out that SFA and DEA do not need behavioral assumptions nor price information. In fact this thesis as discussed above, and in line with Coelli et al. (2005), assumes that the units are not operating under full efficiency, but a priori assumes nothing about behavior.

There are many ways that governments can enforce regulations in order to encourage cost containment in the public sector, meaning, enforce a behavior. In the typology of Nordic Health Care Systems, Magnussen et al. (2009) discusses some of the most widely used mechanisms in all of the Nordic countries. Thus, even though the individual units might not act according to the usual economic notion of cost minimization, the Nordic health care sector per se is assumed to attain to cost containment. Consequently, given the above discussion I can justify the use of a cost function in my modeling.

4.5 Estimation

In discussing how to best estimate the parameters in an SFA model, Coelli et al. (2005) propose the estimation technique called maximum likelihood estimation (ML) because it exhibit asymptotic properties.

A good explanation of ML is provided in Thomas (2005). Assume a sample size $= n$, and an unknown population mean which is expected to be $\mu = 40$. When observing the sample mean it is detected that it has a mean of $\bar{x} = 10$. Thereby, it seems unlikely that the sample stems from a population with $\mu = 40$. But what if it is suspected that $\mu = 12$? Then the probability that our sample with $\bar{x} = 10$ stems from a population with a mean of $\mu = 12$ seems reasonable. As Thomas (2005) puts it “*The maximum likelihood estimator (MLE) of μ [and variance σ^2] is simply that value of μ [and variance σ^2] that maximizes the probability of generating the sample values actually obtained*” (Thomas, 2005, p. 346).

Without going into technical details, the notion of ML requires an explanation. In discussing the properties of MLE, Andersen (1970) notes that the ML estimation technique dates back to 1922. He points out that throughout the years multiple proofs have been set forward that ML displays true asymptotic properties. By the 1980's ML was established as a solid method when wanting to estimate parameters, just like Ordinary Least Squares (OLS). However, unlike the OLS discussion, concern was devoted to the residuals. It was a necessity to develop methods so that the residual(s) also could be proven to exert asymptotic properties.

An early proof of ML showing asymptotic properties was achieved with a quadratic function and exponential distribution of the residuals in 1968 by Aigner and Chu (1968). In a two pages long paper, Schmidt proved how ML under the same deterministic conditions as set forward by Aigner and Chu, was valid with half-normal distribution (Schmidt, 1976). In discussing ML estimation, Greene (1980) put forward proof for the Gamma distribution. The Gamma distribution proved in theory to hold what is termed the usual desired properties of MLE. That is, to be consistent, asymptotically normally distributed and asymptotically efficient (Greene, 1980). The latter implies that if the residuals are symmetric, MLE approaches OLS. Thus, ML outperforms OLS when the residual distribution can be assumed to be skewed away (asymmetric) from the frontier (Greene, 1980), thereby explaining Coelli's statement above in regards to why ML is preferred.

However, as discussed in Schmidt et al., (2008) the statistical properties of the two first proofs are virtually unknown. It has later been revealed that Greene's (1980) gamma distribution shows inconsistency in its estimation regardless of sample size (Jacobs et al., 2006).

Unfortunately, as indicated in chapter 4.2 above, there is no consensus in the literature which may guide the researcher on which distribution should be applied when estimating the u_i (Jacobs et al., 2006). In discussing the problem of which distribution to choose, Coelli et al., (2005) mentions that it is a matter of "*computational convenience*" (Coelli et al., 2005, p. 252). As evident below, the log likelihood ratio test can be applied for sensitivity analysis in order to assert which distributional assumption is best suited.

4.5.1 Hypothesis testing

According to Coelli et al. (2005) the assumption of technological substitution can be tested for correctness. For example, the researcher expects the technology to exhibit constant return to scale (CRS) but wants to find out if this is true. By imposing equality assumptions on the unknown coefficients and running a restricted least square (RLS) estimator against an unrestricted estimator (e.g. OLS) hypothesis testing on the RLS is possible by using the F -test. This however requires that both the constraints and the regression models need to be linear in their parameters (Coelli et al., 2005). As noted by the very same authors, Greene

(2003) presents methods where one only needs to run the unrestricted model and then conduct the Wald (W) test, or the Lagrange multiplier (LM) test for a restricted model.

An alternative to the F -test is the likelihood ratio statistic (LR). It is still needed to conduct an RLS and an unrestricted regression as in the F -test. However, it is somewhat easier to calculate and is applicable to the ML technique used in this thesis. If the LR test has an advantage in calculating the test, it has a disadvantage if only a small sample is available. The F -test is warranted on small a sample under fulfilled requirements as discussed above (Coelli et al., 2005).

The LR-statistic is

$$LR = -2[\ln L_R - \ln L_U] \sim \chi^2(J) \quad (20)$$

where the values from the log-likelihood function of the restricted and unrestricted models are used under a chi-square distribution. Reject the $H_0 : \mu = 0$ hypothesis if the LR statistic $> \chi^2_{1-\alpha}(J)$, where J is the number of restrictions, and μ represents the parameter(s). The LR test statistic can be used for all the distributions (17), (18) and (19) above, and for parameter testing (Coelli et al., 2005). In order to use the LR test statistic to compare which distribution on the residual is best suited, a necessary condition is that the parameters are the same (Greene, 2003), given they are nested. The same applies to the deterministic function. By nested, I mean that they share properties. From the above discussion it should be clear that (5) and (8) share properties, and (6) and (7) share properties. Furthermore, (13) and (14) do not share properties.

Vuong's test can be applied on MLE for comparing competing models, no matter if they are overlapping, non-nested, nested or wrongly specified (Vuong, 1989). However, using this test, and checking its validity, would be to push the boundaries of this thesis. Furthermore, its applicability is not found in the general literature other than when having two competing models, while this study has several (Schmidt et al., 2008). Instead I will use (20) to explore

whether the assumption - there exists inefficiency - is correct or not. Likewise, I will test which parameters should be included or excluded in the model.

4.5.2 Scale elasticity

The notion of elasticity of scale is a well defined concept in fundamental economics. Thus, I will not elaborate further on the notion here. However, if ending up with a translog function, the scale elasticity will be estimated, as given by Coelli et al. (2005):

$$\varepsilon(\mathbf{y}_i) = \sum_{n=1}^N (\beta_n + \sum_{m=1}^N \beta_{nm} \ln y_{im}) \quad (21)$$

where ε is the cost elasticity evaluated at the output levels of observation i . If $\varepsilon(\mathbf{y}_i) > 1$, the scale elasticity is below zero because;

$$\text{Scale elasticity} = \frac{1}{\varepsilon} \quad (22)$$

5 The Data

The dataset used for the analysis in this thesis stems from the above mentioned SINTEF report (Kittelsen et al., 2009a). It consists of data on all public financed somatic hospitals in Sweden for the years 2005 and 2006 and for Finland, Denmark and Norway throughout the years of 2005, 2006 and 2007. Also included are private non-profit hospitals (except Finland), and private hospitals financed in a prospective payment scheme as they are publicly financed. However, privately financed activity at private hospitals is not included as to avoid biased expenditure. One Danish, three Finnish and one Norwegian observation are left out due to outlier conditions. In total there are 316 observations in the dataset (Kittelsen et al., 2009a).

Naturally the data stems from different sources since four countries are under comparison. However, all the data is based on official statistical sources in each country, and matched as closely as possible to the Norwegian form of organizing the statistical data. Still, the SINTEF report (Kittelsen et al., 2009a) faces a lot of different challenges in trying to match the data. This is much due to the differences in how each specific country has chosen to organize its health care, and why it was decided to only compare the countries somatic public funded hospitals. I will now give an example of a challenge and its solution, explain the variables as well as describe how activity and costs are measured.

As mentioned, by law there is a clear definition of what hospital care responsibilities consists of in Norway, but this is not the case for Finland and Sweden. Thus, for Norway it is easier to separate specialist care data from primary care than in the rest of the other countries as they have a different owner structure (see chapter 3 above). As an example, it can be mentioned that the Swedish and Norwegian data stems from regional level.

Finland and Sweden both have specialist care activity out in district wards that are not administered by hospitals. Norway also has this type of district activity, but data from Norway is only included if the district ward is run by hospital administration. This is not the case for Sweden and Finland, where there is potential overlap in the data from primary care when considering the activity in the health wards. For Denmark this is not a consideration since they do not have this type of activity. The issue has been addressed and corrected for Sweden and Finland in the data (Kittelsen et al., 2009a).

The discussion above suggests that if activity at one hospital is not conducted at others, it can bias the efficiency measure in the DEA method. For this reason the SINTEF report (Kittelsen et al., 2009a) excluded some of the activity measured in DRG points, such as radiotherapy. Specifically, data on activity in the following DRG groups were excluded in the data for the specialist care:

- Main diagnostic group (MDG) 15: Newborns
- DRG 462: Rehabilitation
- DRG 317: Dialysis
- DRG 409 and 410: Radiotherapy and chemotherapy

(Kittelsen et al., 2009a).

5.1 Activity and costs

Activity at hospital level is measured along three dimensions (Kittelsen et al., 2009a). It is differentiated between 1) outpatient activity, 2) inpatient and 3) daycare patients without overnight stay.

Outpatient activity is regular consultancies, in total number. Inpatients are defined as release date minus in-date plus 1. The inpatients are then grouped based on the DRG system. Measurement of the last type of category, daycare patients, is also differentiated based on the DRG system, but defined as patients who require a bed, but not an overnight stay. In addition, patients who have received specialist care that is separated from normal outpatient care, such as day surgery are also included in the definition. (Kittelsen et al., 2009a). At the end of chapter 5.1.1, table 5.1 gives descriptive on the variables.

Costs are measured by taking into consideration the expenditures and wages. The data on wages are collected through official statistical sources except in Denmark where it is collected

from the office of centralized municipality wage³. Wages include data on all “*employees, regular and variable increases, pension and employers contributions*” (Kittelsen et al., 2009a, p. 94).

While wages are on a detailed level, the expenditure costs are only available on an aggregate level in the form of resources used and its attached monetary value. Unfortunately, capital costs are non-comparable due to considerable differences in how the countries depreciate capital costs (Kittelsen et al., 2009a).

5.1.1 The variables

realcost is real costs in Norwegian 2007 Kroner expressed in billions, adjusted by a deflator subject to wage differences and price differences. I will return to the deflator in chapter 5.2.

outpatients are number of outpatients in polyclinic activity.

drg inpatients are total number of DRG points related to inpatients, that is category 2 in chapter 5.1 above.

drg daycare is total number of DRG points related to daycare patients, that is category 3 in chapter 5.1 above.

region Norway has 4 health care regions; 1) South Eastern, 2) Western 3) Middle and 4) Northern health region. These are included in the analysis to see if there is any discrepancy within Norway.

capital city the capital city variable is included in order to differentiate if city hospitals characteristics are different from non capital areas.

university hospital variable is included in order to see if the university hospitals are different from other types of hospitals. Literature, such as Medin et al. (2010), suggests this might be the case.

³ Fælleskommunale løndatakontor

los deviation is a variable that encompasses length of stay (LOS) deviation from the average patient stay sorted by DRG specificity.

share of outpatients is included in case the policlinic activity contains measurement error. This is not adjusted for DRG or case mix.

cmi (case mix index) is a ratio of total DRG points over number of inpatients plus daycare patients, and is thought to reflect the case mix.

Table 5.1 Descriptive statistics

Variables	Country			
	Finland	Sweden	Denmark	Norway
Number of observations	96	40	105	75
Real costs in Billions #	1.112	4.812	1.516	1.864
Outpatients	150 128	368 134	178 620	129 609
DRG Inpatients	22 516	65 262	22 516	31 447
DRG Daycare	3 119	18 000	2 651	4 044
Share of University hospital	0.156	0.250	0.381	0.200
Share of Capital City Hospitals	0.0313	0.050	0.257	0.160
CMI	0.848	0.655	0.915	0.918
LOS Deviation	0.968	1.118	1.017	0.859
Share of Outpatients	0.841	0.731	0.865	0.773

Real costs in NOK 2007

5.2 Selecting deflator

Nominal values change from year to year due to differences in, for instance, inflation and local wage negotiations. Thus, a deflator is needed in order to provide solid comparability over time between the countries in real terms and real change.

The deflator OECD has constructed could be used. However, the OECD deflator is adjusted based on a nation's gross domestic product (GDP). In the SINTEF report (Kittelsen et al., 2009a) it is argued that this is erroneous since by considering the GDP one includes other sectors than the labor intense health care sector, thereby overestimating the discrepancies of input usage between the choice of base country compared to the other countries.

The deflator used in the SINTEF report (Kittelsen et al., 2009a) is calculated by holding the quantity of resource usage fixed in the year 2007. The average nominal cost of the same quantity is then divided by the fixed quantum, which makes the deflator. The nominal costs are consequently divided by the deflator for each year, and thus adjusted to real costs, comparable across the nations across time, as was done in chapter 3 above. The report uses the available statistics on wages and constructs a wage deflator, and a purchasing power parity index from OECD in order to deflate the other inputs (Kittelsen et al., 2009a).

Through constructing their own deflator, not based on the GDP, but rather from the sector which is analyzed, the validity of the SINTEF report results are improved. According to Coelli et al. (2005, p. 155) *“the deflator selected must relate to the commodities that constitute the aggregate as closely as possible..., the domain of the price index should be the same as the domain of the value aggregate”*. This requirement is fulfilled, thus I keep the same deflator in this thesis as in the SINTEF report.

5.2.1 Other considerations

In SFA modeling, one of the underlying assumptions is that the technology is equal for all firms and that the environment facing the firms is similar (Coelli et al., 1999). Yet, we know

that technology changes over time. One common way of taking into account the effect time has on technology is to include a time trend in the functional form of the model. A linear model would in the margin create an inverse relationship to the output, while a C-D function gives a constant technology, and a trans log opens the possibility of both increased and decreased effect of the technology over time (Coelli et al., 2005).

6 The model(s)

I have chosen to continue on the work of Battese and Coelli (1993) and Battese and Coelli (1995). In 1993 and later in the paper from 1995, Battese and Coelli derived an SFA with a production function using panel data in a time-variant efficiency model, thereby extending previous cross-sectional models. Besides accounting for time in both the deterministic function and efficiency term, they also made assumptions regarding the use of parameters directly in the efficiency term using a truncated distribution. For example, in the 1995 model they included parameters which explain the farmers' age and level of schooling beside time, as explanatory variables for differences in efficiency among the paddy farmers' over a 10 year period. However, as this dataset does not have many years of observations their time trend model is not directly applicable to this thesis.

In the random effects model, I will use parameters to account for time differences, countries, years and differences inward the Norwegian health regions. This is a convenient method when there are only a few years of data available and it is less restrictive than using time trends (Jacobs et al., 2006). Thus, I make no assumptions about whether the technology change over time or not, or if it changes between countries; I also let the efficiency of each unit as well as the magnitude of efficiency differ. However, I will include parameters that can explain the efficiency differences in the u_i . By doing these two moments simultaneously, heterogeneity (observable and unobservable effects) is accounted for in both the deterministic model and in the inefficiency model as Greene advises (Schmidt et al., 2008).

Still, this thesis will draw upon the work of Battese and Coelli (1995) and use dummies for the years 2005 and 2006, while 2007 becomes the base year. In this, the otherwise strong assumption of independence between the y_i , u_i and v_i is relaxed (Schmidt et al., 2008). The variables for university hospital, capital city, LOS deviation, CMI and share of outpatient activity are expected to influence the inefficiency component⁴, and include them in the u_i accounts for heterogeneity. For ease of notation, let me abbreviate these variables as *environ*. The LR test will be applied in order to test if the expectation of the *environ* variables effect is correct.

⁴ For explanation of the variables see chapter 5.1.1 above.

In addition this thesis will control for the parameters representing country, regions and time both in the deterministic part as well as in the efficiency part. Output variables are as previously mentioned. Given the different functional forms discussed in chapter 4.4, a generic model can be expressed as:

$$\ln C_i = \alpha_0 + f(y_{i1}, \dots, y_{im}) + \sum_j^N \gamma_j z_{ij}' + v_i - u_i \quad (23)$$

$$\text{and,} \quad u_i = \sum_k^N \delta_k z_{ik}'' + w_i \quad (24)$$

(23) reflects the deterministic model, while (24) is the inefficiency model. In (23) α_0 is the constant. In (23) and (24) γ_j and δ_k are the parameters to be estimated, while z represents non-product explanatory variables, such as country, region and the time dummies. The other variables are as previously explained. Equation (24) is the equal of (15) and expresses the efficiency component as a function of variables. As discussed earlier in this thesis (23) is altered depending on if a C-D, Quadratic or translog model is assessed.

6.1 Estimating the models

Extensive work has been done on constructing the models to assure validity and reliability. They were estimated using STATA 11.0, and its functions for Frontier (StataCorp, 2009). Manipulation was often necessary with respect to algorithms and tolerance of the log likelihood in order to meet the requirements for an ML model, as stated in StataCorp (2009, p. 1012-1015). Often the default Newton-Raphson algorithm works itself into a bad region of the likelihood, and thus, did not find convergence or declared a premature model at a local

point instead of a global. By altering the algorithm⁵ to alternatives or combinations of algorithms provided in STATA, a global maximum was found.

As mentioned both in Coelli (2005) and in StataCorp (2009), an important requirement to secure that a global maximum have been reached, is that the first derivatives of the log likelihood function are close or equal to zero. This can be controlled for in STATA by the gradient function. Furthermore, STATA offers the possibility to specify that the log likelihood tolerance with respect to the gradient should be equal to zero or as close as possible (StataCorp, 2009). The zero tolerance restriction has been applied if necessary, and gradient control exercised on all models.

A further requirement for a valid model is that it does not fail the regularity conditions discussed on page 20, like displaying negative marginal costs which would break a condition. The procedure presented in Salvanes and Tjøtta (1998) have been applied to make sure the second order flexible models fulfill the regularity conditions⁶. One procedure involves calculating a consistency region where the function adheres to the regularity conditions. In this thesis however, it has only been calculated whether each of the hospital observations is in the consistency region.

Furthermore, in conducting the LR test(s) I have used both the Chi-square distribution, $\chi^2_{\alpha=0,05}$ table provided in Thomas (2005), and the less restrictive mixed chi-square distribution table provided in Kodde and Palm (1986) to determine the critical significance level of $\alpha = 0.05$ value. In this way type 1 error is controlled for. Thus, if the value of the test statistic is between the critical values of the two distributions I will use the more conservative in order to avoid type 1 error - rejecting a true $H_0 : \mu = 0$.

⁵ Available in STATA 11.0 are / or in combination(s), the Newton-Raphson, Berndt-Hall-Hausman, Davidon-Fletcher-Powell and Broyden-Fletcher-Goldfarb-Shanno algorithms.

⁶ Calculations executed by Sverre A. C. Kittelsen

7 Results

Table 7.1 Assessment of different models. LR test under: $H_0 : \sigma_u^2 = 0$ against $H_1 : \sigma_u^2 \neq 0$.

Model	Cobb Douglas. Half Normal	Cobb Douglas. Exponen- tial	Cobb Douglas. Truncated Normal	Translog. Half Normal	Translog. Exponential	Translog. Truncated	Quadratic. Half Normal	Quadratic. Exponen- tial	Quadratic. Truncated
Log Likelihood value	139.105	160.420	No conver- gence achieved	No conver- gence achieved	186.393		Unable to calculate		Unable to calculate
<i>environ</i> included	No	No	No	No	No	No	No	No	No
u_i	-2.610	-3.780			-3.813				
Gradient value	2.21e- 08	7.32e-06			1.63e-07	0.082		246.97	
LR test statistic value	62.82; Reject H_0	110; Reject H_0			75.53; Reject H_0				
Valid	Yes	Yes	No	No	Yes	No	No	No	No

Table 7.1 above describes the models onset. First I tested if the deterministic function with all the dummy variables was significant when assessing $\varepsilon = v_i + u_i$. Put in another way, I tested if the technical inefficiency component in the model was different from zero. If not, the model would be equal to an OLS. Secondly, as discussed in chapter 6.1 above, the gradient should be as close as possible to zero. Two models were unable to converge. STATA was unable to calculate the initial values for two of the models, in another two models the gradient was not close enough to zero. This was in spite of trying to force the gradient to zero and using different algorithms. Note that no model under the assumption of truncated normal distribution on the residuals managed to complete. As can be seen, only three models pass the initial stage.

In the next stage the LR test statistic will be used to test if the *environ*⁷ parameters in u_i are significantly different from zero.

The LR test statistic is as previously explained in (19) above, while the actual hypotheses that have been tested for are: $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$.

In table 7.2 below, the three models are shown, and whether the u_i is significant when conditioned with the *environ* variables. All three models pass the LR test for further modeling. Thus, H_0 : that the *environ* parameters are equal to zero, is rejected for all three models.

Table 7.2 Assessment of *environ*

Model	Cobb Douglas Half Normal	Cobb Douglas Exponential	Trans Log Exponential
Log Likelihood value	154.198	170.324	198.739
<i>environ</i> included in u_i	Yes	Yes	Yes
Gradient value	4.22e-07	2.68e-08	1.89e-08
LR test statistic value	30.186	19.808	24.692
Critical value, 5 degrees of freedom (Kodde and Palm, 1986)	10.371	10.371	10.371
Valid	Yes	Yes	Yes

Table 7.2 above illustrates the *environ* variables contribution to explain inefficiency. This is shown by the large value of the LR-test statistic value, and the change in the log likelihood value from table 7.1

⁷ See page 35 above

The testing is continued by first investigating the deterministic part. The question is whether the parameters contribute significantly to the model. This testing is done on group based level. That is, without years against with, without regions against with, etc.

Years prove to be significantly different from zero for one of the three models. When controlled for regions, they prove to be insignificantly different from zero except in the translog model with exponential distribution. The country parameters are significantly different from zero in all models. Below is the third table reflecting the LR-tests of the deterministic part.

Table 7.3 Regularity and parameter test. Deterministic component.

Model	Regularity test and LR test statistic value. Critical Kodde and Palm (1986) $\chi^2_{\alpha=0,05}$ value in (.)		
	Cobb Douglas Half Normal	Cobb Douglas Exponential	Translog Exponential
Years tested for $H_0 : \gamma_j = 0$ against $H_1 : \gamma_j \neq 0$. Degrees of freedom equals 2.	6.791 (5.138)	5.134 (5.138)	5.850 (5.138) [5.991]
Gradient value	4.84e-06	1.94e-08	8.13e-09
Regions tested for $H_0 : \gamma_j = 0$ against $H_1 : \gamma_j \neq 0$. Degrees of freedom equals 3.	2.907 (7.045)		14.005 (7.045)
Gradient value	1.30e-06	1.043	3.47e-06
Countries tested for $H_0 : \gamma_j = 0$ against $H_1 : \gamma_j \neq 0$. Degrees of freedom equals 3.	96.838 (7.045)		182.799 (7.045)
Gradient	1.82e-07	0.000112	4.56e-08
Pass the regularity test	Yes	Yes	68 out of 316 observations are outside the consistency region.
Valid	Yes	No	Yes

From table 7.3, it is clear that the C-D model with a half-normal distribution is a valid model. The year and country parameters are undoubtedly estimated as significant, but not regions. The gradient resembles zero in all three cases. However, it is a different story with the C-D model in the case with Exponential distribution. Unable to push the gradient towards zero in two of the three tests despite extensive manipulation, the latter model is discarded.

The Translog model is also considered valid. Only 68 out of 316 observations are outside the consistency region. The translog model is therefore not considered to break the regularity conditions. At first glance, the latter model agrees with the C-D model in considering the year parameters, and its gradient is close enough to zero. However, as can be observed the test statistic for years falls in between the two critical values. As discussed above, I will therefore use the conservative Thomas (2005) distribution in order to avoid type 1 error. In other words, years are insignificant in the translog model. It disagrees with the Cobb Douglas model when it comes to regions. The region variables are therefore not rejected from the cost function in the translog case. As in the case with the Cobb Douglas model, the translog model considers the country parameters significantly different from zero.

Observe how the country parameters affect both models. As can be seen from the LR test statistic value, the countries contribution to the model is substantial.

Consequently, this thesis is left with two models. Let me term them model CD and model T. In order to avoid any confusion, I will be explicit, model CD is the Cobb-Douglas model, and model T is the Translog model. In view of the fact that the two models are not nested, the LR test cannot be applied in order to differentiate the most significant model. Instead further sensitivity analysis must be undertaken. The LR-tests for the sensitivity analysis of the efficiency component are presented in table 7.4.

Table 7.4 Sensitivity analysis of the efficiency component in model CD and T, including *environ* in u_i . Hypothesis $H_0 : \delta_k = 0$ against $H_1 : \delta_k \neq 0$

Model	Log likelihood value		LR statistic test		Critical values		Degrees of freedom
	CD	T	CD	T	Kodde and Palm (1986)	Thomas (2005)	
Years included in u_i	153.19	196.83	0.890	2.038	5.138	5.991	2
Regions included in u_i	164.90	206.54	24.317	Gradient =0.827	7.045	7.815	3

In table 7.4 the full hypothesis testing for whether to include years and regions in the efficiency component is presented with the LR statistic as a decision making instrument. The first column, log likelihood value, gives the log likelihood value for model CD and T respective. The second column, LR statistic test, gives the LR statistic value for model CD and T respective. The third column, critical values, gives the critical Kodde and Palm (1986) and the Thomas (2005) value for rejecting the hypothesis H_0 : parameters value is equal to zero. The final column gives the value on the degrees of freedom.

From the horizontal perspective, it can be seen that I firstly test the significance of the year variables, secondly the significance of the region variables in the u_i component. As evident, for both models, the year variables are insignificant in the u_i component, and thus H_0 : that the parameters are equal to zero, cannot be rejected.

Observe that for both models the log likelihood value increases when including the region variables in the u_i component. However, it is only model CD that is valid since model T has a gradient > 0 .

This leaves the Cobb Douglas model, model CD, with regions included in the efficiency component along with the *environ* variables, and from table 7.3, without regions in the deterministic function.

The translog model, model T, is left as before the tests in table 7.4. In other words, the sensitivity analysis did not improve model T any further.

It is therefore argued; based on the parsimonious criteria above I ought to use model CD as the “best” model since “it get’s the job done”. However, there is a problem. Cobb Douglas imposes constant scale of elasticity. I believe that the scale elasticity varies between the units. In other words, a larger hospital must either have diseconomies of scale or economies of scale compared to a smaller hospital, they cannot be equal.

According to the flexibility criteria it is preferred with a second order function since it is more exact in estimating several points in space. In other words, the second order flexibility of the translog functional form allows for both varied scale elasticity and varied efficiency between the units.

Though the models are not nested, they have the same dependent and independent variables. Therefore I can compare the log likelihood values, where the value of model T is the larger of the two. But lacking a LR test I cannot conclude that it is significantly better. Adding the log likelihood value to the flexibility criteria, I therefore conclude that model T, the translog model, is the “best” model. Below I present model T in full length, in addition to model CD.

Table 7.5 The final models T and CD. Norway in 2007 is base reference country. Values in the inefficiency part u_i represent percentage distance from the frontier where negative values are increase in efficiency.

Dependent variable is real costs.	Model T		Model CD		
	Coef.	z	Coef.	z	
Ln outpatients	0.083	0.883	0.579	0.000	***
Ln DRG inpatients	0.566	0.122	0.381	0.000	***
Ln DRG daycare	0.294	0.273	0.068	0.010	**
Product of Ln outpatients × Ln DRG inpatients	-0.451	0.002			***
Product of Ln outpatients × Ln DRG daycare	0.103	0.093			*
Product of Ln DRG inpatients × Ln DRG daycare	-0.305	0.001			***
(1/2)Ln outpatients ²	0.337	0.016			**
(1/2)Ln DRG inpatients ²	0.784	0.000			***
(1/2)Ln DRG daycare ²	0.200	0.010			**
Year 2005			-0.024	0.109	
Year 2006			0.006	0.698	
Finland	-0.313	0.000	-0.524	0.000	***
Sweden	-0.019	0.691	-0.019	0.594	
Denmark	-0.094	0.023	-0.300	0.000	***
Western Health region Norway	0.041	0.212			
Middle Health region Norway	0.089	0.021			**
Northern Health region Norway	0.124	0.000			***
Constant	-9.949	0.000	-10.893	0.000	***
$\ln \sigma_v^2$	-5.752	0.000	-6.784	0.000	***

Table continues

Table 7.5 continued

	Model T		Model CD		
	Coef.	z	Coef.	z	
$\ln \sigma_u^2$					
Western Health region Norway			-2.377	0.000	***
Middle Health region Norway			-1.900	0.000	***
Northern Health region Norway			-0.844	0.072	*
Share of outpatients	-4.917	0.139	-6.851	0.000	***
LOS deviation	3.441	0.000	1.502	0.002	***
CMI	0.564	0.668	3.482	0.000	***
University Hospital	0.248	0.615	0.378	0.187	
Capital City	-0.580	0.225	-1.280	0.000	***
Constant	-3.729	0.132	-1.296	0.399	
Average Scale Elasticity	0.931		0.972		

***, ** and * indicate significance level $\alpha = 0.01, 0.05$ and 0.1 respectively

In table 7.5 above, the final translog model T is presented together with the Cobb Douglas model CD for comparison. Firstly, note that the estimated output coefficients can be interpreted as output cost elasticities in the CD case, although in T these elasticities also depend on the second order terms. Though the outpatients variable is insignificant in model T, it is significant in model CD. From model CD it can be observed that the increase is estimated to be 0.579 %.

Secondly, at the bottom of the table is the summarized scale elasticity. As can be seen, the translog model exhibit an average elasticity below 1, indicating diseconomy of scales of about 7 %. In model CD the average scale elasticity is constant, exhibiting diseconomies of scale in the order of 2.8%.

The year variables are insignificant at the 5 and 1 % level in the CD model. The fact that there seems to be no significant difference throughout the years 2005 – 2007 in both models could indicate that the technological pathway appears to have been more or less equal for all countries.

In both models the Finland and Denmark variables are observed to be significantly different from Norway, which is the reference country. Actually it provides information that Finland, Denmark and Norway have their own country specific frontiers. The estimated coefficients suggest that Finland and Denmark has a lower cost level than Norway in providing their specialist health care. Sweden is in both models estimated to be insignificantly different from Norway.

From model T it can be seen that there exist cost differences inward the Norwegian frontier. Norway's frontier includes the health region south east as reference. Two health regions are estimated to have larger costs than the south eastern region, that is, the Norwegian Middle and Northern Health Regions. It can therefore be argued that these regions have a different productivity than the South Eastern Health Region, thereby their own frontiers. The variables were excluded in the C-D model since their contribution was insignificant. I will return to the % difference in cost expenditure between the countries and regions below.

From the σ_v^2 , it can be observed that the noise component in the residual is significant, while this is not the case for the constant in σ_u^2 . The insignificant constant in σ_u^2 suggests that the variation in the efficiency component is well described by the parameters. As evident in the C-D model and perhaps not surprising, outpatients tend to increase efficiency on average, while e.g. increased length of stay decreases efficiency.

As opposed to the C-D model, the second order flexible model does not leave much unexplained efficiency for the u_i . Only the LOS variable is significant. This suggests that the parameters in the cost function, i.e. the DRG weights, are suited to describe the efficiency differences as well as the cost differences. Consequently, model T suggests that there is variation in the country specific productivity defining the frontiers, but not the efficiency level behind the frontier. The difference in the level of the country specific frontier can be thought of as a cost penalty for production in that specific country. A lengthy discussion of the parameters in the inefficiency component follows. Let me first present one more table, with

the percentage difference in cost levels, reflecting the cost penalty between the countries and the regions.

Table 7.6 Percentage deviation between the countries and regional frontiers. Positive values indicate lower costs. 95 % confidence intervals.

	Country			Region		
	Finland	Sweden	Denmark	Western Health region	Middle Health region	Northern Health region
Model T	0.269 ***	0.019	0.090 **	-0.042	-0.094 **	-0.132 ***
Confidence interval	(0.318-0.216)	(0.106-(-0.077))	(0.161-0.013)	(0.023-(-0.110))	(-0.013-(-0.180))	(-0,059-(-0.210))
Model CD	0.408 ***	0.019	0.259 ***			
Confidence interval	(0.437-0.378)	(0.084-(-0.051))	(0.303-0.213)			

***, ** and * indicate significance level $\alpha = 0.01, 0.05$ and 0.1 respectively

As observed in table 7.6 the frontiers become more explicit. Given the size of the hospitals in model T, Finland has an expenditure related to their specialist healthcare that is 26.9 % lower than the Norwegian. Similar interpretation is valid for Denmark with the magnitude of 9 % lower costs compared to Norway. In other words, there is a cost penalty for providing specialist health care in Norway by the magnitude of 26.9 % as compared to Finland, and 9 % as compared to Denmark. Mind that the base reference is the South Eastern health region of Norway.

In model CD, where the reference case is Norway as a whole, we see that the results are even more vigorous. Keep in mind though that the C-D model implies constant elasticity and is not considered to be the “best” model. Nevertheless, given the hospital sizes in the C-D model,

Finland emerges as having incredible 40.8 % lower costs in the delivery of specialist care. Estimation of Denmark suggests 25.9 % lower costs.

Let me return to model T. From table 7.6 it can also be observed differences within Norway. The health regions Middle and Northern are statistical different from the South Eastern region. In fact, given the hospital sizes, the South Eastern health region delivers its specialist health care in a cheaper way. The Middle Health region is estimated to have 9.4 % higher costs associated to its provision of health care, while the Northern Region is estimated to have 13.2 % higher costs. Thus the region specific frontiers lie above that of the Norwegian and there is a cost penalty associated with providing specialist health care in these specific regions. An illustration of the country specific frontiers is provided in the next chapter.

8 Discussion

As can be seen from the preceding chapter, extensive work was done to estimate the models. Out of nine models, three were excluded as the first derivatives never approximated zero, two did not converge and two were incalculable. No model was excluded based on the regularity test of Salvanes and Tjøtta (1998). Further work was put in sensitivity analysis by omitting and including parameters and testing for their relevance with the LR test statistic.

The result, model T, is very interesting and further research could be done to validate the outcome. For example, one could stepwise add the parameters one by one and run the LR test as opposed to what I did. The flexibility of the translog function makes the estimation of the frontier closer to the observations, that is, the translog function resembles DEA estimation. The results from model T are therefore not unexpected. As reported in chapter 3.2 above, similar results have previously been found in Linna et al., (2006) when estimating cost efficiency using the DEA method. The findings also support the SINTEF report (Kittelsen et al., 2009a).

Furthermore, the results explicitly demonstrate that it is important to use SFA in estimating efficiency since the disadvantages of DEA is the strength of SFA, and vice versa. Recollect that DEA assumed no measurement error and that all deviation from the frontier was regarded as inefficiency, while, SFA assumed that the residuals consisted of both. On the other hand, SFA must assume distributional form of the residual and a functional form of the deterministic component, whereas DEA need not. Thus, both methods ought to be used as they complement each other.

Kooreman (1994) and Zuckerman (1994) both discuss the complementarities of the two methods. They point to the fact that DEA can estimate technical efficiency while SFA can estimate a combination of technical efficiency and allocative efficiency. The duality problem applies to the results in this thesis, as with all models where price on output is non-existent.

The tests were controlled for type 1 error as discussed above, and the more conservative chi-square distribution of Thomas (2005) was used once. The findings point in the same direction given the differences between model T and CD. There are considerable - some would say extreme - differences in the cost efficiency between the countries, as observed in table 7.5.

By including explanatory variables in the u_i component, this thesis allows the technical efficiency to vary with the environmental variables in u_i . Since the u_i contains a vector of explanatory variables, the inefficiency effects in the model are no longer identically distributed, but once estimated they are constant. Note that the significance of the explanatory variables in u_i are tested for, and not included at the onset. Thus, heteroskedasticity, time and environmental variables are correctly accounted for.

Because the data provides observations at several points in time, the assumption of independence is relaxed as discussed in chapter 6. Instead, SFA imply strong assumptions regarding the distribution of the inefficiency term. Furthermore, I have tested for both time-invariant and time-varying inefficiency in model CD and T by assessing the significance of the year variables. It turned out that the year variables were insignificant when included in the deterministic as well as the inefficiency component in model T. In model CD, years were included in the deterministic component but not in the efficiency component. Thus, the final model is a time-invariant (constant inefficiency) model in assessing the efficiency behind the frontiers.

The result confirms that the level of technical efficiency has been more or less the same throughout the three years with available data. In the literature (Jacobs et al., 2006), it has been argued that if it is believed that efficiency vary over time it must be modeled. I here show that it is, in my opinion, more correct to test for it.

The productivity can be decomposed into differences in scale efficiency, cost efficiency (or technical efficiency) and the differences in the level of the frontiers. In this thesis I have estimated the two latter of these three. An attempt to illustrate this relationship is provided in figure 5.

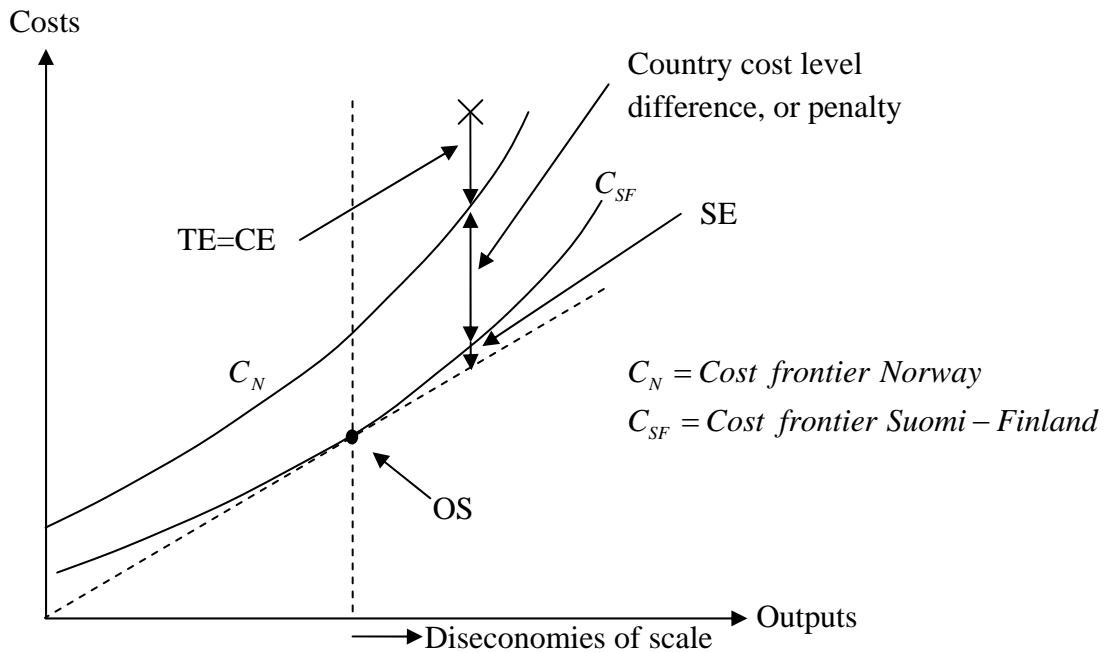


Figure 5. The cost frontiers of Norway and Finland.

Figure 5 illustrates the country specific frontiers. The frontiers are included to provide a deeper understanding, but under no circumstances do they reflect reality. Costs are measured on the vertical axis, while outputs are measured on the horizontal axis. The stippled line emerging from the origin is the tangent to the minimum average cost of the Finnish frontier, or just a hypothetical constant return to scale cost function.

The vertical stippled line indicates the point at which optimal scale efficiency is reached. Any increase in outputs beyond this line (to the right) is associated with a proportionate increase in costs which is larger than the output value. The point indicated by OS, the optimal scale, illustrates the relation. Since the scale elasticity indicated diseconomies of scale and given the hospital sizes in the dataset, my observations are to the right of this line.

The convex line tangent in the OS point represents Finland, illustrating best practice. The fact that the frontiers are convex illustrates variable return to scale. The country specific frontier with increased costs compared to Finland is the frontier of Norway. As can be seen the frontier of Norway indicates higher costs. I have here only illustrated Finland and Norway, as leaving Denmark and Sweden out of the figure does not interfere with the analytical point.

The distance between the Norwegian, C_N , and the Finnish, C_F , frontier is illustrated in figure 5. The percentage deviation was expressed as the cost penalty between the countries, as

in table 7.6. In the same way, scale efficiency (SE) is expressed, but as the distance from a country specific frontier to its minimum average cost line, or the hypothetical CRS cost function, here illustrated by Finland. Scale efficiency is not estimated in this thesis.

The technical efficiency in the u_i component is measured from an observation in space, below the country specific frontier, illustrated by the cross above the frontier of the Norwegian frontier. In model T the variable LOS deviation was positive for all countries, indicating that it is further to “north east” of the country specific frontier and the stippled line representing diseconomies of scale. In other words, it is above the frontier and further to the right of the stippled line, resulting in increased costs and decreased efficiency.

The stochastic error component is not included in the illustration.

8.1 The deterministic component

First a clarification; if a variable is significant, it implies that the coefficient is significantly different from the hypothesis, $H_0 : \mu = 0$. In other words, the variable provides important information at the 5 % level. If it is insignificant, the coefficient is not significantly different from the hypothesis $H_0 : \mu = 0$, where μ represents the coefficient in both instances.

From the coefficients values it can be seen that there is great difference from the Cobb Douglas function to the translog function. The findings contradict the conclusion in Rosko and Mutter (2008) discussed above, and is more in line with Greene (Schmidt et al., 2008), suggesting it is case specific.

In line with previous findings, e.g. in the SINTEF report (Kittelsen et al., 2009a), the frontier for Sweden turns out to be insignificantly different from the Norwegian frontier in both models.

As opposed to previous findings, the Danish frontier is significantly different from the Norwegian. In fact, the CD model estimates that Denmark uses 25.9 % less resources than Norway in supplying their specialist health care. The finding is in addition highly significant

at the 1 % level. The picture is somewhat moderated in model T, the “best practice model”. Here the function represents the data better because of the increased flexibility. In model T Denmark is estimated to have 9 % lower costs compared to Norway. Notice that the costs are deflated to 2007 NOK. The fact that Denmark is estimated to be more cost efficient in providing their specialist health care compared to Norway is in line with Kittelsen et al., (2007) but contradicts the findings in Kittelsen et al., (2009a).

Lastly, the Finnish frontier is also significantly different from the Norwegian. In the CD model, Finland uses 40.8 % less resources in providing specialist health care than Norway does. Again, the results are somewhat moderated in model T but still substantial, to a cost penalty of 26.9 %.

The Northern and Middle health regions are significantly different from the South Eastern health region. In model T they are estimated to use 13.2 % and 9.4 % more resources than the South Eastern health region. In fact, they are so different in their productivity that they have their own frontiers. Considering the budget of 58.3 billion NOK in 2007 (Iversen et al., 2010) for somatic health care in Norway (executed in the health regions), the discrepancy between the regions’ resource usage in providing specialist health care seems enormous.

8.2 The inefficiency

The intuition behind the u_i is that any deviation from a value of 0, that is the frontier, is a result stemming from the behavior of the unit. Thus, if the unit is engaged in activity in an inefficient way, the observed behavior is above the frontier. This was discussed above in relation to figure 5, with regards to technical efficiency.

The variable “share of outpatients” is significant at the 1 % level in the CD model. The magnitude is rather strong, by -6,851. This suggests that a high outpatients share is associated with a high efficiency of hospitals. Perhaps more importantly, it could indicate that a lot of patients are being treated at the wrong point of service. Mind though, the variable is insignificant in model T. One could argue, based on model CD, that outpatients should to a

larger extent be treated elsewhere, at a lower resource demanding level – such as in primary care. However, the argument does not hold for model T which is the preferred model.

Furthermore, the LOS deviation parameter is significant at the 1 % level in both model CD and T. Not surprisingly, the results suggest that the longer the patient stay in hospital the less efficient is the treatment. Longer patient stay is not to be confused with wrong treatment. Nor should it be confused with need of treatment. The deviation from average length of stay is also reflecting patient mix. Thus, the more demanding the patient is, the longer is the tendency for staying in hospital beyond the average patient.

As opposed to model T, it can be observed in model CD that the case mix index parameter is significant and positive. Model CD suggests there might be correlation to the length of stay variable. However, it can also be the case that the DRG points do not fully compensate the patient mix. As mentioned earlier, Medin et al., (2010) finds similar results, in addition to discussing the efficiency of university hospitals.

The university hospital parameter is positive suggesting that these types of hospitals are inefficient compared to non-university hospitals. However, in this thesis the parameter is insignificant in both models.

In contrast to model T, the CD model estimates the capital city parameter as significant. In the latter model metropolitan hospitals are efficient compared to non-metropolitan hospitals. This is opposed to the findings in for instance Zuckerman et al., (1994) who find inefficiency in both city and non-city hospitals. In addition, the capital city parameter is highly significant. One possible explanation is that patients living in urban areas are different from rural areas, while another reason could be connected to the fact that urban hospitals might have lower costs connected to logistics. If so, this could indicate that the scale elasticity of capital city hospitals is different from that of other hospitals, thus be an argument for large metropolitan hospitals since urban conditions seem to increase efficiency.

Two of the region variables are strongly significant in the CD model. Interestingly, the Western and Middle Health Regions seem to be more efficient than the reference region, South East. This is in contrast to model T where the Middle Health Region indicated increased costs. A paradox thus seems to emerge. If the two models are combined, the Western region is seen to have high expenditures but at the same time it is considered to be

more efficient in treating its patients. It would be interesting if future research was to investigate this matter further.

Let us return to model CD, with the Western and Middle health region in the u_i . If the model was to include variables that could explain the combination of rurality and size of hospitals it might have provided a better explanation. Future research could try and account for the underlying organization of providing specialist health care – that is variables which in addition can explain allocative efficiency. Alternatively, further research could be conducted to see if this dataset can be applied to shadow price models in order to estimate allocative efficiency, if there exists any. A good start to investigate this possibility would be to consult Schmidt et al., (2008, p. 40, 187, 199).

Thus, the result in table 7.5 suggests that the major part of the differences in expenditure is due to the productivity specific to each country, and not the distribution of efficiency behind the frontier. For Norway regional differences within the country are found, as evident from the separate frontiers of the Middle and Northern Health Regions. In addition, each country (except Sweden) has its own specific frontier, and the technical efficiency, or cost efficiency, is the difference to each country's frontier.

If the differences had been due to discrepancy in incentives, the efficiency behind the frontier would have been different in model T. Nor are the differences due to the level of costs as wage, nominal values or up-coding of DRG points (incorrect DRG codes selected to obtain higher compensation level) since such differences have been eliminated in the dataset. Thus, the differences must be understood to originate from other unexplained factors such as performance, culture, or organization.

8.3 The elasticity of scale

The above discussion of the inefficiency section has some common suggestions. Firstly, the share of outpatients contributed positively to efficiency in both models. However, it was argued that the outpatients might be treated at the wrong point of service since they contribute

to increased costs. Secondly, Norway emerge as less cost efficient than Denmark and in particular Finland, which represented “best practice”. Together these two points, suggest that an increase in decentralized health care with smaller specialist health care units, as are found in Finland are preferable.

This is also reflected in the average scale elasticity which in model T had a value of 0.931. As mentioned, it indicates diseconomies of scale. In standard economics this implies that the average units are too large. If considering one firm owning all the units in the dataset, one implication would be that it needs to split some of its units, or change its way of doing business. Most likely both actions would be needed in order to reverse the diseconomies of scale and benefit increased efficiency. In policy it can be translated to decentralization, moving both decision making level and point of service to a lower level.

In health care there are many more issues than efficiency to consider, such as adhering to the national goals of health care provision. Thus, the results in model T should not be taken as absolute, rather as signals. In the end, the question of how to deal with the fact of diseconomies of scale in the specialist care sector throughout the Nordic countries is a matter of policy. However, it suggests that public funds are used inefficiently as it is now.

8.4 Limitations

The observations in this dataset stem from four sources. With increased number of sources potential for misspecification is amplified compared to the case of one source. Limitations were connected to the completeness in the data for the countries due to differences in reporting. The data for Sweden are only for the years 2005 and 2006 which might contribute to the insignificant difference between Norway and Sweden.

Furthermore, all the countries, especially Denmark who has their own DK-DRG system, did not use the NorDRG system in the same way. But because DK-DRG is built on the same overarching structure as NorDRG, comparability is possible on aggregated level. In addition, the structural organization of providing health care in the four countries differs.

As discussed in chapter 5, all of these factors could impair the results of this thesis. However, this is also why the study was limited to compare only public funded somatic hospitals. Thus, the results should be valid.

9 Conclusion

The purpose of this thesis has been to examine whether the Nordic countries differ in costs and efficiency in providing specialist health care. Previous studies have considered the issue using data enveloped analysis (DEA), which is the current norm of methodology. However, they have established a joint frontier, something which can be termed a mutual global frontier. By doing so, all countries contribute in establishing one frontier, only providing bits to the larger puzzle of efficiency, production and / or cost differences, with a potential to bias policy decisions as the estimated differences might be erroneous. Consequently, the global frontier in the SINTEF report (Kittelsen et al., 2009a) is not fully representative.

It is argued in this thesis, that the methodology of Stochastic Frontier Analysis (SFA) is preferred when wanting to separate the country specific effects. The method provides means to differentiate (in)efficiency from noise, as contrary to DEA. I have used SFA to decompose the cost efficiency into country specific frontiers.

As previous findings (Kittelsen et al., 2009a) suggest, Finland is exceptionally efficient in providing specialist health care. In fact, in this thesis Finland emerge as 26.9 – 40.8 % more cost efficient than Norway depending on the model used. Previously, Linna et al., (2006) have suggested around 17-25 % for Finland. The findings in Kittelsen et al., (2009a) suggested Finland as significantly different and Sweden and Denmark as insignificantly different from Norway when comparing productivity. The results in this thesis suggest that not only Finland but also Denmark is significantly different from Norway. Just as important, they strengthen the findings in Kittelsen et al., (2009a). Consequently, it is argued that both methods, DEA and SFA, should be considered as complementary and hence, both should be applied when assessing efficiency estimates as to avoid bias in policy.

Heteroskedasticity is accounted for in the models. Among several findings, model CD suggests that many outpatients are treated at the wrong point of service. Even though they enhance the efficiency of specialist care, they do so in a cost driven way. Furthermore, rural regions were more efficient than the South Eastern Region of Norway, but at increased costs. Metropolitan areas on the other hand are seen to increase efficiency. The scale elasticity suggests diseconomy of scale.

Since model CD is not the preferred model, the suggestions in the preceding paragraph are not the final conclusions. The translog model was considered the best model, but I was unable to confirm this by a statistical test. In model T it is argued that there are country specific frontiers. Albeit there is a cost penalty of providing specialist health care in Norway as compared to Finland and Denmark, it seems that each unit within each country is close to its own frontier, indicating cost efficiency under prevailing conditions. The closeness can be observed from the fact that it is only LOS deviation which is significantly different from zero of the *environ* variables.

Model T also indicates that there is large discrepancy inward the Norwegian regions. In fact, it is suggested that two of the four regions in Norway (Middle and Northern) are so different from the South Eastern region that they have their own frontiers. These regions production is at a higher costs level than the South Eastern region, and consequently there is a cost penalty for providing health care in the Middle and Northern Health Region.

It seems reasonable to suggest that there are certain disease specific procedures that Norway might be more efficient in conducting than Finland, but are there quality differences? Future studies of efficiency differences which include variables explaining quality discrepancies are of interest. Since there is a cost penalty of providing specialist health care in Norway it is interesting to investigate if the provision of the Finnish specialist health care is qualitative different from the Norwegian. Other forms of best practice could be learned from combining quality assessment in for instance cost of care of specific diseases, or quality assessment in differences of length of stay. In addition, variables that can explain allocative efficiency are of interest.

Much attention is needed in finding ways of comparing functional forms and distributional assumptions within future research when having more than two competing models. Although this thesis provides exhaustive testing for estimating the optimal functional form of the deterministic component and the best distributional assumption of efficiency, it remains a case specific concern. I consider it beneficial if future research was to differentiate the optimal functional form and distribution in a generic test procedure.

The findings in this thesis suggest that the country specific frontiers differ due to the productivity conditions prevailing in each country, as indicated by the differences in cost efficiency, while the distribution of efficiency behind the frontier does not differ among the

countries. The conditions prevailing in each country remains unexplained. Consequently, Finland seems to exhibit best practice and Norway could reduce its expenditure by learning from the Finnish way of organizing and providing the specialist health care.

References

- AIGNER, D., LOVELL, C. & SCHMIDT, P. 1977. Formulation and estimation of stochastic frontier production function models,. *Journal of Econometrics*, 6, 21-37.
- AIGNER, D. J. & CHU, S. F. 1968. On Estimating the Industry Production Function. *The American Economic Review*, 58, 826-839.
- ANDERSEN, E. B. 1970. Asymptotic Properties of Conditional Maximum-Likelihood Estimators. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32, 283-301.
- BATTESE, G. E. & COELLI, T. 1993. *A stochastic frontier production function incorporating a model for technical inefficiency effects*, University of New England. Department of Econometrics.
- BATTESE, G. E. & COELLI, T. J. 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical economics*, 20, 325-332.
- BOOTH, W. C., COLOMB, G. G. & WILLIAMS, J. M. 2008. *The craft of research*, Chicago, University of Chicago Press.
- CARLTON, D. W. & PERLOFF, J. M. 2005. *Modern industrial organization*, Boston, Pearson/Addison Wesley.
- CAVES, D. W. & CHRISTENSEN, L. R. 1980. Global properties of flexible functional forms. *The American Economic Review*, 70, 422-432.
- COELLI, T., PERELMAN, S. & ROMANO, E. 1999. Accounting for environmental influences in stochastic frontier models: with application to international airlines. *Journal of Productivity Analysis*, 11, 251-273.
- COELLI, T. J., RAO, D. S. P., O'DONNELL, C. J. & BATTESE, G. E. 2005. *An introduction to efficiency and productivity analysis 2nd edition*, Springer Verlag.
- DST.DK. 2001. *Befolkning og valg. Befolkningsprognoser 2001-2040* [Online]. Statistics of Denmark. Available: <http://www.dst.dk>, <http://www.dst.dk/pukora/epub/Nyt/2001/NR224.pdf> [Accessed March 8 2011].
- FRANK, R. H. & GLASS, A. J. 2006. *Microeconomics and behavior*, 6th edition, McGraw-Hill.
- GREENE, W. H. 1980. Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics*, 13, 27-56.
- GREENE, W. H. 2003. *Econometric analysis*, 5th edition, Prentice hall Upper Saddle River, NJ.
- HUANG, C. J. & LIU, J. T. 1994. Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis*, 5, 171-180.
- IVERSEN, T., BARANE, A-M., LILLEENG, B., KALSETH, J., NYLAND, K., SKRIBELAND, N.O., ØKLAND, S. 2010. *Årsrapport. Beregningsutvalget for spesialisthelsetjenesten 2010* [Online]. Available: <http://www.regjeringen.no>, <http://www.regjeringen.no/upload/HOD/SHA/BUS-rapport-2010.pdf> [Accessed May 9 2011].
- JACOBS, R. 2001. Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. *Health care management science*, 4, 103-115.
- JACOBS, R., SMITH, P. C. & STREET, A. 2006. *Measuring efficiency in health care: analytic techniques and health policy*, Cambridge Univ Pr.

- JOHNSON, N. L. & KOTZ, S. 1970. *Continuous univariate distributions*, New York, NY, USA., John Wiley & Sons.
- KITTELSEN, S., ANTHUN, K., KALSETH, B. & KALSETH, J. 2009a. En komparativ analyse av spesialisthelsetjenesten i Finland, Sverige, Danmark og Norge: Aktivitet, ressursbruk og produktivitet 2005-2007. *SINTEF-rapport A*, 12200.
- KITTELSEN, S. A. C. 1997. *Using data envelopment analysis to measure production efficiency in the public sector*, Oslo, Sosialøkonomisk Institutt.
- KITTELSEN, S. A. C., MAGNUSSEN, J. & ANTHUN, K. S. 2007. Sykehusproduktivitet etter statlig overtakelse: En nordisk komparativ analyse. *HERO On line Working Paper Series 2007:1*.
- KITTELSEN, S. A. C., MAGNUSSEN, J., SARHEIM ANTHUN, K., HÄKKINEN, U., LINNA, M., MEDIN, E., OLSEN, K. R. & REHNBERG, C. 2009b. Hospital productivity and the Norwegian ownership reform—A Nordic comparative study. *HERO On line Working Paper Series*.
- KODDE, D. A. & PALM, F. C. 1986. Wald Criteria for Jointly Testing Equality and Inequality Restrictions. *Econometrica*, 54, 1243-1248.
- KOOREMAN, P. 1994. Data envelopment analysis and parametric frontier estimation: complementary tools. *Journal of Health Economics*, 13, 345-346.
- LINNA, M., HÄKKINEN, U. & MAGNUSSEN, J. 2006. Comparing hospital cost efficiency between Norway and Finland. *Health Policy*, 77, 268-278.
- LINNA, M., HÄKKINEN, U., PELTOLA, M., MAGNUSSEN, J., ANTHUN, K. S., KITTELSEN, S., ROED, A., OLSEN, K., MEDIN, E. & REHNBERG, C. 2010. Measuring cost efficiency in the Nordic Hospitals—a cross-sectional comparison of public hospitals in 2002. *Health care management science*, 1-12.
- MAGNUSSEN, J., VRANGBAEK, K. & SALTMAN, R. B. 2009. *Nordic Health Care Systems: Recent Reforms and Current Policy Challenges*, Open University Press.
- MEDIN, E., ANTHUN, K. S., HÄKKINEN, U., KITTELSEN, S. A. C., LINNA, M., MAGNUSSEN, J., OLSEN, K. & REHNBERG, C. 2010. Cost efficiency of university hospitals in the Nordic countries: a cross-country analysis. *The European Journal of Health Economics*, 1-11.
- MEEUSEN, W. & VAN DEN BROECK, J. 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18, 435-444.
- RITTER, C. & SIMAR, L. 1997. Pitfalls of Normal-Gamma Stochastic Frontier Models. *Journal of Productivity Analysis*, 8, 167-182.
- ROEMER, M., I. 1989. National Health Systems as Market Interventions. *Journal of Public Health Policy*, 10, 62-77.
- ROSKO, M. D. & MUTTER, R. L. 2008. Stochastic frontier analysis of hospital inefficiency. *Medical Care Research and Review*, 65, 131.
- SALVANES, K. G. & TJØTTA, S. 1998. A Note on the Importance of Testing for Regularities for Estimated Flexible Functional Forms. *Journal of Productivity Analysis*, 9, 133-143.
- SCB.SE. 2010. *Befolkningsframskrivningar 2010-2060* [Online]. Statistics of Sweden. Available: <http://www.scb.se>, http://www.scb.se/Pages/PressRelease____291786.aspx [Accessed March 8 2011].
- SCHMIDT, P. 1976. On the Statistical Estimation of Parametric Frontier Production Functions. *The Review of Economics and Statistics*, 58, 238-239.
- SCHMIDT, S. S., FRIED, H. O. & LOVELL, C. A. K. 2008. *The measurement of productive efficiency and productivity growth*, New York, Oxford University Press.

- SSB.NO. 2010a. *Befolkningsframskrivninger. Nasjonale og regionale tall, 2010-2060*
[Online]. Statistics of Norway. Available: <http://www.ssb.no>,
<http://www.ssb.no/emner/02/03/folkfram> [Accessed March 8 2011].
- SSB.NO 2010b. Public finances. Statistics of Norway.
- STATA CORP 2009. Stata base reference manual 11. *College Station, TX: StataCorp LP*, 11.
- THOMAS, R. L. 2005. *Using statistics in economics*, McGraw-Hill Education.
- VUONG, Q. H. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57, 307-333.
- ZUCKERMAN, S., HADLEY, J. & IEZZONI, L. 1994. Measuring hospital efficiency with frontier cost functions. *Journal of Health Economics*, 13, 255-280.

