Delft University of Technology

Enhancing hospital productivity

van Hulst, Bart

DOI
10.4233/uuid:9582db34-2d6a-4c14-951e-5f35e92e56ab

Publication date
2016

Document Version
Final published version

Citation (APA)
van Hulst, B. (2016). *Enhancing hospital productivity*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:9582db34-2d6a-4c14-951e-5f35e92e56ab

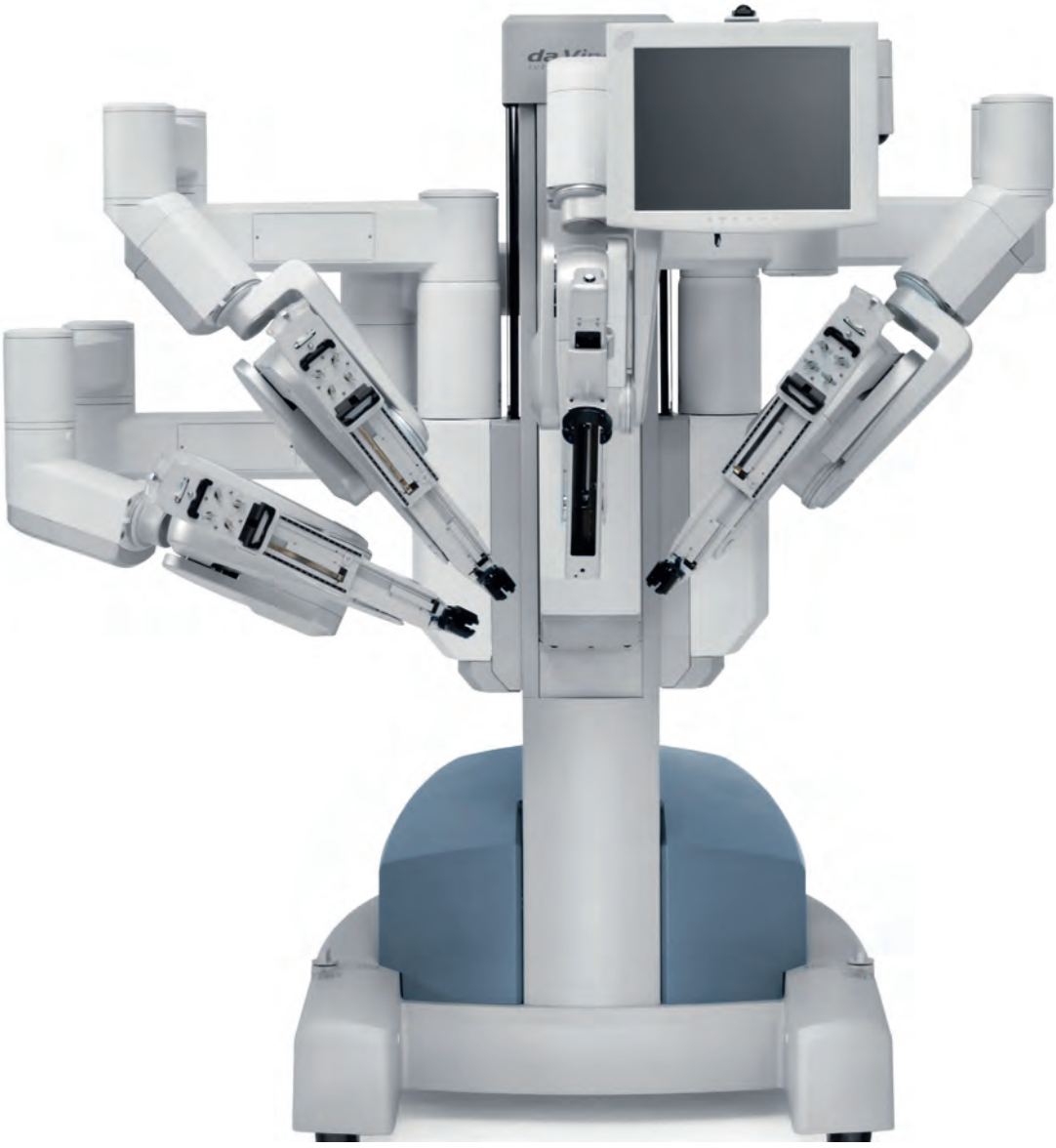
Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Enhancing hospital productivity



Bart van Hulst

Enhancing hospital productivity

Bart van Hulst

Enhancing hospital productivity

B.L. van Hulst

Phd Thesis, Delft University of Technology, Delft, The Netherlands.

ISBN: 978-94-6186-653-0

Copyright © 2016 B.L van Hulst, Delft

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission of the author.

IPSE Studies

TU Delft

Enhancing hospital productivity

Proefschrift
ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
woensdag, 25 mei, 2016 om 15:00 uur
door

Bart Laurens van HULST
Doctorandus econometrie, geboren te Hellevoetsluis, Nederland

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. J.L.T. Blank

Prof. dr. C.P. van Beers

Samenstelling promotiecommissie:

Rector magnificus, voorzitter

Prof. dr. J.L.T. Blank

Prof. dr. C.P. van Beers

Onafhankelijke leden:

Prof. dr. B.M. Balk	RSM, Erasmus University
Prof. mr. dr. J.A. de Bruijn	TU-Delft
Prof. dr. R.T.J.M. Janssen	Tilburg University
Prof. dr. M. Smet	KU Leuven
Prof. V.G. Valdmanis	Grand Valley State University
Prof. dr. M.J.G. van Eeten	TU-Delft

"To improve is to change; to be perfect is to change often."

-Winston Churchill

Acknowledgements

Acknowledgements are a non-consequential part of the thesis in that it is not evaluated. It does not make the acknowledgements less important because here I have the opportunity to show my gratitude to all that have contributed to this thesis.

First and foremost, I would like to thank Jos Blank for offering me the opportunity to do my PhD research under his supervision. I can easily write a full page of acknowledgements dedicated to his contributions. However, I will do it more efficient: without Jos this theses would not have been possible. Thank you for the pleasant cooperation!

I would also like to thank Cees van Beers, who in his own words "er met zijn haren bij is gesleept". I rather would say that an additional supervisor is profiting from economies of scale. Although you were not directly involved from the start, I owe much gratitude for the useful comments and suggestions in the last phase of writing the thesis. It has cost me part of my holiday, but I think it was worth the effort.

Some people have contributed to the thesis or parts of it by commenting on draft versions. Vivian Valdmanis has commented on almost all parts of this thesis two times. First as regular informal reviewer, since most of the individual chapters of this thesis correspond with papers. And a second time as a committee member. Dear Vivian, thank you for all your comments. Many thanks also to Bert Balk, for commenting two times on earlier versions of the thesis, especially for your critical examination of formulas. And thanks to Mike Smet for his comments on chapter 2 of this thesis.

There are a couple of organizations which I would like to thank for their direct or indirect contribution to this thesis. The TU-delft funded the initial

phase of the thesis. Later on the Ministry of the Interior and Kingdom Relations subsidized a research program on governance, innovations and productivity in the public sector. As the programme also include research on Dutch hospitals, the programme enabled me to put economies of scope into practice. Furthermore I would like to thank the Dutch hospital association (NVZ) and the Council for Health and Society (the former Council for Public Health and Health Care) for providing essential data used in this thesis.

Although there might not always be a direct connection to the thesis, I want to thank all my colleagues and former colleagues at IPSE-studies. Thanks for cooperating in the research projects, lunch walks, conferences and all other activities we did together. Some of the (former) colleagues contributed directly to thesis. Patrick Koot has to be thanked for his initial lobby for my PhD position, his function as an informal vade mecum of productivity research and the pleasant conversations traveling from Rotterdam to Delft and back. Alex van Heezik contributed as a sounding board for all topics that did not involve econometrics, advising on lay-out and readability. Also special thanks to Thomas Niaounakis for all his ICT support, critical discussion and functioning as a substitute for Patrick as travel companion. I want thank Anne van der Kloet for thoroughly proof reading of the complete thesis at its final stage. Of course Rachel Keivit deserves thanks for all her practical support when things have to be organized and for being such a good neighbor. I thank Janneke Wilschut for her mental support. Furthermore I also want to thank Leslie Zachariah of the TPM Graduate School for helping me out in taking an unexpected bureaucratic hurdle in the final stage of the PhD project.

Traditionally, the final paragraph of the acknowledgement is dedicated to the author's family and friends. Their contribution is notoriously difficult to measure, in productivity analysis we would probably refer to this as an environmental or quality. I thank all family and friends for your support as well for being family or a friend. Special thanks of course for my lovely wife and fantastic children, especially the twin for providing the empirical evidence for

proposition number nine. Dear Thohngdie, Pepijn and Saskia finally we can go to the "ballenbak".

Bart van Hulst
Delft
April 2016

Contents

List of Tables

List of Figures

Summary

Introduction

Healthcare spending in Western countries is substantial. Not only does this expenditure form a major part of the economy, it also outpaces economic growth. Furthermore, there are no indicators that increasing healthcare spending will slow down. For most Western countries, health expenditure is expected to continue to increase due to aging populations and broader possibilities for medical treatment. Both the extent of spending and the predicted growth make cost containment in healthcare one of the most important policy issues in these countries. At the same time, it is undesirable to sacrifice the accessibility and quality of healthcare through interventions. However, without intervention, the sustainability of healthcare systems may be affected.

One feasible intervention is to increase productivity in healthcare, or in other words: do more with less. Changes in productivity can be decomposed in three factors: scale, efficiency and technical change. An understanding of how each of these factors increases productivity can be used for cost containment. This thesis examines these three factors, using the Dutch hospital industry as a case study.

Optimal scale

First we address the optimum scale of hospitals. For convenience, the number of beds is used as indicator for the scale of a hospital. Economies of scale, in terms of cost per unit of output, decreases as scale increases. As scale grows, diseconomies of scale can eventually prevail. This thesis identifies the point at which economies of scale turn into diseconomies of scale. For that purpose, we conducted a meta-analysis of 41 parametric studies that apply a flexible cost function for the cost structure of hospitals and include results on scale effects. These 41 studies generate 95 observations. Besides parametric

studies non-parametric models are popular tool to study the cost structure of hospitals. In addition to the parametric studies, 19 non-parametric studies are analysed. Since reported results for parametric studies and non-parametric studies differ, both type of studies are analysed separately.

In general, parametric studies do not report the optimal scale, but rather the scale elasticity at the sample mean. The scale elasticity measures the proportional change in output that follows from a proportional change of costs; a scale elasticity greater than one indicates economies of scale and a scale elasticity of less than one indicates diseconomies of scale. The meta-analysis relates the scale elasticity to study characteristics including the scale for which the scale elasticity applies. If non-parametric studies include results on the optimum size, the optimum scale is included directly. However there are only few studies that report the optimum scale.

The results of the analysis indicate that the optimum scale lies around 320 beds, in case of a parametric reference study. There are a couple of factors that have an impact on the optimum scale. In case of a frontier study the optimum is about 240 beds. This compares well with a lower bound found with the non-parametric studies of 220 beds.

This thesis also includes specific results regarding the scale of Dutch hospitals. The average scale efficiency for Dutch hospitals in 2007 is 87.5%, meaning that there is a theoretical efficiency gain possible of 12.5%. The average scale efficiency results from a combined effect of over- and undersized hospitals; however, the vast majority (80%) of the Dutch hospitals is oversized. This was not always the case. In 2002, the average hospital operated at around the optimal scale, but hospital size has increased and the average Dutch hospital now operates at diseconomies of scale.

Efficiency

Next we address the efficiency of hospitals and its governance. Efficiency is a measure of a hospital's productivity compared to a 'best practice hospital'.

Hospitals that perform on par with the best practice hospital have a maximum efficiency score of 100%. Efficiency scores less than 100% indicate conversely how much productivity could improve by raising performance to the best practice level. There are several methods for estimating hospital efficiency, this study applies a non-parametric technique known as data envelopment analysis (DEA).

The average efficiency of Dutch hospitals in 2007 is 78% inclusive of scale inefficiency and 89% after accounting for scale effects. Since efficiency is a relative measure, this indicates that productivity differences between Dutch hospitals are fairly small. Furthermore, these results are in line with the findings of international studies on hospital efficiency.

More relevant than efficiency itself are the reasons why efficiency differs, i.e. which characteristics increase hospital efficiency. In this thesis, we investigate how differences in efficiency relate to governance characteristics of hospitals. It appears that higher remuneration for the supervisory board correlates with lower efficiency. Furthermore, increasing the board's remuneration does not affect efficiency. In general, other governance characteristics appear to correlate with the size of the hospital and it is therefore difficult to make a statement as to how these characteristics affect efficiency.

Technical change

Last but not least we have technical change. Technical change is a collective noun for productivity changes resulting from the overall process of invention, innovation and diffusion of technology. In productivity analyses, technical change is often measured by means of a proxy, namely a time trend. All changes in productivity through time, other than changes due to scale or efficiency, are ascribed to technical change. As a result, we only know how much productivity changes through time, but not what caused the change. Moreover, each hospital adopts technology at its own pace, which means that

hospitals may be operating with different technologies at the same point in time.

In this thesis, we use technology indices for each hospital to gain more insight into the effect of innovations. We explicitly inventory specific and well-known innovations in the Dutch hospital industry. These innovations are aggregated into seven homogenous technology indices, which are measured by means of a set of technology index numbers. The index numbers are included in a cost function specification and estimation.

It appears that some, but not all, innovations increased productivity. In particular, innovations in the field of ICT and chain care have positively contributed to productivity. Productivity loss, on the other hand, is associated with innovations aimed at improving quality (because quality is not measured as output). However, the results are rather ambiguous since the effects of innovations vary across different outputs (i.e. discharges and outpatients).

The results here are derived from a static model that measures effects from one time period to another. Although the technology indices accumulate over time, there is still friction with the inter-temporal effects of adopting innovations. Decisions on the adoption of innovations are inter-temporal, it requires a trade-off between short-term adjustment costs and long-term (future) savings. This has modelling consequences. This thesis shows how an additional equation can be added to cost models to provide insight into the optimal amount of innovation to adopt and makes estimates more reliable.

We do not only investigate the role of innovations but also calculate the productivity changes that result from technical change. Three chapters are dedicated to calculating productivity changes over slightly different time periods. If we combine the results of these three chapters, we find that over the period of 1995–2011, productivity increased at an average of 2% per year due to technical change.

Technical change not only influences productivity, but it may also affect the optimal mix of inputs and outputs. We refer to the first case as input-biased technical change and the second case as output-biased technical change. A combination of both types of technical change is also possible. If the optimal mix of both input and output is unaltered, technical change is neutral. The results indicate that technical change in Dutch hospitals is not neutral. However, the results are not consistent through time. If we combine the results, we see that technical change first is output biased, it then is both input and output biased and finally, towards the end of the period, it is input biased.

Input-biased technical change indicates that the optimal input mix changes, and therefore some inputs are substituted for others. This thesis shows how factor productivity can be calculated, adjusting for substitution effects. It appears that, taking these effects into account, the factor productivity for labour outpaced total productivity. Finally, productivity associated with materials was lower than other inputs.

Policy recommendations

Before addressing policy recommendations, it should be noted that productivity research has some limitations. The first issue is the measurement of output. Here, output is measured as the number of patients treated, which is merely an indicator for the desired outcome: improved health. The effects of treatments that improved health outcomes or that were themselves better quality treatments are not included in the results. A second issue relates to the observation that innovations made only a minor contribution to productivity growth in recent years. It is not unlikely that this observation is partly a result of the particular innovations that we analysed, which are mainly characterised as medical procedures and treatment methods. It is also likely that process innovations (which are underrepresented in this study) made a substantial contribution to productivity growth. Furthermore, as fair warning to policymakers, it should be noted that productivity improvements often have side effects. The substantial increase in productivity in recent years has

coincided with an increase in output. Increased productivity and cost containment are therefore not synonymous. Additionally, it should be noted that hospital productivity is an isolated effect; for example, productivity could be increased by discharging patients early to nursing homes or homecare, thus increasing costs in other sectors.

A first recommendation for policymakers concerns the optimal scale for hospitals. Economies of scale are only found in small hospitals, but are quickly exhausted as hospitals grow and eventually diseconomies of scale prevail. It seems that beyond 320 beds diseconomies of scale will prevail. It is however likely that the optimum scale for a hospital is even smaller than 320 beds. For studies that identify efficient hospitals, the optimum scale lies around 220-240 beds. In the Dutch hospital industry, the scale of operation of all most all hospitals is already beyond the optimal scale. From an economic perspective it is unwise to increase the scale of Dutch hospitals.

The governance of a hospital appears to have only a limited impact on its efficiency. Interestingly, the remuneration of the board of directors has little impact on efficiency and higher remuneration at supervisory board level is associated with lower efficiency. From a policy perspective, this invalidates the claim that competitive remuneration alone attracts capable administrators. The average efficiency score of Dutch hospitals is in line with other (international) studies. Theoretically, there seems to be reasonable potential for improving efficiency; in practice, however, it is likely that improvement of only a few percentage points will be possible.

In recent years, technical change has raised productivity by 2% per year in Dutch hospitals. This is impressive, especially because it has been consistent. Therefore, from the three factors that increase productivity, technical change is the most promising option for enhancing future productivity in the Dutch hospital industry. For policy makers, it seems wise to stimulate productivity-enhancing technical change. This thesis shows that to some extent, innovations contribute to productivity. However, still for a major part it is unknown what

technical is. Therefore some effort has to be made to identify the technologies that really increase productivity.

Samenvatting

Inleiding

De kosten van gezondheidszorg hebben in de westerse landen een aanzienlijke omvang. Niet alleen beslaan de zorgkosten een groot deel van de economie, de zorgkosten groeien ook nog eens sneller dan de economie. Een einde aan de toenemende zorgkosten lijkt voorlopig niet in zicht. Voor de meeste westerse landen is de verwachting dat de kosten van de gezondheidszorg verder toenemen als gevolg van onder andere vergrijzing en de toenemende medische behandelmogelijkheden. Als gevolg hiervan is kostenbeheersing in de gezondheidszorg een beleidsopgave met hoge prioriteit in de meeste de westerse landen. Zonder interventies komt de houdbaarheid van de gezondheidszorg mogelijk in het geding. Tegelijkertijd is het onwenselijk dat door interventies wordt ingeboet op toegankelijkheid en kwaliteit van de gezondheidszorg.

Een alternatief om de toenemende zorgkosten in toom te houden, is het realiseren van productiviteitsgroei in de gezondheidszorg. Meer doen met dezelfde of minder middelen. Productiviteit kan worden ontbonden in drie factoren: schaal, efficiëntie en technologische ontwikkeling. Inzicht in deze factoren kan helpen bij kostenbeheersing in de zorg. Dit proefschrift bestudeert deze drie onderwerpen en gebruikt daarvoor de Nederlandse ziekenhuissector als casus.

Optimale schaal

Allereerst gaan we in op de optimale schaal van ziekenhuizen. Schaalvoordelen bestaan als de kosten per eenheid productie dalen bij een toename van de schaal. Als de schaal steeds verder toeneemt, kan er eventueel een punt zijn waar schaalnadelen de overhand krijgen. Dit proefschrift gaat na bij welke omvang de schaalvoordelen omslaan in schaalnadelen. Daartoe is een meta-analyse gemaakt van 41 parametrisch studies, die een kostenfunctie

gebruiken en resultaten hebben ten aanzien van de schaal van het ziekenhuis. Daarnaast zijn 18 non-parametrische studies geanalyseerd.

Over het algemeen rapporteren parametrische studies niet de optimale schaal, in plaats daarvan wordt vaak de schaalelasticiteit voor het gemiddelde van de onderzoekspopulatie gerapporteerd. De schaalelasticiteit geeft aan hoe groot de proportionele verandering van de productie is bij een proportionele verandering van de kosten. Een schaalelasticiteit groter dan één betekent schaalvoordelen, een schaalelasticiteit kleiner dan één betekent schaalnadelen. In de meta-analyse is de schaalelasticiteit in verband gebracht met de schaal van het ziekenhuis en de kenmerken van de studie waarmee de schaalelasticiteit bepaald is. Daarbij is het aantal bedden gebruikt als indicator van de schaal van een ziekenhuis. Bij non-parametrisch studies wordt de optimale schaal direct gerapporteerd, helaas is maar een beperkt aantal studies waarin dit gebeurt.

De resultaten van de analyse laten zien dat de optimale schaal 320 bedden is voor een parametrische referentiestudie. Er is een aantal factoren die van invloed zijn op de optimale schaal. Zo is de optimale schaal voor een frontier studie 240 bedden. De omvang komt overeen met een ondergrens van 220 bedden als optimale omvang voor non-parametrische studies.

Het proefschrift bevat eveneens specifieke resultaten over de schaal van Nederlandse ziekenhuizen. De gemiddelde schaalefficiëntie van Nederlandse ziekenhuizen in 2007 is 87,5% , dit betekent dat 12.5% efficiëntieverbetering mogelijk is. De gemiddelde schaalefficiëntie is een gecombineerd effect van ziekenhuizen die te klein en te groot zijn. Het overgrote merendeel van de ziekenhuizen is overigens te groot; in 2007 is ongeveer 80% van de Nederlandse ziekenhuis te groot. Dit is niet altijd zo geweest, in 2002 opereert het gemiddelde Nederlandse ziekenhuis rond de optimale schaal, daarna is de schaal van Nederlandse ziekenhuizen te ver doorgeschoten en opereert het gemiddelde ziekenhuis onder schaalnadelen.

Efficiëntie

Efficiëntie is een maatstaf die aangeeft hoe goed een ziekenhuis presteert ten opzichte van de beste praktijk. Ziekenhuizen die hetzelfde presteren als de beste praktijk hebben een efficiëntie van 100%. Een efficiëntie van minder dan 100% geeft aan hoeveel verbeterpotentieel aanwezig is door even productief te worden als de beste praktijk. Er zijn verschillende methoden om de efficiëntie te schatten, deze studie gebruikt data envelopment analysis (DEA).

De gemiddelde efficiëntie van Nederlandse ziekenhuizen in 2007 is 78% inclusief het effect van schaalnadelen en 89% na correctie voor schaalnadelen. Efficiëntie is een relatieve maat en laat zien dat de verschillen in efficiëntie tussen ziekenhuizen in de Nederlandse ziekenhuissector klein zijn. De resultaten zijn overigens in lijn met de internationale literatuur over de efficiëntie van ziekenhuizen.

Interessant zijn de factoren die efficiëntieverschillen tussen ziekenhuizen verklaren; door aanpassing van welke factoren kan de efficiëntie worden verhoogd? In dit proefschrift zijn efficiëntieverschillen verklaard met governance kenmerken van ziekenhuizen. Het blijkt dat een hogere beloning voor de raad van toezicht gepaard gaat met een lagere efficiëntie. Een hoge beloning van het bestuur heeft geen effect op de efficiëntie. Een aantal andere governance kenmerken blijkt sterk te correleren met de omvang van het ziekenhuis, waardoor voor deze kenmerken geen eenduidige uitspraak over effecten op de efficiëntie gedaan kan worden.

Technologische ontwikkeling

Technologische ontwikkeling is een verzamelnaam voor verandering van de productiviteit door uitvindingen, innovatie en diffusie van technologie. In productiviteitsanalyses wordt het effect van technologische verandering gemeten met de productiviteitsverandering tussen twee perioden. Alle productiviteitsverandering door de tijd heen, uitgezonderd veranderingen van schaal en efficiëntie, is het effect van technologische ontwikkeling. Als gevolg weten we alleen hoeveel het effect van technologische ontwikkeling is geweest,

niet wat en in welke mate heeft bijgedragen aan de technologische ontwikkeling. Daarnaast wordt verondersteld dat de technologische ontwikkeling voor ieder ziekenhuis hetzelfde is, terwijl in de praktijk ieder ziekenhuis technologie in een eigen tempo adopteert.

Dit proefschrift maakt gebruik van individuele technologie indices voor ieder ziekenhuis om meer inzicht te krijgen in het effect van innovaties. Voor een lijst van bekende innovaties is nagegaan op welk moment individuele ziekenhuizen deze innovaties adopteren. De verschillende innovaties zijn ingedeeld in zeven homogene clusters, voor ieder ziekenhuis is per cluster een technologie index geconstrueerd. De technologie indices zijn vervolgens toegevoegd aan een kostenmodel van ziekenhuizen. Uit de schattingsresultaten van het kostenmodel kan het effect van innovaties op de productiviteit worden afgeleid.

Het blijkt dat een deel van de innovaties hebben bijgedragen aan de productiviteit. Vooral innovaties op het vlak van ICT en ketenzorg hebben een positieve bijdrage geleverd aan de productiviteitsontwikkeling van ziekenhuizen. Productiviteitsverlies wordt geassocieerd met innovaties gericht op verbetering van de kwaliteit (omdat kwaliteit niet wordt gemeten als productie). Effecten zijn overigens niet altijd eenduidig, een innovatie kan de kosten van de ene output verlagen en van een andere output verhogen.

De voorgaande analyse gebruikt een statisch model, waarbij wordt gekeken naar het effect tussen twee perioden. Ondanks dat technologie-indices zijn gebruikt, die cumuleren, is er een frictie met de inter-temporele effecten van de adoptie van een innovatie. De adoptie van een innovatie is een afweging tussen aanpassingskosten op de korte termijn en toekomstige besparingen die voor een meerdere perioden gelden. Dit impliceert dat er een inter-temporele beslissing wordt genomen. Dit heeft analytische consequenties voor de modellering van de kostenstructuur van ziekenhuizen en de rol van innovaties. Dit proefschrift laat zien hoe een additionele vergelijking aan een kostenmodel

toegevoegd kan worden. Dit geeft inzicht in de optimale omvang van het aantal te adopteren innovaties en resulteert in betrouwbardere schattingen.

Niet alleen de rol van innovaties is onderzocht, ook het effect van technologische ontwikkeling is berekend. Drie hoofdstukken van dit proefschrift bevatten resultaten over het effect van de technologische ontwikkeling voor verschillende perioden. Als de resultaten worden gecombineerd, vinden we een productiviteitsontwikkeling van 2% per jaar over de periode 1995-2011 als gevolg van de technologische ontwikkeling.

De technologische ontwikkeling kan ook de optimale mix van inputs en de optimale mix van outputs veranderen. De technologische ontwikkeling is dan niet neutraal, in het eerste geval is sprake van input-biased, in het tweede geval output-biased. Een combinatie van beide is ook mogelijk. De resultaten van dit proefschrift laten zien dat de technologische ontwikkeling voor de Nederlandse ziekenhuizen niet neutraal is geweest. Aanvankelijk is er sprake van output-biased, daarna van zowel input-biased als output-biased, en tenslotte van input-biased.

Veranderingen in de optimale input-mix impliceert substitutie. Bij het berekenen van de factor productiviteit van inputs kan rekening worden gehouden met substitutie-effecten. Het blijkt dat, rekening houdend met substitutie-effecten, de factorproductiviteit van arbeid sneller is gegroeid dan de totale productiviteit. De productiviteit van materiaal is lager dan die van andere inputs.

Beleidsaanbevelingen

Voor het op waarde schatten van de resultaten en beleidsmatige toepassingen van dit proefschrift moet worden bedacht, dat productiviteitsonderzoek een aantal beperkingen kent. Allereerst is de productie geoperationaliseerd met het aantal (voor case-mix gecorrigeerde) behandelde patiënten. Dat is slechts een indicator voor datgene waar het werkelijk om gaat: verbetering van de gezondheid. Het effect van een

behandelmethode met verbeterde gezondheidsuitkomsten of betere kwaliteit blijft buiten het zicht van de productiviteitscijfers. Een tweede kanttekening is een nuancering van de constatering dat slechts een beperkt deel van de productiviteitsstijging van de afgelopen jaren is toe te schrijven aan innovaties. Dit heeft waarschijnlijk mede te maken met de geanalyseerde innovaties die vooral het karakter hebben van medische procedures en de behandeling van patiënten. Het is niet onwaarschijnlijk dat juist procesinnovaties (ondervertegenwoordigd in deze studie) een substantiële bijdrage hebben geleverd aan de productiviteitsverbetering. Een laatste kanttekening is een waarschuwing aan beleidsmakers: productiviteitsverbetering is een middel dat bijwerkingen kan hebben. De forse productiviteitsverbetering van de afgelopen jaren is gepaard gegaan met een forse groei van de productie, derhalve zijn de kosten niet afgenomen, maar alleen maar minder snel gegroeid. Verder is hier de productiviteit van ziekenhuizen geïsoleerd onderzocht; het is goed mogelijk dat de productiviteitsstijging van ziekhuizen gerealiseerd is door patiënten eerder over te dragen aan verpleeghuis of thuiszorg en is er dus sprake van het verplaatsen van kosten.

De eerste aanbeveling betreft de optimale schaal van ziekenhuizen. Schaalvoordelen zijn vooral in kleine ziekenhuizen te behalen, maar zijn snel uitgeput en bij een bepaalde omvang krijgen schaalnadelen de overhand. Profiteren van de optimale schaal kan ook door juist op een kleinere schaal te opereren. De optimale schaal van ziekenhuizen ligt rond de 300 bedden. Specifiek voor Nederlandse ziekenhuizen lijkt de schaalvergroting te ver doorgeschoten, het gemiddelde ziekenhuis in Nederland is een stuk groter dan 300 bedden. Verdere schaalvergroting in de Nederlandse ziekenhuissector ligt niet voor de hand.

De tweede aanbeveling betreft de efficiëntie van ziekenhuizen en in het bijzonder de rol van de governance. Theoretisch lijkt er bij een efficiëntie van 89% een redelijk verbeterpotentieel te bestaan, in de praktijk is slechts een paar procent haalbaar. De gemiddelde efficiëntie in de Nederlandse ziekenhuizen is

overigens in lijn met de bevindingen van internationale studies naar de efficiëntie van ziekhuizen. Ten aanzien van de governance kenmerken is het interessant dat de beloning van de ziekenhuisbestuurders er weinig toe doet en een hogere beloning van de toezichthouder gepaard gaat met een lager efficiëntie. Vanuit beleidsperspectief ontkracht dit de stelling dat alleen met een beloning die concurrerend is met de private sector capabele bestuurders gevonden kunnen worden.

De afgelopen jaren is de productiviteit van ziekenhuizen met gemiddeld 2% per jaar toegenomen door technologische veranderingen. Dat is een behoorlijk prestatie, zeker omdat het over een langere periode gaat. Voor toekomstige productiviteitsverbeteringen mag het meest verwacht worden van technologische verandering. Voor beleid is de meest kansrijke optie het stimuleren van technologische veranderingen om de productiviteit verder te verhogen. Dit proefschrift laat zien dat voor een deel productiviteitsgroei is toe te schrijven aan innovaties. Voor een groot deel blijft echter onbekend welke veranderingen precies hebben bijgedragen aan de verbeterde productiviteit. Daarom moet een inspanning geleverd worden om de technologieën te identificeren die de productiviteit echt verhogen.

General introduction

1.1 Policy background

Healthcare expenditures in Western countries are substantial and are increasing rapidly. In 2012, healthcare spending in the EU Member States averaged 10.1% of GDP (OECD/European Union, 2014)[1]. Health expenditures in the Americas are even higher, at an average of 14.1% of GDP in 2011. Within the Americas, the United States tops the league, spending well above average at 17.7% of GDP in 2011 (World Health Organization, 2014). Furthermore, healthcare expenditures are outpacing economic growth. Between 2000 and 2012, healthcare spending as a percentage of GDP increased by 1.6 percentage points in the EU countries. In the Americas things went even faster, with healthcare spending increasing by 2.6 percentage points between 2000 and 2011. Again, things are bigger in the United States: healthcare spending there increased by 4.1 percentage points over the past decade.

Still there is no end, it is expected that healthcare spending will continue to increase in most Western countries. This is a result of an ageing population and increasing possibilities for (more costly) medical treatment. Long-term forecasts for Western countries predict that healthcare expenditures as a percentage of GDP will have increased by 3.5 to 6 percentage points in 2050 (OECD, 2010).

The high level of healthcare expenditures and their predicted increase make cost containment one of the most pressing policy challenges for Western countries. Without intervention, the sustainability of healthcare comes under pressure. At the same time, it is undesirable that interventions affect the accessibility or quality of care.

---

[1] Weighted average; the unweighted averages in 2000 and 2012 were, respectively, 7.3% and 8.7% of GDP.

On top of this, the health industry is a labour-intensive sector with skilled personnel, like physicians and nurses, the supply of whom is inelastic in the short term. Due to an ageing labour force and fierce competition in the labour market with other economic sectors, there are also concerns about a looming crisis in the health workforce (OECD, 2008). These concerns have become less urgent since the global financial crisis of 2008, but are still slumbering in both European (European Commission, 2012) and non-European countries (Health Workforce Australia, 2012a). So not only is cost containment a challenge, but a sufficient supply of qualified health personnel also needs policy attention.

Increasing productivity might be a solution for cost containment and at the same time avert the looming crisis in the health workforce. There are several ways to increase productivity. One way that might achieve increased productivity in healthcare is the use of cost-saving technologies. Not without reason, one of the EU's strategic objectives supports dynamic health systems and new technologies (European Commission, 2007). According to the European Commission, new technologies can contribute to the efficiency and sustainability of healthcare and at the same time improve access to safe and high-quality care. Other options that can be used to increase productivity are producing at an optimal scale and increasing efficiency.

1.2 Scope of the thesis

This thesis focuses on the possibilities of increased productivity in the hospital sector. The choice for the hospital sector is motivated by the large proportion of the healthcare budget allocated to the hospital sector. Rumbold et al. (2014) state, in a review study on potential efficiency improvement, that, "The acute hospital sector is one of the largest areas of expenditure within a health system (Jones & Charlesworth, 2013; Orosz & Morgan, 2004), and hence it is likely to be an important area of interest for policy makers seeking to make gains in productive efficiency." Ludwig (2008) also uses the financial

size of the hospital sector as a rationale for research into that sector. In general, the hospital sector can count on the warm interest of productivity researchers. In a survey of efficiency and productivity studies in healthcare, 52% of the studies included in the survey examined the hospital sector (Hollingsworth, 2008).

1.3 Productivity

Input and output

Productivity is the ratio of outputs to inputs. In the case of one output and one input, calculating this ratio is quite straightforward. However, an important issue is how to define the output and input. Defining input is the easier of the two, since it should be clear what is required to produce the output. Defining output is harder, because this can be done in various ways. In the case of hospitals, there are various possibilities for the measurement of output: number of surgeries, admissions, bed days, treated patients or perhaps an improvement in health. The results of productivity analysis depend on the definition of input and output used. So the first question in productivity analysis is how input and output are measured.

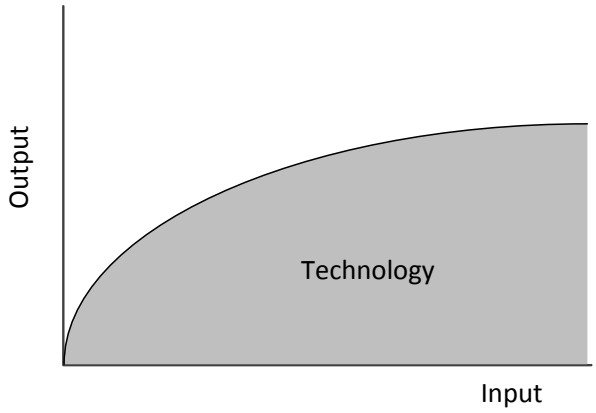
Things become more complicated if there are multiple inputs and multiple outputs. In that case, calculating productivity means that inputs as well as outputs somehow have to be aggregated. Aggregation is more than simply adding up the various inputs or outputs. For example, an admission uses more resources then an outpatient visit, so in productivity analysis it would be incorrect simply to add up admissions and outpatients. Furthermore, in this example admissions are already an aggregate of heterogeneous treatments. When aggregating, therefore, we want to weigh the outputs. A similar reasoning applies to inputs, although aggregating them is usually easier since we can use their costs. At any rate, having determined what the various inputs and outputs are, the next question is how to aggregate them.

Defining output and input, and aggregating them, illustrate that productivity analysis comes with some choices. This implies that the results of productivity analysis have a certain degree of subjectivity. Absolute measurement of productivity therefore has little value. The added value of productivity analysis lies in the comparison of the productivity. Why is one hospital more productive than another? What makes productivity increase over time? In principle, differences in productivity are explained by the following factors: differences in production technology, differences in scale, differences in efficiency and differences in environmental characteristics (Fried et al., 2008).

Production technology

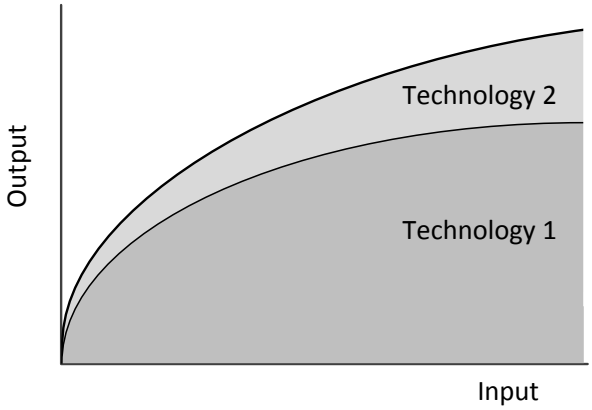
In productivity analyses, a production technology is represented by a system of relations between inputs and outputs. These relations can be represented in different ways. A simple representation is that of a collection of combinations of inputs and outputs, where the combinations are such that, with the inputs, the production of outputs is feasible. Figure 1-1 shows a simple example for one input that produces one output. The shaded area represents all feasible input-output combinations for a technology.

Figure 1-1 Example of a technology with feasible input-output combinations



Differences in productivity between hospitals might be the result of a difference in production technology. If we compare two production technologies, for example, one might have a feasible input-output combination that is not feasible for the other technology. Figure 1-2 extends the example and visualizes the differences in productivity due to differences in technology. A hospital that uses technology 2 can be more productive than a hospital that uses technology 1. This is because, given the inputs under technology 2, the production of more outputs is possible. Or, conversely, the same amount of output can be produced with less input.

Figure 1-2 Example of technologies with different feasible input-output combinations



Productivity analysis often makes the implicit assumption that, at a certain point in time, each hospital has the same production technology at its disposal. Differences in technology occur through changes to the production technology over time. The resulting changes in productivity can be calculated and are the subject of research. Productivity changes over time are referred to as technical change.

Efficiency

Koopmans (1951) gives the following definition of technical efficiency: a firm is technically efficient if it is not possible to increase the production of an output without reducing the production of at least one of the other outputs or

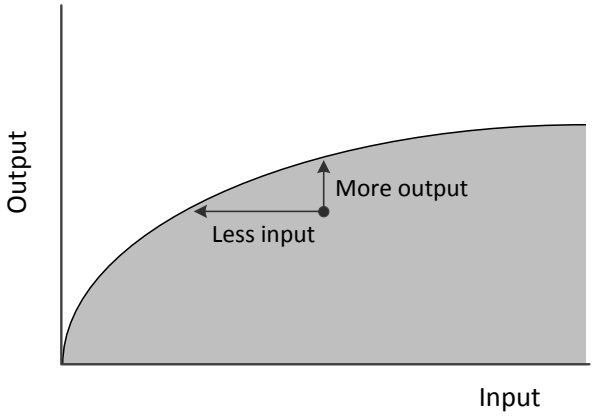
increasing the use of at least one input. Or, reasoning from the input side: a firm is technically efficient if it is not possible to reduce the use of an input without increasing at least one other input or reducing the production of at least one of the outputs. A more practical definition of technical efficiency is that given by Debreu-Farrell (Debreu, 1951; Farrell, 1957): technical efficiency equals the proportional increase of all outputs with no additional inputs required. Or, from the input side: technical efficiency is one minus the proportional reduction in all inputs without reducing output. One of the advantages of the Debreu-Farrell definition is that it gives a direct measure for the efficiency.

Efficiency is a concept that relates the actual productivity of a hospital to its highest possible productivity. A production technology provides insight into all feasible combinations of inputs and outputs, but not every combination has the same productivity. Given the production technology, there is a set of combinations of inputs and outputs, ceteris paribus, that has the highest productivity. The combinations of inputs and outputs with the highest productivity are the so-called efficient combinations. Figure 1-3 shows an example of an inefficient hospital. Given the technology, the input-output combination used by this hospital is feasible. However, it could produce the same amount of output with less input or it could produce more output with the same amount of input. For this reason, the hospital is inefficient. The efficient combinations in Figure 1-3 are on the boundary of the technology.

So a feasible combination is not necessary an efficient one. Now we want to know how inefficient a hospital is. This is done by comparing its productivity with that of a corresponding input-output combination which is efficient. We want to compare the productivity of the hospital with an efficient equivalent, also known as best practice. In determining the efficiency of the hospital, then, it is also necessary to determine the corresponding efficient combination. Once we know that, we can apply the Debreu-Farrell definition

(proportional increase of outputs or proportional decrease of inputs) to find a measure for the efficiency of the hospital.

Figure 1-3 Example of inefficiency where either less input or more outputs is feasible



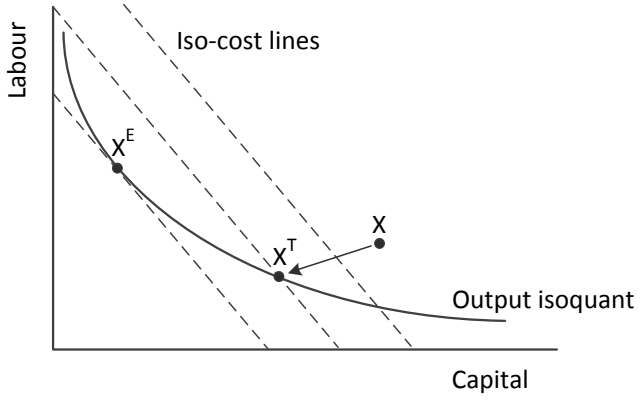
There are two types of efficiency: technical and economic. Most production technologies have several combinations of inputs and outputs that are technically efficient. For example, a given level of output can be attained through either a capital-intensive method or a labour-intensive one, both of which are technically efficient. Although there are several combinations of inputs and outputs that are technically efficient, not all are equally preferable. That has to do with the economic objective – the economic behaviour – of the hospital in question. For example, a hospital might strive for cost minimization. In that case, within the set of technically efficient combinations, the one with the lowest cost is preferred. Besides cost minimization, other possible economic objectives are production maximization, input minimization, revenue maximization, cost minimization and profit maximization (Blank & Valdmanis, 2013).

Economic efficiency imposes an additional restriction on the efficient input-output combination: the allocation of the inputs and/or the outputs has to be optimal. What is optimal here depends on the economic objective of the hospital. Economic efficiency indicates the extent to which the economic

objectives are realized. In case of cost minimization, for example, economic efficiency is the ratio of the minimum feasible costs to actual costs. Note that this is different from a proportional decrease of inputs.

Figure 1-4 sheds some more light on technical and economic efficiency with an example. The figure maps two inputs, labour and capital, which are needed to produce output. The output isoquant is a curve of the minimum amounts of labour and capital required to produce a certain amount of output. Each iso-cost line represents combinations of labour and capital that have the same costs, with the line closest to the origin having the lowest costs. Note that the slope of the iso-cost lines is determined by price ratio of capital and labour. Now suppose that hospital X produces the same amount as hospitals at the output isoquant. Hospital X is inefficient since it can reduce inputs and still produce the same amount of output. To become technically efficient, hospital X reduces labour and capital proportionally, in line with the definition from Debreu-Farrell. By doing so, the hospital ends up at X^T, where it technically efficient. Now assume that the economic objective of the hospital is cost minimization; this means that at X^T the hospital is still not economically efficient since it can reduce costs by substituting capital for labour (i.e. move along the output isoquant). To become economically efficient, the hospital has to move to X^E, where the iso-cost line is tangential to the output isoquant.

Figure 1-4 Example of technical efficiency and economic efficiency



The foregoing shows that efficiency is a measure of how much better a hospital can perform, given the production technology. Now that we know how efficient we are, the follow-up question is: what needs to be changed to become more efficient? What needs to be changed to attain the productivity level of the best practice? Answers to these questions are the subject of numerous productivity and efficiency studies. In general, these studies identify factors that explain a high efficiency.

Scale

An increase or decrease in production does not necessarily imply a proportional change to inputs. The consequence of this is that productivity changes. Productivity is the ratio of production to inputs and, since the size of the numerator and denominator vary independently, the ratio also changes. If output increases faster than inputs, there are economies of scale; if the opposite applies, there are diseconomies of scale. Figure 1-5 illustrates scale effects with an example. In this case input doubles, while output increases by only 30%, meaning that there are diseconomies of scale.

Figure 1-5 Example of diseconomies of scale

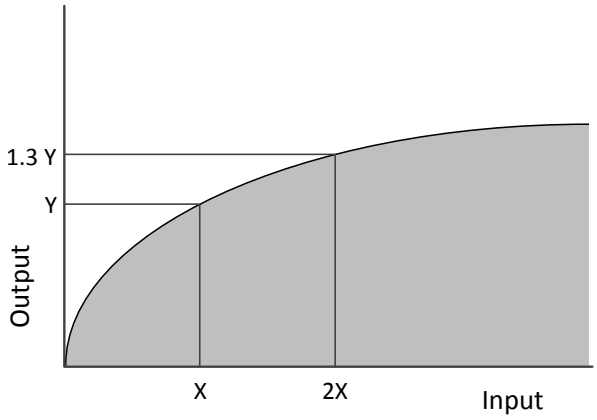
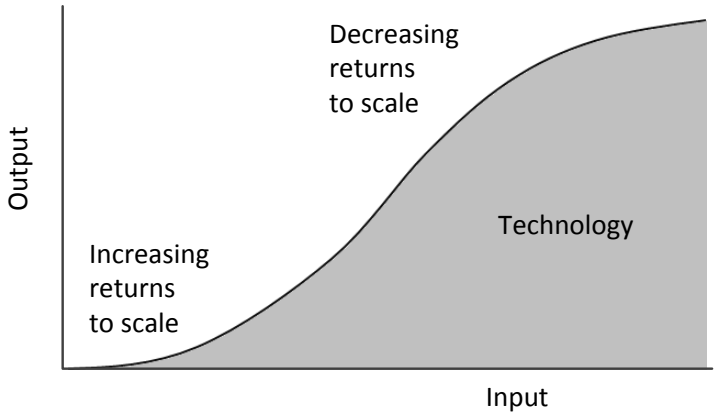


Figure 1-5 exhibits diseconomies of scale at all levels. But it is not necessarily the case that a technology exhibits only economies of scale or only diseconomies of scale. Often, economies of scale also depend on the scale

itself. Figure 1-6 shows an example of a technology with both economies of scale and diseconomies of scale. At low levels of input, additional input leads to a more than proportional increase in output. As input increases, an optimum point is eventually reached, after which diseconomies of scale prevail. From that point an increase in input results in a less than proportional increase in outputs. The intuition behind technologies with both increasing and decreasing returns is that small hospitals have increasing economies of scale since their fixed costs spread over a larger amount of outputs as production increases. Other sources for economies of scale are for example better opportunities for the division of labour (making specialization possible) and purchasing power. However, effects decrease as a hospital becomes bigger, there are still economies of scale, but the effect is smaller. At the same time, as a hospital becomes bigger, another effect occurs: diseconomies of scale. Diseconomies occur for example, because of increased bureaucracy, a lack of communication and less commitment from employees. Eventually the diseconomies of scale become greater than the economies of scale.

Figure 1-6 Example of increasing and decreasing returns to scale



Environmental factors

Environmental factors might also affect productivity. For example, if a hospital is located in an area with a relatively older population, with increased

chances of comorbidity, this will affect its productivity. When comparing the productivity of hospitals, we want to ensure we account for environmental factors. If we identify something as an inefficiency, we want to be sure that this is actually due to factors that are under the control of management and not due to environmental factors that cannot be influenced. Consequently, productivity analysis often considers environmental factors.

The difference between inefficiency and environmental factors resembles the difference between endogenous and exogenous factors. Although the modelling and interpretation of exogenous and endogenous factors are different, the distinction between the two is not always clear-cut. Often, it is debatable whether a factor is endogenous or exogenous. The short and long terms play a role, as does perspective. For example, ownership is exogenous in the short run but might be endogenous in the long run. In most cases ownership is regulated which means it will take some time to change ownership. Furthermore, a hospital might not have the authority to change its own ownership status, whereas governments do have that power, so perspective matters. It should be clear that including environmental factors ensures that inefficiency is due to endogenous factors.

Empirical methods

Efficiency is determined by comparing actual productivity with the productivity of a best practice. Each hospital has a reference best practice; in fact, there is a best practice for each combination of inputs and outputs. These can be described by means of a mathematical function. The function which describes the various best practices is the so-called frontier function. The frontier function is unobservable, but fortunately we can proxy it with empirical estimates. Estimating frontier functions is the core business of productivity and efficiency analysis.

The concept of a frontier is easily linked to the concept of production technology: the frontier encloses all feasible combinations of inputs and outputs of a production technology and shows what is maximally possible in

terms of productivity. The frontier can be seen as the boundary of what is possible given a certain technology (see also Figure 1-1). Furthermore, a new production technology can be viewed as a frontier shift (Figure 1-2). While the distance to the frontier is a measure of efficiency (Figure 1-3).

Estimating a frontier implies a number of choices. Previously mentioned the measurement of inputs and outputs including aggregation issues. Next there has to be a decision on the production structure. Are we interested in technical or economic efficiency? And, in the case of economic efficiency, which economic behaviour applies? Furthermore, there has to be a decision on orientation: is efficiency measured in terms of a reduction of the inputs at a given level of output (input orientation) or in terms of an increase in the outputs given the inputs (output orientation), or does a combination of input reduction and output increase apply?

Once these choices have been made, we have to decide on the estimation method. Basically, there are two main groups of methods: parametric and non-parametric. The parametric methods use a functional form for which parameters are estimated by econometric techniques. The non-parametric methods use linear programming to determine the shape of the frontier. Both main groups include numerous sub variants, creating a range of options for frontier estimation.

Both methods emerged at the end of the 1970s. The origin of the parametric method dates back to 1977, with the papers of Aigner, Lovell and Schmidt (1977) and Meeusen and Van den Broeck (1977). The non-parametric method is a year younger and has its origin in the seminal article by Charnes, Cooper and Rhodes (1978). It goes beyond the scope of this thesis to give full details for both methods and their various sub variants. For an introduction, several standard works are available. For example, Fried et al.(2008) provides a thorough introduction to both. There are also introductions aimed specifically at the healthcare sector, such as Jacobs et al. (2006) for both methods, Ozcan

(2008) for the non-parametric method and Blank and Valdmanis (2008) for a number of applications of the two methods with a focus on the hospital sector.

Applications of the two methods in the hospital sector have their origin in studies by Grosskopf and Valdmanis (1987) for the non-parametric method and Zuckerman et al. (1994) for the parametric method. Since then, a vast number of articles with more applications for the sector have been published. Hollingsworth (2003, 2008) provides an overview of applications in healthcare, with the 2008 article counting 317 papers that apply frontier analysis. More than half of these are about the hospital sector, which makes this clearly the most popular sector for researchers. There are also review studies focusing specifically on hospitals. Rosko and Mutter (2011) provide an overview of the lessons learned from applications of the parametric method in the hospital sector. O'Neill et al. (2008) give an overview of the non-parametric method and the various ways in which these studies are applied. Hadji et al. (2014) provide a systematic overview of the input and output indicators used in productivity studies for hospitals.

1.4 Literature review

Efficiency and environmental factors

Although the modelling and interpretation of exogenous and endogenous factors are different, for the sake of convenience we do not explicitly differentiate between exogenous and endogenous factors in this section. Besides, as noted previously, the distinction is not always clear-cut.

Improving efficiency means catching up with the frontier. Moving towards the frontier requires an insight in the determinants of efficiency: what are the common characteristics of hospitals with high efficiency? The literature on determinants of hospital efficiency is extensive. Here we give a brief overview of determinants studied. A complete literature review of all determinants is beyond the scope of this introduction, the primary goal of which is to provide

a general overview of relevant research topics. More extensive literature reviews can be found in some of the studies already mentioned: O'Neill et al. (2008) for an overview of the non-parametric method, Rosko and Mutter (2011) for an overview of the parametric method and Hollingsworth (2008) for both.

Reimbursement systems are a popular topic of productivity and efficiency studies. The results are consistent on reimbursement incentives. In the US the Prospective Payment System (PPS) applies to a part of the patient population at US hospitals. With only an occasional exception, most studies on the effects of PPS the have indeed found its expected positive effects (Rosko & Mutter, 2011). Critical Access Hospitals – rural hospitals funded on the basis of actual costs to ensure accessibility in sparsely populated areas – have proven less efficient (Rosko & Mutter, 2011). Jakobsen (2010), in a review of studies analysing the effects of Activity-Based Reimbursement in Scandinavian hospitals, reports mixed results: the numbers of studies finding positive and non-positive effects are about equal.

The organization of the market for hospital services has several angles that have been researched. Health Maintenance Organizations (HMOs) are supposed to stimulate the efficiency of hospitals by applying various instruments, including Managed Care, to curb the cost of healthcare. Several studies have found that a higher penetration of a HMO does indeed induce greater efficiency (Rosko & Mutter, 2011). Bates et al. (2006) examine the effects of HMOs more in detail, finding that HMO penetration does possibly correlate with other explanatory variables that also affect efficiency. They demonstrate this by estimating several models and show that additional explanatory variables reduce the estimated effect of HMO penetration on efficiency. Competition or concentration of market power are another aspect of the market that has been researched. In the case of competition, there seems to be no consistency in the results in respect of efficiency: both positive and negative effects have been found (Rosko & Mutter, 2011).

Another popular research topic in the hospital sector is the effect of ownership (commercial, not for profit, public). Review studies such as Hollingsworth (2008) and Rosko and Mutter (2011) pay explicit attention to the effects of ownership. Rosenau and Linder (2003) and Tiemann et al. (2012) are both review studies that focus exclusively on the effects of ownership on efficiency. Overall, however, the findings in this area are inconclusive: some studies identify for-profit hospitals as the most efficient, whereas others identify not-for-profit hospitals as the most efficient.

Related to ownership is system membership or participation in a network. The impact of this factor on the efficiency is the research topic of a couple of studies. Based on a limited number of studies, Rosko and Mutter (2011) conclude that participation in a network or system has a positive effect on efficiency. However, they emphasize that the results of the studies reviewed need to be put into context – the point being that not every network or system is the same. Membership of a system alone does not tell the whole story. This conclusion is based on Rosko et al. (2007), which categorizes systems based on the degree of centralization within them. A system that is less centralized turns out to be more efficient. However, this result is at odds with Bazzoli et al. (2000), which finds that moderate decentralization is most efficient. That is followed by centralization, with full decentralization as the least efficient option.

The effect of mergers is a topic with two angles to it. Most obvious are the scale effects, but there are also efficiency effects. Often, the scale effects are studied ex ante – as, for example, by Preyra and Pink (2006) for Canadian hospitals and by Azevedo and Mateus (2014) for Portuguese hospitals. Bazzoli (2008) provides an overview of studies on the price and cost effects of consolidation in the American hospital market. With regard to efficiency, Bazzoli notes that there are two possible relevant aspects in the case of mergers. First, there might be a merger effect: an efficiency difference between merged and non-merged hospitals. And second, there might be a pre-merger

effect: an improvement in efficiency derived from low efficiency prior to the merger. The general trend, according to Bazzoli, is that small improvements in efficiency are possible through mergers.

Because it is unlikely that the output mixes of the merging hospitals are identical, the output mix of the merged hospital will change. A merger therefore involves not only economies of scale, but also economies of scope. These are the effects of the joint production of different products. Wagstaff and Lopez (1996), Sinay (1998a; 1998b), Preyra and Pink (2006) and Kristensen et al. (2010) study the scale effects as well as the scope effects of merging. Of course, economies of scope are not exclusive to mergers. There are a few studies that include results on economies of scope; examples include Vita (1990), Wholey et al. (1996), Prior (1996), Menke (1997), Li and Rosenman (2001a, 2001b) and Smet (2007). Carey et al. (2008, 2015) study a related topic, focusing on specialization within the hospital industry by comparing hospitals with specialized clinics. Linna and Häkkinen (1999) use the degree of specialization as an explanation for differences in efficiency.

The comparability of the measured production of hospitals is always an issue in productivity analysis. Homogeneity of output is a basic requirement. This is not always the case, though, due to differences in case-mix and differences in delivered quality. Zuckerman et al. (1994) emphasize inclusion of direct measures of illness severity, output quality, and patient outcomes to reduce the likelihood that the inefficiency estimates are capturing unmeasured differences in hospital outputs. Depending on the data available, models include different case-mix and quality indicators. Rosko and Mutter (2011) provide an overview of the following applied case-mix and quality indicators: proportion of inpatient days on intensive care units, births as a proportion of admissions, intra-DRG severity of illness index, the number of high-technology services, teaching status, full-time equivalent resident physicians, joint Commission on Accreditation of Healthcare Organizations accreditation, number of board-certified medical staff per bed, existence of a transplant

programme, (risk-adjusted) mortality rates, risk-adjusted patient safety event rates and patient burden of illness. A frequently used indicator for case-mix is teaching versus non-teaching status of the hospital (Hollingsworth, 2008; Rosko & Mutter, 2011). Moreover, there are also studies that explicitly examine the efficiency differences between teaching and non-teaching hospitals; see for example, Lopez-Casasnovas and Saez (1999), Grosskopf et al. (2001) and Farsi and Filippini (2008).

In addition to the aforementioned factors, there are several other factors for which the effect on efficiency has been studied. Some are specific, but there are also factors included in multivariate models. To illustrate, we give a brief overview.

First, there are a number of factors that deal with the management of operations and the characteristics of the patient population. Zuckerman et al. (1994), for example, studies the impact of occupancy rates, personnel per admission, salary costs, assets and the average income of households in the service area. Ferrier and Valdmanis (1996) study rural hospitals and, besides factors already mentioned – such as ownership and quality – also include occupancy rates, intensity of care, outpatient care and regional differences. Smet (2007) focuses on occupancy rates by taking demand uncertainty into account. Demand uncertainty induces a certain amount of spare capacity. Linna and Häkkinen (1999) test a large number of explanatory factors, including the degree of specialization, use of modern technology (e.g. treating patients in outpatient clinics), input allocation, quality control, patient transfers to other care facilities and demographic characteristics of the patient population, such as the need for care and average income. Carey (2000) and Herr (2008) show the effect of the average length of stay. Martinussen and Midttun (2004) show the effect of the share of day-care treatments. Farsi (2008) looks at both occupancy rates and average length of stay. Vitikainen et al. (2010) study the effects of substituting clinical care with outpatient care.

Another topic of research is regional or national differences in efficiency. Examples include Steinmann et al. (2004), comparing German and Swiss hospitals, Dervaux et al. (2004), comparing French and American hospitals, and Linna et al. (2006), Medin et al. (2011) and Medin et al. (2013), all comparing hospitals in the Nordic countries. One difficult issue in an international comparison, though, is interpreting the results because the institutional characteristics of different countries vary in many ways. One exception is the study by Pilyavsky et al. (2006), which addresses differences within the Ukraine. Here, institutional characteristics are fixed for the regions analysed. This study draws a distinction between the east and the west of the country, and relates efficiency to different spheres of influence (i.e. Soviet-style planned economies versus western management and medical "business" practice).

The internal organization is also subject of research. Rosko (1996) provides a literature review of factors explaining differences in efficiency. In addition to the aforementioned factors, this includes participation of doctors on the hospital board and the personal characteristics of physicians (gender and experience) as explanatory factors for differences in efficiency. Cuellar and Gertler (2006) study effects of the integration of the medical specialist in the hospital. Others investigate the "make or buy" decision. Coles and Hesterly (1998) and Ludwig et al. (2009) examine the effects of outsourcing. Carey and Dor (2008) examine the effects of outsourcing the management (contract management). Ludwig et al. (2010) use the principal-agent theory to explain differences in efficiency, including that of the various departments within the hospital.

Scale

Implicitly, all productivity studies deal with scale because a decision has to be made on the presence or absence of scale effects (either constant or variable returns to scale). Many productivity studies pay explicit attention to scale effects. Roughly speaking, there are three types of such study: those in which

the central question is about the scale, those on mergers and those in which scale is one factor amongst many.

Research on the optimum scale of hospitals is not the exclusive domain of efficiency studies. Such studies were being undertaken long before the parametric and non-parametric methods of efficiency research existed. Early examples are those by Carr and Feldstein (1967) and Berry (1967). Ever since that time, moreover, there has been disagreement about scale effects at hospitals.

Posnett (1999) indicates that, despite the large number of studies on scale and the diversity of the methods used, the evidence on scale effects for hospitals is consistent. Those with up to 200 beds have economies of scale, those with between 200 and 400 beds have an optimum scale and at 400 to 600 beds there are diseconomies of scale. Nevertheless, the debate about the optimum scale continues.

Differences in research results are partly due to methodological choices. A recurring point here is the ex-ante specification of the functional form needed in applications of the parametric method. Vitaliano (1987) applies several specifications to the same sample, in different cases finding both a U-shaped pattern of average costs and a downward trend. The non-parametric method does not require an ex-ante specification: the data define the shape of the cost function. Banker et al. (1986) compare the parametric and non-parametric method by applying both methods to a set of hospitals in North Carolina. The parametric method finds constant returns of scale, whereas the non-parametric method is more flexible and finds both increasing and decreasing returns to scale. Wilson and Carey (2004) deal explicitly with scale effects. Their conclusion is that, due to possible misspecifications, studies that apply the parametric method conclude too early that diseconomies of scale may occur.

Technical change

Studies on the effects of technical change date back to Solow (1957). He distinguishes a shift from a production function and a movement along the production function. Since then, a large number of studies have been published in which technical change has been investigated. Studies of technical change usually compare productivity at various points in time. The difference in productivity in different time periods is the effect of technical change.

A number of studies decompose hospital productivity growth into efficiency changes and technical change, and in some cases scale effects as well. In general, these studies use Malmquist indices for the decomposition. Burgess Jr and Wilson (1995) decompose the productivity growth of US hospitals between 1985 and 1989. One of their conclusions is that new technology makes new or better medical treatments possible, and in that sense new technology means progress, but at the same time these new technologies reduce productivity and so – from an economic perspective – imply regression. Maniadakis et al. (1999) and Maniadakis and Thanassoulis (2000) make a decomposition of productivity growth for hospitals in the United Kingdom after the reform of the NHS in 1991. Both studies find a similar pattern: an initial decline in productivity, followed by growth. Remarkably, the first study attributes this growth to technical change while the second attributes it to (allocative) efficiency. A non-exhaustive list of other decompositions includes Finnish hospitals (Linna, 1998), Northern Irish hospitals (McCallion et al., 2000), Austrian hospitals (Sommersguter-Reichmann, 2000), Portuguese hospitals (Barros et al., 2008; Dismuke & Sena, 1999), Greek hospitals (Dimas et al., 2012) and Australian hospitals (O'Donnell & Nguyen, 2013).

In a frontier-model framework, technical change can be viewed as a shift of the frontier. The magnitude of the shift can be determined by including a time variable in the frontier model. Doing this means that changes to productivity over time are absorbed by the time variable. The underlying assumption is that changes in productivity over time are due to technical change. The time

variable can be modelled as a trend variable or by adding dummy variables for each time period. The former is the option most frequently applied, with examples including Rosko (2001), Blank and Merkies (2004), Farsi (2008) and Ludwig et al. (2009). Examples of time modelled with annual dummy variables are Zwanziger and Melnick (1988), Linna (1998), Biorn et al. (2003), Farsi and Filippini (2006, 2008), Herr (2008) and Herr et al. (2011).

Modelling with a trend variable assumes that productivity change is a smooth process over time, with productivity growth being more or less the same in each time period. Dummy variables allow for productivity growth that varies in each period. Blank and Vogelaar (2004) compare a model with a trend variable with a model with dummy variables. Their conclusion is that, in the case of Dutch hospitals between 1993 and 2000, technical change occurred shock-wise and so dummy variables are preferable.

Modelling technical change by adding a time trend or annual dummies is practical, but at the same it makes a bold assumption. All changes over time which have an effect on productivity, except changes of scale and efficiency, are labelled as technical change. Because of this, the computed technical change is a mishmash of all kinds of changes over time. In that perspective, even policy changes that have an impact on productivity are viewed as technical change. Furthermore, a time variable only provides insight into the magnitude of the productivity change and not into what enabled the productivity growth.

Studies on Dutch hospitals

Because the empirical applications in this thesis concern Dutch hospitals, this literature review ends with a brief overview of productivity studies on Dutch hospitals. We concentrate here on the most important studies. It should be noted that some have only been published in Dutch.

The studies by van Aert (1977) and van Montfort (1980) are the precursors of productivity studies on Dutch hospitals. Van Aert demonstrates the use of

cost functions in the Dutch hospital sector, while van Montfort examines production functions. Both, however, neglect the possibility of differences in efficiency. Rather, their aim is more or less to describe the production structure of hospitals. One of their main contributions is the measurement of inputs and output in the Dutch hospital sector.

Blank et al. (1998) studies the Dutch hospitals extensively. It contains applications of both parametric and non-parametric methods, providing an insight into production structures as well as efficiency. A wide variety of model specifications is applied and tested, making this study a rich source (if not the standard work) for modelling the Dutch hospital industry. Blank et al. (2002) is less extensive and uses the parametric method to pinpoint the efficiency and inefficiency effects of bureaucracy. The thesis by Ludwig (2008), on the efficiency of Dutch hospitals, uses the parametric method to determine the efficiency of Dutch hospital and then looks in depth at three issues to explain differences in that efficiency: quality, efficiency at the department level and outsourcing. Of more recent date are three studies, in a series on the productivity of Dutch hospitals. These studies use the parametric method to shed more light on productivity changes since the last decade (Blank et al., 2011), the role of innovations in efficiency (Dumaij et al., 2012) and the effects of governance on efficiency (Blank et al., 2013).

1.5 Aim of this thesis

This thesis researches the possibilities to increase productivity of hospitals. As discussed, there are three sources of productivity growth: scale, efficiency, and technical change. Efficiency and scale are both topics that have already been studied extensively, so only a minor part of this thesis focuses on them. Technical change is also studied regularly, but for the most part only in terms of its magnitude. This thesis models technical change in another way and ascribes productivity changes to innovations. Furthermore it develops a

framework to decompose productivity changes into productivity changes for individual inputs.

The scale of hospitals is a frequent research topic. However, often only the scale effects for the average hospital are reported, the optimum scale is neglected. Furthermore, it is remarkable that, with such a wide range of studies presenting results on scale effects for hospitals, there is no recent review study that systematically uses the results to find an optimum scale. This thesis provides that systematic review and answers the following research question:

1. **What is the optimum scale for hospitals?**

The literature on the efficiency of hospitals is extensive, and the effect of a wide variety of endogenous and exogenous factors has been studied – especially factors like the influence of funding, market environment, ownership and mergers. The influence of operational management and internal organization, including the influence and role of medical specialist members of boards of directors, has also been the subject of several studies. This thesis extends this topic, focusing on the relationship between the governance of a hospital and its efficiency:

2. **How does the corporate governance of hospitals contribute to their efficiency?**

A gap in productivity research is the modelling of factors underlying technical change. Currently technical change is modelled with a time variable, results are limited to the amount of productivity change through time. This is unsatisfactory because it remains unknown what technical change really is and what actually enabled productivity change. Despite the numerous studies on factors explaining differences in hospital efficiency, little is known about the factors underlying technical change. This thesis tries to develop greater insight in this area by modelling technical change using technology indices. This allows us to answer the following question:

3. **What is the contribution of new technologies to productivity of hospitals?**

Using technology indices might give an answer to the contribution of new technologies, it also raises new questions. The amount of new technology that a hospital adopts is endogenous. The effects of a new technology on productivity are not limited to one period, once the hospital has a technology it also profits from it in the future. At the same time adoption might come with adjustment costs, that are limited to a short period. The question is how to deal with technologies with short-term adjustment costs and long-term benefits. The thesis investigates modelling consequences:

4. **How can we model adjustment costs and the inter-temporal effects of new technology ?**

Technical change is regarded to affect the production process in a neutral way, implicitly it is assumed that technical change is Hicksian neutral. Technical change only causes a shift of the frontier. However, technical change may have many appearances. For instance, technical change may affect the optimal allocation of inputs. This implies that substitution of inputs takes place, which in turn affects factor productivities. For example, labour productivity can be raised by substituting capital for labour. From a policy perspective, factor productivities are interesting because they can be used in predicting future demand for various inputs. The final question of this thesis relates to the productivity development of individual inputs:

5. **How can we decompose total factor productivity into factor productivities?**

1.6 Thesis outline

Chapter 2 answers the first research question and investigates the optimum scale of hospitals. For that purpose a meta-analysis is performed on parametric studies that include results on the scale of hospitals.

Chapter 3 answers the question on the contribution of the governance to the efficiency of hospitals. A non-parametric method is applied to determine

the efficiency of Dutch hospitals in a first stage. In a second stage differences in efficiency are related to governance characteristics by applying the bootstrap procedures proposed by Simar and Wilson (2007).

Chapter 4 investigates the impact of new technologies on productivity. A parametric cost function is estimated for Dutch hospitals. Instead of modelling technical change using a time variable, technology indices are added to the model.

Chapter 5 models inter-temporal effects of new technology. The chapter uses a parametric cost function and develops a framework that incorporates trade-offs between long-term benefits and short-term adjustment costs. As a consequence, an additional equation for the optimum amount of new technology is added to the model. An application of the extended model is demonstrated for Dutch hospitals.

Chapter 6 shows how total factor productivity can be decomposed into productivity change per factor. The decomposition in factor productivities is demonstrated on the Dutch hospital industry using a parametric estimated cost model.

Chapter 7 contains the conclusion of this thesis. It summarizes the main findings on scale, efficiency, and technical change, discusses the policy implications of the findings and ends by discussing limitations of the study and opportunities for future research.

References

Aigner, D., Lovell, C.A.K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6(1), 21-37.

Azevedo, H., & Mateus, C. (2014). Cost effects of hospital mergers in Portugal. European Journal of Health Economics, 15(9), 999-1010.

Banker, R.D., Conrad, R.F., & Strauss, R.P. (1986). A comparative application of data envelopment analysis and translog methods: an illustrative study of hospital production. Management Science, 32(1), 30-44.

Barros, C.P., de Menezes, A.G., Peypoch, N., Solonandrasana, B., & Vieira, J.C. (2008). An analysis of hospital efficiency and productivity growth using the Luenberger indicator. Health Care Management Science, 11(4), 373-381.

Bates, L.J., College, B., Mukherjee, K., & Santerre, R.E. (2006). Market structure and technical efficiency in the hospital services industry: A DEA approach. Medical Care Research and Review, 63(4), 499-524.

Bazzoli, G.J. (2008). Hospital consolidation and integration activity in the United States. In J.L.T. Blank & V.G. Valdmanis (Eds.), Evaluating hospital policy and performance: contributions from hospital policy and productivity research (Vol. 18, pp. 45-61). Oxford: Elsevier.

Bazzoli, G.J., Chan, B., Shortell, S.M., & D'Aunno, T. (2000). The financial performance of hospitals belonging to health networks and systems. Inquiry-the Journal of Health Care Organization Provision and Financing, 37(3), 234-252.

Berry, R.E. (1967). Returns to Scale in the Production of Hospital Services. Health Services Report, 2(2), 123-139.

Biorn, E., Hagen, T.P., Iversen, T., & Magnussen, J. (2003). The effect of activity-based financing on hospital efficiency: a panel data analysis of DEA efficiency scores 1992-2000. Health care management science, 6(4), 271-283.

Blank, J.L.T., Dumaij, A.C.M., & van Hulst, B.L. (2011). Ziekenhuismiddelen in verband. Een empirisch onderzoek naar productiviteit en

doelmatigheid in de Nederlandse ziekenhuizen 2003-2009 IPSE Studies Research Reeks. Delft: TU Delft.

Blank, J.L.T., Eggink, E., & Merkies, A.H.Q.M. (1998). Tussen Bed en Budget. Rijswijk: Sociaal en Cultureel Planbureau.

Blank, J.L.T., Koolmees, W., Vogelaar, I., & Waaijers, J.L. (2002). Tussen bureau en bed: een empirisch onderzoek naar de relatie tussen overhead en productie in Nederlandse ziekenhuizen. Rotterdam: ECORYS-NEI.

Blank, J.L.T., & Merkies, A. (2004). Empirical assessment of the economic behaviour of Dutch general hospitals. Health Economics, 13(3), 265-280.

Blank, J.L.T., & Valdmanis, V.G. (2008). Evaluating hospital policy and performance: contributions from hospital policy and productivity research. Oxford: Elsevier JAI.

Blank, J.L.T., & Valdmanis, V.G. (2013). Principles of productivity measurement: an elementary introduction to quantative research on the productivity, efficiency, effectiveness and quality of the public sector. Maastricht: Shaker Publishing B.V.

Blank, J.L.T., van Hulst, B.L., & Dumaij, A.C.M. (2013). Besturen sturen: Empirisch onderzoek naar het effect van interne sturing op de doelmatigheid van de Nederlandse ziekenhuizen in de periode 2003-2010 IPSE Studies Research Reeks. Delft: IPSE Studies.

Blank, J.L.T., & Vogelaar, I. (2004). Specifying technical change: a research on the nature of technical change in Dutch hospital industry. The Quarterly Review of Economics and Finance, 44, 448-463.

Burgess Jr, J.F., & Wilson, P.W. (1995). Decomposing hospital productivity changes, 1985–1988: a nonparametric Malmquist approach. Journal of Productivity Analysis, 6(4), 343-363.

Carey, K. (2000). Hospital cost containment and length of stay: An econometric analysis. Southern Economic Journal, 67(2), 363-380.

Carey, K., Burgess, J.F., Jr., & Young, G.J. (2008). Specialty and full-service hospitals: A comparative cost analysis. Health Services Research, 43(5), 1869-1887.

Carey, K., Burgess, J.F., & Young, G.J. (2015). Economies of scale and scope: the case of specialty hospital. Contemporary Economic Policy, 33(1), 104-117.

Carey, K., & Dor, A. (2008). Expense preference behavior and management "outsourcing": a comparison of adopters and non-adopters of contract management in US hospitals. Journal of Productivity Analysis, 29(1), 61-75.

Carr, W.J., & Feldstein, P.J. (1967). The relationship of cost to hospital size. Inquiry, 45-65.

Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2(6), 429-444.

Coles, J.W., & Hesterly, W.S. (1998). Transaction Costs, Quality, and Economies of Scale: Examining Contracting Choices in the Hospital Industry. Journal of Corporate Finance, 4(4), 321-345.

Cuellar, A.E., & Gertler, P.J. (2006). Strategic integration of hospitals and physicians. Journal of Health Economics, 25(1), 1-28.

Debreu, G. (1951). The Coefficient of Resource Utilization. Econometrica, 19(3), 273-292.

Dervaux, B., Ferrier, G.D., Leleu, H., & Valdmanis, V. (2004). Comparing French and US hospital technologies: a directional input distance function approach. Applied Economics, 36(10), 1065-1081.

Dimas, G., Goula, A., & Soulis, S. (2012). Productive performance and its components in Greek public hospitals. Operational Research, 12(1), 15-27.

Dismuke, C.E., & Sena, V. (1999). Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portuguese public hospitals? An empirical analysis using parametric and non-parametric methods. Health care management science, 2(2), 107-116.

Dumaij, A.C.M., Blank, J.L.T., & van Hulst, B.L. (2012). Zorg voor versnelling: Empirisch onderzoek naar het effect van innovaties op de doelmatigheid van Nederlandse ziekenhuizen in de periode 2003-2009 IPSE Studies Research Reeks. Delft: IPSE Studies.

European Commission (2007). White paper–together for health: A strategic approach for the EU 2008–2013. European Commission Brussels.

European Commission (2012). Commission Staff Working Document on an Action Plan for the EU Health Workforce

Strasbourg.

Farrell, M.J. (1957). The Measurement of Productive Efficiency. Journal of the Royal Statistical Society. Series A (General), 120(3), 253-290.

Farsi, M. (2008). The temporal variation of cost-efficiency in Switzerland's hospitals: an application of mixed models. Journal of Productivity Analysis, 30(2), 155-168.

Farsi, M., & Filippini, M. (2006). An Analysis of Efficiency and Productivity in Swiss Hospitals. Schweizerische Zeitschrift fur Volkswirtschaft und Statistik/Swiss Journal of Economics and Statistics, 142(1), 1-37.

Farsi, M., & Filippini, M. (2008). Effects of ownership, subsidization and teaching activities on hospital costs in Switzerland. Health Economics, 17(3), 335-350.

Ferrier, G.D., & Valdmanis, V. (1996). Rural hospital performance and its correlates. Journal of Productivity Analysis, 7(1), 63-80.

Fried, H.O., Lovell, C.A.K., & Schmidt, S.S. (2008). The measurement of productive efficiency and productivity growth. New York: Oxford University Press.

Grosskopf, S., Margaritis, D., & Valdmanis, V. (2001). Comparing teaching and non-teaching hospitals: a frontier approach (teaching vs. non-teaching hospitals). Health care management science, 4(2), 83-90.

Grosskopf, S., & Valdmanis, V. (1987). Measuring hospital performance: A non-parametric approach. Journal of Health Economics, 6(2), 89-107.

Hadji, B., Meyer, R., Melikeche, S., Escalon, S., & Degoulet, P. (2014). Assessing the Relationships Between Hospital Resources and Activities: A Systematic Review. Journal of Medical Systems, 38(10).

Health Workforce Australia (2012). Health Workforce 2025. Doctors, Nurses and Midwives (Vol. 1). Adelaide.

Herr, A. (2008). Cost and technical efficiency of German hospitals: Does ownership matter? Health Economics, 17(9), 1057-1071.

Herr, A., Schmitz, H., & Augurzky, B. (2011). Profit efficiency and ownership of German hospitals. Health Economics, 20(6), 660-674.

Hollingsworth, B. (2003). Non-parametric and parametric applications measuring efficiency in health care. Health care management science, 6(4), 203-218.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. Health Economics, 17(10), 1107-1128.

Jacobs, R., Smith, P.C., & Street, A. (2006). Measuring efficiency in health care. Analytic techniques and health policy. Cambridge and New York: Cambridge University Press.

Jakobsen, M.L.F. (2010). The Effects of New Public Management: Activity-based Reimbursement and Efficiency in the Scandinavian Hospital Sectors. Scandinavian Political Studies, 33(2), 113-134.

Jones, N., & Charlesworth, A. (2013). The anatomy of health spending 2011/12: Nuffield Trust.

Koopmans, T.C. (1951). An analysis of production as efficient combination of activities. In T.C. Koopmans (Ed.), Cowless commission for research in economics, monograph no. 13. New York: John Wiley and Sons, Inc.

Kristensen, T., Bogetoft, P., & Pedersen, K.M. (2010). Potential gains from hospital mergers in Denmark. Health Care Management Science, 13(4), 334-345.

Li, T., & Rosenman, R. (2001a). Cost inefficiency in Washington hospitals: a stochastic frontier approach using panel data. Health care management science, 4(2), 73-81.

Li, T., & Rosenman, R. (2001b). Estimating hospital costs with a generalized Leontief function. Health Economics, 10(6), 523-538.

Linna, M. (1998). Measuring hospital cost efficiency with panel data models. Health Economics, 7(5), 415-427.

Linna, M., & Häkkinen, U. (1999). Determinants of Cost efficiency and Finnish Hospitals: A Comparison of DEA and SFA.

Helsinki/Edmonton: National Research and Development Centre for Welfare and Health.

Linna, M., Hakkinen, U., & Magnussen, J. (2006). Comparing hospital cost efficiency between Norway and Finland. Health Policy, 77(3), 268-278.

Lopez-Casasnovas, G., & Saez, M. (1999). The impact of teaching status on average costs in Spanish hospitals. Health Economics, 8(7), 641-651.

Ludwig, M. (2008). Efficiency of Dutch hospitals. Maastricht: Datawyse.

Ludwig, M., Groot, W., & Van Merode, F. (2009). Hospital efficiency and transaction costs: A stochastic frontier approach. Social Science & Medicine, 69(1), 61-67.

Ludwig, M., Van Merode, F., & Groot, W. (2010). Principal agent relationships and the efficiency of hospitals. European Journal of Health Economics, 11(3), 291-304.

Maniadakis, N., Hollingsworth, B., & Thanassoulis, E. (1999). The impact of the internal market on hospital efficiency, productivity and service quality. Health care management science, 2(2), 75-85.

Maniadakis, N., & Thanassoulis, E. (2000). Assessing productivity changes in UK hospitals reflecting technology and input prices. Applied Economics, 32(12), 1575-1589.

Martinussen, P.E., & Midttun, L. (2004). Day surgery and hospital efficiency: empirical analysis of Norwegian hospitals, 1999-2001. Health Policy, 68(2), 183-196.

McCallion, G., Glass, J.C., Jackson, R., Kerr, C.A., & McKillop, D.G. (2000). Investigating productivity change and hospital size: a nonparametric frontier approach. Applied Economics, 32(2), 161-174.

Medin, E., Anthun, K.S., Hakkinen, U., Kittelsen, S.A.C., Linna, M., Magnussen, J., . . . Rehnberg, C. (2011). Cost efficiency of university hospitals in the Nordic countries: a cross-country analysis. European Journal of Health Economics, 12(6), 509-519.

Medin, E., Hakkinen, U., Linna, M., Anthun, K.S., Kittelsen, S.A.C., Rehnberg, C., & EuroHope Study, G. (2013). International hospital productivity comparison: Experiences from the Nordic countries. Health Policy, 112(1-2), 80-87.

Meeusen, W., & Van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review(8), 435–444.

Menke, T.J. (1997). The effect of chain membership on hospital costs. Health Services Research, 32(2), 177-196.

O'Donnell, C.J., & Nguyen, K. (2013). An econometric approach to estimating support prices and measures of productivity change in public hospitals. Journal of Productivity Analysis, 40(3), 323-335.

O'Neill, L., Rauner, M., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. Socio-Economic Planning Sciences, 42(3), 158-189.

OECD (2008). The Looming Crisis In the Health Workforce. How can oecd countries respond? OECD Health Policy Studies. Paris.

OECD (2010). Health care systems. Efficiency and policy settings. In OECD_Publishing (Ed.). Paris: OECD.

OECD/European Union (2014). Health at a Glance: Europe 2014: OECD Publishing.

Orosz, E., & Morgan, D. (2004). SHA-based National Health Accounts in Thirteen OECD Countries. Document de travail de l'OCDE sur la santé(16).

Ozcan, Y.A. (2008). Health care benchmarking and performance evaluation. New York: Springer.

Pilyavsky, A.I., Aaronson, W.E., Bernet, P.M., Rosko, M.D., Valdmanis, V.G., & Golubchikov, M.V. (2006). East-west: does it make a difference to hospital efficiencies in Ukraine? Health Economics, 15(11), 1173-1186.

Posnett, J. (1999). The hospital of the future - Is bigger better? Concentration in the provision of secondary care. British Medical Journal, 319(7216), 1063-1065.

Preyra, C., & Pink, G. (2006). Scale and scope efficiencies through hospital consolidations. Journal of Health Economics, 25(6), 1049-1068.

Prior, D. (1996). Technical efficiency and scope economics in hospitals. Applied Economics, 28(10), 1295-1301.

Rosenau, P.V., & Linder, S.H. (2003). Two decades of research comparing for-profit and nonprofit health provider performance in the United States. Social Science Quarterly, 84(2), 219-241.

Rosko, M.D. (1996). Understanding variations in hospital costs: An economics perspective. Annals of Operations Research, 67, 1-21.

Rosko, M.D. (2001). Cost efficiency of US hospitals: A stochastic frontier approach. Health Economics, 10(6), 539-551.

Rosko, M.D., & Mutter, R.L. (2011). What have we learned from the application of stochastic frontier analysis to U.S. hospitals? Medical care research and review : MCRR, 68(1 Suppl), 75S-100S.

Rosko, M.D., Proenca, J., Zinn, J.S., & Bazzoli, G.J. (2007). Hospital inefficiency: What is the impact of membership in different types of systems? Inquiry-the Journal of Health Care Organization Provision and Financing, 44(3), 335-349.

Rumbold, B.E., Smith, J.A., Hurst, J., Charlesworth, A., & Clarke, A. (2014). Improving productive efficiency in hospitals: findings from a review of the international evidence. Health Economics, Policy and Law, 1-23.

Simar, L., & Wilson, P.W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 136(1), 31-64.

Sinay, U.T. (1998a). Hospital mergers and closures: Survival of rural hospitals. Journal of Rural Health, 14(4), 357-365.

Sinay, U.T. (1998b). Pre- and Post-Merger Investigation of Hospital Mergers. Eastern Economic Journal, 24(1), 83-97.

Smet, M. (2007). Measuring performance in the presence of stochastic demand for hospital services: an analysis of Belgian general care hospitals. Journal of Productivity Analysis, 27(1), 13-29.

Solow, R.M. (1957). Technical change and the aggregate production function. Review of Economics and Statistics, 39(3), 312-320.

Sommersguter-Reichmann, M. (2000). The impact of the Austrian hospital financing reform on hospital productivity: empirical evidence on efficiency and technology changes using a non-parametric input-based Malmquist approach. Health care management science, 3(4), 309-321.

Steinmann, L., Dittrich, G., Karmann, A., & Zweifel, P. (2004). Measuring and comparing the (in)efficiency of German and Swiss hospitals. The European journal of health economics : HEPAC : health economics in prevention and care, 5(3), 216-226.

Tiemann, O., Schreyoegg, J., & Busse, R. (2012). Hospital ownership and efficiency: A review of studies with particular focus on Germany. Health Policy, 104(2), 163-171.

van Aert, J.H. (1977). Ziekenhuiskosten in econometrisch perspectief. Stenfert Kroese, Leiden.

van Montfort, A.P.W.P. (1980). Production functions for general hospitals : an econometric analysis. van Montfort, Tilburg.

Vita, M.G. (1990). Exploring hospital production relationships with flexible functional forms. Journal of Health Economics, 9(1), 1-21.

Vitaliano, D.F. (1987). On the estimation of hospital cost functions. Journal of Health Economics, 6(4), 305-318.

Vitikainen, K., Linna, M., & Street, A. (2010). Substituting inpatient for outpatient care: what is the impact on hospital costs and efficiency? European Journal of Health Economics, 11(4), 395-404.

Wagstaff, A., & Lopez, G. (1996). Hospital costs in Catalonia: A stochastic frontier analysis. Applied Economics Letters, 3(7), 471-474.

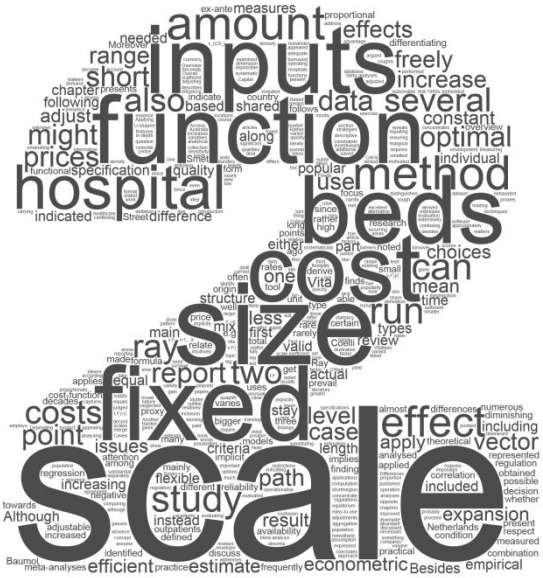
Wholey, D., Feldman, R., Christianson, J.B., & Engberg, J. (1996). Scale and scope economies among Health Maintenance Organizations. Journal of Health Economics, 15(6), 657-684.

Wilson, P.W., & Carey, K. (2004). Nonparametric analysis of returns to scale in the us hospital industry. Journal of Applied Econometrics, 19(4), 505-524.

World Health Organization (2014). World health statistics 2014. Geneva: WHO press.

Zuckerman, S., Hadley, J., & Iezzoni, L. (1994). Measuring hospital efficiency with frontier cost functions. Journal of Health Economics, 13(3), 255-280.

Zwanziger, J., & Melnick, G.A. (1988). The effects of hospital competition and the Medicare PPS program on hospital cost behavior in California. Journal of Health Economics, 7(4), 301-320.



Economies of scale

Economies of scale: a meta-analysis on the scale of hospitals

2.1 Introduction

There is an abundant supply of studies that examined the cost structure of hospitals. In an overview article, Hollingsworth (2008) identified 165 published journal articles and book chapters on the efficiency and productivity of hospitals. Hollingsworth and Street (2006) noted that the popularity of efficiency and productivity studies was increasing, most likely as a result of increased demand for adequate information for decision makers and of lower research barriers resulting from the improved availability of data and easy-to-use software. The numerous efficiency and productivity studies have in turn led to several systematic review studies (see for example Hollingsworth, 2003, 2008; Hollingsworth & Street, 2006; Nguyen & Coelli, 2009; O'Neill et al., 2008; Rosko & Mutter, 2008, 2011; Worthington, 2004).

These reviews discuss the quality and reliability of the productivity and efficiency analysis, including in-depth analysis of modelling choices such as specification and estimation strategies. If there is one thing that the numerous efficiency and productivity studies have revealed, it is that modelling choices do matter. An excellent illustration of this is the study by Nguyen and Coelli (2009), who used meta-analysis to quantify the effects of modelling choices on hospital efficiency. Based on 253 estimated models reported in 95 studies, they found the following three significant effects on efficiency: sample size has a negative (and diminishing) effect; dimension (number of variables) has a positive (and diminishing) effect; and the imposition of constant returns to scale has a negative effect. Besides review studies, there are empirical studies that demonstrated the effects of modelling choices on efficiency scores either as research topic or through the inclusion of sensitivity analysis. The focus of both review and empirical studies is mainly on efficiency and productivity; other results, such as scale, are neglected. At the same time, all productivity studies implicitly deal with scale, because a decision has to be made about the presence or absence of scale effects (either constant returns to scale or variable returns to scale).

Therefore, the focus of this chapter is on the scale of hospitals. To get an idea of hospital size, Aletras (1997) defined the size of an average hospital as a hospital with 200 to 300 beds. A rough point estimate based on the literature is 240 beds. There is, however, a wide variety in scale for individual hospitals, as the scale of hospitals varies according to the type of hospital. For example, in the USA, Critical Access Hospitals (CAHs) are rural hospitals with no more than 25 beds, whereas an academic healthcare centre can be 100 times larger with almost 2,500 beds, scattered over several locations. As a result, the average hospital size in a country depends on the mix of hospitals types. For example, a country like Australia, which has many relatively low population density areas, has many small hospitals resulting in an average size of 75 beds, whereas in the densely populated Netherlands, the smallest hospital has nearly 200 beds. Besides the mix of types of hospitals, regulation also matters. In the Netherlands, regulation incentivised hospitals to merge, resulting in an average hospital size of about 480 beds, compared to 350 beds three decades ago.

Overview studies on the scale of hospitals are rare. The study by Aletras (1997) is one of the few extensive overview studies, if not the only one, on the economies of scales of hospitals. Aletras identified five types of studies: econometric ad hoc analysis (mainly before the mid-1980s), flexible econometric cost studies, econometric production function studies, data envelopment analysis (DEA) and survival analysis (hospitals with an optimal size will gain a bigger market share). He reviewed approximately 100 studies and systematically judged their reliability on the basis of several quality of modelling criteria. The main finding was that for the flexible econometric cost studies, constant returns to scale or even decreasing returns to scale prevail for the average hospital (roughly 200 to 300 beds). Although an optimum size is impossible to pinpoint, it is clear that existing economies of scale are quickly exhausted as the size of a hospital increases. Similar results hold for DEA studies, although the optimum tends to be a slightly bigger hospital. Aletras (1997) indicated that there is general agreement between DEA studies that hospitals with fewer than 220 and more than 620 beds are scale-inefficient.

However, he also indicated that there is one conflicting study. Furthermore, this range is broad, so that most studies are included; for individual studies, the range is smaller. Aletras therefore argued that a valid range for an optimum scale for DEA studies would be 220 to 400 beds.

There are also theoretical and empirical papers that address economies of scale extensively. A recurring point is the ex-ante specification of the functional form needed in applications of the parametric method. Vitaliano (1987) applied several specifications to the same sample, finding both a U-shaped pattern for the average costs and a downward trend in average costs. For DEA, an ex-ante specification is not needed. Banker et al. (1986) compared results obtained from a translog cost function with the results obtained with DEA. Both methods were applied to the same dataset of hospitals. Their findings on scale is that the parametric method finds constant returns to scale, whereas DEA finds both increasing and decreasing returns to scale. A second point of attention is that several methods are used to derive scale effects from the estimates of a parametric cost function (Vita, 1990). Applying these methods means that some rather subtle, mostly implicit assumptions have to be made. Smet (2002) reviewed eleven studies and showed that only a few explicitly take into account the implicit assumptions and restrictions embedded in the models used.

This present study extends the work by Aletras (1997). There are two reasons to do so. First, Aletras carried out his study two decades ago, and since then a substantial number of new studies have appeared. Secondly, Aletras indicated the range where economies of scale apply. The results were based on studies that met certain quality criteria. However, his study lacked a formal analytical tool to obtain results. With an increased number of studies on the cost structure of hospitals, it is possible to apply meta-analysis, which has become a popular tool to distil general results from a collection of studies. Although the present study did not fully apply the techniques used in meta-

analyses, it borrowed some important features of meta-analyses, including regression analysis on the results of several studies.

Meta-analysis can only proceed if there is a measure – the effect size – that is shared among studies. However, the type of results on scale reported by parametric studies and non-parametric studies differs. There is no shared effect size for scale. Therefore, the analysis for parametric and non-parametric studies were performed separately. Furthermore, parametric studies far more frequently report results that can be used to derive the optimum scale. However, these results need some additional computation. We therefore paid more attention to parametric studies.

The outline of this chapter is as follows. Section 2 addresses the measurement of economies of scale for parametric studies. Section 3 describes the database of parametric studies, presents the criteria on which studies were included in the analysis and gives the descriptive statistics of the studies that were used in the analysis. Section 4 presents the results of a regression analysis. Section 5 concentrates on non-parametric studies and the results on the optimum size for hospitals that can be derived from these studies. Section 6 concludes the chapter.

2.2 Measuring economies of scale

Meta-analyses require the use of an effect size that is shared among studies. Ideally, the effect size for this study would be the optimum size for hospitals. However, there are only a few studies that report the optimum size. Furthermore, there is a difference between parametric studies and non-parametric studies. The latter rarely report results on the optimum size, but if they do, the optimum size or a range for the optimum size is reported directly. Parametric studies more frequently report results on the scale. In general, this is the scale elasticity for an average hospital. The combination of scale elasticity

and the scale where the elasticity applies is a decent alternative for the effect size. It is less rare and still reveals something about the optimum size.

The scale elasticity measures the proportionate increase in outputs resulting from an increase in inputs by a given proportion. Depending on the functional form, the scale elasticity usually depends on the scale. Parametric studies often report the scale elasticity at the sample mean, indicating whether an average hospital is too small, too big or the optimum size. Besides reporting the scale elasticity at the sample mean, a few studies also report the scale elasticity at other points in the sample (i.e. first and third quartiles).

The use of the scale elasticity in combination with the scale implies that we have to operationalize the measurement of scale. There are several possibilities such as production, admissions, turnover, costs and number of beds. Of these options, number of beds is a practical measure when comparing studies. It is included in almost all studies and compares well between different studies (no deflation or currency issues, no aggregation issues and only minor definition issues). The downside of beds as a measure of scale is that it might introduce some confusion, since beds are also a popular measure of capital. Capital is assumed to be fixed in the short run and most probably not at its cost minimising level. Moreover, capital set at its cost minimising level is not a sufficient condition for an optimal scale of production. Therefore, it should be kept in mind that the number of beds is used here as a proxy for size of production (flexible in the short run) and not as a measure of capital (fixed in the short run). Although other measures might do more justice to the measurement of scale and might be less confusing, these measures are less practical.

Intuitively, there is a high correlation between beds and admissions. In the case of equal occupancy rates and equal average length of stay, the number of beds are a linear function of the number of admissions. In practice, differences between occupancy rates and average length of stay can be expected. Differences in the average length of stay might be a result of case mix

differences. In that case, the number of beds absorbs this difference in production. Another factor that can blur the proxy is the number of outpatients in relation to the number of admissions. But for outpatients there is also a high correlation with the number of admissions. It is therefore reasonable to assume that the number of beds proxies the size of production rather well.

The scale elasticity arose here as the main variable of analysis. In the remainder of this section, we discuss the derivation of the scale elasticity. What follows here is based on Brautigam and Daughety (1983), Vita (1990) and Smet (2002). For parametric studies the most popular variant is estimating a cost function; we therefore concentrate on the cost function. Overall economies of scale are measured with the concept of ray economies of scale, as defined by Baumol et al. (1982). Ray scale economies are measured as the elasticity of cost along a ray of output emanating from the origin holding output bundles fixed.

One of the main issues in estimating the economies of scale is whether all inputs are freely adjustable. The question is: are hospitals in the short run able to adjust all inputs, including capital, to their cost minimising long-run equilibrium, or are they in the short run able to adjust only some of their inputs (e.g. only the variable inputs)? If freely adjustable inputs is a valid assumption, it is appropriate to estimate a long-run cost function. It should be noted, however, that estimating the long-run cost function also requires the availability of price data on all inputs. If on the other hand for some inputs it takes more time to adjust, which in the case of capital is not unusual, it is more appropriate to estimate a short-run cost function.

For a long-run cost function, for which all inputs can be adjusted freely, ray economies of scale can be expressed as (Baumol et al., 1982):

$$S = \frac{C(y, \mathrm{w})}{\sum_{i=1}^{n} y_i \frac{\partial C(y, \mathrm{w})}{\partial y_i}} = \frac{1}{\sum_{i=1}^{n} \varepsilon_{Cy_i}} \tag{1}$$

With:

$C(y, w)$ = long-run cost function (total costs);

y = output (vector);

y_i = output i;

w = input prices (vector);

ε_{Cy_i} = output cost elasticity of output i.

However, as stated, most of the time capital cannot be freely varied, either because of the nature of capital, which in essence is long run, or because of regulations that prevent hospitals from freely adjusting the amount of capital. Therefore, instead of analysing a long-run cost function, short-run cost functions are often analysed. A short-run cost function distinguishes a variable and a fixed part. The variable part is represented by a variable cost-function that has to be estimated. The fixed part are the fixed costs represented as the product of price and volume of fixed inputs:

$$C^s(y, w^v, w^f, F) = C^v(y, w^v, F) + \sum_{i=1}^{k} w_i^f F_i \tag{2}$$

With:

$C^s(y, w^v, w^f, F)$ = short-run cost function (total costs);

$C^v(y, w^v, F)$ = variable cost function (variable costs);

y = output (vector);

w^v = input prices of variable inputs (vector);

w^f = input prices of fixed inputs (vector);

F_i = fixed inputs.

If the production structure of hospitals is analysed with a short-run cost function, there are two methods to assess the scale elasticity. Both methods were discussed and compared by Brautigam and Daughety (1983). The two methods differ in the measurement of the fixed inputs. The appealing approach from a theoretical point of view is to apply the envelope condition to find the optimal amount of fixed inputs, that is, deriving the long-run optimum by differentiating the short-run function with respect to its fixed factors and equating to zero:

$$\frac{\partial\, C^v(y, w^v, F)}{\partial F_i} + w_i^f = 0 \qquad \forall\ i = 1, .., k \tag{3}$$

This implies that at the point of long-run cost minimisation, the following applies: the variable cost saved by substituting the last unit of a fixed input for variable inputs equals the marginal input cost of that unit of fixed input. Note that, besides the estimates of the variable cost-function, for this method it is necessary to have data on the input prices of the fixed inputs. The exercise is done so that the economies of scale are estimated for an optimum amount of fixed inputs, ensuring that the economies of scale relate to the efficient expansion path. For ray scale economies we get (Brautigam & Daughety, 1983):

$$S = \frac{1 - \sum_{i=1}^{k} \frac{\partial \ln C^v(y, w^v, F)}{\partial \ln F_i}\Big|_{F=F^*}}{\sum_{i=1}^{n} \frac{\partial \ln C^v(y, w^v, F)}{\partial \ln y_i}\Big|_{F=F^*}} = \frac{1 - \sum_{i=1}^{k} \varepsilon_{CX_i}}{\sum_{i=1}^{n} \varepsilon_{Cy_i}}\Big|_{F=F^*} \tag{4}$$

With F^* = the optimal amount of F.

The second method was used by Caves et al. (1981), and employs the actual amount of fixed input instead of the optimum level of fixed inputs. Rather than evaluating economies along the efficient expansion path, they are evaluated along a ray from the origin that passes through an actual point of

operation (e.g. an observation, the sample mean). The advantage of this method is that no data on the prices of the fixed inputs are needed.

As pointed out, the difference between the two methods is the measurement of the fixed output. The first method uses the optimal amount of fixed inputs for F^*, whereas the second uses the actual amount of fixed inputs of each observation or the sample mean of the fixed inputs for F^*. Since the use of the short-run cost function is motivated by the belief that hospitals are not operating on their efficient expansion path, it is expected that the evaluation points used by the two methods coincide only very rarely. Furthermore, Brautigam and Daughety (1983) showed that both methods yield different estimates of the economies of scale. Only in the case of a homothetic cost function are the estimated economies of scale equal. Moreover, in general it is unknown how the estimates of the two methods relate to each other. It is, however, possible to show how the scale elasticity varies with the level of fixed inputs. Brautigam and Daughety (1983) demonstrated this by differentiating the scale economies with respect to the log of the fixed inputs.

Which method is appropriate depends on its use and on beliefs about the flexibility of hospitals to adjust towards the efficient expansion path. If adjustments towards that path are fairly rapid, hospitals should of course use the first method (optimal amount of fixed input is used). If adjustment is slow, however, or there is hardly any adjustment, for example due to regulatory constraints, the second method might be more appropriate (average amount or individual amount of fixed input is used). In practice, the second method might prevail because of a lack of data on the prices of the fixed inputs.

Finally, there are studies that estimated a short-run cost function and used formula (1) to estimate the scale economies. However, this results not in a true measure of scale economies, but in a measure that captures the effect on outputs of varying the variable inputs at a certain level of fixed inputs. In other words, it analyses the short run instead of the long run (Vita, 1990).

Ray economies of scale indicate how much proportional increase in outputs would result from a proportional increase in inputs. Therefore, ray economies of scale greater than 1 (S>1) indicate that there are increasing economies of scale: outputs increase faster than inputs. The opposite is valid for ray economies of scale less than 1 (S<1): there are diseconomies of scale and inputs increase faster than outputs. And last but not least, there is the situation in which the economies of scale equals 1 (S=1). In that case, inputs and outputs increase and decrease at the same pace.

2.3 A database on economies of scale for hospitals

To investigate the optimum scale of hospitals by meta-analysis, a literature database was constructed. This section describes the process of selecting literature and the variables included in the database.

Literature included

As argued in the previous section, this study used a combination of scale elasticity and the scale where the scale elasticity applies, with the scale measured in terms of the number of beds. Therefore, only studies that report the scale elasticity, or for which it is possible to derive the scale elasticity, were included. A consequence of the use of scale elasticity is that the non-parametric studies were excluded, since there are hardly any non-parametric studies that report findings on the scale elasticity. The results that can be derived from non-parametric studies are therefore discussed in a separate section.

An important feature of meta-analysis is assigning weights to individual studies. Rather than compute an unweighted effect from the included studies, a weighted effect assigns more weight to studies that are more precise. In practice, the reliability of results will vary across studies. In general, results from studies with a big sample will be more precise than studies with a small sample. It is most common in meta-analysis to use the inverse of the variance of the effect size as weight (inverse variance weighting). The inverse variance is

roughly proportional to sample size, and is a more nuanced measure than weighting with the sample size (Borenstein et al., 2007).

If the scale elasticity is not reported, it is sometimes possible to derive it from the parameter estimates of the cost function and the sample characteristics. It is far less common to report the variance of the scale elasticity or its standard error. Therefore, for most studies the variance of the individual output cost elasticities are used to compute the variance for the scale elasticity. For this computation, the delta method is applied (see for example Greene, 2003).

The database of literature was built up by carrying out an extensive search of the Web of Science and PubMed using the key words 'scale economies hospitals', 'hospital cost function scale', 'hospital production function scale' and 'hospital frontier'. A search for additional literature included the use of Google Scholar, references from earlier studies and cross references from included literature. The results of the search were first judged as being relevant or not, namely whether the study was an empirical study about the cost structure or production structure of hospitals. At the second stage, the studies were systematically reviewed. The following were the inclusion criteria:

1. The study uses a parametric framework to estimate the cost or production structure for hospitals;
2. The study reports the scale elasticity or present results that can be used to derive the scale elasticity;
3. The study reports the variance of the scale elasticity or it is possible to derive the variance;
4. The study includes data on the size of the hospitals, i.e. data on the average number of beds is included or it is possible to get a fair estimate of the average number of beds;
5. The study was published after 1990 and is written in English.

The final database of literature comprises 41 relevant studies. Since some studies present more than one result on economies of scale, the 41 studies generated 95 observations. There are several reasons for studies to present multiple results on economies of scale. There are studies that included economies of scale at several points in the sample. For example, Sinay and Campbell (1995), Sinay (1998a), Sinay (1998b) and Carey et al. (2015) also included the scale elasticity for the first and third quartiles. Secondly, some studies split the sample into subsamples and analysed the latter. For example, Sinay and Campbell (1995), Sinay (1998a) and Sinay (1998b) analysed a control group and a group of mergers, and Carey et al. (2015) performed additional analyses for for-profit hospitals. Other studies derived results for each year in the sample (each year is a subsample). The differences between the results for each year were considered trivial and included as one observation. Finally, there are studies that analysed several specifications. Carey (2000), Blank and Merkies (2004), Kojima (2004), Farsi and Filippini (2006) and Carey and Stefos (2011) are examples of studies that presented results from different specifications, estimation methods and output measurements.

The appendix of this chapter gives a full overview of the studies included in the database. The overview also shows how many results per study are included in the database.

Variables included

The first two variables included are, of course, the reported scale elasticity and the average number of beds. We also wanted to include variables that characterise a study, since modelling a cost function includes a wide variety of choices. In the study by Nguyen and Coelli (2009), the authors analysed and discussed the impact of choices on the efficiency scores for hospitals. In line with that study, we were faced with a variety of modelling choices that might have an impact on the scale elasticity, but were not sure how.

First of all, we looked at the estimation aspects. There are several aspects that relate to the method of estimation, such as the choice between a frontier

or a non-frontier estimation procedure (ordinary least squares, maximum likelihood, panel data techniques), including cost shares and assumptions on the error term. It is hard to come up in advance with valid hypotheses about how these factors influence the estimates of the scale effects. Furthermore, some studies do not provide all details of the estimation. However, we made an exception for the frontier cost function and non-frontier cost function. This study characteristic was included in the database. Not only is it a characteristic that is easily identified, but there might also be some logic, since some studies use the same explanatory variables for efficiency as economies of scale. If this is true (i.e. efficiency and economies of scale are influenced by the same source), we might expect that a frontier study will result in lower estimates of the economies of scale.

Secondly, there is the aspect of specification. When it comes to specification there is a list of usual suspects, such as translog, generalised translog, Cobb–Douglas and cubic. In addition, there are also some specifications that are used incidentally, that is, log-linear, Leontief and quadratic. It is not clear whether the model specification might have an impact.

Then there is the method that was used to derive the scale elasticity (see discussion in 2.2). In general, we have the difference between the estimation of the long-run cost function and the estimation of the short-run cost function. In the case of the short-run cost function, there is an additional decision whether to evaluate at the optimal or the actual level of fixed inputs. Although this choice is quite relevant and can have major implications, in practice the most common method is the actual level of capital to derive scale effects from the short-run cost function. There are hardly any good examples of use of the envelope condition to derive the long-run cost function from the short-run cost function. Therefore, we only distinguished the difference between direct estimation of the long-run cost function and the use of the short-run cost function.

Next, there were the details of the model, such as the measurement of inputs and outputs, accounting for differences in case mix and environmental differences. Almost all studies use admissions and outpatients as outputs. Besides these two outputs, it is not uncommon for additional outputs to be included, for example patient days or more detailed admission categories like surgery versus non-surgery. Related to this last example is the way a study accounts for case mix differences. In general, there are three possibilities: not accounting for case mix, the use of case mix weighted admissions and including a separate case mix index in the model. In our analysis, we included the number of outputs and the way the model accounts for case mix.

In addition to case mix, environmental differences also have an impact. One variable frequently used to control for this is the teaching status of a hospital; another common control variable is the characteristics of the patient population. We included a variable that indicates whether the study accounts for environmental differences. This variable is quite comprehensive, since many different variants have been used to account for environmental differences.

We also included the number of inputs. It is most common to use two inputs: labour and materials. There are, however, studies that included more inputs. Aletras (1997) emphasised the disaggregation of inputs: more disaggregation implies less bias in the estimates of economies. In addition to the number of inputs, we had a special interest in the measurement of capital. For the long-run cost function, the price of capital should be included otherwise it has to be clear that it is reasonable to assume that there is no price variation for capital. For the short-run cost function, the volume of the fixed input (capital) should be included.

Finally, we included variables that clarify the data used. Besides the average number of beds, we included the standard deviation for the number of beds. We also included the sample size, number of years analysed, sample year and publication date of the study. Finally, we distinguished studies that focused on

the USA and those that focused on other countries. Table 2-1summarises the variables and gives the descriptive statistics.

Table 2-1 Descriptive statistics parametric studies

	Studies (N=41)		Observations (N=95)	
Variable	Mean	Standard Deviation	Mean	Standard Deviation
Scale elasticity	1.10	0.20	1.10	0.27
Long-run cost function	0.34		0.22	
Stochastic frontier	0.39		0.25	
Specification				
Translog	0.41		0.30	
Generalised translog	0.12		0.29	
Cobb–Douglas	0.24		0.18	
Cubic & quadratic	0.12		0.14	
Log linear	0.05		0.06	
Other specification	0.05		0.02	
# inputs	2.5	2.2	2.4	2.0
# outputs	3.3	1.7	3.2	1.4
Case mix correction				
Case mix index	0.56		0.67	
Weighted output	0.29		0.27	
No case mix correction	0.15		0.06	
No environment	0.24		0.19	
USA	0.41		0.55	
Sample year	1996	6.45	1995	6.52
# beds	240	94	225	98
S.d. #beds[a]	216	74	197	66
Sample size	1,505	4,087	1,069	2,822

a, N=23 for studies and N=40 for observations

The average scale elasticity found in studies is 1.1, suggesting that the average study found economies of scale for the average hospital. The short-run cost function was more popular than the long-run cost function. This indicates that there is more support for the assumption that hospitals do not operate at their long-run equilibrium. About a third of the studies applied stochastic frontier analysis; this share drops as we look at the level of observations,

suggesting that studies that applied stochastic frontier analysis more often reported a single result. This is not surprising, since in general the focus of stochastic frontier analysis is, of course, on the efficiency. The well-known translog specification seems to be the most popular specification for researchers. More than 85% of the studies accounted for case mix; the most popular way to do so was to include a case mix index. The average number of beds for the average hospital is 240.

This study relates scale elasticity and scale operationalized by the number of beds. Figure 2-1 gives an idea of the correlation between the two variables. The figure also distinguishes the main results found in studies and the additional results that come from analysis at the quartiles, different model specifications and use of subsamples.

Figure 2-1 Number of beds and scale elasticity found in studies

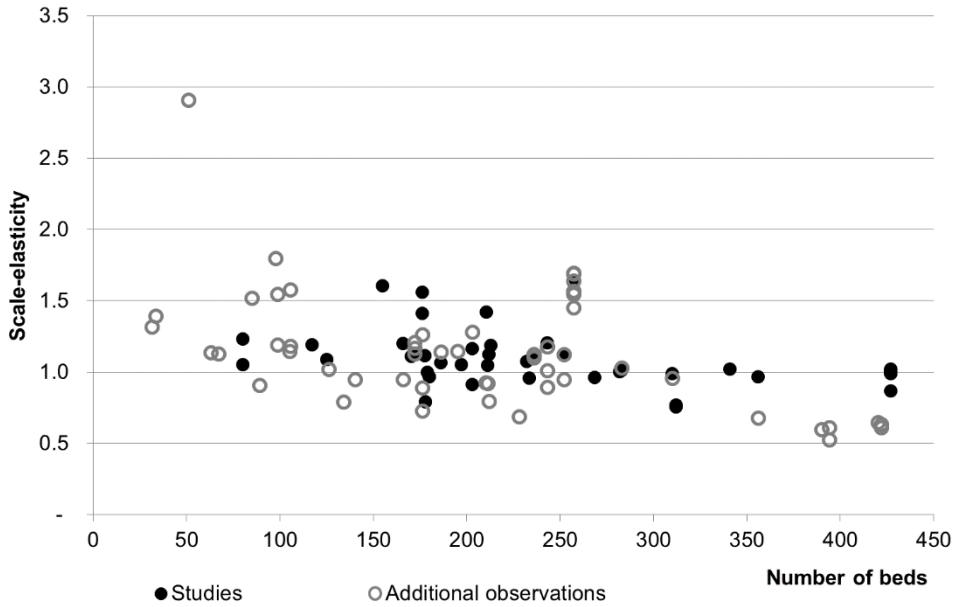


Figure 2-1 shows that an increased number of beds is indeed accompanied by a decrease in scale elasticity. For hospitals with up to 150 beds, almost all studies found economies of scale; there are exceptions merely for additional observations. For hospitals with more than 150 beds, the results are mixed: both economies of scale and diseconomies of scale were found. With an increased number of beds, results tend more towards diseconomies of scale, until finally above 300 beds there is hardly any study that still found economies of scale.

2.4 Regression model

The scale elasticity was regressed on explanatory variables with a weighted multivariate regression. The only real continuous variables in the analysis were the number of beds and the sample size. It was very well possible that the effects of the number of beds and the sample size would not be linear, but rather diminishing as they increase. We therefore also tested models that included the logarithm of these variables instead of a linear term. For the sample size, the logarithm is preferred; for the number of beds, linear is the best option. We originally included the standard deviations of the number of beds in the sample. Since we found no significant results for this variable and including it entails a loss of observations, the variable was omitted. Furthermore, Figure 2-1 shows one outlier at the level of observations (scale elasticity of 2.9 for hospitals with 50 beds). This observation is an additional observation of a study that also presents results at quartiles (in this case, the first quartile). The observation was therefore omitted from the dataset.

In meta-regression analyses it is common practice to weight the observations. Construction of the weights depends on the choice between a fixed effects model and a random effects model. The two methods differ in assumptions concerning the precision of the studies incorporated in the analysis. The fixed effects model is the simplest of the two methods and assumes that there is one true effect size that is shared by all observations.

Therefore, the only source of error is the random error within studies. With a large enough sample size, the error will tend towards zero. For fixed effect, the inverse of the effect size variance is commonly used as weight, such that larger, more precise studies tend to contribute more than smaller studies to the weighted average.

The random effects model assumes that there is a distribution of true effects and estimates the mean of this distribution. In this case, large studies might still be more precise than small studies, but each study estimates a different effect size and all these effect sizes should be included in an estimate of the mean. Therefore, the random effects model applies another, more moderate weighting scheme. For random effects there are two levels of sampling and two levels of error. First, each study is used to estimate the true effect in a specific population; second, all true effects are used to estimate the mean of the distribution of the true effects. Besides within variation, the weight assigned to each study also incorporates between-studies variance. The estimates for a random effects model are derived from either the method of moments or a maximum likelihood approach (Raudenbush, 1994). Alternatively, it is possible to apply a weighted regression with weights: $1/(\mathrm{Var}_i + \hat{\tau}^2)$, where $\hat{\tau}^2$ is the random-effects variance and accounts for the between variation. From these weights it immediately becomes clear that when $\hat{\tau}^2$ increases, the weighting flattens out. That is, with an increased between-studies variance, large studies become less dominant and small studies become less trivial.

The choice between fixed effects and random effects can be formally tested. However, it is advocated not to decide on a formal statistic, but to use prior information about the existence of between variation (Borenstein et al., 2007). Besides, if there is little between variation, $\hat{\tau}^2$ will be small, such that random effects will yield the same estimates as fixed effects. Because we expected a lot of between variation here, random effects were the most appropriate.

Table 2-2 presents the results of the regression analysis.

Table 2-2 Regression results for the scale elasticity of hospitals

	Studies		Observations	
Variable	Estimate	Standard error	Estimate	Standard error
Constant	1.111 ***	0.199	1.338 ***	0.162
Long-run cost function	0.119 *	0.065	0.092 *	0.054
Stochastic frontier	-0.078	0.057	-0.164 **	0.047
Translog (reference)				
Generalised translog	-0.090	0.112	-0.197 **	0.097
Cobb–Douglas	-0.005	0.077	-0.039	0.057
Cubic & quadratic	0.153	0.125	-0.106	0.089
Log linear	0.077	0.058	0.098 *	0.070
Other specification	-0.008	0.253	-0.188	0.182
# inputs	0.041 *	0.021	0.052 **	0.016
# outputs	-0.005	0.021	0.004	0.020
Case mix index (reference)				
Weighted output	0.076	0.059	0.151 **	0.054
No case mix correction	-0.091	0.089	-0.087	0.111
No environment	-0.001	0.058	0.062	0.052
USA	0.007	0.090	0.014	0.080
Sample year (minus 1990)	0.001	0.006	0.001	0.005
Beds (per 100)	-0.096 ***	0.030	-0.170 ***	0.020
Ln (Sample size)	0.015	0.024	0.006	0.025
R2 / adjusted R2	.58	.31	.59	.50

***= significant at 1%, **=significant at 5%, *=significant at 10%

It can be concluded from Table 2-2 that the explanatory variables explain a reasonable part of the variation in scale elasticity. Furthermore, we see that the number of parameters included is a little overdone; this is especially the case for the analyses at the study level, for which the adjusted R^2 is 0.31 compared to 0.58 for the unadjusted R^2. As mentioned, we also estimated the models where the logarithm of the number of beds was included (instead of linear). In general, the results for the parameter estimates and standard errors are similar, except of course for the number of beds. The fit decreases slightly, respectively

to R^2=0.56 (studies) and R^2 =0.57 (observations). A side effect of the model with the logarithm of the number of beds is that the statistics show that the probability of heteroscedasticity increases.

The method used to derive the scale elasticity seems to have an impact on the scale elasticity: estimates of the scale elasticity directly derived from the long-run cost function are 0.12 higher. Vita (1990) discussed the scale elasticity derived from the short-run cost function and argued that true scale economies (i.e. along the efficient expansion path) might be higher than scale elasticities obtained at the actual level of capital if there is overemployment of capital, since the results of the long-run cost function are true scale-economies. Therefore, the result here indicates that there is a tendency towards the overemployment of capital in the average hospital.

On average, stochastic frontier analysis finds lower scale elasticities than standard regression. The effect, however, is not significant at the study level.

Specification can have an impact on the scale elasticity. In the regression, the translog specification is the reference group. This implies that the results in Table 2-2, including tests on significance, are relative to the translog specification. With this in mind, we can roughly conclude that on average the generalised translog specification has a considerable lower scale elasticity. At the other end of the spectrum, we have the log linear specification with on average higher scale elasticities. Besides an effect on the estimated value of scale elasticities, it is also possible to conclude about the validity of certain specifications. As we discuss later, the scale elasticity varies with the size of the hospital; this subverts the validity of a specification that assumes a constant scale elasticity (i.e. Cobb–Douglas).

Aletras (1997) discussed the importance of disaggregating the inputs when it comes to computing scale effects. Here, we found that the number of inputs has a significant impact on the scale elasticity: each additional input included in the analysis increased the scale elasticity by 0.04. For the number of outputs,

there is no significant effect. There seems to be an effect for the use of weighted output compared to the use of a case mix. However, the result is only significant for the analysis on the observations.

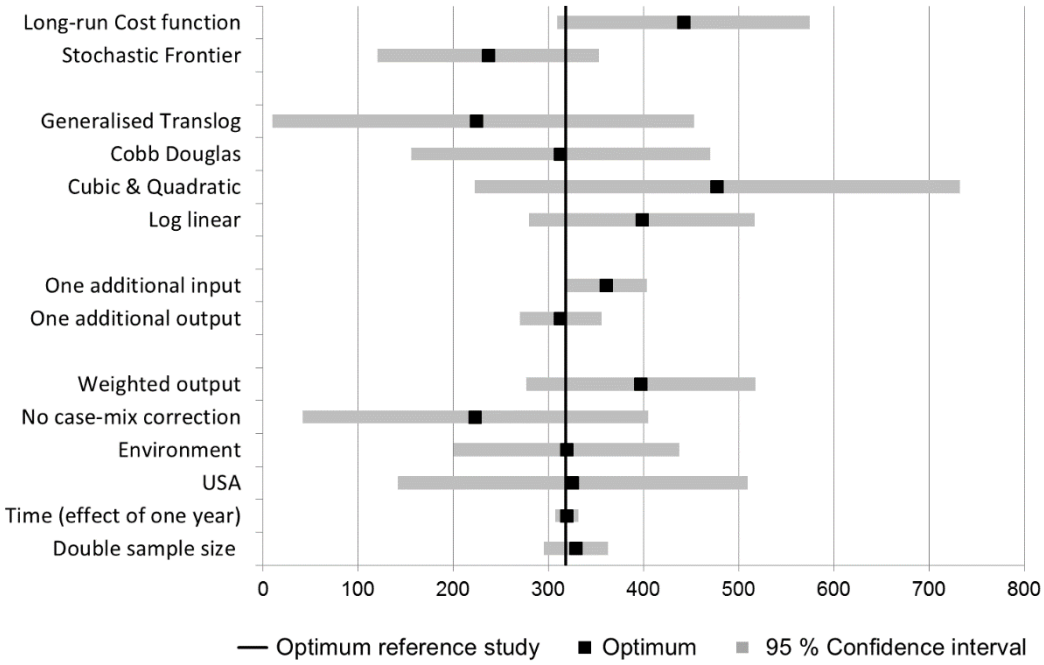
As discussed earlier, we had a special interest in the measurement of capital as input. When we examined the literature we had collected, we found that there was little variation in the measurement of capital or that the measurement was rather specific. In general, the studies that applied a long-run cost function included the price of capital, or it was not included assuming no variation for the price of capital between hospitals. For short-run cost functions, most studies (two out of three) used the number of beds as a proxy for capital. Other studies included depreciation, fixed assets or an index (these measures might also include the number of beds). We tested a model that included the measurement of capital in which we distinguish between the measurement of capital with beds and other measures, but found no significant result.

The parameter estimates for the number of beds are clearly significant. Since we modelled a linear effect, the parameter estimates represent the decrease in scale elasticity for an additional 100 beds. For the analysis based on studies, the scale elasticity decreases by -0.1 per 100 additional beds. For observations, the effect of 100 additional beds is a scale elasticity decrease of -0.17. The parameter estimate for the effect of beds can be used to calculate the optimum scale.

It is rather easy to compute the optimal scale, namely the number of beds for which the scale elasticity equals 1. In order to calculate the optimum it is necessary to have values for the independent variables of the regression analysis. Therefore, we had to specify a reference study for which the optimum applies. Here, we took the most common value for the binary variables, that is: short-run, no frontier, translog, case mix index, non-USA. For the other variables, we took the median value (number of inputs=3, number of outputs=3, logarithm of the sample size=5.4, and sample year=1997). For this reference study, the estimated optimum size is 320 beds.

The reference function is, of course, an arbitrary construct. Instead of most common and median values, we could have used average values, in which case the optimum is 329 beds. The use of alternative values demonstrated that an optimum applies for a set of characteristics. Therefore, the reference study was compared with alternatives. For alternative reference functions, we took the opposite value for binary variables, continuous linear variables were increased by one unit and the sample size was doubled. Here, we calculated the effect for each characteristic, given that all other characteristics remain the same. Figure 2-2 show the optimum and a 95% confidence interval for alternative values.

Figure 2-2 Optimum size (in number of beds) for each characteristic ceteris paribus



The results in Figure 2-2 are, of course, in line with the discussion of the regression results. The most striking aspect of Figure 2-2 is that for all characteristics, the 95% confidence interval contains the optimum of the

reference study. This is, of course, not unexpected, since none of the characteristics in Table 2-2 is significant at the 5% level. In addition to significance issues, there are some other interesting results. The use of a long-run cost function shifts the optimum to 444 beds, whereas the use of a frontier cost function shrinks the optimum to 239 beds. For the specification characteristics, the confidence intervals are wide and the effect of specification is diverse. For the generalised translog, the optimum decreases to 226, whereas for the cubic and quadratic specification the optimum increases to 480 beds. An additional input increases the optimum by 43 beds. Furthermore, the use of weighted output instead of a case mix index increases the optimum by 70 beds, while not accounting for case mix decreases the optimum by 95 beds. All other characteristics have marginal effects on the optimum scale.

The previously described results are based on the study level; the results at the level of observations are similar: the optimum is 321 beds for the reference study, which is almost identical to the optimum at the level of studies. In the case of an average study, the optimum differs: it is 280 beds at the level of observations. Sensitivity analysis on the effect of the study characteristics generally shows the same picture as Figure 2-2, although the confidence intervals are tighter.

Thus, based on parametric studies it can be concluded that the optimum scale for hospitals is around 300 beds. There are hardly any studies that found economies of scale above 300 beds. Furthermore, based on the results of regression analysis, an optimum can be calculated. For a reference study, based on the most common characteristics, the optimum size is 320 beds. However, it is hard to really pinpoint the optimum scale since the optimum scale depends on several factors, including study characteristics.

2.5 Non-parametric studies

This section discusses the results on the optimal scale that can be derived from non-parametric studies. Non-parametric studies generally do not present a scale elasticity; instead, other type of scale results are normally reported. A data envelopment analysis (DEA) study typically reports the average scale efficiency of the sample. Scale efficiency is a measure that indicates how much efficiency can be gained by producing at the optimal scale. A value of 1 for the scale efficiency implies that a hospital is fully scale efficient and that it operates at the optimal scale. Besides the scale efficiency, studies often also report the number of hospitals that have an optimal size; these are known as hospitals that produce at constant returns to scale (CRS). Hospitals that are too small have increasing returns to scale (IRS) and hospitals that are too big have decreasing returns to scale (DRS). Some studies also report the number of hospitals with IRS and DRS.

A literature search on DEA studies resulted in 102 studies that had examined the hospital sector since 1990. Table 2-3 gives an overview of the type of scale results that these studies report. A handful of studies (9%) included results on the optimum scale or reported a range for the optimum scale. Other results on scale, such as average scale efficiency and number of scale efficient hospitals, are reported far more frequently. The number of hospitals that are too small or too big are less frequently reported. Determining the number of hospitals that are either too big or too small requires some additional computation, which is probably why these results are far less frequently reported.

Table 2-3 Scale results reported in DEA studies (N=102)

Reported result	% studies
Optimal scale / range for optimal scale	9%
Average scale efficiency	60%
Number of hospitals with an optimal scale (CRS)	50%
Number of too small (IRS) and too big (DRS) hospitals	25%

Studies that derived the optimal scale are scarce. Ferrier and Valdmanis (1996) used the scale-efficiency of individual hospitals to estimate the optimum scale. The scale efficiency was regressed on the number of beds and the number of beds squared. The optimum scale was found by equating the resulting parabola to 1. The study found an optimal scale of 95 beds for rural hospitals. Butler and Li (2005) presented the average range for the most productive scale size (MPSS). The MPSS, introduced by Banker (1984), is the size where constant returns to scale applies for a specific input–output mix. In line with the results of Ferrier and Valdmanis, Butler and Li found a mean MPSS of 95 beds for rural hospitals. The average MPSS was also presented by Webster et al. (1998) for Australian private acute hospitals, which is 22 beds.

Wilson and Carey (2004) developed a method that incorporates ray scale economies in non-parametric studies. A bootstrap method is used to provide inferences regarding ray scale economies and expansion path scale economies. Wilson and Carey's conclusion was that studies that apply the parametric method, conclude too early that diseconomies of scale may occur due to possible misspecifications. They found evidence of increasing returns to scale among hospitals above the median size, extending to the largest decile in terms of size. These results, however, are not pure ray scale economies, but also include expansion path scale economies, resulting from a changed output mix. Furthermore, the study divided the sample into subsamples to reduce heterogeneity. From the results we can conclude that optimum scale depends on the characteristics of the hospitals in the sample, that is, the optimum scale for teaching hospitals is higher than for non-teaching hospitals.

Besides studies that made a point estimate of the optimum scale, some studies reported a range for the optimal scale. Linna and Häkkinen (1999) found that for Finnish hospitals, scale efficiency is maximised in the range of 40 to 250 beds. It is, however, not clear how this range was derived; it is probably the range where constant returns to scale apply. McKillop et al. (1999) found a range of 222 to 358 beds for scale-efficient Irish hospitals. Lee

et al. (2009) concluded that hospitals in Florida with fewer than 250 beds can profit from economies of scale; however, it seems that the authors related technical efficiency to scale, and did not use scale efficiency to derive their results. Bilsel and Davutyan (2014) computed the average scale efficiency for several ranges and concluded that the scale efficiency is the highest in the range of 100 to 150 beds. This result was further supported by a test on constant returns to scale that cannot be rejected for the range of 100 to 150 beds. Czypionka et al. (2014) applied a similar method for Austrian hospitals and found a range of 200 to 300 beds for a model that included both inpatients and outpatients, and a range of 300 to 400 beds that included only inpatients. Aletras (1997) indicated that Byrnes and Valdmanis (1994) used the MPSS to find a range of 220 to 260 beds.

Because there is abundant information about the average scale efficiency and the number of hospitals with an optimal scale, and to lesser extent the number of hospitals that are either too small or too big, we wondered whether it would be possible to use this information to get an insight into the optimum scale of hospitals.

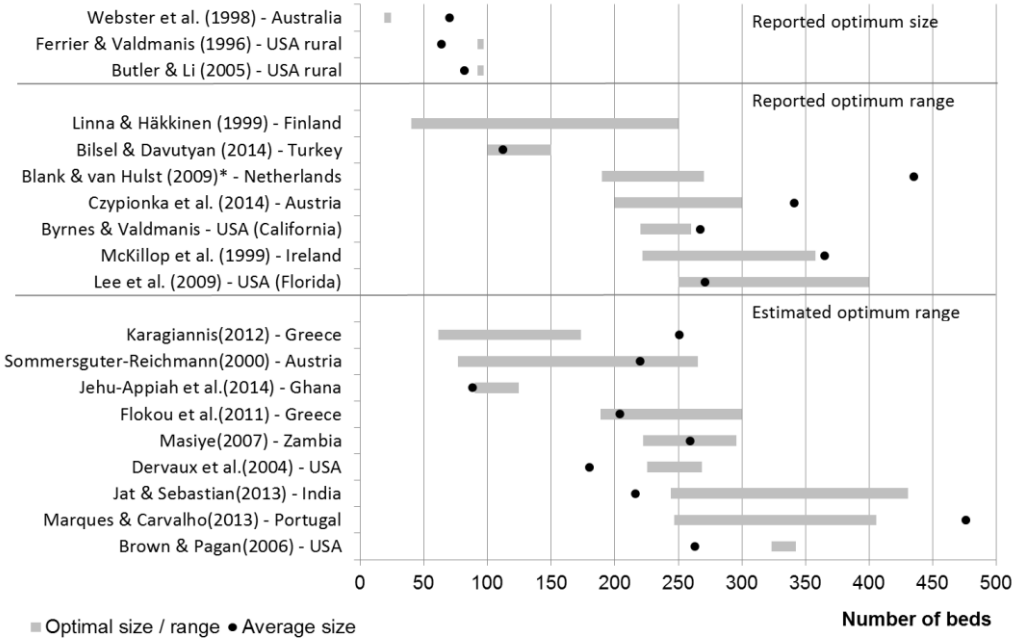
The average scale efficiency is not suitable. This is because average scale efficiency is an average of hospitals with either IRS or DRS. So we do not know how much inefficiency is due to hospitals that have IRS and how much inefficiency is due to hospitals that have DRS. Or, to put it differently, the average scale efficiency is something different from the scale efficiency of the average hospital.

The information about the number of hospitals with IRS, CRS and DRS does not indicate the optimal scale. By making some additional assumptions, however, this information could be used to get some insight into the range of the optimal scale. Basically, we wanted to get an estimate of the range where CRS begins and ends. For that purpose, we first assumed that IRS, CRS and DRS hospitals are neatly ordered, such that IRS hospitals are smaller than CRS hospitals and CRS hospitals are smaller than DRS hospitals. Note that in

practice this will seldom be true: there will normally be an overlap between the range of IRS and CRS, CRS and DRS, and possibly also IRS and DRS hospitals. Second, we assumed that in each study hospital sizes are normal distributed. The decision to use the normal distribution was arbitrary; it is, of course, very well possible that another distribution applies, for example a log-normal distribution of hospital sizes. If a study reports the average size and its standard deviation, it is fairly easy to calculate a range for the optimum scale (the range where CRS applies). From the first assumption it is known at which percentile CRS starts and ends, and with the second assumption we could derive a range by applying these percentiles to the assumed distribution.

The estimated optimum range varies across studies (see Figure 2-3, estimated optimum range). For six out of nine studies, the CRS range stays under 300 beds. There are three studies for which the CRS range stretches beyond 300 beds. For these studies the middle of the CRS ranges lies at 330 beds. The results are of course sensitive for both assumptions. For the assumption on the strict order of IRS, CRS and DRS, it is clear that the estimated CRS is too tight. Most probably for each study there will be CRS hospitals that lie outside the estimated CRS range. The distributional assumption was checked by repeating the exercise assuming a log-normal distribution for hospital sizes. Especially in the tails of the normal distribution and the log-normal distribution, there are differences. Therefore, different results are found for studies with relative low percentages of IRS and DRS; especially for studies that were based on only a few observations, significant shifts of the boundaries occur. As a consequence, the presented results for Jat and Sebastian (2013) and Flokou et al. (2011) should be treated with some caution.

Figure 2-3 Optimal size and average size



* = the range for the optimal scale concerns additional calculations.

Figure 2-3 summarises the DEA results on the optimal scale. The figure shows the optimum scale or range for the optimum size (either reported or estimated); it also shows the average size of the sample for each study. There is a wide variation for the optimum scale: it varies from 22 beds to a range that stretches to 430 beds. The upper bound is consistent with the overview study by Aletras (1997), who found a range for DEA studies of between 220 and 400 beds. The lower bound differs from the lower bound found by Aletras. It should be noted that Aletras's overview was limited to five DEA studies. At the time Aletras conducted his research, there were no empirical results that reported an optimum scale with fewer than 220 beds.

Since the bounds are extreme cases, both the lower and the upper bound can be narrowed down. For instance, the lower bound was found in only one study; without this anomaly, a lower bound of approximately 100 beds is

defendable. In the case of the upper bound, only a few of studies found an optimum range that includes hospitals with 400 or more beds. Besides that, it should be noted that for these studies the reported range is quite broad (over 200 beds wide).

With only 19 studies it is a hazardous exercise to calculate an optimum. At the same time, it is interesting that the optimum scale is the ideal effect size. For some studies, the optimum scale was a point estimation; for other studies, we have a range for the optimum scale. For the purpose of calculating an average optimum scale, we used the midpoint of the latter studies. In analogy with the study on the parametric studies, we used a more sophisticated average by weighting the individual studies. The individual studies were weighted with the square root of the sample size used in the study. The weighted average optimum scale for the 19 studies is 220 beds (the unweighted average is 208 beds).

It was possible to further analyse the factors that might have an impact on the optimum scale. Here, we used the same relevant factors used in the analysis of the parametric studies. However, there are only a few observations, which reduces the number of explanatory variables that can be used. Luckily there are some explanatory variables that are blatantly irrelevant. For example, all non-parametric studies were frontier studies and there was no ex-ante specification. Furthermore, in this case, all studies included one or more capital inputs, implying that we did not have to distinguish between long run and short run. However, the optimum scale was not derived uniformly. We therefore included a dummy variable indicating whether the optimum scale was reported or estimated from the distribution of hospital sizes. Table 2-4 shows the descriptive for the explanatory variables.

Table 2-4 Descriptive statistics non-parametric studies

Variable	Mean	Standard error
Optimum	208	96
# inputs	4.2	1.6
# outputs	3.8	1.9
Case mix	0.20	0.41
Average number of beds	223	118
Ln (sample size)	5.12	1.72
Range is estimated	0.45	0.51

Table 2-5 shows the results of an unweighted and a weighted regression (weighted with the square root of the sample size).

Table 2-5 Regression results optimum scale DEA studies

	Unweighted			Weighted		
Variable	Estimate		Standard error	Estimate		Standard error
Constant	-195.5		137.3	-236.3		144.0
# inputs	-0.5		11.9	-2.0		9.5
# outputs	18.5	*	9.8	14.3		11.3
Case mix	71.0	*	39.0	82.8	*	40.1
Average number of beds	0.72	***	0.14	0.72	***	0.15
Ln (sample size)	25.3	*	13.6	37.2	**	13.2
Range is estimated	71.3	*	37.4	68.9	*	34.8
R2 / adjusted R2	0.72		0.58	0.80		0.70

*** significant at 1%, ** significant at 5%, * significant at 10%

The only parameter that is significant at the 1% level, is the parameter estimate for the average number of beds. There is a strong correlation between the average number of beds and the optimum. The regression results indicate that for each additional bed, the optimum scale increases by 0.72 beds. In practice, things are a bit more nuanced. The average number of beds also captures differences in context. The results indicate that an optimum is sensitive to the context. Therefore, we should keep in mind that for specific hospital types, the optimum scale might very well differ from the general

optimum as calculated previously. Furthermore, one might argue that the analysed studies are not a representative sample. If we compare the average hospital size of the studies analysed with the average hospital size of the 102 studies found, it becomes clear that the average hospital size of the studies analysed is significantly smaller (the difference is 20 beds), implying that the calculated optimum is probably a lower bound.

2.6 Conclusion

In 1997, Aletras conducted an extensive literature research on the economies of scale of hospitals. Since then, we have lacked an updated review on the subject. The present study analysed both parametric and non-parametric studies to gain an insight into the optimum scale of hospitals. The results on scale reported by parametric studies are normally about the scale elasticity at the sample mean, whereas non-parametric studies usually report scale efficiency and the number of hospitals that are too small, too big or the optimal size. Incidentally, the non-parametric studies also report an estimate for the optimum scale or a range for the optimum scale. Since both types of studies report different types of results, both types of studies were analysed separately in the present study.

For the parametric studies, 41 studies were included; these were good for 95 observations on the economies of scale of hospitals. The results on economies of scale were used to perform regression analyses in which the scale elasticity is regressed on the study characteristics, including the number of beds (as a proxy for the scale) where the scale elasticity applies.

From the regression results, we conclude that specification has no significant influence on results for the scale elasticity. However, since we found that the scale elasticity varies with size, a Cobb–Douglas specification is less suitable, especially if there is a wide variation in size in the sample. Some attention has to be paid to the method used to derive the scale elasticity. In

general, the use of a long-run cost function leads to higher estimates for the scale elasticity. The same applies to an increased number of inputs included in the model.

Aletras did not pinpoint an optimum scale, but found that for flexible econometric cost functions there are constant returns to scale or even diseconomies of scale for the average hospital, with an average hospital roughly defined as one with 200–300 beds. The use of regression analysis in our study made it possible to get a point estimate of the optimal scale. However, this estimate offers a deceptive accuracy, because the calculation was based on a reference study. Nevertheless, the results offer something to hold on to as long as we bear in mind that we are dealing with a reference study. For a reference study, based on the most common characteristics, the optimum is 320 beds.

Aletras used only six studies to find a range of 220 to 400 beds for the optimum scale for non-parametric studies. For the non-parametric studies, results on the optimal scale are scarce. Although non-parametric studies are quite popular and we were able to identify over 100 studies, only 19 could be used to get information about the optimal scale. About half of these studies directly reported the optimum scale or a range for the optimum scale. With some additional assumptions for the other half, we estimated a range for the optimum scale. The midpoint of a range is used as point estimate of the optimum scale.

The optimum scale is the preferred effect size and was used to compute a weighted average for the optimal scale for non-parametric studies, resulting in an estimate for the optimum scale of 220 beds. However, there are some caveats. Regression analysis revealed that for non-parametric studies, the optimum heavily depends on the context. Furthermore, the sample of 19 studies is not representative of all non-parametric studies. As a result, the 220 beds can be regarded as a lower bound.

Comparing the results from parametric studies with those from non-parametric studies leads to three conclusions. First, it is striking that non-parametric studies generate so little information about the optimum scale. Although non-parametric studies are quite popular, only a handful of such studies provide an estimate of the optimal scale or have sufficient information to estimate the optimum scale. This contrasts with parametric studies, which do not include an optimum scale, but frequently report a scale elasticity or have results that make it possible to derive a scale elasticity. Secondly, for non-parametric studies the optimum scale found in a study depends on the average scale size of the hospitals being studied. This is not the case for parametric studies. For parametric studies, model characteristics are more relevant: especially the use of a long-run or short-run cost function and the number of inputs included in the model have an impact on the scale elasticity and therefore the optimum scale. Finally, the optimum is similar for both types of studies. At first glance this seems a bold proposition, but since we are dealing with frontier studies for the non-parametric studies, we should compare with a stochastic frontier study for which the optimum is 239 beds, which compares well with a lower bound of 220 beds found for non-parametric studies.

What becomes clear from both parametric and non-parametric studies is that economies of scale only apply in a limited range, they become exhausted and eventually diseconomies of scale prevail. Based on frontier studies one might carefully conclude that the optimum scale lies around 238 beds – that is, the production associated with 238 beds. In the case of a Dutch hospital, that means about 25,000 admissions (including day-care) and 75,000 outpatients.

Furthermore, we should recognise that all kind of factors have an impact on the optimum scale and that the optimum certainly does not apply to all types of hospitals. Policymakers should realise that economies of scale is not synonymous with increasing the scale: if a hospital is already at its optimal scale, there is no point in expanding it. Furthermore, policymakers might also consider the opposite, namely reducing the scale if a hospital is too big –

although in this context the argument of Maindiratta (1990) applies, namely that decreasing returns set in very gradually so that hospitals have to be a lot bigger before it pays to apportion tasks to smaller units. In other words, it might be better to have one hospital that is too big, than two hospitals that are too small.. Finally, it should be noted that all of these conclusions are based on an economic perspective. There might be other arguments to concentrate on a suboptimal scale, such as a small size to provide accessibility in less populated areas.

Appendix: overview of literature

Author year	*Country*	*Specification*	*# Results*
Aletras (1999)	Greece	CD & TL	(2) specification
Azevedo and Mateus (2014)	Portugal	TL	(1)
Barbetta et al. (2007)	Italy	TL frontier	(1)
Barros et al. (2013)	Portugal	TL frontier	(1)
Blank and Eggink (2004)	Netherlands	TL	(1)
Blank and Merkies (2004)	Netherlands	TL	(2) modelling
Blank and Vogelaar (2004)	Netherlands	TL	(1)
Blank and Van Hulst (2009)	Netherlands	TL	(1)
Carey (1997)	USA	Cubic	(2) method of estimation
Carey (1998)	USA	Cubic	(1)
Carey (2000)	USA	Cubic	(3) method of estimation
Carey (2003)	USA	TL frontier	(1)
Carey et al. (2008)	USA	TL	(1)
Carey and Dor (2008)	USA	TL frontier	(1)
Carey and Stefos (2011)	USA	Log-linear	(4) specification
Carey et al. (2015)	USA	Cubic	(6) Q1 and Q3. General & for-profit
Custer and Willke (1991)	USA	Cubic	(1)
Daidone and D'Amico (2009)	Italy	TL frontier	(1)
Ennis et al. (2000)	USA	TL	(1)
Farsi and Filippini (2006)	Switzerland	CD frontier	(4) specification
Farsi and Filippini (2008)	Switzerland	CD frontier	(1)
Filippini et al. (2004)	Switzerland	CD frontier	(2) specification
Folland and Hofler (2001)	USA	CD frontier	(1)
Gaynor and Anderson (1995)	USA	TL	(1)
Kojima (2004)	Japan	TL	(8) specification
Kristensen et al. (2008)	Denmark	TL, QU	(3) specification
Li and Rosenman (2001)	USA	G. Leontief	(1)
Ludwig (2008)	Netherlands	CD frontier	(1)
Ludwig et al. (2010)	Netherlands	CD frontier	(1)
O'Donnell and Nguyen (2013)	Australia	CD frontier	(1)
Preyra and Pink (2006)	Canada	QU	(1)
Romley and Goldman (2011)	USA	TL	(1)
Sari (2003)	USA	Log-linear	(2) specification
Scuffham et al. (1996)	N.-Zealand	TL	(1)

Author year	*Country*	*Specification*	*# Results*
Sinay and Campbell (1995)	USA	GTL	(6) Q1 and Q3. Merger & control
Sinay (1998a)	USA	GTL	(12) Q1 and Q3. Closure & control, merger & control
Sinay (1998b)	USA	GTL	(6) Q1 and Q3. Merger & control
Smet (2004)	Belgium	GTL	(1)
Smet (2007)	Belgium	TL frontier	(1)
Vitikainen et al. (2010)	Finland	CD, CD frontier	(2) specification
Zhao et al. (2011)	Australia	Avg. costs	(1)

GTL=generalised translog, TL=translog, CD=Cobb–Douglas, QU=quadratic

Author year	*Country*	*Specification*	*# Results*
Blank and van Hulst (2011)	Netherlands	DEA	(1)
Bilsel and Davutyan (2014)	Turkey	DEA	(1)
Brown and Pagan (2006)	USA	DEA	(1)
Butler and Li (2005)	USA	DEA	(1)
Byrnes and Valdmanis (1994)	USA	DEA	(1)
Czypionka et al. (2014)	Austria	DEA	(1)
Dervaux et al. (2004)	USA	DEA	(1)
Ferrier and Valdmanis (1996)	USA	DEA	(1)
Flokou et al. (2011)	Greece	DEA	(1)
Jat and Sebastian (2013)	India	DEA	(1)
Jehu-Appiah et al. (2014)	Ghana	DEA	(1)
Lee et al. (2009)	USA	DEA	(1)
Linna and Häkkinen (1999)	Finland	DEA	(1)
Karagiannis (2012)	Greece	DEA	(1)
Masiye (2007)	Zambia	DEA	(1)
Marques and Carvalho (2013)	Portugal	DEA	(1)
McKillop et al. (1999)	Ireland	DEA	(1)
Sommersguter-Reichmann (2000)	Austria	DEA	(1)
Webster et al. (1998)	Australia	DEA	(1)

References

Aletras, V. (1997). The relationship between volume and the scope of activity and hospital costs. Concentration and choice in the provision of hospital services Part II. York: University of York

Aletras, V.H. (1999). A comparison of hospital scale effects in short-run and long-run cost functions. Health Economics, 8(6), 521-530.

Azevedo, H., & Mateus, C. (2014). Cost effects of hospital mergers in Portugal. European Journal of Health Economics, 15(9), 999-1010.

Banker, R. (1984). Estimating the most productive scale size using Data Envelopment Analysis. Eur J Oper Res, 17, 34 - 44.

Banker, R.D., Conrad, R.F., & Strauss, R.P. (1986). A comparative application of data envelopment analysis and translog methods: an illustrative study of hospital production. Management Science, 32(1), 30-44.

Barbetta, G.P., Turati, G., & Zago, A.M. (2007). Behavioral differences between public and private not-for-profit hospitals in the Italian National Health Service. Health Economics, 16(1), 75-96.

Barros, C.P., de Menezes, A.G., & Vieira, J.C. (2013). Measurement of hospital efficiency, using a latent class stochastic frontier model. Applied Economics, 45(1), 47-54.

Baumol, W.J., Panzar, J.C., Willig, R.D., Bailey, E.E., Fischer, D., & Fischer, D. (1982). Contestable markets and the theory of industry structure.

Bilsel, M., & Davutyan, N. (2014). Hospital efficiency with risk adjusted mortality as undesirable output: the Turkish case. Annals of Operations Research, 221(1), 73-88.

Blank, J.L.T., & Eggink, E. (2004). The decomposition of cost efficiency: an empirical application of the shadow cost function model to Dutch general hospitals. Health care management science, 7(2), 79-88.

Blank, J.L.T., & Merkies, A. (2004). Empirical assessment of the economic behaviour of Dutch general hospitals. Health Economics, 13(3), 265-280.

Blank, J.L.T., & Van Hulst, B.L. (2009). Productive innovations in hospitals: an empirical research on the relation between technology and productivity in the Dutch hospital industry. Health Economics, 18(6), 665-679.

Blank, J.L.T., & van Hulst, B.L. (2011). Governance and Performance: The Performance of Dutch Hospitals Explained by Governance Characteristics. Journal of Medical Systems, 35(5), 991-999.

Blank, J.L.T., & Vogelaar, I. (2004). Specifying technical change: a research on the nature of technical change in Dutch hospital industry. The Quarterly Review of Economics and Finance, 44, 448-463.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). Meta-analysis: fixed effect vs. random effects. Meta-Analysis. com.

Brautigam, R., & Daughety, V. (1983). On the Estimation of Returns to Scale using Variable Cost Function. Economics Letter, 11, 25-31.

Brown, H.S., 3rd, & Pagan, J.A. (2006). Managed care and the scale efficiency of US hospitals. International journal of health care finance and economics, 6(4), 278-289.

Butler, T.W., & Li, L. (2005). The utility of returns to scale in DEA programming: An analysis of Michigan rural hospitals. European Journal of Operational Research, 161(2), 469-477.

Byrnes, P., & Valdmanis, V. (1994). Analyzing technical and allocative efficiency of hospitals Data Envelopment Analysis: Theory, Methodology, and Applications (pp. 129-144): Springer.

Carey, K. (1997). A panel data design for estimation of hospital cost functions. Review of Economics and Statistics, 79(3), 443-453.

Carey, K. (1998). Stochastic demand for hospitals and optimizing 'excess' bed capacity. Journal of Regulatory Economics, 14(2), 165-187.

Carey, K. (2000). Hospital cost containment and length of stay: An econometric analysis. Southern Economic Journal, 67(2), 363-380.

Carey, K. (2003). Hospital cost efficiency and system membership. Inquiry-the Journal of Health Care Organization Provision and Financing, 40(1), 25-38.

Carey, K., Burgess, J.F., Jr., & Young, G.J. (2008). Specialty and full-service hospitals: A comparative cost analysis. Health Services Research, 43(5), 1869-1887.

Carey, K., Burgess, J.F., & Young, G.J. (2015). Economies of scale and scope: the case of specialty hospital. Contemporary Economic Policy, 33(1), 104-117.

Carey, K., & Dor, A. (2008). Expense preference behavior and management 'outsourcing': a comparison of adopters and non-adopters of contract management in US hospitals. Journal of Productivity Analysis, 29(1), 61-75.

Carey, K., & Stefos, T. (2011). Controlling for quality in the hospital cost function. Health Care Management Science, 14(2), 125-134.

Caves, D.W., Christensen, L.R., & Swanson, J.A. (1981). Productivity growth, scale economies, and capacity utilization in US railroads, 1955-74. The American Economic Review, 994-1002.

Custer, W.S., & Willke, R.J. (1991). Teaching hospital costs: the effects of medical staff characteristics. Health Services Research, 25(6), 831.

Czypionka, T., Kraus, M., Mayer, S., & Roehrling, G. (2014). Efficiency, ownership, and financing of hospitals: The case of Austria. Health Care Management Science, 17(4), 331-347.

Daidone, S., & D'Amico, F. (2009). Technical efficiency, specialization and ownership form: evidences from a pooling of Italian hospitals. Journal of Productivity Analysis, 32(3), 203-216.

Dervaux, B., Ferrier, G.D., Leleu, H., & Valdmanis, V. (2004). Comparing French and US hospital technologies: a directional input distance function approach. Applied Economics, 36(10), 1065-1081.

Ennis, S., Schoenbaum, M., & Keeler, T. (2000). Optimal prices and costs for hospitals with excess bed capacity. Applied Economics, 32(9), 1201-1212.

Farsi, M., & Filippini, M. (2006). An Analysis of Efficiency and Productivity in Swiss Hospitals. Schweizerische Zeitschrift fur Volkswirtschaft und Statistik/Swiss Journal of Economics and Statistics, 142(1), 1-37.

Farsi, M., & Filippini, M. (2008). Effects of ownership, subsidization and teaching activities on hospital costs in Switzerland. Health Economics, 17(3), 335-350.

Ferrier, G.D., & Valdmanis, V. (1996). Rural hospital performance and its correlates. Journal of Productivity Analysis, 7(1), 63-80.

Filippini, M., Farsi, M., Crivelli, L., & Zola, M. (2004). An analysis of efficiency and productivity in swiss hospitals.

Flokou, A., Kontodimopoulos, N., & Niakas, D. (2011). Employing post-DEA Cross-evaluation and Cluster Analysis in a Sample of Greek NHS Hospitals. Journal of Medical Systems, 35(5), 1001-1014.

Folland, S.T., & Hofler, R. (2001). How reliable are hospital efficiency estimates? Exploiting the dual to homothetic production. Health Economics, 10(8), 683-698.

Gaynor, M., & Anderson, G.F. (1995). Uncertain demand, the structure of hospital costs, and the cost of empty hospital beds. Journal of health economics, 14(3), 291-317.

Greene, W.H. (2003). Econometric analysis, 5th. Ed.. Upper Saddle River, NJ.

Hollingsworth, B. (2003). Non-parametric and parametric applications measuring efficiency in health care. Health care management science, 6(4), 203-218.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. Health Economics, 17(10), 1107-1128.

Hollingsworth, B., & Street, A. (2006). The market for efficiency analysis of health care organisations. Health Economics, 15(10), 1055-1059.

Jat, T.R., & Sebastian, M.S. (2013). Technical efficiency of public district hospitals in Madhya Pradesh, India: a data envelopment analysis. Global health action, 6, 21742-21742.

Jehu-Appiah, C., Sekidde, S., Adjuik, M., Akazili, J., Almeida, S.D., Nyonator, F., . . . Kirigia, J.M. (2014). Ownership and technical efficiency of hospitals: evidence from Ghana using data envelopment analysis. Cost effectiveness and resource allocation : C/E, 12(1), 9-9.

Karagiannis, R. (2012). Technical and scale efficiency of public hospitals in Greece.

Kojima, Y. (2004). Estimating the Cost Function of Japanese Public Hospitals - an Evaluation of Health Care Provision in Japan. Osaka, Japan: Graduate School of Economics, Osaka University.

Kristensen, T., Olsen, K.R., Kilsmark, J., & Pedersen, K.M. (2008). Economies of scale and optimal size of hospitals: Empirical results for Danish public hospitals: Syddansk Universitet.

Lee, K.-H., Yang, S.-B., & Choi, M. (2009). The Association between Hospital Ownership and Technical Efficiency in a Managed Care Environment. Journal of Medical Systems, 33(4), 307-315.

Li, T., & Rosenman, R. (2001). Estimating hospital costs with a generalized Leontief function. Health Economics, 10(6), 523-538.

Linna, M., & Häkkinen, U. (1999). Determinants of Cost efficiency and Finnish Hospitals: A Comparison of DEA and SFA. Helsinki/Edmonton: National Research and Development Centre for Welfare and Health.

Ludwig, M. (2008). Efficiency of Dutch hospitals. Maastricht: Datawyse.

Ludwig, M., Van Merode, F., & Groot, W. (2010). Principal agent relationships and the efficiency of hospitals. European Journal of Health Economics, 11(3), 291-304.

Maindiratta, A. (1990). Largest size-efficient scale and size efficiencies of decision-making units in data envelopment analysis. Journal of Econometrics, 46(1), 57-72.

Marques, R.C., & Carvalho, P. (2013). Estimating the efficiency of Portuguese hospitals using an appropriate production technology. International Transactions in Operational Research, 20(2), 233-249.

Masiye, F. (2007). Investigating health system performance: An application of data envelopment analysis to Zambian hospitals. Bmc Health Services Research, 7.

McKillop, D.G., Glass, J.C., Kerr, C.A., & McCallion, G. (1999). Efficiency in northern Ireland hospitals: A non-parametric analysis. Economic and Social Review, 30(2), 175-196.

Nguyen, K.-H., & Coelli, T. (2009). Quantifying the effects of modelling choices on hospital efficiency measures: a meta-regression analysis: School of Economics, University of Queensland, Australia.

O'Donnell, C.J., & Nguyen, K. (2013). An econometric approach to estimating support prices and measures of productivity change in public hospitals. Journal of Productivity Analysis, 40(3), 323-335.

O'Neill, L., Rauner, M., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. Socio-Economic Planning Sciences, 42(3), 158-189.

Preyra, C., & Pink, G. (2006). Scale and scope efficiencies through hospital consolidations. Journal of Health Economics, 25(6), 1049-1068.

Raudenbush, S.W. (1994). Random effects models. The handbook of research synthesis, 301-321.

Romley, J.A., & Goldman, D.P. (2011). How Costly is Hospital Quality? A Revealed-Preference Approach. Journal of Industrial Economics, 59(4), 578-608.

Rosko, M.D., & Mutter, R.L. (2008). Stochastic frontier analysis of hospital inefficiency. Medical Care Research and Review, 65(2), 131-166.

Rosko, M.D., & Mutter, R.L. (2011). What have we learned from the application of stochastic frontier analysis to U.S. hospitals? Medical care research and review : MCRR, 68(1 Suppl), 75S-100S.

Sari, N. (2003). Efficiency outcomes of market concentration and managed care. International Journal of Industrial Organization, 21(10), 1571-1589.

Scuffham, P.A., Devlin, N.J., & Jaforullah, M. (1996). The structure of costs and production in New Zealand public hospitals: An application of the transcendental logarithmic variable cost function. Applied Economics, 28(1), 75-85.

Sinay, U.A., & Campbell, C.R. (1995). Scope and Scale Economies in Merging Hospitals Prior to Merger. Journal of Economics and Finance, 19(2), 107-123.

Sinay, U.T. (1998a). Hospital mergers and closures: Survival of rural hospitals. Journal of Rural Health, 14(4), 357-365.

Sinay, U.T. (1998b). Pre- and Post-Merger Investigation of Hospital Mergers. Eastern Economic Journal, 24(1), 83-97.

Smet, M. (2002). Cost characteristics of hospitals. Social Science & Medicine, 55(6), 895-906.

Smet, M. (2004). Multi-product costs and standby capacity derived from queuing theory: the case of Belgian hospitals. Applied Economics, 36(13), 1475-1487.

Smet, M. (2007). Measuring performance in the presence of stochastic demand for hospital services: an analysis of Belgian general care hospitals. Journal of Productivity Analysis, 27(1), 13-29.

Sommersguter-Reichmann, M. (2000). The impact of the Austrian hospital financing reform on hospital productivity: empirical evidence on efficiency and technology changes using a non-parametric input-based Malmquist approach. Health care management science, 3(4), 309-321.

Vita, M.G. (1990). Exploring hospital production relationships with flexible functional forms. Journal of Health Economics, 9(1), 1-21.

Vitaliano, D.F. (1987). On the estimation of hospital cost functions. Journal of Health Economics, 6(4), 305-318.

Vitikainen, K., Linna, M., & Street, A. (2010). Substituting inpatient for outpatient care: what is the impact on hospital costs and efficiency? European Journal of Health Economics, 11(4), 395-404.

Webster, R., Kennedy, S.K., & Johnson, L. (1998). Comparing techniques for measuring the efficiency and productivity of Australian private hospitals.

Wilson, P.W., & Carey, K. (2004). Nonparametric analysis of returns to scale in the us hospital industry. Journal of Applied Econometrics, 19(4), 505-524.

Worthington, A.C. (2004). Frontier efficiency measurement in health care: A review of empirical techniques and selected applications. Medical Care Research and Review, 61(2), 135-170.

Zhao, Y., Foley, M., & Eagar, K. (2011). Assessing Economies of Scale in Public Hospitals. Economic Papers: A journal of applied economics and policy, 30(3), 341-347.

Governance and performance

Governance and performance: the performance of Dutch Hospitals explained by governance characteristics

---

This chapter has been published as:

Blank, J.L.T., & van Hulst, B.L. (2011). Governance and Performance: The Performance of Dutch Hospitals Explained by Governance Characteristics. Journal of Medical Systems, 35(5), 991-999

3.1 Introduction

Developments in health care policy in western countries during the last two decades have been characterised by liberalising markets, financial reforms and deregulation. Many health care systems are being transformed from centrally governed systems into regulated competitive markets (see e.g. Chang et al., 2004; Kittelsen et al., 2008; McKee & Healy, 2002; Saltman et al., 2007; Walford & Grant, 1998). Although there is a large variety of systems from country to country, there are some common characteristics. Generally, there has been a shift in competencies from ministries and central authorities to health care suppliers and insurers with regard to service price setting, capacity planning, investments, business conduct, strategic decisions and property rights. The impact of these changes has been analysed extensively. Most of this research is focused on issues concerning the relation between the characteristics of suppliers on the one hand, (issues such as scale, scope, property rights and market concentration) and efficiency and quality on the other hand. The policy reforms were, however, accompanied by substantial change to the way suppliers were managed and controlled. Whereas previously a single, uniform management and control model was used, now a variety of models have been established. Large variations were developed in the size of the management board, the size of the supervisory boards, the remuneration of board members, the intensity of the supervisory board's control, the application of integrity codes and the transparency of decision-making. So far researchers have not paid much attention to these aspects of management and control, which can be summarised in the term 'corporate governance', or to their quantitative effects on productivity and efficiency. Although literature on this issue is scarce, some interesting approaches can be found (see e.g. Azizi et al., 2007; Bozec & Dia, 2007; Diboky & Ubl, 2007; Ditzel et al., 2006; Eldenburg et al., 2004). This chapter therefore focuses on the relationship between corporate governance and efficiency.

An interesting case is the Dutch hospital industry, where the corporate governance is embedded in a governance code. A governance code provides guidelines for good governance, adequate supervision, accountability and justification and is an instrument for self-regulation. The urge to have guidelines for good governance has arisen as there is less supervision from the government. A governance code fills the gap of deregulation. The practice of governance codes in the Dutch hospital industry started a decade ago. In 1999 the Health Care Governance commission published recommendations and guidelines for good governance. The recommendations gave momentum to the debate over good governance and supervision. This resulted in several different governance codes in the Dutch health care system. This situation lasted until 2005, when a single governance code was developed for almost the entire health care system. The academic hospitals, due to regulation, have their own specific governance code.

Summarizing, most of the research on the impact of health care reforms directly focuses on the relationships between the main aspects of reform, such as non-regulated prices, free market entry and competition, and on the efficiency of health care providers. No attention is paid to more indirect effects through changes in corporate governance due to the reform. To our knowledge no research has been carried out yet on the effects of corporate governance structures on efficiency in a health care industry. Since large variation exists in corporate governance structures and relevant data are available, the Dutch hospital industry provides a unique case to establish this type of relationships.

In this chapter we quantify the effects of corporate governance structure on efficiency of Dutch hospitals. To do so, we apply Data Envelopment Analysis (DEA) to derive cost-efficiency scores and follow up by a second stage of the analysis. Bootstrapping techniques are applied to deal with consistency and bias-correction (see Simar & Wilson, 2007). The effect (if any) of corporate governance structure factors on the cost-efficiency scores is

identified. We apply this approach to a set of Dutch hospital data since we wish to provide Dutch hospitals with relevant information on establishing productive corporate governance structures. Earlier work on the efficiency of Dutch hospitals can be found in Blank and Valdmanis (2009), Blank and van Hulst (2009) and Blank and Merkies (2004).

A two-stage analysis that assesses the impact of explanatory factors on efficiency scores derived using DEA has garnered attention in the literature. Whereas some have used Ordinary Least Squares (OLS), Tobit analysis has been the most popular analytical method. In this method the output-based or the reciprocal of the input-based efficiency score is regressed on a variety of variables thought to affect efficiency (see e.g. Kooreman, 1994). Simar and Wilson (2007) challenge this approach by demonstrating that in the second stage:

- Serial correlation arises and explanatory variables are correlated with error terms, which disappear at a slow rate of convergence;
- The efficiency score, which is the dependent variable in the second stage, has a bias.

As an alternative to simply using the efficiency measure as a discrete point with a bias, Simar and Wilson (2007) advocate the use of bootstrapping techniques in order to obtain unbiased and consistent estimates. Various studies have recently applied interesting applications of the Simar and Wilson technique to hospital data (see e.g. Pilyavsky et al., 2006; Puenpatom & Rosenman, 2008; Staat, 2006).

The outline of this chapter is as follows: Section 2 describes economic theory on corporate governance; in Section 3 we define the DEA model and the bootstrapping procedure; in Section 4 describes the available data; Section 5 presents the empirical results and Section 6 concludes.

3.2 Economic theory on corporate governance

There is no general economic framework for evaluating corporate governance structure and the efficiency of business entities. Some elements refer directly to the principal agent problem, which reflect the differing goals of various stakeholders. Government wants to maximise public values, whereas members of the management board or the supervisory board strive to maximise individual goals, such as remuneration or status. In the case of hospitals, patients want to maximise accessibility and quality of care. The extent to which each stakeholder succeeds in achieving these goals strongly depends on their relative position with regard to information, market power, and the instruments available to influence the outcomes of the business process (institutional context). It also depends on the personal characteristics of stakeholders (quality, experience and ethics). Consequently, theory on corporate governance includes aspects of agency theory, production theory, industrial economics and institutional economics. With particular respect to agency theory, Bozec and Dia (2007) present an interesting overview.

We present a rather heuristic theoretical approach in discussing various characteristics of corporate governance structure. We distinguish four major clusters of characteristics: the management board, the supervisory board, the external stakeholders and a cluster of institutional relationships between various stakeholders.

Management board

The management board can be regarded as a resource in the production process. Due to the production structure, size and quality should be in accordance with the level and composition of the services and goods provided. Deviations from the optimal level of management are considered to be allocative inefficiencies. For example, Rodríguez-Alvarez and Lovell (2004) present an application to the Spanish public hospital sector and observe persistent allocative inefficiency in variable inputs and overcapitalisation in these hospitals. A focus on the continuing improvement of employee skills

through training and education is also seen as part of the quality of management (see e.g. Azizi et al., 2007). Evidence for the hypothesis that board size and efficiency are negatively correlated can be found in Eisenberg et al. (1998) and Yermack (1996). The size of the board can be measured as the number of board members, whereas quality can be measured by the members' level of education, the number of years of board experience and the number of new board members. The composition of the board in terms of profession (e.g. economist, lawyer or doctor) may also reflect management board quality. A rather indirect measure of quality is remuneration.

Supervisory board

The supervisory board's main assignment is to act as a countervailing power to the management board. The supervisory board audits and advises the management board. Based on legislative instruments they approve the annual accounts and budgets, monitor the integrity of the hospital and have a say in strategic decisions (such as mergers). Their activities can also be regarded as part of a production process, in which resources are transformed into a number of audits, checks and advice. The size of the board and the quality of its members are therefore relevant characteristics. One should bear in mind that board members may also have personal preferences that conflict with public goals. Personal characteristics of the board may therefore also reflect the ability to accomplish these personal preferences. To accurately reflect the supervisory board's size and quality, the same type of variables used for the management board should be included.

Other stakeholders

Other stakeholders include the central government, insurance companies and patients. It is obvious that central and local governments dictate the regulatory environment, which consequently determines the playing field for commercial enterprises. Issues such as capacity planning, price setting, budget allocation, profit/not-for-profit and so forth also affect the corporate governance structure. Interesting examples of research on the effects of

ownership and profit/not-for-profit entities on efficiency can be found in Diboky and Ubl (2007) (also with bootstrapping techniques) and Mutter and Rosko (2008) (US hospitals). Since these issues are not a part of our research, we will exclude them from further discussion.

Since insurance companies are hospitals' major clients they have a certain influence on the corporate governance structure. The way in which insurance companies are able to use their influence, and the degree to which they do, differs not only according to the regulatory environment but can also differ across organisations: compare, for instance, a HMO and a for-profit hospital. In the Dutch case insurance companies and hospitals are strictly independent, however through their regional market power insurance companies have informal influence on corporate governance.

The role of patients in the corporate governance structure will be expressed in the way patients are able to affect business conduct. Some firms, for instance, provide client (patient) representation on a statutory basis.

Multi-actor dependencies

Multi-actor dependencies refer to the formal and informal relationships between various actors. The relationship between the management board and the supervisory board is one example. We can distinguish two types of relations; first the management board with final responsibility and second the supervisory board or management subordinated by a board of governors of the foundation. In case of a supervisory board the management board has a maximum of competence power and will be executive. In case of a board of governors the board of governors has less competence power and will determine the policy, management has the role of the executive.

Another relevant factor is board independence. Outside managers are supposed to fulfil their monitoring function better than executive managers because they are concerned about their reputation (see e.g. Fama & Jensen, 1983). Baysinger and Hoskinson (1990), on the other hand, claim that inside

managers have inside information and are therefore in a better position to evaluate business conduct and performance. Another aspect affecting boards is internal relations: a board might have a chairperson but it is also possible to have a collegial board. Finally the members of the board can be either internal, in case they are employees of the hospital, or external, i.e. an external interim manager.

3.3 Model and method

As mentioned in the introduction, we have applied a two-stage estimation procedure with bootstrapping to investigate the effect of governance variables on hospital performance. We followed the methodology indicated as algorithm 1 in Simar and Wilson (2007, page 41-42), the methodology of algorithm 1 is described in this chapter. The first stage of the procedure is to estimate the cost efficiency of hospitals. The second stage consists of explaining cost efficiency with governance variables. In the second stage a bootstrap procedure is applied. The second stage bootstrap procedure leads to more accurate estimators for the explanatory variables.

In the first step of our analysis, we conducted standard cost efficiency DEA as described in Färe et al. (1994), for an overview of DEA literature see Emrouznejad et al. (2008). Since we have information on input prices for our sample of Dutch hospitals, we used the cost-efficiency model rather than the technical-efficiency DEA model that does not require input prices. In this standard cost-efficiency DEA model the cost efficiency of hospital 'A' equals the ratio of minimum cost to actual cost. In other words, we gauge the minimum expenditure required to produce service levels given resource prices. The actual cost efficiency measure (CE) is derived by the radial distance between the observed hospital's resources-services correspondence to the 'best practice' frontier. This best practice frontier is constructed by the linear combination of hospitals producing the same levels of services as hospital A but at a lower level of cost.

The mathematical formulation is:

$$
\begin{aligned}
& CE = min_{z,x} \frac{w^A x^A}{w^A x} \;\; subject \;\; to \\
& \textstyle\sum_j z^j y^j \geq \; y^A \\
& \textstyle\sum_j z^j x^j \leq \; x^A \\
& z^j \; \geq 0 \\
& and \; \textstyle\sum_j z^j = 1 \quad (in\ case\ of\ VRS)
\end{aligned}
\tag{1}
$$

With:

CE = cost efficiency;
w^A = vector of resource prices of hospital A;
x^A = vector of resources of hospital A;
y^A = vector of services of hospital A;
x = best practice vector of resources;
z = vector of weights.

The model described here assumes constant returns to scale (CRS). CRS has been justified as long-term equilibrium, since a hospital can adjust its size over time. However, it makes sense to correlate some of the explanatory variables with the size of the hospital, e.g. remuneration of the board and size of the board. If we assume CRS, this could lead to conclusions about the governance variables, which are ambiguous. This is because the estimated parameters also tell us something about the relation between size and efficiency and whether the assumption of CRS is valid. Variable returns to scale (VRS) deals with scale effects. In general it is easier to be cost efficient under VRS then under CRS. Instead of choosing between CRS and VRS we applied both and discussed the results (e.g. Wilson & Carey, 2004). In the model VRS means that we have to add the restriction that the sum of z^j equals one.

After solving for the cost-efficiency scores for hospitals in our dataset, we regressed the reciprocal of the cost-efficiency scores on a set of explanatory variables. Our cost-efficiency scores are obtained as results which theoretically range from zero to one, so the reciprocal varies from one to infinity. We do so because we want to apply a truncated regression. A higher reciprocal therefore implies greater cost inefficiencies. The (explanatory) variables account for the governance variables for each hospital. The regression equation is given by:

$$\delta = \beta_0 + \sum_k \beta_k Q_k + \varepsilon \geq 1 \tag{2}$$

With:

δ = reciprocal of the cost efficiency score;
Q_k = k-th environmental feature;
β_k = parameters to be estimated;
ε = error term.

However, note that δ is unobserved and is replaced by the estimates. So the actual equation is given by:

$$\hat{\delta} = \beta_0 + \sum_k \beta_k Q_k + \varepsilon \geq 1 \tag{3}$$

There are several methods to estimate (3), including OLS, Tobit analysis and truncated regression. However Simar and Wilson (2007) show that there are some issues in estimating equation (3). First, the estimates of the cost-efficiency scores have bias. Second, because the cost-efficiency scores measured by the DEA approach are measured non-parametrically, there is no error term associated with the measure which when used as a dependent variable in the second stage analysis could lead to inconsistent estimators.

To address the bias we begin by specifying the equation given in (3):

$$\hat{\delta} = E(\hat{\delta}) + u \tag{4}$$

with $\mathrm{E(u)} = 0$. The bias of the estimator is defined by:

$$bias(\hat{\delta}) \equiv E(\hat{\delta}) - \delta \tag{5}$$

Substituting (2) and (4) into (5) and rearranging terms yields:

$$\hat{\delta} - bias(\hat{\delta}) - u = \beta_0 + \sum_k \beta_k Q_k + \varepsilon \geq 1 \tag{6}$$

Even though the u's have a zero mean, the bias term does not, it is always strictly negative in finite samples. Although the u's are unknown and cannot be estimated, the bias term can be estimated by bootstrap methods (for a detailed discussion of the bootstrapping approach, see Efron and Tibshirani (1993) and Simar and Wilson (2000b)). The bootstrap estimates of the bias can be used to obtain a bias-corrected estimator of:

$$\hat{\hat{\delta}} = \hat{\delta} + bias(\hat{\delta}) \tag{7}$$

Next we address the consistency of the estimators. As we have already noted δ is unobserved and is replaced by the estimates. However $\hat{\delta}$ has serial correlation, because it depends on all the observations of (w^A, x^A, y^A). Furthermore, the explanatory variables are correlated with $\hat{\delta}$, otherwise there would be no reason for a second stage. Asymptomatically the serial correlation and the correlation between explanatory variables and error terms disappears, but at a slow rate. This means that a maximum likelihood of β is consistent, however the usual parametric rate of convergence $(1/\sqrt{\mathrm{n}})$ does not apply. Therefore using a bootstrap procedure in the second stage may be more appropriate than a simple multiple regression approach since a benefit of bootstrapping is that it leads to consistent estimates of β_{k}.

We next describe the bootstrapping procedure used. We use the algorithm indicated as algorithm 1 in Simar and Wilson (2007). Basically, the algorithm

consists of a procedure in which estimates of δ are obtained and a bootstrap procedure in which estimates of β_k are obtained through truncated regression:

1. Compute the DEA scores using (1) to obtain $\hat{\delta}_i$.
2. Use maximum likelihood to obtain estimates $\hat{\beta}_k$ and $\hat{\sigma}_\varepsilon$ for the truncated regression of the efficiency scores on the governance variables using (3), use only the observation for which $\hat{\delta}_i > 1$.
3. Apply the next three steps L times to obtain a set of bootstrap estimates $\{(\hat{\beta}^*, \hat{\sigma}^*_\varepsilon)_b\}_{b=1}^{L}$.
   3.1. For each draw i = 1,..,n draw ε_i from the $N(0, \hat{\sigma}_\varepsilon)$ distribution with left truncation at $1 - \hat{\beta}_0 - \sum_k \hat{\beta}_k z_k$.
   3.2. For each i = 1,.., n compute $\delta_i^* = \hat{\beta}_0 + \sum_k \hat{\beta}_k z_k + \varepsilon_i$.
   3.3. Use maximum likelihood to estimate the truncated regression of δ_i^* on the z_k's yielding estimates $(\hat{\beta}^*, \hat{\sigma}^*_\varepsilon)$.
4. Use the bootstrap values and the original parameter estimates to construct estimated confidence intervals for the parameters of interest.

To construct the confidence interval for β_j the following procedure can be applied. If the distribution of $(\hat{\beta}_j - \beta_j)$ were known, the confidence interval follows from finding values a_α and b_α such that:

$\Pr[-b_\alpha \leq (\hat{\beta}_j - \beta_j) \leq -a_\alpha] = 1 - \alpha$, for small values of α>0. However, the distribution is unknown and therefore we use the j-th element of each bootstrap value instead to find values a^*_α and b^*_α such that:

$\Pr[-b^*_\alpha \leq (\hat{\beta}^*_j - \beta_j) \leq -a^*_\alpha] \approx 1 - \alpha$. Finding a^*_α and b^*_α involves sorting the values $(\hat{\beta}^*_j - \beta_j)$ in increasing order and then deleting $(\alpha/2 \times 100)$ percent of the elements at either end of the sorted list. After the sorted list is determined we set $-a^*_\alpha$ and $-b^*_\alpha$ equal to the endpoints of the truncated, sorted array. The estimated $(1\text{-}\alpha)$ percent confidence interval is then given by: $[\hat{\beta}_j + a^*_\alpha, \hat{\beta}_j + b^*_\alpha]$.

Note that in the procedure the parameters are bootstrapped. It is also possible to bootstrap the DEA-scores direct (Simar & Wilson, 1998, 2000a), our focus is however on the parameters.

3.4 Data

General

In this study, we used hospital data for the year 2007. Our data come from two sources. First of all detailed information on inputs and outputs were obtained from the Ministry of Health, Welfare and Sport and were collected by the Institute for Health Care Management using numerous surveys, such as financial, patient and personnel surveys. Secondly, for the data on governance, we used data from the annual reports for hospitals. The annual reports are compulsory and are systematically collected by the Central Information point Healthcare Professions (CIBG). The collected data are freely obtainable in a practical digital dataset. The dataset with inputs and outputs is merged with a dataset with governance variables. For the purposes of this study, observations on hospitals with missing or unreliable data and academic hospitals were excluded from the dataset. Academic (7) hospitals have a very different cost structure due to their teaching and research activities such that comparing them to general hospitals is unreliable. Our final dataset contains 75 observations. Since there were 86 hospitals in 2007 in the Netherlands, 11 hospitals are excluded. This was due to missing data, unreliable data or in one case the hospital was excluded because it is a military hospital.

Inputs and Outputs

Table 3-1shows the descriptive statistics of the inputs and outputs.

Table 3-1 Descriptive Statistics, Dutch General Hospitals 2007

	Mean	Std. Deviation	Minimum	Maximum
Output				
Discharges group 1	13.088	7.403	1.754	36.094
Discharges group 2	12.741	6.591	2.152	32.048
Discharges group 3	7.066	3.311	1.344	17.469
First-time visits	69.146	31.544	18.429	152.017
Input prices (in euro)				
Staff and administrative personnel	44.730	4.346	32.907	62.493
Nursing personnel	47.606	3.120	39.367	58.872
Paramedical personnel	111.684	36.368	54.524	249.741
Other personnel	34.191	4.186	10.220	44.695
Inputs (x 1000 euro)				
Staff and administrative personnel	11.090	7.050	1.566	30.093
Nursing personnel	37.789	21.683	9.474	109.699
Paramedical personnel	8.229	7.174	0	45.876
Other personnel	7.449	4.183	90	20.053
Material supplies	38.267	23.666	8.823	111.894
Variable cost (x 1000 euro.)	102.824	60.879	25.978	290.839

Since the main objective of hospitals is patient care, we define the services of hospitals as the number of first-time visits (i.e. the number of patients treated by physicians without an admission) and the number of discharges. Discharges have been separated into medical specialties in order to capture case-mix differences. The dataset distinguishes over 30 specialties, so for computational ease, we aggregated these medical specialties into three categories on the basis of average length of stay (LOS) of a specialty and whether or not patients had surgery. Our first group of patients were treated by a doctor in a specialty with a LOS less than the general LOS. Our second and third group of patients were treated by a specialty with an above-average LOS. The distinction between the second and third group is whether the patient was treated by a surgical specialty or not.

Inputs include staff and administrative personnel, nursing personnel, paramedical personnel (such as lab technicians), other personnel (including

maintenance, security and cleaning), and material supplies. Material supplies include medical supplies, food and heating. Personnel and material supplies are treated as variable resources since the hospital can change these in the short term. Regarding personnel we have data on the volume in terms of fulltime equivalents as well as salary costs, input prices are obtained by dividing costs and volume. It is possible to distinguish several inputs, but as we apply DEA we wish to reduce the number of variables. Because our focus is not really on allocation of inputs in the model we reduced the number of variables by aggregating the inputs

Governance variables

We have information about several variables that provide information on governance; information on governance is publically available through the annual accounts, which are compulsory for hospitals. Section 2 distinguishes three major clusters of characteristics, with variables that provide information about the management board, the supervisory board and a third set of variables that refers to external stakeholders. Information about the management board includes:

- the size of the board, measured with two dummy variables. One dummy for two members and one dummy for three or more members, which leaves one member as the reference group (constant term) ;
- remuneration of board members. Since not all members receive the same remuneration, for practical reasons we have taken the remuneration figure for the chairperson;
- the type of contract (internal or external, with external referring in most cases to interim management).

For the supervisory board our information includes:

- remuneration of board members. For practical reasons we have taken the remuneration figure for the chairperson;

Multi-actor dependencies:

- statutory changes, measured by a dummy variable if statutory changes took place in 2007. Statutory changes refer to a change of the legal form, a change in the way care is organized or a change in the competence of the internal organs.

Table 3-2 shows the descriptive statistics for the governance variables.

Table 3-2 Descriptive Statistics, Governance variables Dutch Hospitals 2007

	Mean	Std. Dev.	Minimum	Maximum
Included in analyses				
dummy, size of the board = 1	0.28	0.45	0	1
dummy, size of the board = 2	0.56	0.50	0	1
dummy, size of the board => 3	0.16	0.37	0	1
remuneration chairperson of the board	243.127	94.063	65.934	644.778
dummy, external contract (int. man.)	0.16	0.37	0	1
remuneration chairperson supervisory board	8.341	3.965	2.500	24.000
dummy, statutory changes	0.49	0.50	0	1
Excluded from analyses				
average # of years members of the board	4.1	4.3	0	25.0
dummy, member leaving the board	0.52	0.78	0	4
the size of the supervisory board	6.7	1.7	4.0	12.0
independence of the supervisory board	0.27	0.45	0	1
statutory provision of client representatives	0.11	0.31	0	1
use of NVZD-code for remuneration	0.79	0.41	0	1

There are some other interesting governance variables available, such as whether the board has a chairperson or the type of administration. However, they are not usable due to a lack of variation across hospitals. For instance, all general hospitals used the same particular governance code in 2007. Some other variables were tested but did not lead to significant results, i.e. number of members that left the board and the size of the supervisory board.

3.5 Empirical Results

For the empirical results we used FEAR (DEA results) and TSP (truncated regressions and bootstrap procedure), the DEA results were also checked with Onfront. The first step of the algorithm results in DEA-scores, Table 3-3 presents the statistics of the DEA-scores. The table contains the results for CRS as well VRS. The lowest score is one, representing the hospitals that are efficient.

Table 3-3 DEA result, reciprocal of the cost efficiency under CRS and VRS

	CRS	VRS
Mean	1.28	1.12
Std. Deviation	0.21	0.13
Minimum	1.00	1.00
Maximum	2.12	1.82
95% percentile	1.67	1.34

Under the CRS assumption the average efficiency is 1.28, the maximum ranges up to 2.12. Under the VRS assumption the average is 1.12 while the maximum is 1.82. These outcomes are very common. Ozcan (2008) summarizes the efficiency scores of a number of hospital studies. Most of these studies report scores near 90 % (meaning the reciprocal is 1.11), depending on the DEA-variant chosen, the distinct services and resources and sample. Also note that the scores under VRS are lower than CRS, this is due to the fact that scale effects are absorbed under VRS.

The efficiency scores provide an overview of the general cost efficiency in the Dutch hospital sample. The variability of performance can be explained by differences in governance. Hence we continue with regressing the efficiency score on the governance variables using the algorithm as discussed in section 3. Recall that we use a bootstrap procedure to generate the results, therefore we have no point estimates of the parameters. Our results are presented as the lower and upper bound of a 95% confidence interval of the estimates. Table 3-4 presents the results of the second stage bootstrap estimates.

Table 3-4 Bounds for 95 % confidence intervals for the parameter estimates

	Median	Lower bound (α=2.5 %)	Upper bound (α=97.5 %)
CRS			
Constant	**1.185**	**0.933**	**1.282**
dummy, size of the board = 2	0.020	-0.064	0.103
dummy, size of the board > 3	**0.141**	**0.0001**	**0.247**
remuneration chairperson of the board	**0.145**	**0.044**	**0.253**
remuneration chairperson sup. board	**0.095**	**0.013**	**0.183**
dummy, external contract (int. man.)	-0.019	-0.060	0.146
dummy, statutory changes	-0.014	-0.108	0.042
sigma	**0.151**	**0.127**	**0.175**
VRS			
Constant	**1.099**	**1.041**	**1.296**
dummy, size of the board = 2	-0.039	-0.103	0.018
dummy, size of the board > 3	-0.032	-0.131	0.037
remuneration chairperson of the board	0.089	-0.010	0.144
remuneration chairperson sup board	**0.052**	**0.001**	**0.124**
dummy, external contract (int. man.)	**0.040**	**0.010**	**0.162**
dummy, statutory changes	-0.005	-0.103	0.001
sigma	**0.110**	**0.088**	**0.122**

Bold: significant at the 5% level

The results can be interpreted as follows. A positive sign means that the variable does not lead to better performance in terms of cost efficiency; likewise a negative sign means that the variable indicates better performance in terms of cost efficiency. Furthermore we included the confidence intervals. A long confidence interval means that there is more uncertainty about the actual value of the parameter. If the borders of a confidence interval have opposite signs, it means we are not sure if there is an interrelationship between the variable and better performance in terms of cost efficiency.

From our modelling, we find some interesting differences between the results under CRS and VRS. While under CRS and VRS we find several parameters to be significant at the 5% level, the significant parameters are not

always the same parameters. In fact only the remuneration of the supervisory board is significant under both assumptions. Some of the explanatory variables under CRS are not merely a measure for governance but also a measure of scale. On the other hand if we are less restrictive with our level of significance we get more significant parameters under VRS. For example if the confidence interval is 90% instead of 95% the remuneration of the board, extern board members and statutory changes are also significant under VRS.

Under CRS we find that remuneration of the board and the size of the board are significant with a positive sign. This is partly so because of the earlier mentioned correlation with the size of the hospital for these variables. However, if we are less strict with the level of significance we find that remuneration of the board is also significant under the VRS assumption. A higher remuneration of the board does not lead to better performance in terms of cost efficiency. One can suggest all kinds of explanation for this varying from overambitious management to the case were the lack of performance is identified and expensive management is hired to get back on track.

The remuneration of the supervisory board is significant with a positive sign under both CRS and VRS. This means that when the remuneration of the supervisory board increases, the performance of the hospital, in terms of cost efficiency, gets worse. This implies that remuneration of the supervisory board is not a sufficient condition for a professional supervisory board that is able to guard the performance of the hospital.

External members in the board, in most cases interim management, leads to significant estimates with a positive sign under VRS. A plausible hypothesis is that interim management has a knowledge gap about the hospital, which may result in lower performance. However, caution is required in interpreting this result, since interim management may also reflect serious organizational problems. In that case causality between interim management and efficiency should be reversed. More details, for instance about the point of time the

interim manager joined the hospital or about temporarily filling-in of a regular vacancy, may shed some light on this.

The variable statutory changes is only significant under VRS and it is only the case when we are less restrictive with our confidence interval. The sign is negative meaning that a statutory change is correlated with a more efficient score.

3.6 Conclusions

This chapter investigates the effect of the corporate governance structure of hospitals on cost efficiency using the method of Data Envelopment Analysis (DEA) with a bootstrapping procedure. We use the DEA measure of cost-efficiency on the hospital level. Our focus then turns to explaining variations in cost inefficiency which is due to a hospital's corporate governance.

A popular way to conduct such analysis has been a Tobit analysis wherein the efficiency score is regressed on a variety of variables thought to affect efficiency. However, in this second stage, DEA-scores are derived relative to a best practice frontier which does not have an associated error term. Without this error term, it is possible that bias may arise leading to measurement error in the dependent variable problem i.e., biased and inconsistent estimates. Simar and Wilson (2007) suggest using a bootstrapping procedure in order to obtain consistent estimates.

Following this suggestion, we proceed to an analysis of a sample of Dutch hospitals using several steps. In the first stage, DEA results indicate, that on average, cost efficiency for general hospitals is 1.28 under the CRS assumption and 1.12 under the VRS assumption. The second stage shows that the cost efficiency scores can be explained by variables that measure aspects of the corporate governance of hospitals. Whether the explanatory variables are

significant depends on the assumption on the returns to scale. Especially under CRS there are some governance variables, i.e. the size of the board, that are correlated with the scale and therefore also explain scale effects. Under VRS scale effects are of course absorbed.

Developments in health care policy in the last two decades in western countries have a tendency towards deregulation. As a result management and control of the health care provider has changed. Justification, transparency and good governance are import elements in the deregulated environment. It is therefore surprising that in productivity analysis not too much attention has been paid to the relation between corporate governance and cost efficiency. Here we investigate this relation for the Dutch hospital industry and contributes in the discussion over good governance. It is important to note that not all differences in cost efficiency are a result of governance. However, the board and the supervisory board call the shots in an organization, quality of governance will therefore have an impact on the performance. This shows that the relation between governance and cost efficiency exists and how they relate. From the viewpoint of the policymaker it is therefore important to keep on monitoring the governance, keep searching for best practices and stimulate governance structures that lead to better performance.

References

Azizi, K.H., Lotfi, F.H., Saati, A., & Vahidi, A.R. (2007). Manpower Efficiency in Electricity Producer Companies in the View of Economic Management; Some Approaches for Increment of Efficiency in Inefficient Companies by DEA. International Journal of Mathematical. Analysis,, 1, 229-236.

Baysinger, B.D., & Hoskisson, R.R. (1990). The composition of board of directors and strategic control: effects on corporate strategy. Academy of management review, 15, 72-87.

Blank, J., & Valdmanis, V. (2009). Environmental factors and productivity on Dutch hospitals: a semi-parametric approach. Health Care Management Science.

Blank, J.L.T., & Merkies, A.H.Q.M. (2004). Empirical assessment of the economic behaviour of Dutch general hospitals. Health Economics, 13(3), 265-280.

Blank, J.L.T., & van Hulst, B.L. (2009). Productive Innovations in Hospitals: An Empirical Research on the Relation between Technology and Productivity in the Dutch Hospital Industry. Health Economics, 18(3), 665-679.

Bozec, R., & Dia, M. (2007). Board structure and firm technical efficiency: Evidence from Canadian state-owned enterprises. European Journal of Operational Research, 177(3), 1734-1750.

Chang, H., Chang, W.-J., Das, S., & Li, S.-H. (2004). Health care regulation and the operating efficiency of hospitals: Evidence from Taiwan. Journal of Accounting and Public Policy, 23(6), 483-510.

Diboky, F., & Ubl, E. (2007). Ownership and Efficiency in the German Life Insurance Market:A DEA Bootstrap Approach.

Ditzel, E., Štrach, P., & Pirozek, P. (2006). An inquiry into good hospital governance: A New Zealand-Czech comparison. Health research policy and systems(4:2).

Efron, B., & Tibshirani, R.J. (1993). An Introduction to the Bootstraps. London: Chapman and Hall.

Eisenberg, T., Sundgren, S., & Wells, M.T. (1998). Larger board size and decreasing firm value in small firms. Journal of financial Economics, 48(1), 35-54.

Eldenburg, L., Hermalin, B.E., Weisbach, M.S., & Wosinska, M. (2004). Governance, performance objectives and organizational form: evidence from hospitals. Journal of Corporate Finance, 10(4), 527-548.

Emrouznejad, A., Parker, B.R., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. Socio-Economic Planning Sciences, 42(3), 151-157.

Fama, E.F., & Jensen, M.C. (1983). Separation of ownership and control. Journal of law and economics, 26, 301-349.

Färe, R., Grosskopf, S., & Lovell, C.A.K. (1994). Production Frontiers. Cambridge: Cambridge University Press.

Kittelsen, S.A.C., J. Magnussen, Anthun, K.S., Häkkinen, U., Linna, M., Medin, E., . . . Rehnberg, C. (2008). Hospital productivity and the Norwegian ownership reform – A Nordic comparative study Working paper 2008. Oslo: University of Oslo.

Kooreman, P. (1994). Nursing home care in the Netherlands: A nonparametric efficiency analysis. Journal of Health Economics, 13(3), 301-316.

McKee, M., & Healy, J. (2002). Hospitals in a changing Europe. Buckingham.

Mutter, R.L., & Rosko, M.D. (2008). The impact of ownership on the cost-efficiency of U.S. hospitals. In J.L.T. Blank & V.G. Valdmanis (Eds.), Evaluating hospital policy and performance: contributions from hospital policy and productivity research (Vol. 18, pp. 113-138). Oxford: Elsevier.

Ozcan, Y.A. (2008). Health care benchmarking and performance evaluation. New York: Springer.

Pilyavsky, A.I., Aaronson, W.E., Bernet, P.M., Rosko, M.D., Valdmanis, V.G., & Golubchikov, M.V. (2006). East-West: Does It Make a Difference to Hospital Efficiencies in Ukraine? Health Economics, 15(11), 1173-1186.

Puenpatom, R.A., & Rosenman, R. (2008). Efficiency of Thai Provincial Public Hospitals during the Introduction of Universal Health Coverage Using Capitation. Health Care Management Science, 11(4), 319-338.

Rodríguez-Álvarez, A., & Lovell, C.A.K. (2004). Excess capacity and expense preference behaviour in National Health Systems: an application to the Spanish public hospitals. Health Economics, 13, 157-169.

Saltman, R.B., Bankauskaite, V., & Vrangbæk, K. (2007). Decentralization in health care. Berkshire.

Simar, L., & Wilson, P.W. (1998). Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. Management Science, 44(1), 49-61.

Simar, L., & Wilson, P.W. (2000a). A general methodology for bootstrapping in nonparametric frontier models. . Journal of Applied Statistics, 27(6), 779-802.

Simar, L., & Wilson, P.W. (2000b). Statistical Inference in Nonparametric Frontier Models: The State of the Art. Journal of Productivity Analysis, 13, 49-78.

Simar, L., & Wilson, P.W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 136(1), 31-64.

Staat, M. (2006). Efficiency of Hospitals in Germany: A DEA-Bootstrap Approach. Applied Economics, 38(19), 2255-2263.

Walford, V., & Grant, K. (1998). Improving health efficiency. London: DFID Health Systems Resource Centre.

Wilson, P.W., & Carey, K. (2004). Nonparametric Analysis of Returns to Scale in the US Hospital Industry. Journal of Applied Econometrics, 19(4), 505-524.

Yermack, D. (1996). Higher market valuation of companies with a small board of directors. Journal of financial Economics, 40(2), 185-211.

Productive Innovations in Hospitals

Productive Innovations in Hospitals: An Empirical Research on the Relation between Technology and Productivity in the Dutch Hospital Industry

This chapter has been published as:

Blank, J.L.T., & van Hulst, B.L. (2009). Productive Innovations in Hospitals: An Empirical Research on the Relation between Technology and Productivity in the Dutch Hospital Industry. Health Economics, 18(3), 665-679.

4.1 Introduction

There is a large body of literature on the efficiency and productivity of hospitals (see e.g. Grosskopf et al. (2001), Hofmarcher et al. (2002)), Prior and Sola (2000), Gruca and Nath (2001), Athanassopoulos et al. (1999), Hollingsworth et al. (1999), Rosko (1999), Blank and Merkies (2004), Rodríguez-Álvarez and Lovell (2004), Dismuke and Sena (1999), Sommersguter-Reichmann (2000), Blank and Valdmanis (2008), Linna et al. (2006) and Li and Rosenman (2001a). Most studies focus on the effects of environmental pressures on hospital efficiency, such as payment systems, competition, and property rights. Other studies pinpoint their attention on economic phenomena, such as economies of scale, economies of scope, economic behaviour, and expense preference. Less attention is paid to the influence of technological developments and innovation. However, in particular in the hospital industry major technical changes can be expected (see e.g. Maniadakis et al. (1999), Okunade (2001), Blank and Vogelaar (2004), McCallion et al. (2000) and Baker and Spetz (1999). From these studies, the developments that are particularly interesting include innovations that may be either saving or pushing up costs. A clear insight in the relationship between technology and cost may provide policymakers with pertinent information that could influence long term cost growth by controlling the availability and diffusion of new technologies.

Studies of hospital efficiency and productivity often neglect explicit treatment of innovations. This raises three issues. The first issue focuses on the measurement of innovations. The second relates to the appropriate choice of specifying productivity and the third issue concerns the extent to which innovations are affecting outcomes in productivity. We therefore suggest using a set of technology index numbers, and estimate and test a cost model to explain productivity differences from innovations. The models are applied to a dataset of Dutch general hospitals that operated during the period 1995-2002.

The chapter is organised as follows: in Section 2 we discuss the Dutch hospital industry. Special attention is paid to the characteristics of the Dutch hospital industry. The appropriate economic model is specified in Section 3. It is argued that considering relevant characteristics of a cost function model is essential in determining the most appropriate model. We also address approach used for specifying innovations. In Section 4, we summarise the data and in Section 5 the econometric estimation procedure and the estimates are given. The last section closes by briefly summarising and drawing conclusions for further analysis.

4.2 Dutch hospital industry

In this section we describe the Dutch general hospital industry. These hospitals comprise about 80% of hospital beds and almost 70% of the cost of the Dutch hospital sector. The remainder of total hospital cost is absorbed by academic and specialty hospitals (such as eye clinics and rehabilitation clinics). In order to give an impression of the size of the Dutch hospital sector Table 4-1 contains some quantitative characteristics of general hospitals between 1995 and 2002.

Table 4-1 Characteristics of the Dutch hospital industry, 1995 and 2002

	1995	2002
Number of general hospitals	109	89
First-time visits (x 1,000)	5,491	7,462
Discharges (x 1,000)	1,353	1,329
Day care patient days (x 1,000)	568	888
Inpatient days (x 1,000)	12,619	9,627
Total cost (x € mln.)[a]	5,504	8,553
Variable cost (x € mln.)[a]	4,578	7,515
Capital cost (x € mln.)[a]	927	1,038
Personnel (x 1,000 ftes)[b]	100	112

a, € 1 ≈ $ 1 (exchange rate 2002).
b, including physicians.
c, in 1993 a fte is 38 working hours a week, in 2000 this is 2 hours less due to a shortening of working hours.
Source: Prismant, Statistics Netherlands.

In 1995 there were 109 general hospitals in the Netherlands. As a result of closures and mergers of hospitals, the number of general hospitals decreased to 89 in 2002. This was accompanied at the same time by an increase in the size of hospitals. In 1995 the hospitals "produced" 5.5 million first-time visits, and over 1.3 million hospital discharges. In 2002 the number of outpatients had increased to 7.5 million but with a slightly decreased number of discharges, increasing the number of outpatient treatments. In 1995, inpatient days in general hospitals numbered 12.6 million with an additional 600,000 day care patient days were produced. In 2002 the number of inpatient days had decreased substantially. The total cost of general hospitals in 1995 equalled € 5.5 billion. Sixteen percent of total cost was spent on capital inputs, the remaining 84% on personnel and material supplies and is seen as variable cost.

In 2002 costs had risen by more than 50% in nominal terms (30% in real terms) with an increase in the share of variable cost. Between 1995 and 2002 the number of FTE employees in hospitals rose by 12% to 112,000. As a result of the hospital consolidation the number of physicians per hospital increased substantially (not in Table), indicating potential opportunities for more specialization within hospitals. A minority of physicians were employed by general hospitals; the others were self-employed, but associated with a particular hospital.

In addition to the description of the Dutch hospital industry above, there are some special characteristics. Capacity is regulated by the central government as well as fully reimbursed by the central government on a prospective basis. Budgets consist of a fixed component related to capacity and a variable component related to production. The fixed component is based on the so-called adherence (the number of patients potentially using the hospital), the number of beds and the number of associated physicians. The production related component is based on regional agreements on the numbers of first-time visits, in-patient days, day-care patient days and the number of discharges. To some extent budgets are based on the severity of cases, since larger

hospitals, which are assumed to treat more severe cases, receive higher budget rates per case. Some budget rates depend on the types of specialties supplied by the hospital, also indicating differences in care need. The hospitals only receive this budget virtually; they have to earn their revenues by producing medical treatment or procedures. For each medical procedure a price is fixed by the Central Tariffs Health Care, and this price is paid by the insurance companies.

If total revenues of the hospital exceed the budget, this is balanced in the next time period. Thus, in the long run revenues always coincide with the budget. A surplus in the operating results can only be generated by keeping expenses (cost of production) lower than the budget. Although it is not allowed to make profits, surpluses do remain available to the hospital and are added to the capital assets. Since wages are regulated, the possibilities for management to increase their own salaries are very limited. Hence, not-for-profit cannot be a matter of 'for-profit in disguise'. However, incentives such as maintaining market share exist and may be realized in terms of non-price, i.e. quality measures. However, if general hospitals have deficits and negative capital assets they will be subjected to budget cutbacks and finally closed down. Through closures of inefficient hospitals the average efficiency of the health care sector can be improved.

Another important feature of the Dutch hospital sector is that hospitals cannot choose their patients. Patients are referred to a hospital by general practitioners. They choose a hospital with a convenient location compared to other hospitals and based on availability of the appropriate specialties. Hospitals are obliged to treat any patient presented to them, provided that they have the medical knowledge required for the treatment. In practice hospitals can attract patients by supplying particular specialties or a high quality of care. This implies that expansion of high tech medical treatments may be another goal.

Since capital is also strongly regulated and some of the innovations are connected with housing and medical equipment technology diffusion is also affected by regulation.

4.3 Economic model

Since in the Netherlands service prices are regulated by a semi-governmental organisation, (Central Tariffs Health Care) services prices are assumed to be exogenous. Resource prices, for instance wages for nursing personnel, are determined by market conditions and can also be assumed to be exogenous. Yet the theory of the firm is not fully applicable, as Dutch hospitals are in general not allowed to close a production line completely. Usually it is even argued that production is exogenous to Dutch hospitals because they are not allowed by law to refuse patients requiring treatment. Capital inputs are also exogenous to hospital management. In such settings, the economic model of (variable) cost minimising subject to a technology constraint is probably the most appropriate one.

Clearly the model thus obtained is equivalent to the direct cost function model:

$$VC = c(y, w, K, t_1, \dots, t_N) \tag{1}$$

Following Shephard's Lemma we derive the cost share equations from the cost function:

$$S_i = \frac{\partial \, ln \, VC}{\partial \, ln \, w_i} \tag{2}$$

With:

VC = variable cost;
y = services delivered;
w = resource prices;
K = capital input;
t_i = technology at time period i.

In applications a (translog) specification of the cost function $c(y, w, K, t_1, \dots, t_N)$ is acceptable if it satisfies the following conditions:
(i) monotonic non-decreasing in input prices;
(ii) concave in input prices.

Of particular interest is modelling the influence of the technology. In general the first- and second-order time trends in economic models tend to dominate, producing a smooth and slowly changing characterisation of the pace of technical change. However, from other studies we also know that the introduction of new technologies and innovations show highly variable rates of adaptation, as suggested by Kopp and Smith (1983). However, Baltagi and Griffin (1988)) advocated another process for estimating a general index of technical change within the context of a quite general production technology. Their procedure yields a general index that may be both non-neutral and scale augmenting. The technology index is a weighted sum of time dummies.

In all the above mentioned studies, technical change is measured by a proxy, namely a time trend of a series of year dummies, instead of a variable that actually measures the technology used. In practice however, innovations slowly spread over all hospitals in the sector and so different hospitals are operating under different technologies at the same point in time.

In this study we therefore suggest a more explicit measurement of technical change. We inventory general and well known innovations in the Dutch hospital industry over the past ten years, such as specialised mamma clinics (clinics for women with breast cancer, with integrate medical treatment,

nursing support, social work and counselling), and add them to the cost function. Very specific technologies for very specific patient groups, which can only be applied in a limited number of hospitals, are not included in the analysis (in the Data section we present the entire list). So the cost function is shifted when a different technology is used. By estimating the parameters of the cost function we simultaneously measure the influence of these innovations on cost and how they contribute to productivity.

We operate on the assumption that the number and types of patients directed to each hospital is exogenous, on the basis of the requirement that hospital care is provided for whichever patients are sent to them. Since we have defined rather general types of innovation we have diminished the eventual problem of selectivity bias. In case of very specific technologies the hospital might attract certain types of patients. Here we assume that the innovations only affect the way patients are treated, the way the medical and administrative process is organised and for the way the hospital is being managed. We do not assume that the innovations affect the composition of the types of patients.

The decision to adopt or not adopt a new technology is likely be the result of a number of factors, one of them being the effect on productivity of the hospital. This might complicate the interpretation of the empirical outcomes, since technology is not strictly endogenous in our model. In particular, efficient hospitals may generate extra resources to adopt new technologies and the estimation may reveal spurious correlations. However, once the technology has been implemented the technology is exogenous in the years to come. Taking this endogeneity into account will complicate the model substantially. Therefore we abstract from this problem and assume that the presence of the technology is exogenous.

4.4 Specifying innovations

According to Spetz and Maiuro (2004) measures of technologies are limited in a number of ways. They state that measurement must be foremost driven by the research question at hand. There is no one-size-fits-all solution. In case of cost function estimation or efficiency measurement an aggregate index of single technologies is to be preferred. In line of their suggestion we therefore introduce our concept of technology measurement based on three notions of innovations:

- Single (or individual) innovations;
- Clusters of innovations;
- Innovation index.

Examples of single innovations are specialised mamma clinics or specialised chronic obstructive pulmonary disease (COPD) nurses. Innovations are present or not and therefore measured by a set of dichotomous variables $T = \{t_1, \dots, t_N\}$. A large number of innovations are strongly related. Aside from the specialised mamma clinics, hospitals also include specialized clinics for people with a sleeping disorder, pain relief etc. These related innovations are aggregated into one group and designated as a cluster of innovations, for instance the "specialised clinics" cluster. The corresponding value to an innovation cluster L is an aggregation function of corresponding single technologies $a(T_L)$ of a subset of all technologies T, say $T_L = \{t_i | \ i \ belongs \ to \ technology \ cluster \ L\}$.

Spetz and Maiuro (2004) also point at the drawbacks of using a general technology index. The index does not distinguish between various heterogeneous technologies. To avoid this problem we introduce the concept of technology clusters with rather homogenous technologies. As mentioned earlier, we further diminish the heterogeneity problem by selecting rather general technologies and excluding academic and so called "top clinical" hospitals from the analysis.

We distinguish two aggregation functions. In the first function the value of the function equals the unweighted sum of the corresponding technologies of T_L.

$$a_h(T_L) = \sum_{i \in L} t_{hi} \tag{3}$$

With:
$a_h(T_L)$ = number of innovations present in cluster L for hospital h;
t_{hi} = technology i present in hospital h.

The second aggregation function is based on a concept of Baker and Spetz (1999), referred to as the Saidin index, which is a weighted sum of various technologies in a base year. Each weight reflects the percentage of hospitals that do not possess the technology or service in a base year. For example, technologies that were rather rare at the beginning of the period —whether they are rare because they are new, expensive, or difficult to implement— receive higher weights in this measure. Technologies that are common receive low weights. This weighting scheme corresponds with most people's idea of what defines "high technology": that which is rarely found, whether it is rare due to newness, expense, or difficulty of operation. When a technology becomes common, it is no longer perceived as being of a high level. To ensure a consistent comparison over time we define indices using a set of technologies and weights that are defined in a base year and held fixed for subsequent years.

$$a_h(T_L) = \sum_{i \in L} r_i t_{hi} \tag{4}$$

With:
$a_h(T_L)$ = index of innovations present in cluster L for hospital h;
r_i = share of hospitals not possessing technology i in a base year;
t_{hi} = technology i present in hospital h (1= present; 0 = not present).

The index has two properties. First, it accurately reflects the degree of technology advancement across hospitals at a single point in time. That is, in any given year, hospitals with higher values of the index are "more advanced".

Adding technologies will increase the index value, especially if the technologies are relatively rare rather than more common. In general, hospitals that have more, rarer technologies will have higher index values than hospitals with fewer, more common technologies.

The second property of the index is the ability to identify changes in technology over time. That is, the index increases over time with increases in the degree of technology advancement. If a hospital has a higher index value this year than last year, we may conclude that the hospital has become more advanced.

4.5 The Data

General

Data for this study covering the period 1995-2002, was obtained from the Ministry of Health, Welfare and Sport and from a separate survey amongst hospitals based on a questionnaire about innovations. The financial, patient and personnel data were collected by the Institute for Health Care Management. The surveys contain information on almost all general hospitals yielding approximately 100 observations each year, situated in 27 health care regions. The data on innovations were collected by ECORYS and the Public Health Council. This survey contains information on 63 innovations from 66 general hospitals over the period 1995-2004. For the purposes of this study, observations on hospitals with missing or unreliable data were excluded from the dataset. Various consistency checks were performed on the data to ensure that changes in average values and the distribution of values across time were not excessive. After eliminating observations containing inaccurate or missing values in the dataset, an unbalanced panel data set of 362 observations over the 8 years of study remained.

Since we have data on several variables over a complete set of hospitals we are able to investigate the representativeness of the sample with respect to

these variables. We have analysed whether a hospital is or is not present in the sample with respect to these variables. The most appropriate statistical method for analysing a dichotomous dependent variable is a logit analysis. Explanatory variables are the size, productivity and type of hospital (general, top clinical and university hospital). The outcomes of the logit analysis show that, based on t-statistics at the 5% significance level, none of these characteristics "explain" the presence/no presence in the sample. In other words, the presence/no presence in the sample is random, at least not depending on one the independent variables. We conclude that, based on these characteristics, the sample is representative.

Production

The main service delivery of hospitals is treating patients. The production of hospitals is therefore measured by the number of discharges and outpatients (not followed by an admission)[2]. The discharges have been separated into over 30 medical specialties in order to measure case-mix. Since it is not possible to use such a large number of categories, these have been aggregated into four categories on the basis of average stay homogeneity and the distinction between surgery/non-surgery specialties.

We distinguish therefore the following groups of specialties:

- Non-surgery with average stay less than 4 days;
- Non-surgery with average stay more than 4 days;
- Surgery with average stay less than 4 days;
- Surgery with average stay more than 4 days.

Although four types of discharges and outpatient explain a very large part – as we shall see later – of variations in cost, services are much more nuanced than just the number of outpatients and discharges. The health outcomes of

[2] If an outpatient was admitted to the hospital later that year, we count this as an admission and not as an outpatient.

patients seem to be a particularly important component of hospital production. A new treatment technology that increases the costs of treatment could also improve health outcomes quite a bit. More research and data on quality of production are therefore important. Nevertheless it seems reasonable to assume that new technologies do not decrease quality and the estimates of productivity can be regarded as a lower bound.

Resources

Resources include staff, administrative and maintenance personnel (including security and cleaning), nursing personnel, paramedical personnel (such as lab technicians), material supplies and capital. Physicians are not included in these personnel variables, to ensure that hospitals with physicians on their payroll and hospitals with physicians who are self-employed are treated equally. The costs of physicians (wages) are not included in the cost or price variables either.

Material supplies include such aspects as medical supplies, food and heating. Personnel and material supplies are treated as variable resources since the hospital can change these in the short run. Capital is included as a fixed resource, since the capital assets such as buildings and medical equipment can only be changed in the long run.

There are data on the costs and the quantity for each resource personnel category. For each region and time period wages are defined as the average wage per full time equivalent. This is considered as the market price for labour; qualitative differences between hospitals are included in the volume of labour.

Since there is no natural unit of measurement for material supplies, a circumventing construction was used. The price of material supplies is a weighted index based on components of the consumer index calculated for the Netherlands by Statistics Netherlands. The weights are derived from cost shares.

Innovations

Table 4-2 includes a complete list of technologies corresponding to each type of innovation. These innovations are clustered the following seven primary types of innovations:

- multidisciplinary diagnostics and treatment (14);
- technical (medical) quality (14);
- nursing consulting hours (13);
- chain care (11);
- logistic optimisation (5);
- hospital transferred care (4);
- information and communication technology (3).

Table 4-2 List of innovations

Multidisciplinary diagnostic and treatment	Technical quality	Nurse consulting hour
Pelvis policlinic	Laparoscopic gallbladder removement	COPD nurse
Diabetes foot policlinic	Laparoscopic intestine neoplasm section	CVA consultant
Mamma policlinic	Laparoscopic kidney removement	Decubitus nurse
Constipation and wee-wee policlinic	Seal equipment at intestine surgery	Diabetes nurse
Mother child unit	MRI instead of muelografics	cardiac nurse
Proctologic policlinic	Shaver blades at endonasal surgery	Mamma care nurse
Vascular or risk policlinic	Stroke care unit	MS nurse
Cardiac policlinic	Thermo therapy gynaecology	Stoma nurse
Pain policlinic	TVT devices	Wound consultant
Sleep disorder policlinic	Preoperative nutrition	Rheumatic consultant
Lung revalidation	Decubitus prevention	Oncology consultant
Down policlinic	Pre Operative screening by anaesthetist	Function differentiation
Protocol of reference by general practitioner	(Postoperative) pain registration	Other innovation
Other innovation	Other innovation	

Chain care	Logistics	ICT
Stroke service	Cataract line	Process support ICT
Total knee (reduction of hospital stay duration)	Filtering of patients (elective, emergency/ focussed care)	Electronic data consultation room & ward
Total hip (reduction of hospital stay)	One stop visit (MRI, varicose, Hernia)	Other innovation
Integrated psycho geriatric care	Joint care for orthopaedics	
Integrated diabetes care	Other innovation	
Integrated COPD care		
Transmural care for oncology patients	**Outside hospital care**	
Transmural care for palliative care	Home monitoring of pregnancy	
Cooperation with general practitioner(f.aid)	Self-measurement thrombotic care	
Transmural care	Night home dialysis	
Other innovation	Other innovation	

For calculation of the Saidin index we defined a list of technologies available in 1990, and determined their relative rarity in this year, and then computed index values for all hospitals in all years using the 1990 list and the 1990 weights. Descriptive statistics of the variables are given in Table 4-3.

In order to simplify the interpretation of the estimated parameters all variables in the analysis are standardized at their arithmetical means. The first-order parameter estimates represent the elasticity of cost with respect to the corresponding service, resource price or fixed input for the "average" firm.

Table 4-3 Descriptive Statistics, Dutch General Hospitals 2002 (N=66)

Variable	Mean	Standard dev.
Discharges 1	8,360	3,891
Discharges 2	6,210	3,012
Discharges 3	5,329	3,016
Discharges 4	5,895	3,223
Outpatients	65,769	31,189
Price auxiliary personnel	1.00	0.04
Price nursing personnel	1.00	0.02
Price paramedical personnel	1.00	0.03
Price material supplies	1.00	0.00
Cost (x € million)	102	88
Cost share auxiliary personnel	0.20	0.02
Cost share nursing personnel	0.28	0.04
Cost share paramedical personnel	0.18	0.03
Cost share material supplies	0.34	0.03
Multidisciplinary diagnostics and treatment	5.62	2.62
Technical (medical) quality	7.21	3.00
Nursing consulting hours	8.03	2.11
Chain care	5.87	2.02
Logistic optimisation	1.25	1.14
Hospital transferred care	0.77	0.82
Information and communication technology	1.13	0.92

4.6 Estimation and Evaluation

Estimation of parameters

In the Dutch context, as we reasoned in Section 3, a model based on cost minimizing behaviour and services, resource prices and capital inputs as exogenous variables, is the most appropriate one. Accordingly, we estimate a direct cost function model. The cost function model constitutes a system with a cost function and a number of cost share equations (see appendix for a full description).

The specification of the direct cost function model contains five output quantities, four input prices, one fixed input quantity, and a technology index based on innovations. These variables are discussed in the previous section. In the direct cost function model the four input quantities are endogenous. Their optimal values are derived from the four share equations obtained by applying Shephard's lemma to the cost equation.

The cost function is specified as a translog function and the share equations are derived from it. Homogeneity of degree one in prices and symmetry is imposed by putting constraints on some of the parameters to be estimated (see Appendix).

The models are estimated as multivariate regression systems using various equations with a joint density, which we assume to be normally distributed. Because disturbances are likely to be cross-equation-correlated, Zellner's Seemingly Unrelated Regression method is used for estimation (Zellner, 1962). As usual, because the shares add up to one causing the variance-covariance matrix of the error terms to be singular, one share equation in the direct cost function model is eliminated.

Evaluation and testing

We have estimated each model. We evaluate the models based on:

- standard statistical properties such as R2 and T-values;
- theoretical requirements (monotonicity and concavity, see Section 3);
- a formal likelihood ratio test;
- economic properties such as productivity change due to innovations.

For the sake of space and since the estimates do not differ substantially between the models we only present the estimates of the most appropriate model, which was chosen by conducting some formal tests.

First of all, we test the nature of technical change. We distinguish 4 models, each with a different type of technical change. Model I represents a disembodied technical change. It is also the most parsimonious model and only includes the technology index A(tech). Model II assumes output biased technical change: aside from A(tech) the model also includes cross terms with produced services. Model III is an input biased technical change model and includes aside from A(tech) cross terms with resource prices. Model IV is an input and output biased model and contains aside from A(tech) cross terms with services as well as resource prices. It's quite clear that model I is nested in models II, III and IV and models II and III are nested in model IV.

We also test the hypothesis that the models excluding the technology index perform as well as the models including the technology index. In other words, we test whether or not a time trend variable is a good proxy for measuring differences in technology.

Similarly we test the hypothesis that the models with a technology index excluding the time trend variable perform as well as the models with a technology index including a time trend variable. Here we actually investigate whether the distinct technology clusters cover technical changes through time adequately, or whether there are remaining (unmeasured) technical changes affecting resource usage.

We also test the difference between the models based on an unweighted sum of innovations and models based on a weighted sum of innovations with weights depending on rarity.

Table 4-4 presents the log likelihoods of the various models.

Table 4-4 Log likelihoods various models (N=362)

Variable	No trend	Neutral	Output Biased	Input Biased	Input, Output Biased
Unweighted index, including trend	3,229	3,258	3,269	3,262	3,275
Trend only	3,229	3,235	3,243	3,239	3,247
Saidin, including trend	3,229	3,254	3,270	3,261	3,276

From Table 4-4 we conclude that the most appropriate model is the output biased model, including a time trend variable and a technology index based on an unweighted sum of underlying technologies. The likelihood of the various models shows that the fit of this model is clearly better than the other models. Deriving the likelihood ratios and the usage of a critical value of 0.025, all mutual model tests favour the output biased model with a trend and an unweighted technology index[3]. The model with the rarity index does not perform better than the model with the unweighted sum of innovations. So there is no explanatory power coming from this rarity approach in contrast with the study by Baker and Spetz (1999).

Table 4-5 presents the parameter estimates of the output biased model with unweighted technology indices.

[3] Strictly speaking, the models with the underlying technologies based on the unweighted sum and the Saidin index are not nested and can therefore not be tested by a likelihood ratio test. However, both models can be seen as restricted models of a model in which the parameters of all individual technologies are being estimated.

Table 4-5 Parameter estimates model with output biased technical change

	Variable	Estimate	T-value
A0	Constant	-0.079	-4.796
A1	Multidisciplinary diagnostics and treatment	0.001	0.159
A2	Technical (medical) quality	0.027	3.322
A3	Nursing consulting hours	0.024	3.355
A4	Chain care	-0.010	-1.465
A5	Logistic optimisation	0.006	0.493
A6	Hospital transferred care	0.078	2.917
A7	Information and communication technology	-0.051	-2.929
A8	Time	-0.019	-5.094
B1	Discharges group 1	0.189	2.887
B2	Discharges group 2	0.067	0.854
B3	Discharges group 3	0.183	4.376
B4	Discharges group 4	0.058	1.015
B5	Outpatients	0.510	8.190
B11	Discharges group 1 * discharges group 1	0.530	2.184
B12	Discharges group 1 * discharges group 2	0.609	3.058
B13	Discharges group 1 * discharges group 3	-0.193	-1.362
B14	Discharges group 1 * discharges group 4	-0.272	-1.673
B15	Discharges group 1 * outpatients	-0.674	-3.894
B22	Discharges group 2 * discharges group 2	-0.757	-2.147
B23	Discharges group 2 * discharges group 3	0.235	1.365
B24	Discharges group 2 * discharges group 4	-0.368	-2.150
B25	Discharges group 2 * outpatients	0.188	1.109
B33	Discharges group 3 * discharges group 3	-0.106	-0.802
B34	Discharges group 3 * discharges group 4	0.087	0.969
B35	Discharges group 3 * outpatients	-0.056	-0.520
B44	Discharges group 4 * discharges group 4	0.230	1.521
B45	Discharges group 4 * outpatients	0.221	1.643
B55	Outpatients * outpatients	0.594	3.838
C1	Price auxiliary personnel	0.198	112.066
C2	Price nursing personnel	0.297	160.733
C3	Price paramedical personnel	0.178	83.647
C4	Price material supplies	0.326	149.074
C11	Price auxiliary personnel * price auxiliary personnel	-0.022	-0.829
C12	Price auxiliary personnel * price nursing personnel	-0.064	-3.118
C13	Price auxiliary personnel * price medical personnel	0.133	6.090

	Variable	Estimate	T-value
C14	Price auxiliary personnel * price material supplies	-0.047	-2.460
C22	Price nursing personnel * price nursing personnel	0.059	2.240
C23	Price nursing personnel * price medical personnel	-0.018	-0.815
C24	Price nursing personnel * price material supplies	0.023	1.327
C33	Price medical personnel * price medical personnel	-0.015	-0.458
C34	Price medical personnel * price material supplies	-0.100	-4.362
C44	Price material supplies * price material supplies	0.123	4.836
E11	Discharges group 1 * price auxiliary personnel	0.002	0.309
E12	Discharges group 1 * price nursing personnel	-0.016	-3.172
E13	Discharges group 1 * price medical personnel	0.019	2.771
E14	Discharges group 1 * price material supplies	-0.004	-0.706
E21	Discharges group 2 * price auxiliary personnel	-0.024	-3.789
E22	Discharges group 2 * price nursing personnel	0.005	0.912
E23	Discharges group 2 * price medical personnel	0.016	2.054
E24	Discharges group 2 * price material supplies	0.002	0.347
E31	Discharges group 3 * price auxiliary personnel	0.015	3.496
E32	Discharges group 3 * price nursing personnel	0.004	0.920
E33	Discharges group 3 * price medical personnel	0.000	0.062
E34	Discharges group 3 * price material supplies	-0.019	-3.856
E41	Discharges group 4 * price auxiliary personnel	0.001	0.129
E42	Discharges group 4 * price nursing personnel	0.004	0.902
E43	Discharges group 4 * price medical personnel	-0.015	-2.643
E44	Discharges group 4 * price material supplies	0.010	2.048
E51	Outpatients * price auxiliary personnel	-0.006	-1.318
E52	Outpatients * price nursing personnel	-0.002	-0.382
E53	Outpatients * price medical personnel	-0.006	-0.974
E54	Outpatients * price material supplies	0.013	2.450
I11	Technology * discharges group 1	-1.331	-1.545
I12	Technology * discharges group 2	4.342	3.536
I13	Technology * discharges group 3	-0.727	-1.200
I14	Technology * discharges group 4	1.447	2.266
I15	Technology * outpatients	-1.872	-2.261

R^2 cost equation	0.97	LR test cost equation (df=55)	1424.4
R^2 cost share nursing personnel	0.08	LR test cost share nursing personnel (df=9)	27.9
R^2 cost share medical personnel	0.16	LR test cost share medical personnel (df=9)	91.7
R^2 cost share material supplies	0.13	LR test cost share material supplies (df=9)	55.6

Table 4-5 shows that in a statistical sense the cost function model fits the data rather well. Results derived from this cost function are plausible. The cost equation has a high R2, i.e. 0.97. About 60% respectively of the estimated parameters are significant at the 5% level. Most R2 s of the share equations are reasonable. The requirements on monotonicity and concavity are also fulfilled to a large extent. The monotonicity property tells us that input demand is always positive, which is the case for all observations and in particular for the "average" hospital[4]. A necessary condition for concavity is the negativity of the "own" elasticity's of substitution. This condition also holds for all observations and the "average" hospital. However, the condition of negative semi-definite of the matrix of elasticity's of substitution only holds for 20% of the observations[5]. For the "average" hospital one of the eigenvalues is slightly positive (=0.012). We also tested the "significance" of each equation in the system separately by imposing the restriction that all the parameters (except the constant) equal zero. Based on likelihood ratio tests all the null hypotheses were overwhelmingly rejected.

From Table 4-5 we conclude that changes in technology affect cost. Innovations in multidisciplinary diagnostics and treatment, technical medical quality, nursing consulting hours, logistic optimisation and hospital transferred care require more resources, whereas chain care and information and communication technology are cost reducing. The t-values show that from the cost enhancing innovations only the effects of innovations in technical quality, nursing consulting hours and hospital transferred care are significant at the 5% level; from the cost reducing innovations only information and communication technology is significant. Other non-measured technical changes approximated by the time trend variable have a significant cost reducing effect. These

---

[4] The average hospital is a fictitious hospital with values for all variables set at the sample (arithmetic) mean.

[5] For most observations there is only one eigenvalue slightly greater than zero.

outcomes were recognized as plausible by the Dutch Council for Public Health and Health Care and were reported in a consulting document to the Dutch minister of Health care[6].

The i_{1m} parameters measure the effect of technology index on the (marginal) cost of specific products. Since the interpretation of the technology index is slightly unclear –it's an unweighted sum of underlying technologies- we present the separate effects of the technology clusters by calculating:

$$i_{1m}a_k = i_{1m} \cdot a_k \tag{5}$$

With $i_{1m}a_k$ being the product specific cost flexibility for product m of technology cluster k. Table 4-6 presents the results of the specific cost flexibilities of each technology cluster.

From Table 4-6 we note that the marginal cost is affected by the technology used. For example, technical (medical) quality affect discharges 2 and discharges 4 in a significantly positive way. In other words, technical (medical) quality make discharges 2 and 4 more expensive to produce. On the other hand, outpatient visits become significant less expensive when more of these technologies are utilised. So the results are rather ambiguous. However, it seems that hospital transferred care in general is a cost pusher, whereas ICT is a cost saver.

[6] The Council for Public Health and Health Care is the independent body which advises the government on public health and care. The Council consists of nine members, each member has her or his own particular expertise and background: a doctor, a hospital director, a managing director, an ethicist and so on.

Table 4-6 Product specific cost flexibility

Estimate (T-value)	Discharge type 1	Discharge type 2	Discharge type 3	Discharge type 4	Outpatient
Multidisciplinary diagnostics	-0.001	0.005	-0.001	0.002	-0.002
	(-0.16)	(0.16)	(-0.16)	(0.16)	(-0.16)
Technical (medical) quality	-0.04	0.12	-0.02	0.04	-0.05
	(-1.46)	(3.06)	(-1.19)	(2.04)	(-2.31)
Nursing consulting hours	-0.03	0.10	-0.02	0.03	-0.05
	(-1.40)	(3.01)	(-1.24)	(2.01)	(-2.28)
Chain care	0.01	-0.04	0.01	-0.01	0.02
	(1.10)	(-1.45)	(1.03)	(-1.31)	(1.30)
Logistic optimisation	-0.01	0.03	0.00	0.01	-0.01
	(-0.48)	(0.50)	(-0.47)	(0.49)	(-0.49)
Hospital transferred care	-0.10	0.34	-0.06	0.11	-0.15
	(-1.40)	(2.62)	(-1.12)	(1.73)	(-2.05)
ICT	0.07	-0.22	0.04	-0.07	0.10
	(1.43)	(-2.89)	(1.24)	(-1.88)	(2.07)
Time	0.03	-0.08	0.01	-0.03	0.04
	(1.56)	(-4.36)	(1.27)	(-2.32)	(2.61)

Figure 4-1 depicts average productivity growth as a result of technical change over time. These growth rates are computed by averaging the fitted technology index in year T for each hospital and comparing with the averaged fitted technology index assuming base year technology.

$$TC(t, t_0) = \left[\frac{\sum_h \hat{A}_{ht}}{\sum_h \hat{A}_{ht0}}\right]^{-1} \times 100 \tag{6}$$

With:

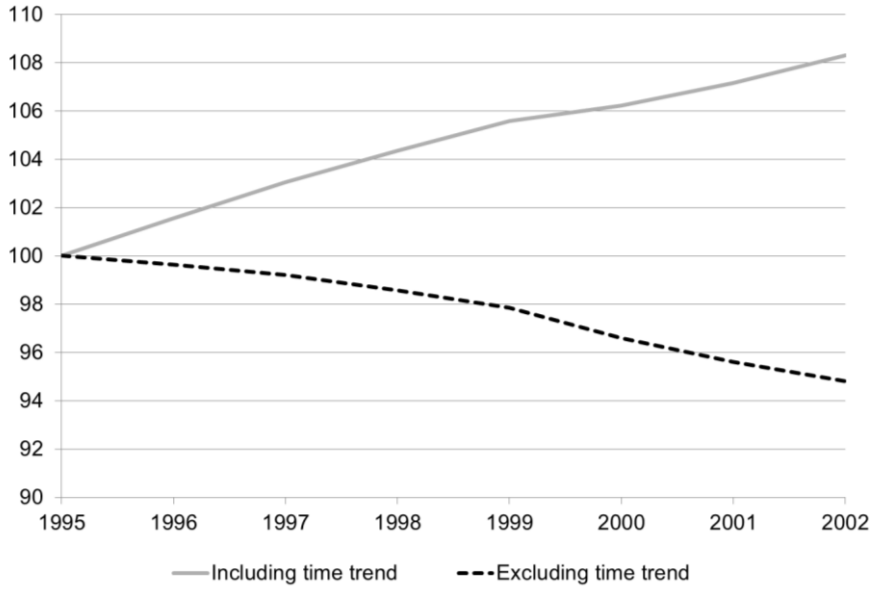
$\mathrm{TC}(\mathrm{t}, \mathrm{t}_0)$ = technical change year t related to base year t0;

$\widehat{\mathrm{A}}_{\mathrm{ht}}$ = technology at hospital h in year t.

Figure 4-1 includes two lines. The upper line represents productivity change according to the above formula. The lower line represents productivity

change purely based on the measured innovations. So this line excludes the impact from the time trend variable in the equation of A(tech).

Figure 4-1 Productivity growth technical change, including and excluding time trend



According to the upper line, Figure 4-1 shows a steady growth in productivity. In the seven year period productivity grows by more than 8%. This growth is dominated by the time trend variable in the estimates. When we exclude the time trend effect and recalculate productivity based on the parameters of the innovations we observe a different picture. There is technical regress and productivity slows down by 5% over the seven years period. All innovations under inspection here are in some way related to the "processing of patients". These innovations mainly influence medical procedures and treatments. The introduction of these technologies is probably not motivated by productivity reasons, but merely by quality reasons. The productivity growth has obviously been realized on the "input side". More qualified and trained

personnel, efficient working procedures, better IT in administrative procedures and outsourcing are likely to have increased productivity substantially.

Blank and Vogelaar (2004) also estimate technical change for Dutch general hospitals. They establish a substantial productivity growth as well, although for a slightly different time frame (1993-2000). In contrast with this study they conclude that technical change is input biased. This may support the idea that technical change is both input and output biased. In this study the input output biased model is not rejected at a critical value of 5 %. However, in Blank and Vogelaar (2004) the input output biased model is even rejected at a critical level of 10 %.

In other studies on Dutch hospital industry, based on earlier time frames, Blank and Eggink (2004) and Blank and Merkies (2004) report technical regress, respectively negligible technical change. Since time frames do not overlap between these two studies and this study it is difficult to draw any conclusions from this contradiction.

4.7 Summary and Concluding Remarks

This chapter studies the relationship between technology and productivity in Dutch hospitals. In most previous studies technical change is measured by a proxy, namely a time trend. In practice however, innovations slowly spread over all hospital, therefore different hospitals are operating under different technologies at the same point in time. In this study we explicitly inventory specific and well known innovations in the Dutch hospital industry over the past ten years. These innovations are aggregated into a limited number of homogenous innovation clusters, which are measured by a set of technology index numbers. The index numbers are included in the cost function specification and the estimation.

The estimates indicate that some technology (clusters) increase cost, while others reduce cost. Productivity gains are merely realised in the production process (e.g. ICT and chain care). Productivity "losses" are more connected with innovations in services, which can be regarded as product innovations. In general, these innovations are implemented for reasons of quality (in terms of better health outcomes or less stressful treatments for patients). Examples are innovations in medical procedures and treatments.

The outcomes also indicate that technical change is non-neutral and output biased. This means that the marginal cost of services is affected by the technology used. For example, technical (medical) quality make type 2 and type 4 discharges more expensive to produce. On the other hand, inpatient visits become less costly when more of these technologies are utilised. In general, hospital transferred care is a cost increasing innovation, whereas ICT is a cost saving innovation.

A methodological side result from this research concerns the use of a weighted technology index (the Saidin index) instead of an unweighted index in the model specification. It is shown that a model including the Saidin index does not perform better than the model including an unweighted technology index.

Appendix: Cost model specification

The cost function model consists of a translog cost function and the corresponding cost share equations:

$$\begin{aligned}\ln VC = a_0 &+ \sum_{m=1}^{M} b_m \ln y_m + \sum_{n=1}^{N} c_n \ln w_n + \sum_{o=1}^{O} d_o \ln z_o + \\ &+ \frac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M} b_{mm'} \ln y_m \ln y_{m'} \\ &+ \frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N} c_{nn'} \ln w_n \ln w_{n'} \\ &+ \frac{1}{2}\sum_{o=1}^{O}\sum_{o'=1}^{O} d_{oo'} \ln z_o \ln z_{o'} + \sum_{m=1}^{M}\sum_{n=1}^{N} e_{mn} \ln y_m \ln w_n \\ &+ \sum_{o=1}^{O}\sum_{n=1}^{N} f_{on} \ln z_o \ln w_n + \sum_{o=1}^{O}\sum_{m=1}^{M} g_{om} \ln z_o \ln y_m \\ &+ A(tech) \\ &+ \sum_{m=1}^{M} i_{1m} A(tech) \ln y_m + \sum_{n=1}^{N} j_{1n} A(tech) \ln w_n\end{aligned} \tag{7}$$

With:

VC = variable costs;

y_m = output m (m = 1,.., M);

w_n = price input n (n = 1,.., N);

z_o = fixed input o (o = 1,.., O);

$A(tech) = \sum_k a_k \, tech_k$ (technology indices);

$tech_k$ = number of innovations of type k;

$a_0; a_k; b_m; c_n; d_o; b_{mm'}; c_{nn'}; d_{oo'}; e_{mn}; f_{on}; g_{om}; h_0; h_1; i_{1m}; j_{1n}$ parameters to be estimated.

With Shephard's lemma the optimal cost share functions can be deducted:

$$S_n^0 = c_n + \sum_{n'=1}^{N} c_{n'n} \ln w_n + \sum_{m=1}^{M} e_{mn} \ln y_m + \sum_{o=1}^{O} f_{on} \ln Z_o + j_{1n} A(tech) \qquad n = 1,..,N \tag{8}$$

With: S_n^0 = optimal cost share input n (n = 1,.., N).

Homogeneity of degree one in prices and symmetry is imposed by putting constraints on some of the parameters to be estimated. In formula:

$$b_{mm'} = b_{m'm} \; ; \quad c_{nn'} = c_{n'n} \; ; \quad d_{oo'} = d_{o'o} \; ;$$

$$\sum_{n=1}^{N} c_n = 1 \; ; \qquad \sum_{n=1}^{N} c_{nn'} = 0 \;\; (\forall n') \; ; \qquad \sum_{n=1}^{N} e_{mn} = 0 \; (\forall m) \; ; \tag{9}$$

$$\sum_{n=1}^{N} j_{1n} = 0 \; ; \qquad \sum_{n=1}^{N} f_{on} = 0 \; (\forall o).$$

References

Athanassopoulos, A.D., Gounaris, C., & Sissouras, A. (1999). A Descriptive Assessment of the Production and Cost Efficiency of General Hospitals in Greece. Health Care Management Science, 2(2), 97-106.

Baker, L., & Spetz, J. (1999). Managed care and medical technology growth: National Bureau of Economic Research.

Baltagi, B.H., & Griffin, J.M. (1988). A general index of technical change. The Journal of Political Economy, 20-41.

Blank, J., & Valdmanis, V.G. (2008). Efficiency in hospital industry: summary and conclusions. In J.L.T. Blank & V.G. Valdmanis (Eds.), Evaluating hospital policy and performance: contributions from hospital policy and productivity research (Vol. 18, pp. 231-240). Oxford: Elsevier.

Blank, J.L., & Eggink, E. (2004). The decomposition of cost efficiency: an empirical application of the shadow cost function model to Dutch general hospitals. Health care management science, 7(2), 79-88.

Blank, J.L., & Vogelaar, I. (2004). Specifying technical change: a research on the nature of technical change in Dutch hospital industry. The Quarterly Review of Economics and Finance, 44(3), 448-463.

Blank, J.L.T., & Merkies, A.H.Q.M. (2004). Empirical assessment of the economic behaviour of Dutch general hospitals. Health Economics, 13(3), 265-280.

Dismuke, C.E., & Sena, V. (1999). Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portuguese public hospitals? An empirical analysis using parametric and non-parametric methods. Health Care Management Science, 2(2), 107-116.

Grosskopf, S., Margaritis, D., & Valdmanis, V. (2001). Comparing teaching and non-teaching hospitals: a frontier approach (teaching vs. non-teaching hospitals). Health Care Management Science, 4(2), 83-90.

Gruca, T.S., & Nath, D. (2001). The Technical Efficiency of Hospitals under a Single Payer System: The Case of Ontario Community Hospitals. health Care Management Science, 4, 91-101.

Hofmarcher, M.M., Paterson, I., & Riedel, M. (2002). Measuring hospital efficiency in Austria–a DEA approach. Health Care Management Science, 5(1), 7-14.

Hollingsworth, B., Dawson, P., & Maniadakis, N. (1999). Efficiency measurement of health care: a review of non-parametric methods and applications. Health Care Manag Sci, 2, 161 - 172.

Kopp, R.J., & Smith, V.K. (1983). Neoclassical modeling of nonneutral technological change: an experimental appraisal. The Scandinavian Journal of Economics, 127-146.

Li, T., & Rosenman, R. (2001). Cost inefficiency in Washington hospitals: a stochastic frontier approach using panel data. Health care management science, 4(2), 73-81.

Linna, M., Hakkinen, U., & Magnussen, J. (2006). Comparing hospital cost efficiency between Norway and Finland. Health Policy, 77(3), 268-278.

Maniadakis, N., Hollingsworth, B., & Thanassoulis, E. (1999). The impact of the internal market on hospital efficiency, productivity and service quality. Health Care Management Science, 2(2), 75-85.

McCallion, G., Glass, J.C., Jackson, R., Kerr, C.A., & McKillop, D.G. (2000). Investigating productivity change and hospital size: a nonparametric frontier approach. Applied Economics, 32, 161-174.

Okunade, A.A. (2001). Cost-output relation, technological progress, and clinical activity mix of US hospital pharmacies. Journal of Productivity Analysis, 16(2), 167-193.

Prior, D., & Sola, M. (2000). Technical efficiency and economies of diversification in health care. Health Care Management Science.

Rodríguez-Álvarez, A., & Lovell, C.A.K. (2004). Excess capacity and expense preference behaviour in National Health Systems: an application to the Spanish public hospitals. Health Economics, 13, 157-169.

Rosko, M. (1999). Impact of internal and external environmental pressures on hospital inefficiency. Health Care Manag Sci, 2, 63 - 74.

Sommersguter-Reichmann, M. (2000). The impact of the Austrian hospital financing reform on hospital productivity: empirical evidence on

efficiency and technology changes using a non-parametric input-based Malmquist approach. Health Care Management Science, 3(4), 309-321.

Spetz, J., & Maiuro, L.S. (2004). Measuring levels of technology in hospitals. The quarterly review of economics and finance, 44(3), 430-447.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. Journal of the American Statistical Association 57, 348–368.

Adjustment costs and inter-temporal savings

Adjustment costs and inter-temporal savings: endogenous modelling of the adoption of innovations

Submitted for publication to: Applied Mathematical Modelling

5.1 Introduction

There is a large body of literature on the efficiency and productivity of firms (for an overview see Fried et al., 2008), in particular in the hospital industry (for overviews see Blank & Valdmanis, 2008; Jacobs et al., 2006; Ozcan, 2008). Most studies try to establish static relationships with productivity. Popular productivity related topics are competition (Chu et al., 2011; Chua et al., 2011), scale economies (Harrison, 2011; Kristensen et al., 2010; Zhao et al., 2011), network economies (Granderson, 2011), management quality (Besstremyannaya, 2011) and ownership (Eggleston & Shen, 2011; Herr et al., 2011; Rego et al., 2010). These studies are rather static and leave adjustment cost and intertemporal savings out of consideration. Changes in technology are being regarded as a more or less autonomous process ("manna from heaven") that can be modelled as a shift in the production or cost frontier.

This simplified approach is particularly striking since the work of Morrison and Berndt (1981) in the eighties of the last century has provided an excellent framework to deal with dynamics (inter-temporal decisions) in productivity analysis. Morrison and Berndt (1981) describe the adjustment process with respect to fixed resources. Changes in production levels, product prices and resource prices may affect the allocation of resources, in particularly fixed resources. The adjustment to the new optimal allocation of resources may be accompanied by adjustment costs and by time lags to accomplish a new efficient allocation of resources. In other words, there is a trade-off between adjustment costs and productivity gain over time.

Literature that builds on the ideas of Morrison and Berndt is scarce. For an overview of dynamic efficiency in a non-parametric framework we refer to Fallah-Fini et al (2014). Based on the literature they identify five types of inter-temporal relations of inputs and outputs that have been researched: production delays, inventories, capital or generally fixed factors, adjustment costs and

incremental improvement and learning models. For a parametric framework Luh & Stefanou (1991) extend the measure of total factor productivity (TFP) growth to a dynamic measure of productivity growth adjusted for deviations from the long-run equilibrium within an adjustment-cost framework. Rungsuriyawiboon and Stefanou (2007) formulate a model of dynamic efficiency by generalizing the shadow cost approach in a context of the dynamic duality model of intertemporal decision making. The model incorporates adjustment costs for quasi-fixed factors for a firm's dynamic production decision problem. Rungsuriyawiboon and Stefanou (2008) extend this furthermore with a method to estimate the decomposition of dynamic total factor productivity growth in the presence of inefficiency, as an extension of the dynamic total factor productivity growth, adjusted for deviations from the long-run equilibrium within an adjustment-cost framework.

Research on dynamics of efficiencies has in common that it takes quasi-fixed factors into account for which in the dynamic framework an optimal allocation can be derived. However there is also a measurement issue at stake here. Change in technology is not simply an equivalent of new equipment or devices. If it was, prices and quantities of new technologies could be well measured and be included in the dynamic models of Morrison and Berndt. However, new technologies or more specific innovations can also be linked with subtle changes in (the quality of) resources and production processes (logistics), which cannot be measured in a straightforward manner. Instead, the innovations may merely be identified as such but may not be expressed in terms of quantities and prices of devices or equipment, simply due to lack of information.

This chapter shows how to deal with adjustment costs and intertemporal savings. We will focus on innovations and their impact on costs in the short run through adjustment costs and their impact on costs in the long run through the use of improved technology. In a static approach the adjustment cost may be interpreted as inefficiency, whereas they represent inevitable costs

to attain (future) higher productivity. We therefore present a model that treats innovations as endogenous (in the previous chapter innovations were exogenous). The model is particularly useful in case innovations are hard to measure, like process innovations cannot be expressed in terms of physical input quantities. One may think of innovations in terms of changes in production lines, organisational structures or specific changes in employees' skills, typically these can generally only be measured in terms of a dichotomous variable (implemented/not implemented). The model is applied to a dataset of Dutch general hospitals operating during the years 1995-2005.

The chapter is organised as follows. In Section 2 we discuss the theoretical model, followed by a formal mathematical representation. Section 3 gives the details of an empirical model. Section 4 applies the empirical model to the Dutch hospital industry. Section 5 contains the outcomes of our econometric analysis. The last section closes by briefly summarising and drawing conclusions.

5.2 Economic model

As mentioned in the introduction, there is large body of literature on the cost structure of hospitals. We apply and adapt the cost function model to describe the economic behaviour of firms (see e.g. Blank & Lovell, 2000, pp. 6-11). The cost function model has been well documented, in case of hospitals (see e.g. Blank & Vogelaar, 2004; Blank & Merkies, 2004; Blank & Van Hulst, 2009). The modification that we make to the cost function model is that we explicate the effects of the level of technology present as well as the adoption of new technology on costs.

It is assumed that a firm is a long-term cost minimiser, i.e. the firm minimises its total discounted costs over an infinite time horizon at a given set of services delivered, a given set of resource prices and at a given technology. The firm operates at the technology frontier (technical efficient) and allocates

its resources in such a way that it produces at the lowest cost (allocative efficient). However, in practice firms are not operating on a long run equilibrium, but rather on a short run equilibrium. The short run equilibrium implies that even though the firm is minimising its cost providing given services and resource prices, it is not operating at the frontier technology (technical inefficient). The reason behind this technical inefficiency is that the frontier is permanently shifting due to technical changes and the adoption of new technologies, firms that lag behind are deemed inefficient. The central idea is that the adoption of new technologies is accompanied by extra costs; the so-called adjustment costs. Adjustment costs are the temporary costs for switching from one technology to another. Aside from these costs, new technologies may also result in extra costs due to new equipment or higher salaries for more skilled labour. However, these costs are an integral part of the producer costs and must be distinguished from the temporary adjustment costs. The point is that the adjustment costs are there, but that we cannot exactly pin-point these costs.

Note that we could also have started from another economic objective. However, our line of reasoning would be the same in case of profit or revenue maximisation and similar derivations could be made. Product prices are particularly lacking in public sector applications and it would not make any sense to derive the optimal behaviour for service delivery allocation. However, in the appropriate context, there are no objections whatsoever to follow a similar approach.

Our formal starting point is thus a standard cost function $c(y,w,A)$ in a static world with y representing a set of services, w a set of resource prices and A a measure for the state of technology. The cost function $c(\cdot)$ is a twice differentiable function with respect to w, which satisfies the requirements concerning monotonicity and concavity (Färe & Primont, 1995). In a static world and under the assumption of cost minimisation we may derive the

optimal resource demand functions by Shephard's Lemma (Shephard, 1953, 1970).

However, in a dynamic world - aside from establishing the optimal inputs- the firm also has to decide on the optimal level of technology. Instead of "manna from heaven" the change in technology (and thus also the level) is regarded as an endogenous variable in the model. Since the introduction of new technologies is accompanied by adjustment costs, new technologies will not be placed at their maximum level. Instead, the firm has to balance productivity gains generated by new technologies with the adjustment costs of new technologies. Since data on the costs of new technologies and in particular adjustment costs are unavailable and hard to collect, we will follow an alternative approach that is totally based on empirics. It is stated that on the short run minimum cost depends on service quantities, resource prices, technologies implemented and the recent implementation of new technologies (let's say in the past year). Therefore, we have extended the cost function to:

$$C = c(y, w, A, A') \tag{1}$$

With:
C = cost;
y = vector of services;
w = vector resource prices;
A = technology index;
A' = change in technology index in the past year.

For this cost function we have derived the standard input demand equations by applying Shephard's lemma:

$$x = \nabla_w \, c(y, w, A, A') \tag{2}$$

With:
x = vector of resources;
∇_w = gradient with respect to w.

Note that the input demand is also affected by the technology index and by the change in technology index.

However, since technology is not "manna from heaven" but is a result of strategic decisions made by the management, we need to include the decision on new technologies as well. This is, however, not a static decision since a new technology also has an impact on future costs, by deciding on A' the present level of A changes as well the future level of A. Hence, deciding on new technology is a dynamic decision. We model this dynamic decision by constructing a function that includes all future cost functions. The function represents the lifetime costs of producing each periodical output at resource prices. Since the present decision on the use of different types of technology may also affect future cost flows, the dynamic optimisation rule can be derived by minimising total discounted future costs. Since we do not know what future developments in production levels and resource prices will entail, we assume them to be constant through time. Deciding what level of implementation of new technology is optimal in terms of cost minimisation is mathematically represented by:

$$LTC = \min_{A'} \int_0^\infty e^{-rt}\ c(y_t, w_t)dt = \min_{A'} \int_0^\infty e^{-rt}\ c(y_t, w_t, A_t, A'_t)dt \qquad (3)$$

With:

LTC = lifetime cost of producing y at resource prices w;

r = discount rate;

t = time.

To solve the minimisation problem of equation (3) we applied the Euler Lagrange equation (Sydsaeter et al., 2005, p. 111) and solved it. The Euler Lagrange equation states that $\frac{\partial F}{\partial x} = \frac{d}{dt}\frac{\partial F}{\partial \frac{dx}{dt}}$ is a necessary condition for the solution of $\max \int_{t0}^{t1} F(t, x, \frac{\partial x}{\partial t})dt$. Applying this to equation (3) yields:

$$e^{-rt}.\frac{\partial C}{\partial A} = \frac{d}{dt}\left[e^{-rt}.\frac{\partial C}{\partial A'}\right] \tag{4}$$

Implying:

$$e^{-rt}.\frac{\partial C}{\partial A} = -r.e^{-rt}.\frac{\partial C}{\partial A'} + e^{-rt}.\frac{d}{dt}\left[\frac{\partial C}{\partial A'}\right] \tag{5}$$

Furthermore, we assume that the change in technology is in itself not time-dependent indicating that the second term on the RHS equals zero. Equation (4) reduces to:

$$\frac{\partial C}{\partial A} = -r.\frac{\partial C}{\partial A'} \tag{6}$$

Equation (6) can easily be interpreted as follows. The left hand side reflects the structural cost savings per year by expanding the technology by one unit. Over an infinite time horizon the structural cost saving equalises $1/\big(1-(1-r)\big) = 1/r$ times the annual savings. De facto we have derived the discrete analogue of equation (6). Intuitively this is clear, since the marginal costs of an additional unit of change in technology should be equal to the discounted structural savings.

After rearranging and taking logarithms, equation (6) in terms of elasticity can be written as:

$$A' = -r.A.\frac{d\ln C/d\ln A'}{d\ln C/d\ln A} \tag{7}$$

Note that we also need to check the first order conditions with respect to A and A'. Outcomes are consistent with our theoretical considerations if the cost function is monotonously non-increasing in A and monotonously non-decreasing in A'. Furthermore, the cost function needs to be concave in A and convex in A'.

So, starting from a standard cost function $c(y, w, A)$ in a static world where firms decide on production and allocation of costs, we extended the cost function with the decision on the amount of new technology resulting in the cost function $c(y, w, A, A')$. We then added the inter-temporal dimension of the decision-making process on the amount of new technology A'. This implies an additional equation for the model that is given by (7).

5.3 Empirical model

For an empirical application of the economic model we use the well-known translog cost function model (Jorgenson et al, 1973). The cost function model consists of a translog cost function and the corresponding cost share equations. The model includes first and second order terms, cross terms between outputs and input prices, a time trend and cross terms of outputs and input prices with the time trend. The cross terms with the time trend represent the possible different natures of technical change. Cross terms with outputs refer to output-biased technical change and cross terms with input prices to input-biased technical change. Furthermore the technology index A and change in technology index A', as well cross terms are included. The translog cost function is given as:

$$
\begin{aligned}
\ln C = a_0 &+ \sum_{m=1}^{M} b_m \ln y_m + \sum_{n=1}^{N} c_n \ln w_n \\
&+ \frac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M} b_{mm'} \ln y_m \ln y_{m'} \\
&+ \frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N} c_{nn'} \ln w_n \ln w_{n'} + \sum_{m=1}^{M}\sum_{n=1}^{N} e_{mn} \ln y_m \ln w_n \\
&+ h_1 \text{time} \\
&+ \sum_{m=1}^{M} i_{1m} \ln y_m \text{time} + \sum_{n=1}^{N} j_{1n} \ln w_n \text{time} + l_1 \ln A \\
&+ l_2 \ln A' + \frac{1}{2} l_{11} \ln A \ln A \\
&+ \frac{1}{2} l_{22} \ln A' \ln A' + \sum_{m=1}^{M} n_{1m} \ln A \ln y_m \\
&+ \sum_{n=1}^{N} m_{1n} \ln A \ln w_n \\
&+ \sum_{m=1}^{M} n_{2m} \ln A' \ln y_m + \sum_{n=1}^{N} m_{2n} \ln A' \ln w_n
\end{aligned}
\tag{8}
$$

With:

C	= total costs;
y_m	= output m (m = 1,.., M);
w_n	= price of input n (n = 1,.., N);
time	= trend;
A	= technology index;
A'	= change in technology index in past year;

$a_0, b_m, b_{mm'}, c_n, c_{nn'}, e_{mn}, h_1, i_{1m}, j_{1n}, l_1, l_2, l_{11}, l_{22}, m_{1n}, m_{2n}, n_{1m}, n_{2m}$ parameters to be estimated.

Applying Shephard's Lemma we found the optimal cost share functions:

$$S_n^0 = c_n + \sum_{n'=1}^{N} c_{n'n} \ln w_{n'} + \sum_{m=1}^{M} e_{mn} \ln y_m + j_{1n} T + m_{1n} \ln A + m_{1n} \ln A' \qquad n = 1,..,N \tag{9}$$

With:

S_n^0 = optimal cost share for input n ($n = 1,.., N$).

Homogeneity of degree one in prices and symmetry is imposed by putting constraints on some of the parameters to be estimated. In formula:

$$\mathrm{b_{mm'} = b_{m'm}} \; ; \quad \mathrm{c_{nn'} = c_{n'n}} \; ; \quad \mathrm{d_{oo'} = d_{o'o}} \; ;$$

$$\sum_{n=1}^{N} \mathrm{c_n} = 1 \; ; \qquad \sum_{n=1}^{N} \mathrm{c_{nn'}} = 0 \;\; (\forall \mathrm{n'}) \; ; \qquad \sum_{n=1}^{N} \mathrm{e_{mn}} = 0 \; (\forall \mathrm{m}) \; ; \tag{10}$$

$$\sum_{n=1}^{N} \mathrm{j_{1n}} = 0 \; ; \qquad \sum_{n=1}^{N} \mathrm{m_{kn}} = 0 \; (\mathrm{k} = 1,2)$$

For reasons of convenience, technical change is modelled by including a time trend, implicitly assuming that technical change is a rather smooth process in time. However, previous research on Dutch hospitals (Blank & Vogelaar, 2004) shows that technical change appears intermittently. So including a time trend in a cost function model is a rather restrictive way of incorporating technical change. Moreover the time trend is included in addition to A en A'.

In the empirical model we expect the time trend to capture that part of the technical change that is not captured by A or A'.

Finally, we have an additional innovation equation that models the decision on the optimal amount of new technology. This is of course equation (7), which for the empirical application is written as:

$$A' = -rA \cdot \frac{l_2 + l_{22} \ln A' + \sum_{n=1}^{N} m_{2n} \ln w_n + \sum_{m=1}^{M} n_{2n} \ln y_m}{l_1 + l_{11} \ln A + \sum_{n=1}^{N} m_{1n} \ln w_n + \sum_{m=1}^{M} n_{1n} \ln y_m} \tag{11}$$

5.4 Application to Dutch hospitals

General

Data for this study covers the period 1995-2005. The financial, patient and personnel data were collected by the Institute for Health Care Management and obtained from the Ministry of Health, Welfare and Sport. The data contain information on almost all general hospitals yielding approximately 90 observations each year, situated in 11 health care regions. Data about the adoption of innovations come from an additional survey amongst hospitals collected by ECORYS and the Public Health Council. The survey contains information on 63 innovations in 66 general hospitals. For the purposes of this study, observations on hospitals with missing or unreliable data were excluded from the dataset. Various consistency checks were performed on the data to ensure that changes in average values and the distribution of values across time were not excessive. After eliminating observations containing inaccurate or missing values in the dataset, an unbalanced panel dataset of 539 observations over the 10 years of study remained (1995 is omitted from the dataset because we use a lag to measure the change in the technology index).

Since the hospitals that participated in the survey on innovations is a sample of the complete population of Dutch general hospitals we checked this

sample for representativeness. We have analysed whether a hospital is or is not present in the sample with respect to several variables for which we have information for the complete population of Dutch general hospitals. The most appropriate statistical method for analysing a dichotomous dependent variable is a logit analysis. Explanatory variables are the size, productivity, region and type of hospital. The outcomes of the logit analysis show that, based on t-statistics at the 5% significance level, none of these characteristics "explain" the presence/no presence in the sample. In other words, the presence/no presence in the sample is random, or at least does not depend on one of the independent variables. We conclude that, based on these characteristics, the sample is representative. The sample consists of almost 75% of the complete population.

Production

The main service rendered by hospitals is treating patients. The health outcomes of patients are a particularly important component of hospital production. However data on health outcomes are not available. Instead we use a more common approach in which the production of hospitals is measured by the number of discharges and outpatients (see for example Blank & van Hulst, 2009). The data on discharges cover 30 medical specialties in order to measure case-mix. Since it is not possible to use such a large number of specialties, the specialties have been aggregated into four categories based on average stay homogeneity and the distinction between surgery/non-surgery specialties. We therefore distinguish the following groups of specialties:

- Non-surgery with average stay less than 4 days;
- Non-surgery with average stay more than 4 days;
- Surgery with average stay less than 4 days;
- Surgery with average stay more than 4 days.

In total five products, four discharge groups and outpatients. These five products explain– as we shall see later – variations in cost to a very large extent.

Resources

Resources include four categories of labour, material supplies and capital. The sum of costs of these resources adds up to total costs. Furthermore, data on prices for these resources are needed for successful estimation of equation (8) and (9). We discuss the approach of determining the input price for each resource.

The following four categories of labour are distinguished: management and administrative personnel, nursing personnel, paramedical personnel (e.g. lab technicians, psychologist) and auxiliary personnel (e.g. hotel personnel, security, cleaning). Physicians are not included in the model. The reason for this, is that in the Dutch system a hospital can have physicians on their payroll as well self-employed physicians. Excluding physicians ensures that hospitals with physicians on their payroll and hospitals with self-employed physicians are treated equally. Note that the implicit assumption here is that there are no substitution possibilities between physicians and the other personnel, which makes sense, since physicians practice a protected profession to conduct procedures which other personnel is not licensed for. Substitution may only occur in case of administrative tasks and simple medical procedures, which are limited tasks for the physicians.

For all hospitals, data are available on the costs and the quantity for each personnel category. The price of each personnel category is computed as the quotient of cost and volume. We then use these prices to estimate a regional price for each time period by regressing the prices on regional dummies and time dummies. Hence, regional prices are considered exogenous, differences from the regional prices are considered endogenous. The estimate prices are considered as the market prices for labour.

Material supplies include such aspects as medical supplies, food and heating Since there is no natural unit of measurement for material supplies, a circumventing construction was used. For the first year in the dataset the price of materials is set as a unit price. In the following years the price of material

supplies is derived from the consumer price index for the Netherlands as calculated by Statistics Netherlands.

Capital consists of capital assets such as buildings and medical equipment. There are data available on the costs of capital and a couple of indicators that represent the volume of capital. The price of capital is derived from the cost of capital divided by the volume of capital. The latter is a volume index based on the weighted aggregation of the number of beds, intensive care beds, radiotherapists (proxy for the number of linear accelerators and cobalt machines) and surgery rooms. The weights for the volume index are derived from a regression of capital costs on the variables that make up the volume index.

Technology

According to Spetz and Maiuro (2004) measures of technologies are limited in a number of ways. They state that measurement must be foremost driven by the research question at hand. There is no one-size-fits-all solution. In case of cost function estimation or efficiency measurement, an aggregate index of single technologies is preferred. In line with the suggestion of Spetz and Maiuro we therefore introduce our concept of technology measurement based on an innovation index. Examples of single innovations are specialised mamma clinics (breast cancer tests and diagnosis) or specialised cardiac nurses for consultation. In a hospital an innovation is present or not and therefore measured by a set of dichotomous variables $[I_1, .. ,I_J]$. The technology index equals the unweighted sum of number of implemented innovations:

$$A_{h,t} = \sum_{j=1}^{J} I_{j,h,t} \tag{12}$$

With :

$A_{h,t}$ = technology index for hospital h at time period t;

$I_{j,h,t}$ =1 if innovation j is present in hospital h at time period t, 0 elsewhere.

From the technology index A we also derive the change in technology by taking the first differential of A. The change in technology for hospital h at time period t is therefore:

$$A'_{h,t} = A_{h,t} - A_{h,t-1} \tag{13}$$

Spetz and Maiuro (2004) also refer to the drawbacks of using a general technology index. The index does not distinguish between various heterogeneous technologies. We diminish the heterogeneity problem by selecting rather general technologies and excluding academic hospitals from the analysis. However, this is only a limited correction for the heterogeneity problem. The index therefore must be regarded as a global proxy measuring the innovation propensity of a firm. In constructing the technology index Spetz and Maiuro (2004) also suggest to use a weighted sum of technologies, where the weights are based on the rarity of the innovation. Analysis on an earlier dataset on innovations of Dutch hospitals (Blank & Van Hulst, 2009) show that measuring innovations by the unweighted or the weighted sum hardly affects the productivity outcomes. For this reason and the lack of relevance for the core aim of our research we further ignore this option.

The survey on innovations originally contains information on a set of 63 innovations. However, not all of these 63 innovations are aimed at enhancing productivity, instead some of the innovations are implemented for quality or marketing reasons. Innovations that are not aimed at raising productivity are filtered out, leading to a set of 31 innovations. Table 5-1 shows the complete list of 31 innovations used in this study to construct the technology index.

Table 5-1 List of innovations

Diabetes foot policlinic	Cardiac nurse
Mamma policlinic	MS nurse
Constipation and wee-wee policlinic (children)	Wound consultant
Mother child unit	Oncology consultant
Cardiac policlinic	Total knee (reduction of hospital stay duration)
Lung revalidation	Transmural care for oncology patients
Down policlinic	Cooperation with general practitioner (first aid)
Protocol of reference by general practitioner	Transmural care
Use of seal equipment at intestine surgery	Cataract line
MRI instead of muelografics	Other logistic innovation
Shaver blades at endonasal surgery	Home monitoring of pregnancy
Stroke care unit	Self-measurement thrombotic care
TVT devices	Other outside hospital care innovation
Pre-Operative screening by anaesthesiology	Electronic data at consultation room & ward
CVA consultant	Other ICT innovation
Decubitus nurse	

Interest rate

Since we estimate a model that also discounts future productivity, we need to include an interest rate (r). Since future interest rates are unknown, several values of the interest rate are picked (r=0.05, r=0.10, r= 0.15) and analysed.

Descriptive statistics

Table 5-2 shows the descriptive statistics of the variables for 2005 (the last year of observation in our dataset).

In order to simplify the interpretation of the estimated parameters all variables in the analysis are standardised at their arithmetical means. The first-order parameter estimates represent the elasticity of cost with respect to the corresponding service or resource price for the "average" firm.

Table 5-2 Descriptive Statistics, Dutch General Hospitals 2005 (N=51)

Variable	Mean	Standard dev.
Discharges 1	9,021	4,150
Discharges 2	6,851	3,326
Discharges 3	6,173	3,300
Discharges 4	6,716	3,327
Outpatients	65,173	32,253
Price management & administrative personnel (in €)	43,582	5,244
Price nursing personnel (in €)	46,024	3,007
Price paramedical personnel (in €)	75,535	21,726
Price auxiliary personnel (in €)	33,696	3,072
Price material supplies (index)	1.25	0
Price capital (index)	1.46	0.401
Cost (x € million)	99,431	57,643
Cost share management & administrative personnel	0.103	0.019
Cost share nursing personnel	0.341	0.027
Cost share paramedical personnel	0.031	0.016
Cost share auxiliary personnel	0.095	0.014
Cost share material supplies	0.308	0.022
Cost share capital	0.119	0.025
Technology	19.6	4.6
Change in technology	1.1	1.22

5.5 Estimation and evaluation

Specification

In the Dutch context, a model based on cost minimising behaviour and services, resource prices and capital inputs as exogenous variables is the most appropriate one (see e.g. Blank & Van Hulst, 2009). Accordingly, we estimate a direct cost function model. The cost function model constitutes a system with a cost function and a number of cost share equations (see section 2). In addition to regular models we also have the Euler equation which gives the optimum amount of new technology as derived in section 2. Our base model is a full model with no additional restrictions on the parameters. Other

specifications of the model, with additional restrictions on parameters, are tested against this base model.

The models are estimated as multivariate regression systems using various equations with a joint density, which we assume to be normally distributed. Because disturbances are likely to be cross-equation-correlated, a minimum distance estimator is used. As usual, because the shares add up to one causing the variance-covariance matrix of the error terms to be singular, one share equation in the direct cost function model is eliminated.

Various estimated model specifications are evaluated based on:

- a formal likelihood ratio test;
- standard statistical properties such as R^2 and T-values;
- theoretical requirements (monotonicity and concavity of the cost function);
- economic plausibility of estimated parameters such as productivity change due to innovations.

The estimates of the various specifications generate robust results. Most of the parameters are significant and have the expected sign. A formal test of the models is the log likelihood ratio test. Table 5-3 presents the log likelihoods of various specifications, for several values of the interest rate r. Table 5-3 only includes the models with the highest likelihoods. For instance the models with restrictions on the parameters for the cross terms of "trend * input price" are omitted because these models lead to inferior results.

Table 5-3 Results of the log likelihood for various models (N=539)

Restrictions	# param	Log Likelihood r=0.05	Log Likelihood r=0.10	Log Likelihood r=0.15
none	101	8.927	9.056	9.059
Technology * input price	96	8.923 *	9.048	9.048
Technology * output	96	8.922 *	9.052 *	9.055 *
Change in technology * input price	96	8.910	9.047	9.050
Change in technology * output	96	8.923 *	9.047	9.050
Technology * input, Change in technology * input price	91	8.907	9.039	9.040
Technology * output, Change in technology * output	91	8.918 *	9.038	9.041
Technology * output, Change in technology * output, Technology * input price, Change in technology * input price	81	8.898	9.022	9.022

*= The likelihood-ratio test rejects the hypothesis that the unrestricted model is a better model.

Table 5-3 shows that the unrestricted model behaves well. However, for all the distinct values of *r* the likelihood-ratio test rejects the hypothesis that the unrestricted model is preferred over the model with restricted parameters for "technology * output". So we conclude that parameters corresponding to the cross terms of "technology * output" can be omitted from the model. Furthermore, Table 5-3 demonstrates the influence of *r*. For a relatively low value of 0.05 more cross terms could be dropped from the model. For *r* is 0.05 both "technology * input" as "change in technology * output" could be dropped from the model.

Technically it is not possible to perform a likelihood test between a dynamic model and a static model since the static model is not nested in the dynamic model. In particular the technology equation (11) cannot be estimated in the static case. However, it is possible to estimate a semi-dynamic model, a model that includes the parameters of A and A' without the additional technology equation. Note that the results of the semi-dynamic model and static model both are independent of the assumptions on the interest rate. The semi-dynamic model tested against the static model leads to the conclusion that the semi-dynamic model outperforms the static model (likelihood values

of respectively 8,167 for the static model and 8,135 for the semi-dynamic model).

Results

For the sake of space we only present the estimates of the most appropriate model, this is the model from which the cross terms of "technology*output" are eliminated (n1m parameters).

Table 5-4 Parameter estimates, fully specified model

Parameter	Variable	Estimate	T-value
A0	Constant	0.310	12.93
B1	Discharges group 1	0.008	0.12
B2	Discharges group 2	0.425	5.25
B3	Discharges group 3	0.100	1.68
B4	Discharges group 4	0.305	4.98
B5	Outpatients	0.414	6.93
B11	Discharges group 1 * discharges group 1	0.053	0.30
B12	Discharges group 1 * discharges group 2	0.721	5.19
B13	Discharges group 1 * discharges group 3	-0.282	-2.69
B14	Discharges group 1 * discharges group 4	-0.143	-1.30
B15	Discharges group 1 * outpatients	-0.453	-4.01
B22	Discharges group 2 * discharges group 2	-0.661	-2.61
B23	Discharges group 2 * discharges group 3	0.390	2.72
B24	Discharges group 2 * discharges group 4	-0.291	-2.16
B25	Discharges group 2 * outpatients	-0.228	-1.62
B33	Discharges group 3 * discharges group 3	-0.363	-2.79
B34	Discharges group 3 * discharges group 4	0.211	2.55
B35	Discharges group 3 * outpatients	0.126	1.32
B44	Discharges group 4 * discharges group 4	-0.031	-0.27
B45	Discharges group 4 * outpatients	0.312	3.07
B55	Outpatients * outpatients	0.374	3.04
C1	Price man. & adm.	0.086	16.63
C2	Price nursing personnel	0.360	53.62
C3	Price paramedical personnel	0.035	14.19
C4	Price auxiliary personnel	0.113	24.04
C5	Price material supplies	0.252	53.47
C6	Price capital	0.154	63.58

Parameter	Variable	Estimate	T-value
C11	Price man. & adm. * price man. & adm.	0.015	2.04
C12	Price man. & adm.* price nursing personnel	0.018	2.22
C13	Price man. & adm. * price medical personnel	-0.004	-1.86
C14	Price man. & adm.	0.024	3.95
C15	Price man. & adm. * price material supplies	-0.042	-4.92
C16	Price man. & adm.	-0.011	-1.98
C22	Price nursing personnel * price nursing personnel	0.005	0.32
C23	Price nursing personnel * price medical personnel	-0.002	-0.77
C24	Price nursing personnel * price auxiliary personnel	0.010	1.10
C25	Price nursing personnel * price capital	-0.025	-1.86
C26	Price nursing personnel * price material supplies	-0.005	-0.79
C33	Price medical personnel * price medical personnel	0.017	10.63
C34	Price medical personnel * price auxiliary personnel	0.0003	-0.13
C35	Price medical personnel * price capital	-0.004	-1.51
C36	Price medical personnel * price material supplies	-0.006	-3.41
C44	Price auxiliary personnel * price auxiliary personnel	-0.030	-3.24
C45	Price auxiliary personnel * price capital	0.001	0.06
C46	Price auxiliary personnel * price material supplies	-0.004	-0.60
C55	Price material supplies * price material supplies	0.125	7.75
C56	Price material supplies * price capital	-0.054	-22.33
C66	Price capital * price capital	0.081	42.22
E11	Discharges group 1 * price man. & adm.	0.006	1.87
E12	Discharges group 1 * price nursing personnel	-0.009	-1.46
E13	Discharges group 1 * price medical personnel	0.001	0.23
E14	Discharges group 1 * price auxiliary personnel	-0.005	-1.23
E15	Discharges group 1 * price material supplies	-0.001	-0.12
E16	Discharges group 1 * capital	0.007	2.43
E21	Discharges group 2 * price man. & adm.	0.001	0.27
E22	Discharges group 2 * price nursing personnel	0.019	2.77
E23	Discharges group 2 * price medical personnel	0.007	1.77
E24	Discharges group 2 * price auxiliary personnel	-0.025	-5.10
E25	Discharges group 2 * price material supplies	0.007	1.25
E26	Discharges group 2* capital	-0.009	-2.61
E31	Discharges group 3 * price man. & adm.	0.005	1.73
E32	Discharges group 3 * price nursing personnel	0.009	1.71
E33	Discharges group 3 * price medical personnel	0.002	0.87
E34	Discharges group 3 * price auxiliary personnel	0.0003	0.07

Parameter	Variable	Estimate	T-value
E35	Discharges group 3 * price material supplies	-0.005	-1.22
E36	Discharges group 3 * capital	-0.012	-4.35
E41	Discharges group 4 * price man. & adm.	-0.014	-4.77
E42	Discharges group 4 * price nursing personnel	-0.003	-0.65
E43	Discharges group 4 * price medical personnel	-0.006	-2.11
E44	Discharges group 4 * price auxiliary personnel	0.010	2.82
E45	Discharges group 4 * price material supplies	0.0003	-0.11
E46	Discharges group 4 * capital	0.014	5.17
E51	Outpatients * price man. & adm.	0.006	1.92
E52	Outpatients * price nursing personnel	-0.026	-4.58
E53	Outpatients * price medical personnel	0.012	4.29
E54	Outpatients * price auxiliary personnel	0.013	3.36
E55	Outpatients * price material supplies	0.003	0.60
E56	Outpatients * capital	-0.009	-3.19
H1	Trend	-0.025	-9.79
I11	Trend * discharges group 1	0.007	0.96
I12	Trend * discharges group 2	-0.018	-1.86
I13	Trend * discharges group 3	0.010	1.32
I14	Trend * discharges group 4	-0.023	-3.26
I15	Trend * outpatients	0.011	1.39
J11	Trend * price man. & adm.	0.002	3.16
J12	Trend * price nursing personnel	-0.001	-0.68
J13	Trend * price medical personnel	0.0005	-0.67
J14	Trend * price auxiliary personnel	-0.001	-2.27
J15	Trend* price material supplies	0.004	7.56
J16	Trend * price capital	-0.004	-15.67
L1	Technology	-0.013	-4.40
L2	Change in Technology	0.039	4.38
L11	Technology * technology	-0.005	-4.26
L22	Change in Technology * Change in Technology	0.008	4.35
M11	Technology * price man. & adm.	0.001	1.30
M12	Technology * price nursing personnel	-0.003	-1.68
M13	Technology * price paramedical personnel	-0.001	-1.80
M14	Technology * price auxiliary personnel	-0.002	-1.28
M15	Technology * price material supplies	0.005	3.02
M16	Technology * price capital	-0.001	-2.24
M21	Change in Technology * price man. & adm.	0.0002	0.79

Parameter	Variable	Estimate	T-value
M22	Change in Technology * price nursing personnel	0.001	1.77
M23	Change in Technology * price parammed. personnel	0.001	1.45
M24	Change in Technology * price auxiliary personnel	0.001	1.14
M25	Change in Technology * price material supplies	-0.002	-2.90
M26	Change in Technology * price capital	-0.001	-2.68
N21	Change in Technology * discharges group 1	-0.001	-0.64
N22	Change in Technology * discharges group 2	0.003	1.51
N23	Change in Technology * discharges group 3	-0.005	-2.72
N24	Change in Technology * discharges group 4	-0.002	-1.30
N25	Change in Technology * outpatients	0.003	1.68
	R2 cost equation	0.97	
	R2 cost share nursing personnel	0.08	
	R2 cost share medical personnel	0.35	
	R2 cost share price auxiliary personnel	0.15	
	R2 cost share material supplies	0.28	
	R2 cost share capital	0.82	
	R2 Innovation equation	0.73	

Table 5-4 shows that in a statistical sense the cost function model fits the data rather well. Results derived from this cost function are plausible. The cost equation has a high $R^{2,}$ i.e. 0.97. More than 75% of the estimated parameters are significant at the 5% level. Most R^2 's of the share equations are in line with previous results (Blank & Van Hulst, 2009). The R^2 of the technology equation is 0.73 and seems to be a good result. The requirements on monotonicity and concavity are also fulfilled to a large extent. The monotonicity property tells us that input demand is always positive, which is the case for all observations and in particular for the hypothetical "average" hospital[7]. A necessary condition for concavity is the negativity of the "own" elasticities of substitution. This

[7] The "average" hospital is designed wherein values for all variables set at the sample (arithmetic) mean.

condition also holds for the "average" hospital and is valid for 81% of the observations. The invalid observations for the monotonicity property are mainly due to the input paramedic personnel, for which the cost share can be quite small in particular cases. Finally the condition of negative semi-definite of the matrix of elasticity's of substitution holds for the average hospital and is also valid in 65% of the observations[8]. We also tested the "significance" of each equation in the system separately by imposing the restriction that all the parameters (except the constant) equal zero. Based on likelihood ratio tests all the null hypotheses were overwhelmingly rejected.

Of course we are especially interested in the parameters estimated for the technology index (A) and the change in technology (A`). Both parameter estimates are significant and have plausible signs for the dynamic model. The parameter of the technology index has a negative sign, meaning more adopted innovations lowers the costs. The opposite applies for the change in technology; the positive sign - implying higher costs - reflects the adjustment costs of adoption. The results of the other specifications of the model reveal that the estimates are quite robust, the parameter estimates for the technology index (A) are consequent in the range from -0.023 up to -0.006, while the parameter estimates for the change in technology range from 0.007 to 0.029.

Furthermore, we can calculate the cost elasticities for innovations. At an interest level of r = 0.1 we found an average cost elasticity for the technology index of -0.008, implying that 10% more adopted innovations lead to a reduction of the costs of 0.08%. Changing the assumptions about the interest rate leads to plausible results. A higher interest rate leads to a lower elasticity, meaning less cost reduction. For a lower interest rate the opposite is valid.

---

[8] For most observations there is only one value slightly greater than zero.

5.6 Summary and Conclusions

One of the striking aspects of the recent productivity literature is the lack of attention to the dynamics of productivity. Little attention is paid to the costs and savings of adopting new technology. Changes in technology are being regarded as a more or less autonomous process ("manna from heaven") that can be modelled as a shift in the production or cost frontier. In this chapter we have presented a dynamic approach by introducing adjustment costs and inter-temporal decisions. In our approach, service levels and resource prices may affect the adoption of new technologies. The adjustment to the new technology may be accompanied by adjustment costs and by time lags to accomplish a new efficient allocation of resources.

Aside from the theoretical framework there is also a measurement issue at stake here. Change in technology is not simply an equivalent of new equipment or devices. If they were, prices and quantities of new technologies could be well measured and be included in the dynamic models of Morrison and Berndt (1981). Change in technology can also be linked with changes in (the quality of) resources and production processes (logistics), which cannot be measured in a straightforward manner. One may think of innovations in terms of changes in production lines, organisational structures or specific changes in employees' skills. These changes or innovations may merely be identified as being present or not, but cannot be expressed in terms of quantities and prices, simply due to lack of information.

This chapter presented a model that treats innovations as endogenous and puts productivity in a dynamic framework. Instead of only including a time trend in the model to deal with technical change, we adapted the model by incorporating the technology available to a firm and the new technologies that the firm adopts. We then note that the decision on the adoption of new technologies has two sides, present adjustment costs and future savings. The adoption of new technologies is an inter-temporal decision in which adjustment costs of adoption are weighted against future savings. The

implication of this inter-temporal decision is that there is an optimum amount of new innovations, for which we derived an equation that can be added to (frontier) models.

The model is applied to a dataset of Dutch general hospitals operating during years 1995-2005. In order to stress the relevance of a dynamic approach, we also applied a static version of the model and a semi-dynamic version of the model. The static version compared with the semi-dynamic version shows that including a technology index and a change in the technology improves the estimates, i.e. the (semi-) dynamic model is superior to the static model. The application to Dutch hospitals here is a rather simple one used to demonstrate the mechanism and implication of a dynamic model. Further research can extend the model with more sophisticated techniques and probably more detailed measurement of technology and change in technology.

References

Besstremyannaya, G. (2011). Managerial Performance and Cost Efficiency of Japanese Local Public Hospitals: A Latent Class Stochastic Frontier Model. Health Economics, 20, 19-34.

Blank, J.L.T., & Lovell, C. (2000). Performance assessment in the public sector: contributions from efficiency and productivity measurement. In J.L.T. Blank (Ed.), Public provision and performance. Amsterdam: Elsevier.

Blank, J.L.T., & Merkies, A.H.Q.M. (2004). Empirical assessment of the economic behaviour of Dutch general hospitals. Health Economics, 13(3), 265-280.

Blank, J.L.T., & Valdmanis, V.G. (2008). Evaluating hospital policy and performance: contributions from hospital policy and productivity research. Oxford: Elsevier JAI.

Blank, J.L.T., & van Hulst, B.L. (2009). Productive Innovations in Hospitals: An Empirical Research on the Relation between Technology and Productivity in the Dutch Hospital Industry. Health Economics, 18(3), 665-679.

Blank, J.L.T., & Vogelaar, I. (2004). Specifying technical change: a research on the nature of technical change in Dutch hospital industry. The Quarterly Review of Economics and Finance, 44, 448-463.

Chu, C.-L., Chiang, T.-L., & Chang, R.-E. (2011). Hospital Competition and Inpatient Services Efficiency in Taiwan: A Longitudinal Study. Health Economics, 20(10), 1268-1280.

Chua, C.L., Palangkaraya, A., & Yong, J. (2011). Hospital Competition, Technical Efficiency and Quality. Economic Record, 87(277), 252-268.

Eggleston, K., & Shen, Y.-C. (2011). Soft Budget Constraints and Ownership: Empirical Evidence from US Hospitals. Economics Letters, 110(1), 7-11.

Fallah-Fini, S., Triantis, K., & Johnson, A. (2014). Reviewing the literature on non-parametric dynamic efficiency measurement: state-of-the-art. Journal of Productivity Analysis, 41(1), 51-67.

Färe, R., & Primont, D. (1995). Multi-Output Production and Duality: Theory and applications. Dordrecht: Kluwer Academic Publishers.

Fried, H.O., Lovell, C.A.K., & Schmidt, S.S. (2008). The measurement of productive efficiency and productivity growth. New York: Oxford University Press.

Granderson, G. (2011). The Impacts of Hospital Alliance Membership, Alliance Size, and Repealing Certificate of Need Regulation, on the Cost Efficiency of Non-profit Hospitals. Managerial and Decision Economics, 32(3), 159-173.

Harrison, T.D. (2011). Do Mergers Really Reduce Costs? Evidence from Hospitals. Economic Inquiry, 49(4), 1054-1069.

Herr, A., Schmitz, H., & Augurzky, B. (2011). Profit Efficiency and Ownership of German Hospitals. Health Economics, 20(6), 660-674.

Jacobs, R., Smith, P.C., & Street, A. (2006). Measuring efficiency in health care. Analytic techniques and health policy. Cambridge and New York: Cambridge University Press.

Kristensen, T., Bogetoft, P., & Pedersen, K.M. (2010). Potential Gains from Hospital Mergers in Denmark. Health Care Management Science, 13(4), 334-345.

Luh, Y.-H., & Stefanou, S.E. (1991). Productivity growth in US agriculture under dynamic adjustment. American Journal of Agricultural Economics, 73(4), 1116-1125.

Morrison, C.J., & Berndt, E.R. (1981). Short-run labor productivity in a dynamic model. Journal of Econometrics, 16(3), 339-365.

Ozcan, Y.A. (2008). Health care benchmarking and performance evaluation. An assessment using data envelopment analysis (DEA). International Series in Operations Research and Management Science. New York: Springer.

Rego, G., Nunes, R., & Costa, J. (2010). The Challenge of Corporatisation: The Experience of Portuguese Public Hospitals. European Journal of Health Economics, 11(4), 367-381.

Rungsuriyawiboon, S., & Stefanou, S. (2008). The dynamics of efficiency and productivity growth in U.S. electric utilities. Journal of Productivity Analysis, 30(3), 177-190.

Rungsuriyawiboon, S., & Stefanou, S.E. (2007). Dynamic efficiency estimation: an application to US electric utilities. Journal of Business & Economic Statistics, 25(2), 226-238.

Shephard, R.W. (1953). Theory of Cost and Production Functions. Princeton, New Jersey

Shephard, R.W. (1970). Theory of cost and production functions. Princeton: Princeton University Press.

Spetz, J., & Maiuro, L.S. (2004). Measuring levels of technology in hospitals. The quarterly Review of Economics and Finance, 44(3), 430-447.

Sydsaeter, K., A. Strom, & P. Berck (2005). Economists' Mathematical Manual (4th ed.). Berlin: Springer.

Zhao, Y., Foley, M., & Eagar, K. (2011). Assessing Economies of Scale in Public Hospitals. Economic Papers, 30(3), 341-347.

Factor technical change

Healthcare productivity: decomposing overall technical change into factor technical changes. An empirical application to the Dutch hospital industry

Submitted for publication to: Human resources for health

6.1 Introduction

One of the health policies on which the OECD focuses, is an adequate planning of the health workforce. According to the OECD (2015) well-trained, well distributed and productive health workers are crucial for ensuring access to high quality and cost-effective health care in OECD countries. The challenge for policy-makers is to balance between a shortage and overabundance of health workers. Training too many health professionals is expensive, since this implies unnecessary cost for training facilities. On the other hand, training too few health professionals leads to access problems and delayed treatment with severe consequences. Especially for health professions a shortage of health professionals has no short run solutions, since the supply of health professionals is rather inelastic. Training health professionals takes a couple of years.

In order to make adequate forecasts of the demand for health professionals, workforce planning models are used. Aside from the demand side aspects of these models, knowledge of the production structure and productivity are of great relevance. In particular, the aspects of how shifts in health care demand are related to the required inputs and how technical changes effects the demand for inputs should be well addressed in these models. To be more specific, besides demand and supply, productivity also changes. Productivity growth might bridge a possible gap between demand and supply of the health labour workforce. Therefore incorporating productivity growth in planning models seems obvious. In practice, however, this is easier said than done

Productivity is the relationship between one or more inputs and one or more outputs that can be produced with the input. Measuring the productivity in the health sector means relating inputs (e.g. physicians, nurses, medical equipment) with output (e.g. number of doctor consultations or number of hospital discharges). Basically the concept of productivity is quite simple.

However, Evans et al. (2010) point out that "the concept of 'productivity' is very simple in principle, but rather slippery to pin down in practice".

The measurement of productivity and productivity growth in health care is complicated by the fact that the nature of production is multiple-output-multiple-input. A patient going to a primary care centre may first be examined by a nurse and then, depending on the nature of his disease, may be referred to a doctor. In a hospital a patient require services of several professionals (doctors, nurses) material supplies (bandages, pharmaceutics) and capital (beds, medical equipment). To make things more complicated each patient requires another mix of inputs. We are therefore faced with a multi-input, multi-output production process. Still, we want to relate the inputs and the outputs, for that purpose the production process can be represented through a production or cost function. The amount of needed inputs, such as the number of doctors, the number of nurses, material and capital are a function of different outputs such as the number of hospital discharges and outpatients inputs. Various combinations on the level of outputs (i.e. case-mix of patients) influence the amount of inputs needed. Productivity might change over time, implying that the relation between inputs and outputs has changed. For example productivity grows by working smarter, therefore the same amount of a outputs can produced with less inputs.

Given the uncertainty about the factors affecting future productivity, most health workforce planning models ignore the possibility of productivity growth or assume some arbitrary growth rate. Ono et al. (2013) illustrate this with two examples: Health Workforce Australia assumes in one of their scenarios a 5% productivity gain for doctors and nurses over the period 2010 and 2025, without specifying the sources of these productivity gains (Health Workforce Australia, 2012b); the Canadian Nurse Association assumes a 1% productivity growth per year over the period 2007 and 2022 (Canadian Nurses Association, 2009). Productivity growth is assumed rather then that it is determined and that is if productivity growth is addressed.

At the same time there is a huge amount research available that addresses the topic of productivity and productivity growth. Total Factor Productivity (TFP) and its decomposition is a popular research topic in the productivity literature (see e.g. Balk, 2003; Diewert, 1981; Grifell-Tatjé & Lovell, 1999; Kopp & Diewert, 1982). The central issue in this type of research is breaking down TFP growth into changes in scale, pure technical and allocative efficiencies and technical change, based on the seminal work of Diewert (1981). Based on his earlier work, research has been extended to cover a myriad of topics and methodological advances (Balk, 2001; O'Donnell, 2010). However, TFP is a general measure that applies to the aggregate of all inputs, while our interest is on the productivity of individual inputs. More specific the productivity of professionals. For that case it is questionable whether it is appropriate to apply a general measure for the productivity growth to all inputs. Each input will have its own specific productivity development. Therefore TFP has to be decomposed for the various inputs. For example into labour, material supplies and capital, but a more detailed decomposition might be preferred as labour consists of various inputs. Despite the increase of literature on TFP, far less attention has been paid to further decomposing technical change into productivity measures for individual inputs.

The limited attention paid to the phenomenon is striking, since in particular innovations with respect to working conditions, absenteeism policy, education and other forms of human resource management (HRM) may be seen as affecting the partial productivity of labour. From a policy point of view this might be an intriguing question, since it may provide an opportunity to establish the effects of aforementioned HRM measures on productivity. More generally, innovations are most likely to be linked with a particular input, such as an improved usage of floor space (capital) or a decreased waste of materials. This chapter therefore focuses on the (partial) factor productivity changes that are directly related to technical change and which are controlled for the influences of changes in output and input prices on partial productivity. This results in what we call factor technical change (FTC) for each input. We

establish these FTCs from an integral framework in such a way that the derived measures are completely consistent with overall technical change.

The approach presented here is strongly connected with the research on the various types of technical progress or digress. This type of research goes back to Schumpeter (1942) and also Hicks (1932) and Solow (1970). Generally, in the literature three types of technical change are being distinguished: neutral technical change, input biased technical change and output biased technical change. Technical change is called Hicks neutral if the ratios of the various marginal outputs to inputs are unchanged due to technical change. Note that, in that case further decomposition of TFP into FTCs is unnecessary, since FTCs will be equal for all inputs. Technical change is called input-biased if the relationship between input demand and input prices is affected by the introduction of new technologies. Putting it differently, this implies that input demand, controlled for changes in outputs and input prices, changes in time due to technical change. Note that FTCs may differ accordingly. Technical change is called output-biased if the relationship between input demand and levels and mix of outputs is affected by the introduction of new technologies. A combination of input and output biased is also possible.

It is important to note that FTC differs from (partial) factor productivity growth. Factor productivity growth is the change of output divided by the volume change of an input. The volume change of a input depends on the change in output, technical change and substitution effects due to changes in relative prices. Thus factor productivity might include effects from substitution (for instance due to changes in the level or mix of output or input prices). It is expected that, for instance, in a market with increasing wages (partial) labour productivity will be enhanced by substituting capital for labour. The same holds for changing output levels and mix. The FTC measures to what extent factor productivity changes due to technical change and can be regarded as factor productivity after controlling for output levels and input prices. To put it

differently, we decompose non-neutral technical change into components related to various distinct inputs.

This chapter is organized as follows. Section 2 gives a brief review of previous literature. Section 3 contains a description of the model and derives the relevant expressions for the decomposition of technical change into various FTCs. Section 4 and 5 give an application of the model for the Dutch hospital industry. It discusses the data, the estimation techniques and the outcomes of the model. Section 6 concludes.

6.2 Literature review

There is a large body of literature on the efficiency and productivity of hospitals. Most studies on hospitals focus on the effects of environmental pressures on hospital efficiency and productivity, such as payment systems, competition, and property rights (Arocena & Garcia-Prado, 2007; Barros, 2003; Bates et al., 2006; Farsi & Filippini, 2008; Ferrier & Valdmanis, 2008; Mutter & Rosko, 2008; Pilyavsky et al., 2006; Puenpatom & Rosenman, 2008; Sari, 2008; Varkevisser et al., 2008). Other studies pinpoint their attention to economic phenomena, such as economies of scale, economies of scope, economic behaviour, and expense preference (Bazzoli, 2008; Blank & Eggink, 2004; Blank & Merkies, 2004; Dervaux et al., 2004; Linna & Häkkinen, 2008). The influence of managerial and organizational aspects, such as outsourcing or the size of departments, is the central focus in a number of other studies (Ancarani et al., 2009; Bilodeau et al., 2004; Hikmet et al., 2008; Ludwig et al., 2009; Ludwig et al., 2010).

Only a limited number of studies focus on estimating technical change of hospitals (Barros, 2007; Blank & Vogelaar, 2004; Blank & Van Hulst, 2009; Ferrier & Valdmanis, 2008; Kittelsen et al., 2008). From these studies only Blank & Vogelaar (2004) and Blank & Van Hulst (2009) explicitly refer to the different natures of technical change (i.e. input- and/or output-biased). Even

less attention is paid to the increase of labour productivity. From a policy point of view, however, the labour productivity growth is at the centre of attention. With an aging population, the demand for healthcare is growing, while a decreasing labour force makes it hard to tap sufficient labour. An interesting exception is a contribution of Ozcan et al. (1996), in which they address the development of labour efficiency in hospitals with an explicit reference to the labour market of hospital personnel.

As said, this chapter introduces a method for decomposing TFP growth into factor technical changes (FTCs) of distinct inputs. The decomposition is based on a cost function framework. It is argued that the cost minimizing first order conditions derived from the cost function provide the necessary information for the decomposition. The method is elaborated for a translog cost function (see Christensen et al., 1973) and applied to Dutch hospitals.

6.3 Model description

We define the FTC of input *n* as the relative change in usage of input *n* at a given level of output and given input prices due to technical change. For reasons of convenience the decomposition is based on a cost function framework. Furthermore we assume that the firm is a cost minimizing decision making unit. Rather similar derivations of the decomposition can be made in case of other representations than the cost function, such as revenue functions or indirect cost functions (Färe & Primont, 1995).

Assume the cost structure can be represented by a well-defined cost function:

$$C = c(y, w, T) \tag{1}$$

With:

C = total cost;

y = output (vector of dimension M);

w = input prices (vector of dimension N);

T = time (reflecting technical change).

And $c(y, w, T)$ is a twice differentiable function with respect to w and T, which satisfies the requirements concerning monotonicity and concavity (see e.g Färe and Primont 1995).

The monotonicity requirement refers to non-decreasing costs at a price increase: when the price of an input increases (at equal other prices) the costs cannot decrease (non-decreasing). The concavity requirement states that the total costs cannot increase by a higher percentage than the original amount of an input multiplied by the price increase of that input.

By definition the volume of input n is given by:

$$x_n \equiv \frac{S_n \cdot C}{w_n} \qquad n = 1, .., N \tag{2}$$

With:

x_n = volume of input n;

S_n = cost share n;

w_n = input price for input n.

For reasons of convenience we rewrite equation (2) in logarithms:

$$\ln x_n = \ln S_n + \ln C - \ln w_n \tag{3}$$

The total differential of (3) with respect to T yields:

$$\frac{d \ln x_n}{dT} = \frac{\partial \ln S_n}{\partial T} + \frac{\partial \ln C}{\partial T} - \frac{\partial \ln w_n}{\partial T} \qquad n = 1, .., N \tag{4}$$

Input prices are exogenous, and not depending on technical change or any other time related variables, consequently the last term on the right hand side drops from the equation. Note that the independency between input prices and technical change or time does not mean that input prices cannot change over time, in fact in an empirical application it is most likely that input prices will change over time. We only assume that the changes in input prices are not a result of technical or other time related changes. It is however valid to question whether we can ignore effects similar to skill biased technical change (a new technology causes a rise in the demand for highly skilled labour, which in turn leads to a rise in earnings inequality), see Card and DiNardo (2002).

In the empirical application input prices are included and costs and cost shares are controlled for changes in input price, the volume of inputs is controlled for substitution effects. So after eliminating the last term of (4) the expression states that the relative change in input n equals the sum of the relative change in the cost share corresponding to input n and the relative change in cost, both with respect to the factor time.

Since we further assume that the firm is cost minimizing, Shephard's lemma holds (Shephard, 1953). From this we derive equation (5) for the identification of the optimal input demands:

$$S_n = \frac{\partial \ln C}{\partial \ln w_n} \qquad n = 1, .., N \tag{5}$$

Substituting (5) into (4) yields:

$$\frac{d \ln x_n}{dT} = \frac{1}{S_n} \frac{\partial \left[\frac{\partial \ln C}{\partial \ln w_n} \right]}{\partial T} + \frac{\partial \ln C}{\partial T} \qquad n = 1, .., N \tag{6}$$

Which is the basic expression for a FTC. The percentage change in input n equals the percentage change in cost due to overall technical change (second

term on the right-hand side), corrected for the relative annual change in the corresponding cost share (first term right-hand side).

We now apply the well-known translog cost function (Jorgenson et al, 1973). The cost function model consists of a translog cost function and the corresponding cost share equations. The model includes first and second order terms, cross terms between outputs and input prices on the one hand and a time trend on the other hand are also included. These cross terms with a time trend represent the possible different natures of technical change. Cross terms with outputs refer to output-biased technical change and cross terms with input prices to input-biased technical change.

$$\begin{aligned}\ln C = a_0 &+ \sum_{m=1}^{M} b_m \ln y_m + \sum_{n=1}^{N} c_n \ln w_n + \sum_{o=1}^{O} d_o \ln z_o \\ &+ \frac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M} b_{mm'} \ln y_m \ln y_{m'} \\ &+ \frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N} c_{nn'} \ln w_n \ln w_{n'} \\ &+ \frac{1}{2}\sum_{o=1}^{O}\sum_{o'=1}^{O} d_{oo'} \ln z_o \ln z_{o'} + \sum_{m=1}^{M}\sum_{n=1}^{N} e_{mn} \ln y_m \ln n \\ &+ \sum_{o=1}^{O}\sum_{n=1}^{N} f_{on} \ln z_o \ln w_n + \sum_{o=1}^{O}\sum_{m=1}^{M} g_{om} \ln z_o \ln y_m \\ &+ h_0 T + h_1 T^2 + \sum_{m=1}^{M} i_{1m} T \ln y_m + \sum_{n=1}^{N} j_{1n} T \ln w_n\end{aligned} \tag{7}$$

With:

C = total costs;

y_i = output i ($i = 1,.., M$);

w_i = price of input i ($i = 1,.., N$);

z_i = fixed input i ($i = 1,.., O$);

T = time;

$a_0, b_m, c_n, d_o, b_{mm'}, c_{nn'}, d_{oo'}, e_{mn}, f_{on}, g_{om}, h_0, h_1, i_{1m}, j_{1n}$ parameters to be estimated.

And applying Shephard's Lemma we find the optimal input demands :

$$S_n^0 = c_n + \sum_{n'=1}^{N} c_{n'n} \ln w_{n'} + \sum_{m=1}^{M} e_{mn} \ln y_m + \sum_{o=1}^{O} f_{on} \ln z_o + j_{1n} T \quad n = 1,..,N \tag{8}$$

With:

S_n^0 = optimal cost share for input n ($n = 1,.., N$) .

Homogeneity of degree one in prices and symmetry is imposed by putting constraints on some of the parameters to be estimated. In formula:

$$b_{mm'} = b_{m'm} \; ; \quad c_{nn'} = c_{n'n} \; ; \quad d_{oo'} = d_{o'o} \; ;$$

$$\sum_{n=1}^{N} c_n = 1 \; ; \qquad \sum_{n=1}^{N} c_{nn'} = 0 \;\; (\forall n') \; ; \qquad \sum_{n=1}^{N} e_{mn} = 0 \; (\forall m) \; ; \tag{9}$$

$$\sum_{n=1}^{N} j_{1n} = 0 \; ; \qquad \sum_{n=1}^{N} f_{on} = 0 \; (\forall o).$$

We differentiate the translog cost function with respect to input price and time to find an expression for the first right-hand side term of expression (6). This is fairly simple since the translog cost function (7) contains only one term that depends on both input price and time:

$$\frac{\partial\left[\frac{\partial \ln C}{\partial \ln w_n}\right]}{\partial T} = j_{1n} \tag{10}$$

Substituting (10) into (6) yields:

$$\frac{d \ln x_n}{dT} = \frac{j_{1n}}{S_n} + \frac{\partial \ln C}{\partial T} \qquad n = 1, \dots, N \tag{11}$$

Equation (11) shows that in case of a translog cost function the FTC of input *n* equals the autonomous growth of cost and a correction factor for the extent of input- and output-biased technical change. It can easily be verified that a weighted aggregate of FTCs using cost shares as weights equals overall technical change. This follows from the notion that the sum of j_{1n} equals zero and the sum of the cost shares (S_n) equals one.

Both right-hand side terms of (11) can be elaborated. For the first term it is possible to substitute S_n with (8). For the second right-hand term is the first derivative with respect to T, in case of the translog cost function this yields the following expression:

$$\frac{\partial \ln C}{\partial T} = h_0 + 2h_1 T + \sum_{m=1}^{M} i_{1m} \ln y_m + \sum_{n=1}^{N} j_{1n} \ln w_n \tag{12}$$

6.4 Application to Dutch hospitals

The Dutch hospital industry

This section starts with a brief description of the Dutch hospital industry, some choices made in the application are a result of the specific Dutch context and data availability. In the Dutch hospital sector there are three types of hospitals: general hospitals, academic hospitals and categorical hospitals. Here the research is limited to the general hospitals; the characteristics of academic hospitals and categorical hospitals differ too much from general hospitals. Including these hospitals implies adding heterogeneity to the data. Moreover the vast majority of hospitals in the Netherlands are general hospitals, comprising about 80% of hospital beds and almost 70% of Dutch hospital costs. A general hospital is a concentration of facilities for diagnostics, treatment and nursing of patients. Other activities of general hospitals include training of physicians and nurses.

Since 2005, the Dutch hospital industry has a system with product classification. Patients are classified based on their diagnoses and treatments. The products that are derived from this classification is the so called DTC (diagnose treatment combination). From an economic perspective there are two types of DTCs, the A-segment for which the price is regulated and set by the government and the B-segment for which prices are negotiated by hospital and health insurance providers. Hospitals negotiate with health insurers on price, volume and quality.

In modelling the costs of hospitals we have to pay special attention to the costs of physicians. Some physicians are employed by the hospital, but most physicians are entrepreneurs who cooperate with the hospital. Costs and funding of the physicians and the hospital are strictly separated. One drawback of this arrangement is that the data on physicians are incomplete and including physicians in an empirical application needs some special attention. This also means that data on costs should be corrected for the costs of the physicians who are employed by the hospitals.

All hospitals are required to present annual reports containing information on costs, output and some specific hospital characteristics. Besides the annual reports there is a yearly survey containing information on specific inputs and also some hospital characteristics. Data from the annual reports are freely available; the additional data from the survey were obtained from the Dutch general hospital association (NVZ). The data from both sources are combined into one dataset suitable for analysis. The combined dataset contains information on Dutch hospitals during 2003 to 2011. The dataset has been checked for observations with unreliable or missing data; this resulted in the removal of 15 observations. Furthermore, during the period of analysis there have been a couple of mergers. As a result of mergers and removal of observations an unequal number of observations for each year remain. The final dataset has 672 observations, about 75 hospitals per year.

Outputs

Hospital output consists of the health outcomes of patients. However data on health outcomes are not available. Instead we use a more common approach in which the output of hospitals is measured by the number of admissions and outpatients. Including only admissions and outpatients as output measures lacks the notion of heterogeneity of hospital output. The case-mix of small general hospitals deviates from the bigger teaching hospitals. We therefore apply a hedonic-index (Lancaster, 1966) that accounts for the characteristics of the hospital. The hedonic-index is constructed from the following elements: relative size of surgery and orthopaedics, expected length of stay (based on the mix of specialties available in the hospital), number of intensive care (IC) beds, the presence of a psychiatric ward, the presence of neurosurgery and the presence of cardiothoracic surgery. The hedonic-index is a straightforward tool that accounts for case-mix differences among hospitals. The admissions included in the cost function are weighted by the hedonic index and credits hospitals with a more severe case-mix in accounting for cost differentials. Besides treating patients, hospitals deploy research activities and other services. This output has been measured by the revenues that these

activities and services generate; the revenues are deflated (i.e. adjusted for price-effects trough time).

A shortcoming of the measurement of output is that health outcomes and the number of treated patients are no synonym, to put it more strongly health outcome is poorly captured by output indicators such as admission and outpatients. Nevertheless there are reasons to assume that quality of Dutch hospital care has not decreased, as quality is constantly monitored, for instance by the health inspectorate, by patient associations and media, and is subject to interventions that raise quality.

Inputs and input prices

A common classification of inputs is labour, materials and capital. Since our aim is to obtain the FTC for inputs it makes sense to use a more detailed classification of the main input categories. Moreover, we want to distinguish inputs that are rather homogenous and have some characteristics that make them different form the other inputs. At the same time we want to be parsimonious with the number of categories, since each category induces several extra parameters to be estimated. Besides the classification of input categories is also limited by the data that are available.

Classification of the labour categories is based on cost homogeneity and matching of professions. Dutch hospitals use a standardized job classification system, the system provides main categories and subcategories of personnel. The classification including main categories and subcategories is available in the data, and suits the matching of professions. The following four labour categories result:

- Management and clerical staff;
- Nursing personnel;
- Paramedical personnel;
- Auxiliary personnel (such as maintenance, kitchen staff, security and cleaning personnel).

For each of the four distinguished labour categories there are data on the costs and the number of full time equivalents. Unit values are used as prices, for each category prices are calculated by dividing the corresponding costs and full time equivalents. The implicit assumption made here is that prices are exogenous for hospitals and all price variation comes from exogenous factors. In the sensitivity analysis we will pay attention to this assumption and present an alternative.

One labour input is missing: the physicians. As mentioned physicians are often entrepreneurs, as a result data on the costs of physicians are not included in the hospital data. Therefore physicians are omitted from the model. In the sensitivity analysis a circumventing construction is used to include physicians in the model.

For the material costs we distinguish two categories: costs that are directly related to patients (such as medical supplies, food and hotel cost) and other material costs (such as energy and general costs). We acknowledge that a more detailed categorization for medical supplies, especially including a separate category for medicine, would be more sophisticated. However a more detailed categorization is limited by the data availability, besides we want to be sparse. Since there is no natural unit of measurement for material supplies, a circumventing construction was used. For both categories we use price indices that are calculated and published by Statistics Netherlands. Price indices vary only over time, for our purpose this is adequate since there are no reasons to assume variations in prices of materials between regions. For both main categories of materials we have subcategories for which we have data on the costs at the hospital level and a price-index at the national level. The price-indices for both main categories are constructed as the weighted average of the price-indices of the subcategories. Price indices of subcategories are weighted by matching hospital specific cost shares.

For capital we use only one input category, hard as it is to find appropriate measures for capital. Capital refers to the capital assets such as buildings and

medical equipment. In the available data there are some indicators for the volume of capital stock. The volume of capital is measured as a weighted aggregate of beds, intensive care beds, psychiatric beds, square meters and the number of linear accelerators and cobalt units. The weights for each capital stock indicator are obtained by a regression of the capital cost on the indicators. The price of capital is defined as a unit value, derived from capital costs and the aforementioned volume of capital.

Table 6-1 Descriptive statistics, Dutch general hospitals 2011 (N=69)

Variable	Mean	Standard dev.
Admissions	45,055	20,834
Outpatient	76,967	29,281
Other revenues (in million €)	14.7	10.8
Surgery and orthopaedics (%)	11.7	2.4
Psychiatric beds per 1000 admissions.	0.27	0.40
IC-beds per 1000 admissions	0.23	0.09
Expected length of stay	3.3	0.3
Neurosurgery (%)	0.8	1.9
Cardiothoracic surgery (%)	0.3	0.9
Price management and clerical staff	53,898	2,799
Price nursing personnel	54,476	1,297
Price paramedical personnel	90,142	6,662
Price auxiliary personnel	40,491	1,637
Price patient related material costs	1.12	0.005
Price general material costs	1.16	0.008
Price capital	2.1	0.64
Total Cost ((in million €)	148.3	77.9
Cost share management and clerical staff	0.10	0.02
Cost share nursing personnel	0.34	0.04
Cost share paramedical personnel	0.03	0.02
Cost share auxiliary personnel	0.09	0.02
Cost share patient related material costs	0.21	0.03
Cost share general material costs	0.13	0.03
Cost share capital	0.09	0.02

Table 6-1 contains the descriptive statistics for the variables used in the cost function. The descriptive statistics are related to 2011, the most actual year in the dataset.

6.5 Estimation

The model is specified as a translog long-run cost function and cost share equations, which are derived from the cost function (see 7 and 8). The model is estimated as a multivariate regression system with various equations with a joint density, which we assume to be a normal distribution. The specification of the model has three output variables (with a correction for case-mix), seven inputs and year dummies to measure technical change. The choice for these variables is discussed in the previous section.

Because disturbances are likely to be cross-equation-correlated, Zellner's Seemingly Unrelated Regression method is used for estimation (Zellner, 1962). As usual, because the shares add up to one, causing the variance–covariance matrix of the error terms to be singular, one share equation in the direct cost function model is eliminated.

Since we are dealing with a relatively large number of cross sectional units and a limited number of periods, we ignore the fact that we are dealing with a panel data (with respect to intra firm correlations). Little harm is done here, since the between variance is far more relevant for the estimation than the within variance. We therefore pool all the data in one data set and control for the time varying effects by including a technology variable. For reasons of convenience technical change is often being modelled by using a time trend, implicitly assuming that technical change is a rather smooth process in time.

Previous research on Dutch hospitals (Blank & Vogelaar, 2004) shows that technical change appears shock wise, they argue that including a time trend in a cost function model is a rather restrictive way of incorporating technical

change. We follow the time trend approach, since that approach is in line with the way that the model is mathematically derived. The shock wise approach, with year dummies instead of a time trend, is included in the sensitivity analysis.

Homogeneity of degree one in prices and symmetry is imposed by adding restriction to the model. Aside from these imposed theoretical requirements, a few other requirements need to be fulfilled as well, such as monotonicity and concavity in input prices (Färe & Primont, 1995). These requirements can be tested posteriorly. An estimated cost function is monotonic in input prices if the fitted cost shares are positive. Concavity can be tested by exploring necessary and sufficient conditions for concavity.

Table 6-2 presents the parameter estimates.

Table 6-2 Estimates translog cost function model (N = 682)

Variable	Estimate	St. Error	T-value
Constant	0.270	0.016	17.14
Admissions	0.656	0.022	29.61
Outpatients	0.383	0.024	16.14
Other revenues	0.072	0.009	7.96
Admissions * Admissions	-0.160	0.069	-2.32
Admissions * Outpatients	0.162	0.067	2.42
Admissions * Other revenues	0.061	0.026	2.35
Outpatients * Outpatients	-0.239	0.092	-2.59
Outpatients* Other revenues	-0.012	0.027	-0.44
Other revenues* Other revenues	-0.021	0.009	-2.30
Price management and clerical staff	0.101	0.002	42.13
Price nursing personnel	0.335	0.004	87.37
Price paramedical personnel	0.049	0.002	26.87
Price auxiliary personnel	0.109	0.002	44.37
Price materiel patients	0.231	0.004	51.54
Price material general	0.078	0.003	25.18
Price capital	0.097	0.002	61.26
Price man. & staff. * price man. & staff.	0.055	0.006	9.36
Price man. & staff.* price nursing personnel	0.004	0.007	0.61

Variable	Estimate	St. Error	T-value
Price man. & staff. * price medical personnel	-0.004	0.002	-1.92
Price man. & staff. * price auxiliary personnel	-0.010	0.002	-5.44
Price man. & staff. * price materiel patients	-0.030	0.008	-3.78
Price man. & staff. * price material general	-0.007	0.006	-1.13
Price man. & staff.* price capital	-0.008	0.002	-3.69
Price nursing personnel * price nursing personnel	0.088	0.015	5.98
Price nursing personnel * price medical personnel	0.011	0.003	3.26
Price nursing personnel * price auxiliary personnel	-0.023	0.003	-7.57
Price nursing personnel * price materiel patients	-0.021	0.014	-1.55
Price nursing personnel * price material general	-0.034	0.011	-3.14
Price nursing personnel * price capital	-0.025	0.003	-7.66
Price medical personnel * price medical personnel	0.012	0.002	5.82
Price medical personnel * price auxiliary personnel	0.003	0.001	2.10
Price medical personnel * price materiel patients	-0.009	0.004	-2.42
Price medical personnel * price material general	-0.010	0.003	-3.74
Price medical personnel * price capital	-0.003	0.001	-2.26
Price auxiliary personnel * price auxiliary personnel	0.026	0.002	10.90
Price auxiliary personnel * price materiel patients	0.006	0.003	2.09
Price auxiliary personnel * price material general	0.002	0.002	1.02
Price auxiliary personnel * price capital	-0.004	0.001	-3.17
Price medical supplies *price materiel patients	0.024	0.044	0.55
Price materiel patients * price material general	0.034	0.041	0.83
Price materiel patients * capital	-0.004	0.004	-1.18
Price material general * price material general	0.028	0.040	0.69
price material supplies * price capital	-0.013	0.003	-4.53
price capital * price capital	0.057	0.002	34.27
Admissions * price man. & staff.	0.00003	0.003	0.01
Admissions* price nursing personnel	-0.018	0.004	-4.11
Admissions * price medical personnel	0.011	0.002	4.51
Admissions * price auxiliary personnel	-0.010	0.003	-3.12
Admissions * price materiel patients	0.031	0.005	6.82
Admissions * price material general	-0.003	0.003	-1.15
Admissions * capital	-0.009	0.002	-5.07
Outpatients * price man. & staff.	0.006	0.003	1.87
Outpatients * price nursing personnel	0.012	0.005	2.31
Outpatients * price medical personnel	0.003	0.003	1.17
Outpatients * price auxiliary personnel	0.007	0.004	1.80

Variable	Estimate	St. Error	T-value
Outpatients * price materiel patients	-0.033	0.005	-6.24
Outpatients * price material general	-0.008	0.004	-2.30
Outpatients * capital	0.012	0.002	5.87
Other rev. * price man. & staff.	-0.001	0.001	-0.47
Other rev. * price nursing personnel	-0.001	0.002	-0.59
Other rev. * price medical personnel	0.005	0.001	5.49
Other rev. * price auxiliary personnel	-0.004	0.001	-3.05
Other rev. * price materiel patients	0.0005	0.002	0.27
Other rev. * price material general	0.003	0.001	2.36
Other rev. * capital	-0.003	0.001	-3.99
Time	-0.030	0.005	-6.24
Time squared	0.001	0.001	1.11
Trend * price man. & staff.	-0.0005	0.0003	-1.41
Trend * price nursing personnel	-0.001	0.001	-2.45
Trend * price medical personnel	-0.001	0.0003	-2.19
Trend * price auxiliary personnel	-0.002	0.0004	-5.60
Trend * price materiel patients	0.003	0.001	4.87
Trend * price material general	0.002	0.0004	4.11
Trend * price capital	-0.001	0.0002	-2.44
Admissions * Surgery and orthopaedics	-0.158	0.026	-5.98
Admissions * Psychiatric beds	0.013	0.003	3.99
Admissions * IC-beds	0.018	0.011	1.67
Admissions * Expected length of stay	0.241	0.063	3.80
Admissions * Neurosurgery	0.028	0.004	6.57
Admissions * Cardiothoracic surgery	0.098	0.010	9.64
Admissions * Cardiothoracic surgery	0.140	0.012	11.920
R2 cost function	0.98		
R2 management and clerical staff	0.21		
R2 nursing personnel	0.29		
R2 paramedical personnel	0.37		
R2 auxiliary personnel	0.30		
R2 medical supplies	0.46		
R2 material supplies	0.10		
R2 capital	0.66		

Almost three quarters (74%) of the parameter estimates are significant at the 5% level. Estimated parameters also have the expected signs. Since the fitted cost shares are positive for all firms, the theoretical condition for monotonicity is satisfied for all inputs. A necessary condition for concavity of the cost function is that the own partial elasticities of substitution is less than zero for all inputs. This necessary condition for concavity of the cost function holds for all the inputs. A sufficient condition is that the matrix of partial elasticity's of substitution is negative semi-definite. A matrix is negative semi-definite if all eigenvalues are all less than or equal to zero. The eigenvalues of the matrix of partial elasticity's of substitution can be calculated for each observation. The matrix of partial elasticity's of substitution is negative semi-definite for 87% of the observations, 13% of the observations fail the sufficient conditions. For these observations the sufficient condition is too tight.

A quick inspection of the estimated parameters of the output variables shows that the average hospital faces diseconomies of scale ($\sum b_m = 1.11$). An increase of one per cent in outputs for the average hospital corresponds to a 1.11% increase in total costs. Furthermore the cost flexibilities, the responsiveness of the costs to changes in output, can be calculated for the individual hospitals.

Also note that the cross parameters of time and input prices (j_{in}) are significant in six out of seven cases. These parameters are a key element in the formula of the FTCs (see equation 11). Furthermore we tested the model against the alternative model with the cross parameters of time and input prices set to zero. Based on the likelihood ratio test the alternative model is overwhelmingly rejected (likelihood of 11,712 and 11,672). We therefore find that technical change is input biased.

Results factor technical change

Applying (10) and (11) and using the parameter estimates of Table 6-2 yield the following overall technical change and FTCs per type of input (see Table 6-3).

Table 6-3 Index factor technical change per type of input (2003=100)

	Overall	Management and staff	Nursing personnel	Paramedic personnel	Auxiliary personnel
2003	100.0	100.0	100.0	100.0	100.0
2004	102.8	103.3	103.2	105.4	104.8
2005	105.6	106.6	106.4	110.2	109.9
2006	108.4	109.9	109.6	115.2	115.3
2007	111.2	113.3	112.9	120.6	120.9
2008	114.0	116.8	116.2	124.4	126.9
2009	116.8	120.2	119.5	128.9	133.1
2010	119.4	123.4	122.7	133.3	139.9
2011	122.1	126.7	125.9	138.0	146.8
	Material patients	**Material general**	**Capital**		
2003	100.0	100.0	100.0		
2004	101.5	100.8	103.5		
2005	102.9	101.4	107.1		
2006	104.3	102.0	110.6		
2007	105.6	102.7	114.0		
2008	107.0	103.3	117.7		
2009	108.4	103.8	121.1		
2010	109.6	103.9	124.9		
2011	110.7	104.2	128.2		

Table 6-3 shows that the technical change between 2003 and 2011 was 22.1%, with an annual technical change of 2.8%. FTC per type of input differs from this measure. The FTC for nursing personnel and management and clerical staff are slightly higher than the general productivity measure, with an increase of respectively 26% and 27% during 2003-2011. The increase of the FTC for capital is even a little bit higher than this with an increase of 28%.

Significant differences between FTC and total technical change exist for both paramedical personnel and auxiliary personnel. The FTC for auxiliary personnel is substantial higher than the FTC of all other inputs, during 2003-2011 the FTC increase is 47%. In the past period hospitals have been reducing the costs of auxiliary personnel, during the period of analysis the volume of auxiliary personnel dropped roughly with 15%. For paramedical personnel there is also an increase of FTC that is higher than total technical change, 38% during 2003-2011. For paramedical personnel we should keep in mind that this is a rather small group for which fluctuations can have an high impact.

Most notable is the development of the FTC for both medical supplies and material supplies, for both material inputs FTC lags behind. The lack of FTC gains for an input itself is not a bad thing. Merely, it reflects the increased or decreased relative importance of an input due to technical change.

Sensitivity analysis

We performed a couple of sensitivity analyses to get insight into the reliability of the results. In the main analysis a detailed classification of the inputs has been introduced. In this sensitivity analysis we use the common classification in tree inputs: labour, materials and capital. The results are shown in Table 6-4.

Table 6-4 Index factor technical change per type of input (2003=100)

	Overall	Labour	Materials	Capital
2003	100.0	100.0	100.0	100.0
2004	102.6	103.3	101.2	103.3
2005	105.2	106.7	102.2	106.6
2006	107.8	110.1	103.3	109.8
2007	110.3	113.5	104.3	112.9
2008	112.8	116.9	105.2	116.3
2009	115.2	120.3	106.1	119.3
2010	117.6	123.8	106.9	122.6
2011	119.8	127.1	107.6	125.5

From Table 6-4 it becomes clear that aggregating inputs has a minor impact on the results. The estimated total technical change is a little bit lower when aggregating inputs. The annual difference is about 0.3 percent point and accumulates to a difference of 2.2 percent point for the whole period. Capital is the only input that is not further aggregated in this sensitivity analysis. For this input the difference of FTC is in line with the differences for total technical change, it is only a little lower than the model with seven inputs.

For labour the sensitivity analysis demonstrates the added value of disaggregating; the FTC of labour as an aggregate is a crude measure for the FTC underlying labour categories. For most labour categories deviations are small, but for auxiliary personnel the deviation is huge. For this category the development has been quite different than for other labour categories. For materials the FTC lies between the FTC of both underlying categories. This is of course not totally unexpected since the FTC will approximately equal the weighted average of its underlying inputs and in this case a little bit lower since general productivity change is estimated lower.

As mentioned the data does not allow us to distinguish physicians as an input directly. In this sensitivity analysis we use a circumventing approach. Data on the costs of independent physicians is obtained by multiplying average profits of independent physicians by the number of physicians (since independent physicians have no wage, profits are a good indication). The national Statistics Netherlands has data on average profits of 15 different specialisms. For physicians on the payroll of the hospital we take a fixed amount based on the collective labour agreement for these physicians. Doing so we have rather crude estimates on costs an prices of physicians that can be incorporated in the model. The data are sufficient for this sensitivity analysis, however analysis of FTC of physicians needs more accurate and detailed information. Table 6-5 shows the results of including physicians.

Table 6-5 Index factor technical change per type of input (2003=100)

	Overall	Management and staff	Nursing personnel	Paramedic personnel	Auxiliary personnel
2003	100.0	100.0	100.0	100.0	100.0
2004	102.6	103.2	103.1	103.6	104.8
2005	105.1	106.4	106.2	107.1	109.8
2006	107.5	109.5	109.3	110.5	115.1
2007	109.9	112.7	112.4	114.2	120.6
2008	112.2	115.9	115.5	117.5	126.5
2009	114.4	119.1	118.6	120.7	132.7
2010	116.7	122.0	121.6	122.9	139.2
2011	118.9	125.1	124.7	125.6	146.0
	Material patients	**Material general**	**Capital**	**Physicians**	
2003	100.0	100.0	100.0	100.0	
2004	101.3	100.8	103.6	101.7	
2005	102.5	101.6	107.1	103.3	
2006	103.6	102.2	110.5	104.8	
2007	104.7	102.8	113.8	106.3	
2008	105.7	103.4	117.4	107.8	
2009	106.6	103.7	120.6	109.3	
2010	107.6	104.0	124.4	110.2	
2011	108.6	104.4	127.6	110.9	

Including physicians results in less productivity growth, this follows from the FTC of physicians that lags behind (total productivity is a weighted average). Furthermore we see a remarkable difference for the FTC of paramedic personnel that dropped with 12.5 percent point. This suggests that the FTC calculated for paramedics is less robust than other FTCs and should be interpreted with care.

Another sensitivity analysis is performed by modelling the prices of labour differently. As stated before we use unit values, assuming that all price variations are exogenous. Although it is common practice to use unit values, there is something to say for an approach that assumes (partly) endogenous price variation. The Dutch labour market is regulated, therefore in case of wages equal exogenous conditions are plausible. However, some regional

variation might be expected as there are differences between regional labour markets. We now drop the assumption of complete exogenous price differences and assume that price differences are a result of regional exogenous factors and endogenous factors. We therefore calculate annual regional prices, or market price, for labour as the average price in a region in a year. Differences between the regional price and the actual observed price of a hospital are ascribed to endogenous factors. The model is estimated with the regional prices.

Not only do the results for the FTCs change, the statistics of the model also change. Most noticeable is that the sufficient condition for concavity fails for all observations (either the model is invalid or possibly the sufficient condition for concavity is too tight). If we look at the FTCs we observe clear changes. For instance the FTC for management and clerical personnel is estimated 10 percent point higher, while the FTC of nurses drops with 7 percent point. Other FTCs also change to some extent. One possible explanation for the different results is that the FTCs as calculated in the base model include trade-offs between the price of inputs and its productivity. For example a hospital can decide to hire more expensive but also more experienced staff, assuming that higher wages pay of in better productivity by more experienced staff. The sensitivity analysis corrects for these trade-offs since it excludes endogenous price effects. From a model perspective the changes are due to changed parameter estimates for the interaction term of time and input ($j_{11} - j_{17}$). However we have to be careful to conclude here since we are not sure about the statistical properties of the model (concavity). What we can conclude is that modelling is sensitive for data on prices, in interpreting results one has to take in account how prices are incorporated in the model.

Finally we tested a model in which the time trend is replaced by a so called technology index, represented by a set of dummy variables for each year and weighted by (estimated) weights (Blank & Vogelaar, 2004). Based on the log

likelihoods (11,711 for the trend model and 11,716 for the shock wise model) we cannot conclude that the shock wise model has to be preferred over the trend model. The big difference between the trend model and the shock wise model is that the shock wise model shows a higher productivity growth in 2004, while in 2010 the productivity growth is less. Aside from that there are a couple of small deviations. As expected over the whole period of analysis the deviations balance each other out. For the FTCs there are yearly differences that are in line with the differences as observed for total technical change. For the whole period the deviations are rather small with differences with a magnitude of 1 percent point for most inputs and maximum of 1.6 percent point.

In general the sensitivity analysis supports the results of the base model. Table 6-6 summarizes the results of the sensitivity analysis and shows the average annual technical change and the average annual FTC for nursing personnel. The absolute magnitude of the differences of average annual technical change are small. An exception is the model with adapted prices, for this model the differences are noticeable. The sensitivity analysis shows that some care has to be taken in statements about the magnitude of FTC.

Table 6-6 Average annual FTC (Overall and nursing personnel)

	Overall	Nursing personnel
Base Model	2.8	3.2
Aggregated inputs [a]	2.5	3.4
Physicians included	2.4	3.1
Regional Prices	2.1	2.4
Shock wise technical change	2.8	3.4

a, the FTC presented for nursing personnel is the FTC for labour

6.6 Conclusions

This chapter presents a framework for the decomposition of technical change into FTCs. The FTC of input n is defined as the relative change in usage of input n at a given level of output and given input prices due to technical change. FTC differs from (partial) factor productivity growth, it is a pure measure of the effect of technical change instead of a ratio of output and input that also includes substitution effects. An expression for the FTCs has been derived from the cost function. The FTC of an input equals the autonomous growth of cost and a correction factor for the extent of input-biased technical change.

We applied the model to the Dutch hospital industry. The empirical model consists of a cost function with three outputs and an additional case-mix indictor and seven inputs. The results show that due to technical change productivity increased yearly with an average of 2.8%. Furthermore the results show that technical change is input-biased and that FTC differs amongst the various inputs. The FTC of nursing personnel has increased at a slightly higher rate than total technical change and increased with 3.2% per year. Interesting are the results for the FTC of materials. For materials we distinguished two categories, one category of costs that are directly related to patients and one more general category. For both categories we observe that the FTC lags behind with an annual average increase of respectively 1.3% and 0.5 %. Besides that we have to be careful with interpretation here, since prices used for materials are a crude approximation, we can speculate that this is a compound result of two opposite forces. In the period of analysis we observe a remarkable decline of the average length of stay[9]. Since one of our output indictors is the number of admissions it is clear that a shorter length of stay is

[9] During 2003-2011 the average length of stay per admission dropped from 7 days to 5 days. Including day care the decline was even faster with an averages of 4.5 days in 2003 and 2.9 days in 2011.

an important explanation of increased productivity (an admission consists of less nursing's days). At the same time an admission becomes more intensive (more materials are used per nursing day). Apparently for materials this results only in small gains in FTC, suggesting that of these inputs relatively more has been used.

Productivity is often neglected in workforce planning models, sometimes completely ignored, sometimes assumed as some arbitrary growth rate. And that is a pity, since future demand for health professionals will be tempered by raised productivity. Therefore FTCs can help policymakers to improve forecasts of the demand for health professionals. The computation of FTCs in a cost function framework does justice to the multi-input multi-output production process of hospitals, without using ex ante weights for the products. Furthermore the decomposition of FTCs is consistent with the concept of total technical change. It incorporates input-biased technical change and therefore it takes into account that changes in productivity might vary between inputs. And finally FTCs purely measure the effect of technical change, it is adjusted for substitution effects resulting from price effects.

References

Ancarani, A., Di Mauro, C., & Giammanco, M.D. (2009). The impact of managerial and organizational aspects on hospital wards' efficiency: Evidence from a case study. European Journal of Operational Research, 194(1), 280-293.

Arocena, P., & Garcia-Prado, A. (2007). Accounting for Quality in the Measurement of Hospital Performance: Evidence from Costa Rica. Health Economics, 16(7), 667-685.

Balk, B.M. (2001). Scale Efficiency and Productivity Change. Journal of Productivity Analysis, 15(3), 159-183.

Balk, B.M. (2003). The Residual: On Monitoring and Benchmarking Firms, Industries, and Economies with Respect to Productivity. Journal of Productivity Analysis, 20, 5-47.

Barros, C.P. (2007). An analysis of hospital efficiency and productivity growth using the Luenberger indicator. Health Care Management Science, 2007.

Barros, P.P. (2003). Random Output and Hospital Performance. Health Care Management Science, 6(4), 219-227.

Bates, L.J., Mukherjee, K., & Santerre, R.E. (2006). Market Structure and Technical Efficiency in the Hospital Services Industry: A DEA approach. Med Care Res Rev, 63(4), 499-524.

Bazzoli, G.J. (2008). Hospital consolidation and integration activity in the United States. In J.L.T. Blank & V.G. Valdmanis (Eds.), Evaluating hospital policy and performance: contributions from hospital policy and productivity research (Vol. 18, pp. 45-61). Oxford: Elsevier.

Bilodeau, D., Cremieux, P.-Y., Jaumard, B., Ouellette, P., & Vovor, T. (2004). Measuring Hospital Performance in the Presence of Quasi-fixed Inputs: An Analysis of Quebec Hospitals. Journal of Productivity Analysis, 21(2), 183-199.

Blank, J.L.T., & Eggink, E. (2004). The Decomposition of Cost Efficiency: An Empirical Application of the Shadow Cost Function Model to Dutch General Hospitals. Health Care Management Science, 7, 79-88.

Blank, J.L.T., & Merkies, A.H.Q.M. (2004). Empirical assessment of the economic behaviour of Dutch general hospitals. Health Economics, 13(3), 265-280.

Blank, J.L.T., & van Hulst, B.L. (2009). Productive Innovations in Hospitals: An Empirical Research on the Relation between Technology and Productivity in the Dutch Hospital Industry. Health Economics, 18(3), 665-679.

Blank, J.L.T., & Vogelaar, I. (2004). Specifying technical change: a research on the nature of technical change in Dutch hospital industry. The Quarterly Review of Economics and Finance, 44, 448-463.

Canadian Nurses Association (2009). Tested solutions for eliminating Canada's registered nurse shortage: Canadian Nurses Association.

Card, D., & DiNardo, J.E. (2002). Skill biased technological change and rising wage inequality: some problems and puzzles: National Bureau of Economic Research.

Christensen, L.R., Jorgenson, D.W., & Lau, L.J. (1973). Transcendental Logarithmic Production Frontiers. The Review of Economics and Statistics, 55(1), 28-45.

Dervaux, B., Ferrier, G.D., Leleu, H., & Valdmanis, V.G. (2004). Comparing French and US hospital technologies: A directional input distance function approach. Applied Economics, 36(10), 1065-1081.

Diewert, W.E. (1981). The Theory of Total Factor Productivity Measurement in Regulated Industries. New York: Academic Press.

Evans, R.G., Schneider, D., Barer, M.L., & Foundation, C.H.S.R. (2010). Health human resources productivity: what it is, how it's measured, why (how you measure) it matters, and who's thinking about it: Canadian Health Services Research Foundation.

Färe, R., & Primont, D. (1995). Multi-Output Production and Duality: Theory and applications. Dordrecht: Kluwer Academic Publishers.

Farsi, M., & Filippini, M. (2008). Effects of Ownership, Subsidization and Teaching Activities on Hospital Costs in Switzerland. Health Economics, 17(3), 335-350.

Ferrier, G.D., & Valdmanis, V.G. (2008). Efficiency and productive changes in large urban hospitals 1994-2002: ownership, markets, and the uninsured. In J.L.T. Blank & V.G. Valdmanis (Eds.), Evaluating hospital policy and performance: contributions from hospital policy and productivity research (Vol. 18, pp. 157-176). Oxford: Elsevier.

Grifell-Tatjé, E., & Lovell, C. (1999). A generalized Malmquist productivity index. TOP, 7(1), 81-101.

Health Workforce Australia (2012). National health workforce productivity modelling -Final Report Adelaide: Health Workforce Australia.

Hicks, J.R. (1932). Theory of wages. London: MacMillan.

Hikmet, N., Bhattacherjee, A., Menachemi, N., Kayhan, V.O., & Brooks, R.G. (2008). The role of organizational factors in the adoption of healthcare information technology in Florida hospitals. Health Care Management Science, 11, 1-9.

Kittelsen, S.A.C., J. Magnussen, Anthun, K.S., Häkkinen, U., Linna, M., Medin, E., . . . Rehnberg, C. (2008). Hospital productivity and the Norwegian ownership reform – A Nordic comparative study Working paper 2008. Oslo: University of Oslo.

Kopp, R.-J., & Diewert, W.E. (1982). The Decomposition of Frontier Cost Function Deviations into Measures of Technical and Allocative Efficiency. Journal of Econometrics, 19(2-3), 319-331.

Lancaster, K.J. (1966). A new approach to consumer theory. Journal of Political Economy, 74(2), 132-157.

Linna, M., & Häkkinen, U. (2008). Benchmarking Finnish Hospitals. In J.L.T. Blank & V.G. Valdmanis (Eds.), Evaluating hospital policy and performance: contributions from hospital policy and productivity research (Vol. 18, pp. 179-190). Oxford: Elsevier.

Ludwig, M., Groot, W., & Van Merode, F. (2009). Hospital efficiency and transaction costs: A stochastic frontier approach. Social Science & Medicine, 69(1), 61-67.

Ludwig, M., Van Merode, F., & Groot, W. (2010). Principal agent relationships and the efficiency of hospitals. The European Journal of Health Economics, 11(3), 291-304.

Mutter, R.L., & Rosko, M.D. (2008). The impact of ownership on the cost-efficiency of U.S. hospitals. In J.L.T. Blank & V.G. Valdmanis (Eds.), Evaluating hospital policy and performance: contributions from hospital policy and productivity research (Vol. 18, pp. 113-138). Oxford: Elsevier.

O'Donnell, C.J. (2010). Nonparametric Estimates of the Components of Productivity and Profitability Change in US Agriculture Working Paper Series. Brisbane: The University of Queensland.

OECD (2015). http://www.oecd.org/els/health-systems/workforce.htm Retrieved 13-1-2015, 2015

Ono, T., Lafortune, G., & Schoenstein, M. (2013). Health workforce planning in OECD countries: a review of 26 projection models from 18 countries: OECD Publishing.

Ozcan, Y.A., Yeh, S.C., McCollum, D., & Begun, J. (1996). Trends in labor efficiency among American hospital markets. Annals of Operations Research, 67(1), 61-81.

Pilyavsky, A.I., Aaronson, W.E., Bernet, P.M., Rosko, M.D., Valdmanis, V.G., & Golubchikov, M.V. (2006). East-West: Does It Make a Difference to Hospital Efficiencies in Ukraine? Health Economics, 15(11), 1173-1186.

Puenpatom, R.A., & Rosenman, R. (2008). Efficiency of Thai Provincial Public Hospitals during the Introduction of Universal Health Coverage Using Capitation. Health Care Management Science, 11(4), 319-338.

Sari, N. (2008). Competition and market concentration. In J.L.T. Blank & V.G. Valdmanis (Eds.), Evaluating hospital policy and performance: contributions from hospital policy and productivity research (Vol. 18, pp. 139-156). Oxford: Elsevier.

Schumpeter, J.A. (1942). Capitalism, Socialism and Democracy. New York: Harpers Collins.

Shephard, R.W. (1953). Theory of Cost and Production Functions. Princeton, New Jersey

Solow, R.M. (1970). Growth Theory. An Exposition. New York: Oxford University Press.

Varkevisser, M., Capps, C.S., & Schut, F.T. (2008). Defining hospital markets for antitrust enforcement: new approaches and their applicability to The Netherlands. Health Economics, Policy and Law, 3, 7-29.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. Journal of the American Statistical Association 57, 348–368.

7

Conclusions

7.1 Introduction

Raising productivity is a promising alternative in curbing the rising costs of healthcare. Improvement of productivity in healthcare can be achieved by applying cost-saving technologies. Besides advances in technology, scale and efficiency have their impact on productivity. This thesis examines how scale, efficiency and new technology enhance productivity of hospitals.

This chapter presents the conclusions of this thesis. First the main findings are presented. The first topic of the main findings is the optimal scale of hospitals. Next are the effects of corporate governance on efficiency. Then the effect of new technology on productivity is discussed. Since effects of new technology are expected to be inter-temporal, there is special attention for this topic. Next there is attention for factor productivities. Finally there are some general results on technical change in the Dutch hospital industry. After the main findings, the policy implications follow. The chapter concludes with opportunities for future research.

7.2 Main findings

The optimum scale

This thesis includes a meta-analysis to get insight on the optimum scale, 41 parametric studies on the cost structure of hospitals (95 model outcomes) and their results on scale are analysed. The main characteristics of the studies are inventoried systematically. The optimum scale is derived from the scale elasticity, which varies with the scale. The scale elasticity relates a proportional growth in costs to a proportional growth in production. A scale elasticity of one means constant returns to scale, at that point the hospital is at its optimal scale.

First regression analysis is applied, to estimate how the scale elasticity depends on study characteristics, including the number of beds as an indicator

for scale. Second, from the estimates we determine the number of beds for which the scale elasticity equals one. Since the regression analysis includes various study characteristics, we have to choose values for these characteristics. In other words we need a reference study for which the optimum applies. The reference study is constructed bases on the most frequent used characteristics For the reference study the optimum lies around 320 beds.

Sensitivity analysis shows how the optimum shifts as different characteristics are applied. For the analysis on studies only a limited number of modelling choices have a significant effect at the 10% level. For the analysis on observations confidence interval are more tight, resulting in more significant effects even at more severe significance level. Especially the number of inputs included in the model and the use of a long-run cost function (instead of indirect results derived from the short-run function) have an impact on the optimum scale.

Besides parametric studies, non-parametric studies are also analysed. Non-parametric studies rarely report the scale elasticity, the optimum scale or a range for the optimum scale is reported directly. However only few non-parametric include these results, we only found 10 studies that reported these results. For another 9 studies we were able to get a rough estimate for the optimum scale. Based on these 19 studies we find an under bound of 220 beds for the optimum scale. Furthermore the non-parametric studies reveal that context matters. The results for non-parametric studies compare rather well with the results for the parametric studies. This is because we have to compare with parametric frontier studies, for which the optimum scale is estimated at 239 beds.

The thesis also provides specific results on the scale of Dutch hospitals. The results indicate that the average Dutch hospital nowadays is too big. A decade ago, the average hospital operated at around the optimum scale. Since then scale has increased, resulting in hospitals that, from an economic perspective, are too big and operate under diseconomies of scale. The average

scale efficiency is a measure that indicates how much can be saved by operating at the optimum scale. The average scale efficiency for Dutch hospitals in 2007 is 87.5%. This is a combined effect of hospitals that are too small (4% of the population) and too big (80% of the population).

Governance and efficiency

The efficiency of a hospital is determined by comparing its performance with a so-called best practice. Efficiency gives an indication of the potential savings to be made, showing either how much more can be produced with the same amount of resources or how the same amount can be produced with fewer resources. Efficiency can be calculated using several methods. This thesis uses the non-parametric method of data envelopment analysis (DEA). More interesting than the efficiency scores themselves, however, are the factors which explain differences in efficiency. In a second stage this study uses the bootstrap method from Simar and Wilson (2007) to analyse the effect of corporate governance on efficiency.

The average efficiency score of Dutch hospitals is 78% under constant returns to scale (CRS) and 89% under variable returns to scale (VRS), respectively. This is in line with general results from efficiency research on hospitals. Nguyen and Coelli (2009) use meta-analysis to compute an average efficiency score of 84% for hospitals. However, efficiency is something relative; it tells us how good a performance is compared with that of an estimated best practice. Furthermore, that estimated best practice is derived from observations in a sample. There is little point, therefore, in comparing the average efficiency score from one study with that from another. The average efficiency score does, however, tell us something about the dispersion within the sample. An average efficiency score of 89% indicates that the dispersion in this study is rather moderate, compared to other efficiency studies and considering that maximum efficiency is 100%.

Differences in efficiency can be related to hospital characteristics, doing so provides an insight in the determinants of efficiency. This thesis relates the

differences in efficiency to the governance characteristics of a hospital. There is no general economic framework for evaluating governance structure and the efficiency of business entities. This thesis distinguishes the following clusters of governance characteristics: the management board, the supervisory board and multi-actor dependencies. Operationalization of governance is based on available data, such as such as the size of the board, its remuneration, et cetera.

The results show that part of the cost efficiency can be explained by governance characteristics. However, most governance characteristics as measured in this study correlate with the size of the hospital. We therefore find more significant effects of the governance characteristics if we assume constant returns to scale (CRS) rather than variable returns to scale (VRS). In fact, the only characteristic which has an effect under both assumptions is the remuneration of the supervisory board: higher remuneration corresponds with a worse performance in terms of efficiency. The frequently suggested professionalization of the supervisory board will not be accomplished through a higher remuneration.

Furthermore, the remuneration of the executive board also has little to do with the efficiency of the hospital. Under the CRS assumption, a higher remuneration even implies less efficiency. However, part of this result is due to scale effects: a higher remuneration corresponds with a bigger hospital. Under the CRS assumption, part of the inefficiency results from unaccounted scale effects. Under the VRS assumption, a higher remuneration also has a tendency to lessen efficiency, although the effect is not significant. Nevertheless, this result also shows that the opposite effect – higher remuneration of the executive board leads to higher efficiency – is invalid.

New technology and productivity

The effect of technical change on productivity is usually computed by adding a time trend or year dummies to a model. All changes in productivity through time, excluding scale and efficiency effects, are absorbed by the time trend or year dummies. Therefore, the estimated effect of technical change is a

mishmash of all kinds of changes over time. This study focuses on new technologies and the reorganization of processes, which – for the sake of convenience – are called innovations here. Instead of modelling technical change with time only, this study adds innovations.

The study uses information on 69 innovations adopted by Dutch hospitals during the period 1995-2005. Because hospitals do not adopt innovations at the same time – there are earlier adopters and laggards – the adopted innovations are used to construct individual technology indices for hospitals. The innovations have a wide range and are quite heterogonous, and so are grouped into six homogenous clusters (six technology indices). The technology indices are added to a cost-model, to model technical change in a more sophisticated way. From the estimates of the cost model it is possible to derive the impact on productivity for each technology index.

Innovations that improve productivity are typically found in the ICT cluster and the chain-care cluster (all activities in the treatment programme are geared to another). Productivity losses are associated with product innovations aimed at raising quality (in terms of better health outcomes or less stressful treatment for patients). Note that the negative effect on productivity from some innovations is partly a consequence of how production is measured. In this study, that is done by output rather than outcome, an approach that does not fully capture the quality factor.

Inter-temporal effects

So far, innovations have been regarded as exogenous: we have ignored the fact that hospitals actually decide on the innovations they adopt. The adoption of an innovation has two effects: adjustment costs and inter-temporal savings. Typically, adjustment costs are temporary, whereas cost savings are structural. Adopting an innovation is therefore an inter-temporal decision, in which adjustment costs are weighed against future savings. This has modelling consequences: an additional equation on the optimum amount of innovations

can be added to models in order to obtain more reliable estimates and estimates for the optimum amount of innovations.

The model including this additional equation is applied to a dataset of Dutch general hospitals operating during years 1995-2005. In order to stress its relevance, this model has been compared with a traditional version and with one that includes adoption but has no additional equation. Models including adoption outperform the traditional one, and the model with the additional equation performs even better.

Factor productivity

Technical change not only influences productivity, but it may also affect the optimal mix of inputs. This is referred to as input-biased technical change. Input-biased technical change indicates implicates that some inputs are substituted for others. This thesis shows how factor productivity can be calculated, adjusting for substitution effects. The thesis includes an application for the Dutch hospitals during 2003-2011. During 2003-2011 technical change is input biased. It appears that, taking the substitution effect into account, the factor productivity for labour outpaced total productivity, productivity associated with materials was lower than other inputs.

General results on technical change

Throughout this thesis there are results on technical change in Dutch hospitals during the last decade. Here the results are summarized. First we look at the effect of technical change on productivity:

- Chapter 4 analyses the period 1995-2002 with a time trend. The annual improvement in productivity found is 1.9%.
- Chapter 5 analyses the period 1995-2005 with a time trend. The annual improvement in productivity found is 2.5%.
- Chapter 6 analyses the period 2003-2011 with year dummies. The annual improvement in productivity ranges from 0.9 % to 3.5%, with an average of 2.1%.

Differences in results are explained by the period of analysis, small differences in the measurement of production and additional assumptions on technical change. This last difference includes assumptions on the individual technology index. Therefore, the results are not fully comparable. Overall, the results can be summarized as consistent with a persistent productivity growth of about 2% per year from 2003 until 2011.

Next we look at the nature of technical change (neutral, input-biased, output-biased). Technical change is non-neutral, it affects the optimum allocation of inputs as well outputs. From our results, it is clear that technical change is non-neutral, although mixed results are found regarding input and output bias. The results from Chapter 4 indicate that technical change is output-biased, while those from Chapter 6 indicate that technical change is input-biased. Chapter 5 finds that technical change is both input and output-biased. This may support the idea that technical change is both input as well output-biased.

The results found for the output bias in Chapter 4 and Chapter 5 are consistent. In general, the output bias implies that the marginal costs of admission for specialities with an above-average length of stay are rising. The opposite applies to the treatment of outpatients, for whom the marginal costs decline due to technical change.

The results regarding the input bias found in Chapter 5 and Chapter 6 differ on a few points. There are a couple of reasons for these differences. Chapter 6 has a specific focus on the inputs, modelling eight of them, compared with six in Chapter 5. Furthermore, Chapter 5 has a different period of analysis and uses a subsample of hospitals. From Chapter 6, we conclude that the input-biased technical change for administrative personnel, nursing personnel and capital implies relatively more use of these resources. There was relatively less use of physicians and none of medical materials (i.e. energy and general costs).

7.3 Policy implications

Productivity gains in healthcare enable policymakers to get a grip on the rising costs of healthcare, without compromising on accessibility and quality. Therefore, productivity gains in healthcare are a policy objective. This thesis examines the three sources of productivity growth: scale, efficiency and technical change. In addition the factor productivity of inputs is analysed. The research is applied to Dutch hospitals, so policy implications from this study essentially apply first and foremost to the Dutch hospital industry.

Some of the data used for the applications in this thesis might be qualified as somewhat outdated, therefore todays relevance and actuality of the results are addressed. Chapter 3 relates efficiency of hospitals to the governance of the hospital. The results on efficiency are stable, Blank et al. (2011) show that over a period of seven years the average efficiency of Dutch hospitals fluctuates, however there are no structural changes for the average efficiency. Furthermore the annual fluctuations of the average efficiency are small, in general about 0.5 percentage point. As for the governance of healthcare organisations, Stoopendaal and van de Bovenkamp (2015) note that this is a topic that is on the agenda in many countries. It is however not exclusively the relation between efficiency and governance that is the topic of research, the relation between governance and quality of care is also a popular research topic. Chapter 4, 5 and 6 deal with technological change. Section 7.2 summarizes and compares the results on technological change from the three chapters. For different periods there are small differences in technological change. There are several explanations for these differences, one is that technological changes occurs shock-wise, see also Blank and Vogelaar (2004). At the same time the results of the three chapters compare well and show that over a period of 16 years, productivity increased with a about 2 per cent per year as a result of technical change. It is unsure whether this growth recently continued. From Blank and van Heezik (2016) it can be concluded that after 2011 the productivity of Dutch hospitals slightly decreased. The differences in

technological change per period supports the relevance of chapter 4 and 5, these chapters not only measure productivity growth, but also focus on the underlying factors of productivity growth. Chapter 6 identifies a higher productivity growth for labour than for the other inputs, from Blank and van Heezik (2016) it can be concluded that this trend continued in recent years, since volume growth of materials and capital outpaced the volume growth of labour. All together there are no reasons to doubt about the relevance of the results for today's policy.

Before policy implications are addressed, a warning is in order: productivity gains are not synonymous with cost savings. Productivity growth is a medicine with possible side-effects. Productivity growth tells us only that the ratio of outputs to inputs has grown, implying more output per unit of input. If we are interested in cost savings, however, we should also keep an eye on growth in production. The productivity of Dutch hospitals has grown by 2% a year, but during the same period there has also been a sizeable growth in production (more than one would expect on demographic grounds alone). In real terms, therefore, the total costs of hospital care have increased in spite of productivity gains. It is true that costs would have been even higher had there been no productivity gains. But there is a possibility that what induced productivity improvement also induced more production, resulting in a simultaneous growth of productivity and costs. If the objective is cost saving, stimulating productivity growth might overshoot its original goal if production increases even faster in its slipstream.

The first recommendation is about producing at an optimum scale. The optimum scale is often mistaken for producing on a larger scale. However, there are limits to economies of scale: at about 300 beds, diseconomies of scale prevail. Policymakers might therefore also consider reducing the scale of hospitals, or limiting their size, as an instrument to increase productivity. In the Dutch hospital industry, scales have increased beyond the optimum level, so

that most Dutch hospitals are experiencing diseconomies of scale. Instead of increasing the scale, downsizing might have a positive impact on productivity.

The second policy recommendation relates to the governance of hospitals, in particular the remuneration of the board and supervisory board. Recent reforms in the Dutch hospital industry have implied less regulation and the introduction of regulated competition. As a side-effect of these reforms, the relationship between government and hospitals has changed. In this new setting, governance has gained in importance. Although there is a governance code in the healthcare sector that provides guidance, individual hospitals still retain a high degree of freedom in arranging their own governance. In the case of the governance characteristics examined here, there is no great need for further regulation as most have no effect on efficiency. In part, this is because the differences in characteristics are small and because most of the time they are adapted to the size of the hospital.

However, one finding is rather interesting from a policy perspective. That is the result concerning the remuneration of the executive and supervisory boards. Since the reforms, a heated debate has arisen about their remuneration. The national government has tried to restrain the remuneration of executives in the public sector, including hospitals. This has generated opposition from interested parties. The argument they use is that their remuneration has to compete with that in the private sector, otherwise it would be impossible to attract capable executives and supervisors. Or, to put it more simply: better performance requires higher remuneration. The results of this research show that, from the efficiency perspective, this argument is invalid. Higher remuneration of the board does not increase the efficiency of the hospital. In fact, the opposite seems true, although the effect is not statistically significant. In the case of the supervisory board, higher remuneration in general implies less efficiency. This result is in line with Cardinaels (2009), who notes that the monitoring function of the supervisory board is hampered if its remuneration becomes excessive. Consequently, policymakers should not worry too much

about the argument that the substantial remuneration of executives and supervisors is a guarantee of performance. Rather, they should worry about the opposite.

The overall efficiency found in this study is 89%, in 2007. Based on this, one might argue that an improvement of 11% is still possible. However, realizing the full potential is an utopia. And, quite apart from that, there is the fact that we also have to determine what the determinants of inefficiency are. A more realistic goal is an overall improvement by a couple of percentage points, achieved by focusing on those hospitals which are really lagging behind in efficiency. This contrasts with technical change, which has resulted in tremendous productivity gains, with a persistent average growth in productivity of 2% per year. It would thus seem that technical change has the greatest potential.

The logical next question is: what is technical change? This thesis uses innovations to model technical change. One of the findings of this study is that innovations increase productivity. However, not all innovations improve productivity. From a cost-savings perspective, there are innovations that save costs and innovations that push them. Cost saving is only one possible reason; often, an innovation aims to improve quality and/or safety or to enable the treatment of previously untreatable conditions. In addition, it is also possible that new technologies may lead to a medical "arms race". Healthcare firms compete with each other with the latest medical technologies, possibly resulting in underutilization of the technology (Luft et al., 1986). To make things even more complicated, the cost savings induced by innovation might not occur immediately. The thesis identifies ICT and chain-care as innovations that increase productivity. However, this result is ambiguous, because it does not apply for all products. Increasing productivity with innovations is therefore less straightforward, instead some effort has to be made to identify the innovations that really increase productivity.

Securing future accessibility to healthcare requires a sufficient future supply of health workers. The short-term supply of health workers is inelastic, due to the qualification requirements. At the same time, a surplus of health workers is undesirable because of the costs involved in training them. Therefore, an adequate forecast of the future demand for health workers is desirable. Ideally, planning the future number of health workers would take productivity gains into account. Most of the time, though, productivity forecasts are rough approximations that do not differentiate between different resources or occupational groups. But in practice there are big differences in the development of individual resources. Moreover, there are substitution effects due to changing relative prices. Therefore, policymakers should consider using more detailed and specific information on productivity gains in forecasts of the future demand for health workers.

7.4 Limitations and future research

This study has limitations. Some are intrinsic to this field of research, others specific to this study. A couple of them are already mentioned between the lines, the most important are singled out below.

First of all, the measurement of output is – as always – an issue. In this study, the focus lies on that what hospitals produce and not on the underlying objective of that production. In other words, production is measured as the number of patients treated, while the actual goal is better health. The number of patients treated is only a proxy for better health. One implication of this is that new technologies might have contributed more to productivity than is found in this study. A treatment that leads to better health outcomes does not necessarily improve productivity as measured in terms of patients treated, while it is quite possible that it would improve productivity were production to be measured in terms of health outcomes.

A similar reasoning applies to the omission of quality from the measurement of production. Newhouse (1994) states that differences in efficiency between hospitals might be explained by differences in quality. This might very well be true, but it is questionable whether it is problematic. The quality of hospitals is monitored continuously by the healthcare inspectorate. This guarantees that the level of quality meets the standards set by policymakers. The policy objective is to deliver good care rather than to maximize quality of care. In any case, some studies indicate only a tenuous relationship between efficiency and quality (Dismuke & Sena, 1999; Zuckerman et al., 1994). On the other hand, Ludwig et al. (2010) find that efficiency and quality do go hand in hand.

Whether it is wise to concentrate future research on other measurements of hospital output and incorporating quality is debatable. Typically, these limitations are intrinsic to productivity analysis in the health sector. It will be very hard, if not impossible, to find data on health outcomes that can be related to the efforts of hospitals. For that reason, incorporating quality indicators might be moving on less slippery ground, and this is an approach promoted by several studies – for instance, Jacobs and Dawson (2003). But as indicated previously, it is questionable whether there are excessive differences in quality between hospitals. Leaving a comprehensive discussion on incorporating quality to others, the point to be made here is that, where this study finds that some innovations have no effect on productivity, or a negative one, this should be placed in the context of the measurement of production. More obvious research opportunities are studies to map the effects of innovations on quality and health outcome.

A second limitation is that we do not know whether what we observe as technical change is a movement towards an unobserved frontier or a shift of the frontier itself. To underpin this, this study uses the adoption of innovations to model technical change. From this it follows that innovations improve productivity, although the effects found are rather small and do not fully

capture enhanced productivity. This means that, for the remaining part, technical change is either a movement towards an unobserved frontier or that this study has missed some innovations that have pushed the frontier. This last possibility seems quite possible, since the core of the innovations included in the research are related to medical procedures and the treatment of patients. At the same time there is no doubt that the enormous growth in productivity due to technical change is in large part explained by a sizeable decline in the average length of inpatient stays and by a tremendous growth in day-care treatment[10].

Here lies an opportunity for future research. Focusing on process innovations, rather than a complete palette of innovations, might explain more about the factors which have increased productivity. Naturally, this requires a preliminary research stage that identifies those process innovations which aim to boost productivity. To be more specific: if the reduction in the average length of stay and the growth in number of day-care patients explain the productivity growth, how did hospitals accomplish this? What adaptations have been made to reduce the length of stay? An additional argument for future research in this area is that the data on innovations used in this study is up to date only until 2005, while major changes in funding started in 2006. The reforms to the funding system changed incentives at the same time as a sizeable growth in production and productivity is observed. The question is whether the new funding system introduced incentives that stimulated hospitals to move towards an (unobserved) frontier that was already there, or whether it introduced incentives to innovate and push the frontier? An additional research opportunity lies in the measurement of technology. In this study, a plain index of the number of unweighted innovations has been used. It

[10] Between 1995 and 2011, the average length of inpatient stays decreased by approximately 46%, an average decline of 3% per year. In the same period, the number of daycare patients increased by approximately 248%, an average growth of 15% per year.

is not unreasonable to hypothesize that a weighted index would generate better results.

A third limitation relates to the measurement of governance. In this study, that is limited by the available data on governance. Meanwhile, policymakers have been trying to steer governance in the right direction by introducing codes and additional guidance for good governance. There are more aspects to governance than have been explored in this study. Evidence of the impact of these on performance, in terms of efficiency, remains unknown. So extending research in this direction is certainly an option.

Finally, the results found for the factor technical change are retrospective – and, as we all should know, past results are no guarantee of future performance. Care should therefore be taken in applying factor technical change in forecasts of the labour market. The main point made in this study is that the level of aggregation of resources matters, and should be considered in forecasts of productivity gains.

References

Blank, J.L.T., Dumaij, A.C.M., & van Hulst, B.L. (2011). Ziekenhuismiddelen in verband. Een empirisch onderzoek naar productiviteit en doelmatigheid in de Nederlandse ziekenhuizen 2003-2009 IPSE Studies Research Reeks. Delft: TU Delft.

Blank, J.L.T., & van Heezik, A.A.S. (2016). Productiviteit van Overheidsbeleid Deel II: De Nederlandse zorg, 1980-2013. Delft: IPSE Studies.

Blank, J.L.T., & Vogelaar, I. (2004). Specifying technical change: a research on the nature of technical change in Dutch hospital industry. The Quarterly Review of Economics and Finance, 44, 448-463.

Cardinaels, E. (2009). Governance in non-for-profit hospitals: Effects of board members' remuneration and expertise on CEO compensation. Health Policy, 93(1), 64-75.

Dismuke, C.E., & Sena, V. (1999). Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portuguese public hospitals? An empirical analysis using parametric and non-parametric methods. Health care management science, 2(2), 107-116.

Jacobs, R., & Dawson, D. (2003). Hospital efficiency targets. Health Economics, 12(8), 669-684.

Ludwig, M., Van Merode, F., & Groot, W. (2010). Principal agent relationships and the efficiency of hospitals. European Journal of Health Economics, 11(3), 291-304.

Luft, H.S., Robinson, J.C., Garnick, D.W., Maerki, S.C., & McPhee, S.J. (1986). The role of specialized clinical services in competition among hospitals. Inquiry, 83-94.

Newhouse, J.P. (1994). Frontier estimation: How useful a tool for health economics? Journal of Health Economics, 13(3), 317-322.

Nguyen, K.-H., & Coelli, T. (2009). Quantifying the effects of modelling choices on hospital efficiency measures: a meta-regression analysis: School of Economics, University of Queensland, Australia.

Simar, L., & Wilson, P.W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 136(1), 31-64.

Stoopendaal, A., & van de Bovenkamp, H. (2015). The mutual shaping of governance and regulation of quality and safety in Dutch healthcare. Health Services Management Research, 28(1-2), 9-15.

Zuckerman, S., Hadley, J., & Iezzoni, L. (1994). Measuring hospital efficiency with frontier cost functions. Journal of Health Economics, 13(3), 255-280.

Enhancing hospital productivity

Healthcare expenditure in Western countries is substantial and outpaces economic growth, therefore cost containment in healthcare is high on the political agenda. One option is to increase productivity in healthcare, do more with less. This thesis uses the Dutch hospitals as a case-study and examines the three cornerstones of productivity: scale, efficiency and technical change. Based on meta-analysis it is concluded that there are no economies of scale for hospitals beyond 320 beds. Furthermore there are indications that the optimum size is significant smaller. Analysis of the efficiency of Dutch hospitals shows that there are only marginal possibilities for improvement of the efficiency. Technical change is a collective noun for productivity changes resulting from the overall process of invention, innovation, diffusion of technology and institutional changes. Although productivity consistently increased with about 2% per year as a result of technical change, it is difficult to pinpoint the innovations that contributed most to this growth. In general innovations in the field of ICT and chain care have positively contributed to productivity; productivity loss is associated with innovations aimed at improving quality. Furthermore, the thesis shows that innovations have an initial phase in which they hamper productivity; it takes time before hospitals can fully benefit from innovations.