Efficiency Gains from Mergers in the Healthcare Sector

Preface

The three-volume paper “Efficiency Gains from Mergers in the Healthcare Sector” is the tenth in the Research Paper Series by the Dutch Healthcare Authority (NZa). The Research Paper Series aims at the enhancement of knowledge and expertise in the regulation of and competition in health care markets. The papers in this series are written by invited authors and/or NZa staff.

The liberalization of healthcare markets triggers an increase in competition regulation issues. The authorities control the necessary prerequisites of effective competition and intervene in case it is violated. One of the authority’s tasks in this field is to scrutinize merger request of healthcare agents. There is a wide literature on the methods that estimate the expected effects of the integration. The three-volume research paper “Efficiency gains from mergers in the healthcare sector” focuses on the positive effects of integrations and moves from purely economic theory over measurement methods to actual program implementations. Due to the large field covered, it is organized in three coherent volumes (Part A: Economic theory, Part B: Modelling, Part C: Implementation).

The paper discusses potential gains of both horizontal and vertical mergers and considers intermediate network integrations and efficiency changes due to a split of organizations. The developed models measure the total gains, but also offer a distinction between learning, harmony and size effects. There are a number of extensions of the models that enhance the practical relevance of the measures. It is possible to impose restrictions on the controllability or transferability of selected factors and services. One can also adjust for quality dimensions that by their nature or by the dimensionality problem cannot be directly included in the base models. Finally the model allows taking into consideration post merger technical inefficiencies or alternative strategies of the merged entity.

As an illustration, the paper contains preliminary calculations of the potential efficiency gains stemming from horizontal mergers between Dutch hospitals. That part demonstrates how the models and the various assumptions work and gives an interpretation of the results.

The present research paper is the first step in the development of a better underpinned evaluation of mergers. The model and its results still have to be built in the formal evaluation process. NZa intends to describe in a separate document the way NZa can use the estimation results in its argumentation and in formulating its viewpoints on specific cases. The stakeholders can be also involved in the discussions about the general considerations on a merger evaluation.

While the present paper focuses on the methodology, the separately issued document will detail the procedural questions. It will elaborate the precise role of the various authorities (Netherlands Competition Authority, Dutch Healthcare Authority, and Netherlands Health Care Inspectorate) involved in the evaluation process, and the routine of data share between the same authorities.

Part A of the research paper is written by Katalin Katona (Dutch Healthcare Authority, NZa), while the model development and applications (Part B and C) are the work of Professor Peter Bogetoft (Copenhagen Business School, CBS).

The authors would like to thank Rolf Färe and Shawna Grosskopf for their interesting input and collaboration with Peter Bogetoft on these issues, Janneke Everaers for her contribution to write the section on foreclosure and vertical restraints, Rein Halbersma for the many useful suggestions on the research paper, Michiel Verkoulen (project leader), and Misja Mikkers for their regular reading and valuable comments. The authors would also like to thank the reviewer group[1]: Victoria Shestalova (Netherlands Bureau for Economic Policy Analysis, CPB), Gulbahar Tezel and Freek Bruggert (Netherlands Competition Authority, NMa), Cees van Gent (Netherlands Health Care Inspectorate, IGZ), Jan Boone (Tilburg University) and Wolf Sauter (NZa) for their critical reviews of this series of research paper.

Frank de Grave
Chairman of the Executive Board
Dutch Healthcare Authority

[1] All opinions and views expressed in the research paper reflect the personal views of its authors. The members of the reviewer group do not necessarily agree on the results and conclusions of the research paper.

Samenvatting

Efficiëntievoordelen van Fusies in de Zorg

In de afgelopen jaren is op de Nederlandse zorgmarkten meer marktwerking geïntroduceerd. In deze periode hebben ook veel fusies plaatsgevonden, zowel aan de kant van zorgverzekeraars als aan de kant van zorgaanbieders. Ook zijn organisaties steeds meer gaan samenwerken. Doelmatigheidsoverwegingen spelen vaak een grote rol bij zulke beslissingen. Het is voor een overheid van belang deze mogelijke doelmatigheidswinsten te kunnen kwantificeren.

Fusies en samenwerkingsverbanden zijn onder te verdelen in twee groepen. In de eerste groep valt het samengaan van instellingen die actief zijn op dezelfde zorgmarkt, zoals de fusie tussen twee ziekenhuizen, verzekeraars of verpleeghuizen. In de tweede groep valt het samengaan van instellingen die op verschillende zorgmarkten actief zijn, zoals het samengaan van een verzekeraar en een ziekenhuis, of van een ziekenhuis en een verpleegtehuis. In economische terminologie zijn dit respectievelijk horizontale en verticale fusies.

In beide gevallen kan een fusie zowel positieve als negatieve effecten hebben. Bij positieve effecten kan gedacht worden aan verbeteringen in kwaliteit of doelmatigheid bij de negatieve effecten aan hogere prijzen als gevolg van toegenomen concentratie op de markt. Het nut van een fusie voor de consument is afhankelijk van welke effecten domineren. De negatieve effecten van toegenomen concentratie worden door de mededingingsautoriteit beoordeeld, terwijl de te fuseren partijen de positieve effecten moeten onderbouwen. Tot op heden hebben maar weinig partijen de mogelijke baten weten te onderbouwen, maar in de toekomst zijn betere doelmatigheids- en kwaliteitsargumenten te verwachten.

Het vooraf beoordelen van fusies en samenwerkingsverbanden vereist een voorzichtige analyse, omdat de bronnen van voor- en nadelen erg divers zijn. Daarnaast gaat het om voorspellingen die de toekomst betreffen. In een hypothetische situatie moeten twee vragen worden beantwoord: (1) wat zijn de verwachte effecten van de fusie en (2) wat zou er op de markt gebeuren, wanneer de fusie niet doorgaat? Vooral het bepalen van de impact van een fusie op kwaliteit, is een complexe aangelegenheid.

In de mededingingsliteratuur worden veel schattingsmethodes genoemd om de negatieve effecten van een fusie te kwantificeren. Tot op heden is minder aandacht besteed aan de mogelijke positieve effecten. Dit onderzoekspaper richt zich op de positieve effecten van het samengaan van zorgorganisaties. Het doel is het ontwikkelen van analytische en praktische methodes om de positieve effecten van fusies te kwantificeren en hiermee een lacune in de economische literatuur te vullen.

Tijdens het project hebben we een klankbordgroep van wetenschappers en medewerkers van de NMa en IGZ gevraagd om op onze model en onderzoekspaper commentaar te geven. Aandachtpunten waren de robuustheid en aannemelijkheid van het model. Dit is voornamelijk een theoretische studie. De soliditeit van het model wordt afgetast in aparte (pilot) projecten. Als vervolg op dit project zal de NZa in een visiedocument formuleren welke algemene afwegingen NZa in haar zienswijzen op fusies maakt. Daarover zal het veld worden geconsulteerd.

Het project en de resultaten vanuit een economisch perspectief

Met het project "Efficiëntievoordelen van Fusies in de Zorg" probeert de NZa de bronnen van doelmatigheidswinst bij fusies te kwantificeren. Het project vertrekt vanuit de economische theorie en komt, via analytische meetmethodes, uit bij een aantal implementaties. Het onderzoekspaper is op dezelfde manier gestructureerd en bestaat uit drie delen: (A) de theoretische achtergrond van mogelijke doelmatigheidswinsten, (B) de uitwerking van economische modellen van productie en organisatie van zowel zelfstandige als gefuseerde instellingen en (C) het implementeren van deze modellen in specifieke indicatoren en software.

Het hoofddoel van het project is het ontwikkelen van een methode tot het kwantificeren van de mogelijke positieve (doelmatigheids-)winsten van een fusie. Deel A is een economische introductie tot de delen B en C. Het analytische model zelf staat in deel B beschreven; deel C gaat over de geprogrammeerde implementatie van het model en geeft een eerste indicatie van de resultaten die hiermee behaald kunnen worden.

Het is belangrijk om op te merken dat er niet één juiste methode bestaat voor het schatten van de doelmatigheidswinsten van een fusie. Het algemene onderliggende idee is hetzelfde van model tot model, maar doordat er keuzes mogelijk zijn in de manier waarop de onderliggende technologieën gemodelleerd worden en in de manier waarop het model wordt geschat (parametrisch of non-parametrisch) zijn de uitkomsten verschillend tussen de modellen. Natuurlijk zijn ook niet alle mogelijkerwijs te schatten modellen in elk geval even toepasselijk, bijvoorbeeld omdat de (impliciete of expliciete) aannames niet overeen komen met de werkelijkheid.

Door een aantal verschillende modellen voor een bepaalde fusie te schatten, kunnen we analyseren hoe de resultaten van de schatting veranderen naar gelang de aannames. Hoe kleiner de verschillen tussen de schattingen zijn, hoe robuuster de resultaten. Bovendien kunnen we door vergelijking van de geschatte cijfers nagaan of onze aannames houdbaar zijn. Deze robuustheidtesten bevorderen ook de juridische houdbaarheid van de besluiten die op grond van de analyses genomen worden.

Het project en de resultaten vanuit een juridisch perspectief

De Nederlandse Mededingingsautoriteit (NMa) is in Nederland de bevoegde autoriteit ten aanzien van concentraties (fusies en overnames), ook die in de zorg. Voor de NZa is het afgeven van zienswijzen inzake voorgenomen concentraties, gelet op haar rol in de zorgsector, van groot belang voor de uitvoering van de eigen wettelijke taken, met name ten aanzien van de (markt)monitor en het Aanmerkelijke Marktmacht (AMM) instrument. Voor de NMa is samenwerking met de NZa van belang omdat de marktkennis waarover de NZa beschikt vaak van belang is in het kader van een door de NMa te nemen besluit. De zienswijze bevat ondermeer opmerkingen en aandachtspunten bij (de wijze van) marktafbakening en gaat in op de mogelijke gevolgen die de voorgenomen concentratie heeft op de werking van de markt zoals die binnen de doelstellingen van de Wet marktordening gezondheidszorg voor gereguleerde marktwerking is beoogd. Voor de invulling van het onderdeel ‘kwaliteit’ gaat de NZa af op het oordeel van de Inspectie voor de Gezondheidszorg (IGZ). De samenwerking tussen de NMa en de NZa is vastgelegd in een Samenwerkingsprotocol, en in Werkafspraken voor concentratiezaken. Zienswijzen van de NZa zijn openbaar. De NMa gaat in haar besluit op deze zienswijzen in, en zal een eventuele afwijking daarvan in het besluit motiveren.

Ten aanzien van machtsposities is tenslotte Art 18 van de Wet marktordening gezondheidszorg (Wmg) relevant. Hierin wordt bepaald dat bij samenloop in beginsel de NZa aan zet is (op basis van haar bevoegdheden ten aanzien van aanmerkelijke marktmacht (AMM)) tenzij op grond van doelmatigheid het voortouw beter bij de NMa of de NMa en NZa gezamenlijk kan worden gelegd. Een kenmerkend verschil tussen misbruik van economische machtspositie, waar de NMa bevoegd is, en AMM, is dat in het eerste geval ook feitelijk misbruik moet worden aangetoond, en in het geval van AMM alleen het bestaan van prikkels en mogelijkheden om onafhankelijk gedrag ten nadele van andere marktpartijen (uiteindelijk ten nadele van consumenten) te vertonen. In beide gevallen het noodzakelijk om de relevante markt af te bakenen en de mogelijkheid van onafhankelijk gedrag aan te tonen. Een hieraan gerelateerd verschil is dat misbruik van economische machtspositie doorgaan achteraf (ex post) zal worden geadresseerd, terwijl AMM ook ex ante, dus voordat sprake is van misbruik, kan worden vastgesteld.

De huidige studie kan met name van belang zijn voor de advisering van de NZa aan de NMa in het kader van de zienswijzen. Daarnaast kan zij relevant zijn voor eigen AMM analyses van de NZa en bij de monitoring van ontwikkelingen in specifieke zorgmarkten. Tenslotte is deze studie van belang om het debat te voeden over het beleid ten aanzien van verschillende soorten fusies en de rol die doelmatigheidsargumenten spelen bij de toetsing daarvan.

Samenvatting van deel A

Dit rapport behelst de eerste stap in de researchpaper: het identificeren van bronnen van doelmatigheidswinsten bij fusies met behulp van de economische theorie. Alternatieven voor fusies, die vergelijkbare doelmatigheidswinsten kunnen opleveren, worden ook besproken. Dit onderdeel van het project is achtergrondinformatie bij de onderdelen B en C en bij de praktijk van fusiebeoordelingen. Dit rapport geeft een inzicht in de mogelijke baten voor de consument. Verschillende economische theorieën komen aan de orde, zoals de neoklassieke aanpak, de transactiekostentheorie en literatuur op het gebied van management.

Bij het beoordelen van horizontale fusies (bijv. tussen twee ziekenhuizen) zijn schaal- en scope-effecten de voornaamste bronnen voor doelmatigheidswinsten. Zo kan de groei van een zorgorganisatie winst opleveren wanneer productiebenodigdheden ondeelbaar zijn. Een fusie kan bijvoorbeeld zorgen voor een minimumaantal aan patiënten dat nodig is om dure apparatuur doelmatig te benutten. Ook bestaat de mogelijkheid van specialisatie of het benutten van inkoopvoordelen. Door de fusie worden verschillende onderdelen van de zorgorganisatie samengevoegd. Hierdoor kan geld bespaard worden doordat vaste kosten (bijvoorbeeld op gebieden als inkoop en administratie) niet dubbel gemaakt hoeven te worden.

Bij verticale fusies zijn zorgorganisaties actief op verschillende maar vergelijkbare markten. Vaak zijn ze opeenvolgende stappen in de bevoorradingscyclus. De productie van de ene zorgorganisatie is in dat geval een productiemiddel van de volgende zorgorganisatie. In de gezondheidszorg kan een verticale relatie optreden wanneer twee aanbieders opeenvolgende stappen in een behandeling aanbieden. Zo kan na de diagnose- en behandelfases bij een serieuze ziekte een behoefte aan revalidatie of tijdelijke verzorging bestaan.

In dit geval zijn de partijen actief op verschillende markten. Zij zullen qua technologie, input en output minder op elkaar lijken. Schaal- en scopevoordelen zijn dan minder belangrijk. Het coördinatie-effect overheerst; de winst kan hem zitten in de harmonisatie van het productieproces (zo zouden een ziekenhuis en verpleegtehuis de bezetting van bedden beter op elkaar af kunnen stemmen bij een fusie) of van de onderlinge verhoudingen (geen contractpartijen meer, maar collega's). Strategisch biedt zo'n fusie het voordeel van minder risico's in de 'supply chain'. Vanuit de patiënt gezien bieden twee zorgaanbieders opeenvolgende stappen aan, terwijl zij niet op dezelfde markt actief zijn. Voordelen zijn dan te behalen doordat de patiënt maar één keer als klant hoeft te worden binnengehaald en dat de overdracht van informatie gemakkelijker verloopt.

Naast fusies bestaan ook andere vormen van samenwerking. In het geval van verticale relaties werken verticale beperkingen, zoals beperkingen op de prijs, hoeveelheid of een niet-lineaire prijsstelling, in meer of mindere mate alsof de instellingen gefuseerd zijn. In horizontale gevallen bestaan er ook voordelen voor de consument bij samenwerking. Zulke strategische allianties (bijv. in onderzoek en ontwikkeling) kunnen vorm krijgen met het gezamenlijk ontwikkelen van standaarden of via 'cross-licensing'.

In de delen B en C van dit paper bespreken we hoe we fusies kunnen modellen en hoe we de mogelijke doelmatigheidswinsten ook zo snel mogelijk kunnen kwantificeren.

Samenvatting van deel B

Dit rapport behelst de tweede stap in de researchpaper: het modelmatig beschouwen van fusies. Het bouwt voort op de theoretische aanpak uit deel A. Een belangrijke doelstelling is het kunnen kwantificeren van synergie-effecten. Deel C zal een demonstratie geven van de modellen die in dit onderdeel uitgebreid besproken worden.

In dit rapport ontwikkelen we modellen waarin we het productieproces kunnen volgen van instellingen, zowel voor als na een fusie. Een organisatie zet productiemiddelen in om productie of diensten te leveren. Hierbij valt te denken aan een huisarts, die zijn eigen tijd (arbeid), zijn praktijkruimte en materialen (kapitaal en medische technologie) en opleiding (menselijk kapitaal en kennistechnologie) inzet in het behandelen van patiënten (de productie). Met een model dat omzetting van productiemiddelen in diensten omschrijft (de productiefunctie), kunnen we de potentiële baten inschatten van:

- Horizontale fusies
- Verticale fusies
- Samenwerkingsverbanden en andere vormen van integratie
- Het opsplitsen van zorgorganisaties

Deze mogelijke doelmatigheidswinsten kunnen op verschillende manieren worden ingezet. Zo kan een zorgorganisatie hetzelfde blijven produceren met minder productiemiddelen. Zo zou een verzekeringsmaatschappij bij een fusie gebruik kunnen maken van het modernere ICT-systeem van de fusiepartner en zo de kosten per verzekerde drukken. Ook kan een zorgorganisatie er voor kiezen met dezelfde hoeveelheid productiemiddelen meer te gaan produceren. Zo kan de verzekeraar uit het vorige voorbeeld met evenveel administratiekosten als de twee onderdelen voorheen, meer verzekerden administreren. Door rekening te houden met verschillende opties om doelmatigheidswinsten te meten kunnen we effecten schatten als:

- De mogelijkheden tot meer efficiëntie na een fusie
- Een proportionele uitbreiding van de geleverde diensten
- Een proportionele inkrimping van de hoeveelheid gebruikte productiemiddelen
- Een uitbreiding of inkrimping van een deel van de geleverde diensten of gebruikte productiemiddelen
- Een mix van de bovenstaande effecten

De mix van uitbreiding van geleverde diensten en inkrimping van de gebruikte productiemiddelen leent van de gedachte dat bij een fusie het doel vaak zowel het besparen van kosten als groei in omzet is. Zo kan de verzekeraar uit het eerdere voorbeeld zich zowel richten op lagere beheerskosten als op een grotere hoeveelheid aan te trekken klanten. Deze aanpak is het meest algemeen en wordt in deel C dan ook uitgewerkt in software.

Hierna is het mogelijk de doelmatigheidswinsten van fusies te kwantificeren. De uitdaging is daarbij om de mogelijke doelmatigheidswinsten die puur voortkomen uit de fusie te onderscheiden van mogelijke doelmatigheidswinsten die ieder zorgorganisatie op zichzelf zou hebben. In het voorbeeld van de verzekeraar, zou de minder efficiënte verzekeraar ook het softwarepakket voor de administratie kunnen aanschaffen op de markt. Omdat fusies ook zorgen voor meer concentratie op de markt, is het van belang om de winsten die puur uit de fusie voortkomen te wegen tegen de kosten van een gestegen concentratie op de markt. We onderscheiden daarom:

- Leereffecten (doelmatigheidswinsten door zo efficiënt mogelijk te produceren)
- Scope of harmonie-effecten (een betere mix van inputs en outputs kiezen)
- Schaaleffecten (doelmatigheidswinsten doordat de gefuseerde zorgorganisatie groter is)

Bij een horizontale fusie kan een gedeelte van de leereffecten en schaaleffecten ook behaald worden door allereerste de bedrijfsvoering te verbeteren naar de ‘best practice’ en ten tweede door sommige activiteiten tussen verschillende zorgorganisaties te verplaatsen. Om dan de fusie te beoordelen, is kennis van de lokale situatie van belang.

Om de praktische relevantie van deze maatstaven te vergroten, komen ook manieren aan de orde om rekening te houden met:

- Alternatieve strategieën voor gefuseerde zorgorganisatie
- Technische ondoelmatigheid na een fusie
- Beperkte invloed van de gefuseerde zorgorganisatie op productiefactoren en diensten
- Beperkte verplaatsbaarheid van productiefactoren en diensten

Er kunnen omstandigheden bestaan die een organisatie verhinderen alle potentiële doelmatigheidswinsten te realiseren. Zo kan het bouwen van een nieuw gebouw voor gefuseerde zorgorganisatie jaren duren. In de tussentijd kunnen nog niet alle mogelijkheden worden benut. Ook externe factoren zoals regelgeving kunnen noodzakelijke reorganisaties bemoeilijken.

Bij een fusie is de samenstelling en de kwaliteiten van het (nieuwe) management ook iets wat gemodelleerd kan worden. We kunnen bijvoorbeeld veronderstellen, dat de beste managers de zorgorganisatie gaan besturen. Dat zou in een ideale situatie gebeuren. Het is meer realistisch aan te nemen dat het nieuwe management de gemiddelde prestatie gaat tonen. Of dat de prestatie van het nieuwe management gelijk is aan die van de grootse gefuseerde zorgorganisatie (omdat deze doorgaans meer invloed hebben in de onderhandelingen over de fusie)..Met de hierboven genoemde aandachtspunten (externe factoren, vermogen van het nieuwe management) kan het model beter worden toegepast op de realiteit.

Dit rapport beschrijft een van de manieren om de doelmatigheidswinsten van fusies te schatten. De aanpak is zo algemeen mogelijk. Op basis van de situatie bij de fusie kan dan het meest toepasselijke model worden gekozen en toegesneden op de situatie.

De uitwerking in dit onderdeel veronderstelt dat we een beeld hebben van de doelstellingen van zorgorganisaties. In deel C beschrijven we hoe dit met wiskundige en statistische technieken kunnen schatten. Hier wordt ook een implementatie in software besproken.

Samenvatting van deel C

Dit rapport behelst de derde stap in de researchpaper: implementeren van de in deel B besproken modellen en het kwantificeren van potentiële productiviteitsverbeteringen. Het rapport bespreekt ook de complicaties die kunnen optreden bij het toevoegen van kwaliteitsgegevens in de modellen en geeft ook een aantal voorbeeldberekeningen op basis van data over Nederlandse ziekenhuizen.

In deel B is besproken dat de doelmatigheid van een zorgorganisatie te meten is door de onderliggende technologie in de sector te vergelijken met de daadwerkelijke productie van een zorgorganisatie. In dit rapport laten we de aanname los dat de technologie bekend is en ontwikkelen we manieren om de technologie in een sector te schatten met de beschikbare data. We gebruiken hiervoor moderne ‘frontier’-modellen. Meer specifiek bespreken we de volgende modellen:

- Data Envelopment Analyses (DEA) modellen en
- Stochastic Frontier Analyses (SFA) modellen

Een fundamenteel verschil tussen deze twee in algemene methodologische perspectief en vanuit een reguleringsperspectief is de manier waarop zij de afruil maken tussen een flexibele specificatie van de productietechnologie en de mate van gevoeligheid voor onvolkomenheden en ruis in de productiegegevens.

DEA-modellen zijn beter in het eerste aspect; SFA-modellen beter in het tweede aspect. Als we niets weten over de onderliggende technologie, is het beter om DEA te kiezen, omdat deze meer flexibel is in de vorm van de productietechnologie. Als de scheiding tussen onvolkomenheden in de data en ondoelmatigheid belangrijker is, omdat de kwaliteit van de data beperkt is, dan is het beter om voor SFA modellen te kiezen.

Gegeven een schatting van de onderliggende technologie is de volgende uitdaging het meten van het doelmatigheidspotentieel van een fusie, zoals ontwikkeld in deel B. We demonstreren hoe we zelfs de meest algemene variant kunnen implementeren door gebruik te maken van:

- Lineair programmeren en
- Eenvoudige lijn-zoekmethoden in respectievelijk DEA- en SFA-modellen.

Een bijzondere uitdaging is het meewegen van de geleverde kwaliteit. Het effect van kwaliteit is belangrijk, omdat zeker in de zorg kwaliteit een belangrijk motivatie voor reorganisaties kan zijn en tegelijkertijd ook als kostendrijver genoemd wordt. Een implementatie van kwaliteit moet een midden vinden tussen de beschikbare data en de beschikbare vrijheidsgraden in het schatten van het model. We besteden expliciet aandacht aan kwaliteit en bespreken:

- Verschillende kwaliteitsindicatoren
- Het opnemen van kwaliteit in benchmarks
- Het evalueren van kwaliteitsargumenten in fusiezaken.

Er bestaan verschillende manieren om om te gaan met kwaliteit in fusiezaken. De beste aanpak is afhankelijk van de manier waarop de kwaliteitsindicator is opgesteld en de beschikbaarheid van data. Zoals geschetst bestaat er een grens aan het aantal dimensies waarop een model geschat kan worden bij een beperkte hoeveelheid data.

In dit rapport geven we een voorbeeld analyse op basis van data voor de Nederlandse ziekenhuizen. We berekenen de doelmatigheid van alle mogelijke fusies tussen ziekenhuizen die hun hoofdvestiging binnen 10 kilometer van elkaar hebben. Op deze manier demonstreren we hoe de productie van ziekenhuizen binnen het model valt en hoe de potentiële baten van fusies kunnen worden geschat.

De berekeningen zijn slechts bedoeld als indicatie van de mogelijkheden die deze analyses bieden. Om voor toekomstige fusies dergelijke berekeningen in detail te doen, moet, zoals ook in deel B is besproken, het model worden toegesneden op de specifieke situatie en gezocht worden naar zoveel mogelijk goede data. Daarom trekken we voorlopig alleen globale conclusies voor de sector als geheel.

Fusies van de bestaande ziekenhuizen in Nederland bieden de mogelijkheid tot doelmatigheidswinsten, maar in de meeste gevallen blijken dit leereffecten te zijn. Er bestaan weinig winsten als gevolg van de mogelijkheid om inputs en outputs te verschuiven. Schaaleffecten zijn in veel gevallen zelfs negatief, wat een indicatie is dat bij fusies veelal te grote instellingen ontstaan. Wel zijn er winsten in doelmatigheid te behalen voor individuele ziekenhuizen, wat wil zeggen dat er voor veel ziekenhuizen vergelijkbare ziekenhuizen bestaan in Nederland die zelfstandig doelmatiger functioneren. Veel van de bekeken fusies bieden volgens de gehanteerde modellen weliswaar kleine voordelen, maar deze wegen naar verwachting niet op tegen de toegenomen concentratie in de zorg na een fusie. We kunnen de conclusie trekken dat het de moeite waard is om efficiëntieverhogende methoden te onderzoeken, omdat er mogelijke efficiëntiewinsten bestaan in de ziekenhuissector. Fusies tussen naburige ziekenhuizen lijken op het eerste gezicht gemiddeld voordelig maar niet noodzakelijk om de efficiëntievoordelen te behalen. Het lijkt mogelijk de doelmatigheid met andere, minder anticompetitieve methoden te verhogen.

Gelet op de hierboven geschetste omstandigheden moeten we de resultaten vooral zien als een indicatie van de mogelijkheden van het ontwikkelde model. Ook is de ontwikkelde programmatuur vooral bedoeld om fusies in individuele gevallen te beoordelen; met meer specifieke data kunnen meer precieze schattingen worden gedaan. Met behulp van die berekeningen kan NZa haar argumentatie in zienswijzen beter onderbouwen, en kan de NZa meer inzicht krijgen in de efficiëntieverweren van de fuserende partijen.

Het project "Efficiëntievoordelen van Fusies in de Zorg" is een ambitieuze poging geweest een aantal stappen te zetten in de analyse van fusies. Vanuit de theorie en een modelmatige aanpak, is de uitkomst een kader waarin een kwantitatief oordeel gegeven kan worden van de eventuele positieve effecten van een fusie. Hiermee is een betere afweging mogelijk tussen het wegvallen van een concurrent en het ontstaan van een groter zorgorganisatie, dat ook mogelijkheden heeft tot beter functioneren.

Efficiency Gains from Mergers in the Healthcare Sector

Part A, Economic Theory

Contents

Abstract

The three-part research paper “Efficiency gains from mergers in the healthcare sector” focuses on the positive effects of integrations and moves, from purely economic theories about measurement methods to actual program implementations. Due to the wide scope of the research, this paper is organized into three coherent parts (Part A: Economic theory, Part B: Modelling, Part C: Implementation).

The aim of this paper is to offer the economic theory of efficiency gains as a background to the two subsequent parts of the research paper. The sources of gains are collected for horizontal and vertical mergers, respectively, while alternative forms of partnership are also.

With horizontal mergers, the main effects are typically due to efficiencies of scale and scope, while with vertical mergers the coordination and motivation problems dominate. Besides mergers, several alternative forms of partnership also exist that allow efficiency gains.

1. Introduction

1.1 Background

The liberalization of the Dutch healthcare sector has led to a number of mergers between healthcare and related institutions. The mergers of major concern are between hospitals or between a hospital and an insurance company, while the greatest number of mergers take place between nursing homes and home care providers. Although, at present, the merging parties have only put forward unsubstantiated arguments claiming the benefits of their action, more and better efficiency and quality defences are to be expected in the future.

The a priori assessment of mergers and agreements requires careful consideration since the positive and adverse effects stem from a number of different sources, which must be weighed against each other. An additional complication lies in the fact that the calculations and estimates concern the future. We often have to compare two hypothetical situations: (1) what are the probable effects of the merger, and (2) what would happen in the market without a merger? Quantifying the effects, especially those that impact quality, also presents a challenge.

A great deal of literature exists about horizontal mergers and their evaluation methods, while some questions remain about vertical mergers. The latter are more complicated to evaluate, because the participating parties operate in different markets, and consequently more circumstances and effects need to be taken into account. Generally, vertical mergers offer a wider range of possible efficiency gains, and therefore a higher probability that the gains outweigh the adverse, anticompetitive effects. So far, less attention has been devoted to these beneficial consequences. Their evaluation and quantification are not as well-developed as the estimation of anticompetitive effects.

The three-part research paper “Efficiency gains from mergers in the healthcare sector” focuses on the positive effects of integrations and moves from purely economic theory about measurement methods to actual program implementations. It considers both horizontal and vertical mergers and aims to contribute to the literature of merger evaluation. It offers new findings in the measurement of efficiency gains of vertical mergers and the quantification of the effects of those mergers on quality. This is the first part (Part A) in the three-part series, and describes the economic background to and the potential sources of efficiency gains from mergers.

1.2 Research questions and objectives

The main research question of the project Efficiency gains from mergers in the healthcare sector (and therefore the focus of the research paper series) is 1) *How can we measure the efficiency gains that organizations achieve by merging?* By answering this question, it becomes possible to obtain objective data about positive effects of integrations. After translating these organization-level effects into changes in consumer welfare, we can weigh them against the negative consequences of the merger. In this way, the findings and products of this project, i.e. the developed model, serve as input for the merger assessment process. The series of research paper contributes to the further improvement of the decision-making process by authorities.

The main research question of this particular paper of the series is *What manner of efficiency gains might a merger offer?* In order to answer this question, we turn to the literature on economic theory and look for advantages of extending the organization's boundaries. This paper serves as a background and economic introduction to Parts B and C.

1.3 Scope of the research

The series of research paper intends to offer a comprehensive view on positive effects of mergers at the organization level. It embraces (A) the theoretical background to the potential efficiency sources, which we can find in the economic literature, (B) an overview and the development of mathematical models that are appropriate to the evaluation of the effects, (C) and a demonstration of program applications and examples of data runs.

However, Part A, which deals with the economic background to potential gains, considers strategic advantages of mergers; the developed model (Parts B and C) only discusses cost efficiencies and gains of better coordination between the two parties. The model considers the organizations as individual "decision-making units" and the modelled gains of the merger are efficiencies of scale and scope and positive effects of solving moral hazard problems. Potential changes in the competition environment and interactions between market agents are not considered.

The negative welfare effects of mergers and their measurement have been discussed at length in the literature, and as such also fall beyond the scope of this series of research papers.

1.4 Outline

The series of research papers is structured as follows. Part A gives a summary of economic theory relating to sources of efficiencies in mergers. Part B is devoted to model development, and describes the principal ideas and technical details of calculating the overall gains and the decomposition of those gains according their sources. Part C reviews application possibilities.

This paper is structured as follows. Chapter 2 offers a review of efficiency gains which can be achieved by a horizontal or vertical merger. The title "Boundaries of organizations" refers to the approach adopted: we identify the gains as motivations to extend the boundaries. We can assume that not only the integrated parties benefit from the integration, but that under certain circumstances the gains trickle down to consumers. However, it is possible that the merger is not a prerequisite to obtain the positive effects. Chapter 3 describes partnerships and vertical restraints that may serve as efficient alternatives to integrations. Finally, Chapter 4 summarizes the paper's principal conclusions.

2. Boundaries of organizations

The identifying characteristic of integration is that it changes the boundaries of the organization. A relationship of two entities that was contractual before now becomes an internal connection. The conditions and the environment of the partnership alter, which may bring benefits not only the merging parties but to society as well. We start this research paper with a discussion of the factors that influence the boundaries of an organization, to better understand the motives and opportunities of integration.

Mergers take different forms, depending on the relation between the parties before the action. A merger between two healthcare providers or two insurers that are active in the same market is called a **horizontal merger**. Generally, the parties were competitors before the merger. (In the case of multi-product firms, we may interpret the horizontal merger as integration of firms whose actives overlap.) A merger of two hospitals (in the same geographic area) is an example. The benefits stem on the one hand from the reduction of competition in the market and the increased market power of the merged entity, and on the other hand from possible improvements in resource allocation. The two hospitals, for example, may share some capacities, removing the need to maintain double systems (e.g. IT) or they can specialize in certain activities.

Another form of integration is the **conglomerate merger**, which means a merger of two entities which operate in markets that are not related to each other. This situation is similar to the horizontal merger, except that the agents are not close competitors. However, their market power in both markets may increase as a result of the merger, and they will be able to exploit some advantages of task sharing as well. The integration of different specialists' practices is an example of conglomerate merger.

It is also possible that an organization acquires its supplier, extending towards the upstream markets, or decides to merge with a downstream organization and have direct contact with the consumers. We call these actions **vertical integration**, because they take place between members of a vertical chain. An example is when an insurer acquires a hospital (backward integration) so that it can better control the process of treatment, among other benefits. In the healthcare sector, vertical integration often means an alliance between two institutions that were previously linked by patient flow. For example, a hospital may merge with a nursing home and, among other benefits, better plan the occupancy of beds. The harmonization of processes brings advantages to the patients as well.

The common element in these three merger forms is that the boundaries of the organization have been extended: either vertically in the direction of upstream or downstream markets, or horizontally by acquiring competitors or towards other non-related markets. In the following, we summarize a number of theoretical considerations about motivations behind mergers. We start by describing two general lines of thought about the existence of firms, after which we focus on factors influencing vertical and horizontal mergers, respectively.

Ronald Coase discusses the question of boundaries of firms in his article 'Nature of the Firm' (Coase, 1937). He starts by asking for a good definition of 'firm'. In Coase, the most important characteristic of firms is that they do not use the price mechanism in the resource allocation decisions, but an entrepreneur directs the production process. The subsequent question is why firms exist at all: why it is not sufficient to use the price mechanism in every economic relationship? After considering some possible answers, Coase concludes that using the price mechanism may come with a cost which can be superseded by the entrepreneurial direction.

Coase also looks for factors that determine the size of the firm, i.e. *"why does the entrepreneur not organize one less transaction or one more?"* He finds the following factors that make firms become larger:

- "*The less the costs of organizing and the slower these costs rise with an increase in the transaction organized.*
- *The less likely the entrepreneur is to make mistakes and the smaller the increase in mistakes with an increase in the transactions organized.*
- *The greater the lowering (or the less the rise) in the supply price of factors of production to firms of larger size.*" *(Coase, 1937, pp. 396-397)*

Milgrom and Roberts present a similar line of argument in their book (P. Milgrom, J. Roberts, 1992). They start from the idea that specialization in the economy requires some kind of coordination, so as to harmonize the production process and also the demand and supply. In free markets, coordination is based on the price system. However, to ensure that that system performs efficiently, a number of conditions have to be met, such as the existence of market clearing prices and the affordability of tentatively searching for solutions.

When the price system fails in terms of coordination, one needs to look for other possible structures. Sometimes the government intervenes by supplying the goods or services as public goods, yet sometimes an institution such as a firm can also provide a good solution. Within firms, different mechanisms may be used, such as centralized decision-making or incentive systems.

The existence of economies of scale in production technology is one of the circumstances hindering the proper performance of prices. As the average unit cost decreases if the volume of output increases, a firm's optimal strategy is to operate using its entire capacity and to announce the average cost (or any higher number) as the price. If the demand for such a volume is absent at that price, the market does not exist. However, by adding more information, the volume of demand for example, the business would be profitable for both parties. The information transmitted by the price system was not sufficient to ensure an efficient performance.

Another disadvantage to the price system, compared with centralized decision-making, is that smaller inaccuracies may occur and prompt decisions are not feasible. This problem arises mainly when transmitters are used: the prices. The centralized system is a good solution if immediate and perfect coordination is important, for example because slight mismatches in the process cause great damage.

These are some examples of when the market is not the best coordinating mechanism, and the emergence of companies can be explained. Nevertheless, we have yet to answer the question of how extended a firm should be. Below, we describe some situations that influence the ideal size of a company. Our intention is to list every gain that may be reached by better organizing the relationships between market agents, and we will put further emphasis on those that are consequences of mergers.

Parts B and C also discuss the advantages of mergers. We will focus on the quantification possibilities and mathematical solutions to the measurement of gains. This chapter highlights the background to and context of merger efficiencies in economic theory. Its focus is on identifying the potential gains and on explicating the motivations and sources from which they stem.

2.1 Horizontal boundaries

The question of the horizontal boundaries of an organization is related, on the one hand, to the size of the organization, and on the other hand, to the markets in which it operates. The first aspect answers the question of the scale of the production, while for the second we can ask ourselves, "With the given infrastructure, should the organization be engaged in activities outside its core business?", or "How strongly or weakly related activities should be integrated in the organization's field of activity?" A possible third aspect of the horizontal organization of the organization is the location of the subsidiaries.

The most important goal of horizontal mergers is often to obtain the advantages of production in higher volumes or in different related areas (in this paper, we use the term economies of scale). The merging providers may also obtain gains by changing their input or output mix (we use the term economies of scope). The strategic considerations are no less significant, and are based on the exploitation of the greater market power of the merged entity.

We examine the first two aspects (economies of scale and scope) in detail. The third aspect (strategic considerations) is not related to efficiencies, but to anticompetition measures. The firm achieves higher profits by reducing competition in the market, but does not enhance consumer welfare. This aspect is only briefly discussed in the next section.

2.1.1 Advantages stemming from interactions in the market

Horizontal mergers offer advantages to the parties involved, by reducing the competitive constraints in the market and by increasing the likelihood that market agents are able to cooperate (without a formal agreement).

As a consequence of the reduction of competitive constraints, the merged entity may profit by raising its prices. This is more likely if (1) the merging parties have a large market share, if (2) they are close competitors, if (3) the consumers' options for switching are limited or if (4) the merged entity is able to hinder the expansion of the competitors, i.e. if competitors are unlikely to increase, or to be able to increase, the supply if prices rise.

The other possible effect of horizontal integration is that cooperation between the market agents becomes easier. The companies are likely to adjust their behaviour to the industry's common interest if monitoring and credible deterrents are available in the case of non-cooperation, and if the reaction of outsiders does not jeopardise the results expected from the coordination (European Commission, 2004).

All these developments on the market have an anticompetitive effect and the ultimate consequence for consumer welfare is unambiguously adverse. This result may, however, be countervailed by cost efficiencies and other synergies, as discussed in the next sections.

These anticompetitive effects of horizontal integrations emerge in all cases and are easy to demonstrate. The positive, welfare-enhancing effects of the horizontal mergers are more difficult to prove up front. In the practice of merger assessment processes, the burden of proving the existence of efficiency gains lies with the claimant, and the evidentiary standard is high.

Usually, since mergers of significant size take place in oligopoly markets, studies often use an appropriate Cournot or Bertrand model to estimate the price effects and changes in market shares. Calculations using market simulation models generally use the demand function or at least demand elasticities. Another possibility for studying merger efficiencies is to calculate the Werden-Froeb Index. The Werden-Froeb Index ('WFI') is based on the concept of compensating marginal cost reductions. By calculating the necessary decrease in marginal costs to compensate the anticompetitive effects (either by approximation or exactly) and to restore the pre-merger equilibrium, the WFI measures the scope for efficiency defence (Tuinstra, 2008). The model developed in this series of research papers measures the potential efficiency gains from the merger, independent from the anticompetitive effects. At the same time, however, that the model allows for breaking down the overall gains into specific sources (see Part B, Chapter 3).

2.1.2 Size of the organization

The horizontal growth of an organization can lead to cost reductions in various ways. Many of those ways are related to the higher volume of the production, while others are based on the organization's increased bargaining or market power. In this section we summarize the efficiency- enhancing effects of the higher production scale (following Besanko *et al*., 2006).

We may assume that the cost function is not linear, and the average cost decreases as the volume of production increases. The installation of the facility adds a fixed cost to the production cost, which for larger production scales is divided over more product units. This means that the higher the volume produced, the lower the average cost is. Usually larger scale is optimal when the capital investment is significant. The phenomenon of lowering average cost by volume is called **economies of scale**.

The most common reason of economies of scale is the *indivisibility* of certain inputs or investments. It is obvious that due to high initial capital investments producing in insufficient volumes is not rewarding. For example, a hospital will purchase an expensive diagnostic appliance only if it has many patients who need that kind of examination. The investment returns if the appliance brings revenue, therefore a certain level of usage is necessary. Formally speaking: the more the output to spread the fixed costs over, the less the average cost of the production. A firm exploits all its economies of scale if it grows as long as it sees its returns increase with scale. This definition applies to the short run since it takes the technology as given.

In the long run, however, it is possible that a firm switches to a technology that is more suitable for high volume production. Figure 1 shows that the average costs using technology 2 are lower above quantity y; however, below that output level it is more profitable to use technology 1 because it has lower average costs (AC1 is below AC2). When switching to technology 2, further growth is advantageous, since the cost per output declines. Expansion allows exploiting additional scale economies above Y.

At the same time, the firm cannot choose the output level arbitrarily, but has to adjust to the market conditions. A sufficient volume of the market and demand is necessary to permit the exploitation of economies of scale and to allow the switch between technologies to be profitable.

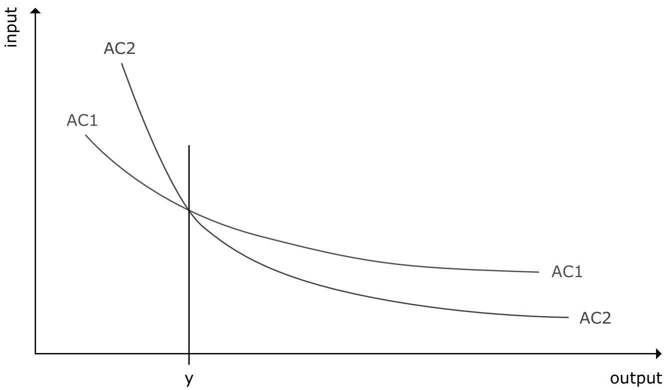

Figure 1: Average costs using different technologies

The cube-square rule, for example, applies to many production technologies. The amount produced increases cubically, while the costs increase only quadratically. This rule is typical for transportation or storage of products. If the amount produced increases (the demand allows the use of pipelines or tanks with a higher volume), the cost rises less than proportionally, i.e. we can exploit the economies of scale.

However, although economies of scale are linked to the indivisibility of capital costs first of all, the *specialization of employees* may also reduce the average costs. A specially educated workforce is able to perform the production in a more professional and more efficient way. However, the education costs of the specialized workforce are higher, and do not return below a certain volume. The previous condition holds true here as well, namely a certain level of demand is required to make the specialization profitable.

Besanko *et al*. gives the example of surgery. They describe the cost and efficiency differences between general surgery and specialists. The education of a specialist takes more time and involves higher costs, but later his/her patients have fewer complications, may recover faster and need shorter stays in the hospital. The higher initial costs of the specialists' education are compensated by a lower cost per patient. However, if only a few people need that operation, the cost is not returned and it is more profitable to employ general surgeons.

A similar concept to scale efficiency is learning efficiency, because it also explains the positive relationship between volume and efficiency. In contrast to scale efficiencies, where the increased efficiency is based on technical principals, here the benefits stem from people participating in the production process. The more is produced, the more experience accumulates, and with time the most efficient practice is reached. This is another reason why firms in the market are able to produce the same product or offer the same service at lower costs. A merger presents opportunities for higher-level production and, in the presence of learning economies, for cost reduction.

The inventory of an organization can also be a source of economies of scale. The role of the inventory is to level off the fluctuations in demand, i.e. the supplier holds extra capacity in order to be able to meet possible unpredictably high demands. The cost of the extra inventory is proportional to the ratio of the normal necessities to spare units. At the same time, a larger organization experiences relatively lower fluctuations in demand since the same amount of necessary extra inventory is a lower percentage of the whole capacity. Consequently, the additional costs are also relatively reduced.

Besanko *et al*. gives the example of blood substitutes here. The hospital cannot plan the need entirely because it is in part contingent on external random events such as road accidents. The hospital has to store blood substitutes to meet the predictable consumption (e.g. blood used during operations), and the possible immediate requisites. Suppose the spare inventory is 5% for an average hospital. However, the probability of external random events does not increase proportionally to the size of the hospital. Moreover, a larger institution has more possibilities for organizing the logistics of the supply and the allocation of resources; consequently, it can manage with less spare capacity. An institution of twice the size may have the same units of extra inventory, which in percentages and in terms of costs means less (e.g. 2.5%).

Besides the production, economies of scale may emerge in other fields as well, such as purchasing, advertising and research and development. These are discussed in the next section, because their characteristic is the simultaneous effect of economies of scale and scope.

2.1.3 Field of activity

What are the motivations of an organization to enter into new activities? Moreover, why do not owners set up another, independent firm to operate in those markets? The answer leads us to the economies of scale and to the potential synergy effect among the input (or output) mixes of the merging parties.

The first reason is very similar to that used for the size of the organization. The indivisibility of investments may require a certain level of production. Suppose a new appliance whose investment returns only if it is used at least 80% of the time, and its utilization is only 50% at the present level of production. One solution, mentioned above in connection with the economies of scale, is to raise the level of production by merging with a competitor (one firm will supply the consumers of both previous entities). The other solution is to produce different products using the same appliance so that it works 80% of the time. The consequence is the same, namely the fixed costs of the infrastructure are shared across more products, and the average cost decreases.

The direction of the horizontal extension remains a question: how can a supplier decide what new markets to enter? Sometimes we see that organizations engage in truly new activities, which previously fell beyond the scope of their activities. These are conglomerate mergers, and they may stem from one of two reasons: (1) the intention to exploit the scale efficiencies in the manner described above, or (2) expectations on joint consumption of goods or services. Conglomerate mergers go beyond the scope of this research paper, and are not discussed it greater detail.

In other cases, the merging organizations have the same (or at least overlapping) activities. The reason for the extension may be related to the features of the production function. It is possible that they can combine their input or output in a favourable manner, creating a more efficient input mix or a more advantageous common production plan. This phenomenon is the **efficiency of scope** – namely it is more profitable to supply a range of products or services together than it is to supply them separately, or, using a more technical phrasing, the sum of the average cost of the separate production is greater than the average cost of the joint production.

Besides production, some other aspects (purchasing, advertising, research and development) also exist where the supplier may exploit economies of scale or scope. In the case of *purchasing*, the supplier generally gives discounts to larger buyers. Besanko et al. lists three reasons: (1) it costs less to sell to one buyer than to several smaller buyers – the fixed costs, such as writing the contract or delivery, arise only once; (2) an organization that buys in high volumes is generally more price sensitive, and therefore presents a higher risk of switching suppliers if prices are unfavourable; (3) the supplier depends greatly on bulk buyers, and so intends to stabilize the business relationship by offering discounts.

In the healthcare sector, we can observe the purchasing scale and scope efficiency. Health insurance companies generally contract for a bundle of treatments with healthcare providers. They can successfully bargain for discounts and therefore they obtain lower prices than the price listed for individuals.

In the case of *advertising*, it is possible to exploit scale efficiencies by spreading the fixed costs over more advertisement units. The fixed costs may include design, placing (in case of advertising multiple products at the same time) or the time and effort of the agent (in the case of marketing campaigns involving personal visits). These benefits are similar to those of the production aspect, but in the case of advertising, we find additional synergy effects. The use of a brand name recalls previous advertisements for other products, for example. It is also possible that the good quality of the advertised product raises the credit of other products of that brand.

In the Netherlands, only a limited segment of the healthcare services, medicines or medical devices may be advertised to non-professionals. Public advertising is allowed with non-reimbursed medicines or treatments. The efficiencies stemming from advertising are relevant only in that segment.

Following theoretical reasoning, economies of both scale and scope are present in *research and development*. The economies of scale are due to the generally costly infrastructure and staff of scientific research. Here, the indivisibility rule applies. Efficiencies of scope may arise if research in one field has positive effects on other research project, because the basic results can be used in both fields or simply because of inspiration. One institution can perform the investigations more efficiently than two separate research programmes.

Unlike the theory, empirical evidence does not show a relationship between the organization's size and its effectiveness in R&D. An argument for the success of small organizations may be that they have better possibilities for motivating their researchers and have different incentives for innovating.

2.1.4 Organizing a network – location

In order to sketch the entire range of economies that influence the efficiency of a company, we describe in brief what other considerations may be taken into account, for example regarding the demand side or the organization's locations.

A company can extend horizontally by entering *new geographical markets*. The difference between expansion toward different product markets (discussed above) and growth into new geographical markets is that in the latter case the field of activity remains the same, but the company will probably have more subsidiaries, and will have to organize them into a network. It is also important that the volume of marketed products will likely rise and more consumers will know and/or use the same brand.

The concept of *network economies* refers to the advantage that an organization gains from an increasing number of users of its good or service. This concept is similar to economies of scale, but focuses on an externality, i.e. a factor independent from the firm's strategy. A higher production volume favours the supplier, but is not due to the lower average costs (as it would be in the case of economies of scale). The users' utility depends on the total number of clients of the firm (e.g. in telecommunications: calls inside the network are usually cheaper than outside, and consequently, if the size of the network grows, the consumers gain). Therefore, a firm with more consumers can demand higher prices for its product or service. This phenomenon is better known as network externality, which is usually analyzed from the users' point of view.

Network economies are characteristics of special industries, often called network industries, and are not relevant in healthcare.

If we think about the number of an organization's locations growing, we can analyze their *relative position in space*. If the outlets or sites are established close to each other, the advantage of faster and cheaper logistics can be exploited, as can the possibility of infrastructure or workforce sharing (or their easy reallocation).

2.1.5 Disadvantages of a larger organization

Besides the positive effects, every merger – and in general terms the growth of an organization – has disadvantages stemming from organizational reasons. The structure becomes more complex and the work of more units has to be harmonized.

We use the term *influence* costs to refer to the effort made by division or department leaders in order to influence the decisions of the central office. Besides the direct cost of lobbying, such as wasted time, we have to take into account the indirect cost as well, such as costs of the bad decisions. These organizational problems increase as the organization grows in size and complexity. Since there are more units competing for the organization's resources, there is more motivation to try to influence the decision-makers.

Experience shows that larger firms generally pay *higher wages*. Reasons may be that they are more likely to be unionized or that they look for workers in a larger area and therefore have to compensate the employees for longer travelling distances. Another possible explication is the greater reliance on efficiency wages. Because of the company's larger size, its monitoring costs increase (it is harder to monitor the employees' work), and management may decide to pay rather higher wages (higher incentive rent) and fire employees having been found shirking than spend on monitoring. This is a negative effect on the supplier's efficiency caused by growth.

The final general disadvantage of the expansion of an organization that we mention here is that the risk of "*conflicting out*" increases. The supplier's clients may require that it not have contracts with their competitors (in order to minimize the risk of leaking information). As an organization supplies services in more and more fields, it is inevitable that it comes into contact with competitors. The above claim of its clients limits the supplier's possibilities.

A horizontally merged company integrates different activity fields, and it is usual that they share some input factors, in many cases being the organization's most important, specialized input factors. They may be computers or special tools for example, but also a highly qualified expert or a successful manager. The many different tasks require a dispersion of attention or capacity, which may lead to the input being overburdened. The consequence is that performance and efficiency decline.

The economies of scale and scope are the most important sources of efficiency gains with horizontal mergers. The company's growth may be beneficial because of, among other reasons, indivisible input factors, the possibility of specialization and inventory advantages. The extension of the organization's boundaries into other related markets makes it possible to avoid duplication of fixed costs of purchasing, in advertising for example. However, the growth of an organization has disadvantages as well, such as motivational problems within the company and influence costs.

2.2 Vertical Boundaries

When considering the production of a certain good, we can list a number of actions that have to be executed, and a list of input factors that have to be used during the production process. By answering the question of how many of these activities should be performed within the organization, and how many of the input factors should be manufactured by the organization, we can define the reasonable vertical boundaries of the organization. Furthermore, we can also identify circumstances that promote the extension of organizations, e.g. by merger. In the context of healthcare, it is sometimes more expressive to think of the steps of a therapy a patient needs. After a serious illness, one may need rehabilitation or temporary nursing services, for example. In this case, the two healthcare providers are the subsequent steps from the patient's point of view. A patient flow exists between the two institutions, but they are not a true supplier/client pair.

A fundamental difference between vertical and horizontal mergers is that in the vertical integration the affected parties had a business relationship (or referred patients to each other) before merging, because they were usually members of the same vertical chain. Besides the economies of scale and scope (which, as we will see, are less significant in the case of vertical mergers) further aspects may also bring efficiency gains to the merging parties.

Various theories have been put forward that aim to explain the sources of merger gains and therefore the motives of firms to integrate. The neoclassical approach focuses on the firm's production or cost function, and attempts to draw inferences from their characteristics. In the following section, we contemplate the problem of the firm's extension in a less theoretical context, from a managerial point of view. We then summarize the transaction cost theory, which examines the relationship between a principal and an agent and focuses on the contract (problems stemming from its costly writing, execution and likely opportunistic behaviour).

The theory of moral hazard problems, discussed in the subsequent section, highlights the motivational questions of the principal/agent relationship (the possibility of incentive contracts between an upstream and a downstream firm that may merge).

Vertical mergers offer advantages that stem from the interaction between the upstream and downstream organizations, for example if the merger makes it possible to avoid the externalities between the parties. Unlike with horizontal mergers, these improvements may benefit the patients as well. Solving the problem of double marginalization not only increases the provider's profits, but also enhances consumer welfare.

In practice, the potential gains from vertical mergers are more evident than the welfare-enhancing effects of horizontal integrations. Furthermore, "*non-horizontal mergers are generally less likely to significantly impede effective competition than horizontal mergers*" (European Commission, 2007), because there is no direct competition between the merging parties. Due to these two facts, the evidentiary standard for efficiency gains from vertical mergers is effectively lower than in the case of horizontal mergers. Vertical mergers are less likely to raise competition policy concerns.

2.2.1 Advantages stemming from interactions in the market

Vertical mergers offer strategic advantages for the parties involved. Integrations may make it possible to eliminate distortions, internalize the external effects emerging through the vertical chain, or simply use the increased market power for strategic reasons. Unlike with horizontal mergers, the strategic behaviour of the firms may also bring additional gains for consumers. The advantages described below always benefit the merging organizations, while in certain circumstances the patients may gain as well.

These considerations describe situations where the agents already possess a degree of market power on one or more levels of the supplier chain before the integration. These assumptions match the present situation in the Dutch healthcare markets.

Double marginalization

An often-cited problem of vertical chains is the double marginalization, which occurs when both parties (supplier and retailer) have market power. By pricing, both firms add their mark-up, which results in excessively high prices (higher than is in the supplier's interest, and certainly higher than any other solution would be). The supplier can influence the behaviour of the retailer and avoid this distortion by using special contracts (vertical restraints) or by simply acquiring it (merger).

A vertical merger can solve this double marginalization problem, because it allows the firms to internalize the externality that they impose on each other (Motta, 2004). A vertically integrated firm can lower its mark-up, which results in a decrease in the final price. As a result, both the consumer and the producer benefit, because the firm will sell more goods at a lower price.

Variable factor proportions
Suppose two upstream markets exist. One is monopolized, while the other is competitive. Their products are both input factors for a downstream market, and to a certain extent they can substitute for each other. The monopolist prices above the marginal costs, unlike the firms in the competitive market. Consequently, the downstream firms will use relatively more of the product of the competitive market. The distortion of the market will cease if the monopolist merges with the downstream firms and they transfer the product at marginal cost within the firm. The final effect for the consumer, however, is ambiguous.

Foreclosure[2]
A vertical relation can disrupt effective competition if it gives rise to foreclosure. Foreclosure means the behaviour of a dominant firm (2 vertically integrated firms) that uses its market power to discourage entry or expansion of rivals or encourage their exit.[3] The rivals are damaged and are unable to compete with the dominant integrated firm anymore (Aghion and Bolton, 1987). The integrated firm is able to profitably increase the final price that it charges to consumers. Two forms of foreclosure exist: input and customer foreclosure. If the merger is likely to raise the costs of downstream rivals by restricting their access to an important input factor, this is called input foreclosure. Customer foreclosure arises if the merger is likely to foreclose upstream rivals by restricting their access to a sufficient customer base.[4] Some economists point out that these vertical contracts damage competitors, while others are more sceptical about the possibility that exclusive vertical contracts lead to foreclosure (Motta, 2004). The various views on foreclosure are discussed below.

---

[2] Everaers, 2008, section 2.3.1

[3] European Commission (2007). G*uidelines on the assessment of non-horizontal mergers under the Council Regulation on the control of concentrations between undertakings*, Note 29.

[4] European Commission (2007). G*uidelines on the assessment of non-horizontal mergers under the Council Regulation on the control of concentrations between undertakings*, Note 30.

Input foreclosure occurs in a market with one hospital (upstream firm) and two insurers (downstream firms). If one of the insurers merges vertically with the hospital, the possibility that the other insurer will be foreclosed from the market exists. The hospital sells its treatments only to insurer 1, and insurer 2 cannot buy treatments from the hospital anymore. Consumers insured with insurer 2 can no longer go to that hospital. If the insured consumers wish to go to the hospital, they must buy insurance from insurer 1. The vertical merger between the hospital and insurer 1 may foreclose insurer 2 from the market.

Customer foreclosure arises if the market consists of two hospitals and one insurer (see Figure 2). Hospital 1 integrates vertically with the insurer, so that the insurer only offers consumers treatments from hospital 1. If consumers wish to go to hospital 2, they are forced to pay the costs themselves, because the insurer will not reimburse the costs. With this vertical merger, hospital 1 forecloses hospital 2 from the market.

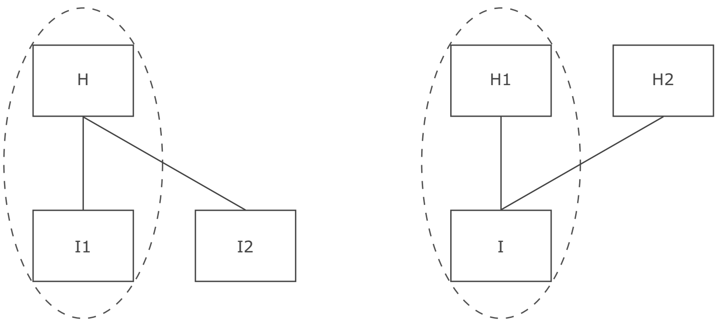

Figure 2: A graphical representation of input foreclosure (left) and customer foreclosure (right)

Input foreclosure only occurs if it is profitable for the upstream firm (hospital). The upstream firm is unwilling to accept the contract without financial or other compensation. The upstream firm sells fewer treatments due to the merger (certainly in the short run), and so the firm wants to be compensated for this loss.

The Chicago School argues[5] that, in the customer foreclosure example, foreclosure will not always occur. Customer foreclosure only occurs if it is profitable for the downstream firm. A downstream firm (insurer) is unwilling to accept an exclusive contract from an inefficient upstream firm (hospital 1) if the market offers a more efficient hospital.[6] By accepting the exclusive contract, the downstream firm is obliged to pay the monopoly price for the good. The upstream firm has to make an attractive offer to the downstream firm, otherwise the downstream firm will not accept the exclusive deal. The upstream firm is unwilling to offer a high compensation, because it will not give up its monopoly profits (and therefore create deadweight loss). This argument implies that exclusive contracts only exist because they entail some efficiency. These exclusive contracts are then beneficial to the firm and to the consumers.

The Post-Chicago School economists argue that there are also circumstances under which foreclosure will have anticompetitive effects. The conditions for anticompetitive impacts of vertical restraints/mergers are, according to the Post-Chicago School: (1) the potential for a vertically related firm to force rivals' costs up and/or foreclose rivals' access to a necessary market and (2) the potential for vertical relations to confer market power by facilitating horizontal coordination or collusion.[7] Firstly, anticompetitive effects might occur if the upstream firm (hospital 1) is able to make an offer attractive enough to persuade the downstream firm (insurer) to accept the exclusive contract. Hospital 1 is able to make such an offer if it also operates in another market where it can make a profit. Hospital 1 makes a profit in both markets, therefore it can afford to give up a part of the monopoly profit in one market. Secondly, the insurer receives information about prices and costs from hospital 1. This gives the insurer a degree of bargaining power in the price negotiations with other hospitals. However, hospital 1 also possesses more information about the insurer, and in response hospital 1 may collude with hospital 2 to keep the prices high. This implies that an increase in the market power due to a vertical relation might result in collusive horizontal behaviour on the upstream level.

---

[5] Posner, R.A. (1976). *Antitrust Law* and Bork, R. (1978). *The antitrust paradox.*
[6] Motta (2004). *Competition Policy, theory and practice.* Chapter 6.
[7] Source: Gaynor, M. and Vogt, W.B. (2000). Antitrust and competition in health care markets. In Cuyler, A. and Newhouse, J., editors, *Handbook of Health Economics*, chapter 27.

Strategic delegation

When two firms integrate vertically, one level of the vertical chain disappears. If we assume a situation with only two levels (upstream and downstream organizations) in the vertical chain, the upstream organization will have direct contact with the consumers, who will be able to set the consumer price for the product/service (see Figure 3). This leads to direct competition between upstream organizations, which – assuming price competition – results in insufficient prices (compared with the joint-profit-maximizing prices).

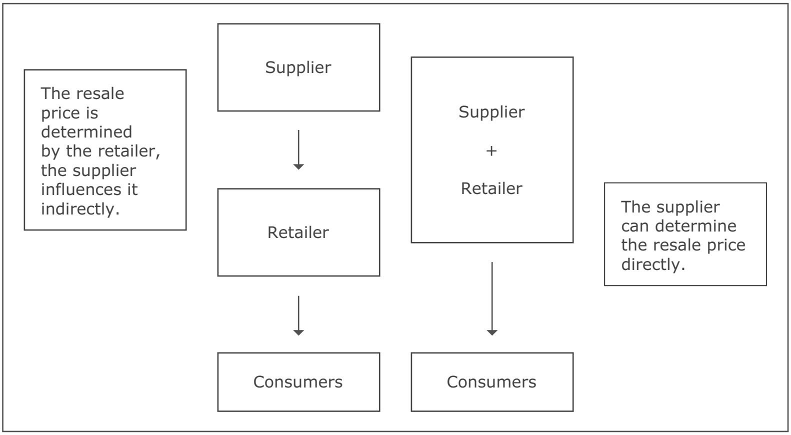

Figure 3: Strategic delegation

However, the upstream firm can delegate the price decision downstream by using vertical restraints (two-part tariffs or exclusive territories) instead of integration. If the two levels of the vertical chain are separated, competition becomes less fierce, since the upstream organizations compete indirectly. The downstream organizations' prices are strategic complements, i.e. if one increases its price, the other will react with higher prices as well. The downstream organizations can base their decisions on the competitor's retail price because they can observe the purchase price. If their wholesale prices increase, they both react with higher retail prices. Due to the two-part tariffs, the upstream organizations are able to extract profits from the downstream organizations, and are better off than they would be with integration.

These results, however, only hold true in the case of price competition, and if the downstream organizations have some market power (e.g. differentiated product or exclusive territories). The low degree of substitutability favours separation, since it increases the downstream organizations’ market power. The uncertainty or the increase of risk aversion of the downstream market agents decreases the benefits of decision delegation. Vertical integration becomes attractive above a certain level of risk aversion.

Backward integration by a monopsonist
Assume that we have a competitive upstream market with an increasing marginal cost function and a monopsonist in the downstream market. The monopsonist may intend to integrate backwards, so that it can capture the rents earned by the supplier of the input and internalize the efficiency loss stemming from underutilization of the input. This second action lowers the consumer prices as well, and therefore enhances welfare.

Price discrimination
Mergers may make price discrimination possible, which again serves the firms’ interests. Imagine a monopolist that supplies its product in two markets with different degrees of price elasticity. If trade between the two markets is possible, the monopolist cannot use different prices in the two markets, and therefore does not maximize its profit (compared with the price discrimination case). A merger with the retailers in the market with higher price elasticity makes it possible to raise the prices in the other market, which may yield additional profits for the monopolist. In certain circumstances, price discrimination is beneficial for the consumers as well.

2.2.2 Neoclassical approach
The neoclassical approach infers the behaviour of market agents by assuming that they maximize their utility or profits, and that they act independently. The decisions are derived from the utility and cost functions and other limitations, such as capacity constraints or relatively low demand. As a result of a merger, the supplier’s options and the conditions of production change. In this section, we identify the characteristics of the cost or production function that make a merger advantageous and, consequently, influence the optimal size of the firm.

With vertical mergers, the scale and learning efficiencies generally arise in the ancillary services. The scale of the main product or service does not change since the merged parties are interested in different markets. However, they can exploit the advantages of an increase in size in services such as building sanitation, they may become more appealing in the labour market, or they may reduce their administrative costs by integrating the systems of the two firms.

Scope efficiency, as defined in section 2.1.2., is also less relevant with vertical mergers. However, we may give another interpretation, and use the term "harmony effect" (as we do in Part B, section 3.3.3). In a vertical chain, we can measure the harmony between the output mix of the upstream firm and the input mix of the downstream firm. The harmony index indicates the suitability of the match of the subsequent links of the chain (or the merged parties).

2.2.3 Organizational and management approach

Another approach to examining the effects of mergers is to consider the processes within the organization. For example, if perfect synchronization of the workflow is indispensable, or information comes to light during production that the firm does not want to share with independent market agents, a merger may be a solution. However, the agency problems can be solved other ways, for example by different types of vertical restraints. We discuss these possibilities in Chapter 3. Here, we focus on the gains that stem from a better-harmonized relationship or from managerial advantages.

Considering the whole process of production, we turn our attention to the link between the sub-technologies. A perfect workflow is highly important if even the slightest mismatches (e.g. in time or in design) result in the loss of a significant proportion of the economic value. In the case of conditions that are hard to define unambiguously in contracts, firms are best off bringing the activity in-house. This solution allows for tighter supervision of execution and therefore more certainty of a precise fit.

In a healthcare institution, the timing of the processes is important. Before an operation, for example, the patient may need certain treatments, and after the surgical intervention, he or she may be treated in the intensive care unit. The hospital must organize the process well and on time. If separate providers were responsible for supplying the various elements, the harmonization could be more difficult.

In strategy literature, the term "strategic fit" refers to the positive interference of strategy elements. We can take the example of an insurer/hospital merger in the Dutch healthcare context. Hospitals are traditionally less competition-oriented organizations than insurers. However, the present reforms in the Netherlands target the introduction of more competition between hospitals and bargaining over product prices between insurers and hospitals. An insurer that takes over a hospital may induce a more profit-oriented approach of company governance and improve the hospital's performance in a competitive environment.

The concept of a strategic fit is very similar to scope efficiency, since it describes a synergy effect, although it is independent from the firm's size. We may obtain efficiency gains by properly coordinating strategy elements (or, as in the previous example, by changing the approach), and in theory the merger does not affect this aspect.

In the situation of special input factors, the product is not sold in the market, it does not have a price, but it is the subject of negotiations between the supplier and the buyer. The evolved price reflects their relative bargaining power, and it is likely less preferable for the principal than the competitive market price (around marginal cost) would be. We discuss the question of transaction costs and hold-up problem in section 2.2.4. Here, we describe a situation where high transaction costs and uncertainty regarding to future outcomes do not factor into the considerations.

A possible strategy for exploiting the beneficial effects of competition on prices, even in case of specific input factors, is to use competitive biddings. However, this option assumes the revelation of the special characteristics of the input for a number of suppliers, which may be undesirable in the case of the firm's distinguishing features. Avoiding the disadvantage of bilateral bargaining, we face the problem of company secrets. Merger may be a solution both for unfavourable prices and for company secrets.

However, we face another problem in the case of integration. In the market, competition presents a strong incentive to become more and more efficient or innovative. In contrast, if the activity is brought in-house the firm has some commitment to continue it, even if other firms supply the product in the market. The department needs less effectiveness or innovation to survive because of historical reasons.

The disadvantages of increasing size, such as influence costs, higher wages and conflicting out (described in section 2.1.4), apply to vertical mergers as well.

2.2.4 Moral hazard problem

The moral hazard problem refers to the consequences of information asymmetry in the relationship between the principal and the agent. The main assumption is that the principal cannot directly observe the actions of the agent, and the results of the agent's work are also contingent upon external factors. In this way, the principal cannot make a distinction between the agent's efforts and the external effects. The agent may exploit the information asymmetry and act contrary to the principal's interests. In order to reduce these intentions, the principal designs incentive contracts, which stimulate the agent to behave in a favourable way.

The objective of *incentive contracts* is to motivate the agent to better efforts. This can be achieved by offering a compensation that is partially contingent upon the outcome. However, the outcome is influenced by external effects as well, and therefore the agent has to bear a risk. The principal can stimulate the agent by making him the residual claimant of profits. At the same time, this means a higher risk for the agent, who requires compensation for that risk.

In the case of a merger, the relationship between the principal and the agent changes, namely it becomes an employer/employee relationship. The motivational power is generally different in the two structures; however they may approximate each other as a result of certain conditions in the contract. In short, the moral hazard problem looks at the trade-off between providing agents with stable income and with effort incentives.

We can identify a number of circumstances under which a merger or disintegration is more advantageous, using a theoretical model. In such a model, we set off the risk premiums required by agents as compensation for bearing the risk against the gains we expect from the increased effort of the agent. The gain that a merger can bring is the difference between the as yet unnecessary risk premium and the consequences of the lower level of the agent's efforts.

Lafontaine and Slade (Lafontaine-Slade, 2007) reviewed articles on practical data and found evidence that under some circumstances a merger is more likely (hence, in those situations they likely offer gains). It is important that the integration and the incentive contracts are not substitutes for the effect that they are efficient under different circumstances. In some cases an incentive contract is more advantageous, in other cases integration is.

Lafontaine and Slade found evidence that mergers are more frequent if the agent's efforts are less important for reaching efficiency than those of the principal. Conversely, the agent would need high incentives for putting forth sufficient efforts, which can be better provided by simple contracts.

In the case of a manufacturer/retailer relationship, the greater the size of the outlet, the more probable the merger is. Theory can explain both this and its opposite, yet evidence shows that an increasing outlet size is accompanied by the greater likelihood of a merger. That likelihood also increases if the costs of outcome monitoring increase, i.e. if measuring the agent's performance (for example assessing sales data) becomes more costly. In contrast, a merger is less likely if the costs of the direct behaviour monitoring rise (measuring sales data becomes relatively cheaper and the principal will rely more on that data and therefore not merge).

The *multi-tasking* problem refers to the numerous tasks of the agent over which he has to divide his capacities. For example, petrol retail involves several "secondary" services. At a petrol station, we find shops, washing and repair services and other goods and services. From an organizational point of view, the problem is that the efforts of the manager and staff are shared over the activities. Assuming that they are employees of the oil company, an appropriate incentive system is needed to make the manager (and the staff) act in accordance with the firm's interests. However, the first step in designing the incentive system is to define the best effort allocation. This requires that the benefits of the various activities can be measured correctly and the incentives can be allocated accordingly. If this is impossible, i.e. if the information between the central office and the gas station's manager is asymmetric, efficient performance is not guaranteed. However, if the manager is the owner of the gas station, he could achieve a greater level of efficiency.

The last situation that we mention is the *free-riding problem* between retailers. We expect this problem if multiple retailers can sell the product, and the efforts put forth to increase the demand is a common good among them. This means that no one can appropriate the entire benefit of their own efforts, and therefore no one will invest sufficiently in that activity. From the suppliers' point of view, this is an undesirable outcome; hence, the supplier would be interested in increasing the demand.

The free-riding problem can be solved again by various restraining contracts or by mergers. The possible equivalent alternatives of the mergers are discussed in Chapter 3.

2.2.5 Transaction costs theory

Transaction costs are the costs of designing, writing and enforcing the contracts between two companies. They play a role in merger decisions because under certain circumstances one or more of these elements may rise so high that the merger (where no contract is needed) is more advantageous. Therefore, the benefits that can be gained from the merger lie in the absence of the transaction costs (although other similar costs may emerge within of the firm as well).

Uncertainty in the market and the inherent complexity of contracts are two circumstances that make it more difficult to write contracts, hence they significantly raise the transaction costs. Lafontaine and Slade (Lafontaine-Slade, 2007) reviewed articles that aimed to demonstrate the effects of these factors on merger decisions. They found that both have a significant positive effect: the higher the uncertainty or complexity, the more likely the merger is.

Transaction costs in combination with specificity
If uncertainty or complexity (or both) emerges in combination with the necessity for a specific input, we likely face the hold-up problem. The specific input means that only the given firm uses it, and consequently it does not have a market and generally its manufacturer has to make an investment (e.g. new appliances or training for the employees). This construction results in an interdependence between the two firms. The lack of market necessitates a bilateral agreement (contract) between the manufacturer and the client. If contracting problems also arise (e.g. due to uncertainty or complexity), the parties have possibilities for opportunistic behaviour[8].

If the two firms both have to make supplementary investments, the parties are equally dependent on each other, and the likelihood of a hold-up problem is reduced. However, if the investments are complementary, so that only one firm has to make the investment, a hold-up problem may occur (Everaers, 2008). The expression “hold-up” refers to the possibility that one party abuses the other’s dependent situation, and forces a renegotiation of the contract for its own benefit.

The production of the input may require different specificities. In some cases the manufacturer has to invest in specific physical or human capital, or has to purchase an asset for the sole purpose of being able to supply the client. We speak of site specificity if the inventory or transaction becomes simpler due to the investment. The final type of specificity is temporal specificity, which is the characteristic of assets that have to be used within a certain time or within a given schedule. All these specificities, combined with contracting problems, lead to hold-up problems, which may be a reason for merger.

Sometimes the input is specialized but only in certain features, while its other characteristics allow for using widely-known technologies. In this case, a supplier may exploit the learning efficiencies of using the basic technology, and produce more efficiently than in-house production would be. The individualized output factors of several buyers differ only partially; therefore, the hold-up problem occurs to a minor extent.

[8] More formally: The specific investment means sunken costs for the investing party, which may incite the other party to abuse the situation in the following way. After the investment the conditions surrounding the production decision are changed: before, the supplier would refuse the contract if the price did not cover its total costs; after the investment, the supplier will supply if the price covers its variable costs. Therefore, the supplier accepts contracts with lower prices after the specific investment.

The example of designing a specific building (e.g. a hospital) makes the previous paragraph more understandable. The given hospital will be a specific one, so it has its own requirements as to the number of wards and operating theatres, it has a site where the building has to fit, etc. All this means that it requires a unique plan, although the hospital does not necessarily have to design the building itself. Several firms exist that specialize in building design, which, although they do not have standard solutions, have a great deal of experience in the field. Unlike hospitals, their core competence is designing. Due to this fact, the specialized firms can solve the assignment more efficiently.

Another consequence of not strictly special input is that scale effects may occur as well. If a product is used by a limited number of firms, it is possible that their combined demand is met by a single supplier. In this way, the firms that opt for market purchase will be dependent on that particular supplier, although they can exploit the advantages offered by economies of scale. The firms that choose in-house production, however, will be independent in terms of production, but at the expense of scale diseconomies. This represents a trade-off between scale efficiencies and agency problems.

In this chapter, we reviewed the advantages and disadvantages of vertical integration, and we organized them by different approaches. We saw that gains can be achieved as a result of the characteristics of the cost function (scale, scope and learning efficiencies), better organization and motivation possibilities (e.g. harmonizing the process, resolving the multi-tasking and free-rider problems), or avoidance of transaction costs and the hold-up problem. The strategic use of vertical integration has both negative and positive welfare effects. It is a remedy for double marginalization, but allows for foreclosure and control over more markets. The disadvantages manifest themselves, and are significant, if the merged firm has considerable market power. Large companies may suffer from inefficiencies because of their size (e.g. influence costs) or lack of competitors.

3. Gains without mergers

In the previous chapter, we summarized the possible gains of vertical and horizontal integration. We compared the two extreme forms of relationships between market agents: (1) when they have only occasional market relationships and (2) when one of them acquires the other one. In reality, many intermediate forms exist, which require some kind of engagement of the parties while they remain independent entities. Different long-term contracts, special conditions in the contract (e.g. parties have to announce their intention to terminate the contract six months in advance) or vertical restraints may also be appropriate to achieve the same effects as with mergers. We discuss the vertical restraints in greater detail in the first part of this chapter.

The examples above all influence the legal relationships between the two parties. However, an appropriate business or market environment is also capable of fomenting cooperation or information flows between the parties. Two examples are (1) frequent and repeated negotiations, which allow for rewarding and penalizing and (2) a relatively tight market with a good information flow, where the company's reputation is an important factor.

Alliances, joint ventures and subcontractor networks are also less formal forms of cooperation. The second part of the chapter describes the characteristics, advantages and disadvantages of these co-operations. Among these types of co-operations, we consider the possibilities of horizontal structures. Horizontal agreements generally tend to reduce competition by harmonizing the behaviour of the competitors, and consequently are prohibited by law. Nevertheless, we can find forms that enhance the welfare.

3.1 Vertical restraints

As with vertical integration, vertical restraints can also serve for both better coordination throughout the chain and enhancement of the market position. Following Motta's structure, we can divide the effects into intra-brand and inter-brand competition questions. Intra-brand competition influences the efficiency of the vertical chain and of the redistribution system, but does not consider the effects on competitors (sections 3.1.2. and 3.1.3). Inter-brand competition deals with the consequences on competition and covers topics such as foreclosure, collusion and strategic price delegation (section 3.1.4)

Generally, the inter-brand effects decrease welfare, as they help the manufacturer to soften the competition and come closer to the integrated monopolist solution (higher prices in the market and presence of deadweight loss). Since the gains of mergers are the focus of this paper, we discuss the inter-brand effects only briefly toward the end of this chapter.

The other sections are structured as follows. Firstly, section 3.1.1 lists a number of vertical restraints; we define them and describe the principal effects on efficiency and the relationship between the market agents. In sections 3.1.2 and 3.1.3, we use two examples (double marginalization and the free-riding problem) to demonstrate how we can solve the problem of externalities by means of vertical restraints.

3.1.1 Types of vertical restraints[9]

There are essentially two forms of vertical restraints: price restraints and non-price restraints. Price restraints per se are illegal in Europe, and non-linear price restraints are assessed case by case. Both price and non-price vertical restraints may increase or decrease welfare. Important factors include the context in which the vertical restraint is used and the goal that it is supposed to achieve (Rey and Vergé, 2005). Resale price maintenance is a price restraint, while the other restraints described below are non-price restraints. Many restraints are substitutes and may have the same effect on welfare.

Resale price maintenance (price restraint)

If the final price charged to consumers is not set by the distributor, but by the producer, we talk about resale price maintenance. Producers set the final price and therefore affect the distributor's price decisions. Several options are available to the producer: a maximum retail price, a price floor or a recommended price. This is a vertical relation, because the distributor and the producer have an agreement about the price. Resale price maintenance per se is forbidden, because it is a price restraint. These restraints are not assessed by the rule of reason, but are always forbidden.

Quantity fixing

Quantity fixing has the same effect as resale price maintenance, provided that no uncertainty exists about the demand curve. If a producer specifies the quantity to be bought and resold by the distributor, the vertical restraint is called quantity fixing. This is only possible if the distributor is unable to sell to or buy from other distributors.

---

[9] Everaers, 2008, section 2.2.1

Non-linear pricing (two-part tariff contracts)
In most cases where an agent buys products (for example medicines) from another agent, he has to pay for all the products proportionally. No matter whether he buys 10 or 100 products, the unit price is the same. This is called linear pricing. Non-linear pricing consists of a fixed amount (also called franchise fee) and a variable component that depends on the number of units bought. The effect of non-linear pricing is that it may be more attractive – depending on the price offered - for a buyer to buy more units from a single producer. An important assumption here is that the buyer is able to distinguish the types of suppliers from each other. The buyer now has a vertical relation with the seller, because it is unappealing for the buyer to switch to another seller due to the fixed amount that the buyer has already paid.

Giving a quantity discount to a distributor has the same effect as non-linear pricing, because the larger the quantity the distributor buys, the cheaper the transaction will be.

Tie-in (bundling)
If a distributor buys one or more goods from a producer on top of the products the distributor ordered, this is called tie-in. The producer offers a bundle of products to the distributor for a fixed price. This may have cost advantages or other benefits for the buyer, because the price of buying both goods separately may be higher. It may be more efficient to buy all the products from a single producer, so that search costs decrease. The seller benefits from a larger volume and can make a higher profit. Healthcare providers can offer full-line service to consumers. This means that a healthcare provider bundles the hospital treatment with the necessary home care the consumer needs after the treatment, for example. The consumer can make use of the home care services immediately after the treatment. The consumer does not have to wait, because the bundling ensures that the process is better organized. The consumer has a limited choice due to bundling. Consumers can only choose the care providers that are bundled with the hospital.

Exclusivity clauses
A producer and a distributor may sign an exclusive agreement, which means that the distributor agrees only to do business with the producer. Such exclusive agreements can soften competition and/or exclude rivals. Various forms of exclusive agreements exist. An exclusive territory implies that only a single distributor may sell a certain brand (of the producer’s) in a certain geographical area. An exclusive deal is an agreement in which the distributor agrees to carry only the brand of one certain producer (also called selective distribution).

Existing literature does not share the same opinion about the total welfare effects of exclusivity clauses. The Chicago School argues that a distributor will only sign the contract if the exclusive agreement is beneficial for him. If a more efficient producer is available, the distributor will not accept an exclusive agreement with an inefficient producer. This leads to the conclusion that parties will only sign exclusive agreements which entail efficiency for both parties. Unlike the Chicago School, Rasmusen *et al*. (1991) and Aghion and Bolton (1987) conclude that exclusive contracts may be a deterrent to entry. Bernheim and Whinston (1998) argue that exclusive agreements can be used to foreclose markets, except if the upstream and the downstream firms can achieve perfect coordination by this vertical relation. These foreclosure effects have been discussed in more detail in paragraph 2.1.5.

A vertical exclusive agreement is not always welfare-enhancing. A vertical exclusive agreement is very binding: the upstream firm sells only to one downstream firm and no longer to other downstream firms. Assume a market with two upstream firms (U1 and U2) and two downstream firms (D1 and D2). One upstream firm (U1) is efficient and the other upstream firm (U2) is an inefficient producer. U1 signs an exclusive contract with D1. This implies that D2 can only buy products from the inefficient upstream firm (U2). Even if D2 wants to buy products from U1, that is impossible because of the exclusive contract. This is not efficient and reduces welfare.

Most-favoured-nation clause

A most-favoured-nation (MFN) clause is a vertical agreement in which the seller agrees not to charge the buyer more than the lowest price it charges to any other buyer (Gaynor and Vogt, 2000). These vertical contracts have often appeared in the United States between healthcare providers and HMOs. An MFN clause has anticompetitive effects and may lead to foreclosure.

3.1.2 Solutions to double marginalization

Double marginalization problems occur in a vertical chain if both the upstream and the downstream organization have market power. The upstream organization sets the wholesale price at marginal cost plus a mark-up, exploiting its market power. The downstream organization is also able to price above marginal cost. Due to this double marginalization, consumers face a higher price than the monopoly price. Besides the negative effects on consumer surplus, the vertical chain could obtain more gains by avoiding double marginalization. We demonstrated in section 2.1.5 that a merger could solve this problem. Below, we summarize other possibilities, namely vertical restraints that are also suitable for eliminating double marginalization.

Retail price maintenance (RPM) is suitable for avoiding the disadvantages of double marginalization (as long as the resale prices are visible). In this case, the upstream organization is able to determine the resale price so that it matches its own interests. The wholesale price, which is a result of negotiations between the upstream and downstream organizations, determines the allocation of the profits between the two parties to the contract. If the upstream organization has all the bargaining power, it is able to obtain all the profits and restore the results under integration.

However, if we assume a risk-averse downstream organization or a significant cost fluctuation, we see that the downstream organization may require an additional risk premium. The retail price is given and therefore a rise in the costs reduces the downstream organization's margin. The downstream organization will require compensation (risk premium) for its uncertainty, and consequently the upstream organization cannot count on the same profit level. The risk that the demand changes, in contrast, does not affect the downstream organization since the resale price is a given.

Another option is **quantity fixing**. This uses the same mechanism as RPM, but guarantees a higher quantity in the market and influences prices indirectly. The upstream organization can determine its share in the profits by choosing the wholesale price.

Non-linear pricing can also solve the problem of double marginalization. The upstream organization can set the variable fee equal to its marginal costs, which offers the same decision situation for the downstream organization as in the case of integration. Subsequently, the upstream organization can determine how much profit its leaves for the downstream organization by defining a fixed fee.

The logic of the system lies in the different ways of risk sharing. Using the two-part tariff, the upstream organization makes the downstream organization the residual claimant of all profits generated in the market. The downstream organization cannot pass on the risk of variable demand; its acts precisely as an integrated entity would act. In this case, the result is exactly the same as with a merger.

Non-linear pricing has precisely the opposite effect under uncertainty to RPM. Variations in the cost level do not affect the downstream firm's margin, since it can pass the risk on to the consumers. The demand fluctuations, however, have a great effect on the downstream organization's profits. We can conclude that RPM is more suitable under demand uncertainty, and non-linear pricing is a better choice in the case of significant cost fluctuations.

A large-scale solution would be to handle the preconditions of double marginalization, namely to increase the competition in the downstream market. Competition forces down the margins of the downstream organizations and weakens the externality. At the same time, the opposite is also true: if the downstream organizations' market power increases, the problem of double marginalization becomes more significant (an example is given in the following section).

The double marginalization problem can potentially be solved by vertical restraints, as discussed above. However, double marginalization will only disappear completely in a world without uncertainty, where marginal costs are constant, contracts are complete and complete information is available to all agents. A vertical relation may solve part of the double marginalization problem, despite the existing uncertainties. So in a real world, with a healthcare market that does not fulfil these conditions, it is unrealistic to solve the inefficiencies entirely using a vertical restraint (Everaers, 2008).

3.1.3 Solutions to the free-riding problem

The free-riding problem emerges among the retailers of a product if it is impossible to fully appropriate the benefits of an investment. The retailer that makes an investment has to bear all the costs, yet the benefits are shared among the firms in the downstream market. Therefore, the incentives to invest are lower than in the case of a vertically-integrated chain. This problem can be solved if the upstream firm acquires all its retailers so that it internalizes the externality, as described in section 2.1.3. Below, we describe other solutions for the problem.

Resale price maintenance **(RPM)** or price-floor setting may be a solution. It avoids the problem of undercutting in price competition, and consumers will not have any reason to go to another shop to buy the product. Even so, RPM does not restore the results of the vertically-integrated case. Externality still exists between the retailers because they cannot fully appropriate the benefit of their investment.

The solution is to give them additional incentives, for example by non-linear pricing. The manufacturer has to set the wholesale prices below its marginal cost, to give an extra incentive. The fixed part of the payment will determine the allocation of the profits between the parties. If the upstream firm has sufficient market power, it can fully restore the integrated solution.

Another vertical restraint that might be used in combination with RPM is quantity fixing. The upstream organization forces the optimal effort level by determining a higher sales level. In this construction, the downstream organization does not have any other decision possibilities, and so the upstream organization can use the wholesale price freely to determine a profit share (extract all the profit from the downstream organization).

Another solution to the free-rider problem is to make it more difficult for consumers to visit multiple shops, so that they are forced to buy at a single shop when they look around. The downstream organizations have more chance to appropriate the benefits of their investment and consequently they tend to invest more. This can be reached using **exclusive territories**, for example.

However, creating local monopolies brings another problem to the surface, namely double marginalization. Strong competition is advantageous because it prevents high mark-ups in the resale market, yet it is disadvantageous considering the free-rider problem. In order to solve the two problems simultaneously, a combination of vertical restraints have to be used, for example exclusive territories and non-linear pricing.

Internalization of the externality does not always mean an increase in welfare. It is possible that as a result of the vertical restraint (or of the merger), the investment in the public good increases excessively, and welfare decreases. However, this only occurs in the absence of any competition and if consumers cannot choose another supplier.

3.1.4 Inter-band competition effects

Firms may use vertical restraints for strategic reasons, as also discussed in the section on price-decision delegation. However, not every restraint has these effects. For example, resale price maintenance does not work this way, since the manufacturer continues to determine the resale price. Moreover, the results are sensitive to the nature of competition as well. The consequence is that it is impossible to formulate general rules about efficiency gains, and so we must consider the unique cases.

Besides their strategic aims, organizations may intend to collude in prices. RPM and the installation of common agencies promote these ambitions. RPM works in a way that increases the visibility of prices, thereby making it easier to reveal possible deviations from the collusive price. The common agency, at the same time, internalizes the externalities, and behaves as the organizations would behave in order to maximize their joint profits.

The third aspect is the foreclosure of potential competitors from the market. Exclusive contracts lead to these consequences, especially if a firm with significant market power applies the contract. Various approaches to foreclosure (Chicago School, Post Chicago School) are discussed in section 2.2.1. Foreclosure.

3.2 Other partnerships between companies

So far, we have focused on vertical relations. Parallel contracts in horizontal relationships would be collusive arrangements. However, these horizontal partnerships always reduce competition in the market; therefore, it is not surprising that on the whole the law prohibit them. Common examples of horizontal cooperation are agreements on sales prices, allocation of quotas, market division and coordination of behaviour along some other dimension (Motta, 2004, p. 137).

In this section, we review less strict forms of cooperation that may be used both in horizontal and vertical relations. These forms of horizontal collaboration are beneficial to welfare.

Strategic alliances and joint ventures are forms of cooperation between organizations that try to exploit synergies while preserving as much independence as possible. The contracts contain less liability than vertical restraints, and the formulation is sometimes general. The precise extent of the cooperation is often subject to informal agreements or norms and reciprocity. While vertical restraints precisely define the obligation of the parties and create a formal relationship, alliances are more flexible and partially informal. Besanko et al. formulate the definition of a strategic alliance as follows:

"*In a* **strategic alliance***, two or more firms agree to collaborate on a project or to share information or productive resources. Firms may rely on a contract to spell out specific responsibilities for investing in assets as well as the distribution of earnings, but the contracts may be largely silent about the details of the collaborative effort.*" *(*Besanko *et al., 2006. p. 150)*

Strategic alliances may occur between institutions that operate in different markets. For example, a healthcare insurance company may reach an agreement with an optical shop. The optical shop offers discounts to the insurer's clients, but there is no closer relationship between them. They exploit the opportunities of their overlapping clientele, and by cooperating they are able to attract more consumers.

A **joint venture** differs only in a small compass. The cooperating parties remain perfectly independent because they create a new firm as a jointly-owned organization. The aim is to work together in certain fields, for example in research and development, resale or production, so that they can enjoy the gains of scale efficiencies or coordination of their behaviour. In this aspect, it is similar to a merger, and its consequences can be analysed in a similar fashion (Motta, 2004, p.203).

Alliances intend to integrate the advantages of cooperation and independence, but they cannot prevent the problems that arise. The risk of information leakage, agency and influence costs and the free-rider problem all may emerge in these relationships.

A special form of alliance between companies is the Japanese **keiretsu**, which is a group of loosely-related firms from every field of the economy. Financial links exist between them, since they usually have a common central bank and they hold shares in each other's businesses. Nevertheless, the informal relationships between the leaders (and employees) of the companies also play an important role.

Cooperative standard setting is important in industries where compatibility of network and other elements is important. In network industries, externalities emerge due to the extension of the network. The more people use the same standard, the more value it offers for everybody. Cooperative standard setting helps avoid evolving a fragmented network and makes it possible to obtain the highest possible value in the industry. Another advantage of an agreement about standards is that consumers do not delay their purchase because of uncertainty. (If a person buys a durable good before the standard is established, he or she risks that ending up with the abandoned good, which will have minimal value).

Cooperative standard setting has a disadvantage as well, in that it is not clear that the market will choose the best standard. It is possible that the most powerful or most influential company's technological solution will be accepted. In the absence of cooperative standard setting, we can still assume that the market itself would develop in the direction of a common technical solution because of network externalities. Consumers choose a provider that uses the most widespread standard, which becomes even more attractive due to the large number of users. At the end of the process, only a single standard remains in the market: that which has won the competition of standards. However, if the "war of standards" lasts long, consumers may delay their purchases for so long that the market's development suffers.

For example, in the healthcare sector it is beneficial to set a standard for the format for recording and storing data about the patient's medical conditions. Using standard electronic patient files, healthcare providers nationwide could query the patient's medical background. If different subsystems existed, the patient's medical data would not be accessible to every healthcare institution, which would make successful treatment more difficult.

It is possible that a technological development requires the know-how of a process that is protected by a patent. Two possibilities are available for acquiring the licence to the patent. The firms negotiate and come to a bilateral solution, or a patent pooling organization exists in the market. In both cases, the payment for the patent may be a fixed amount or a royalty fee per unit. The latter is less advantageous because it increases the variable costs for the producers and reduces quantity.

Bilateral negotiations may lead to **cross-licensing**, which means that two organizations exchange their patents so both of them will be able to perform the development. This has special importance if the two technologies are complementary and the simultaneous possession of the two brings extra benefits. The mechanism and the consequences are similar to complementary input, namely that the prices fall in the market. Cross-licensing may reduce competition if the technologies are substitutes and the payment is a royalty fee per unit.

Patents and development occupy a central place in the pharmaceutical industry. Cross-licensing makes it possible to avoid parallel investigations (or doubled costs invested in investigation) and therefore the developments are more efficient and faster. In the absence of cooperation, cross-licensing still allows for investigations without infringing upon the other company's patent.

Patent pools "*hold the patent rights of two or more firms and licenses them to third parties as a package*" (Motta, 2004, p. 206). They have the advantage compared with cross-licensing that they reduce transaction costs because dealings are only required with a single party instead of multiple bilateral negotiations. If the patents are complementary, the packages offer an extra benefit.

In this chapter, we reviewed a number of possibilities for relationships between organizations that may offer efficiency gains without the need for merging. As we have seen, partnerships have positive and negative welfare effects as well. In the case of vertical chains, those positive and negative effects must always be considered. However, the competition-reducing consequences are significant only if the contracting parties have sufficient market power. In the case of horizontal relationships, some agreements (on prices, on quotas, etc) are automatically prohibited because of their welfare-decreasing effects. Some other forms of cooperation, however, may be beneficial to the whole of society.

The merger assessment process deals not only with the net effects of mergers on consumer welfare, but also considers whether the integration is the proportional and necessary solution for achieving the predicted gains. If another form of cooperation (e.g. long-term contract, bilateral agreements) achieves the same positive effects, while having less anticompetitive and welfare-reducing effects, it is preferable over a merger.

The evaluation of the necessity and proportionality of a merger requires a scrutiny of the sector and of the nature of the announced gain. General and universal classification of the efficiencies (on this basis) is impossible. From a strictly theoretical point of view, every efficiency gain is attainable by designing an appropriate (though in some cases very complex) contract. In a merger assessment, however, the question not only concerns the possibility of another solution, but also the probability of its realization. For example, an organization may achieve economies of scale without a merger, simple by extending its production. It may be feasible (and profitable) in a growing market, but less advantageous in a narrowing market. Farrell and Shapiro (Farrell and Shapiro, 2001) analyses in which circumstances economies of scale are merger-specific in horizontal integrations. A similar analysis would be possible for all types of efficiencies.

In Part B of this research paper, we develop a model that not only quantifies the positive effects, but also offers a breakdown. In a concrete case, it is possible to analyse and decide what efficiencies (cluster of efficiencies) are likely to be merger-specific, and to take this into account in the calculations. The advantage of the breakdown is that it makes it possible to treat several different aspects of the gains separately.

4. Conclusions

In this research paper we focused on the positive effects of mergers, such as efficiency gains and improved coordination. The paper gives an overview of gains that a vertical or horizontal merger may offer and in this way serves as a background to the two other papers of the series (Part B and Part C). Besides horizontal and vertical mergers, alternative forms of integration and cooperation are also discussed.

Considering the main efficiency effects, we find significant differences between horizontal and vertical situations. These stem on the one hand from the different natures of the relationship of the merging parties. In the horizontal aspect, the typical relationship is competition. The vertical aspect, however, leads to supplier/buyer relationship, implying a certain degree of coordination. On the other hand, differences between horizontal and vertical integration stem from the fields of operations (and therefore the similarity of input and output factors) of the merged organizations. In the horizontal case, the two organizations operate in the same market, and so use the same input factors and produce the same output factors. At the same time, vertically-connected organizations operate in different markets, and their input and output factors are also different.

In the case of horizontal integration, the scale and scope efficiencies play an important role in the main activities. As their field of operations is the same, horizontally-merged organizations can exploit the advantages of pooled production. The closer cooperation, however, does not bring efficiency gains, but allows for anticompetitive behaviour and results in a reduction in consumer welfare.

Vertical integration, in contrast, affects more levels of the production and supplier chain; therefore, there is room to improve the efficiencies throughout the chain. Here we refer to the gains that stem from improved coordination and internalization of externalities, for example. Scale and scope efficiencies, however, are significant only in secondary activities, such as administration or building maintenance.

Some efficiency gains may be achieved using alternatives to mergers. The range of possibilities extends from a strict contractual relationship (e.g. vertical restraints) to informal cooperation. The consequence is that the integration is not always proportional and necessary to achieve the gains. In the merger assessment, this is also an important aspect.

5. References

Aghion, P. and Bolton, P: Contracts as a Barrier to Entry.
American Economic Review 77, pp. 388–401, 1987

Besanko, D., Dranove, D., Shanley, M. and Schaefer, S:
Economics of Strategy, *John Wiley & Sons*, New York, 2006

Bernheim, D. and Whinston, M: Exclusive Dealing,
Journal of Political Economy 106, pp. 64–103, 1998

Coase, Ronald H: The Nature of the Firm, *Economica 4*, pp 386-405, 1937

European Commission: Guidelines on the assessment of horizontal mergers under the Council Regulation on the control of concentrations between undertakings,(*2004/C 31/03*)

European Commission: Guidelines on the Assessment of Non-horizontal Mergers under the Council Regulation on the Control of Concentrations between Undertakings, 2007

Everaers, Janneke: Vertical Relations in Healthcare,
Master Thesis Economics, Tilburg University, 2008

Farrel, J. and Shapiro, C.: Scale economies and Synergies in Horizontal Merger Analysis, *Antitrust Law Journal*, 68: pp. 685-710, 2001

Gaynor, M. and Vogt, W.B: Antitrust and Competition in Health Care Markets. In Cuyler, A. and Newhouse, J., editors, *Handbook of Health Economics*, Chapter 27, pp. 1405-1487. Elsevier Science, B.V., Amsterdam, 2000

Lafontaine, F. and Slade, M: Vertical Integration and Firm Boundaries:
The Evidence, *Journal of Economic Literature*, Vol. XLV (2007, September) pp. 629-685, 2007

Milgrom, P.–Robert J: Economics, Organization and Management,
Prentice Hall, Upper Saddle River, New Jersey 07458, 1992

Motta, M: Competition Policy, Theory and Practice,
Cambridge University Press, New York, 2004

Rasmusen, E., Mark Ramseyer, J. and Wiley Jr., J.S: Naked Exclusion. *The American Economic Review* 81, pp. 1137–1145, 1991

Rey, P. and Vergé, T: The Economics of Vertical Restraints, in *Handbook of Antritrust Economics*, Paolo Buccirossi Ed., M.I.T. Press, 2005

Tuinstra, J; Goppelsroeder, M; Schinkel, M.P.: Quantifying the scope for efficiency defence in merger control: the Werden-Froeb Index, *Presentation at the Academic Seminar on Efficiency Gains from Merger*, Dutch Healthcare Authority, Utrecht, 28.02.2008.

Efficiency Gains from Mergers in the Healthcare Sector

Part B, Modelling

Contents[1]

1 The sections marked with a * are more technically formulated. However, these sections can be skipped without losing the general ideas.

Abstract

The three-part research paper "Efficiency gains from mergers" focuses on the positive effects of integrations and moves from purely economic theory over measurement methods to actual program implementations. Due to the large field covered, it is organized in three coherent parts (Part A: Economic theory, Part B: Modelling, Part C: Implementation).

This paper develops models of horizontal and vertical integration that allow quantifying the efficiency gains from the mergers. The paper describes different alternatives of measures of the gains, like improvement set, proportional resource reductions, proportional service expansions and directional improvements. The overall gains are decomposed in efficiencies stemming from learning, economies of scope and economies of scale. In this way, it possible to identify which aspect of the integration (rise of production scale, coordination of different processes, etc) offers the most significant efficiency gains.

It may be the case that some of these efficiencies can be achieved by other forms of cooperation than full integration as well. The paper considers the effects of these partial integrations too and as a by-pass the efficiencies of split of firms. The models are completed with the relaxation of several initial conditions in order to make it possible to adjust the calculations to the constraints of the real environment.

1. Introduction

1.1 Background

The liberalization of the Dutch healthcare sector has lead to a number of mergers between healthcare and related institutions. The mergers of major concern are between hospitals or between a hospital and an insurance company, while the greatest number of mergers takes place between nursing homes and home care providers. Although, at present, the merging parties have only put forward unsubstantiated arguments claiming the benefits of their action, more and better efficiency and quality defences are to be expected in the future.

The a priori assessment of mergers and agreements requires careful consideration since the negative and positive effects stem from a number of different sources, and they have to be weighed against each other. An additional complication lies in the fact that the calculations and estimates concern to the future. We often have to compare two hypothetical situations: (1) what are the probable effects of the merger, (2) what would happen in the market without a merger? Quantifying the effects, especially those that impact quality, also presents a challenge.

A great deal of literature exists about horizontal mergers and their evaluation methods, while some questions remain about vertical mergers. This latter is more complicated to evaluate, because the participating parties operate in different markets, and consequently more circumstances and effects need to be taken into account. Generally, vertical mergers offer a wider range of possible efficiency gains, and therefore a higher probability that they outweigh the adverse, anticompetitive effects. So far, less attention has been devoted to these beneficial consequences. Their evaluation and quantification are not as well-developed as the estimation of anticompetitive effects.

The three-part research paper “Efficiency gains from mergers in the healthcare sector” focuses on the positive effects of integrations and moves from purely economic theory over measurement methods to actual program implementations. It considers both horizontal and vertical mergers and aims to contribute to the literature of merger evaluation. It offers new findings in the measurement of efficiency gains of vertical mergers and the quantification of the effects of those mergers on quality. This is the second part (Part B) in the three-part series, which covers the model development.

1.2 Research questions and objectives

The main research question of the project Efficiency gains from mergers in the healthcare sector (and therefore the focus of the research paper series) is 1) *How can we measure the efficiency gains that organizations achieve by merging?* By answering this question, it becomes possible to obtain objective data about positive effects of integrations. After translating these organization-level effects into changes in the consumer welfare, we can weigh them against the negative consequences of the merger. In this way, the result and products of this project, i.e. the developed model, serves as input to merger assessment process. The series of research paper contributes to the further improvement of the decision-making process by authorities.

The main research questions of this particular paper of the series are *"How can we calculate the efficiency gains from mergers?"* and *"How can we decompose these gains according their source?"* The objective is to find techniques that quantify the efficiency gains which can be attained by merging two (or more) entities, and to identify different elements of these gains. The question is considered theoretically in this paper, and mathematical models are developed as answer.

1.3 Scope of research

The series of research paper intends to offer a comprehensive view on positive effects of mergers at the organization level. It embraces (A) the theoretical background of the potential efficiency sources, which we can find in the economic literature, (B) an overview and development of mathematical models that are appropriate for the evaluation of the effects, (C) and the demonstration of program applications and examples of data runs.

However, Part A, which deals with the economic background of potential gains, considers strategic advantages of mergers; the developed model (Part B and C) only discusses cost efficiencies and gains of better coordination between the two parties. The model considers the organizations as individual "decision-making units" and the modelled gains of the merger are efficiencies of scale and scope and positive effects of solving moral hazard problem. Potential changes in the competition environment and interactions between market agents are not considered.

The negative welfare effects of mergers and their measurement have been discussed at length in the literature, and as such also fall beyond the scope of this series of research papers.

1.4 Outline

The series of research papers is structured as follows, Part A gives a summary of economic theory relating to sources of efficiencies in mergers. Part B is devoted to model development, and describes the principal ideas and technical details of calculating the overall gains and the decomposition of these gains according their sources. Part C reviews application possibilities.

This paper is structured as follows. We compare the transformations of inputs into outputs that are possible with two or more independent entities with the possible transformations that are possible in an integrated entity. In chapter 2 we provide a bit of background information and introduce a minimum of notation that will be handy in the rest of the paper. We also derive the overall measure of potential gains from horizontal and vertical integrations. In Chapters 3-5 we then decomposes and adjust these measures to take into account what can be adjusted in a given context and using types of reorganizations. We close with some alternative views on inefficiency in Chapter 6.

The paper can be read at two levels. One is a conceptual level, and another is a more technical level with a view towards actual measurements and implementation of the ideas into software. The conceptual idea can be explained with a minimum of mathematics while the implementation requires somewhat more details and precision and therefore makes it most efficient with more formalism and mathematics. Throughout the paper we try to make the general ideas clear using verbal explanations and simple graphics. These parts should therefore be accessible to most readers. The more technical parts are marked with a * in the section headings and can be skipped without losing the general ideas.

The focus on the conceptual ideas is of course relevant to readers with less time or background to study the details of the modelling. We suggest that it is also very useful to more technically oriented readers. In many applications limited data is available and this may prohibit specific modelling and make estimations impossible. Experience shows however that the conceptual framework in such cases is useful to get a good idea of the forces and effects at work, and to structure the analyses.

2. Overall gains from mergers

In this chapter we develop models of the overall potential gains from horizontal and vertical mergers. We shall then decompose and refine these measures in the subsequent chapters.

2.1 Technology and efficiency

A simple way to think of an organization is as a transformation of multiple inputs (x) into multiple outputs (y). We may call the organization (or production unit or agent) P^1 and can therefore illustrate the activities of P^1 like in Figure 1 below.

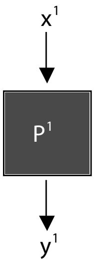

Figure 1: Simple organization

In the case of a hospital, for example, we may think of the inputs as doctors and nurses, while the outputs as number of treatments and the capacity provided as a buffer against uncertainty:

$$x^1 = (\text{doctors, nurses})$$
$$y^1 = (\text{treatments, capacity})$$

To evaluate the performance of P^1 we need some benchmark against which to evaluate its production (x^1, y^1). In general terms, we may think of this benchmark as the *technology*

$$T = \text{set of input - output combinations that are possible}$$

We shall discuss how to estimate T in real situations in Part C of this research paper series. For now it is enough that T describes what we consider to be possible to accomplish in real organizations of the same type as P^1.

For illustrative purposes, it is usually convenient to think of T as input sets L(y) or as output sets P(x). The interpretations of these sets are

L(y) = set of inputs that suffices to produce the outputs in y
P(x) = set of outputs that can be produced from the inputs in x

To evaluate the performance of organization P^1, we shall compare its resource usage and service provision against the technology T. If it is possible to produce more outputs than y^1 using fewer inputs than x^1, we say that P^1 is inefficient. An illustration of this is provided in Figure 2 below. We see that the actual production – as represented here by the blue dots in the input and output space – is using more inputs than what is strictly needed. Indeed, it is possible to move from the blue point x^1 to the black point Ex^1 and still have enough resources to produce the given output. This leads to the so-called *Farrell measure of input efficiency*

E = minimal input / actual input

Small values of E indicate large inefficiency – or improvement potentials. A score of, for example, E=0.7 suggests that it is possible to reduce all inputs with 30% and still produce the present outputs.

Likewise, the inefficiency shows up at the output side since with the used resourced x^1, the blue point y^1 is not on the frontier of the production possibility set. Rather it would be possible to expand the services – e.g. proportionally as in the point Fy^1. This corresponds to so-called *Farrell output (in)efficiency*

F = maximal output / actual output

Large values of F are a sign of large inefficiency – or improvement potentials. A score of for example F=1.2 suggests that it is possible to increase all outputs with 20% without increasing the use of inputs.

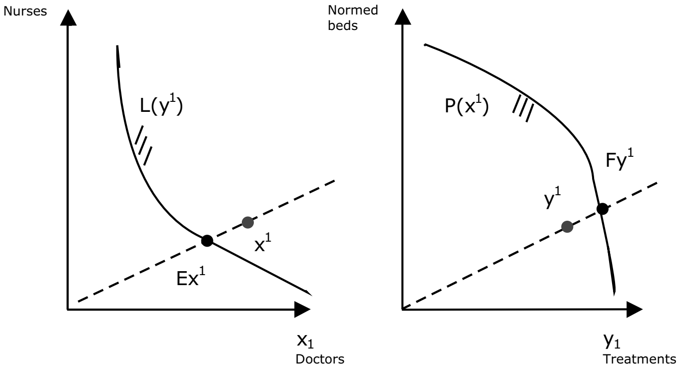

Figure 2: Inefficiency

It is clear that the larger the distance from the frontier of technology, here the isoquants, the more inefficient is P^1.

We shall use the same logic when we evaluate individual entities as when we evaluate merged entities. The larger the distance to the frontier, the more inefficient is the merged unit.

Being inefficient represent a loss. On the other hand, being inefficient also suggest improvement possibilities. This leads to our basic idea.

Basic idea: We capture the synergies from a merger by the increase in improvement potential when we move from independent to joint operations.

Corporate synergy occurs when corporations through their interactions are able to produce more services with given resources, or to produce given services with less resources.

We shall now see how joining the operations, in a horizontal or vertical merger, can create new improvement potentials.

***Formal definitions of technology and efficiency**
In formal terms we can summarize the technology as

$$T = \{(x,y) \in R_0^{p+q} \mid x \text{ can produce } y\}$$

That is: T is all the combinations of p inputs x that are able to produce the q outputs in y.

Likewise we have

$$L(y) = \{x \in R_0^{p} \mid x \text{ can produce } y\}$$

$$P(x) = \{y \in R_0^{q} \mid x \text{ can produce } y\}$$

Simple measures of the efficiency of P^1 is therefore the Farrell input and output efficiencies

$$E = \mathrm{Min}\{E \in R_0 \mid (Ex^1, y^1) \in T\}$$
$$F = \mathrm{Max}\{E \in R_0 \mid (Ex^1, y^1) \in T\}$$

2.2 Horizontal integration

Figure 3 illustrates a classical horizontal merger. Two units (or production entities) P^1 and P^2 individually transforms vectors of inputs (resources), x^1 and x^2, respectively, into vectors of outputs (services), y^1 and y^2. As an example we may think of two hospitals. Observe that we do not assume that they use exactly the same input and output types since we can always allow the value of some of the dimensions of the x and y vectors to be 0.

If the two units integrate but continue to operate as two independent entities, they would transform the vector of inputs $x^1 + x^2$ into the vector of outputs $y^1 + y^2$. To evaluate the potential efficiency gains from the merger, we shall evaluate the efficiency of the latter transformation, i.e. the use of $x^1 + x^2$ to produce $y^1 + y^2$.

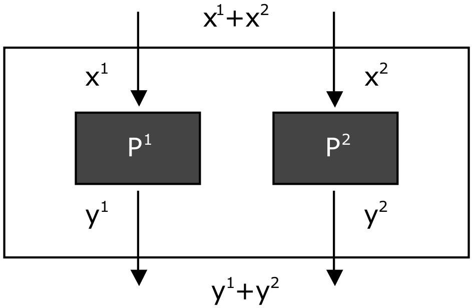

Figure 3: Horizontal integration

Of course, integration of more than two units is possible as well. Hence, in general we shall consider a situation where the units with indices in some set J are integrated. The full set of units we have information about shall be denoted I = {1,…,n}. Usually, J is a subset of I, but this need not be the case.

2.2.1 Resources and controls

It is clear that a good prediction of the likely corporate synergies requires at least two parts

- A delineation of the *resources and services* on which they collaborate as above
- A delineation of the *controls* (governance structures) by which the joint resources and services will be controlled

We shall initially focus on the first aspect, i.e. the production economic gains from integration or collaboration. This corresponds to a neo-classical approach to integration as discussed in Part A section 2.2.1.

The control aspects, i.e. the ability of the merged entity or the collaborating partners to coordinate, motivate and negotiate internally, are of course equally important. This corresponds to the incentive (moral hazard) approach to integration discussed in Part A section 2.2.3. It is intuitively clear, however, that the controls available following a merger may be hard to observe ex ante, and that the post merger reactions may therefore also be difficult to delineate ex ante. In particular, they will depend on a series of organizational and market characteristics. We shall discuss some of these below with the aim to make reasonable predictions, i.e. delineations of the span of possible reactions ex post. The intervals of likely reactions and some key determinant hereof will have to substitute for precise organizational and market analyses in a given situation.

The two aspects do of course interact and there may be a trade-off between the pursuits of production economic effects and the incentive aspects. An example of this could be that a joint operation allows for task specialization that could lead to production economic gains. The control aspects, however, may prohibit this exploitation of such possibilities since it may not empower anyone with the necessary instruments to coordinate the allocation of tasks. Hence, the control and production economic aspects clearly interact.

The same goes of course with the market and the organizational aspects of a merger. If for example the merger leads to less (yardstick) competition in the market, this may feed back into the organizational efficiency. The interaction with market aspects shall largely be ignored here.

2.2.2 Potential overall gains

The idea of making a priori estimates of potential production economic (cost) gains from mergers is developed in Bogetoft, Strange and Thorsen (2003) and in particular in Bogetoft and Wang (2005). We briefly recall the basic ideas here before turning to the development of refined measures below.

Consider a possible merger of the units in a subset J of the available units I= {1, 2, ..., n}. The merged unit is denoted DMU^J. Direct pooling of the inputs and outputs gives a unit that has used $\sum_{j\in J} x^j$ to produce $\sum_{j\in J} y^j$. This corresponds to having a completely decentralized (or divisionalized) organization where the decentralized units correspond to the J-units. Now the inefficiency of the directly pooled production plan

$$\sum_{j\in J} x^j = \text{combined inputs}$$
$$\sum_{j\in J} y^j = \text{combined outputs}$$

is a measure of the improvement potentials in the merged unit and can therefore be interpreted as the overall potential gains form the merger.

To illustrate this logic, consider Figure 4 below. Two hospitals A and B have been producing technically efficient in the past as indicated by the fact that they are located on the efficient frontier, the production function, ex ante. If they integrate but does not utilize the new synergies (in the illustration the economies of scale) they would spend $(x^1 + x^2)$ to produce $(y^1 + y^2)$ as indicated by the point A+B. This is however a technically inefficient combined production since there are feasible productions to the northwest of A+B, i.e. it is possible to find alternative productions that use fewer inputs to produce more outputs as reflected by the Potential Improvement PI set. The possibilities to improve can be summarized in different ways.

The simplest way is to use Farrell measure on the input side. The Farrell measure reduces to a simple comparison of horizontal length between A+B and C. We see that the aggregate input consumption can be scaled down with the factor E.

Instead of focusing on the input (cost reductions) we could use the Farrell measure on the output side. This would stress the possibilities to increase outputs with a factor F.

A more complete picture of the savings potential is to consider all point northwest of A+B. Any such point can be generated by a so-called directional distance function approach by varying the improvement direction. In this case again, we can get a score to measure the possible gains, namely G, but the disadvantage that is more delicate to interpret. Another way to think of the direction approach here is therefore simply as a mechanism by which all possible improvements can be generated. This is the approach we take in the Interactive Benchmarking software developed as part of this project.

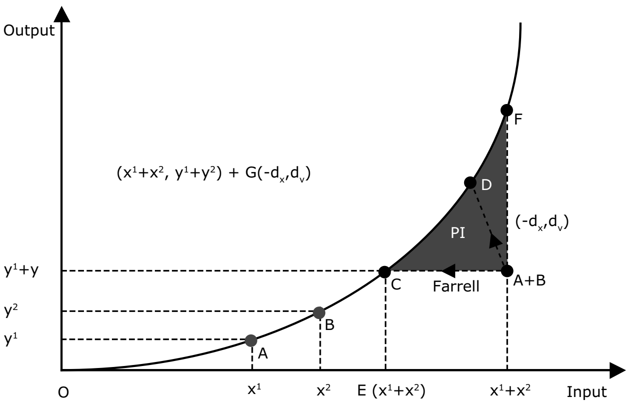

Figure 4: Overall gains from horizontal integration

2.2.3 *Formalizing the improvement potentials

A radial Farrell like input based measure of the *potential overall gains from merging* the J-DMUs is

$$E^J = \text{Min}\{E \in R_0 \,|\, (E[\textstyle\sum_{j\in J} x^j], \sum_{j\in J} y^j) \in T\}$$

E^J is the maximal proportional reduction in the aggregated inputs $\sum_{j\in J} x^j$ that allows the production of the aggregated output profile $\sum_{j\in J} y^j$.
If $E^J < 1$, we can save by merging. If $E^J > 1$, the merger is costly. A score of $E^J = 0.8$ would suggest that 20% of all inputs could be saved by integrating the units in J. Likewise a score of $E^J = 1.3$ would suggest that an integration would necessitate 30% more of all the resources. In some situations, the above problem may have no solution at all. It is however always feasible if T satisfies additivity. In particular, it is therefore feasible in the CRS DEA and FRH models discussed in Part C section 2.4.2.

One could of course measure the potential gains using many other indices. In particular one could do all of the evaluation and decompositions below on the output side. We focus on the input measure here since demand for hospital services are not easy to control at least in the short run.

More importantly, one may question the importance of summarizing the potential gains in a single, simple index. Indeed, the savings potentials are not fully captured by any single number since it is basically a set at possible improvement PI

$$PI^J = \{(x,y) \in R_0^{p+q} \,|\, (x,y) \in T \,,\, x \le \textstyle\sum_{j\in J} x^j \,,\, y \ge \sum_{j\in J} y^j) \in T\}$$

PI^J is the set of feasible productions that use no more inputs to produce no less outputs as illustrated in Figure 4.

An operational way to present this set is interactive benchmarking based for example on a directional distance approach. In this approach, one would look at $(\sum_{j\in J} x^j , \sum_{j\in J} y^j)$ as a preliminary budget and a point in the potential improvement set PI^J as an alternative plan or budgets for the integrated entity. The search for an improved budget and hereby the search for alternative allocations of the potential gains from the merger could then involve the reduction of some costs and the expansion of some services.

The directional distance function approach involved solving a problem like

$$G^J_d = \max \{G \in R_0 \,|\, (\textstyle\sum_{j\in J} x^j, \sum_{j\in J} y^j) + G(-d_x,d_y) \in T\}$$

Here $d_x \in R_0^p$ is the direction in input s pace and $d_y \in R_0^q$ is the direction in output space that we want to reduce and expand. The directional distance G is a measure of the number of times we can introduced the improvements packages ($-d_x$, d_y). The relevant direction in input–output space will depend on a series of factors like regulatory conditions, organizational control and market conditions.

The mathematical program above can be solved using several models of the technology T. We shall return to the DEA and SFA based descriptions of T in Part C, Chapter 2.

2.3 Dis-integrations

We have so far investigated the likely impact of merging two or more units. It depends on the details of the units we merge and the details of the underlying technology whether there are potential positive synergies. This means that it is sometimes less resource consuming to operate two independent rather than one joint unit. This is not surprising since the coordination and motivation tasks may be considerable inside large organizations. This also explains why we sometimes see different divisions of a joint enterprise operate independently, e.g. as individual profit-centres.

In fact we can use the same logic as above to investigate the potential gains from dis-integration of large entities. For the purpose of illustration, assume that we consider splitting up a unit (x^J,y^J) into two units (x^1,y^1) and (x^2,y^2). Now, if it is possible to find feasible plans for the individual units, i.e.

$$(x^1,y^1) \in T$$
$$(x^2,y^2) \in T$$

such that the individual units together use less resources to produce more services

$$x^J \geq x^1 + x^2$$
$$y^J \leq y^1 + y^2$$

we can look at these reduced inputs and expanded outputs as an indication of the potential gains from dis-integration.

***A measure of dis-integration gains**
Ignoring all the complexities of restricted controllability, restricted transferability and possible post dis-integration inefficiency, cf next Chapter, the potential savings from the disintegration can be calculated as

$$
\begin{aligned}
E = \quad & \min_{(x^1, y^1), (x^2, y^2)} E \\
\text{s.t.} \quad & Ex^J \geq x^1 + x^2 \\
& y^J \leq y^1 + y^2 \\
& (x^1, y^1) \in T \\
& (x^2, y^2) \in T
\end{aligned}
$$

That is, we seek to find two feasible production plans (x^1, y^1) and (x^2, y^2) that together are able to produce at least the same output as (x^J, y^J), and to ensure the largest possible proportional reduction of all inputs.

Note that if $E < 1$, there is a potential saving involved. This would typically happen when (x^J, y^J) operates somewhat above optimal scale size. It is of course also possible that $E > 1$ suggesting a net cost of forcing a dis-integration. Such analyses can therefore be used to make trade-offs between required dis-integrations to increase competition and losses in the production economic efficiency.

2.4 Vertical integration

Figure 5 illustrates a very simply and stylized vertical integration. An upstream firm (producer) P^1 uses p^{01} inputs (y^{01}) to produce p^{12} outputs (y^{12}), which is used as inputs in the downstream firm (processor) P^2 to produce p^{23} outputs (y^{23}). In a health context, P^1 could for example be a general practitioner, and P^2 could be a hospital.

Observe that we indicate resources or products here by both their origin and their destination nodes, and that we use 0 as a dummy initial source node and 3 as a final output sink node. The reason for this notation is that the distinction between inputs and outputs is ambiguous in vertical chains as the outputs of some are the inputs of others.

If the two units integrate but continue to operate as two independent entities as before, they would basically transform y^{01} to y^{23}. The question is if this is an efficient production - and if not what can be saved by better integration, coordination and exploitation of economies of scale. Vertical synergy (of the positive type) occurs when the firms through their integration are able to produce more services with given resources, or to produce given services with less resources.

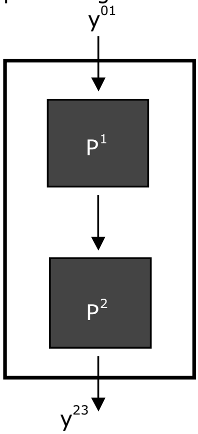

Figure 5: Simple vertical integration

2.4.1 Resources and controls

Like in the case of horizontal integration, to make reasonable predictions of the likely synergies from a vertical integration requires

A delineation of the *resources and services* on which they collaborate
A delineation of the *controls* (governance structures) by which the joint resources and services will be controlled

We shall be particularly focused on the first aspect, i.e. the production economic gains from integration or collaboration. This corresponds to a neo-classical approach to integration.

The control aspects, i.e. the ability of the merged entity or the collaborating partners to better coordinate, motivate and negotiate internally, are of course equally important. As discussed above, however, the control available following a merger may be hard to observe ex ante, and the post merger reactions may therefore also be difficult to delineate ex ante. In particular, they will depend on a series of organizational and market characteristics. In the case of vertical integration, we shall make some of these issues and cases clearer and find ways to measure their impact, by investigating a series of more or less involved integrations. In this way we can also decompose the potential gains into parts depending on what resources can be coordinated and which managerial resources are available following a merger.

2.4.2 Potential overall gains

Consider a closer vertical integration of the two entities P^1 and P^2. If the entities continue to operate as they do now, they effectively transform the primary input y^{01} to the final output y^{23}

$$y^{01} = \text{primary input}$$
$$y^{23} = \text{final output}$$

This is a possible production plan since there is a set of intermediate products y^{12}, which P^1 can produce, given its primary inputs, and which suffices in P^2 to produce the final outputs y^{23}. However, perhaps there are other intermediate outputs that P^1 can produce and perhaps P^2 with these inputs could produce other final outputs. To the extent that more final outputs can be produced with the given primal inputs (or less primal inputs are necessary to produce the same final output), we say that there are potential overall gains from vertical integration.

To illustrate this, consider Figure 6 below. Presently both the upstream entity and the downstream entity operates technically efficient since at the blue dot, the intermediate products are at the production possibility frontier $P(y^{01})$ and input consumption frontier $L(y^{23})$ of P^1 and P^2 , respectively. Still, a better coordination could allow reduced usage of the primal input, namely from y^{01} to Ey^{01} (or increased production of the final output, namely from y^{23} to Fy^{23}) by choosing intermediate productions in one of the two black dots.

As an example one may think of the upstream firm as a GP doing a certain combination of tests on the patients before they enter the downstream hospital, that undertakes treatments based in part on the tests undertaken by the GP. Now by change the test package it may be possible to save costs at the GP and still give as valuable information to the hospital, or it may be possible to make better use of the GP resources by doing an alternative test mix that would allow the hospital to produce more treatments since the patients now arrives with an more useful set of test results.

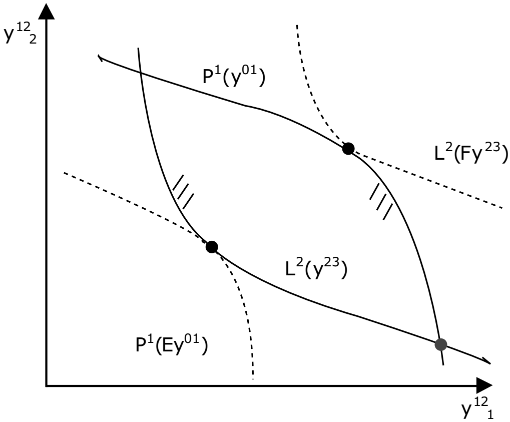

Figure 6 Overall vertical gains

More generally, to measure the potential gains from vertical integration in the simple case, illustrated in Figure 5, we shall evaluate the present (pre-merger observed) combined production plan

$$(y^{01}, y^{23})$$

i.e. the transformation of y^{01} to y^{23}, in the joint technology

$$T = \{(y'^{01}, y'^{23}) \mid y'^{01} \text{ can produce } y'^{23} \}$$

This is similar to the approach taken in the horizontal case. This means also that we can derive measures of the potential improvements in the same way. In particular we can for example use the Farrell base inputs efficiency

E = minimal primary input / actual primal inputs

Now, there are basically two ways to model the technology T. We could call them the aggregate and process oriented approaches, respectively.

In *the aggregate approach*, we would take real observations of actual transformations of inputs y'^{01} into outputs y'^{23}, and we would then construct an approximation of the corresponding technology. This corresponds to looking at the transformation as a black box and ignoring the possible sub-processes involved. The firms that have generated alternative pairs of primary inputs and final outputs (y'^{01}, y'^{23}) may have used different ways to do so – some may have outsourced the first process P^1, some may have out-sourced the second process P^2 and some may have done everything in-house. This is the concern of the aggregated approach. It just considers the transformation as a black-box and can therefore take data generated in different ways. Just like the organization structure or managerial style may differ between the units we compare in a usual benchmarking exercise.

In the *process-oriented approach*, we start instead by modelling the individual processes P^1 and P^2. This requires data on the process level. We may call the corresponding technologies T^1 and T^2 with associated output sets P^1 and P^2 and input sets L^1 and L^2. Now, having modelled the individual processes, we can construct the joint technology T analytically: we can consider a production plan (y'^{01}, y'^{23}) as feasible if the inputs y'^{01} suffices to produce some vector of intermediate goods y'^{12} that are sufficient to produce the final outputs y'^{23}.

It is clear that this approach may be considerably more demanding since it requires us to evaluate the possible effects of all possible intermediate product vectors. In practice, this will complicate the optimization problems needed to evaluate a given plan (y'^{01}, y'^{23}).

Since the latter is the more challenging approach, we will in general describe how to proceed in this case.

Before continuing, we note that it of course also is possible to combine the two approaches. If we have data on both the aggregate level for some units and on the process level for others, we can for example construct a technology that contains both sets and satisfy the usual properties of technologies.

It should be observed also that even though we can find the overall potential gains from vertical integration using the aggregate modelling of T, we will still need some models T^1 and T^2 to decompose this vertical merger gains into technical efficiency, harmony and size effects. We shall examine such decompositions below.

2.4.3 *Joint technology and improvement measures

To be more precise, we can formalize the joint technology derived via process descriptions as

$$T = \{(y'^{01}, y'^{23}) \mid \exists\, y'^{12} : y'^{12} \in P^1(y'^{01}),\ y'^{23} \in P^1(y'^{12})\}$$

It is worthwhile to note that if we construct T from T^1 and T^2, and if our models of T^1 and T^2 are based on best practice modelling, we effectively may generate new data or best practice data for the integrated entity T without having observed any such entities in reality.

In this way, the process-oriented approach extends the conceptual idea underlying the horizontal analyses. Here we use directly observed units to construct the technology. In the horizontal case, however as long as the technology is additive, i.e. if T+T is part of T, as it happens in for example a CRS or FRH technology, we would not get any new points using this procedure.

Consider now the overall evaluation of the possible gains from a vertical integration of P^1 and P^2, i.e. the potential improvements in the joint unit J = {1, 2}.

A radial Farrell like input based measure of the *potential overall gains from vertical integration* is therefore

$$E^J = \text{Min } \{E \in R_0 \mid (Ey^{01}, y^{23}) \in T\}$$

E^J is the maximal proportional reduction in actual inputs y^{01} that allows the integrated entity to produce the final outputs y^{23}. A score of E^J =0.8 would suggest that 20% of all inputs y^{01} could be saved by integrating the units in J. Likewise, a score of E^J =1.3 would suggest that an integration would necessitate 30% more of all the resources.

One could of course measure the potential gains using many other indices. In particular, one could do all of the evaluations and decompositions below on the output side using the output Farrell measure

$$F^J = \text{Max } \{F \in R_0 \mid (y^{01}, Fy^{23}) \in T\}$$

Indeed, the savings potentials are not fully captured by any single number since it is basically a set at possible improvement PI

$$PI^J = \{(y^{01}, y^{23}) \mid (y'^{01}, y'^{23}) \in T, y'^{01} \leq y^{01}, y'^{23} \geq y^{23}\}$$

PI^J is the set of feasible productions that use no more basic inputs to produce no less final outputs.

An operational way to present this set is via interactive benchmarking based for example on a directional distance approach. In this approach, one would look at (y^{01}, y^{23}) as a preliminary budget and a point in the potential improvement set PI^J as an alternative plan or budget for the integrated entity. The search for an improved budget and hereby the search for alternative allocations of the potential gains from the merger could then involve the reduction of some costs and the expansion of some services. As in the horizontal case, the directional distance function approach involved solve a problem like

$$G^J_d = \max \{G \in R_0 \mid (y^{01}, y^{23}) + G(-d_x, d_y) \in T\}$$

Here $d_x \in R_0^{p^{01}}$ is the direction in input space and $d_y \in R_0^{p^{23}}$ is the direction in output space that we want to reduce and expand. The directional distance G is a measure of the number of times we can introduced the improvements packages $(-d_x, d_y)$. The relevant direction in the input–output space will depend on a series of factors like regulatory conditions, organizational control and market conditions.

The advantage of this approach is than we can fine-tune the measurement of potential gains to reflect scarcity of different health care resources and preferences for health care services.

If one has access to input prices, one could instead use cost efficiency

$$CE^J = \text{Min} \{wy'^{01} \mid (y'^{01}, y^{23}) \in T\} / wy^{01}$$

i.e. a measure of the potential cost savings from finding a less expensive way to produce the intermediate products that are needed to produce the final outputs. Likewise, if one has access to prices or priorities on final outputs, one could measure the revenue efficiency as

$$RE^J = \text{Max} \{py'^{23} \mid (y^{01}, y'^{23}) \in T\} / py^{23}$$

i.e. as the possible expansion of the total revenue that the given primary input y^{01} is able to generate. Lastly, if we have both prices (or priorities) on the input and output sides, we could do an evaluation of the profit efficiency (or net-benefit or effectiveness) using

$$PE^J = \text{Max} \{py'^{23} - wy'^{01} \mid (y'^{01}, y'^{23}) \in T\} / (py^{23} - wy^{01})$$

i.e. we could measure the possibility to increase profits (net-benefits or effectiveness) above the present level generated by (y^{01}, y^{23}).

Small values of E and CE and large values of F, RE and PE suggest that the combined vertical production process can be improved considerably.

The mathematical programs above can be solved using several models of the technology T. We shall return to the DEA and SFA based descriptions of T in Part C, Chapter 2.

2.5 A spectrum of integrations

Above, we have discussed some simple horizontal and vertical integrations.

We will now provide a complementary treatment of the spectrum of possible integrations. We will discuss different levels of vertical integration ranging from simple trade of some intermediate products to full integration and sharing of possibly scare managerial resources.

A useful starting point is the more general network framework illustrated in Figure 7. In this we have two production units P^1 and P^2 that both use primary inputs to produce final outputs. In addition, P^1 produces intermediate products and P^2 consumes intermediate products.

The primary inputs used by P^1 and P^2 may without loss of generality be thought of as the same types as long as we make no assumption about strictly positive values. A product only consumed by one of the units can simply be represented by a zero in the corresponding coordinate for the other unit. The same goes for final outputs of course. This implies that additions make sense, i.e. it is meaningful to consider the total consumption of inputs as $y^{01}+y^{02}$ and the total production of final products as $y^{13}+y^{23}$, respectively.

We note also that the general network contains the pure horizontal case ($y^{12} = 0$) and the pure vertical case ($y^{02} = 0$, $y^{13} = 0$) from above as special cases.

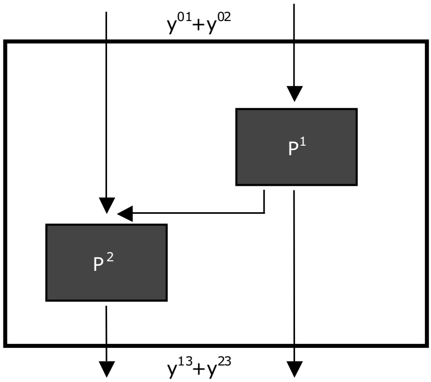

Figure 7: More general network

If we consider now the more realistic general case in Figure 7, we see that the integration may be more or less involved.

The *less involved integration* would assume that we can not reallocate the inputs y^{01} and y^{02} or the final outputs y^{13} and y^{23} among the two processes P^1 and P^2. This means that we can only improve the performance by change the composition of the intermediate product y^{12}. This could be relevant for example if the two entities use different types of inputs and produce different types of outputs, possibly on different geographic locations.

A *more involved integration* would allow the primary inputs to be shared or reallocated. In this case, gains from integration are not only a matter of coordinating on the intermediate products. Gains can therefore be generated by moving primary inputs to the sites with highest marginal products.

This corresponds to a situation where the production factors are *transferable* between the two processes. If for example P^1 is a primary physician and P^2 is a hospital, we can imagine that at least some of the secretarial resources can be transferred. Other resources, say quasi – factors like population (patient base), may be non-transferable due to their location specificity. In the latter case, we would use the less involved integration model above.

As the *most involved integration*, we may introduce also transferability of the outputs. In this case synergies are also possible by relying more on the most productive entities to produce the final outputs.

It is clear that these models are nested. The less involved integration production possibility set is contained in the more involved integration set which again is contained in the most involved integration case. This means that the saving possibilities are increasing the more involved the integration gets.

***Network technologies with varying integration**
In the less involved integration case, the feasible productions from the integrated entity is given by

$$P^J(y^{01}, y^{02}) = \{(y'^{13}, y'^{23}) \mid \exists\, y'^{12} : (y'^{12}, y'^{13}) \in P^1(y^{01}),\ y'^{23} \in P^2(y'^{12}, y^{02})\}$$

In the more involved integration case, the joint technology is given by

$$P^J(y^{01} + y^{02}) = \{(y'^{13}, y'^{23}) \mid \exists\, y'^{01}, y'^{02}, y'^{12}: (y'^{12}, y'^{13}) \in P^1(y'^{01}),\ y'^{23} \in P^2(y'^{12}, y'^{02}),\ y'^{01} + y'^{02} \leq y^{01} + y^{02}\}$$

As the *most involved integration*, we may introduce also transferability of the outputs.

$$P^J(y^{01} + y^{02}) = \{(y'^{13}, y'^{23}) \mid \exists\, y'^{01}, y'^{02}, y'^{12}, y''^{13}, y''^{23}: (y'^{12}, y''^{13}) \in P^1(y'^{01}),\ y''^{23} \in P^2(y'^{12}, y'^{02}),\ y'^{01} + y'^{02} \leq y^{01} + y^{02}\ y'^{13} + y'^{23} \leq y''^{13} + y''^{23}\}$$

2.6 Next steps

In this chapter we showed how to evaluate the overall potential gains from mergers. This is an interesting starting point, a best case, upper limit scenario that can be used by the competition authority as a first test to see if the efficiency gains can possibly outweigh the competitive effects.

The overall measures developed in this chapters are however optimistic and crude, and there is a need to make refinements in several directions.

First of all, some of the gains could possibly be obtained without mergers and can therefore not be associated directly with the mergers. In the next Chapters, We shall decompose the gains into learning, scope and scale effects to account for this.

Secondly, the overall gains may be too optimistic since there may be restrictions in the controllability and transferability of the resources and services. We shall show how to take this into account in Chapter 4.

Thirdly, one can question the assumption that the merged entity will be technically efficient given than firms even in highly competitive industries show inefficiencies. This assumption is implicit in this chapter, but we shall show how to relax it in Chapter 5.

3. Learning, scope and scale effects

Our measures of the potential overall merger gain from a merger encompass several effects. We shall now decompose it into technical efficiency, scale, and mix effects and discuss the organizational relevance of this decomposition. We start by illustrating the ideas in the horizontal case before turning to more formal derivations of the effects in the horizontal and vertical cases.

3.1 The idea

Some of the total improvement potentials in a merged entity are possibly available without a merger as well. One can therefore argue that the competition authority must trade-off the competitive effects against the increase in potential gains compared to other solutions rather than against the absolute level of potentials in a merged entity.

We can identify at least three sources of improvement.

One is *learning* and is associated with the ability to adjust to best practice. Consider a horizontal merger of A and B illustrated in Figure 8 below.

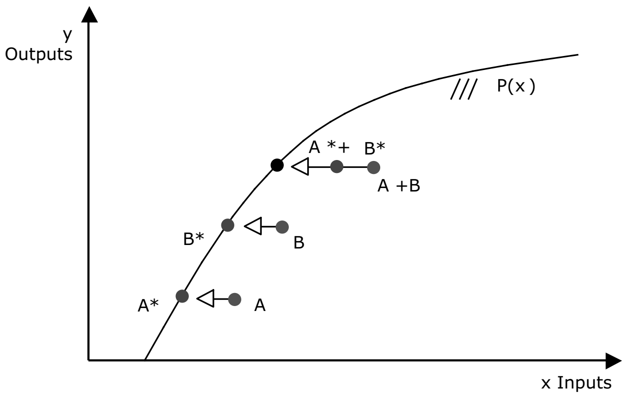

Figure 8: Learning effect

If the organizations merge but operate as they have done in the past, we see that there are considerable saving potentials as represented by the distance of A+B to the production possibility set. One can argue however that a considerable share of these potentials were available also on an individual basis if the individual entities had optimized their businesses as represented by the blue dots A* and B*. If businesses A* and B* integrate this would lead to the blue dot A*+B*, where the saving potentials are quite a bit less than in A+B. We shall think of this as learning or technical efficiency effect and say that this is not – at least not to a full extent – associated with the merger.

Another source of saving potentials is the so-called *scope or harmony effect* associated with the mix of resources used and the mix of services provided. To illustrate this, consider two hospitals with the same level of output and an input requirement set corresponding to the L(y) curve as illustrated in Figure 9 below.

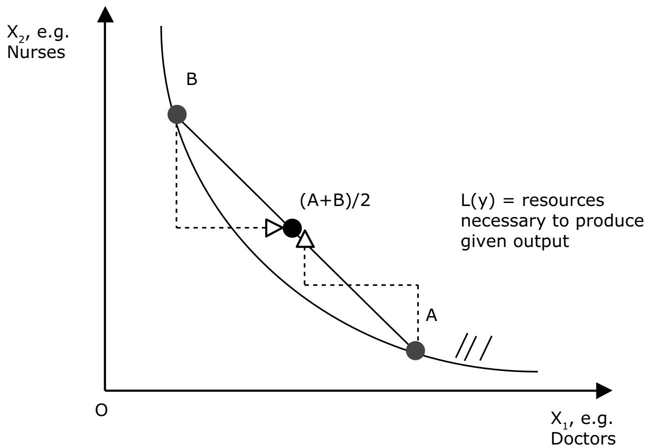

Figure 9: Scope effect

Now, A is quite doctor intensive while B is nurse intensive. It is clear, however, that neither of the factor mixes may be optimal – at least they cannot be optimal simultaneously. We see that the rate of substitution between doctors and nurses is different in the hospitals. In A it takes a large number of doctors to compensate the loss of extra nurses while in B it takes numerous nurses to compensate the loss of one doctor. This means that there are possibilities to improve by moving some nurses from B to A and some doctors from A to B. If we move the factors like indicated both end up in the black point (A+B) / 2. We see that there are now possibilities to save for each of the hospitals. Of course, similar possibilities exist on he output side, i.e. by moving some obligations from A to B and other obligations from B to A, we can get service combinations that requires less resources to produce or that matches the existing factor combinations in a better way. We shall talk about this effect as the harmony or the scope effect. Again, the point is that if independent enterprises just cooperated somewhat, they may gain and improve the pre-merger performance making the pure gains from a merger less.

In addition to these effects, a merger will also impact the scale of operation. This leads to the so-called *scale or size effect*. We have already illustrated Figure 4 the case of a horizontal merger. In case of technology with economies of scale, it is attractive to be large since this allows the organization to produce at lower average costs. Of course, the scale effect need not be positive – it depends on the underlying technology whether the increase in scale is favourable or not.

The tree effects above, technical, harmony and scale efficiency give the combined effect of a merger. We shall formalize the concepts below in a more technical section and then return to the interpretations and the integration of the effects.

3.2 Decomposing horizontal mergers

Using the above notions of learning (technical efficiency TE), scope (harmony H) and scale (size S) effects, we get our basic decomposition

$$E^J = TE^J * H^J * S^J$$

This corresponds to a decomposition of the basic merger index E^J into a technical efficiency index TE^J, a harmony index H^J, and a size index S^J. The technical efficiency measure TE^J captures what can be gained by making the individual units efficient. The remaining potentials to save, E^{*J}, are created by the harmony effect, H^J, and the size effect, S^J. How to calculate the measures exactly in general multiple input – multiple output technologies, will be discussed in the next sub-section.

The decomposition of the potential gains from merging DMUs into a technical efficiency measure, a harmony measure, and a size measure is important because full scale mergers are typically not the only available organizational option, and alternative organizational changes may be easier to implement. In particular, we suggest that the following may guide the *organizational restructuring*:

Low technical efficiency measure TE^J: One could let the inefficient DMUs learn from the practices and procedures of the more efficient ones. If the problem is not lack of skills, but rather lack of motivation, one could improve the incentives, e.g. by using relative performance evaluation and yardstick competition based on the technical efficiency measures, cf. Bogetoft (1994,95,97,00). Of course, if the problem is scarcity of managerial talent, it may still be necessary to make a genuine merger to transfer control to the more efficient administrative teams and hereby improve the managerial efficiency (X-efficiency). Another effect of a genuine merger emphasized by practitioners is the fact that a merger is a change event. Here established rule and procedure are being re-evaluated and improved. The logic of this is that every organization has some slack and it is difficult to reduce this under normal conditions.

Low harmony measure H^J: One could consider reallocating the inputs and outputs among the DMUs to create more "powerful" input mixes and more easily produced output mixes. This can be done a) inside a hierarchy, b) by long term contracts or perhaps c) by creating a market for key inputs and outputs, cf. also Andersen and Bogetoft (2007), Bogetoft e.a. (2007), and Brännlund, Chung, Färe and Grosskopf (1998).

Low size measure S^J: In this case, full scale mergers may be the only alternative. If we need large amounts of fixed capital, highly specialized staff, long run-lengths or simply a critical mass to obtain sufficient returns from scale, it may be relevant to merge. Also, and perhaps most importantly, this may be relevant if the reallocation through contracts or a market is associated with too many transaction costs to make it attractive, cf. the general discussion of the size of the firm in the industrial organization literature, e.g. Tirole (1988).

3.2.1 *Formalizing the effects in horizontal mergers

To avoid compounding the effects, it is useful to adjust the overall merger gains for the *technical efficiency effect*. As a first approach to this, we project the original units to the production possibility frontier and use the projected plans as the basis for evaluating the remaining gains from the merger. Thus, we project (x^j, y^j) into $(E^j x^j, y^j)$ for all $j \in J$, where $E^j = E^{\{j\}}$ is the standard efficiency score for the single DMU^j, and use the projected plans $(E^j x^j, y^j)$, $j \in J$ as the basis for calculating the *adjusted overall gains* from the merger

$$E^{*J} = \text{Min}\ \{E \in R_0 \mid (E[\textstyle\sum_{j\in J} E^j x^j], \sum_{j\in J} y^j) \in T\}$$

Letting

$$TE^J = E^J / E^{*J}$$

we get $E^J = TE^J * E^{*J}$, where $TE^J \in [0, 1]$ indicates what can be saved by individual technical efficiency adjustments in the different units in J.

Assuming that individual technical inefficiencies have been dealt, we are left with the scaling or size effect on the one hand and the harmony, scope or mixture effect on the other hand. Without further assumptions about the technology, we cannot say whether size and harmony effects favor a merger. This issue is discussed in details in Bogetoft and Wang (2005). Here it suffices to note that with a convex technology the harmony effect generally favors a merger, while the size effect may or may not favor a merger.

To formalize the *harmony gains* we examine how much H^J of the average input could have been saved in the production of the average output

$$H^J = \text{Min } \{H \in R_0 \mid (H[|J|^{-1} \textstyle\sum_{j \in J} E^j x^j], |J|^{-1} \textstyle\sum_{j \in J} y^j) \in T\}$$

where |J| is the number of elements in J. We look at the average input and average output, since we do not want the expansion of size to come into play yet. Using the average is most relevant if the units in J are not too different in size to begin with. If the sizes differ considerably, we may be picking up scale effects, e.g. if some units are larger than and some are smaller than the "most productive scale size" as defined by Banker (1984). Note that $H^J < 1$ indicates a savings potential due to improved harmony, while $H^J > 1$ indicates a cost of harmonizing the inputs and outputs.

Next, we capture the *size gains* by asking how much could have been saved by operating at full scale rather than average scale, i.e. by the measure S^J

$$S^J = \text{Min}\{S \in R_0 \mid (S[H^J \textstyle\sum_{j \in J} E^j x^j], \textstyle\sum_{j \in J} y^j) \in T\}$$

The re-scaling is advantageous ($S^J < 1$) if we have economies of scale, and costly ($S^J > 1$) if the return to scale property does not favor larger units.

3.2.2 *The rationale of the harmony measure

The decomposition developed above gives a natural way, we believe, to define and distinguish between the technical efficiency, the size and the harmony effects. In Bogetoft and Wang (2005) additional motivations for these measures are provided. Most importantly, it is shown that with a convex technology, the harmony effect measures are the most that can be gained by any kind of reallocation between the units in J. Assuming that we were to pick new inputs and outputs $(x^{\#j}, y^{\#j})$ for each $j \in J$ such that total inputs and outputs stay feasible, $\sum_{j \in J} x^{\#j} \leq \sum_{j \in J} x^j$ and $\sum_{j \in J} y^{\#j} \geq \sum_{j \in J} y^j$, and such that all of the new production is possible $(x^{\#j}, y^{\#j}) \in T$, the largest possible saving in $\sum_{j\in J} E^j x^j$ is precisely the harmony effect. That is, assuming so-called free disposability and convexity of the technology T, H is also the solution to

$$\begin{array}{ll} \underset{(x^{\#j},\, y^{\#j}),\ j \in I}{\text{Min}} & h \\ \text{s.t.} & h \sum_{j\in J} E^j x^j \geq \sum_{j\in J} x^{\#j} \\ & \sum_{j\in J} y^j \leq \sum_{j\in J} y^{\#j} \\ & (x^{\#j}, y^{\#j}) \in T \end{array}$$

3.2.3 *Decomposition with a cost function

Before closing this discussion of the basic decomposition, it may be useful to illustrate it in the context of a single input (cost) multiple output context. Thus, let x^j be the cost of producing y^j in DMU^i, and let $c(y) = \min \{x \in R_0 \mid (x, y) \in T\}$ be the underlying cost function which gives an alternative representation of the underlying technology. We then have

$$E^J = c(\textstyle\sum_{j\in J} y^j) / \sum_{j\in J} x^j$$
$$E^{*J} = c(\textstyle\sum_{j\in J} y^j) / \sum_{j\in J} c(y^j)$$
$$TE^J = \textstyle\sum_{j\in J} c(y^j) / \sum_{j\in J} x^j$$
$$H^J = c(|J|^{-1}\textstyle\sum_{j\in J} y^j) / |J|^{-1}\sum_{j\in J} c(y^j)$$
$$S^J = c(\textstyle\sum_{j\in J} y^j) / |J| c(|J|^{-1}\sum_{j\in J} y^j)$$

The total potential gains can be decomposed into: (1) the learning effect TE^J measuring the reduction in costs if everyone learns best practice but remains independent entity, (2) the harmony effect H^J measuring the minimal cost of the average output vector compared to the average of the costs corrected for individual learning, and (3) the size effect S^J measuring the cost of operating at the full (integrated) scale compared to the average scale of the original entities.

3.3 Decomposing vertical mergers

In this section, we discuss ways to decompose the overall gains from vertical integration into elements that can be realized with intermediate forms of collaboration and integration.

In the following, we will develop decompositions that distinguish between

- Individual inefficiencies
- Structural inefficiencies

and we will decompose the structural inefficiencies into

- Allocative (mix) inefficiencies
- Scale (size) inefficiencies

The advantage of decompositions according to these types of inefficiencies is that they align with traditional decompositions of individual inefficiencies and that they - in part as a consequence - allow for relative simple interpretations. Thus, we may think of *individual inefficiency* as technical, cost or revenue inefficiency in either of the two individual entities P^1 and P^2, and *structural inefficiency* as inefficiency in the coordination of the two levels. Structural inefficiency may result if the two levels are sub-optimally aligned (mix effect) or if the trade-off between the economies of scale in the two levels are sub-optimal (size effect). Note that the interpretation of mix effect and size effect is different to those in case of horizontal integration.

As in the horizontal case, the idea and motivation of the decomposition is that different inefficiencies may call for different organizational remedies:

- Low individual efficiency would suggest a need for learning, better incentive schemes etc. in P^1 and P^2, respectively.
- Low mix efficiency would suggest better coordination, synchronization and matching between P^1 and P^2, and hereby better supply chain contracts, more appropriate allocation of decision rights etc.
- Low size efficiency would suggest a close cooperation, possible a full scale integration such that the trade-off between possible cost and benefits of size in P^1 and P^2 can be internalized.

It is worth emphasizing that the organizational prescriptions are less clear-cut in the vertical than in the horizontal case. One can argue as above that a lack of coordination could be handled by better supply chain contracts – but one could in principle handle a lack of optimal size adjustments in the same way. In the first case, the contact changes would be focusing on changes in the mix – say how improve synchronization of production processes. In the second case the contract improvements would focus on changes in the production levels and possibly the need to pay one unit to move further ways form optimal scale size. In the horizontal case, the size effect was related to the economies of scale of the fully merged unit and hence required a more close integration of all activities into the same organizational entity.

Even when we restrict ourselves to decomposition according to these tree dimensions there are numerous ways to do the decomposition. The multiplicity of decompositions follow from alternative ways to choose

- Direction of measurement, e.g. in terms of input saving, output expansions, etc (as above)
- Order of measurement, i.e. do we evaluate structural inefficiency presuming that individual inefficiencies have already been eliminated as it is usually done, or do we choose the other order as advocated in for example Bogetoft, Färe and Obel (2006) and Andersen and Bogetoft (2007)
- Sharing of gains, i.e. how do we foresee that the gains shall be shared among P1 and P2 and which "anchor point", cf. below, should we therefore decompose around.

Since the multiplicity of choice above may lead to different estimates of the individual and structural efficiencies, and thereby also to different policy implications, it is important to develop the sensitivity of the decompositions to the choices and the linkage with incentive and bargaining theory. We shall take the analyses some of the way, but we do not want to leave the impression that there is only one right decomposition or that we will provide a complete and ultimate coverage of the possible decompositions.

To get started, we shall now develop a rather simple and straightforward decomposition by presuming that 1) individual inefficiencies are eliminated first, 2) the gains (losses) from the simultaneous change of scale in P^1 and P^2 is determined next and 3) finally the gains from improved coordination (mix) across P^1 and P^2 is determined. We shall do this on the input side, but note that parallel decompositions would be possible on the output side as well.

3.3.1 Individual inefficiency effect (Technical efficiency effect)

Some or all of the units in J may be technically inefficient and this may be captured in E^J. A merger may bring in new management, which may facilitate the elimination of such inefficiencies. However, it is also possible to reduce technical inefficiencies through other means, e.g. by imitating better performers, sometimes referred to as peer units. To avoid compounding the effects, it is useful to adjust the overall merger gains for the technical efficiency effect.

As a first approach to this, we project the original units to the production possibility frontier on the input side. This is we project (y^{12}, y^{23}) to (E^2y^{12}, y^{23}) and we project (y^{01}, y^{12}) to (E^1y^{01}, y^{12}). Here we have used a shorthand notation, E^1 and E^2, for the input reductions in the two entities. To be more precise, we should of course indicate the production plans in which these reductions are calculated, namely $E^1 = E^1(y^{01}, y^{12})$ and $E^2 = E^2(y^{12}, y^{23})$.

Both of the units are now technically efficient, and a simple measure of the effects of this, the *individual inefficiency effect* or the *technical efficiency effect* is

$$TE = E^1 E^2 = E^1(y^{01}, y^{12})\, E^2(y^{12}, y^{23})$$

Thus for example, if we can save 10% in P^2 and 20% in P^1 we would get TE=0.8*0.9 = 0.72 suggesting a total savings potential of 28% - a little less that a 30% saving since we cannot save 10% of the 20% resources we did not spend in P^1 since we already saved them.

The projections involved in the determination of the technical efficiency effect are illustrated in Figure 10 below, where we illustrate all processes in the intermediate product space. We see that each of the production processes P^1 and P^2 are technically inefficient. With the given input use, y^{01}, the intermediate outputs (blue dot in the figure) produced by P^1 is not on the production possibility frontier $P^1(y^{01})$. Indeed, we could reduce the input consumption to E^1y^{01} and still be in the feasible output set $P^1(E^1y^{01})$. Likewise, the second process is not operating efficiently. The given input, y^{12}, is more than what is needed to produce y^{23} efficiently as illustrated by the input iso-quant $L^2(y^{23})$. Hence, we could reduce the consumption of intermediate products to E^2y^{12} and still have enough to produce the present output y^{23}.

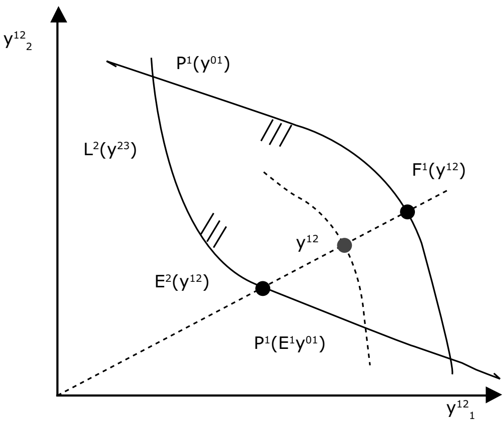

Figure 10: Individual inefficiencies

From the potential total gains $E^J(y^{01}, y^{23})$ and the individual inefficiency impact TE, we can also determined the *structural efficiency*,

$$SE = E^J / TE = E^J(y^{01}, y^{23}) / E^1(y^{01}, y^{12})\ E^2(y^{12}, y^{23})$$

such that $E^J = TE * SE$.

We shall now take a close look at the structural efficiency component.

3.3.2 Size (scale) effect

To capture the effects of size, we note that the reduction in y^{12} resulting from the elimination of technical inefficiency in P^2 would lead to a savings potential in P^1 simply by the reduced size of the total operation and irrespectively of P^1 being efficient or not.

To capture the combined effect of the reduced output requirement to P^1 and the elimination of the incumbent inefficiency in P^1, we can calculate

$$E^{**J} = E^1(y^{01}, E^2 y^{12})$$

The idea of this measure is that we first reduce the output requirement to P^1 by eliminating the inefficiency in P^2. This reduces the required output of P1 to E^2y^{12}. Next, we calculate how much we can reduce the input in P^1 by 1) elimination of any inefficiency in P^1 and by 2) taking into account the need reduced need for outputs from P^1.

The individual inefficiencies as well as the effect of scale are illustrated in Figure 11 below. In Figure 11 we have illustrated the intermediate product on the horizontal axis –in the case of multiple products, we may simply think of this as the length of the intermediate product vector with a given mix of such products. On the vertical axis, we have illustrated the original primary inputs as well as the final outputs

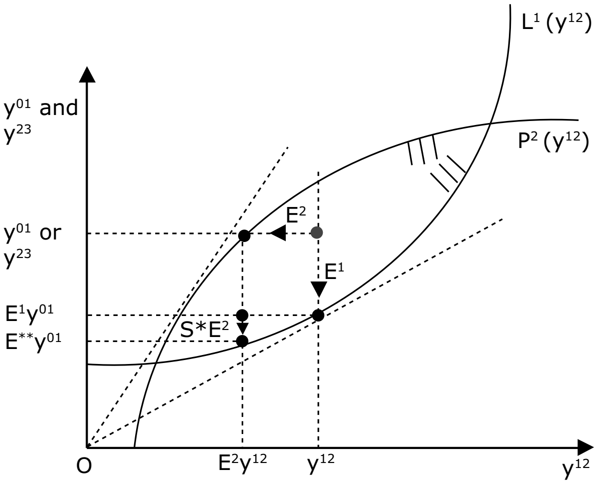

Figure 11: Size effect

Whether a given reduction of the required outputs from P^1 leads to a smaller or larger reduction in the necessary resources in P^1 will depend on the return to scale. If P^1 is operating above optimal scale size, i.e. at decreasing return to scale, the reduced output requirement leads to a high reduction in the necessary inputs. If instead P^1 is operating below optimal sale size, i.e. with increasing economies of scale, the resulting input reductions will be smaller.

A natural measure of the size effect is therefore

$$S = E^{**J} / TE = E^1(y^{01}, E^2y^{12}) / [E^1(y^{01}, y^{12})\ E^2(y^{12}, y^{23})]$$

In the example in Figure 11, we see that the elimination of inefficiency in P^2 leads to a saving of approximately 40% inputs to P^2 (from y^{12} to E^2y^{12}) and hereby a reduction in the output requirements from P^1 with approximately 40%. Unfortunately, P^1 is operating below optimal scale size and the input saving are therefore somewhat less, approximately 20% (from E^1y^{01} to $E^{**}y^{01}$). This means that the scale effects are counteracting the saving from improved individual technical efficiency and we get S of approximate 1.2 corresponding to an increased spending by the scale of approximate 20%.

Another way to think of this is to note that TE measures the combined effects of eliminating the input inefficiency in the two units assuming a constant return to scale technology; when we multiply the effects E^1 and E^2, we implicitly or tentatively assume that a reduction of the output requirements to P^1 of E^2 leads to a proportional reduction in the input requirements. In reality, this may not be the case. If P^1 is below optimal scale size, the reduction is less. If instead P^1 is operating above optimal scale size, the reduction is expanded since production of outputs from P^1 is particularly expensive in this case.

3.3.3 Harmony (mix, scope) effect

The pure (net) gains from better alignment of the productions in P^1 and P^2 can now be determined as the remaining potential to improve the production plan from E**, i.e. as

$$H^J = E(E^{**J}y^{01}, y^{23}) = E(\ (E^1(y^{01}, E^2y^{12})\ y^{01}, y^{23})\ )$$

The idea of this measure is that we have first eliminated any excessive resource usage in an organization when an intermediate goods vector proportional to the original vector accomplishes the coordination across the production levels P^1 and P^2. We now determine the remaining possibilities to save.

Using that the Farrell input distance function is homogeneous in degree -1 in inputs, we have

$$\begin{aligned} H^J &= E(\ (E^1(y^{01},E^2y^{12})y^{01},y^{23})\ ) \\ &= E^1(y^{01},E^2y^{12})^{-1}E^J(y^{01},y^{23}) \\ &= E^J(y^{01},y^{23})/\ E^1(y^{01},E^2y^{12}) \end{aligned}$$

The harmony measure is a measure of the gains from improved coordination across P^1 and P^2. It measures the gains from making sure that the P^1 production stage produces a mix of intermediate products that is particularly useful for the processing stage P^2. This is illustrated in Figure 12 below.

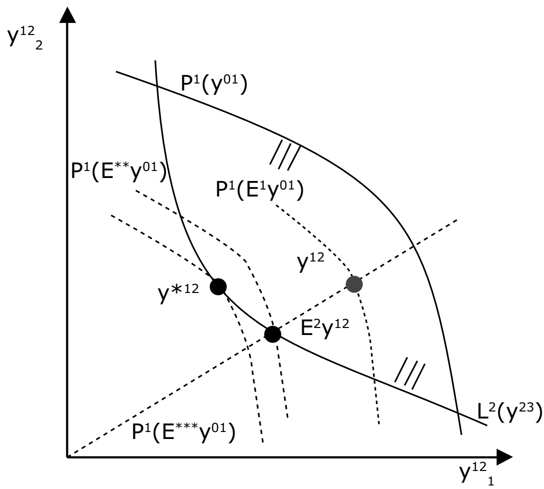

Figure 12: Harmony (mix) effect

As illustrated in Figure 12, having eliminated individual inefficiency and accounted for the resulting change in scale, there are still possibilities to save by moving from point E^2y^{12} to the optimal $y*^{12}$. The idea is that the marginal rate of transformation in P^1 must equal the marginal rate of substitution in P^2 - or if there are several such rates, they contain a common element (the so-called sub-differentials overlap).

The harmony effect resembles the notations of allocative and revenue efficiency known from ordinary production economics. The latter concepts, however, presume prices on the inputs or outputs. In the simple vertical merger case, we do not presume a perfect market for the intermediate goods. Rather, we substitute prices with the marginal values of the intermediate goods as derived from the other production process. In this way, we use endogenous prices instead of exogenous ones.

3.3.4 *The effects as constrained improvements

We can also say that E** measures what can be saved by the integrated entity as long as we do not allow the mix of intermediate products to change. Hence, we have – compared to the potential overall gains from vertical integration $E^J = \text{Min } \{E \in R_0 \,|\, (Ey^{01}, y^{23}) \in T\}$ – added one restriction, namely that the vector of intermediate products shall be proportional to the original vector of intermediate products. That is, we have:

$$
\begin{aligned}
E^{**J} = \quad & \underset{E,y'^{12},k}{\text{Min}} \quad E \\
& \text{s.t.} \quad y'^{12} \in P^1(Ey^{01}) \\
& \qquad\quad y^{23} \in P^2(y'^{12}) \\
& \qquad\quad y'^{12} = ky^{12}
\end{aligned}
$$

The harmony measure can also be interpreted along these lines. We can say that H^J measures the effect of relaxing the

$$y'^{12} = ky^{12}$$

constraint in the E^{**J} program above: If we solve the E^{**J} and then solve it again without the $y'^{12} = ky^{12}$, i.e. solve

$$
\begin{aligned}
E^{***J} = \quad & \underset{E,y'^{12}}{\text{Min}} \quad E \\
& \text{s.t.} \quad y'^{12} \in P^1(Ey^{01}) \\
& \qquad\quad y^{23} \in P^2(y'^{12})
\end{aligned}
$$

and call the resulting value E^{***J}, we have $E^{***J} = H^J * E^{**J}$ and therefore

$$H^J = E^{***J} / E^{**J}$$

Observe also that since the E^{***J} program is a relaxation of the E^{**J} program, we always have $E^{***J} \leq E^{**J}$. It follows therefore that the harmony index is always less than or equal to 1 in the vertical integration case. That is, irrespectively of the characteristics of the underlying process technologies, the harmony effect is never working counter to the integration.

3.3.5 Basic decomposition in the vertical case

Using the definitions above we have $H^J = E^J(y^{01}, y^{23}) / E^1(y^{01}, E^2y^{12})$, $E^J = E^J(y^{01}, y^{23})$ and $E^1(y^{01}, E^2y^{12}) = S^J * TE^J$. We hereby get the *basic decomposition* in the vertical case

$$E^J = TE^J * H^J * S^J$$

like in the horizontal case.

This corresponds to a decomposition of the basic merger index E^J into a technical efficiency index T^J, a harmony index H^J, and a size index S^J. The technical efficiency measure T^J captures what can be gained by making the individual units efficient. The remaining potentials to save, E^{*J}, are created by the harmony effect, H^J, and the size effect, S^J.

The decomposition of the potential gains from merging DMUs into a technical efficiency measure, a harmony measure, and a size measure is important because full scale mergers are typically not the only available organizational option, and alternative organizational changes may be easier to implement. In particular, we suggest that the following may guide the *organizational restructuring*:

Low technical efficiency measure TE^J: The organizational response to low technical efficiency could as in the horizontal case be to facilitate learning, to improve the motivation via performance based contracting, or to bring in more talented management. The latter may be facilitated by a genuine merger if one of the levels have superior management and if the managerial skills can be extend from a one-level to a two-level organization. Also a merger may be a change event where established rule and procedure are being re-evaluated and improved.

Low harmony measure H^J: One should consider ways to better coordinate or align production across the two production processes. This means that the P^1 level should be directed towards the production of outputs that are particularly useful in P^2 and that P^2 should adjust its factor usage towards product combinations that are easier to provide by in P^1. Of course, such alignments can be done both inside a hierarchy using instructions and via a market or transfer pricing arrangement. Careful synchronization are usually best done using instruction, c.f. Bogetoft and Olesen(2004). This would suggest that the harmony effect is most easily harvested in an integrated organization. On the other hand, a lack of perfect synchronization does not necessarily imply a large loss of value and this suggest that also a price-based approach can be used, and that a full integration may not be necessary.

Low size measure S^J: A low size measure essentially means that the input saving in the down stream unit is particularly valuable since it would allow the upstream unit to move closer to optimal scale size. It represents a second order effect of improved efficiency in the down stream firm or an interaction effect between the two levels. Again to harvest the size effect, a genuine merger is strictly speaking not necessary. In this way, the size effect differs in the interpretation from the size effect in the horizontal case. In the horizontal case, the pooling of production leads to operations at a higher scale, and it is the economies of scale at the pooled levels that matters. In the vertical case, the production processes may be of a very different nature, using for example totally different equipment or labor with totally different characteristics, and the pooling effect may therefore be nil. The alignment of scale at the two levels could involve P^1 compensating P^2 for the possibility to operate at a more optimal scale size – or P^2

compensating P^1 if it is forced further way from the optimal scale size. In practical terms, we may expect the unit price charged by P^1 declines when $S < 1$ and increases when $S > 1$.

In the vertical case, we therefore see that the gains in theory may be harvested by improved contracting between independent parties running P1 and P2 respectively. Of course, if contracting is incomplete or in general associated with too many transaction costs to make it attractive, cf. the general discussion of the size of the firm in the industrial organization literature, e.g. Tirole (1988), then we may also need a genuine merger to realize the H and S effects in the vertical case. An example of this could be related to specific investments. It may be that a change of scale or mix would require P^1 to make a specific investment in new technology or procedures. By the hold-up problem discussed in Part A section 2.2.4, such investments are difficult in a supply chain context unless it is possible to make long term contracts – or to internalize the effects via a genuine merger.

4. Restricted controllability and transferability

In the estimates of potential merger gains above, we have assumed that all inputs and outputs can be redistributed in the merged entity J. In many cases, this approach is too restrictive. At least in a short run perspective, some dimensions are easier to change and reallocate than others. It may for example be easier to reduce the labor input than the capital input, which is largely based on sunk investments. Also, some services may have to be provided on location and can therefore not be transferred to another unit, located elsewhere. It may for example be possible to transfer IT, accounting and HR to another location, but the production of emergency room services cannot easily be relocated. Lastly, some variables in actual models typically describe the context rather than choice variables, and they are therefore not transferable. Population density, education level, and age distribution, for example, has limited transferability.

We will now show how to evaluate efficiency and calculate potential gains when only some of the inputs and outputs can be *adjusted* and *transferred* among the members of the new and merged entity. First we consider the relatively straightforward case with restricted possibilities to control inputs and outputs, and then we extend by introducing restrictions in the transferability of some of the resources and services.

4.1 Controllability

In the modern efficiency analyses literature, it is common to account for the *non-discretionary* character of some dimensions by only looking for improvements in the other directions, cf. e.g. Charnes, Cooper, Lewin and Seiford (1994).

Assume that we can split the inputs x and outputs y in two types, $x = (x_V, x_F)$ and $y = (y_V, y_F)$, corresponding to the variable (controllable) V and fixed F (non-controllable) dimensions.

The relevant measures in this case are based on the idea of so-called *sub-vector* efficiency. Focusing on input reductions, we would look for the largest reduction of all controllable inputs that together with the fixed inputs allow the unit to produce given outputs.

To illustrate, consider the case in Figure 13 below. We assume that the integrated entity A+B is using nurses and doctors in its production. Presuming also that doctors cannot be adjusted but that nurses can, e.g. because doctors have more bargaining power or more rigid contracts, the improvement will have to take place along the nurse axis, the vertical axis. The savings potential (improvement potential) in the merged entity will therefore be a fraction E^R of the nursing staff only.

The classical distinction between discretionary and non-discretionary inputs can be generalized by thinking in terms of directional distance functions. Instead of the extreme dichotomy of variables, we can allow that some resources are simply easier or more desirable to adjust than others. Likewise, some outputs are more easy or desirable to expand than others. Assuming that we look for improvements in the direction $(-d_x, d_y)$, we can decompose a measure G^J_d of the total directional improvement into learning, scope and scale effects much like above. The detailed calculations are given in the technical sections below.

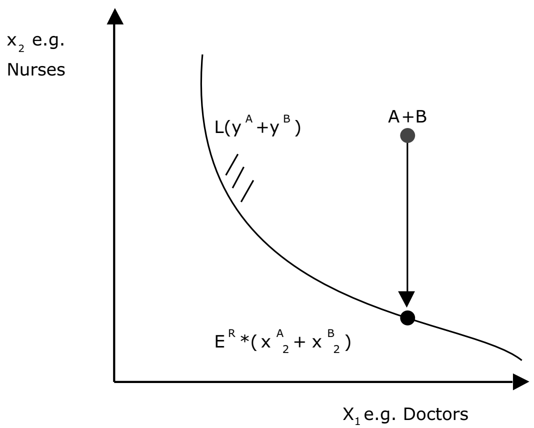

Figure 13: Restricted controllability

The approaches above apply equally well to horizontal and vertical cases, in particular when the joint technology in the vertical case is estimated directly from comparable joint operations.

In the vertical case, there is an additional obstacle, namely when the joint technology is constructed from process technologies. It may be that only some of the intermediate resources are discretionary. In that case, the construction of the reference technology must reflect this. We illustrate this in the technical sub-section below.

4.1.1 *Horizontal decomposition using sub-vectors

To formalize the sub-vector approach, we measure the efficiency of DMU^i as

$$E^i_V = \text{Min } \{E \in R_0 \mid (Ex^i_V, x_F, y^i) \in T\}$$

Likewise the potential gross gains from a horizontal merger of the J units is given by

$$E^J_V = \text{Min } \{E \in R_0 \mid (E[\Sigma_{j\in J}\, x^j_V], [\Sigma_{j\in J}\, x^j_F], \Sigma_{j\in J}\, y^j) \in T\}$$

i.e. E^J_V is the maximal proportional reduction in the variable (discretionary) inputs that together with the fixed (non-discretionary) resources allows the production of the aggregated output profile $\Sigma_{j\in J}\, y^j$. If $E^J_V < 1$, we can save the fraction $(1 - E^J_V)$ of the variable inputs by merging. If $E^J_V > 1$, the merger is costly and requires that the total usage of the variable resources is increased.

As previously, we may also filter out the effects of individual inefficiencies by determining the *adjusted overall gains* in the direction of the variable inputs as

$$E^{*J}_V = \text{Min}\{E \in R_0 \mid (E[\Sigma_{j\in J} E^j_V\, x^j_V], \Sigma_{j\in J}\, x^j_F\, , \Sigma_{j\in J}\, y^j) \in T\},$$

the *harmony effect* as

$$H^J{}_V = \text{Min } \{H \in R_0 \mid (H[|J|^{-1}\Sigma_{j\in J} E^j\, x^j_V], |J|^{-1}\Sigma_{j\in J}\, x^j_{F,}\, |J|^{-1}\Sigma_{j\in J} y^j) \in T\},$$

and the *size effect* as

$$S^J_V = \text{Min } \{S \in R_0 \mid (S[H^J\, \Sigma_{j\in J} E^j\, x^j_V], \Sigma_{j\in J}\, x^j_F, \Sigma_{j\in J} y^j) \in T\}$$

The interpretations and organizational implications of these scores are as previously explained, except that all evaluations are now in savings of only the controllable inputs and they are calculated conditional on the given levels of the non-controllable inputs. Thus, for example, re-scaling is advantageous, $S^J_V < 1$, if we have economies of scale in (x_V, y) for given x_F, and costly, $S^J_V > 1$, if the return to scale property does not favor larger units for the given values of the fixed inputs.

Using the above definitions, we once again get decomposition

$$E^J_v = TE^J_v * H^J_v * S^J_v$$

This corresponds to a decomposition of the basic merger index E^J_V into a technical efficiency index TE^J_V, a harmony index H^J_V, and a size index S^J_V.

4.1.2 *Horizontal decomposition using directional distance functions

Using directional distances, we get a similar although additive decompositions.

First, to correct for individual improvement potentials, we calculate

$$G^{*J}_d = \text{Max} \{G \in R_0 \,|\, (\Sigma_{j\in J}(x^j - G^j d_x),\ \Sigma_{j\in J}(y^j + G^j d_y)) + G(-d_x, d_y) \in T\}$$

i.e. G^{*J}_d measures the number of times we can introduce improvement packages $(-d_x, d_y)$ after each and every DMU has individually introduced the maximal possible improvement. Hence, a measure of the potential impact of individual initiatives is

$$GTE^J_d = (G^J_d - G^{*J}_d)$$

The remaining possibilities to improve can as before be split into two components.

One refers to the impact of a charged *mix* of resources and services and can be calculated by considering what the average, individually adjusted unit could save and then aggregating this to the level of the total entity

$$GH^J_d = |J| * \text{Max} \{H \in R_0 \,|\, ([|J|^{-1}\Sigma_{j\in J}(x^j - G^j d_x)],\ |J|^{-1}\Sigma_{j\in J}(y^j + G^j d_y)) + H(-d_x, d_y) \in T\},$$

What remains then is the impact of the merged unit operating at a larger scale size

$$GS^J_d = G^{*J}_d - GH^J_d$$

In total, this leads to a decomposition as before, just now in an additive way

$$G^J_d = GTE^J_d + GH^J_d + GS^J_d$$

The interpretation of this decomposition is similar to that of the basic decomposition – it splits the savings potential in the direction $(-d_x, d_y)$ into gains from individual learning, gains from improved mix of inputs and outputs and gains from a change in the scale of operation.

4.1.3 *Restricted controllability in the vertical case

Consider the simple vertical case. The integrated entity is assumed to use the technology

$$T = \{(y'^{01}, y'^{23}) \mid \exists\, y'^{12} : y'^{12} \in P(y'^{01}), y'^{23} \in P(y'^{12})\}$$

Now if we assume that only some of the primary inputs are variable (discretionary), y^{01}_V while the rest y^{01}_F are fixed (non-discretionary), we can simply measure the overall gains from integration as

$$E^J_V = \text{Min}\ \{E \in R_0 \mid (Ey^{01}_V, y^{01}_F, y^{23}) \in T\}$$

Proceeding like this, we can also decompose the overall efficiency with non-discretionary resources.

In the vertical case, there is one more obstacle. It may be that only some of the intermediate resources are discretionary. In that case, the construction of the reference technology must reflect this. Assume that the discretionary parts of y^{12} are y^{12}_V while the rest y^{01}_F are non-discretionary. We can then model the relevant technology for the integration of P1 and P2 as

$$\begin{aligned} T = \{(y'^{01}, y'^{23}) \mid \exists\ y'^{12} :\ & \\ & y'^{12} \in P(y'^{01}), \\ & y'^{23} \in P(y'^{12}), \\ & y'^{12}_F = y^{12}_F \} \end{aligned}$$

and next proceed as above with the calculation of E^J_V.

4.2 Transferability

So far we have dealt with the possibility that only some of the variables are actually discretionary within a given time horizon. Another obstacle to the reallocation among DMUs may, however, be *non-transferable (local, L)* resources and services as opposed to *transferable (global, G)* ones.

To illustrate this consider Figure 14 below. We have two service providers, and to make the interpretation easiest, they produce the same globally transferable outputs. Also, we shall assume that they use the same technology. Now, if the nurses are transferable and the doctors are not, we could move some nurses from B, where that have a rather low marginal value compared to the doctors, to A, where their marginal value is higher. In the end locations, B is going to have its outputs reduced (from say y^B to

y^A) but A is going to have its output increased from y^A to some level $y^* > y^B$. The net result is therefore positive even though we cannot reallocate the factors as easily as in the usual harmony calculations.

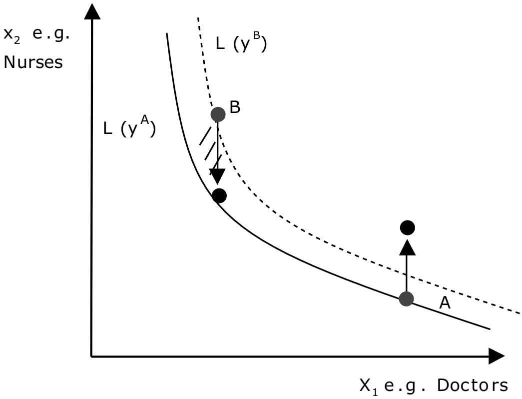

Figure 14: Restricted transferability

Allowing for possibly restricted transferability together with possibly restricted controllability, we get a 2 x 2 taxonomy of all variables: They may be LF (local and fixed, e.g. buildings), LV (local and variable, e.g. cleaning personnel), GF (global and fixed such as perhaps specialized measurement equipment), and GV (global and variable such as for example different types of specialists). To simplify notation, we can in the usual way indicate vectors of such variables by suppressing the sub-scripts that we include. We will for example refer to the L variables as the combination of the LF and LV variables, and to the F factors as the combination LF and GF factors.

In such a setting, it is more demanding to measure the possible gains from a reallocation of resources and services among the otherwise autonomous units in J.

A harmony measure H in the horizontal case could in this case be calculated in the following way. We consider what can be saved – after individual learning - by mere reallocations of the global inputs and outputs among the units in J. Let the new input and output combinations after reallocation be ($x^{\#j}$, $y^{\#j}$) for $j \in J$. We shall then look for such new productions plans for all the units involved such that

$$
\begin{array}{lcll}
H[\sum_{j\in J} E^j_V x^j_V] & \geq & \sum_{j\in J} x^{\#j}_V & \text{: we reduce use of variable factors} \\
x^{\#j}_L & \leq & x^j_L & \text{: local factors must be saved locally} \\
\sum_{j\in J} x^{\#j}_{GF} & \leq & \sum_{j\in J} x^j_{GF} & \text{: global fixed factors not reduced} \\
y^{\#j}_L & \geq & y^j_L & \text{: local service produced on-site} \\
\sum_{j\in J} y^{\#j}_G & \geq & \sum_{j\in J} y^j_G & \text{: global serv. can be prod. off-site} \\
(x^{\#j}, y^{\#j}) \in T \text{ all } j\in J & & & \text{: all plans technically feasible}
\end{array}
$$

***Harmony with limited transferability**
If we measure the savings potential in the variable inputs and stick to proportional improvements, the savings from pure reallocations can be determined by solving the following *pure reallocation problem*:

$$
\begin{array}{ll}
\underset{H,\ (x^{\#j},\ y^{\#j}),\ j\in J}{\text{Min}} & H \\
\text{s.t} & H[\sum_{j\in J} E^j_V x^j_V] \geq \sum_{j\in J} x^{\#j}_V, \\
 & x^{\#j}_L \leq x^j_L \\
 & \sum_{j\in J} x^{\#j}_{GF} \leq \sum_{j\in J} x^j_{GF} \\
 & y^{\#j}_L \geq y^j_L \\
 & \sum_{j\in J} y^{\#j}_G \geq \sum_{j\in J} y^j_G \\
 & (x^{\#j}, y^{\#j}) \in T \text{ all } j\in J \\
 & H\in R_0
\end{array}
$$

The choice variables in this program are the contraction factor H and the new input and output combinations $(x^{\#j}, y^{\#j})$ for $j\in J$. Since the original adjusted productions $(E^j_V x^j_V, x_F, y^j_L, y^j_G)$ $j\in J$ satisfy all the constraints, we always have $H \leq 1$ corresponding to a potential saving.

The constraints in this program ensure i) that we measure proportional reductions of all variable inputs, ii) that the local consumption of local inputs is not increased, and iii) that the global consumption of the transferable fixed inputs does not increase. The output constraints similarly ensure iv) that the local services do not fall and v) that the total production of globally transferable services does not suffer. Finally, we require vi) that the new production plans are feasible on an individual level.

When all resources and services are being global and variable, the above program leads to the original harmony measure. This is a simple application of Proposition 2 in Bogetoft and Wang(2005) as discussed above.

5. Post-merger efficiency

To evaluate a merger, it is important which practice is adopted after the merger. This will depend on a series of factor internal to the organization(s), including the possibility to learn better practices, to introduce better incentive systems and to bring in better managerial skills etc. In addition it may depend on external factors like the post-merger competition among similar entities.

A potential drawback of most of the literature as well the models presented above is that they assume the reallocations to take place at the frontier. This means that all units are assumed to adapt to the best practice before reallocation occurs. Although competition may work to drive out inefficient firms, it is naive to assume technical efficiency up front. Empirical studies have shown that inefficiency is a persistent phenomenon in many industries, and even theoretically one can find economic rationales for maintaining some technical inefficiency. It may, for example, help to compensate employees by making their job more attractive, or it may improve the strategic interactions with other firms on the market by showing excess capacity (see Bogetoft and Hougaard, 2003). Alternatively, the idea of reallocations on the frontier presumes that all entities are profit maximizing which certainly is not the case in the health sector, at least in the short run.

We will now develop some alternative approaches that do not presume technical efficiency. That is, we assume that if units have been inefficient in the past, there are reasons to believe that they will continue to be so in the future. Also, the level of future inefficiency may depend on the extent to which the units have to change the scale and scope of their operations.

A theoretical contribution is Bogetoft, Färe and Obel (2006). The focus in this paper is on allocative efficiency for a technical inefficient DMU. We assume that ex ante and ex post efficiency are the same and provide necessary and sufficient conditions for the inefficiency to be non-consequential for the evaluation of gains from reallocations. It basically requires a ray-homothetic technology. Intuitively, this means that the input iso-quants for different levels of output must be "parallel" such the rate of substitution is not affected by the output level. The approach also assumes the existence of a perfect market for inputs.

The idea of a perfect market for inputs are dispensed with in Andersen and Bogetoft(2007) and Bogetoft, Boye, Neergaard-Petersen and

Nielsen(2007), where we also allow for off-frontier adjustments. The reallocations considered there however is among all entities and no decompositions are derived. We now extend this approach to our setting.

5.1 Set-up

To formalize, and save notation, we ignore again possible lack of controllability or possible location specificity of certain inputs and outputs. We denote the original inputs and outputs x and y. A reallocation will lead to potentially different values hereof. We shall denote these $x^{\#}$ and $y^{\#}$.

Also, as before, let the *ex ante efficiency* of DMU^i be

$$E^i = \text{minimal input ex ante / actual input exante}$$

More precisely, we may define E^i as the Farrell input efficiency,
$E^i = \text{Min } \{E \in R_0 \mid (Ex^i, y^i) \in T\})$

The *ex post efficiency* may depend on whether any reallocation of resources and tasks takes place or whether the resources and tasks are more or less as before the reorganization

$$E''^i = \text{input efficiency ex post with no re-allocation}$$
$$E^{\#i} = \text{input efficiency ex post when reallocation takes place}$$

For a group of DMUs, we use similar notation with J substituted for j.

The size of E''^i compared to $E^{\#i}$ may depend on several factors. One argument could be that when a DMU changes its size and scope, it has to learn new procedures that may be costly at least in the short run. Also, sticking to given procedures that can be refined over time could potentially lead to some improvements in technical efficiency. This would suggest that

$$E^{\#i}_V < E^i_V < E''^i_V.$$

Another line of reasoning could be that a reallocation is a change event that could allow a DMU to reduce historical inefficiency. This would suggest that

$$E^i_V < E''^i_V < E^{\#i}_V.$$

5.2 Alternative hypothesis

There may be several reasons why all or more of the underlying units are not technically fully efficient to begin with. There could be structural reasons, exogenously given conditions and the like, but it could also be lack of management talent. In particular, where firms are quite alike and face similar exogenous conditions, it is tempting and seems reasonable to interpret large variations in technical efficiency as caused by underlying variations in managerial talent in the different firms.

An interesting question is therefore, which management talent the new merged unit may acquire? Of course, the new management may come from outside the units in the merged unit J, but in many cases the management of the new merged unit is usually 'inherited' from the underlying j units. In that case, one could wish to evaluate – also pre-merger – what effects a management change may have on potentials for efficiency gain. Following that line of thought, the next obvious question is: Which management skills will be inherited to the new unit? Again, several answers may be possible: One could hope for the merged unit to install the best performing management team of the underlying j units. Perhaps more realistically the new management team will be some sort of mixture of the underlying units' management teams. If the merger occurs in an environment where, e.g. votes or other kinds of size-related power matters, one could also assume that the new management team will come from the largest unit in the merger.

In the following, we formalize a generalized approach for weighting together the efficiencies of the underlying units into a measure replacing the E^j's. We then give examples corresponding to three different and relevant assumptions,

I *The spread of mediocrity:* the management effect is a mix of the underlying units' and therefore it will only be able to make them reach a joint average level of efficiency,

II *The best man on the prairie wins*: the units are able to reach the efficiency level of the best performing unit, corresponding to the best management team taking over the new unit, and finally

III *Size matters:* the new management team will come from the largest of the underlying units, and thus the skills of this team will define the potentials from a shared management in the merger. The latter assumption may lead to a merged unit, which is producing technically less efficient than the underlying units would have done if not merged. This happens whenever the largest unit is not the most efficient one.

5.2.1 *Formalizing the three hypothesis

For the merged unit J, we define the following weighted average of efficiency scores

$$E^{\#J} = \sum_{j\in J} w^j E^j$$

where E^j is as above and $w^j \in [0,1]$ with $\sum_{j\in J} w^j = 1$ are weights assigned to and potentially varying across each of the j units in J. The exact set of w^j will vary with the assumption made on post-merger management. We apply this measure for all j units in J and evaluate against the underlying technology as previously:

$$E^{Jm} = \text{Min}\ \{E \in R_0 \mid (E[\sum_{j\in J} E^j x^j / E^{\#J}], \sum_{j\in J} y^j) \in T\}$$

That is, we essentially project all the j units in J to the frontier implied by $E^{\#J}$. We hereby assume that they will waste a fraction (1- $E^{\#J}$) of its resources before they are put to effective use – or that it requires $x^j / E^{\#J}$ to be able to really use x^j.

The resulting measure E^{Jm} can be seen as a measure of the *management adjusted* overall efficiency of the merged unit. In comparison with the old measure, we have $E^{*J}=E^{Jm} / E^{\#J}$.

We define the potential gain (or loss) from the *management effect* as $E^J/E^{Jm} = M^J$. We see that if $M^J > 1$ the expected effect of the management resulting from the merger is in fact a loss, and if $M^J < 1$, a direct efficiency gain may result from the management effect of the merger.

The total potential savings from adopting best practices were $TE^J = E^J / E^{*J}$. If we assume a post-merger management effect as captured in M^J, we can calculate the potential left behind by the new management as $TE^{Jm} = TE^J / M^J$, i.e. this becomes a measure of the *technical efficiency potential post management-effects*. Inserting the definitions of TE^J and M^J and rewriting, we get $TE^{Jm} = E^{\#J}$. This makes sense since when the new management is only able to bring the organization to an efficiency level of $E^{\#J}$, then it basically leaves behind a savings potential (1-$E^{\#J}$).

We then have the decomposition

$$E^J = TE^{Jm} * M^J * E^{*J} = TE^{Jm} * M^J * H^J * S^J = E^{\#J} * M^J * H^J * S^J$$

where H^J and S^J are exactly as above. The interpretation is that the total savings potential can be split in a post-merger management effect M^J, a harmony and a size effect. In addition, there is some unused potential, T^{Jm} which capture what the new management team does not activate.

Next we describe three possible hypotheses concerning the way management skills are inherited from the underlying j units of J, all which can be fitted into the above measure.

5.2.1.1 *The spread of mediocrity

If the merged unit inherits a management team, company culture etc., which is a mixture of the j underlying units, we may assign the set of weights w^j, which best reflects the expected mixture. For example, the mixture may reflect the different sizes of the j units involved in merger J. The implication is that the management effect of the merger will make the new unit operate at a lower efficiency level than the best performing ones of the underlying j units. The overall effect may, however, still be a gain if the resulting efficiency is higher for a sufficiently large volume of the underlying units. If a gain is made will depend on the exact form of the mixture expected.

5.2.1.2 *The best man on the prairie wins

A more positive assumption is that the new merged unit will inherit the best performing management team of the underlying j units.

$$E^{\#J} = \text{Max } \{E^j, j \in J\}$$

This will in fact often be one of the immediate gains expected from mergers, where more efficient firms take over less efficient, but otherwise rather similar ones. Hence, they will be able to benefit from the best internal practice and hence reach in all subunits j the level of technical efficiency obtained by the best performing unit.

5.2.1.3 *Size matters

When there are important variations in size of the merging units, and size is a source of power for formal or subtler reasons, one may expect that the management teams of the larger units will eventually be those setting the standard of the new merged unit. To represent this form of management inheritance, we set w^j= 1 for the largest unit only and 0 for all others, where ‘large’ is a measure based on the variables expected to be decisive.

5.3 A more general approach

If we acknowledge that the entities in a horizontal merger may have different efficiencies ex ante, we could also recalculate the likely gains from reallocating resources and tasks without getting into a full-scale merger. That is we could estimate the harmony effect taking into account ex ante differences in technical efficiency, to be denoted $H^{\#J}$. We provide the technical details for this below.

The gains in this case would stem from at least two effects.

One is the mix or scope effect as before. In convex technologies it generally pays that the individual entities are operating closer to the average, cf. above. The other is an efficiency related effect. Since the units no longer are assumed equally efficient, it may pay to move resources from the less to the more efficient units, i.e. from units with low $E^{\#j}$ to units with high $E^{\#j}$.

A second effect may be related to economies of scale. If the technology is not a constant return to scale technology, it will in general be attractive to move individual DMUs closer to so-called most productive scale size. This suggests that large units operating above optimal scale size should allocate some of their resources and services to units operating below optimal scale size. This is a delicate operation however; it interacts with the scope effect since optimal scale size is usually dependent on the mix of inputs and outputs.

One can claim therefore that the harmony effect would compound a scope and size effect. On the other hand, the new measure would captures what can be accomplished by direct reallocations without a genuine merger and from a regulatory perspective this may be precisely what is needed – namely a measure of what can be gained if better allocation of resources and tasks are introduced among the same set of (competing) entities. These are gains that to a large extent can be captured without having to compromise on the competitive effects of a merger. Indeed, pursuing this perspective, one can argue that the deeper rationale for the harmony effect in the original decomposition was precisely the equivalence with the pure reallocation savings.

Taking this approach, we can decompose the overall potential as follows

$$E^J = (E^J / E^{\#J})\, H^{\#J} (E^{\#J} / H^{\#J})$$

where the first term is a measure of the potential left unused by the merged unit being less that fully efficient, the second terms is the potential from pure reallocations, and the last term is the potential from operating at full scale.

***Formalizing the more general approach**
If the units in J merge and their ex post efficiency is $E^{\#J}$ we have

$$(E^{\#J} [\sum_{j\in J} x^j], \sum_{j\in J} y^j) \in T$$

A measure of the overall gains taking into account possible inefficiency ex post is therefore $E^{\#J}$.

To eliminate what could have been gained from individual improvements we note that individually they would have been able to move from $(x^j, y^j) \in T$ to $(E''^j x^j, y^j) \in T$. In total, therefore, they would have produced

$$([\sum_{j\in J} E''^j x^j], \sum_{j\in J} y^j).$$

By comparing $E^{\#J}[\sum_{j\in J} x^j]$ and $[\sum_{j\in J} E''^j x^j]$, we can therefore get a measure of what extra – if anything – the merger could contribute beyond individual learning

$$E^{*\#J} = \text{Min } \{E \in R_0 \mid E[\sum_{j\in J} E''^j x^j] \geq E^{\#J}[\sum_{j\in J} x^j]\}$$

To measure what could be accomplished by pure reallocations, we may reason as follows. Assuming that we were to pick new inputs and outputs $(x^{\#j}, y^{\#j})$ for each $j\in J$ such that total outputs stay feasible, $\sum_{j\in J} y^{\#j} \geq \sum_{j\in J} y^j$, and such that all of the new productions are possible with the given ex post efficiencies, $(E^{\#j}x^{\#j}, y^{\#j}) \in T$. The largest potential saving in $\sum_{j\in J} x^j$ is then a measure of the reallocation or harmony effect

$$\begin{aligned} H^{\#J} = \quad & \min_{H, (x^{\#j}, y^{\#j}), j\in J} H \\ \text{s.t} \quad & H[\sum_{j\in J} x^j] \geq \sum_{j\in J} x^{\#j} \\ & \sum_{j\in J} y^j \leq \sum_{j\in J} y^{\#j} \\ & (E^{\#j}x^{\#j}, y^{\#j}) \in T \text{ all } j\in J \\ & H \in R_0 \end{aligned}$$

5.4 Post-merger efficiency in the vertical case

To refine the evaluation of a vertical integration, it is important – like in the horizontal case – to consider which practice is adopted after a more or less involved integration. The factors affecting the post merger performance of the processes and the ways to model them are analogous to the situation in the horizontal case. We shall therefore simply illustrate how to proceed and then discuss one new approach specific to the vertical or network case.

Consider the simple vertical case. The integrated entity has so far been assumed to use the technology

$$T = \{(y'^{01}, y'^{23}) \mid \exists\, y'^{12} : y'^{12} \in P^1(y'^{01}),\ y'^{23} \in P^2(y'^{12})\}$$

And we have therefore determined the potential overall gains from vertical integration as

$$E^J = \text{Min}\ \{E \in R_0 \mid (Ey^{01}, y^{23}) \in T\}$$

If we assume instead that the post-integration will not lead to technical efficient operations, we could capture this by modelling the technology as

$$T^{\#} = \{(y'^{01}, y'^{23}) \mid (E^{\#J}y'^{01}, y'^{23}) \in T\}$$

where $E^{\#J}$ is a given measure of post-merger managerial efficiency. The idea of $T^{\#}$ is that if we spend resources y'^{01}, only a part of them, $E^{\#J}y'^{01}$, will effectively enter the production. With this approximation of the resulting technology, we can then calculate saving potentials compared to the pre-merger situation, i.e. the modified E^J, as well as the decompositions into learning, harmony and size effects. In the previous formula, we should simply substitute $T^{\#}$ for T.

In addition to such approaches that are parallel to the ones proposed in the horizontal case, a special situation may arise in the vertical case. We have so far assumed that the processes P^1 and P^2 after a merger operate more or less independently except for the interaction modelled via the intermediate product y^{12}. In reality, this may not be a realistic assumption since we may assume that the production processes following integration will have to share managerial resources. In that case, the two processes may compete for the attention of the manager and since his time is limited, this could lead to a reduction in the production possibility sets. In a DEA context, we could model this by presuming that the intensity variables in the two processes should sum to one – rather than presuming that they sum to one in each of the production units. The intensity variables would in this case proxy for managerial effort.

5.5 Choosing the model

The discussion in Chapter 4 and 5 illustrates that there are many possibilities to modify any merger program. One implication of this is that a good understanding of the specific merger and the nature of resources and services produced, including the transferability and controllability of the different resources, is important to get realistic estimates of the likely savings. This means that the general recommendation should be to model the resource and product flows quite detailed in a given situation and then simulate the likely gains using production models and saving measures like above. If it is very difficult to model the different factors or to make reasonable assumptions about for example transferability, one can always start by a rather relaxed formulation with full transferability and controllability and post merger adjustment of technical efficiency, i.e. with the original formulation, and then interpret the potential gains calculated in such a model as an upper bound on the potentials derived from a merger.

6. Rational inefficiency

Although the present project does not allow us to pursue this idea in any details, we should like to point to one more perspective that could lead to more refine predictions of the likely post-merger efficiency and in particular to a prediction of the allocation of post-merger inefficiency.

The starting point for such an approach would be to consider the nature of inefficiency. In this paper we have largely presumed that inefficiency is waste and serves no real purpose. This is in line with much of the literature.

Technical inefficiency is often interpreted as waste following the concept of X- efficiency by Leibenstien (1966, 1978). It means that too many inputs have been used to produce too few outputs. According to Leibenstein, X-efficiency is primarily caused by lack of motivation and lack of knowledge. If an inefficient DMU does not motivate its employees sufficiently to save inputs and expand outputs, performance may be improved by redesigning the incentive structures. If inefficiency is caused by lack of information, performance may be increased by improving the markets for knowledge, learning, etc.

Along similar lines, inefficiency may be related to sub-optimal decision procedures. According to work by Chris Argyris (cf. Leibenstein and Maital, 1994) the main source of technical inefficiency in organizations is ‘‘defensive behavior’’. Employees are often reluctant to admit that their decisions were wrong — even though they themselves are aware of it. Thus the problem is not lack of information on how to optimize performance, but rather lack of willingness to use this information in order to improve on procedures. According to the literature on organizational learning this problem lies in the structure of the organization.

On the other hand, technical inefficiency can also be interpreted as the result of model mis-specifications. Measured waste may simply reflect that not all inputs or outputs are accounted for, that heterogeneous inputs and outputs are pooled or that the assumed relationship between inputs and outputs is flawed. Taking this perspective, we should either refrain from making efficiency judgments or we should improve our modelling.

A third perspective is that inefficiency may somehow be the result of rational behavior. This is the perspective taken in Bogetoft and Hougaard(2003), and it is at least partially in line with Stigler (1976) who argues against Leibenstein's concept of X-efficiency since ''Leibenstein does not attempt to understand the allocation of 'inefficient' resources, and hence does not see the necessity for attributing his X-inefficiency to specific inputs''. Pursuing this idea, one would see inefficiency and in particular the allocation of inefficiency (slack) among different inputs as the result of a rational choice made by an organization. There may be many gains from inefficiency in real organizations.

Measured inefficiency can be part of the (fringe) compensation paid to stake holders. Inefficiency can also contribute to incentives. For example, the firm can pay its employees more than their opportunity costs in order to make them work efficiently out of fear of the harsh penalty associated with a dismissal for poor performance. It can create loyal employees and thereby reduce costly turnovers in the labor force. In the organizational literature, technical inefficiency is also recognized as a possibly useful resource. In the literature there has been numerous discussions of the use of organizational slack as a buffer for uncertainty, as a means of de-coupling activities and hereby diminishing the information flow and coordination among subunits and as a necessity for providing resources for innovation, cf. e.g., Galbraith (1974) and Stabler and Sydow (2002). So the general idea of the rational choice perspective is that there are gains from inefficiency and therefore costs of improving efficiency. The gains are derived from the ability to offer on-the-job complementary payment, to improve incentives, and to ease planning, coordination and innovation in an uncertain environment.

If one takes such a perspective, one would model the organization as one of maximizing both efficiency and slack. From observed production plans, we can make at least partial inference about the DMU's revealed preferences for slack and efficiency and in particular the preferences for one type or slack as opposed to another. Using this, one can the calibrate models that predict the actual slack generation process in an organization. In one interpretation, this could involve bargaining among the different employee groups at for example a hospital. If the calibrated bargaining model shows that the bargaining power of for example doctors is in general twice the bargaining power of nurses, and that doctors have half the bargaining power of share holders or managers, one can then use this information to predict the reaction to changes in the organization, to changes in the allocation of resources etc. We may predict for example that a budget cut will lead to output reductions that are half the size of the reductions in doctor inputs and that the number of nurses may be reduced about twice as much as the doctors.

7. References

Andersen, J. and P. Bogetoft, Gains from Quota Trade: Theoretical models and an application to the Danish Fishery, *European Review of Agricultural Economics*, 34 (1), 105-127, 2007.

Banker, R.D., Estimating Most Productive Scale Size Using Data Envelopment Analysis, *European Journal of Operational Research*, 17, 35-44, 1984.

Bogetoft, P., Incentive Efficient Production Frontiers: An Agency Perspective on DEA, *Management Science*, 40, 959 – 968, 1994.

Bogetoft, P., Incentives and Productivity Measurements, *International Journal of Production Economics*, 39, 67-81, 1995

Bogetoft, P., DEA-Based Yardstick Competition: The Optimality of Best Practice Regulation, *Annals of Operations Research*, 73, 277 - 298, 1997

Bogetoft, P., DEA and Activity Planning under Asymmetric Information, *Journal of Productivity Analysis*, 13, 7 – 48, 2000.

Bogetoft, P. , K. Boye, H. Neergaard-Petersen and K. Nielsen, Reallocating sugar beet contracts: can sugar production survive in Denmark. *European Review of Agricultural Economics*, 34(1), 1-20, 2007.

Bogetoft, P. , J.M. Bramsen and K. Nielsen, Balanced Benchmarking, *International Journal of Business Performance Management*, 8(4), 274-289, 2006.

Bogetoft, P., R. Färe and B.Obel, Allocative Efficiency of Technically Inefficient Production Units, *European Journal of Operational Research*, 168 (2), 450-462, 2006.

Bogetoft, P. R. Färe and S. Grosskopf, Thought about Network DEA, WP, OSU, 2008.

Bogetoft, P., and T.E. Grammeltvedt , Mergers in Norwegian Electricity Distribution: A Cost Saving Exercise? Working Paper, BI, Oslo, 2006.

Bogetoft, P. and J.L. Hougaard, Rational inefficiencies. *Journal of Productivity Analysis* 20: 243-271, 2003.

Bogetoft, P. and K. Nielsen, Internet Based Benchmarking, *Journal of Group Decision and Negotiation*, 14 (3), 195-215, 2005.

Bogetoft, P. and H.B.Olesen, Design of Production Contacts, CBS Press 2004.

Bogetoft, P, N.Strange and B.J.Thorsen, Efficiency and Merger Gains in The Danish Forestry Extension Service, *Forest Science*, 49 (4), pp. 585-595, 2003.

Bogetoft, P. and D. Wang, Estimating the Potential Gains from Mergers, *Journal of Productivity Analysis*, 23, pp. 145-171, 2005.

Brännlund, R., Y. Chung, R. Färe and S. Grosskopf, Emissions Trading and Profitability: The Swedish Pulp and Paper Industry, *Environmental and Resource Economics*, 12, pp. 345-356, 1998.

Charnes, A., W.W. Cooper, A.Y. Lewin and L.M. Seiford (eds) (1944). *Data Envelopment Analysis: Theory, Methodology, and Application*, Klüwer Academic Publishers, 1994.

Färe, R. and S.Grosskopf , Intertemporal Production Frontiers: With dynamic DEA, Kluwer Academic Publishers, Boston, 1996.

Färe, R. and S.Grosskopf, Network DEA, Socio-Economic Planning Sciences, 35-49, 2000.

Farrell, M.J., The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, 120, 253 – 281, 1957

Farrell, J. and C. Shapiro, Horizontal Mergers: An Equilibrium Analysis, *American Economic Review*, 80, 107 – 126, 1990.

Galbraith, J.R. (1974). ‘‘Organizational Design: An Information Processing View.’’ *Interfaces* 4, 28–36.
Leibenstien, H. (1966). ‘‘Allocative efficiency vs. ‘X-efficiency’.’’ *The American Economic Review* 56, 392– 415.

Leibenstein, H. (1978). ‘‘X-inefficiency Xists: Reply to an Xorcist.’’ *The American Economic Review* 68, 203–211.

Leibenstein, H. and S. Maital. (1994). ‘‘The Organizational Foundation of X-inefficiency: A Game Theoretic Interpretation of Argyris’ Model of Organizational Learning.’’ *Journal of Economic Behavior and Organization* 23, 251–268.

Perry, M.K. and R.H. Porter, Oligopoly and the Incentives for Horizontal Merger*, American Economic Review*, 75, 219 –227, 1985.

Ray, S.C., Data envelopment analysis; Theory and techniques for economics and operations research, *Cambrigde Unversity Press*, Cambridge, 2004.

Salant, S., S. Switzer and R. Reynolds, Losses From Horizontal Merger: The Effects of an Exogenous Change in Industry Structure on Cournot-Nash Equilibrium, *Quarterly Journal of Economics*, 98, 185-199, 1983.

Stabler, U. and J. Sydow. (2002). ‘‘Organizational Adaptive Capacity: A Structuration Perspective.’’ *Journal of Management Enquiry*, 11(4), 408–424.

Tirole, J., *The Theory of Industrial Organization*, MIT Press, 1988.

Van Bergeijk, P.A.G. and E. Kloosterhuis(eds), Modelling European Mergers: Theory, Competition Policy and Case Studies, Edward Elgar Publishing, USA, 2005.

Efficiency Gains from Mergers in the Healthcare Sector

Part C, Implementation

Contents[1]

1 The sections marked with a * are more technically formulated. However, these sections can be skipped without losing the general ideas.

Abstract

The three-part research paper “Efficiency gains from mergers in the healthcare sector” focuses on the positive effects of integrations and moves from purely economic theory over measurement methods to actual program implementations. Due to the large field covered, it is organized in three coherent parts (Part A: Economic theory, Part B: Modelling, Part C: Implementation).

This paper discusses the implementation of the models of horizontal and vertical integration developed in Part B. The paper describes different ways to estimate actual production technologies, including Data Envelopment Analyses (DEA) and Stochastic Frontier Analysis (SFA), and it shows how to calculate the measures of potential gains in such technologies using optimization techniques like linear programming and line search methods. In addition, this paper discusses the particular challenges of including quality indicators in technology models and how to estimate likely quality impacts of mergers in real cases. Lastly, the paper illustrates the estimation techniques and the calculation of the potential gains measures in the horizontal case using data from Dutch hospitals.

1. Introduction

1.1 Background

The liberalization of the Dutch healthcare sector has lead to a number of mergers between healthcare and related institutions. The mergers of major concern are between hospitals or between a hospital and an insurance company while the greatest number of mergers takes place between nursing homes and home care providers. Although, at present, the merging parties have only put forward unsubstantiated arguments claiming the benefits of their action, more and better efficiency and quality defences are to be expected in the future.

The a priori assessment of mergers and agreements requires careful consideration since the negative and positive effects stem from a number of different sources, and they have to be weighed against each other. An additional complication lies in the fact that the calculations and estimates concern to the future. We often have to compare two hypothetical situations: (1) what are the probable effects of the merger, (2) what would happen in the market without a merger? Quantifying the effects, especially those that impact quality, also presents a challenge.

A great deal of literature exists about horizontal mergers and their evaluation methods, while some questions remain about vertical mergers. This latter is more complicated to evaluate, because the participating parties operate in different markets, and consequently more circumstances and effects need to be taken into account. Generally, vertical mergers offer a wider range of possible efficiency gains, and, therefore a higher probability that they outweigh the adverse, anticompetitive effects. So far, less attention has been devoted to these beneficial consequences. Their evaluation and quantification are not as well-developed as the estimation of anticompetitive effects.

The three-part research paper “Efficiency gains from mergers in the healthcare sector” focuses on the positive effects of integrations and moves from purely economic theory over measurement methods to actual program implementations. It considers both horizontal and vertical mergers and aims to contribute to the literature of merger evaluation. It offers new findings in the measurement of efficiency gains of vertical mergers and the quantification of the effects of those mergers on quality. This is the third part (Part C) in the three-part series, which describes the implementation and application of the models developed in Part B.

1.2 Research questions and objectives

The main research question of the project Efficiency gains from mergers in the healthcare sector (and therefore the focus of the research paper series) is 1) *How can we measure the efficiency gains that organizations achieve by merging?* By answering this question, it becomes possible to obtain objective data about positive effects of integrations. After translating these organization-level effects into changes in the consumer welfare, we can weigh them against the negative consequences of the merger. In this way, the result and products of this project, i.e. the developed model, serves as input to merger assessment process. The series of research paper contributes to the further improvement of the decision making process by authorities.

The main research questions of this particular paper of the series are *"How can we quantify the technologies using actual data of past productions?"* and *"How can we calculate the potential gains in actually estimated technologies using for example optimization methods?"* In this way, the paper is devoted to the actual implementation of the ideas developed in Part B. This means that we shall discuss as a separate issues *"How can we include quality in the models and predict the quality impact if mergers?"*

1.3 Scope of research

The series of research paper intends to offer a comprehensive view on positive effects of mergers at the organization level. It embraces (A) the theoretical background of the potential efficiency sources, which we can find in the economic literature, (B) an overview and development of mathematical models that are appropriate for the evaluation of the effects, (C) and the demonstration of program applications and examples of data runs.

However, Part A, which deals with the economic background of potential gains, considers strategic advantages of mergers; the developed model (Part B and C) only discusses cost efficiencies and gains of better coordination between the two parties. The model considers the organizations as individual "decision-making units" and the modelled gains of the merger are efficiencies of scale and scope and positive effects of solving moral hazard problem. Potential changes in the competition environment and interactions between market agents are not considered.

The negative welfare effects of mergers and their measurement have been discussed at length in the literature, and as such also fall beyond the scope of this series of research papers.

1.4 Outline

The series of research papers is structured as follows. Part A gives a summary of economic theory relating to sources of efficiencies in mergers. Part B is devoted to model development, and describes the principal ideas and technical details of calculating the overall gains and the decomposition of these gains according their sources. Part C discusses how to implement these models and measures using on actual data to estimate the underlying technologies.

This paper is structured as follows. We first discuss state of the art frontier modelling in Chapter 2, and then discuss how to calculate the merger measures from Part B in relation to these models using optimization techniques in Chapter 3. The particular challenges of quality are discussed in Chapter 4 and some illustrative applications on actual data from Dutch hospitals are provided in Chapter 5. We conclude with some indications of relevant themes for future research in Chapter 6.

The paper can – as Part B - be read at two levels. One is a conceptual level, and another is a more technical level with a view towards actual measurements and implementation of the ideas into software. The conceptual idea can be explained with a minimum of mathematics while the implementation requires somewhat more details and precision and therefore makes it most efficient with more formalism and mathematics. Throughout the paper we try to make the general ideas clear using verbal explanations and simple graphics. These parts should therefore be accessible to most readers. The more technical parts are marked with a * in the section headings and can be skipped without losing the general ideas.

The focus on the conceptual ideas is of course relevant to readers with less time or background to study the details of the modelling. We suggest that it also very useful to more technically oriented readers. In many applications limited data is available and this may prohibit specific modelling and make estimations impossible. Experience shows however that the conceptual framework in such cases is useful to get a good idea of the forces and effects at work, and to structure the analyses.

2. Frontier models

The aim of this section is to give a brief introduction to applied production modelling and state-of-the-art benchmarking techniques.

2.1 Overview

Within the area of productivity and efficiency evaluation, there is by now not only a series of simple ad hoc approaches but also a portfolio of state-of-the-art techniques to estimate production and cost models. Depending on the assumption regarding the data generating process we divide the techniques in deterministic and stochastic, and depending on the functional form into parametric and non-parametric techniques cf.
below.

	Deterministic	Stochastic
Parametric	Corrected Ordinary Least Square (COLS) Greene (1990), Lovell (1993), Aigner and Chu (1968)	Stochastic Frontier Analysis (SFA) Aigner, Lovell and Schmidt (1977), Batesee and Coelli (1992), Coelli, Rao and Battesee (1998)
Non-Parametric	Data Envelopment Analysis (DEA) Charnes, Cooper and Rhodes (1978), Deprins, Simar and Tulkens (1984), Bogetoft(1996)	Stochastic Data Envelopment Analysis (SDEA) Land, Lovell and Thore (1993), Olesen and Petersen (1995)

Table 1: Model taxonomy

To get a first idea of the methods, it is convenient to think of the modelling of the costs of some services, e.g. the costs of hospital production. Corrected ordinary least square (COLS) corresponds to estimating an ordinary regression model and then making a parallel shift to make all units be above the minimal cost line. Stochastic Frontier Analysis (SFA) on the other hand recognizes that some of the variation will be noise and only shifts the line – in case of a linear mean structure – part of the way towards the COLS line. Data Envelopment Analysis (DEA) estimates the technology using the so-called minimal extrapolation principle. It finds the smallest production set (i.e. the set over the cost curve) containing data and satisfying a minimum of production economic regularities. Like COLS, it is located below all cost-output points, but the functional form is more flexible and the model therefore adapts closer to the data. Finally, Stochastic DEA (SDEA) combines the flexible structure with a realization, that some of the variations may be noise and only requires most of the points to be enveloped.

A fundamental difference from a general methodological perspective and from regulatory viewpoint is the relative importance of flexibility in the mean structure vs. precision in the noise separation. This means that there are basically two risks for error that cannot be overcome simultaneously. These are 1) risk of specification error, and 2) risk of data error as illustrated in Figure 1 below.

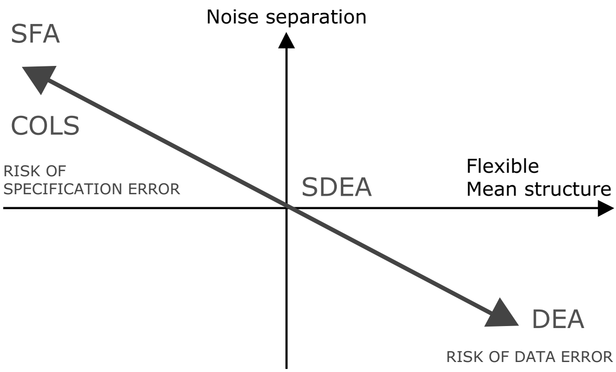

Figure 1: Basic trade-off in benchmarking analyses

2.1.1 Specification error

The inability of the model to reflect and respect the real characteristics of the industry is related to the specification error. Avoiding the risk of specification error requires a flexible model in the wide sense. This means that the shape of the model (or its mean structure to use statistical terms) is able to adapt to data instead of relying excessively on arbitrary assumptions. The non-parametric models are by nature superior in terms of flexibility.

2.1.2 Data error

The inability to cope with noisy data is called data error. A robust estimation method gives results that are not too sensitive to random variations in data. This is particularly important in yardstick regulation with individual targets – and less important in industry wide motivation and coordination studies. The stochastic models are particularly useful in this respect.

It is worthwhile to observe that the two properties may to some extent substitute each other. That is, the flexible structure allowed by non-parametric deterministic approaches like DEA may compensate for the fact that DEA does not allow for noise and therefore assigns any deviation from the estimated functional relationship to the inefficiency terms. Likewise, the explicit inclusion of noise or unexplained variation in the data in SFA may to some extent compensate for the fact that the structural relationships are fixed a priori, i.e. the noise terms may not only be interpreted as a data problem but also as a problem in picking the right structural relationship. As an illustration of this, it has been found that the SFA efficiencies are often larger than the DEA efficiencies as long as the model is somewhat ill-specified, i.e. the inputs and outputs are badly chosen. The reason is that SFA in this case assigns the variations to the noise term while DEA assigns everything to the efficiency term. As the model is extended to include more relevant inputs and outputs, the two methods have been found to produce quite comparable results.

2.2 Technology and distance functions

In the production economic and efficiency analyses literature – as in Part B of this project - we think of an organization as a transformation of multiple inputs (x) into multiple outputs (y). We describe the possible transformations as the technology

T = set of feasible input-output (x, y) combinations

In the estimation of the technology, we combine a priori assumptions about the technology with observations of actual organizations to determine a reasonable description of what can be accomplished in real organizations.

In production and cost theory, classical assumptions are

A1 *(strong) free disposability,* i.e. weakly more input can produce weakly more output. Or put differently, we can freely dispose inputs and outputs in the sense that if we can produce a given output with a given input, we can also produce less inputs with more outputs.

A2 *convexity,* i.e. weighted averages of feasible production plans are feasible as well. This is a technical assumption which is popular in economic theory, and it is a fundamental condition for the use of prices in general. It has applied motivations also, since one can sometime imagine that half of the production is produced using one set of procedures and the other half is produced using another set of procedures. Still, it is fair to say that it is not always an obvious assumption to make and several developments in the efficiency analyses literature is relaxing this assumption or even dispensing with it entirely.

A3 *return to scale*, i.e. if we change the input with a certain amount we can also change the output with a certain amount. If it is possible to scale a given production plan up, we talk about increasing return to scale (irs), or non decreasing return to scale (ndrs). If we only presume that we can scale down a production plan, we talk about decreasing return to scale (drs), or non increasing return to scale (nirs). The strongest assumption is that we can scale both up and down. This is referred to as constant return to scale (crs), and it is a strong assumption at least in the short run. It says that if we change inputs with a given percentage, we also change the maximal output with the same percentage. If we do not make assumptions about the possibilities to re-scale, we sometimes talk about this as an assumption of variable returns to scale (vrs). Again, it is possible to have variants of this assumption, e.g. limited constant return to scale where we only presume that we can scale a given production up and down with say 10%.

A final assumption with some intuitive appeal and a particular role in the analyses of mergers is that of replicability or

A4 additivity, i.e. if two production plans are feasible than their sum is feasible as well. The intuition is that we could always organize the joint production as two independent productions by imitating the original entities.

***Formalizing the technologies using sets and functions**

To generalize and to formalize, we consider entities or so-called Decision Making Units (DMUs), $i \in I = \{1, 2, \ldots, n\}$, that transforms p inputs (x^i) to q outputs (y^i).

The *technology* T is given by

$$T = \{(x, y) \in R_0^{p+q} \mid x \text{ can produce } y\}$$

Classical assumptions about T are

A1 *(strong) free disposability,* i.e. weakly more input can produce weakly more output

$$(x', y') \in T \text{ and } x'' \geq x' \text{ and } y'' \leq y' \Rightarrow (x'', y'') \in T$$

A2 *convexity,* i.e. weighted averages of feasible production plans are feasible as well

$$(x, y) \in T, (x', y') \in T, \lambda \in [0, 1] \Rightarrow \lambda(x, y) + (1 - \lambda)(x', y') \in T$$

A3 *s-return to scale*, i.e. it is possible to scale a given production plan

$$(x', y') \in T \Rightarrow k(x', y') \in T \text{ for } k \in K(s)$$

where s = "crs", "drs", "vrs", or "irs" corresponding to constant, decreasing, varying, or increasing return to scale, and where K(crs) = $\Re_0$, K(drs) = [0,1], K(vrs) = {1}, and K(irs) = [1, +∞), respectively.

A less common, but very relevant assumption here, is the replicability assumption A4 *additivity*, i.e. it is possible to add feasible production plans

$$(x', y') \in T \text{ and } (x'', y'') \in T \Rightarrow (x' + x'', y' + y'') \in T.$$

We shall usually describe the technology by the production possibility set. However, it is sometimes convenient to use one or more alternative formulations. The equivalence of the alternative formulations requires a minimum of regularity on the technology. We shall therefore generally follow Färe and Primont(1995) and assume a *regular technology* in the sense that that 1) T is closed, 2) Inputs and outputs are freely disposable (A1), 3) T is convex (A2) and 4) P(x) is bounded for each x.

Important alternative descriptions of the technology are by *cost functions* $C(y, w) = \min_x \{wx \mid (x, y) \in T\}$, i.e. by the minimal costs of producing any given output for any vector of factor prices, by *revenue functions* $R(x, p) = \max_y \{py \mid (x, y) \in T\}$, i.e. by the maximal revenue that can be generated by any factor combination and for any prices on the final outputs, or by *profit functions* $\Pi(w, p)= \max_{x,y} \{py - wx \mid (x, y) \in T\}$. Cost, revenue and profit functions are sometimes used in the evaluation of efficiency, e.g. by comparing actual costs with estimated minimal costs. The estimation of these functions does however presume data on prices as well as optimizing behavior on part of DMUs. Since neither of these assumptions may be fulfilled in many health applications, we shall not make too much use of them below.

Instead we shall sometimes use yet another characterization of the technology, namely in terms of distance functions. Using the so-called *Shepard* distance here, we have the *input distance function* defined as

$$D_I(x, y) = \max_\rho \{\rho \mid (x/\rho, y) \in T\}$$

i.e. as the maximal contraction of inputs that keeps the production plan feasible. For regular technologies, we have

$$T = \{(x, y) \in R_0^{p+q} \mid D_I(x, y) \geq 1\}$$

i.e. the distance function gives an equivalent delineation of the technology. The input distance function therefore also have a series of parallel properties, most notably $D_I(x, y)$ is 1) decreasing in y, and 2) non-decreasing, positively linearly homogeneous and concave in x. The positively linearly homogenous here means that

$$D_I(\rho x, y) = \rho D_I(x, y)$$

with the natural interpretation that when we double the input we double the distance to the frontier.

In a similar way, we can define an *output distance function* as

$$D_o(x, y) = \min_\rho \{\rho \mid (x, y/\rho) \in T\}$$

and get

$$T = \{(x, y) \in R_0^{p+q} \mid D_o(x, y) \leq 1\}.$$

Under the usual regularity conditions, output distance functions are 1) decreasing in x and 2) non-decreasing, positively linearly homogenous and convex in y.

Instead of using Shephard distance functions, we can of course use the by now more commonly used *Farrell* formulations with input efficiency E(x, y) = 1 / D_I(x, y) and output efficiency F(x, y) = 1 / D_o(x, y). In Part B we relied mainly on Farrell efficiencies, but we shall sometimes use Shepard distances in the implementations.

2.3 Parametric methods

In the parametric SFA approach, the separation of noise and inefficiency is technically done by assuming that the noise is two sided and inefficiency is one sided. Inefficiency makes costs increase and makes production fall short of the best possible, while noise may also lower the observed costs or increase the observed output. In addition to having one- and two sided deviations, the separation of noise and inefficiency is accomplished by making specific assumptions about the nature of the distributions, e.g. normal and half normal.

In the parametric approach, one also makes specific assumptions about the type of relationship between the inputs and outputs. The so-called functional form may for example be linear, log-linear or translog. We shall return to these assumptions below.

2.3.1 Average, deterministic and stochastic frontier

To be more specific, we may distinguish between three combinations of noise and inefficiency. Namely pure noise models, pure efficiency models and combined models. In a cost setting, we may assume that costs (x) depend on a series of output driver (y) as well as on a combination of the inefficiency term u ≥ 0 and the noise term v for each entity i

Pure noise (Ordinary least squares (OLS), average cost function):

$$x^i = C(y^i, \beta) + v^i$$

Pure inefficiency (Deterministic frontier):

$$x^i = C(y^i, \beta) + u^i$$

Combined (Stochastic frontiers):

$$x^i = C(y^i, \beta) + u^i + v^i$$

In the specifications above, C(y, β) is the minimal cost function. It defines the least expensive way to provide the outputs y. The functional form is given, except for some unknown parameter values β. The statistical analysis seeks to estimate the functional relationship, i.e. β, and to estimate the inefficiencies, i.e. u^i.

The first of these specifications (OLS) is the specification in classical statistics. It fits a function to the data in such a way that the positive and negative deviations are as small as possible. The standard measure of goodness-of-fit is the sum of squares of deviations, which is why this approach is often referred to simply as the OLS, ordinary least squares approach. Since the OLS approach does not work with the idea of individual inefficiencies, the usage of OLS in our context is less convenient.

The deterministic frontier approach considers everything as inefficiency. Still, as hinted at above, this approach is generally inferior to the others since it invokes both strong restrictions on the functional forms and ignores noise in the data.

The stochastic frontier approach was introduced independently by Aigner, Lovell and Smith (1977) and Meeusen and Van den Broeck (1977). This section provides a short introduction to the basic characteristics of SFA. A comprehensive introduction to SFA can be found for example in Coelli et al. (1998). Conceptually, it is attractive to allow for the realistic existence of both noise and inefficiency. The drawback of the approach is in turn that we need a priori to justify 1) the distribution of the inefficiency terms and 2) the functional form of the frontier.
The SFA specification is also attractive by allowing the use of classical statistical approaches like maximum likelihood estimation, likelihood ratio testing etc. In a SFA approach, we use the data to come up with a best estimate of the underlying costs function C. Compared to DEA, we have less freedom in our choices since we have to decide already at the outset about a possible classes of such functions. Given a best estimate of the cost function C, we can determine the noise plus inefficiency by comparing the actual cost and the cost function value.

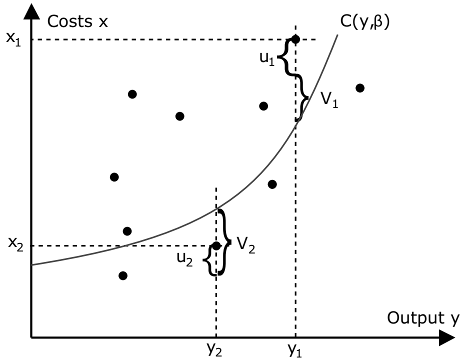

Figure 2: Stochastic cost frontier with noise v and inefficiency u

2.3.2 Distributional assumptions

SFA requires some a priori assumption about the distribution of the inefficiency term in order to separate noise and inefficiency. It is hard to give strong arguments for a specific form; it would generally require a careful modelling of the data generation process and the allocation of, for example, competences over the set of managers in an industry. It is therefore best to start with a rather general and flexible specification and to let the data reveal as closely as possible the correct distribution. A good choice here is to work with u^i as being truncated normal; $N_+(\mu, Ð_u^2)$, i.e. a normal distribution centred around μ and next truncated to be above zero.

2.3.3 Functional forms

The second fundamental problem in a parametric frontier approach is to select a functional form for the frontier.

The selection of functional form is guided by intuition and data as well as by theory. Experienced econometricians may for example be good at choosing functional forms with possibly data transformations and the sufficient degrees of freedom to provide a reasonable goodness of fit of the data at hand. In addition, theory guides the selection by imposing reasonable properties on the estimated function, e.g. that cost function is homogenous in prices or that output sets are convex.

A good general principle is to use the simplest possible representation with the sufficient flexibility to represent data. The simplest possible form is the *linear* form and as a good starting point – and even a starting point used in the iterative procedures used to estimate more advanced forms – it is therefore recommended to do a linear regression of cost on the different outputs. A slightly more complicated specification is the *log-linear* one being linear in the log of the variables, corresponding to a multiplicative relationship in the original variables, well-known from Cobb-Douglas type functions.

Linear specifications correspond to first order approximations and the natural next step towards a workable form is to use quadratic approximations, possibly in the log of the variables. A second order approximation using log variables gives the so-called *translog* form. In a cost function specification with n outputs and no prices (single input), it pictures the relationship as

$$\ln C_i = \beta_o + \sum_{h=1}^{q} \beta_h \ln y_h^i + \sum_{h=1}^{q}\sum_{k=1}^{q} \beta_{hk} \ln y_h^i \ln y_k^i + u^i + v^i$$

where C^i is the total cost of the i-th unit, y^i_h is the quantity of the h-th output of the i-th unit, and the βs are unknown parameters to be estimated.

2.3.4 *Parametric distance functions

To estimate general multiple inputs multiple output technologies as opposed to simple single input cost functions or single output production functions, we shall rely on the distance function characterization of the technology from above.

Thinking again in the translog as a second order approximation (in logs), we can follow Färe et.al. (1993) estimating a translog input distance function as

$$\ln D_I = \beta_o + \sum_{h=1}^{q} \beta_h \ln y_h + \frac{1}{2}\sum_{h=1}^{q}\sum_{k=1}^{q} \beta_{hk} \ln y_h \ln y_k + \sum_{h=1}^{p} \alpha_h \ln x_h + \frac{1}{2}\sum_{h=1}^{p}\sum_{k=1}^{p} \alpha_{hk} \ln x_h \ln x_k + \sum_{h=1}^{p}\sum_{k=1}^{q} \delta_{hk} \ln x_h \ln y_k$$

where D_I^i in the input distance of DMU^i. Of course, we need to put restrictions on the approximation to ensure that it gets the usual properties of an input distance function. Most notably, linear homogeneity in the inputs requires

$$\sum_{h=1}^{p} \alpha_h = 1$$
$$\sum_{k=1}^{p} \alpha_{hk} = 0 \quad h = 1, \ldots, p$$
$$\sum_{h=1}^{p} \delta_{hk} = 0 \quad k = 1, \ldots, q$$

Also, symmetry requires

$$\alpha_{hk} = \alpha_{kh} \quad h, k = 1, \ldots, p$$
$$\beta_{hk} = \beta_{kh} \quad h, k = 1, \ldots, p$$

Further restrictions can be imposed to ensure convexity in x and possible return to scale properties.

To estimate a parametric distance function taking into account both noise and inefficiency, one may make use of the linear homogeneity. Thus, if we let $D_I(x, y) = \exp(u) \geq 1$, we have that

xhDI(x/xh, y) = DI(x, y) = exp(u)

where x_h is one of the input dimensions. We therefore get

xh = (DI(x/xh, y))-1 exp(u)

and therefore taking logarithms, we have

-ln(xh) = ln(DI(x/xh, y)) – u + v

where we have added a symmetric noise term v. This functional form can be estimated as other stochastic frontier functions.

Of course, similar approach is possible on the output side. The function form of D_O in a translog approximation is like the D_I above, but since D_O is homogenous of degree 1 in outputs, we would have the following restrictions on the parameters

$$\sum_{h=1}^{q} \beta_h = 1$$

$$\sum_{k=1}^{q} \beta_{hk} = 0 \quad h = 1,\dots,q$$

$$\sum_{k=1}^{q} \delta_{hk} = 0 \quad h = 1,\dots,p$$

in addition to the same symmetry restrictions as above.
Letting $D_O(x, y) = \exp(-u) \leq 1$ we have that

$$y_h D_O(x, y/y_h) = D_O(x, y) = \exp(-u)$$

where y_h is one of the output dimensions. We therefore get

$$y_h = (D_O(x, y/y_h))^{-1} \exp(-u)$$

and therefore taking logarithms, we have

$$-\ln(y_h) = \ln(D_O(x, y/y_h)) + u + v$$

where we have added a symmetric noise term v.

2.4 Non-parametric methods

Since it was first proposed by Charnes, Cooper and Rhodes (1978,79), Data Envelopment Analyses (DEA) has become a tremendously popular relative performance evaluation tool.

Currently, DEA is also used as the basis for regulation regimes in different areas. In particular, it has been used in incentive regulation of private, semi-private and public utilities. In the regulation of utilities, for example, several countries have introduced DEA based revenue and price cap systems.

The great popularity of DEA among researchers and theorist alike also means that the basics of DEA modelling is described in several text books, cf. e.g. Charnes, Cooper, Lewin and Seiford (1994), Coelli, Rao and Battese (1998) or Cooper Seiford and Tone (2000). We shall therefore not give much of an introduction to DEA. It suffices to introduce a bit of notation and to recall a few basic definitions.

2.4.1 Estimation principle

DEA does not estimate the technology using traditional statistical principles like maximum likelihood or least square estimation. Rather, the underling idea of DEA is one of minimal extrapolation. The estimate of T, the empirical reference technology T*, is constructed according to the *minimal extrapolation* principle: T* is the smallest subset of R_0^{p+q} that contains the actual production plans (x^i, y^i), $i \in I$, and satisfies certain technological assumptions specific to the given approach.

2.4.2 Important DEA technologies

Different models in the DEA family are therefore distinguished by the basic technological assumptions made. The original constant returns to scale (crs) DEA model proposed by Charnes, Cooper and Rhodes (1978,79) assumes A1, A2 and A3(crs) while the decreasing returns to scale (drs) and (local) variable returns to scale (vrs) models developed by Banker (1984) and Banker, Charnes and Cooper (1984) appeal to A1, A2 and A3(drs) and A1, A2 and A3(vrs), respectively.

The classes of DEA technologies are illustrated in the single input single output case in Figure 3 below. We note that the simplicity of these models is exaggerated in the single input single output illustrations. If we for example invoke crs, we do not assume that the input or output sets are linear as one assumes in for example simple linear regression models.

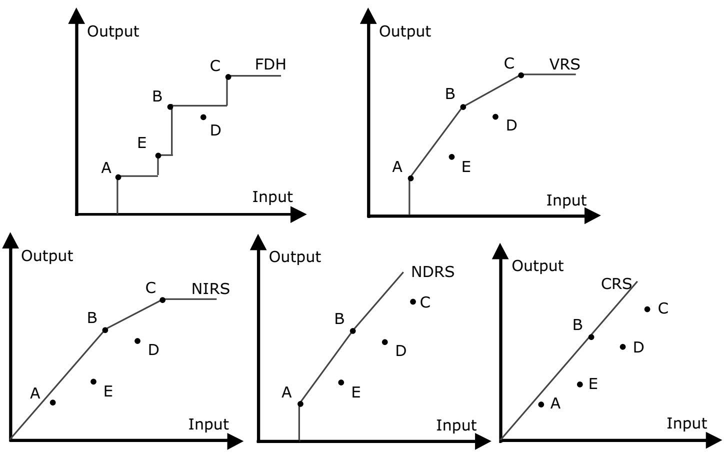

Figure 3: Classical DEA models in single input single output case

2.4.3 *Formalizing the DEA technologies

It is easy to see, cf. e.g. the references above, that A1, A2 and A3(s) together with the idea of minimal extrapolation, lead to the empirical reference technology

$$T^*(s) = \{(x, y) \in R_0^{p+q} \mid \exists\, \lambda \in R_0^n : x \geq \sum_i \lambda^i x^i,\ y \leq \sum_i \lambda^i y^i,\ \lambda \in \Lambda(s)\}$$

where $\Lambda(crs) = R_0^n$, $\Lambda(drs) = \{\lambda \in R_0^n \mid \sum_i \lambda^i \leq 1\}$, $\Lambda(irs) = \{\lambda \in R_0^n \mid \sum_i \lambda^i \geq 1\}$, and $\Lambda(vrs) = \{\lambda \in R_0^n \mid \sum_i \lambda^i = 1\}$.

The assumptions A1 - A3 have been relaxed in the <u>f</u>ree <u>d</u>isposability <u>h</u>ull (fdh) model used by Deprins, Simar and Tulkens (1984), and the <u>f</u>ree <u>r</u>eplicability <u>h</u>ull (frh) model briefly proposed in Tulkens (1993). The fdh model invokes only A1 and T*(fdh) and therefore has the structure above

with $\Lambda(fdh) = \{\lambda \in R_0^n \mid \sum_i \lambda^i = 1,\ \lambda^i \in \{0, 1\}\ \forall i\}$. The frh model presumes A1 and the additivity assumption A4 such that T*(frh) has the structure above

with $\Lambda(frh) = \{\lambda \in R_0^n \mid \lambda^i \text{ integer } \forall i\}$. Technically, the fdh ad frh hull models lead to mixed integer programs. Luckily, however, the fdh model can easily be solved by simple trying all possible pair wise comparisons. DEA models which partially relax the convexity assumptions are suggested in Bogetoft (1996) and Petersen (1990).

2.4.4 Efficiency measurement in DEA

Given an estimated DEA technology T*, the (in)efficiency of a given DMU, DMU^i, reflects the possibility to reduce inputs and expand outputs. In the DEA literature, this is most commonly done using the inverse Shepard distance measure, i.e. the Farrell inefficiency with the interpretation

$$E = \text{minimal inputs / actual inputs}$$

i.e. as the maximal contraction of (all) inputs that is possible when we still want to produce the same outputs.

One of the advantages of DEA is the relative ease of determining the efficiency of a given unit. It reduces to solving a simple LP problem as we shall illustrate now. The estimation of the technology and the measurement of distance relative to the technology is done in one single step.

2.4.5 *Efficient measurement using linear programming

The Farrell measures are by definition given

$$E^i = \text{Min}\{E \in \Re_0 \mid (Ex^i, y^i) \in T^*\} \quad \text{or} \quad F^i = \text{Max}\{F \in \Re_0 \mid (x^i, Fy^i) \in T^*\}$$

where E^i is the maximal contraction of all inputs, and F^i is the maximal expansion of all outputs that are feasible in T*. Since we can approximate T* like above, the calculation of E and F reduces to solving simple Linear Programming problems. In the case of Farrell input efficiency we get

$$\begin{aligned} &\underset{E,\lambda}{\text{Min}} && E \\ &\text{s.t.} && Ex^i \geq \textstyle\sum_{j\in I} \lambda^j x^j \\ &&& y^i \leq \textstyle\sum_{j\in I} \lambda^j y^j \\ &&& \lambda \in \Lambda(k) \end{aligned}$$

And in case of Farrell output efficiency we get

$$\begin{aligned} &\underset{F,\lambda}{\text{Max}} && F \\ &\text{s.t.} && x^i \geq \textstyle\sum_{j\in I} \lambda^j x^j \\ &&& Fy^i \leq \textstyle\sum_{j\in I} \lambda^j y^j \\ &&& \lambda \in \Lambda(k) \end{aligned}$$

In this way, efficiency evaluations in a DEA framework are particularly simple.

2.4.6 Courteous (conservative estimates) and bootstrapping

We note that DEA, by the minimal extrapolation principle and assuming no noise in data, provides an inner approximation of the underlying production possibility set

$$T^* \subseteq T$$

The efficiency estimates are therefore optimistic and the potential input savings and output expansions are underestimated. The extent of this bias can be estimated using bootstrapping, cf. e.g. Simar and Wilson (2000). The bias corrected technology T** is larger than, i.e. contains T*,

$$T^* \subseteq T^{**}$$

and it does provide a closer approximation of T.

In regulatory applications, there is no consensus about the use of the T* or T** as the basis for decisions.

An advantage of using T** is its closer approximation of T and hereby the less rents it tends to leave to inefficient entities. Another advantage is that T** will tend to give a more even approximation of T. In contrast, T* may give a close approximation for units that resembles several other units and a very poor approximation for more special units. In other words, if we do not correct for bias, we will tend to make the standards tougher for some entities than for other.

A disadvantage of the bias corrected technology is that the standards become somewhat harder to explain since we cannot point simple to a group of actual peer units. Another disadvantage is that the bias corrected technology may not be contained in the true technology. That is, we may end up with a too optimistic model of the improvement possibilities. In the case of input efficiency of a given unit, we may possibly get $E^* > E > E^{**}$ where E^*, E and E^{**} is input efficiency of a given unit against the T*, T and T** technologies, respectively. Of course, one can argue that since any data set contains noise, we may run into situations where $E>E^*$ such that even the uncorrected estimates of efficiency are too pessimistic, and the possible regulatory rulings based on T* may be too harsh against the entity in question. If this is the case, it means that even the minimal extrapolation of DEA does not provide enough protection. Note however that this is not an argument in favour of bias corrections; these corrections will simply expand this problem since $E^* \geq E^{**}$ making the unfair evaluation even less fair in this case.

3. Merger gains in DEA and SFA models

The aim of this chapter is to describe how the measures of merger gains developed in Part B can be combined with the actually estimated technologies from the previous chapter. This means that we shall describe how to calculate the measures in actual technologies modeled using non-parametric DEA or parametric SFA models. As we will see, the practical implementation requires application of single and multiple dimensional optimization techniques.

The simple way to look at our approach is as a two-stage approach where we first estimate the technology and next measure the gains from reorganizations relative to the estimated technology. When we estimate the technology using SFA approaches, we do indeed proceed in these two stages. In the case of DEA, however, the estimation of technology and the measurement of improvement potentials relative to the technology are done simultaneously. The reason is that DEA does not provide a simple closed form expression of the underlying technology or the associated distance function but rather an implicit one in terms of a set of linear constraints. These constraints are imposed when we seek to find gains form mergers.

Since this chapter describes the details of the underlying calculations, it is somewhat technical. It can however be skipped by less technically interested readers.

3.1 *Implementation - the horizontal case

In the general formulations in Part B, we evaluated the potential effects of mergers relative to a general technology set T. In the previous Chapter we developed methods to approximate T by an empirical reference technology T*. The estimate of T* is based on the actual observations, say of n DMUs, DMU^i, $i \in I = \{1, 2, \ldots, n\}$, that have transformed p inputs (x^i) to q outputs (y^i).

3.1.1 SFA implementation

Consider first a technology estimated using SFA. To be general, we may assume that we have estimated a Shepard input distance function $D_I(x, y)$ from the actual data. Using this we can approximate the technology as

$$T^* = \{(x, y) \in R_0^{p+q} \mid D_I(x, y) \geq 1\}$$

Inserting this into the different mergers measures from Part B, we can calculate the different measures of potential gains from mergers.

To be specific, let us assume that we want to evaluate a hypothetical entity (x*, y*) against T*. Using the usual Farrell input measure, this then amounts to solving

$$\text{Min } \{E \mid D_I(Ex^*, y^*) \geq 1\}$$

which is particularly simple by the relationship between Farrell and Shepard distance measures, i.e. $E = 1 / D_I$ in this case. More generally, however, the problem can be solved by a line search like below.

Consider next the case where some inputs non-discretionary, i.e. $x = (x_V, x_F)$. In this case we evaluate (x*, y*) using

$$\text{Min } \{E \mid D_I(Ex^*_V, x^*_F, y^*) \geq 1\}$$

Since D_I is monotonous (weakly increasing) in x_V and hereby in the scalar E, this problem can be solved by a simple line search, e.g by a bi-section approach that does not require any further assumptions about the functional form of D_I.

Similarly, consider the case of a general directional distance measure G_d in the direction $(-d_x, d_y)$. Again, to determine G_d we need to solve

$$\text{Max } \{G \mid D_I(x^* - Gd_x, y^* + Gd_y) \geq 1\}$$

As a function of the scalar G, $D_I(x^* - Gd_x, y^* + Gd_y)$ is monotonous (weakly decreasing) and therefore we can again make use of a simple line search to approximate G_d. The simple search problems are illustrated in Figure 4 below.

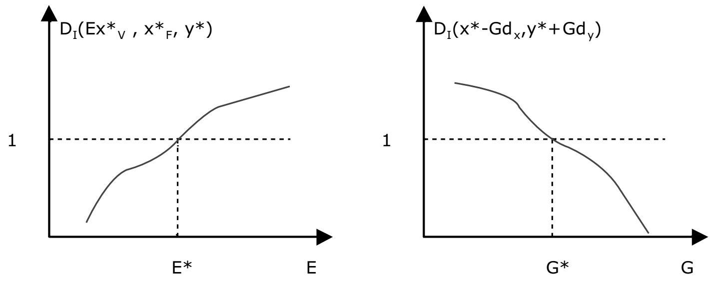

Figure 4: Simple line search

Note also that if the post merger units are working at an efficiency level of $E^{\#}$, then the saving potential for a unit (x*, y*) reduces to

$$\text{Max } \{G \mid E^{\#}\, D_I(x^* - Gd_x,\, y^* + Gd_y) \geq 1\}$$

Hence it is straightforward to calculate improvement potentials while recognizing that even ex post there may be organizational and market conditions that hinder full efficiency.

To solve such line search approach problems, we propose to use simple bi-section methods. Consider the last, general directional distance function approach as illustrated to the right in Figure 4 above. To evaluate G, we can now proceed as follows.

Step Initialization
We start with two values G_L and G_H such that

$$D_I(x^* - G_L d_x,\, y^* + G_L d_y) > 1 \text{ and } D_I(x^* - G_H d_x,\, y^* + G_H d_y) < 1$$

Step 1
Calculate middle point

$$G_M = (G_L + G_H)/2$$

Step 2

$$\text{If } D_I(x^* - G_M d_x,\, y^* + G_M d_y) > 1 \text{ let } G_L := G_M$$
$$\text{otherwise let } G_H := G_M$$

Step 3
If $G_H - G_L < \varepsilon$ let $G = G_L$ and stop, other wise go to Step 1.
Here ε is some small number determining the precision of the numerical method.

When the procedure stops we have a good approximation of G, and we can calculate the projected point (the benchmark) as $(x^* - Gd_x, y^* + Gd_y)$

While all of the above problems are easy to solve for any closed form of the distance function, the more elaborated reallocation problem can become rather complex. Consider for example the reallocation problem to calculate the harmony measure h above, i.e.

$$\begin{aligned}
&\min_{(x^{\#j},y^{\#j}),\ j\in I} && h \\
&\text{s.t.} && \textstyle\sum_{j\in J} x^{\#j} \le h\sum_{j\in J} E^j x^j \\
& && \textstyle\sum_{j\in J} y^{\#j} \ge \sum_{j\in J} y^j \\
& && (x^{\#j}, y^{\#j}) \in T
\end{aligned}$$

When we approximate T with $T^* = \{(x, y)\in R_0^{p+q} \mid D_I(x, y) \ge 1\}$ above, we will for most distance functions get a non-linear constraint set. Presuming that the estimated D_I has all the properties of a theoretical distance function, we would in the case of a regular (convex) technology get a convex optimization problem for which there are still relatively efficient general numerical algorithms available.

Observe also that the dis-integration model from Part B will be non-trial in the case of a SFA modelled technology. The complexity is basically similar to that of the general reallocation model above; the problem is that we have to choose inputs and outputs for two units and that each of them to be feasible units must fulfil non-linear constraints. We will discuss these difficulties in more details in the next Section since the integration of vertical entities involve similar problems as we have to choose feasible intermediate product plans.

3.1.2 DEA implementation

Consider next the case of DEA estimated technologies. In this case we do not have a closed form expression of the distance function and evaluations are therefore slightly more involved than in the parametric case.

From the given observations of DMU^i, $i \in I$, we can construct the technology set as

$$T^*(s) =\{(x, y) \in R_0^{p+q} \mid \exists\, \lambda \in R_0^n : x \ge \textstyle\sum_i \lambda^i x^i,\ y \le \sum_i \lambda^i y^i,\ \lambda \in \Lambda(s)\}$$

cf. Chapter 2.4. We can therefore evaluate a hypothetical entity (x*, y*) against T* by solving the simple LP problem

$$\begin{aligned} \min_{E,\lambda} \quad & E \\ \text{s.t.} \quad & Ex^* \geq \textstyle\sum_{j\in I} \lambda^j x^j \\ & y^* \leq \textstyle\sum_{j\in I} \lambda^j y^j \\ & \lambda \in \Lambda(k) \end{aligned}$$

in the case of all inputs being discretionary and by

$$\begin{aligned} \min_{E,\lambda} \quad & E_V \\ \text{s.t.} \quad & E_V x^*_V \geq \textstyle\sum_{j\in I} \lambda^j x^j_V \\ & x^*_F \geq \textstyle\sum_{j\in I} \lambda^j x^j_F \\ & y^* \leq \textstyle\sum_{j\in I} \lambda^j y^j \\ & \lambda \in \Lambda(k) \end{aligned}$$

when only the x_V part of x is discretionary.

Similarly, if we want to use the general directional distance measure G_d in the direction $(-d_x, d_y)$, we must solve a simple LP problem, namely

$$\begin{aligned} \max_{G,\lambda} \quad & G \\ \text{s.t.} \quad & x^* - Gd_x \geq \textstyle\sum_{j\in I} \lambda^j x^j \\ & y^* + Gd_y \leq \textstyle\sum_{j\in I} \lambda^j y^j \\ & \lambda \in \Lambda(k) \end{aligned}$$

In all cases, these LP problem have p+q ordinary constraints plus possibly one constraint relating to the $\lambda \in \Lambda(k)$ condition. Also, the number of variables equals n+1 in all cases. For usual data sets, these are small LP problems that can easily be solved on standard desktop computers.

The DEA approach leads also to simple although larger LP problems when we seek to evaluate the possible gains from reallocation directly. The h problem above, for example, becomes a LP problem that contains the usual DEA problems for each individual entity plus a series of constraints that bind together the individual entity models. Several such models are illustrated and solved in Andersen and Bogetoft(2007) and in Bogetoft, Boye, Neergaard-Petersen and Nielsen (2007).

If in the DEA context, we are interested in performance relative to the biased corrected technology, this can be done easily by determining the biased corrected original entities, i.e. using $(E^{**i}x^i, y^i)$ instead of (x^i, y^i) where E^{**i} is the bias corrected input efficiency of DMU^i, cf. the discussion in chapter 2.4.6. To normal input based measurement of the hypothetical entity (x*, y*) can therefore be determined by solving

$$\begin{array}{ll} \text{Min} & E \\ E, \lambda & \\ \text{s.t.} & Ex^* \geq \sum_{j\in I} \lambda^j E^{**j} x^j \\ & y^* \leq \sum_{j\in I} \lambda^j y^j \\ & \lambda \in \Lambda(k) \end{array}$$

and similarly for the other measures.

Also, we can assume that the post merger units are only working at efficiency level $E^{\#}$. In that case the saving potential becomes

$$\begin{array}{ll} \text{Min} & E \\ E, \lambda & \\ \text{s.t.} & E^{\#} E\, x^* \geq \sum_{j\in I} \lambda^j E^{**j} x^j \\ & y^* \leq \sum_{j\in I} \lambda^j y^j \\ & \lambda \in \Lambda(k) \end{array}$$

In this case, the saving reduces with a factor $E^{\#}$.

Using DEA models, it will also be easy to solve the dis-integration models. The resulting optimization problem would again become a linear programming problem, only now with three times more rows and twice (+2) times more columns since we would basically have to put up a standard DEA model for each of the new entities we are designing and in addition have rows that ensure that we get the outputs of the original unit covered without using any more input.

In Part B, we also developed general decompositions of the potential gains E^J into learning TE^J, scope (harmony) H^J and scale (size) gains S^J:

$$E^J = TE^J * H^J * S^J$$

To determine these measures, we constructed different new entities and evaluated their efficiency relative to the estimated technology T*. The practical implementation is therefore similar to the above.

3.1.3 Decompositions using R software

A simple way to undertake these calculations is to use the R programming environment.

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

To calculate the potential gains from horizontal mergers – and the decomposition into learning, scope and scale effects, we may proceed as follows:

Let data be given in an input matrix **X** (with DMUs in columns and the different inputs in the rows) and an output matrix **Y** (with DMUs in columns and the different outputs in the rows).

Also, let **M** be a matrix where each column corresponds to a merger and the rows corresponding to the different DMUs to be merged. Thus for example, a column like $(1,0,1,0,0,1,0...0)^t$ would correspond to a merger of DMU^1, DMU^3 and DMU^6.

Also, let the calculation of the Farrell efficiency of some entity using inputs x to produce outputs y be denoted **eff(x,y)**. This function can be defined as the solution to a linear programming problem or a line search problem depending on the underlying model of the technology. In the simple case of a single input multiple outputs technology represented by a cost function C(.) it could also be the cost efficiency eff(x, y) = C(y)/ x.

With these definitions, the following lines of R code will generate the efficiency gains and their decompositions for all possible mergers represented in the M matrix:

```
# CALCULATE THE AGGREGATED UNITS
Xmer <- X %*% M
```

Here we have calculated the merged units' inputs - the first column could contain the sum of the inputs of DMU^1, DMU^3 and

DMU6 in the example above. Note that %*% is matrix multiplication.

Ymer <- Y %*% M

This gives in a similar way the merged units' output vectors.

BASIC POTENTIAL GAINS

E <- eff(Xmer,Ymer)

These are the overall gains. E = (0.9,0.7,..) suggests for example that we can save 10% of all inputs in the first merger, 30% in the second merger, etc.

LEARNING

e <- eff(X,Y)

These are the efficiencies of the individual units.

Xmer_proj <- (X%*%diag(e)) %*% M

These are the input consumptions of the merged units after having eliminated individual inefficiencies

E_star <- eff(Xmer_proj,Ymer)

These are the overall gains when we have eliminated individual learning, i.e. the E*s

TE <- E/E_star

These are the learning measures, i.e. the TEs

HARMONY

Xharm <- Xmer_proj %*% diag(1/colSums(M))

Yharm <- Ymer %*% diag(1/colSums(M))

These are the inputs and outputs of the merged units after having eliminated individual inefficiencies and having eliminated the impact of scale by considering the average sizes.

H <- eff(Xharm,Yharm)

These are the harmony (scope) measures.

SIZE EFFECT

S <- E_star/H

These are the size measures

Having undertaken these calculations, we have in the **E, TE, H** and **S** vectors the values of E^J , TE^J, H^J , S^J for all the mergers J corresponding to columns in the **M** matrix.

3.2 *Implementation - the vertical and network cases

In Part B, we developed several ways to calculate and decompose the potential gains from vertical integrations. The aim of this section is to briefly discuss the practical implementation of these measures using technologies modelled by SFA and DEA models.

To do so, let us consider the general network model of vertical integration from Part B. The situation is illustrated in Figure 5 below. In this we have two production units P^1 and P^2 that both use primary inputs to produce final outputs. In addition, P^1 produces intermediate products and P^2 consumes intermediate products.

The primary inputs used by P^1 and P^2 may without loss of generality be thought of as the same types, as long as we make no assumption about strictly positive values since a product only consumed by one of the units can then simply be represented by a zero in the corresponding coordinate for the other unit. The same goes for final outputs of course. This implies that additions make sense, i.e. it is meaningful to consider the total consumption of inputs as $y^{01}+y^{02}$ and the total production of final products as $y^{13}+y^{23}$, respectively.

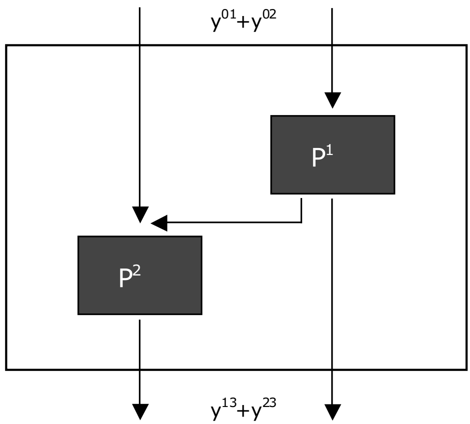

Figure 5: Network model

Now to evaluate the potentials from organizing production in this network as a joint operation, we basically need a model of the underlying technology for the integrated entity, T. Also, to make the decompositions, we need models of technologies for the individual entities P^1 and P^2, T^1 and T^2.

In some situations, we may model T directly. It basically requires that we can find entities that transform $y^{01} + y^{02}$ to $y^{13} + y^{23}$, e.g. companies where the upstream activities have already been integrated (in-sourcing).

In others we may try to model T^1 and T^2 and derive T from these using

$$T = \{(y^{0\cdot}, y^{\cdot 3}) \mid \exists\, y^{01}, y^{02}, y^{12}, y^{13}, y^{23}: \; (y^{01}, y^{12} y^{13}) \in T^1, \; (y^{02}, y^{12}, y^{23}) \in T^2, \; y^{0\cdot} = y^{01} + y^{02}, \; y^{\cdot 3} = y^{13} + y^{23}\}$$

i.e. by looking for an allocation of inputs among P^1 and P^2, a vector of intermediate products to be transferred from P^1 to P^2, and an allocation of output requirements on P^1 and P^2, such that in total we transform y^0 to $y^{\cdot 3}$. In practice, combining T^1 and T^2 to form T may be challenging since it effectively requires optimization as we shall see below. Note that $y^{0\cdot}$is the sum of primary inputs originating in node 0, and $y^{\cdot 3}$ is the sum of final products received in node 3. We have here ignored the possible ex post managerial inefficiency and the possible variation in the controllability of the resources. Corrections to accommodate these effects are easily introduced as illustrated above

Of course we can also pool observations of actual joint operations with analytically derived feasible productions (like in the expression of T above).

To sum up, if we can model T1 and T2, we can derive (or supplement any information we have already on) T.

On the other hand, we cannot derive the processes models from the aggregate model, i.e. we cannot derive T^1 and T^2 from T (unless we make additional assumptions like in Bogetoft and Tind(2006)).

3.2.1 SFA implementation

Consider now the situation where we have estimated the technologies using SFA. To be general, we may assume that we have estimated Shepard input distance functions $D^1_I(x,y)$ and $D^2_I(x,y)$ for the two sub-processes P^1 and P^2.

In that case and by combining with free disposability we get

$$T = \{(y'^{0},y'^{3}) \mid \exists\, y'^{01},y'^{02}, y'^{12},y'^{13},y'^{23}: \\ D^{1}_{I}(y'^{01},y'^{12},y'^{13}) \geq 1, \\ D^{2}_{I}(y'^{02},y'^{12},y'^{23}) \geq 1 \\ y'^{01}+y'^{02} \leq y'^{0} \\ y'^{3} \leq y'^{13}+y'^{23}\}$$

We can therefore find the total potential gains from vertical integration of the two production units P^1 and P^2 that so far has used primary inputs y^{01} and y^{02} and produce final outputs y^{13} and y^{23} , i.e.

$$E^{J} = \mathrm{Min}\{E \in R_{0} \mid (E(y^{01} + y^{02}), (y^{13} + y^{23}) \in T\},$$

by solving the following problem

$$\min_{y'^{01},y'^{02}, y'^{12},y'^{13},y'^{23}, E} \quad E$$
$$\text{s.t.} \quad D^{1}_{I}(y'^{01},y'^{12},y'^{13}) \geq 1, \\ D^{2}_{I}(y'^{02},y'^{12},y'^{23}) \geq 1 \\ y'^{01}+y'^{02} \leq E(y^{01}+y^{02}) \\ y^{13}+y^{23} \leq y'^{13}+y'^{23}$$

The objective function is linear and so are the two last sets of constraints. Unfortunately, the first two constraints will typically be non-linear and the resulting optimization problem is therefore not easy to solve and certainly much more complex than the simple line search problems from the horizontal case.

We see also that the decomposition involves similar problems. Although we developed the decomposition for the simple (purely) vertical case in Part B, we can easily extend it to the network case as follows:

First, we shall calculate individual improvements E^1 and E^2 and hereby the learning effect as

$$TE = E^{1} * E^{2}$$

This only requires models of the individual technologies in P^1 and P^2, T^1 and T^2. To illustrate, we would determine the efficiency E^2 of P^2 that have so far transformed y^{02} and y^{12} to y^{23} by solving

$$\min_{E} \quad E$$
$$\text{s.t.} \quad D^2_I\,(E(y^{02},y^{12}),y^{23}) \geq 1$$

which is a simple line search problem like in the horizontal case.

Next to calculate scale effects, we shall first calculate

$$E^{**J} = E^1(y^{01}, E^2 y^{12})$$

and then

$$S = E^{**J} / TE.$$

To determine E^{**J}, we only need a model of P^1, and the problem therefore reduces to another line search problem

$$\min_{E} \quad E$$
$$\text{s.t.} \quad D^1_I\,(Ey^{01},E^2y^{12}) \geq 1$$

Finally, we can calculate the harmony effect as a residual,

$$H^J = E^J / (TE^J * S^J).$$

In some situations, we may estimate a model of the combined processes directly, i.e. a distance function $D(y^{0\cdot},y^{\cdot 3})$. This may for example be possible if there are several organizations that already have both processes P^1 and P^2 in-house, cf. the discussion above. If we use such a model, the determination of E reduces to a simple line search problem. In such cases, therefore, we can easily solve all the estimation problems.

3.2.2 DEA implementation

Consider next the case of DEA estimated technologies. Due to the approximation of technologies by (polyhedral) convex sets, the calculations are actually easier with such an approach.

To see this, let us assume that we have a series of observation (y^{h01}, y^{h12}, y^{h13}), h = 1,…,H of the first process P^1 and a series of observations (y^{k02}, y^{k12}, y^{k23}), k = 1,…,K of the second process.

We can then approximate the two process technologies by

$$
\begin{aligned}
T^{1}*(s^{1}) = \{(y^{01}, y^{12}, y^{13}) \mid \exists\lambda \in R_0^H :\ & y^{01} \geq \textstyle\sum_h \lambda^h y^{h01}, \\
& y^{12} \leq \textstyle\sum_h \lambda^h y^{h12}, \\
& y^{13} \leq \textstyle\sum_h \lambda^h y^{h13} \\
& \lambda \in \Lambda(s^1) \}
\end{aligned}
$$

$$
\begin{aligned}
T^{2}*(s^{2}) = \{(y^{02}, y^{12}, y^{23}) \mid \exists\delta \in R_0^K :\ & y^{02} \geq \textstyle\sum_h \delta^k y^{k02}, \\
& y^{12} \geq \textstyle\sum_k \delta^k y^{k12}, \\
& y^{23} \leq \textstyle\sum_k \delta^k y^{k23} \\
& \delta \in \Lambda(s^2) \}
\end{aligned}
$$

and the technology T by

$$
\begin{aligned}
T*(s^{1},s^{2}) = \{(y^{0\cdot}, y^{\cdot 3}) \mid \exists\lambda \in R_0^H, \delta \in R_0^K, y^{01}, y^{12}, y^{13}, y^{02}, y^{23}:\ & y^{01} \geq \textstyle\sum_h \lambda^h y^{h01}, \\
& y^{12} \leq \textstyle\sum_h \lambda^h y^{h12}, \\
& y^{13} \leq \textstyle\sum_h \lambda^h y^{h13} \\
& \lambda \in \Lambda(s^1) \\
& y^{02} \geq \textstyle\sum_h \delta^k y^{k02}, \\
& y^{12} \geq \textstyle\sum_k \delta^k y^{k12}, \\
& y^{23} \leq \textstyle\sum_k \delta^k y^{k23} \\
& \delta \in \Lambda(s^2) \\
& y^{01}+y^{02} \leq y^{0\cdot} \\
& y^{13}+y^{23} \geq y^{\cdot 3} \}
\end{aligned}
$$

It follows directly that we can calculate the overall gains from vertical integration E^J as well as its decomposition into technical inefficiency, TE = E^1'E^2, size S = $E^1(y^{01}, E^2 y^{12})$/ TE, and harmony H = E^J/ (TE * S) by simple linear programs. Specifically, if P^1 has used y^{01} to produce y^{12} and y^{23} and P^2 has used y^{02} and y^{12} to produce y^{23}, we can calculate the gain and its elements like:

$$E^J = \text{Min}\{E\in R_0 \,|\, (E(y^{01}+y^{02}) , y^{13} + y^{23}) \in T^*(s^1,s^2)\}$$

$$E^1 = \text{Min}\{E\in R_0 \,|\, (Ey^{01} , y^{12}, y^{13}) \in T^{1*}(s^1)\}$$

$$E^2 = \text{Min}\{E\in R_0 \,|\, (E(y^{02} , y^{12}), y^{23}) \in T^{2*}(s^2)\}$$

$$TE = E^1 * E^2$$

$$E^{**J} = E^1(y^{01}, E^2y^{12},y^{13}) = \text{Min}\{E\in R_0 \,|\, (Ey^{01} , E^2y^{12}, y^{13}) \in T^{1*}(s^1)\}$$

$$S = E^{**J} / TE$$

$$H^J = E^J / (TE^J * S^J)$$

Also, instead of calculating the total gains first and the scope (harmony) effect as a residual in the end, we could of course calculate the scope effect directly as

$$H^J = E(E^{**J} y^{01} + E^2y^{02}, y^{13} + y^{23})$$
$$= \text{Min}\{E \in R_0 \,|\, (E(E^{**J}y^{01} + E^2y^{02}), y^{13} + y^{23}) \in T^*(s^1, s^2)\}$$

and then calculate the total gains like

$$EJ = TE^J * H^J * S^J$$

Net-puts

To simplify and standardize the modelling of alternative networks using DEA, and to facilitate also software development, it is useful to suppress the distinction between inputs and outputs and simply think of products as netputs. Positive netputs are then traditional outputs while negative netputs are (-) traditional inputs.

Consider for example the simple vertical model where one up-stream unit uses inputs y^{01} to produce intermediate products y^{12} that are used in the down-stream unit to produce final outputs y^{23}. The overall potential gains from integrating using the notation so far is

$$\begin{aligned}
&\min_{E,\lambda,\gamma} \quad E \\
&\text{s.t.} \quad Ey^{01} \geq \textstyle\sum_h \lambda^h y^{h01} \\
&\qquad \textstyle\sum_h \lambda^h y^{h12} \geq \sum_k \gamma^k y^{k12} \\
&\qquad \textstyle\sum_{j\in J} \gamma^k y^{k23} \geq y^{k23} \\
&\qquad \lambda \in \Lambda(s^1) \qquad \gamma \in \Gamma(s^2)
\end{aligned}$$

Using netputs instead we get that the overall improvement potential in the netput vector $(-y^{01},y^{23})$ corresponding to the use of y^{01} to produce y^{23} can be determined by solving

$$
\begin{array}{llll}
\min\limits_{E,\lambda,\gamma} & E & & \\
\text{s.t.} & E(-y^{01}) \leq & \sum_h \lambda^h(-y^{h01}) & \\
 & 0 \leq & \sum_h \lambda^h y^{h12} + & \sum_k \gamma^k(-y^{k12}) \\
 & y^{23} \leq & & \sum_k \gamma^k y^{k23} \\
 & & \lambda \in \Lambda(s^1) & \gamma \in \Gamma(s^2)
\end{array}
$$

4. Quality

Quality is important in any sector and in the health care sector in particular. For a given patient, it is important not only to get treated but also to get a treatment of high quality. Indeed, the latter may outweigh the former and patients may choose to delay treatments or pay extra to get access to more experienced providers. It is therefore important to consider possible quality effects of mergers and other types of re-organizations.

Quality is also a particular challenge in the attempt to measure the potential gains from mergers. We are not aware of scientific papers on how this can best be done, and the empirical evidence on the quality implications of mergers is limited with no clear conclusions, as we shall see below.

To address the issue, we will first identify different types of quality. The intension is not to be comprehensive; health care quality is a complex and sizeable subject, which cannot be covered in any single report. Our intension is therefore to raise the issue and to point out some principal ways to handle quality in benchmarking and hereby also in evaluations of the potential effects of mergers. We discuss some key quality indicators and how they differ from the point of views of generation, measurement, and interaction with mergers.

We next give a brief account of the scattered literature on the interaction between quality and mergers in the health sector.

Having prepared in this way, we shall discuss how quality can be included in the costs and production economic models used in benchmarking. The challenge is in particular to avoid the curse of dimensionality.

We finally discuss the inclusion of the quality adjustments in the merger programs of the previous chapters.

4.1 Quality indicators

To get started, we shall discuss some of key quality dimensions and indicators that are generally accepted in the health care sector.

4.1.1 Quality dimensions

The notion of health quality is a multi-dimensional concept. Following OECD(2006), the three most commonly accepted dimensions are

Effectiveness
the extent to which attainable improvements in health are attained. It is the degree to which processes result in desired outcomes, free from error.

Safety
the degree to which health care processes avoid, prevent, and ameliorate adverse outcomes or injuries that stem from the processes of health care itself. Safety is hereby a dimension that is closely related to effectiveness, but distinct by emphasizing the prevention of unintentional adverse events for patients.

Responsiveness (patient-centeredness, patients' satisfaction)
how a system treats people to meet their legitimate non-health expectations. It concerns the degree to which a system actually places the patient/ user at the center of its delivery of healthcare, and how it focuses on the caring, communication and understanding in the clinician-patient relationship. Responsiveness is often assessed in terms of patient's experience of their health care.

Two other important dimensions are obviously

Accessibility
the ease with which health services are reached. Access can be physical, financial or psychological, and requires that health services are a priori available.

Equity (equitability)
the extent to which a system deals fairly with all concerned. Equity, in this context, deals with the distribution of healthcare and its benefits among people.

The last two dimensions are primarily properties of the health care sector at large. We suggest therefore that these dimensions may be of less importance in the evaluation of a single merger unless the units have very large (local) market shares.

4.1.2 Quality indicators in The Netherlands[2]

In practice indicators for these dimensions are in The Netherlands as in most countries, in a phase of development. Different organizations have developed different indicators in the last years. This process had only limited results as the approach toward quality became fragmented.

To create a uniform and nationwide quality indicator system, The Netherlands Health Care Inspectorate launched the programme Transparency in the Health-Care Sector (Zichtbare Zorg). The aim is to develop appropriate indicators in every segment of the healthcare and to organize the regular data collection. The final goal of the programme is to be able to provide reliable, relevant, valid and publicly available information about quality of institutions. The Inspectorate cooperates to reach the goals of the project with several healthcare institutions and professionals (e.g. hospitals, specialists, nurses, health insurance companies, patient organizations).

As an example we demonstrate the type of indicators under development in the hospital sector. The hospital quality indicators can be classified into three groups. The *care-related indicators* measure the safety, effectiveness and efficiency of the treatment of a given disease from a medical point of view. At the moment ten indicator sets are available which cover the most frequent diseases (e.g. diabetes, knee and hip operation, incontinence).

The second group of indicators, *etalage+ questionnaires*[3], describes the provided health services of a hospital. These indicators aim to discover the completeness or quality of provided services to cure a disease. For example in case of diabetes, there is information about the specialists who are available in the hospital, and if they make up a special diabetes team. Other points refer to the supportive services like the availability of nurses for everyday (practical) questions about diabetes via telephone or e-mail. Questionnaires of six diseases are developed so far.

The third group of indicators is the *CQI-questionnaires (Consumer Quality Indexes)*. These indicators measure the quality of healthcare from the patients' point of view. They focus on topics that patients consider important and put the emphasis on the patients' experience. At present two measures are completed: one for cataract operations and one for knee and hip operations.

---

2 Source: http://www.zichtbarezorg.nl

3 Questionnaires completed by medical specialists to rate their hospitals' performance in various fields

The goal of the programme Transparency in the Health-Care Sector is to be able to measure and publish the quality of all healthcare segments as from 2010. The Inspectorate intends to measure the performance related to eighty diseases in total. The developed database could be the as choice information (for patients), as purchase information (for insurers) and as reference information (for doctors).

4.1.3 Indicator types

To understand the nature of alternative more specific indicators, it is useful to distinguish between input (structure), process and outcome oriented indicators.

Input indicators
or structure indicators capture the characteristics of, or inputs to, health care. One may for example measure if doctors have specific qualifications or if hospitals are appropriately equipped. Of course, input indicators are not linked 1-1 to quality but their presence is expected to facilitate the delivery of a desired quality.

Process indicators
represent measures of the delivery of appropriate health care. Compliance with health care protocols is a good example of such indicators. One might for example see if, for those at risk, patients' blood pressure is checked regularly by a physician, or whether certain tests are done sufficiently fast after a patient's admittance to a hospital. Again, one can say that process indicators are only indirect – following specific protocols are likely means of good health care outcomes but hardly an end in itself. Another concern is that such indicators are more vulnerable to gaming than outcome or structure measures.

Process measures represent the closest approximation of the actual health care offered and are typically the most clinically specific of the three types of indicators. The Dutch DBC system in itself can be seen as a very detailed set of process indicators.

Outcome indicators
seek to measure health improvements attributable to medical care. From an incentive provision perspective these are usually the indicators that would be preferred – the idea is that there is asymmetric information about the actual ways to accomplish different outcomes and that one should therefore preferably give incentives on the outcomes and leave the specific implementation to the better-informed provider.

Unfortunately, there are in general several complications in the measurement of the health improvements.

One is that it requires a counterfactual. In particular, it may not suffice to compare health status before and after treatment since the status without treatment (the counter-factual) is likely to change also, c.f. e.g. Jacobs, Smith and Street(2006).

Another complication is that health improvement may be very depended on other factors than quality of care, e.g. patient-level factors like socioeconomic status. Ideally, such other factors should be appropriately accounted for – e.g. via a well-developed risk adjustment.

A third complication that is sometimes suggested is that there may be no 1-1 link between the physical status and the patient's own experience. Some would therefore argue that it is important to attach a capability model to the model of for example a hospital. The capability model will transform changes in physical characteristics to changes in the patient's sense of well-being or quality of life. The medical outcome may be improved eyesight but the capability may be the self-valued ability to read. Improved physical conditions have limited value if they do not transform into changes in what the patient does. Recent research suggests that the latter stage can be modeled much like a production economic model and that changes in capabilities can be proxied using indices related to the Malmquist index for changes in productivity, c.f. e.g. Färe, Grosskopf, Lundström and Roos(2007).

It follows from the above that neither type of indicator – input, process or output – gives a perfect measure of the actual quality. This speaks in favor of using a multiplicity of measures.

4.2 Empirical literature

We now give a brief account of the scattered empirical literature on the interaction of quality and mergers in the health sector.

A rather comprehensive empirical analysis is Wan e.a. (2002). They investigate the role of different integration mechanism used in integrated health care delivery systems (IDSs) nationwide in the US and conclude that high efficiency in hospital care can be achieved by employing proper integration strategies in operations.

Integration mechanisms are categorized into six related domains:

*Informatics integration**
Informatics integration, is characterized as the use of a variety of automated application systems by an IDS to integrate its administrative, management and clinical functions

Case management
The application of case/ disease management refers to the clinical operational integration, care coordination, and teamwork in an IDS.

*Hybrid physician–hospital integration**
Hybrid physician hospital integration is measured by whether or not an IDS has any forms of physician groups affiliated with its operation —including MSOs, PHOs, IPAs, and other arrangements — that are established to offer broad geographic coverage, reduce costs, develop name recognition in the market, and compete on the basis of differentiation.

Forward integration
Forward integration is the establishment of linkages among home health services, hospices, physical rehabilitation outpatient services, rehabilitation care, skilled nursing care, and other long-term care.

Backward integration
Backward integration refers to the formation of freestanding outpatient centers, hospital-based outpatient care centers, outpatient surgery, primary care department, and psychiatric outpatient services within a health care system.

*Services differentiation**
Services differentiation involves the offering of high tech medical services and integration across a number of high tech medical services.

Informatics integration, hybrid physician – hospital integration, and service differentiation (marked with a * above) are found to have a significantly positive impact on hospital efficiency. The other integration mechanisms did not have a significant relation to efficiency.

The qualitative case studies by Meyer e.a. (2004) of US hospitals identify four elements of successful quality improvement strategies, namely

- Developing the right culture for quality to flourish
- Attracting and retaining the right people to promote quality
- Devising and updating the right in-house processes for quality improvement
- Giving staff the right tools to do the job.

This suggests that the quality impact of a merger is likely to depend on the details of the merger and the post merger culture, processes and tools.

4.3 Quality in benchmarking

We continue by discussing some principal ways to include quality into the efficiency analyses models. The challenge here is to give a proper inclusion without running into dimensionality problem, i.e. to give a reasonable approximation without having a number of output indicators that will prohibit econometric calibration given the data sets that are available or is likely to become available within foreseeable future.

In fact, the discussion here is not only relevant for quality. It can be extended also to cope with other complicating factors and properties, i.e. local conditions for, and local properties of for example the hospital activities that should ideally be taken into account, e.g. patients' education, age, etc. It hereby illustrates ways to account for environmental and context variables in general. The challenge is to account for these details without having too many dimensions in the model to prohibit comparisons and hereby the ability to discriminate the managerial and organizational efficiency of the different entities.

We will now formalize the discussion on how quality can be included in a benchmarking model T. In addition to the usual inputs x and outputs y, we shall in this section allow for a vector z of some r quality dimensions or complicating factors or properties with possible values in Z.

One way to distinguish between the different ways to include z is to think of factors that affect the transformation of inputs into outputs in an integrated or in a separate manner. In the case of quality, the question is if we can think of quality improvements, as being done in a separate process or whether it is intimately integrated with the core processes such that the products we get out are really different products.
Is a particularly safe heart operation just a particular form of general heart surgery, or is it as distinct from general heart procedures as it is from

other types of surgery? In the latter case, it would call for a particular product dimension. In the other cases, we might be able to handle it by modifying the cost impact according to the safety level.

4.3.1 Quasi inputs and outputs

The usual way to handle complicating factors is as quasi inputs (if they facilitate the outputs) or outputs (if they require resource to cope with). If they are furthermore non-controllable, as with complicating factors but not complicating properties, this is handled by avoiding contractions (or expansions) in the direction of these factors.
The modified input-based measure becomes

$$E^j = \text{Min}\{E \in R_0 \mid (Ex^j, z^j, y^j) \in T\}$$

The advantage of this approach is that the dimensionality of the problem does not expand too much although there will be some loss in discriminatory power. The primary disadvantage of this approach is that several factors affect multiple inputs or multiple outputs but are not inputs or outputs in the usual sense. This means that the usual production economic properties, like convexity and scaling up and down, may make little sense with regard to the z factors.

There are several possible variations of this approach. The straightforward application of the approach would in a DEA context require that the reference units have been using no more of the z inputs than the evaluated unit. This is the old idea of Banker and Morey(1986). A possible non-attractive feature of this approach is that the reference unit may contain units that have been working under both tougher (which is fine) and easier conditions (which may be problematic). Ruggiero(1996) argues that no entities with easier conditions (higher z) should be allowed in a (virtual) dominating combination.

Another difficulty of the approach is that the units in a dominating combination may be of very different size. This means that a large share of outputs in a reference unit can potentially be produced in units working under easier conditions. Recently, Olesen and Petersen (2007) have therefore proposed a volume-weighted variant of the approach by using (in the single output case) constraints like

$$\Sigma_j \lambda^j y^j (z^j - z^i) \leq 0$$

instead of the usual

$$\Sigma_j \lambda^j z^j \leq z^i$$

in the DEA problem. This approach can also be related to the idea of using handicap functions to effectively increase inputs or decrease outputs in the case of favorable conditions, cf. Paradi, Vela and Yang (2004)

A particularly important variant of the approach is to work with z variables that are ordinal or even categorical. In such cases, we would effectively split the general model in a series of (in the case of ordinal variables, nested) sub-models, one for each quality level. If for example we distinguish between easy and difficult patients (corresponding to for example their general health status), hospitals with difficult patients should be allowed to dominate hospitals with easy cases, but hospitals with easy cases can only help span the frontier for other easy cases. This is similar to the idea of Ruggiero(1996). In a DEA context, this implies that $z^j > z^i$ implies $\lambda^j = 0$ when we evaluate DMU^i.

4.3.2 Contingent inputs and outputs

Complicating factors affecting the nature of the inputs or outputs shall ideally be dealt with by redefining the inputs and outputs according to the z values. That is, one distinguishes between for example nurses trained in one school compared to another school. Similarly, one may distinguish between emergency capacity produced in rural and urban areas, respectively. Formally, this approach means that we use $(x_z, z \in Z)$ as inputs and $(y_z, z \in Z)$ as outputs where x_z and y_z are the number of the inputs and outputs with properties z.

The advantage of this approach is that it is theoretically sound as it corresponds to the idea of state-contingent goods etc. The disadvantage is that the dimensionality explodes as the number of inputs and outputs increases from p + q to |Z|(p + q) where |Z| is the cardinality of Z.

4.3.3 Adjustment coefficients

One way to deal with the complications as properties while at the same time avoiding the explosion of dimensionality is by the use of adjustment coefficients on the input and output side. This is similar to the familiar correction for variation in salaries or currencies.

Formally, we simply redefine the input and output vectors into W(z)x and P(z)y where W(z) and P(z) are pxp and qxq diagonal matrices, respectively. This effectively adjusts the output iso-quants for given inputs depending on the quality level – or it adjusts in the input iso-quants for given outputs.

If for example there is a general increase in the costs of a% from higher quality, one can adjust the outputs using P(z) with (1 + a/ 100) in all diagonal cells or one can adjust the inputs using W(z) with 1/ (1 + a/ 100) in all diagonal cells. The correction factors can be derived from experts or from underlying models.

The advantage of the adjustment factor approach is that it does not increase the dimensionality of the problem. Of course, this is accomplished by hiding the problems of determining the aggregate impact of the complicating properties inside an expert or a sub-model.

4.3.4 Factorized impacts

Between the extremes of using aggregate impact coefficients and full scale contingent input-output models, one can consider the use of factorized impacts by assuming one of the following regularities

$$\text{Eff}((x_z,\ z\in Z),\ (y_z,\ z\in Z))=E(x,\ y)G(z)$$

$$\text{Eff}(x,\ y,\ z)=E(x,\ y)G(z)$$

where Eff is one or the other efficiency measure discussed so far.

4.3.5 Two Stage Approaches

The most common approach is probably to leave out most complicating factors in a first analysis and then to examine in a second stage if the complicating factors and properties may contribute to explaining the variation in efficiency. The second stage may involve regressing the first stage results on the multiple complicating factors and properties. Also, it may involve building a non-parametric (DEA-like) model linking the factors and properties to the first stage efficiency scores.

The second stage models can then be used to correct the first efficiency measures by using a corrected efficiency measure:

$$\text{Efficiency}^{\text{FIRST STAGE}} / f(\text{Efficiency},\ z)$$

where $\text{Efficiency}^{\text{FIRST STAGE}}$ is the efficiency of the hospital in the first stage model and f(Efficiency, z) is the predicted first stage efficiency of a unit with complicating factors and properties z. This prediction is determined by the model estimated in the second stage. The advantage of this approach compared to a direct inclusion of the complicating factors in the first stage is once again to save degrees of freedom.

4.4 Quality in merger programs

4.4.1 General approach

Following the analyses above, the natural way to account for quality in the analyses of merger cases is to include quality in the underlying production economic models that we apply to determine the potential effects. That is, quality should be integrated in a DEA, an SFA or in any other representation we establish of the production possibilities. In this way the analyses of the likely effects of a merger or other form of horizontal or vertical integration will include an analysis of the possible quality effects.

This is in line with general production economics. Quality is a characteristic of processes and products, and it should ideally be included directly into the modeling like any other characteristic, e.g. time, location etc. The ideal solution is therefore to model using quality contingent products as opposed to ordinary product, cf. also the discussion above.

The idea of considering quality improvement in line with other managerial objectives can also be supported by applied management studies. The elements of successful quality improvement strategies (right culture, right people, in-house learning procedures and good IT tools) identified by Meyer ea. (2004), for example, suggest that there is considerable concurrence between the strategies used for quality improvement and the strategies used to improve the more traditional production economic efficiency of an organization.

Hence, the managerial tools are similar but of course quality improvements like other objectives require explicit managerial commitment. This can be illustrated by the Meyer ea. (2004) recommendation to employ "dashboard" indicators for every department with specific targets related to quality, service, people, and finances.

From an overall perspective, we would therefore *speculate* that a merger has potential to impact quality much the same way as it has potential to impact other characteristics, including quantity.

This implies that learning effects, harmony effects and size effects may impact quality.

The learning effect

Individual units, e.g. hospitals that under-perform on quality given the resources available, may improve by individual learning of best practices (protocols) and by incentive schemes rewarding quality.

A merger itself may also enhance the learning. This can happen for example if historical underperformance is due to a lack of managerial talents – and if "the best quality standards" are allowed or encouraged to dominate the new organization, cf. the discussion in Part B as to ex post efficiency. On the other hand, quality provision like other performance dimensions may suffer if the staff lacks commitment to a new organization or if the new managerial talents are not directed towards quality dimensions.

As discussed, a merger is also a change-event where historical procedures are re-evaluated and possibly improved. On the other hand, a merger may lead to declining efficiency in general and therefore declining quality since extra resources may be spend on the integration of unlike managerial cultures.

Harmony effects
In a horizontal integration, there are likely gains from being able to reallocate inputs and outputs so as to get more effective combinations of production factors and to get easier to produce service combinations.

In production economic models, it is customary to assume convexity of the production possibility set. This is a convenient assumption that is often well motivated but which in some contexts may seem artificial. In particular, convexity may conflict with gains from specialization or with (globally) positive economies of scale.

In a convex technology, positive harmony effects are the result of getting less specialized combinations of production factors and less specialized output requirements. In a non-convex technology on the other hand, the harmony effects may be negative suggesting that gains can be obtained by becoming more specialized.

When it comes to quality, we suggest that both situations are indeed possible. One can image that good quality requires an appropriate mix of production factors, e.g. nurses and doctors. Patients' satisfaction is a likely example. A unit with relatively few nurses may therefore benefit from harmonizing its production factors with a department with a relatively high number of nurses. When it comes to the services produced, one can also imagine that the marginal cost of some quality dimension, e.g. a marginal decline in mortality, is higher in one unit than in another. It may therefore be beneficial to harmonize and hereby reallocate the effort with an aim of equalizing the mortality rates. What can be saved in the high quality department by a small decline in survival rates may more than outweigh the cost of increasing survival in the low quality department.

In other situations, however, we would expect the quality dimensions to call for more specialized labor and more specialized services. The obvious example of this is that doing a large number of similar procedures tends to improve the skills of a health provider. That is, there may be economies of scale arguments that effectively limit the gains for producing a harmonious or balanced mix.

The exchange of quality related resources and services could be arranged in many ways as discussed earlier. It can be done inside a hierarchy, as it is likely to take place in a genuine merger, or it can be done with long-term contracts or a binding cooperation. The creation of genuine markets for inputs and outputs may also be an option in some cases. In principle, one could have procurement auctions where the state or an insurance company request bids for a certain set of services; say a basket with a certain set of DBC's.

Size
The effects of size may – like the effects of a more harmonious mix – have both positive and negative impacts on quality.

If we think of quality dimensions that are related to the routine of the providers, we are likely to find increasing return to scale favoring large scale entities and hereby mergers.

If we think instead of quality dimensions like patient satisfaction, risk of widespread infections etc, one may on the other hand expect decreasing return to scale. This would work against mergers.

4.4.2 Simpler approaches

Our general approach to quality is to model it as an integrated part of the production processes and hereby to capture the quality impacts of a merger like we would capture the quantitative impacts. The logic of this approach is that organizations spend time and effort on both the level and the quality of their production.

To understand the application of this general principle, we close by discussing the use of second stage adjusts, adjustment coefficients and factorization to capture quality aspects. Modeling quality using these approaches is conceptually inferior to the general contingent product approach, but they are in many situations more applicable given the limited data. It is therefore useful to comment on the conceptual ideas of these approaches and their usage in horizontal and vertical integration analyses.

Initially, we note that the prediction of the quality in a merged unit depends on the type of quality indicators we use – input, process or output oriented.

Input oriented measures are particularly easy to use in a merger case. In a merger, we know which resources are pooled and therefore we know also the resulting level of input quality - at least immediately after the merger.

The impact on process and output indicators is more difficult to predict – it requires hypotheses about what standards and cultures (e.g. openness about mistakes and willingness to share these, to learn and to change accordingly) will dominate the new organization. In such cases, we will have to somehow relate such qualities to other and easier to predict variable, including input and structural variables like scale, scope and size of specialties.

Second stage
A good starting point would be to collect quality indicators for a large number of hospitals and try – using econometrics and the benchmarking methods of this paper - to explain quality variations by factors like the scale and the scope of the health care provider, the size of the individual specialties and the time since major changes in the organization, for example.

Having done this one can use the relationships to evaluate a merger since this will have a direct and quantifiable impact on the same explanatory variables.

This approach can be used on all the principal quality dimensions, effectiveness, safety and responsiveness. As an example, consider effectiveness. Imagine that there is a positive relationship between effectiveness and the size of hospitals. Some possible explanations of this could be that larger hospitals have resources for systematic analyses of treatment effects and for a systematic change of treatment procedures in view of such analyses. Another explanation could be that larger hospitals can have more specialties of a certain size and that there are positive synergies across specialties, e.g. because many patients are in fact requiring multi-specialty treatment. We do not need to know exactly what the relationship is to make use of it – at least as long as there is a well-established relationship and as long at there is likely to be a direct or indirect casual effect from size to effectiveness. It may of course also be the case that size only has a positive impact until a certain level after which it has an adverse effect by coordination and motivation costs getting larger.

The size of the merged unit can be calculated directly and using an estimated second stage relationship like above, we can therefore predict the effectiveness of the merged unit.

Adjustment coefficients or factorization
Another practical approach is to use adjustments of physical outputs to reflect quality differences.

Consider for illustrative purposes a hospital that produces two outputs y_1 and y_2, and let us assume that we can summarize the quality q of the services as either H(igh) or L(ow). The situation can then be illustrated in output space as in Figure 6 below where we assume that all units and all production iso-quants are illustrated for the same level of inputs. We see that a unit like A is inferior since it produces too low outputs even disregarding its low quality. Exactly how to distinguish (decompose) the effects of too low physical production and too low quality provision is unclear since in the illustration, quality and quantity are competing properties of the output – if we forgo some outputs, we may increase quality and if we forgo quality we may increase the physical output levels. We can therefore say that it is lacking on either or both dimensions. The same situation is illustrated in the right panel of Figure 6, only now we assume that we have some more aggregate production and quality indices.

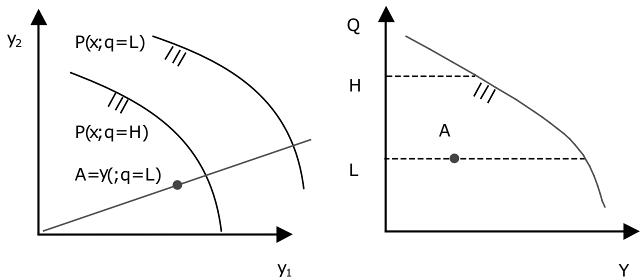

Figure 6: Simple quality models

It is important to note that if one observes the quantitative and qualitative properties of several hospitals, there might well be a positive correlation between quantity and quality. That is, hospitals that are good at producing large volumes may well be good at producing quality as well. Such phenomena are observed quite often – the economically most efficient farmers, for example, are also more environmentally efficient in many cases. In average terms, the two properties may therefore not look as if they compete.

The explanation however is most likely a slightly different one. The well-run hospitals may be better at ensuring both large volumes and good quality levels because they have managers and employees that are more skillful or work harder to create an efficient organization. If this were the explanation, we would still expect to see a trade-off between quality and quantity. If fact, this is often what we do. There is a trade-off between quality and quantity on the frontier as opposed to off the frontier - such that high quality comes at the cost of lower quantity and vice versa for given skills and effort. On the frontier, i.e. for the most efficient organization, there is still a need to allocate attention between the two objectives. At the general level we would therefore expect quantity and quality to compete.

Consider now the impact of mergers on the general quality measures like Q.

How would we expect the quality level of a horizontal integration to depend on the quality levels of the individual units P^1 and P^2 ? One possibility is that the qualities interact positively. Hence, everything equal, if we combine pre-merger qualities Q^1 and Q^2 we get at least the (weighted) average level for the integrated unit

$$(Q^1 + Q^2) / 2$$

If the integrated entity can gain from a larger scope of activities – e.g. when a patient requires considerable cross departmental effort – or the scale of activities - e.g. when the provider gets better training on a larger volume - the resulting quality level may also be above the average level and possible at least

$$\mathrm{Max}(Q^1, Q^2).$$

If we think of a vertical integration, we may expect similar resulting qualities. For some characteristics, the qualities supplied by P^1 and P^2 may be substitutes – a high quality diagnosis in P^1 may for example substitute for high quality screening in P^2. In that case, the integrated unit will have quality level

$$\mathrm{Max}(Q^1, Q^2)$$

In other cases, we might see complementary qualities with a resulting quality level of

$$\mathrm{Min}(Q^1, Q^2)$$

This could happen for example if the patients tend to remember the worst communication or if the treatments in P^2 are entirely dependent on the diagnosis done in P^1.

4.4.3 Applied predictions

To sum up, we see that quality is likely to be affected by mergers and that there are different ways to handle this in a pre merger analysis. The best approach depends on the nature of the quality indicator as well as the availability of information and the possibility of including quality information into the production economic models without running into a dimensionality problem.

The general approach is to look at qualities as other product characteristics, i.e. to include quality directly into the underlying frontier models. Ideally, we would use a contingent product approach but this is likely to lead to a curse of dimensionality problem and it may therefore be more realistic to use the other approaches outlined in Section 4.3

Instead of integrating quality directly in the quantitative production modeling, we can also work with add-on quality predictions.

Input indictor can be predicted directly.

Process and output indicators are probably best included in the merger predictions using simple methods like second stage analyses. In these we would determine a statistical relationship between key characteristics of health providers and their quality provision. We would focus on provider characteristics that we can predict in a merger case, e.g. the scale and scope properties of a merged entity. Having done this, we can take a specific merger case and determine its scale and scope and hereby its likely quality characteristics.

If we want to go further, we need to investigate the individual quality types in more details to determine if they are likely substitutes or complements. With this information we can then make informed prediction of the post merger quality intervals based on the pre merger qualities as described above.

5. Illustrative program application

The aim of the chapter is to give a first illustration of some of the models developed above. In view of available data, we shall concentrate on the horizontal integration of Dutch hospitals. Specifically, we have used cost and production data from 97 hospitals in 2006 to estimate best practice DEA and SFA models. Also, we have used physical distance information to determine all potential pairs of merges of two hospitals with a maximal distance of 10 km. The numbers of such mergers are 37. We have then for each of these pairs evaluated the total potential gains, and its decomposition into technical efficiency, harmony and size potentials.

We emphasize that the aim is illustrative and not to advocate a definitive model of Dutch hospital activities nor to analyze any particular merger case. To develop a more authoritative model, one would engage in more comprehensive analyses of a large set of alternative model specifications and tests for the inclusion and exclusion of different parameters. Here we have primarily been guided by the data available and the conceptual properties of the model. Also, we have not included any quality considerations in the models. Likewise, in the analysis of a specific merger case, one would both develop a good model but also analyze the potential merger gains in different directions of the input-output space and under different assumptions about uncertainty, controllability and transferability. Neither of this is attempted in this small illustration.

5.1 Data and basic models of the technology

The model we have used depicts a hospital as transforming costs into six output categories. The categories are closely related to the proposals in Varkevisser, Capps and Schut(2008) for how to define economically homogeneous specialty clusters, although a few adjustments has been necessary to align with the available data. In effects, then, the groups cover the following specialties:

Group	Specialty	Definition / Reference
1	Anesthesiology	ANAESTHESIOLOGIE
	Cardiology	CARDIOLOGIE
	Surgery	CHIRURGIE
	Dermatology	DERMATOLOGIE
	Internal medicine	INWENDIGE GENEESKUNDE
	Paediatrics	KINDERGENEESKUNDE
	Ophtalmology	OOGHEELKUNDE
	Orthopaedics	ORTHOPAEDIE
	Cosmetic surgery	PLASTISCH CHIRURGIE
	Radiology	RADIOLOGIE (RADIODIAGNOSTIEK)
2	Cardio-pulmonary surgery	CARDIO-PULMONALE CHIRURGIE
	Pulmonary medicine	LONGZIEKTEN
	Neurology	NEUROLOGIE
3	Geriatrics	GERIATRIE
	Clinical genetics	KLINISCHE GENETICA
	Neurosurgery	NEUROCHIRURGIE
	Radiotherapy	RADIOTHERAPIE
4	Gastroenterology	GASTRO-ENTEROLOGIE (MAAG-DARM-LEVER-ARTS)
	ENT (ear-nose-throat)	KEEL-, NEUS- EN OORHEELKUNDE
	Rehabilitation	REVALIDATIE
	Urology	UROLOGIE
	Gynaecology and obstetrics	VERLOSKUNDE EN GYNAECOLOGIE
5	Allergology	ALLERGOLOGIE
	Rheumatology	REUMATOLOGIE
6	Centers in hearing problems	AUDIOLOGISCHE CENTRA
	Psychiatry	PSYCHIATRIE

Table 2: Grouping of hospital output

For each of the groups we have looked at the total product of regulated DBCs. The value hereof, as evaluated with the DBC weights in use, are calculated to determine the group “turnover” at regulated prices at a given hospital (OMZET_LND_PRIJS)

$$y_J = \sum_{h=1} q_i \; p_i$$

where q_i = number of DBC_i produced, p_i = regulated price of DBC_i, and J = a group of DBCs. These turnovers are the outputs and cost drivers in our model of hospital service production

The use of such weighted combinations of underlying heterogeneous productions is common and a useful way to reduce the degrees of freedom in any estimation approach. It basically implies that we accept the intra-group calibration. Of course, one could continue like this and aggregate across the groups to give the total weighted output of a given hospital. This would mean that we also accept a priori the inter-group calibration implied by the DBC weights. Instead, we shall calibrate the relative importance of the cost drivers across groups using frontier models and the data available on total costs and total service productions from the 97 hospitals. A possible intermediate approach would be to add weight restrictions on the inter-group calibration. The inclusion of weight restrictions is commonly used in DEA models to reflect partial information.

On the cost side, we have tried different aggregated costs, most importantly estimated budgets (KOSTEN) and specialist cost, i.e. doctor salary except for academic hospitals (LUMPSUM). The results are not too sensitive to the choice of cost measure and we shall therefore only refer to the outcomes from using the sum of the two, denoted COSTS, below.

In effect then we have modeled the cost of producing the regulated products in Dutch hospitals in 2006. The A segment is regulated by the regulator setting DBC prices directly while in the B segment prices are set freely in the hospital - insurance company negotiations. Some special activities like emergency rooms and education are not included in any of these segments. We used data only about the A segment in the calculations.

Tables 3 and 4 below give a few summary statistics for the data in our sample.

Cost	KOSTEN	LUMPSUM	COST (SUM)
Average	107,955,361	18,881,781	124,890,566
Std.dev.	72,876,887	20,508,449	79,301,843
Min	21,074,264	0	23,598,146
Max	363,747,063	143,616,575	363,747,063

Table 3 Cost summary

Services	Y_1	y_2	y_3	y_4	y_5	y_6
Average	66,985,799	15,229,866	4,375,212	16,454,254	1,539,218	131,379
Std. dev	36,691,570	11,041,302	6,928,448	8,676,438	1,292,590	96,709
Min	8,115,414	1,771,211	1,542	1,803,759	128,730	1,944
Max	171,331,944	52,774,503	29,582,039	37,703,119	4,898,941	405,609

Table 4 Service summary

OLS estimation of average cost model
Using these data, we can estimate a linear average cost model as

$$\text{Cost}_h = \beta_O + \sum_{j=1}^{6} \beta_j y_{hj} + \varepsilon_h$$

where Cost_h is the (relevant part of the) cost of hospital h, y_{hj} is the

production level for group j in hospital h and ε_h is a random noise terms. A simple linear regression analyses give an adjusted R-squared of 0.8656, i.e. the regression is able to explain 86% of the variation of costs by the six cost-drivers in the model. The parameter estimates are given in Table 5 below.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
Estimate	1.36E+07	7.79E-01	1.66^{E}+00	3.06E+00	4.36E-01	6.37^{E}+00	8.93E+01
Std. dev	6.89E+06	2.77E-01	6.27E-01	5.86E-01	8.20E-01	3.32^{E}+00	3.70E+01
Err.t value	1.971	2.817	2.655	5.227	0.531	1.916	2.415
Pr(>|t|)	0.05184	0.00596**	0.00939**	1.11e-06***	0.59649	0.05849	0.01775*

Table 5 Average cost model

This simple regression suggests that it may be worthwhile to consider the weight calibration in the DBC system. In particular, it suggests that the weight in group 2, 3, 5 and in particular 6 may be set somewhat below the real costs while the other groups have slightly boosted values. We acknowledge of course that the cost and product definitions and the data set used here do not suffice to make final conclusions as to relevance of he actual DBC weights.

5.2 Individual efficiencies

We have next estimated a series of frontier models of the cost function, i.e. the costs, Cost, as a function of the 6 outputs or cost drivers, $y_1,..,y_6$:

$$\text{Cost} = C(y_1,\dots,y_6)$$

The model specification, i.e. the inputs and outputs defined, has been tested using both parametric (SFA) and non-parametric (DEA) frontier (best practice) approaches. In each class we have estimated a range of possible specifications to get an impression of the sensitivity of the results to the specification of the model. In the SFA framework, we have estimated linear, log-linear, translog, normed linear and normed loglinear specifications of the means structure and truncated normal distribution for the inefficiency error term. In the DEA framework, we have estimated using the scale assumptions CRS (constant), DRS (decreasing), NDRS (non-decreasing) and VRS (variable) returns to scale. Specific runs were also made with a bias-corrected DEA model, including confidence interval [c1, c2] for the bias corrected efficiencies. A summary of the preliminary results about Farrell input efficiencies E, here simply

$$E = \text{modeled minimal costs / actual costs}$$

for the sample data is provided in Table 6 below.

Model	E Average	E Std.Dev.	Efficient # (E≥1)	Worst E Min.
d_cols_far	0.547	0.276	1	-0.678
d_fdh_far	0.981	0.081	87	0.386
d_dea_far_vrs	0.887	0.137	34	0.227
d_dea_far_drs	0.865	0.151	32	0.136
d_dea_far_ndrs	0.848	0.133	17	0.227
d_dea_far_crs	0.825	0.141	15	0.136
d_dea_sup_far_vrs	0.934	0.286	34	0.227
d_dea_sup_far_drs	0.898	0.288	32	0.136
d_dea_sup_far_ndrs	0.900	0.270	17	0.227
d_dea_sup_far_crs	0.869	0.268	15	0.136
d_dea_far_vrs_biascorr	0.829	0.119	0	0.213
d_dea_far_vrs_biascorr_c1	0.751	0.115	0	0.191
d_dea_far_vrs_biascorr_c2	0.884	0.136	0	0.226
d_dea_far_crs_biascorr	0.768	0.125	0	0.124
d_dea_far_crs_biascorr_c1	0.722	0.116	0	0.112
d_dea_far_crs_biascorr_c2	0.819	0.140	0	0.135
d_dea_far_ndrs_biascorr	0.788	0.118	0	0.205
d_dea_far_ndrs_biascorr_c1	0.740	0.109	0	0.188
d_dea_far_ndrs_biascorr_c2	0.842	0.132	0	0.225
d_orderm_far	0.989	0.084	87	0.388
d_sfa_linear_far	0.739	0.169	2	0.141
d_sfa_loglinear_far	0.819	0.145	0	0.083
d_sfa_translog_far	0.831	0.138	0	0.336
d_sfa_normedlinear_vrs_far	0.618	0.222	6	0.008
d_sfa_normedlinear_crs_far	0.623	0.187	0	0.158
d_sfa_normedloglinear_far	0.818	0.144	0	0.082
d_dea_far_se	0.929	0.080	15	0.594

Table 6; Comparison of frontier models

As we can see from the summary of different estimations, the level of cost inefficiency (1-E) in the Dutch hospital sector is 10-20 percent in most specifications. The interpretation of this is that if everyone learned best practice the total costs could be reduced with 10-20 % without changing the organization of the sector.

The scale inefficiency is approximately 7% in the DEA models suggesting that some 7% could be saved if everyone adjusted to optimal scale size.

Although our aim is not to develop an authoritative cost model of Dutch hospitals, a few notes on these levels are useful. As a first quantification of cost inefficiency, and by comparison with other sectors, the estimated cost inefficiency is not alarming. Truly, the results suggest considerable possibilities to save due to the large underlying costs in euro, but in relative terms, one finds similar saving potential in many other sectors, both regulated and more competitive. It should on the other hand be observed that this level of estimated inefficiency may also reflect the way the DBCs are priced. Since they are intended to reflect actual costs and since there are much more DBCs than cost pools (hospitals), the DBC prices can easily be set to make everyone look efficient.

5.3 Potential gains from mergers

We have used the DEA-crs model to evaluate the 37 potential mergers of pairs of hospitals with a maximal distance of 10 km.

The summary statistics of overall potential gains E^J and its decomposition into learning effects TE^J, gains after individual learning E^{*J}, harmony effects H^J and size effects S^J is reported in Table 7 below. Recall that

$$E^J = TE^J * E^{*J} = H^J * S^J$$

and that the size effect S^J is 1 (corresponding to no gains from resizing) when we presume constant return to scale. The calculation and interpretation of these measures are discussed at length in Part B.

	E	E*	TE	H	S
Average	0.82	0.97	0.84	0.97	1.00
Std.Dev.	0.08	0.03	0.07	0.03	0.00
Max	1.00	1.00	1.00	1.00	1.00
Min	0.64	0.88	0.69	0.88	1.00

Table 7: Potential gains from mergers in DEA-CRS model

At an overall scale, we see that the average saving potential in the 37 mergers is 18% (1 – E = 1 – 0.82). Indeed, in the detailed results 17 out of the 37 pairs has an improvement potential of more than 20% and 32 out of 37 can save more than 10%.

An important part of this, namely 16% is learning potentials. Some of the learning potentials can no doubt be activated by benchmarking across hospitals, and by developing better incentive schemes relying for example on cross hospital relative performance evaluations. A merger may however also have a positive effect on learning by increasing the scale of process development and by being a change event where past procedures are re-evaluated and changed.

Ignoring the learning effect, however, the average saving potential is only 3 % (the harmony saving) and only 9 out of 37 (or about 24 % of the mergers) can generate a saving of more than 5% by simply reallocating resources and tasks. Again, this is theoretically possible without a genuine mergers, e.g. by creating inter-hospital DBC markets, but the reallocation of resources and tasks may be easier inside a merged hospital where asymmetric information and the fight over profit shares may be less.

This results suggest that the underlying estimated technology is rather linear, i.e. that not only do we have constant return to scale by assumption, but also output iso-quants that are rather linear corresponding to approximately constant rates of substitution between the outputs. This is not entirely surprising; the linear SFA model gives average efficiencies that are quite similar to the DEA models suggesting that the inability to have curved iso-quants in this technology does not lead to too much deviations of the actual performance to the estimated best practice frontier.

If we assume instead a VRS technology, the corresponding results are given in Table 8 below

	E	E*	TE	H	S
Average	1.00	1.12	0.89	0.93	1.20
Std.Dev.	0.26	0.22	0.08	0.06	0.21
Max	1.94	1.94	1.00	1.00	1.94
Min	0.72	0.95	0.72	0.83	0.99

Table 8: Potential gains from mergers in DEA-VRS model

In the VRS calculations, several mergers lead to LP problems with no solutions. The explanation is that when two hospitals are merged they will in many cases become rather large compared to the existing hospitals (with similar mix of resources and services) and consequently be above the estimated optimal scale size for this mix. In that cases, the existing best practice does not even show that the resulting production plans are feasible.

If we believe firmly in the estimated VRS technology, the interpretation is that it will be impossible to operate hospitals of that size. Or, in the case where a solution is found but the score is above 1, that it will be more costly to operate the hospitals jointly than individually. In one concrete merger for example we can find that the estimated net effect is a cost increase of some 19 %. This cost increase is the result of three effects. First, since the underlying units are technically inefficient, there is a learning potential of 12 %. Also, by reallocating resources and services, some 2% can be saved. The return to scale however is rather unfavorable to this merger with corresponding to a cost increase of 38%. The net-effect – when correcting for the fact that these different effects are multiplicative and not additive - is a cost increase of 19%.

Another and more likely explanation of these findings is of course that the estimated technology is flawed or at least heavily biased for large units. The bias of the DEA estimated technology is well-know; DEA makes a conservative (cautious) inner approximation of the production possibility set and in the parts of the production space which is more sparse on observations, this bias is larger. Hence, if there are only few large units comparable to the size of a merged one, the best practice model is most likely too pessimistic – and more so than in other areas of the production space. This may explain the rather modest improvement potentials identified in the VRS case. (We could of course do similar corrections of the technology above.)

Even more fundamentally, one may of course question the VRS assumption using similar reasoning in a theoretical framework: A large entity must be able to do at least as well as any two smaller units that it could be decomposed into since it could simple (re-) organized as two independently run divisions.

This suggests that we should either use the bias corrected technology or the irs technology – or both. Alternatively, we could make parallel evaluations using SFA estimate models. The results (summary statistics) of doing this is shown below.

	E	E*	TE	H	S
Average	0.82	0.96	0.85	0.96	0.99
Std.Dev.	0.08	0.03	0.06	0.03	0.01
Max	1.00	1.00	1.00	1.00	1.00
Min	0.64	0.88	0.71	0.88	0.95

Table 9 Potential gains from mergers in DEA-IRS model

We see Table 9 that the irs case gives quite similar results on average as the crs case. The saving potential from individual learning is 15%, from scope (harmony) 4% and from scale (size) 1%. It is interesting to see that even if we acknowledge the possibilities of small units being disadvantaged by their scale, the gains from the merged units operating at larger scale is in general limited and only about ¼ of the gains from better economies of scope.

	E	E*	TE	H	S
Average	0.73	0.96	0.76	0.97	0.99
Std.Dev.	0.06	0.03	0.05	0.03	0.01
Max	0.89	1.00	0.89	1.00	1.00
Min	0.58	0.89	0.63	0.89	0.95

Table 10: Potential gains from mergers in Bias Corrected DEA-IRS model

We see from Table 10 that the bias correction increases the overall potentials to improve, E, but that it is in general the learning affect that picks up all the changes in the cost frontier. The scope (harmony) and scale (size) effects are largely unchanged.

In the vrs case the impact is also mainly in the learning effect although the negative impact is also lowered a little as expected. Only under special circumstances does the boot-strapping eliminate the LP (no-solution) problem. This is illustrated in Table 11 below.

	E	E*	TE	H	S
Average	0.892	1.107	0.802	0.933	1.182
Std.Dev.	0.23	0.22	0.07	0.06	0.21
Max	1.73	1.94	0.92	1.00	1.94
Min	0.65	0.94	0.65	0.82	0.99

Table 11: Potential gains from mergers in Bias Corrected DEA-VRS model

To illustrate the parametric approach, consider the log-linear model. Since we are estimating a cost function and not a production function, the log linear specification conflicts with the usual convexity properties, i.e. the set T may not be convex. Rather, the log-linear specification allows for gains from specialization as well as potentially genuine global economies of scale. As a supplement, these properties can be interesting to allow for. Assuming a truncated normal inefficiency distribution (with underlying mean μ) and normal distributed noise, the maximum likelihood estimates are as shown in Table 12 below.

	Coefficient	Std.Dev	t-Ratio
beta_0	0.696	0.923	0.754
beta_1	0.651	0.141	4.608
beta_2	0.204	0.089	2.297
beta_3	0.006	0.006	1.048
beta_4	0.149	0.103	1.456
beta_5	0.000	0.005	0.011
beta_6	0.005	0.004	1.198
sigma-squared	0.476	0.105	4.538
gamma	0.985	0.007	142.700
mu	-1.370	0.313	-4.376

Table 12: Loglinear parametric function

It is worthwhile to note that the sum of the beta 1 to 6 coefficients is 1.01 suggesting a more or less constant return to scale technology.

Using the log-linear specifications we can calculate and decompose the gains from mergers as in Table 13 below.

	E	E*	TE	H	S
Average	0.77	1.03	0.75	1.02	1.01
Std.Dev.	0.11	0.02	0.11	0.02	0.00
Max	1.00	1.10	0.92	1.09	1.01
Min	0.51	1.01	0.48	1.00	1.01

Table 13: Potential gains from mergers in loglinear SFA model

We see that the log-linear model is suggesting that the economies of scale are largely neutral to the mergers as are the economies of scope. In the loglinear specification, even the scope economies are speaking (slightly) against the mergers corresponding to a cost increase of 2% on average. The log-linear model suggests that the gains are primarily from learning.

The lack of gains from larger scale, in many cases even losses from the merged units operating at larger scale, has been a consistent finding in the models above. Of course, it should be emphasized also that our analyses build on existing practices only. If a new merger leads to new facilities and new organizations that have not been implemented in other hospitals in the data set, the estimated models cannot capture the potentials that these may generate. This would require a much more detailed organizational and engineering approach. The network approach of this project could potentially be developed in this direction. Specifically, if one can define hospital processes and allocate not only activities but also costs to these processes, it is possible to create new pseudo-observations by making new combinations of old processes as explained in Section 3.2 above.

We note also that the Spearman correlation between individual efficiencies calculated in the loglinear model and the DEA-irs model is 0.62 while it is 0.68 in relation to the DEA-irs bias corrected model and 0.68 in relation to the DEA-crs model. In general, then, these models suggest correlated although not perfect agreement in the individual evaluations. This illustrates what was also emphasized in the introduction to this Chapter: Our aim is not to suggest an authoritative cost models for Dutch hospitals but rather to illustrate the merger analyses. In full-scale real applications, therefore, it is important to develop good underlying production and/ or cost models of the technology in place. It is likely however that even after such efforts, there may be several, reasonable model candidates. The best approach in this situation may be to evaluate the merger gains in the different models – as we have done here - and to look at the results as interval estimates established in this way.

6. Extensions

In this research project on Efficiency Gains from Mergers, we have covered a large territory. We have moved from purely economic theory over measurement methods to actual program implementations and numerical illustrations. We have covered not only simple horizontal mergers but vertical and network situations as well. Moreover, we have discussed full-fleshed mergers as well as alternative forms of integration like simple learning of best practices, sharing of production factors, and in- and out-sourcing of selected services. The total potential gains of integrations have been decomposed accordingly.

There are of course a series of relevant extensions of this research.

It is important to further investigate the horizontal integration case using the existing and other data sets. It is in particular relevant to investigate how sensitive the merger gains and their decompositions are to the underlying estimation technique (DEA, SFA, Bias corrected DEA etc) as well as to the assumptions about post-merger technical inefficiency.

It is important to test the vertical and network models also. This can be done in due time by combining for example the hospital data with data on up- or downstream parts of the Dutch heath care sector. An obvious candidate is the long-term care sectors for which NZa presently are establishing data sets.

It would also be important to make separate investigations of the relationship between key quality indicators and the characteristics of the different hospitals. For the merger analyses it would be particularly important to try to explain quality from key structural characteristics of the hospitals like the scale, output scopes and inputs mixes since this would allow for rather simple predictions of the impact of mergers on quality.

We also suggest that it would be useful to apply the framework of this project to make sector wide investigations of the structural efficiency of the Dutch health care. This could be done by more systematic analyses of the potential gains from reallocating resources and services among providers that are horizontally or vertically close. Also, in many cases, accessibility would suggest geographical restrictions on the transferability of resources and services much like in the illustrative program application of this paper.

Another extension of this project would be to combine the methods developed here to capture the potential gains with industrial organization models of the market impact of mergers. Specifically, it would be interesting to compare estimated potential changes in marginal costs from a merger with thresholds of the so-called compensating marginal cost decreases, i.e. the changes in marginal costs that would make a less competitive post-merger sector with lower marginal costs produce at least the same services as a pre-merger more competitive sector with higher marginal costs.

Taking this step even further, an interesting but also more intense project could be to apply the frontier models of this project to also describe the market competition and hereby to use the same set of approaches to evaluate both the cost and competitive effects of mergers.

7. References

Aigner, D.J. and S.F. Chu (1968), On Estimating the Industry Production Function, *American Economic Review*, 58, pp. 826-839.

Aigner, D.J., C.A.K. Lovell and P. Schmidt (1977), Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics*, 6, pp. 21-37.

Andersen, J. and P. Bogetoft, Gains from Quota Trade: Theoretical models and an application to the Danish Fishery, *European Review of Agricultural Economics*, 34 (1), 105-127, 2007.

Banker, R.D., Estimating Most Productive Scale Size Using Data Envelopment Analysis, *European Journal of Operational Research*, 17, 35-44, 1984.

Banker, R.D., A. Charnes and W.W. Cooper, Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, 30, 1078-1092, 1984.

Banker, R.D. and Morey, E.C, Efficiency analysis for exogenously fixed inputs and outputs, *Operations Research*, 34, 513-521, 1986.

Battese, G.E. and T.J. Coelli (1992), Frontier Production Functions, Technical Efficiency and Panel Data: With Applications to Paddy Farmers in India, *Journal of Productivity Analysis*, 3, pp. 153-169

Bogetoft, P., DEA on Relaxed Convexity Assumptions, *Management Science*, 42, 457 – 465, 1996.

Bogetoft, P. , K. Boye, H. Neergaard-Petersen and K. Nielsen, Reallocating sugar beet contracts: can sugar production survive in Denmark. *European Review of Agricultural Economics*, 34(1), 1-20, 2007.

Bogetoft, P. , J.M. Bramsen and K. Nielsen, Balanced Benchmarking, *International Journal of Business Performance Management*, 8(4), 274-289, 2006.

Bogetoft, P. R. Färe and S. Grosskopf, Thought about Network DEA, WP, OSU, 2008.

Bogetoft, P., and T.E. Grammeltvedt , Mergers in Norwegian Electricity Distribution: A Cost Saving Exercise? Working Paper, BI, Oslo, 2006.

Bogetoft, P. and K. Nielsen, Internet Based Benchmarking, *Journal of Group Decision and Negotiation*, 14 (3), 195-215, 2005.

Bogetof, P. and J. Tind, Process Aggregation and Efficiency, WP, Copenhagen University, 2006.

Charnes, A., W.W. Cooper and E. Rhodes, Measuring the Efficiency of Decision Making Units*, European Journal of Operational Research*, 2, pp. 429-444, 1978.

Charnes, A., W.W. Cooper and E. Rhodes, Short Communication: Measuring the Efficiency of Decision Making Units*, European Journal of Operational Research*, 3, 339, 1979.

Charnes, A., W.W. Cooper, A.Y. Lewin and L.M. Seiford (eds) (1944). *Data Envelopment Analysis: Theory, Methodology, and Application*, Klüwer Academic Publishers, 1994.

Coelli, T., D.S.Prasada Rao, and G. Battese, An Introduction to Efficiency and Productivity Analysis, Kluwer Academic Publishers, 1998.

Cooper, W.W., L.M.Seiford, and K.Tone, Data Envelopment Analysis, Kluwer Academic Publishers, 2000.

Deprins, D., D. Simar, and H. Tulkens, *Measuring Labor Efficiency in Post Offices,* 243 - 267 in M. Marchand, P. Pestieau, and H. Tulkens, "*The Performance of Public Enterprises: Concepts and Measurements*", North Holland, 1984.

Färe, R. and S.Grosskopf , Intertemporal Production Frontiers: With dynamic DEA, Kluwer Academic Publishers, Boston, 1996.

Färe, R. and S.Grosskopf, Network DEA, Socio-Economic Planning Sciences, 35-49, 2000.

Färe, R., S. Grosskopf, C.A.K. Lovell and S. Yaisawatng, S., Derivation of shadow prices for undesirable outputs: A distance function approach, Review of Economics and Statistics, 75, 374-380, 1993.

Färe, R., S. Grosskopf, M. Lundström and P. Roos, Evaluating Health Care Efficiency, Scientific Report 1: 2007, R.R. Institute of Applied Economics, 2007

Färe, R. and D. Primont, Multi-Output Production and Duality: Theory and Applications, Kluwer Academic Publishers, Boston, 1995.

Farrell, M.J., The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, 120, 253 – 281, 1957

Farrell, J. and C. Shapiro, Horizontal Mergers: An Equilibrium Analysis, *American Economic Review*, 80, 107 – 126, 1990.

Greene, W.H. (1990). 'A Gamma-Distributed Stochastic Frontier Model." *Journal of Econometrics* 46, 141-164

Jacobs, R., P.C. Smith and A. Street, Measuring Efficiency in Health Care, Cambrigde University Press, Cambridge, UK, 2006.

Land Kenneth C, C. A. Knox Lovell and Sten Thore: Chance-Constrained Data Envelopment Analysis*, Managerial and Decision Economics*, Vol. 14, No. 6 (Nov. - Dec., 1993), pp. 541-554

Lovell, C.A.K. (1993) Production Frontiers and Productive Efficiency*, in: H. Fried, C.A.K. Lovell, S. Schmidt (eds) The Measurement of Productive Efficiency: Techniques and Applications*, New York, Oxford University Press

Meeusen, W. and J. van den Broeck (1977), Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 18, pp. 435-444.

Meyer, J.A., S. Silow-Carroll, T. Kutyla, L.S. Stepnick, and L.S. Rybowski (2004), Hospital Quality: Ingredients for Success – Overview and Lessons Learned, Publication 761, The Commonwealth Fund, New York.

OECD (2006), Health Care Quality Indicators Project Conceptual Framework Paper, OECD Health Working Papers, Nr. 26.

Olesen, O. B. and N. C. Petersen (1995). Chance constrained efficiency evaluation. *Management Science* 41(3): 442-457.

Olesen, O.B. and N.C: Petersen, Target and Technical Efficiency in DEA – Controlling for Environmental Characteristics, Working Paper, The University of Southern Denmark, 2007.

Paradi, J.C., S. Vela and Z. Yang (2004), Assessing Bank and Bank Branch Performance: Modeling Considerations and Approaches*, in Cooper, W.W., L. M. Seiford and J. Zhu (eds), Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, Boston.

Perry, M.K. and R.H. Porter, Oligopoly and the Incentives for Horizontal Merger*, American Economic Review*, 75, 219 –227, 1985.

Petersen, N.C., Data Envelopment Analysis on a Relaxed Set of Assumptions*, Management Science*, 36, 305-314, 1990.

Ray, S.C., Data envelopment analysis; Theory and techniques for economics and operations research, *Cambrigde Unversity Press*, Cambridge, 2004.

Ruggiero, J., On the measurement of technical efficiency in the public sector, *European Journal of Operations Research,* 553-565, 1996.

Salant, S., S. Switzer and R. Reynolds, Losses From Horizontal Merger: The Effects of an Exogenous Change in Industry Structure on Cournot-Nash Equilibrium, *Quarterly Journal of Economics*, 98, 185-199, 1983.

Simar, L. and P. Wilson, A general methodology for bootstrapping in nonparametric frontier models, *Journal of Applied Statistics*, 27, 779-802, 2000.

Stubbe J.J., W. Brouwer and D.M.J. Delnoij (2007), Patients' experiences with quality of hospital care: the Consumer Quality Index Cataract Questionnaire, *BMC Ophthalmology*, 7:14.

Tulkens, H., On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit*, Journal of Productivity Analysis*, 4, 183 – 210, 1993.

Van Bergeijk, P.A.G. and E. Kloosterhuis(eds), Modelling European Mergers: Theory, Competition Policy and Case Studies, Edward Elgar Publishing, USA, 2005.

Varkevisser, M., Capps, C.S., Schut, F.T.: Defining hospital markets for antitrust enforcement: new approaches and their applicability to The Netherlands, Health Economics, Policy and Law (2008), 3: 7-29

Wan, T.T.H., D. Y-J Lin, and A. Ma (2002), Integration Mechanisms and Hospital Efficiency in Integrated Health Care Delivery Systems*, Journal of Medical Systems*, Vol 26, pp. 127-143, 2002