

Effektivitetsanalyse av somatiske sengeposter ved Universitetssykehuset Nord-Norge HF

- en prestasjonsmåling ved bruk av DEA og SFA

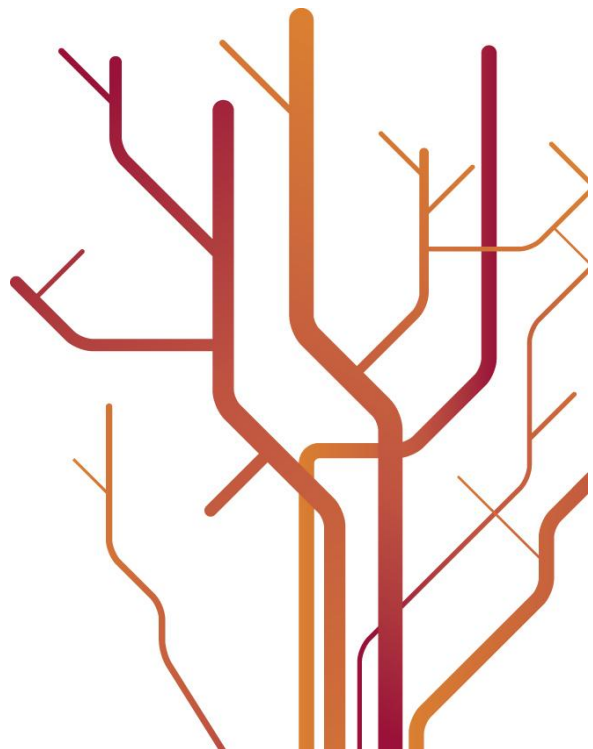


Morten Espejord

Helge Jenssen Lægreid

Masteroppgave i økonomi og administrasjon
-studieretning bedriftsøkonomi (30 stp)

Juni 2012



Forord

Valg av oppgave kom etter et initiativ fra økonomiledelsen ved Universitetssykehuset Nord-Norge som ønsket å få belyst eventuelle problemer innad i organisasjonen. Oppgaven er en obligatorisk del av mastergradstudiet i økonomi og administrasjon, og er et resultat av vår interesse for benchmarking. Oppgaven markerer avslutningen av studiet.

Vi ønsker å rette en stor takk til ledelsen og de ansatte ved Økonomi- og analysesenteret ved UNN, da spesielt Stig Bakken, for å ha bidratt med data og forslag gjennom hele prosessen.

Vi vil rette en stor takk til vår veileder, professor Terje Vassdal, for gode innspill og faglig diskusjon. Vi ønsker også å takke stipendiat Sverre Braathen Thyholdt for nyttige innspill.

Til å løse modellene og svare på våre problemstillinger har vi brukt DEA-Frontier (Zhu) i kombinasjon med salterprogrammet til Dag Fjeld Edvardsen og programmeringspakken R versjon 2.14.2 med tilleggspakkene Benchmarking(Bogetoft og Otto) og FEAR(Wilson og Walker). I tillegg benytter vi oss av Shazam Professional edition og STATA for å gjennomføre alle økonomiske beregninger.

Til slutt vil vi takke hverandre for et meget bra samarbeid gjennom studietiden, og vi må også rette en takk til våre kontorkamerater Martin Skorstad og Marcus Dons for samarbeidet gjennom hele studieperioden.

Tromsø, Mai 2012

Morten Espejord

Helge J. Lægreid

Sammendrag av oppgaven

I oppgaven analyseres effektiviteten ved de somatiske sengepostene ved UNN for årene 2010 og 2011. Analysen foretas på enhetsnivå, og fokuserer først og fremst på den indre effektiviteten, altså hvor godt sengepostene utnytter sine ressurser. I oppgaven prøver vi også å finne sammenhenger mellom effektivitet og variabler som ikke inngår direkte i sengepostenes produktfunksjon.

Opgaven baserer seg primært på *DEA* (data envelopment analysis) som verktøy for å måle effektivitet. Som et supplement til disse beregningene benyttes *SFA* (stokastisk frontanalyse) for å støtte opp om de primære resultatene. Begge metodene rapporterer effektiviteten som et enkelt tall i intervallet $[0,1]$ (der 1 er effektiv). Valg av modell og variabler testes ved hjelp av Banker-tester. Datasettet siles også ved hjelp av to ulike outlier-metoder for å lete etter enheter som bør fjernes.

Begge modellene måler effektiviteten basert på en ”lett” justert sykehusmodell. Variablene som inngår i modellen er *lønnskostnader*, *andre driftskostnader*, *sengekapasitet* og *justerte DRG-poeng*.

Resultatene våre viser at den gjennomsnittlige effektiviteten for 2010 ligger i intervallet 64-80 prosent, mens den i 2011 ligger i intervallet 65-81 prosent. Individuelt finner vi to effektive sengeposter når vi forutsetter konstant skalautbytte og fem når vi antar variabelt skalautbytte. *SFA*-estimatene viser seg å ligge en plass i mellom *DEA*-estimatene. De rapporterte effektivitetene viser et stort potensial for kostnadsbesparelse. Av variablene som ikke inngår direkte i produktfunksjonen ser det ut til at beleggsgrad påvirker effektiviteten negativt mens kompleksiteten hos pasientene ser ut til å påvirke effektivitet positivt.

Nøkkelord:

Benchmarking, effektivitet, DEA, SFA, sykehus, sengeposter

Innholdsfortegnelse

Forord	iii
Sammendrag av oppgaven.....	iv
TABELLER.....	vii
FIGURER	viii
1 Innledning.....	1
1.1 Bakgrunn.....	1
1.2 Avgrensing og begrepsdefinisjoner	1
1.3 Problemstilling.....	2
1.4 Oppgavens struktur	4
2 Helsesektoren.....	6
2.1 Sykehusenes inntekter og finansiering.....	8
2.2 Om UNN.....	12
2.3 Om sengeposter.....	13
2.4 Oppsummering.....	13
3 Litteratursammendrag.....	15
3.1 Effektivitetsundersøkelser ved hjelp av parametriske metode	17
3.2 Undersøkelser ved hjelp av ikke-parametriske metoder	18
4 Teori og metode.....	23
4.1 Effektivitet og produktivitet.....	23
4.1.2 Teknisk effektivitet	24
4.2 Distansefunksjoner.....	24
4.3.1 Måling av teknisk effektivitet	26
4.3 Data envelopment analysis	27
4.3.1 CCR-modellen.....	28
4.3.2 Omhyllingsmodellen	29
4.3.3 BCC-modellen.....	31
4.4 Outliers.....	33
4.4.1 Wilson-rutinen.....	34
4.4.2 Supereffektivitet	35
4.5 Stokastisk frontanalyse	36
4.6 Sensitivitet og robustheten til effektivitetsestimater	38
4.7 DEA og hypotesetester	39
5 Datasett - input, output og andre variabler	41

5.1 Datasettet.....	42
5.1.1 Valg av variabler	43
5.2 Input	44
5.2.1 Faste innsatsfaktorer.....	46
5.3 Output	46
5.4 Kontrollvariabler.....	50
5.5 Valg av modell.....	52
5.6 Outliers.....	54
6 Resultat	57
6.1 Sengepostenes effektivitet 2010	58
6.2 Sengepostenes effektivitet 2011	60
6.4 Stokastisk frontanalyse – 2010 og 2011	62
6.5 Sensiviteten og robustheten til effektivitetsestimaterne.....	65
6.6 Testing av hypotesene.....	66
6.6.1 Regresjon 2010.....	67
6.6.2 Regresjon 2011	68
6.7 Effektivitet og sammenhenger	70
6.8 Sammenlikning av analysemetodene	72
6.8.1 Sammenligning mot tradisjonelle mål	75
7 Oppsummering og konklusjon.....	77
7.1 Oppsummering.....	77
7.2 Konklusjon.....	79
7.3 Veien videre	80
8 Litteraturliste	81
Vedlegg.....	84
Vedlegg 1 Eksempeldata.....	84
Vedlegg 2 Spesifikasjonstest 1	84
Vedlegg 3 Spesifikasjonstest 2	86
Vedlegg 4 Skalatest.....	88
Vedlegg 5 Datasett 2010 og 2011	89
Vedlegg 6 Heteroskedastisitet, SFA og Wald-test.....	90

TABELLER

Tabell 1: Kliniske avdelinger og seksjoner ved ulike sykehustyper	7
Tabell 2: Deskriptiv statistikk av antall DRG og HDG ved sengepostene	10
Tabell 3: Eksempel på DRG.....	10
Tabell 4: Eksempel kostnadsfordeling	11
Tabell 5: Eksempel på justering av DRG-poeng.....	48
Tabell 6: Deskriptiv statistikk av sengepostenes pleie -og grunnandel	49
Tabell 7: Deskriptiv statistikk input og output 2010.....	50
Tabell 8: Deskriptiv statistikk input og output 2011	50
Tabell 9: Korrelasjon mellom input, output og kontrollvariabler 2011	52
Tabell 10: Urban Hospital Model.....	52
Tabell 11: Vår modell	53
Tabell 12: Oversikt over hvilke enheter som fjernes i hvert steg 2010.....	55
Tabell 13: Oversikt over hvilke enheter som fjernes hvert steg 2011	55
Tabell 14: Supereffektivitet.....	56
Tabell 15: Effektivitet 2010	58
Tabell 16: Mål på optimal reduksjon 2010	59
Tabell 17: Effektiviteter 2011	60
Tabell 18: Mål på optimal reduksjon 2011	61
Tabell 19: SFA-data 2010	62
Tabell 20: SFA-data 2011	62
Tabell 21: SFA-effektiviteter	64
Tabell 22: Sensitivitetsestimater basert på effektivitetsstiger	65
Tabell 23: Regresjonsresultater CRS 2010	67
Tabell 24: Regresjonsresultater VRS 2010	67
Tabell 25: Regresjonsresultater SFA 2010.....	67
Tabell 26: Korrelasjon mellom kontrollvariablene 2010	68
Tabell 27: Regresjonsresultater CRS 2011	68
Tabell 28: Regresjonsresultater VRS 2011	69
Tabell 29: Regresjonsresultater SFA 2011.....	69
Tabell 30: Korrelasjon mellom kontrollvariablene 2011	69
Tabell 31: Korrelasjon mellom effektivitetsestimaterne	75
Tabell 32: Korrelasjon mellom CRS 2011 og tradisjonelle prestasjonsmål	76

FIGURER

Figur 1: Distansefunksjon	26
Figur 2: Eksempel på en effektivitetsfront	32
Figur 3: Outlieranalyse med medfølgende analyseobjekter som fjernes i hvert steg.....	34
Figur 4: Eksempel på effektivitetsstige	39
Figur 5: Outlieranalyse 2010 og 2011	55
Figur 6: Salterdiagram med total kostnad og effektiviteter.....	73
Figur 7: Tetthetsfunksjonen til 2010-estimatene.....	74
Figur 8: Tetthetsfunksjonen til 2011-estimatene.....	74

1 Innledning

1.1 Bakgrunn

Det stilles stadig større krav til effektivitet og produktivitet i offentlig sektor. Helsesektoren er intet unntak, og det foregår kontinuerlige prosesser for å forbedre måten sektoren styres på og hvordan ting blir gjort på institusjonsnivå. Kostnadene forbundet med helse øker årlig mer enn den generelle inflasjonen, hvilket kan tyde på at helsesektoren ikke drives *effektivt* nok. Det kan imidlertid være andre årsaker til dette. Nye og kostnadskrevede behandlinger og medisiner har forbedret folkehelsen, samtidig som levealderen i befolkningen forlenges. Den norske demografien viser at befolkningen i gjennomsnitt blir eldre, noe som kan føre til økte sykehuskostnader. Sykehusene måles også kontinuerlig på kvaliteten i behandlingstilbudet; det handler om å *gjøre de rette tingene* på en best mulig måte. Tidligere undersøkelser viser at sykehusene har forbedringspotensial både når det gjelder kapitalutnyttelse og kvaliteten på helsetjenestene. Når kostnadene øker er det interessant å se på hvordan sykehusene utnytter sine ressurser, og dette ønsker vi å belyse i oppgaven vår.

Etter initiativ fra Universitetssykehuset i Nord-Norge, har vi valgt å analysere effektiviteten ved helseforetakets døgnbemannede somatiske sengeposter. Sengepostene er et interessant utgangspunkt for en slik analyse da aktiviteten her kan gjøres sammenlignbar på tvers av pasientkarakteristika. Både i Norge og internasjonalt, har helsesektoren vært gjenstand for flere ulike typer effektivitetsanalyser. Uavhengig av om undersøkelsene fokuserer på forskjeller i effektiviteten ved pleiehjem, hjertekirurgiske avdelinger ved ulike sykehus eller effektene av prospektiv finansiering, er analysene nesten utelukkende gjennomført på sykehusnivå.

1.2 Avgrensning og begrepsdefinisjoner

Effektivitet og produktivitet er begreper det er lett å blande sammen. Som nevnt innledningsvis er studien vår en effektivitetsanalyse, hvilket vil si at vi rangerer de ulike sengepostene ved sykehuset etter hvor effektivt de utnytter innsatsfaktorene. For en institusjon som et sykehus er det viktig å yte best mulig, gitt de ressursene man rår over. Ytelsen er en blanding av *efficiency* og *effectiveness* (Ozcan, 2008). Effektivitet er ofte assosiert med å ”*gjøre tingene rett*”. Slik er også *efficiency* definert: det handler om å bruke minst mulig innsatsfaktorer i produksjonen av en gitt mengde output. For helsesektoren betyr

det at det produseres en tilstrekkelig kvalitetssikret helsetjeneste ved å bruke minimalt med ressurser. *Effectiveness* har en litt annerledes assosiasjon: ”Å gjøre de rette tingene”. Når man måler *effectiveness* ser man på helsetjenestens bidrag til folkehelsen: En friskere befolkning, lavere dødelighetsrate og lengre levetid er indikasjoner på høy *effectiveness*. *Efficiency* og *effectiveness* vil påvirke hverandre. *Effectiveness* er et spørsmål om hvorvidt valgene av innsatsfaktorer er optimale i forhold til produksjon av helsetjenesten. *Efficiency* vil her påvirkes negativt dersom man ikke velger den ideelle inputsammensetningen. På samme måte kan dårlig utnyttelse av innsatsfaktorer gi negative utslag på *effectiveness*.

Effektiviteten kan deles inn i to kategorier: Indre og ytre. Den indre effektiviteten er videre delt inn i teknisk- og kostnadseffektivitet, mens den ytre kalles allokeringseffektivitet. Den indre effektiviteten kan sammenstilles med *efficiency* og den ytre med *effectiveness*. Vi forutsetter at UNN utnytter ressursene sine i samsvar med de gjeldende helsepolitiske mål og måler dermed ikke allokeringseffektiviteten. Har man gode indikatorer for helsevesenets kvalitet, vil den ytre effektiviteten kunne måles.

Mange undersøkelser fokuserer på effektivitetsforskjeller mellom sykehus på både internasjonalt og nasjonalt nivå. Vi har avgrenset området vi ønsker å undersøke til den somatiske behandlingen ved et enkelt helseforetak. Somatisk stammer fra gresk og betyr “fra legemet”. Somatiske sykehus behandler kroppen i motsetning til psykisk helsevern og rusbehandling. Videre går vi inn på et enkelt helseforetak med tre tilhørende sykehus, der vi har valgt å analysere de ulike sengepostene. En sengepost er en sykehusavdeling med døgnkontinuerlig pleiebemannning. Sengepostene er gjerne delt inn etter hvilken klinikk de tilhører og om det utføres medisinske eller kirurgiske konsultasjoner. På grunn av ulikheter i spesialisering ser vi bort fra intensivavdelinger.

1.3 Problemstilling

Formålet med oppgaven er å gjennomføre en effektivitetsanalyse av de somatiske sengepostene ved UNN for årene 2010 og 2011, basert på de variablene som kan knyttes til sengepostene. Inntektene til sykehusene baserer seg delvis på aktiviteten ved sykehuset. Således kan ineffektivitet påføre sykehusene uønskede merkostnader. I Samdata-rapporten (2011e) viser beregninger at UNN har en kostnad på 40 862 NOK per DRG-poeng i 2010. Diagnoserelaterte grupper (DRG) er et grupperingssystem som homogeniserer aktiviteten ved

et sykehus. På de tre andre universitetssykehusene i Norge kostet et DRG-poeng i gjennomsnitt 34 000 NOK. Det faktum at aktiviteten påfører UNN et potensielt tap på 6000 kroner per DRG-poeng i forhold til de sammenlignbare sykehusene¹, tyder på at UNN ikke opererer optimalt, eller like effektivt som mulighetene tilsier. Vi vil innlede arbeidet med å avdekke årsakene til de høye kostnadene med å gjennomføre en internundersøkelse av sammenlignbare enkeltposter innad i organisasjonen. Et alternativ til denne framgangsmåten er å sammenligne enkeltposter ved UNN med liknende poster ved de andre universitetssykehusene, noe som faller utenfor rammene til denne oppgaven.

Vi har valgt å reise problemstillinger knyttet til effektivitetsforskjellene mellom sengepostene på Universitetssykehuset i Nord Norge. Hovedformålet er å estimere effektivitetsforskjellene mellom sengepostene på en god og konsistent måte. Ved å bruke effektivitetsestimater vil vi undersøke om det finnes noen særegne karakteristikk ved de mest effektive sengepostene. Vi vil også forsøke å avgjøre hvorvidt størrelsen ved de ulike sengepostene påvirker effektiviteten, noe Biørn, Hagen, Iversen, og Magnussen (2003) avviser på sykehusnivå. Som nevnt innledningsvis vil det i tillegg være et overordnet mål å undersøke i hvilken grad det refusjonsbaserte finansieringssystemet fungerer som aktivitetsmål i en intern analyse. Gjennom relevant litteratur og etter ønske fra UNN, har vi kommet fram til ulike eksogene variabler knyttet til sengepostene, og vi ønsker også å undersøke om disse influerer effektivitetsestimater.

UNN ønsket primært å avdekke effektivitetsforskjeller knyttet til beleggsgrad og variasjoner i pleieressurser ved sengepostene. Beleggsgrad er ikke i seg selv et mål på effektiviteten, men heller et endimensjonalt produktivitetsmål. Den forteller noe om utnyttelsen av kapasitet, men fanger ikke opp viktig informasjon som for eksempel kostnader og priser i prosessen. Det kan likevel være en sammenheng mellom beleggsgrad og effektivitet.

Kompleksitet er en betydelig kostnadsdriver ved norske sykehus (Nyland & Bjørnenak, 2000). Funnene fra arbeidet til Nyland og Bjørnenak tyder på at de store sykehusene i Norge har et høyere relativt kostnadsnivå enn de små, og forfatterne hevder at dette skyldes større kompleksitet i både behandlingstilbudet og blant pasientene hos de store sykehusene. Hos UNN er kompleksitet ved en sengepost definert som en indeks av hvor krevende en

¹ Organisatorisk sett er de sammenlignbare, selv om de geografiske (og demografiske) forholdene er noe annerledes.

enkelpasient er i forhold til gjennomsnittspasienten (gjennomsnittspasienten=1). Et annet interessant poeng vil være å se om effektivitet avhenger av hvorvidt det er en medisinsk sengepost eller kirurgisk sengepost. Holmeide, Seierstad, og Eimot (2008) har påvist at DRG-systemet fører til en overdekning av kirurgiske DRGer og en underdekning av medisinske DRGer. Store deler av kostnadene ved et sykehus er faste på kort sikt, og sykehusene får betalt etter et aktivitetsbasert belønningssystem. Pasientmengde og aktivitet er en driver for å få ned kostnadene per pasient. Siden store deler av kostnadene er faste, vil kostnaden av en ekstra pasient (marginalkostnaden) synke når man øker pasientmengden/aktiviteten (dette gjelder om man opererer innenfor den gitte kapasitetsrammen). Poengene ovenfor bunner ut i fire forskningslhypoteser.

Hypotese 1: Belegg/utnyttelse av kapasitet har ingen innvirkning på effektivitet

Hypotese 2: Det eksisterer ingen effektivitetsforskjeller mellom kirurgiske og medisinske sengeposter

Hypotese 3: Andel av utskrevne pasienter har ingen påvirkning på effektivitet

Hypotese 4: Kompleksiteten har ingen påvirkning på effektivitet.

Hovedoppgaven er å analysere hvilke poster som ikke opererer effektivt i forhold til de andre, og samtidig undersøke om noen faktorer som i andre undersøkelser har vist seg å være årsak til ineffektivitet, også er effektivitetsdrivere for sengepostene på UNN.

1.4 Oppgavens struktur

Oppgaven er delt inn i syv hovedkapitler. Det foregående kapitlet inneholdt en kort bakgrunn, avgrensning og presentasjon av problemstillingene vi ønsker å undersøke. Kapittel 2 gir en generell presentasjon av helsesektoren og hvordan sykehusene i Norge er organisert og finansiert. Til slutt presenteres UNN HF. Kapittel 3 inneholder litteratursammendrag av relevant fagartikler. Her gjennomgår vi forskjellige effektivitetsanalyser av helsesektoren, både nasjonale og internasjonale. Teorien og metoden vi baserer oss på i analysedelen er presentert i kapittel 4. I kapittel 5 diskuteres datasett og det gis beskrivelser av input, output og kontrollvariabler i vår analyse. Her foretar vi også analyser av datasettet for å kunne gjøre oss opp om en mening om valg av modell og indikatorvariabler. Analysene av selve

problemstillingen presenteres i kapittel 6, med følgende resultater og en diskusjon rundt disse. Avslutningsvis, i kapittel 7, oppsummerer vi resultatene våre og konkluderer på bakgrunn av det vi har observert. Her foreslås det også muligheter for utdypende undersøkelser.

2 Helsesektoren

Det var kirken og klostre som stod for den første typen sykehus i Norge i form av hospitser og fattighus. Det eldste kjente sykehuset i Norge er Trondhjem Hospital som har vært i drift siden 1277. Første tegn til et offentlig organisert helsevesen viste seg først tidlig på 1500-tallet, i form av at staten Danmark-Norge ansatte universitetsutdannede leger til å ha ansvar for hvert sitt stift (dagens bispedømme). Utviklingen videre var at hver kommune skulle ha henholdsvis lege og jordmor. Dette, i tillegg til en egen kirurgtjeneste, utgjorde i unionstiden med Danmark helsevesenet i Norge. Rikshospitalet ble åpnet i 1826 og i 1868 ble den første sykepleierskolen åpnet i Norge. Den store utbyggingen av sykehus i Norge skjedde i perioden 1900-1940 (Store Norske Leksikon, 2012).

Helsesektoren er en stor og viktig "næring" i alle sammenhenger. I 2009 var helseutgiftene i Norge på cirka 229 milliarder kroner, noe som tilsvarer om lag 12 prosent av fastland-Norges brutto nasjonalprodukt. Om lag 84 prosent av dette var offentlig finansiert. Dette tilsvarer et forbruk på 47 350 kroner per innbygger. I 2009 sysselsatte man 490 000 personer innenfor helse og sosialsektoren (Helsedirektoratet, 2012). Det øverste organet for den norske helsesektoren er Helse- og omsorgsdepartementet, som igjen har fordelt ansvar til Helsedirektoratet. Helse- og omsorgsdepartementet er ansvarlig for folkehelsearbeid, primærhelsetjenesten i kommunene, spesialisthelsetjenestene, tannhelsetjenesten, hjelp til personer med psykiske lidelser, helsetjenester til rusmiddelavhengige og tilgang på legemidler. Det er Helse- og omsorgsdepartementet som "eier" helseforetakene, og sitter som øverste myndighet for driften av offentlige helsetjenester i Norge. I 2002 ble sykehusene og helseregionene overført fra fylkeskommunene til staten, slik at sykehusene og helseregionene er underlagt statlig eierskap.

Det er per dags dato fire helseregioner i Norge delt inn i nord, sør, øst og vest. Disse "eier" 25 helseforetak, som igjen administrer og styrer 32 offentlige sykehus fordelt på kategoriene regions- og universitetssykehus, sentralsykehus og lokalsykehus.

Helse Nord er helseregionen med størst geografisk ansvarsområde, og strekker seg fra helt sør i Nordland til Kirkenes helt øst i Finnmark. Det er også helseregionen med færrest innbyggere.

Tabell 1: Kliniske avdelinger og seksjoner ved ulike sykehustyper

Kliniske avdelinger og seksjoner ved ulike sykehustyper			
	Lokalsykehus	Sentralsykehus	Regionsykehus
	Generell kirurgisk avd.	Barneavdeling	Nevrokirurgisk avd.
	Generell indremedisinsk avd.	Nevrologisk avd.	Kjevekirurgi og
	Føde- /gynekologisk avd.	Øre-nese-hals avd.	munnhulesykd.
Avdelinger		Revmatologisk avd.	Hud- og veneriske
		Øyeavdeling	Plastikkirurgi
		Ortopedi (skadekirurgi) ¹	sykdommer
		Psykiatrisk avd.	Onkologisk avd.
Seksjoner	Innenfor kirurgisk avd.:	Urologi	Barnekirurgi
		Gastroent. kirurgi	Hjertekirurgi
		Karkirurgi	
		Thoraxkirurgi	
	Innenfor indremedisinsk avd.:	Hjertesykdommer (Cardiologi)	Hematologi
		Nyresykdommer	Geriatrici
		Lungesykdommer	Endokrinologi
		Fordøyelsesykdommer	
		Infeksjonssykdommer	
<p>1. Kan også være organisert som seksjon ved kirurgisk avdeling.</p> <p>(Regjeringen.no, 1997)</p>			

2.1 Sykehusenes inntekter og finansiering

I 1997 ble det introdusert et nytt finansieringssystem for den somatiske spesialhelsetjenesten. Modellen er basert på det amerikanske systemet *Health Care Financing Administration-DRG* (HCFA) der man deler inn pasientopphold i ulike diagnosegrupper med tilhørende ressursbruk. HCFA ble utviklet ved Yale-universitet på slutten av 1970 tallet. Formålet var å skape et grupperingssystem som både var medisinsk meningsfullt og ressursmessig homogent. I 1984 startet Norsk institutt for Sykehusforskning (nå SINTEF Helse) et forprosjekt der de studerte HCFA for videre å fortsette arbeidet med å lage en norsk versjon av systemet. Prosjektet var todelt, der det ble utviklet et grupperingsprogram og et vektsystem basert på norske forhold. Dataene ble samlet inn fra de 20 norske sykehusene med best datakvalitet. Tidlig på nittitallet ble systemet prøvd ut, først ved Haukeland sykehus, for så i 1994-1995 å bli prøvd ut som fordelingsystem for deler av øremerkede midler fra Helse- og omsorgsdepartementet (Helsedirektoratet, 2011a), og på slutten av 1990-tallet var systemet innført i alle 19 fylker.

Innsatsstyrt finansiering (ISF) er en todelt konstruksjon som består av en basisdel og en aktivitetsbasert del. Forholdet delene i mellom er 60 prosent basisbevilgning og 40 prosent aktivitetsbevilgning. Formålet med ISF er å bidra til de regionale helseforetakenes ”sørge for”-ansvar og deres oppfyllelse av de aktivitetsmål som er gitt av Helse- og omsorgsdepartementet. Dermed vil ISF fungere som et styringssystem som gir alle pasientgrupper lik oppmerksomhet (Helsedirektoratet, 2010).

Før innføringen av ISF var basisbevilgningen på 100 prosent. I motsetning til ISF var dette systemet konstruert slik at kompliserte og kostnadskrevenende operasjoner kunne bli nedprioritert i forhold til de relativt enklere behandlingene. En følge av dette kunne være fare for økte behandlingsskøer både totalt og kanskje særlig for kompliserte diagnoser. Et annet problem med den gamle finansieringsmodellen var at fylkene kunne overføre helsebevilgninger til andre offentlige formål (Biørn et al., 2003). Med innsatsstyrt finansiering vil behandlingsprioriteringen kunne balanseres gjennom aktivitetsbevilgningsdelen. Den refunderbare andelen på 40 prosent er basert på et system kalt diagnoserelaterte grupper. De diagnoserelaterte gruppene er delt inn i mange forskjellige vektklasser, avhengig av hvilken type behandlingsaktivitet som kreves. Slik vil man kunne differensiere mellom behandlinger av ulik kompleksitet og dermed skape en mer rettferdig prioritering av de ulike

behandlingene. Der det før ble bevilget en gitt sum som skulle fordeles over alle diagnoser (og dermed lettere kunne prioritere de minst kostnadsbelastede), vil det nå tildeles en høyere refusjon for en komplisert behandling sammenlignet med behandlinger av mindre kompleks art. Det er dog viktig å tilse at det ligger et rettferdig vektsystem til grunn, slik at det ikke gis insentiver til å prioritere refusjonsoptimale diagnoser. For å sikre et rettferdig system er det innført en årlig revisjon av vektene basert på gjennomsnittsdata for behandlingskostnader. Utvalget vektarbeidet baseres på består av en gruppe samarbeidssykehus. Disse velges ut av de regionale helseforetakene, og i 2010 sto samarbeidssykehusene for 60 prosent av aktivitetsgrunnlaget for ISF-refusjonen. Utarbeidelsen av refusjonssystemet er et nordisk samarbeid som gjør det mulig å sammenligne pasientbehandling på tvers av landene (Helsedirektoratet, 2011f).

Helsedirektoratet poengterer at de diagnoserelaterte gruppene (heretter DRG) skal fungere som pasientklassifiseringssystem der man grupperer sykehusopphold og polikliniske konsultasjoner slik at de blir medisinsk meningsfulle og ressursmessig (tilnærmet) homogene. Ved behandling grupperes pasienten i en DRG basert på ulike variabler, slik som alder, kjønn og medisinsk behandling. Man kan dermed ikke selv velge hvilken DRG pasienten skal tilskrives. DRGene deles også inn etter medisinske og kirurgiske tilfeller. Dermed kan to like pasienter med samme type problemer grupperes ulikt hvis den ene krever operasjon og den andre ikke.

Både medisinsk og økonomisk informasjon gis gjennom klassifiseringen og systemet er konstruert slik at man får en oversikt over pasientsammensetningen og aktiviteten på sykehusnivå. Slik vil det kunne gjennomføres effektivitetsanalyser på tvers av sykehus. Siden DRG beregnes ved hjelp av gjennomsnittstall i populasjonen, vil det alltid være avvik mellom faktisk ressursbruk per enkeltpasient og den beregnede DRG-vekten. Et annet moment er at alle sykehusets kostnadssteder er representert i hver enkelt DRG. Dermed er det ikke kun avdelingen pasienten skrives ut av som har en "eierandel" i den aktuelle DRG-produksjonen.

Det finnes 24 hoveddiagnosegrupper (HDG) og om lag 870 selvstendige DRG-er. Fordelingen av kostnader følger to ulike modeller. Dag/døgnmodellen fordeler kostnadene som hører til innlagte pasienter mens poliklinikkmodellen er basert på kostnadene ved dagkonsultasjoner. DRG 1~521 er knyttet opp mot dag/døgnmodellen, mens DRG 187E, 187O og 462O og DRGene med prefiks 7**, 8** og 9** er beregnet ved hjelp av poliklinikkmodellen.

Tabell 2: Deskriptiv statistikk av antall DRG og HDG ved sengepostene

	Antall DRG	Antall HDG
Gjennomsnitt	239	21
Min	59	15
Maks	409	24

I tabell 2 vises en deskriptiv statistikk over hvor stort antall DRG-er sengepostene ved UNN opererer innenfor, og hvor mange HDG-er disse DRG-ene er fordelt på. Selv om sengepostene er spesialiserte etter sykdomstype, vil man ha innslag av de fleste hoveddiagnosetyper i pasientbehandlingen. Majoriteten av sengepostene vil ha en hovedvekt på HDG-er knyttet opp mot egen spesialisering.

Følgende er et eksempel på to diagnoserelaterte grupper med ulik kompleksitet:

Tabell 3: Eksempel på DRG

DRG	Navn	Kostnadsvekt
235	Lårbeinsbrudd	0,950
480	Levertransplantasjon	26,718

Kostnadsvektene som tillegges hver DRG, er en relativ størrelse som uttrykker hva alle opphold i en DRG i gjennomsnitt koster i forhold til gjennomsnittspasienten. I beregningen av kostnadsvektene benytter man en ”*top down*” modell der sykehusenes driftskostnader fordeles til ulike DRG-er ved hjelp av fordelingsnøkler knyttet opp mot ulike kostnadsgrupper.

Fordelingsnøklerne i dag/døgnmodellen er fordelt ved hjelp av disse kostnadsgruppene:

- Operasjon
- Intensiv
- Bildediagnostikk
- Laboratorier
- Cytostatika

- Strålebehandling
- Dialyse
- Direkte pleie
- Grunnkostnad

Gjennomsnittsberegninger av hver enkelt DRG bestemmer hvor mye av kostnadene som skal fordeles til hver gruppe. I tabell 4 nedenfor viser vi kostnadsfordelingen av DRG 235 – Lårbeinsbrudd og DRG 214A – Kombinert fremre/bakre spondylodese.

Tabell 4: Eksempel kostnadsfordeling

	DRG 235		DRG 214A	
	Lårbeinsbrudd		Kombinert fremre/ bakre spondylodese	
Kostnadsgrupper	Vekt	Enhetsrefusjon	Vekt	Enhetsrefusjon
	0,95	36 968	5,903	36 968
Grunnkostnad	76 %	26 715	20 %	44 323
Pleiekostnad	17 %	5 995	4 %	8 270
Intensiv	1 %	521	2 %	5 288
Operasjon	1 %	419	72 %	157 873
Stråle	0 %	-	0 %	-
Bilddiagnostikk	2 %	598	1 %	1 434
Laboratorium	2 %	872	0 %	1 035
Cytostatika	0 %	-	0 %	-
Implantat	0 %	-	0 %	-
Refusjonsgrunnlag		35 120		218 222

I 2010 hadde UNN en DRG-produksjon på nærmere 60 000. Av disse kom omtrent 47 000 fra sengepostene. Ved UNN kommer med andre ord en overvekt av DRG-produksjonen fra behandlingen av døgnpasienter. Som nevnt over er det flere avdelinger på et sykehus som har et eierforhold til en og samme DRG-produksjon. Legges det inn en pasient som går i gjennom tre avdelinger før vedkommende skrives ut, vil det kun registreres på siste avdeling. Et DRG-poeng har en enhetsrefusjon på 36 968 kroner, og av dette er den aktivitetsbaserte refusjonen 40 % (Helsedirektoratet, 2010). Ved UNN i 2010 utgjorde den aktivitetsbasert delen cirka 17 prosent av totalinntekten.

På makronivå vil sammenligninger med bruk av DRG-poeng som aktivitetsmål kunne gi effektivitetsforskjeller på grunnlag av forskjeller i omgivelsene til de ulike sykehusene. Når systemet er basert på gjennomsnittskostnader, vil det være nærliggende å anta at kostnadene ved nordlige sykehus vil ligge på et høyere nivå enn resten av landet. En stor geografisk spredning av befolkningen, og lavere befolkningstetthet enn resten av landet, vil gi høye transportkostnader og høyere kostnader knyttet til beredskap for å kunne betjene en spredd befolkning. På avdelingsnivå må man også være forsiktig med å tolke for mye ut av resultatene i en sammenligning, da det presiseres at ISF som finansieringssystem er rettet mot de regionale helseforetakene, og ikke ment som et system for treffsikker finansiering av enkeltpasienter eller avdelinger (Helsedirektoratet, 2011b).

Finansieringen av helseforetakene har også innslag av andre elementer som refusjon for utdanning, kompensasjon for forskning, ”raskere tilbake”² og andre prosjektspesifikke tilskudd.

2.2 Om UNN

Universitetssykehuset i Nord-Norge HF (UNN) er det desidert største helseforetaket innenfor regionen. Det hadde et budsjett på 5.8 milliarder kroner i 2011, og i overkant av 6000 årsverk (HelseNord, 2012). I tillegg til sykehuset i Tromsø inngår også lokalsykehusene i Harstad og Narvik i UNN. UNN er delt inn i 12 klinikker og 7 sentre, som igjen består av poliklinikker, sengeposter, operasjonsstuer og støttefunksjoner som laboratorier og røntgenavdelinger. UNN er et universitetssykehus med ansvar for opplæring av studenter innen helsefag som fysioterapi, sykepleie og medisin.

Helse Nord sier at deres viktigste styringskrav knyttet til effektiv drift er ”Å *organisere spesialisthelsetjenesten i Helse Nord på en kostnadseffektiv måte, samtidig som de kombinerte krav til kvalitet og tilgjengelighet og utdanning ivaretas*”(HelseNord, 2006). Overskudd er ikke et krav, men heller en forutsetning for å kunne investere i nye bygg og anlegg. I 2009 gikk regionforetakene med samlet overskudd, og i 2010 hadde alle de regionale foretakene overskudd i sin drift. Helse Nord viser i en bærekraftanalyse fra oktober 2011 at det forventede overskuddet ved UNN i 2012 vil bli 34 millioner, noe som avviker positivt fra resultatkravet, som for dette året er 24 millioner (HelseNord, 2011).

² Et tilbud om arbeidsrettede tiltak til sykmeldte for å hindre unødig langvarig sykefravær.

2.3 Om sengeposter

En sengepost kan ses på som oppholdsstedet for innlagte pasienter. Dette er et sted for rekonvalesens samtidig som det er en venteplass i forkant av operasjon. Sengepostene driver primært med pleie av pasienter, og er dermed de viktigste arbeidsplassene for sykepleiere og hjelpepleiere. Organisatorisk er sengepostene underlagt sine respektive avdelinger etter medisinsk spesialitet, men det finnes også sengeposter som betjener pasienter fra ulike spesialiseringer. Større sykehus er ofte divisjonsorganisert, hvilket vil si at man har en felles avdeling som betjener både medisinske og kirurgiske tilfeller. Her vil det kunne være sengeposter som betjener både kirurgiske og indremedisinske pasienter. Driftsmessig er sengepostene oftest betjent døgkontinuerlig, syv dager i uken.

2.4 Oppsummering

Tidligere effektivitetsanalyser av det norske helsevesenet har vært gjennomført på sykehusnivå. Det vil si at den aggregerte bruken av innsatsfaktorer og den totale produksjonen for hvert enkelte sykehus har blitt sammenlignet opp mot de andre. Effektivitetsbegrepet kan ofte feilaktig bli brukt i betydningen produktivitet. Produktiviteten er gjerne et øyeblikksbilde av hvor mye man produserer, gitt forbrukte innsatsfaktorer. Effektivitet er derimot et mål på mengden innsatsfaktorer man faktisk benytter, sett i forhold til andre, sammenlignbare enheter. Når man skal foreta en sammenligning av flere like enheter er det ikke gitt at den med høyest produksjon bruker innsatsfaktorene sine på best mulig måte. En effektivitetsrangering vil dermed kunne rangere to produktive enheter forskjellig når det antas en felles produksjonsteknologi.

I vår oppgave har vi valgt å fokusere på et enkelt helseforetak med tre sykehus. Videre ønsker vi å sammenligne de tilhørende somatiske sengepostene ved UNN. I 2011 kom nesten 80 prosent av den totale DRG-produksjonen ved UNN fra de tilhørende sengepostene. Dette tilsier at pasienttyngden kommer fra innlagte pasienter. Økonomiavdelingen ved UNN presenterte selv aktuelle problemstillinger ved sykehuset. I tillegg har vi funnet andre interessante problemstillinger gjennom litteraturgjennomgangen. Ved å analysere sengeposter direkte må det tas noen forholdsregler siden mange pasienter har vært innom operasjonsstuer og støttefunksjoner ved sykehuset, samtidig som det er forskjeller i pleiebehov på tvers av pasientgrupper.

For å kunne gjøre en meningsfull analyse av de forskjellige sengepostene må det innhentes tall for innsatsfaktorer og aktivitetsmål ved de ulike avdelingene ved sykehuset. Vårt primære mål er å finne robuste indikatorer på disse målene, slik at vi kan gjennomføre en meningsfull effektivitetsanalyse av undersøkelsesobjektene. Primært vil vi benytte oss av *Data Envelopment Analysis*, men også støtte oss på *Stokastisk Frontanalyse*.

Da det ikke er gjennomført sammenlignbare studier nasjonalt, og de internasjonale avdelingsstudiene gjerne har omhandlet klinikker eller legearbeid, må oppgaven dermed sees på som en forberedende, eksplorativ studie, med lite veiledning fra litteraturen.

3 Litteratursammendrag

Effektivitetsstudier benytter seg enten av parametriske eller ikke-parametriske modeller for å analysere datasettene. Flere benytter seg også av begge typer for å støtte opp om resultatene fra den foretrukne modellen. Forskjellen på en parametrisk og ikke-parametrisk modell ligger i estimeringen av effektivitetsresultatene. En ikke-parametrisk modell tar utgangspunkt i de faktiske observasjonene, og lager en effektivitetsfront basert på analyseobjektene relative faktorforbruk og produksjon. Man behøver heller ingen spesifisert produktfunksjon og står (relativt) fritt i valg av variabler. I motsetning til ikke-parametriske modeller er en parametrisk modell basert på statistiske beregninger. Man leter etter gjennomsnittet i observasjonene og beregner tilhørende feilledd basert på disse. Dette forutsetter en *a priori* antagelse om den funksjonelle formen, i form av produktfunksjon eller kostnadsfunksjon.

Det finnes mange rapporter og artikler der sykehus og helsetjenester blir beskrevet og analysert. Tidligere undersøkelser har i hovedsak fokus på effektivitetsforskjeller mellom sammenlignbare sykehus på regionalt og nasjonalt nivå, og det finnes også en mengde undersøkelser som ser på effektivitetsforskjeller mellom land. Det finnes enkelte undersøkelser som tar for seg samme type problemstilling som den vi ønsker svar på. Det nærmeste vi kommer interne undersøkelser av sykehus er arbeidene til Chilingirian((1994) og (1995)), og senere i samarbeid med Sherman ((1996),(1997)) samt Lyroudi, Glaveli, Koulakiotis og Angelidis (2006) som omhandler greske klinikker. Chilingirian og Sherman analyserer flere forhold knyttet til legepraksisene innad i sykehus. Legens alder, erfaring og ulike typer legepraksis undersøkes, men studieobjektene er ikke helt av samme art som de vi ønsker å undersøke. Den greske undersøkelsen er noe nærmere vår problemstilling, men sengepostene ligger igjen et nivå under klinikker. Vi skal se nærmere på slike typer analyser senere i litteraturgjennomgangen. Om lag 2/3 av alle vitenskapelige undersøkelser av effektivitet i helsesektoren, bruker ikke-parametriske metoder i sine analyser, og omtrent halvparten av alle undersøkelsene omhandler sykehus (Hollingsworth, Dawson, & Maniadakis, 1999). Den resterende tredjedelen bruker parametriske metoder som regresjon og stokastisk frontanalyse.

Bjørnenak og Nyland (2000) har særlig analysert faktorer som driver kostnadene ved norske somatiske sykehus. I tradisjonell kostnadsanalyse er produksjonsvolum den eneste kostnadsdriveren, men Bjørnenak og Nyland argumenterer for at kompleksitet er en vel så

viktig kostnadsdriver. Kompleksitet kan være en av hovedårsakene til at det ikke finnes stordriftsfordeler i norske sykehus (Magnussen & Solstad, 1994). Deres analyser viser også tegn til dette, og resultatene heller mot at større sykehus er dyrere enn små. De store sykehusene oppviser kompleksitet i både behandlingstilbud og pasientsammensetning. Forfatterne deler kompleksiteten inn i to dimensjoner, dybde og bredde. Ved et sykehus utgjør den pasientrelaterte ressursbruken dybdedimensjonen og resten av ressursbruken breddedimensjonen. Breddedimensjonen inneholder både funksjonsrelatert ressursbruk og sykehusrelatert ressursbruk. Kostnader for uutnyttet kapasitet/beredskap inngår i den siste gruppen, og kan enkelt forveksles med ineffektivitet, dette da avdelinger/sykehus har for store kostnader/kapasitet i forhold til forventet mengde. Kompleksitet vil finnes både i forhold til pasienter og i forhold til funksjonene. Den største årsaken til kompleksitet i dybdedimensjonen knytter forfatterne til forskjellen i pasientenes behov. Andre årsaker til avvik fra standardene (DRG) kan stamme fra forskjellig behandlingspraksis og ineffektivitet. Kompleksitetskostnadene i bredden knytter forfatterne til stordriftsulempene, der det pekes på koordinering, kompromiss og fleksibilitet som drivere i breddedimensjonen. Funnene fra denne artikkelen vil trekkes inn når vi analyserer ineffektivitet, bl.a. fordi DRG-systemet og DRG som aktivitetsmål ikke kompenserer for kompleksitet utover den opprinnelige refusjonen. Alle negative avvik fra DRG-standarden vil påvirke effektiviteten negativt. Systematiske avvik kan sykehuset gjøre noe med, dette kan gjelde behandlingsmetoder eller ineffektivitet. Usystematiske avvik som spesialtilfeller eller ekstra pleietrengende pasienter, ligger utenfor sykehusenes kontroll og vil dermed i en analyse kunne framstå som ineffektivitet. Tallmaterialet som finnes ved sykehusene i dag gjør det vanskelig å fastslå hva som er ineffektivitet og hva som er nødvendige og pasientspesifikke avvik fra standarden.

Helsedirektoratet gir årlig ut *Samdata spesialhelsetjenesten*. Rapporten inneholder informasjon om ressursinnsatsen, bruken av tjenestene og utnyttelsen av ressursene i det foregående året. Den nyeste rapporten ble utgitt september 2011 og omhandler året 2010. Helsedirektoratet (2011e) gjennomfører en form for benchmarking av de store sykehusene og helseregionene ved hjelp av nøkkeltall, og rapporten beskriver også utviklingen de fem siste årene. Basert på tallmateriale fra perioden 2006-2010 beregnes ulike forholdstall som kostnad per innbygger, kostnad per DRG og kostnad per liggedøgn. Samdata-rapporten inneholder funn som tyder på det er variasjon mellom de fire helseregionene i Norge, og også store variasjoner mellom sykehusene. Rapporten forsøker imidlertid ikke å forklare årsakene bak variasjonene.

Vi vil presentere syv studier som omhandler effektivitet i helsesektoren, der alle analyserer effektivitetsforskjeller knyttet til sykehus. Fire av undersøkelsene bruker DEA som forskningsmetode. En ser på sykehusrelaterte problemstillinger ved hjelp av parametriske metode og ikke-parametriske metode, og en bruker kun parametriske metoder for å evaluere effektivitet ved 1600 amerikanske sykehus. En av undersøkelsene omhandler norske sykehus alene, mens to andre inkluderer Norge i en undersøkelse av effektivitetsforskjeller mellom de nordiske landene. De siste undersøkelsene omhandler sykehus i USA og Hellas, en av undersøkelsene omhandler legepraksiser innad i et sykehus i USA, og en annen ser på intern kliniskeffektivitet ved et offentlig sykehus i Hellas.

3.1 Effektivitetsundersøkelser ved hjelp av parametriske metode

Zuckerman, Hadley og Iezzoni (1994) analyserte utvalgte amerikanske sykehus ved hjelp av stokastisk frontanalyse. Utgangspunktet var et datasett med 5322 sykehus fra 1987, der en til slutt endte opp med et utvalg på 1600 sykehus. Det ble lagd et estimat på totalkostnader (avhengig variabel), inkludert inputpriser, generelle output, sykehusspesifikke output, og til slutt de to feilleddene som karakteriserer stokastisk frontanalyse (et tilfeldig feilledd og et systematisk feilledd som også kan sees på som teknisk ineffektivitet). I analysen brukte de naturlige logaritmer av alle tallene. Videre utvidet de analysen til å ta hensyn til en rekke dummyvariabler, og de prøvde også å definere sub-utvalg basert på et skille mellom offentlig og private sykehus, og mellom universitetssykehus og “vanlige” sykehus. Fra sub-utvalgene utviklet de egne effektivitetsfronter for de spesielle gruppene, og ved hjelp av statistiske tester viser det seg at de ulike gruppene har ulike kostnadsfunksjoner. Forfatterne hevder at korrelasjonen mellom ineffektiviteten til sykehusene i sub-utvalgene og ineffektiviteten i hele utvalget er såpass høy at det ikke er nødvendig med denne oppdelingen. De finner at mange av variablene i modellen påvirker ineffektivitet, og hovedtrekket er at sykehusene som er ineffektive har overforbruk av årsverk, kapital/kapasitet og dyre ansatte. Gjennomsnittlig ineffektivitet i denne undersøkelsen var på 13,6 prosent. Forfatterne hevder avslutningsvis at SFA egner seg bedre enn DEA fordi SFA både ser på ineffektiviteten som ligger innenfor bedriftens kontroll gjennom det systematiske feilleddet, men også den del av ineffektiviteten som ligger utenfor (det tilfeldige feilleddet).

Chirikos og Sear (2000) gjorde en undersøkelse av effektiviteten mellom sykehus i Florida i perioden 1982-1993, ved å sammenligne DEA og SFA-estimatene. Funnene fra denne undersøkelsen er at de to metodene gir nogen lunde samme resultater på "industrinivå". Når man ser på karakteristikene til de enkelte sykehusene, gir de to framgangsmåtene noe avvikende resultater. Metodene virker å håndtere lengde på sykehusopphold, personalmiks og eierskap på tilnærmet samme måte, men metodene avviker signifikant fra hverandre når det kommer til effekten av antall senger, dekningsgrad og antall saker/tilfeller. SFA antyder at små sykehus og de med lav dekningsgrad er mest ineffektive, mens DEA indikerer det motsatte. I DEA-modellen er det også signifikant sammenheng mellom ineffektivitet og antall senger, mens denne sammenhengen ikke eksisterer i SFA-modellen. Noe forskjell i effektivitet mellom DEA og SFA er forventet da SFA også tar hensyn til stokastisk støy som kan forekomme. Men det viser seg at innslag av tilfeldig støy ikke er nok til å forklare forskjellene mellom metodene, og årsakene til dette er noe forfatterne mener at det bør forskes mer på.

3.2 Undersøkelser ved hjelp av ikke-parametriske metoder

Den analysen som har et tema som ligger nærmest vårt er Lyroudi et al. (2006). Denne artikkelen omhandler effektivitet ved klinikker ved et sykehus i Hellas. Analyseobjektene var 10 ulike klinikker ved sykehuset. Formålet med studien var å se på effektivitetsforandringer/forbedringer ved klinikkene. Dette ble gjort ved å estimere en DEA-modell for klinikkene på månedsnivå, for så å bruke komponentene fra denne til å estimere Malmquist-indekser. Som input i modellen har forfatterne valgt senger og ulike kostnadstyper, og som output har de sykehusdager, konsultasjoner og ulike former for inntekter. Totalt inneholder modellen 6 outputs og 6 inputs. Formålet med undersøkelsen var ikke å sammenligne effektiviteten mellom de ulike klinikkene, men å se på endringene i hver enkelt klinikk effektivitet over tid. Resultatet er effektivitetstall for periodene og Malmquist-indekser for endringer mellom periodene. På grunn av lavt antall DMUer og høyt antall input og output, blir etter hvert 9 av 10 enheter effektive, og forfatterne dokumenterer en produktivitetsøkning i perioden.

En artikkel om ISF-innføringen (Biørn et al., 2003) ble skrevet på oppdrag fra Helsedirektoratet. Formålet var å undersøke hvordan innføringen av ISF-systemet påvirket

effektiviteten i det norske helsevesenet. Forfatterne så her etter endringer både i teknisk- og kostnadseffektivitet. I deres gjennomgang av lignende studier var det vanskelig å dokumentere effektivitetsforbedring etter en innføring av prospektiv finansiering. Studien baserer seg på data fra 48 ulike somatiske sykehus i perioden 1992 til 2000. Innsatsstyrt finansiering ble innført i 1997, hvilket ga mulighet til å analysere effekten over en lengre periode.

Artikkelen er tredelt: Først presenteres hovedhypotesen: Innsatsstyrt finansiering vil forbedre effektiviteten i forhold til en situasjon der sykehus er finansiert via grunnbevilgninger.

Hypotesen er utformet ved hjelp av en stilisert beslutningsmodell for norske sykehus (produktfunksjoner og nyttefunksjoner). De presenterer nyttefunksjoner for sykehusledelsen, der kostnadsreducerende tiltak har en negativ innvirkning på ledelsen. Selv om analysen er på sykehusnivå, poengteres det at en intern kilde til ineffektivitet kan være kødannelse på sykehusets støttefunksjoner (laboratorier, røntgenavdeling). En kødannelse her kan føre til økt liggetid på sykehusets sengeposter. Etter å ha presentert nytte- og kostnadsfunksjonene oppsummerer forfatterne sine antagelser.

- Økning i budsjettet har en ubestemmelig innvirkning på ytelse og dermed også på effektiviteten.
- Et skifte fra faste budsjetter til en kombinasjon av fast bevilgning og refusjon per behandling er antatt å gi en økning i arbeidsinnsats, og dermed også en økning i effektivitet.

I del 2 forklares valgene av inputs og outputs, for så å gå over til selve DEA-analysen. Det stadfestes at det ikke eksisterer et felles akseptert mål på output i helsesektoren. Siden den konseptuelle output, relativ helseforbedring i befolkningen, ikke er observerbar, måler de kvaliteten på helsetjenesten istedenfor eventuelle forbedringer i folks helse. De har brukt følgende outputs:

- *Behandling av innlagte pasienter (døgnopphold)*: Denne kategorien måler behandling i form av vektjusterte DRG-poeng.
- *Poliklinisk behandling*: Outputen er målt som antall polikliniske konsultasjoner justert etter refusjonen staten gir for hvert enkelte besøk.

Inputs er gitt ved

- *Legeårsverk*: Dette er kun en approksimering. Det er ikke skilt ut forskjeller i bruk av overtid fra sykehus til sykehus.
- *Årsverk andre yrkesgrupper*: En samlekategori av alle andre yrkesgrupper. En mer detaljert inndeling av disse yrkesgruppene endret ikke resultatet av analysen.
- *Medisinske kostnader*: Direktekostnaden av behandling.
- *Totale driftskostnader*: Driftskostnadene ble brukt som alternativ input i en av analysene, hvor formålet var å analysere kostnadseffektivitet.

DEA ble valgt istedenfor en parametrisk analyse på grunn av at den lettere håndterer flere inputs og outputs. Analysen er gjort ved hjelp av to inputminimerende modeller. Den ene er basert på årsverk (volum) for å måle teknisk effektivitet. Den andre modellen måler inputs i kostnader for å få et mål på kostnadseffektiviteten. Resultatmessig viser det seg at den tekniske effektiviteten økte gjennom perioden. Det største effektivitetsskiftet finner man i det første året etter reformen. Den positive trenden er lik for kostnadseffektiviteten helt frem til ISF-reformen. Etter innføringen observerte de et betydelig fall. Den siste seksjonen inneholder den empiriske analysen av hypotesen. Det stadfestes at effektiviteten (bestående av teknisk- og kostnadseffektivitet) påvirkes av ulike variabler. Blant disse variablene finner man *budsjettstørrelse*, *langtidsinnleggelse* og *antall senger*. Ved å utføre vanlig multippel regresjon mellom den avhengige variabelen DEA-effektivitet og de uavhengige variablene, fant de ulike sammenhenger. For *budsjettstørrelse* fant de en signifikant negativ sammenheng. For *antall senger* fant de ingen signifikant sammenheng, noe som støtter bruken av DEA analyse med konstant skalautbytte. *Langtidsinnleggelse* har også en positiv sammenheng, men det ses bort fra den siden den kan tilegnes eldre mennesker utenfor sykehusets kontroll. Videre har de foretatt mer avanserte økonometriske analyser, hvor det inkluderes en dummyvariabel ABF (Activity based financing[0,1]: før og etter ISF-innføringen). Her finner de en signifikant positiv sammenheng mellom innføringen av ISF og teknisk effektivitet, mens sammenhengen mot kostnadseffektivitet er signifikant negativ. Avslutningsvis konkluderer de med at innsattsstyrt finansiering har ført til en økning i den tekniske effektiviteten, mens det ser ut til at kostnadseffektiviteten har falt noe.

Kittelsen et al. (2009) undersøkte effekten av at norske helseforetak ble overført fra fylkene til staten i 2002. De analyserer årene 1999-2004. Formålet med helsereformen var

kostnadskontroll, effektivitetsforbedring og reduksjon av ventetiden. De undersøker effekten over 6 år og bruker i tillegg andre skandinaviske sykehus som kontroll- og sammenligningsgrunnlag. I undersøkelsen bruker de sykehuskostnader som input; output er DRG-poeng fordelt på gruppene medisin, kirurgi og dagpasienter. Resultatene fra undersøkelsen viser at effektivitetsforbedringen av reformen har vært på om lag 4 prosent ved norske sykehus i denne perioden. Mens "casemix" og ISF kun har små og ikke-signifikante effekter, er lengden på sykehusopphold signifikant negativt knyttet til produktivitet, men ingen av disse kan sies å virke gjennom innføringen av sentraliseringsreformen. Det påpekes at bedre administrasjon og tettere kontroll er mulige årsaker til forbedringen, men det er ikke testet for dette. Ved å bruke de andre nordiske landene i en analyse av norske forhold får man tilstrekkelige observasjoner/frihetsgrader for å gjøre analysen signifikant, samt at man på denne måten kan vise at effektene stammer fra reformen og ikke fra strukturelle eller teknologiendringer som er felles for de nordiske landene.

Medin et al. (2011) studerer effektivitetsforskjeller mellom universitetssykehus i Norden. I tillegg til å se på effektivitetsforskjeller mellom landene, brukes datasettet til å undersøke effektene av utdanning og forskning på kostnader og den optimale "størrelsen" på universitetssykehusene. Dette gjøres ved å først estimere effektivitet ved hjelp av DEA med rene sykehusinput/output. De bruker driftskostnader som input, og bruker DRG-justerte medisin, kirurgi og dagpasienter som outputs. Deretter estimeres det en ny effektivitetsfront, der effektene av utdanning og forskning inkluderes. De inkluderer utdanningskostnader som ny input, samt tre nye outputs: studentårsverk, publikasjonsmål og siteringer. Resultatene viser at det er store forskjeller i kostnadseffektivitet både mellom og innad i landene. Andrestegsanalysen deres forsøker å forklare hvor ineffektiviteten stammer fra. Geografisk lokalisering og andel pasienter med høy kompleksitet forklarer deler av ineffektiviteten, men mye ineffektivitet forble uforklart, og det viste seg at effektivitetene endret seg da utdanningsdelen ble hensyntatt.

Chilingerian (1995) hadde som formål å finne ut hvilke faktorer som var de viktigste hos legene som drev på best mulig måte (beste praksis). Undersøkellesobjektene var 36 leger fra samme sykehus som ble fulgt over en periode på 6 måneder. Chilingerian ønsker å teste effektivitet opp mot 7 "forklarende" variabler, der blant annet alder, case mix, antall DRG-er som behandles og om legepraksisen tilhører en HMO (Health Maintenance Organization, en organisasjon som opererer på kontrakter fra forsikringsselskaper). Effektivitetsundersøkelsene gjøres gjennom fire DEA-modeller med litt ulik spesifisering. Input i alle modellene er lengde

på opphold og utstyrskostnader, og output er utskrevne pasienter med høy kompleksitet og pasienter med lav kompleksitet. Chilingierian stiller spørsmåltegn ved hvorvidt input og output faktisk er homogene i dette tilfellet, og gjør først en inndeling der han skiller medisinske og kirurgiske praksiser. Deretter argumenterer han for at også leger med "høy" casemix må operere under en annen type produksjonsteknologi enn de med lav, og lager en inndeling basert på dette. Resultatet av dette er fire DEA-modeller, der han tester både VRS og CRS med først kun kirurg/medisin-inndelingen, før han deretter tester det samme inkludert casemix-grupperingen. Chilingierian bruker TOBIT for å se på effekten til de 7 forklarende variablene, og ender opp med to karakteristikker som symboliserer de teknisk effektive praksisene. De effektive gruppene tilhører en HMO, eller så spesialiserer de seg på få DRGer. Årsaker til at HMO gjør praksiser mer effektive enn de som ikke tilhører en slik organisasjon, har ikke forfatteren noen formening om, mens når det kommer til spesialisering finnes det to mulige årsaker. Den ene er at legene fokuserer på færre produkter, og dermed lærer seg disse "bedre". Den andre mulige årsaken er at leger som fokuserer på færre DRGer har lettere for å avvise/videreformidle pasienter som ligger utenfor deres kunnskapsfelt og dermed unngår pasienter eller saker som de ikke er egnet til å håndtere.

Vi har sett på et utvalg av vitenskapelige artikler og arbeider som tar for seg effektivitet og produktivitet i helsesektoren. Et vesentlig problem ved analyser av sykehus er at det ikke er enighet om hvilke variabler som er "riktig" å bruke i analysene. Det finnes derfor få eller ingen artikler som kan validere funn gjort i tidligere arbeider. Funnene fra undersøkelsen over er riktignok viktig å ha med seg i det videre arbeidet. En del av resultatene kan direkte knyttes til vår analyse, mens andre funn er viktig bakgrunnsstoff og kunnskap som vil hjelpe oss videre i oppgaven.

4 Teori og metode

Effektivitet og produktivitet er for mange synonyme begreper. Begrepsdefinisjonene er ofte ikke nok for å skille dem, da effektivitetsdefinisjonen alene kan virke noe ugripelig.

Produktivitet kan egentlig ses på som en av flere deler innenfor effektivitetsbegrepet. Støtter en seg på produksjonsteori blir det ofte lettere å se forskjellene i begrepene, og hvorfor produktivitet er en del av effektiviteten. Teori og metode presenteres i følgende rekkefølge:

- Effektivitet og produktivitet
- Distansefunksjoner
- DEA
 - o CCR-modellen
 - o BCC-modellen
 - o Outliers
 - o Supereffektivitet
- Stokastisk Frontanalyse
- Sensitivitetsanalyse
- Hypotesetesting

Analyseobjektene i en bransje kaller man gjerne for *decision making units* (Coelli, Prasada Rao, O'Donnell, & Battese, 2005). En DMU kan både være en egen bedrift, eller som i vårt tilfelle, en autonom avdeling eller sengepost på et sykehus som har kontroll over eget forbruk av innsatsfaktorer. Selv om sengepostene er underlagt egne avdelinger, velger vi å analysere dem som autonome enheter.

4.1 Effektivitet og produktivitet

Produktivitet er et mål på hvor mye man får igjen i forhold til hvor mye man forbruker (her) i en produksjonsprosess. I et enkelt tilfelle der man benytter seg av en innsatsfaktor x for å fremstille et produkt y , vil produktiviteten måles som en rate av produsert produkt (heretter output) i forhold til benyttet innsatsfaktor (heretter input).

$$\text{Produktivitet} = \frac{y}{x}$$

Den konstruerte produktivitetsraten forteller oss hvor mye ferdigprodusert output y man får for hver forbrukte input x . Produktivitetsmål alene kan gi et godt øyeblikksbilde av hvordan bedriften presterer i forhold til en egen satt standard, eller andre sammenlignbare bedrifter.

4.1.2 Teknisk effektivitet

I motsetning til produktivitetsbegrepet som ikke har noen antagelse om bransjeteknologi eller sammenligning, vil teknisk effektivitet avhenge av produksjonsnormen i bransjen og hva som er teknisk mulig å oppnå i produksjon. Ved hjelp av isokvanter³ som viser den optimale sammensetningen av inputs x_i i produksjon av output y_i definerer Farrell (1957) teknisk effektivitet som $\frac{\text{avstanden mellom origo og isokvanten}}{\text{avstanden mellom origo og } x}$, der x er et punkt på isokvanten eller utenfor. En bedrift som ligger på isokvanten vil dermed være teknisk effektiv. I dette punktet kan man ikke redusere en input x_i uten å måtte redusere en annen input eller output, y_i . Ved å bruke en slik beregning kan man rangere bedriftene. De som ligger på utsiden av isokvanten vil sammenlignes mot de som ligger på, noe som er analogt med benchmarking, en sammenligning mot "best practice" – de beste i bransjen. Hvordan de kan bevege seg mot de beste i bransjen kommer vi tilbake til i de neste avsnittene.

4.2 Distansefunksjoner

Distansefunksjoner er et verktøy man benytter i måling av effektivitet, og de ble i følge Coelli et al. (2005) introdusert (uavhengig) av både Shepard og Malmquist i 1953. Man beskriver produksjonsteknologien gjennom ulike sett slik at man kan måle distansen mellom ineffektive observasjoner og den gjeldende produksjonsmulighetsfronten/isokvanten.

Distansefunksjonene kan spesifiseres både i input- og outputretning, de håndterer multiple inputs og outputs, og de er grunnleggende for utviklingen av DEA. Siden vi måler effektivitet

³ Isokvantene er, gitt antagelsene til produktfunksjonen, konveks mot origo. De forteller oss noe om bytteforholdet inputene i mellom.

i input-retning, presenterer vi kun den input-orienterte distansefunksjonen slik dette er presentert i Coelli et al. (2005)

For å konstruere en input-orientert distansefunksjon må man først definere et input-sett.

$$L(\mathbf{y}) = [\mathbf{x}: \mathbf{x} \text{ kan produsere } \mathbf{y}] = [\mathbf{x}: (\mathbf{x}, \mathbf{y}) \in S]$$

Dette settet består av alle input-vektorene \mathbf{x} som kan produsere en gitt output-vektor \mathbf{y} . Gitt antagelsene rundt produksjonsteori har man følgende:

- (i) $L(\mathbf{y})$ er lukket for alle \mathbf{y}
- (ii) $L(\mathbf{y})$ er konveks for alle \mathbf{y}
- (iii) Inputs er svakt utbyttbare hvis $\mathbf{x} \in L(\mathbf{y})$ da, for alle $\lambda \geq 1$, $\lambda\mathbf{x} \in L(\mathbf{y})$
- (iv) Inputs er sterkt utbyttbare hvis $\mathbf{x} \in L(\mathbf{y})$ og hvis $\mathbf{x}^* \geq \mathbf{x}$, da er $\mathbf{x}^* \in L(\mathbf{y})$

Input-distansefunksjonen som skalerer vektoren \mathbf{x} , er definert på bakgrunn av det presenterte settet $L(\mathbf{y})$.

$$d_i(\mathbf{x}, \mathbf{y}) = \max_{\rho} [\rho: \left(\frac{\mathbf{x}}{\rho}\right) \in L(\mathbf{y})]$$

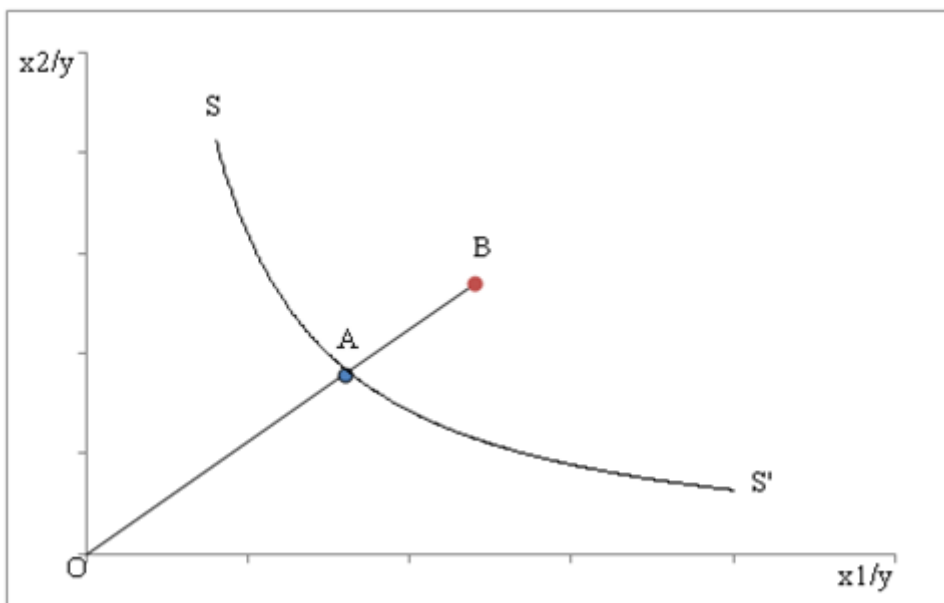
Der $L(\mathbf{y})$ er input-settet som representerer alle vektorene \mathbf{x} som kan produsere vektorene \mathbf{y} og ρ er en skaleringsfaktor. Gitt antagelsene ovenfor kan man stadfeste følgende:

- (i) Hvis \mathbf{x} tilhører input-settet til \mathbf{y} så er $d_i(\mathbf{x}, \mathbf{y}) \geq 1$.
- (ii) Distansen er $d_i(\mathbf{x}, \mathbf{y}) = 1$ hvis \mathbf{x} tilhører fronten til inputsettet.

Den gjeldende ”produksjonsmulighetsfronten” er egentlig input-settets tilhørende isokvant i forhold til ”best practice”. Fronten kan spesifiseres teoretisk, men den er enklere å estimere i forhold til faktiske observasjoner (Farrell, 1957), noe som førte til utviklingen av DEA.

4.3.1 Måling av teknisk effektivitet

For å illustrere hvordan distansefunksjoner kan benyttes i måling av teknisk effektivitet viser vi et eksempel med to inputs og en output i figur 1.



Figur 1: Distansefunksjon

Her er SS' enhetsisokvanten til de teoretisk (sett) effektive bedriftene. Hadde man konstruert fronten basert på virkelige observasjoner, ville dette resultere i flere punkter langs SS' knyttet sammen av rette linjer. I figuren har vi en enhet som opererer i punkt B. Gitt bransjens beste sammensetning av inputs ser vi at B kan redusere inputs til han er i punkt A uten at det reduserer produksjonen. Distansen mellom det effektive punkt A og det ineffektive punkt B gir oss den tekniske ineffektiviteten til enhet B. Målet gis som faktisk inputforbruk over optimalt inputforbruk.

$$\text{Teknisk effektivitet (TE)} = \frac{OA}{OB}$$

Så lenge en bedrift ligger utenfor fronten, vil distansen fra origo til bedriften være større enn distansen fra origo til fronten. Dermed vil de teknisk effektive ha en $TE = 1$ og de ineffektive vil ha $TE < 1$. Merk at $d_i(x, y) \geq 1$ når $TE \leq 1$. Det vil si at $TE = \frac{1}{d_i(x, y)}$.

4.3 Data envelopment analysis

Farrells distansefunksjoner var forløperen til utviklingen av *data envelopment analysis* (DEA). DEA er en ikke-parametrisk metode som benytter seg av matematisk programmering for å konstruere en front man kan måle effektivitet opp mot. Metoden blir på en måte en form for benchmarking, da analyseobjektene måles opp i mot de(n) beste i bransjen. For hvert analyseobjekt konstruerer man rater for output/input-forholdet som angis vekter man videre optimaliserer i konstruksjonen av en effektivitetsrangering. Vektene vil også forklare hvordan et ineffektivt analyseobjekt kan bevege seg mot fronten gjennom å endre på eget output/input-forhold. Det er viktig å poengtere at effektivitetsmålene er relative størrelser, basert på faktiske observasjoner innenfor datasettet. Den viktigste forutsetningen ved bruk av DEA er at analyseobjektene (DMUene) er homogene, det vil si at de opererer innenfor sammenlignbare omgivelser og under like teknologiske forutsetninger. Utenom dette trengs ingen *a priori* antagelse om funksjonell form (Zhu, 2004). DEA-metoden har vært i utvikling gjennom flere tiår og er basert på arbeidet til Farrell (1957). DEA fikk først navnet sitt i 1978, da Charnes, Cooper og Rhodes ga ut en artikkel der de antok en modell med konstant skalautbytte, CCR-modellen (Charnes, Cooper, & Rhodes, 1978). I 1984 kom Banker, Charnes og Cooper med en ny modell der de foreslo variabelt skalautbytte, noe som ga navnet BCC-modellen (Banker, Charnes, & Cooper, 1984).

DEA er dual av form. Det vil si at man har to sammenfallende former av hver modell. Dette er et resultat av at det originale lineære matematiske programmeringsproblemet har to løsninger; en primal og en dual. Vi presenterer den primale formen som en multiplikatormodell og dualformen som en "omhyllingsmodell". Forskjellen går ut på å maksimere eller minimere effektiviteten, gitt ulike betingelser. Den beregnede effektiviteten vil være lik uavhengig av hvilken modell man benytter seg av, men valg av modellform kan bidra til forskjellig tolkning av resultatene.

4.3.1 CCR-modellen

CCR-modellen er den første navngitte DEA-modellen. Modellen vi ser på er inputorientert, hvilket vil si at den skaper fronten basert på inputminimering. Selv om modellen er basert på rater, kan man sette den opp som et lineært programmeringsproblem. Vi har da følgende primalform, også kalt multiplikatormodell (1)

$$\begin{aligned} & \underset{u,v}{\text{maksimer}} E_o = \sum_{r=1}^S u_r y_{ro} \\ & \text{når} \\ (1) \quad & \sum_{i=1}^M v_i x_{io} = 1 \\ & \left(\sum_{r=1}^S u_r y_{rj} - \sum_{i=1}^M v_i x_{ij} \right) \leq 0 \quad (j = 1, \dots, N) \\ & u, v \geq 0 \end{aligned}$$

E_o er effektivitetstallet til DMU_o , DMUen som undersøkes, og dette maksimeres ved å finne den optimale verdien for inputvekten u_r ($r = 1 \dots, S$) og outputvekten v_i ($i = 1 \dots, M$). x_{ij} og y_{rj} angir hvor mye DMU_j bruker av input i og produserer av output r .

Det lineære programmeringsproblemet løses for hver DMU_o , $o \in (1 \dots, N)$, slik at alle kombinasjoner av de respektive DMUers inputs og outputs måles mot hverandre.

Beskrankingene i modellen fordrer at den observerte DMU_o sitt vektete forbruk av innsatsfaktor i er normalisert til 1. I tillegg rangeres vektene slik at outputvekten v_i er størst mulig mens inputvekten u_r er minst mulig. Disse vektene er også pålagt å være større eller lik null, og er ikke lik for alle DMUer. I evalueringen av ulike DMUers effektivitet vil den normaliserte inputvekten sørge for at dersom to DMUer har lik produksjon, vil den med lavest

forbruk av input rangeres høyere. I (Cooper, Seiford, & Tone, 2001) presenterer forfatterne følgende definisjon av CCR-effektivitet:

CCR-effektivitet

1. DMU_o er CCR-effektiv hvis $E_o = 1$ og det eksisterer minst en optimal kombinasjon av v og u der begge vektene er større enn null.
2. I alle andre tilfeller er DMU_o CCR-ineffektiv.

Hvis DMU_o har $E_o < 1$ må det finnes en kombinasjon (en av de andre DMUene) i beskrankningssettet der vektene v og u skaper likhet mellom summen av vektete outputs og summen av vektete inputs. Antallet observasjoner som har denne likheten skaper referansesettet for effektivitetsfronten

$$E'_o = [j: \sum_{r=1}^S u_r y_{rj} = \sum_{i=1}^M v_i x_{ij}]$$

Dette settet eksisterer for hver observerte DMU og sørger for at de som ikke oppfyller likheten rangeres som ineffektive.

4.3.2 Omhyllingsmodellen

Basert på teori om dualitet i lineær programmering, finnes det en dual form av multiplikatormodellen. Den duale formen kalles omhyllingsmodellen (2). Modellen innehar færre restriksjoner, og har en intuitiv tolkning (Coelli et al., 2005) som forklares etter at modellen er presentert.

$$\min \theta_o$$

når

$$(2) \quad \sum_{j=1}^N \lambda_j x_{ij} \leq \theta_o x_{io} \quad (i = 1, \dots, M)$$

$$\sum_{j=1}^N \lambda_j y_{rj} \geq y_{ro} \quad (r = 1, \dots, S)$$

$$\lambda_j \geq 0 \quad (j = 1, \dots, N)$$

Her er θ_o en effektivitetsskalar og λ_j er en vektor med konstanter som er større eller lik null. Verdien man får på θ_o er et mål på effektiviteten til den observerte DMUen. Størrelsen på λ_j er avhengig av datasettet under evaluering. I tilfeller der $\theta_o = 1$ ser man av den første restriksjonen at den minimale bruken av input x_i er lik inputforbruket til den observerte DMU. Slik vil man konstruere et referansesett der de effektive er sin egen referanse, mens de ineffektive har en annen DMUs λ som referanse. Dette blir et mål på hvor mye de bør redusere sitt forbruk av inputs. I et tilfelle der man har mer enn én input vil man kunne ha flere ulike DMUer som referanse på ulike inputs. Målet for hvor mye man må redusere sine innsatsfaktorer er gitt ved

$$x_i^* = \sum_{ij=1}^N \lambda_j x_i$$

Her er x_i^* et mål på optimalt forbruk av input i . Man kan også se det optimale forbruket ved å multiplisere eget effektivitetstall med forbruket av input i .

$$x_i^{**} = \theta_o x_i$$

Er det forskjell mellom x_i^* og x_i^{**} vil det være tegn på slakk i input i for DMU_o.

4.3.3 BCC-modellen

I 1984 presenterte Banker et al. (1984) et alternativ til CCR-modellen som tillot analyseobjektene å operere med annen skala enn konstant utbytte. Modellen kalles BCC-modellen. Skalaforskjellene kan komme av at analyseobjektene har konkurranseulikheter, er utsatt for forskjellig statlig regulering (Coelli et al., 2005) eller i vårt tilfelle, har ulikheter i pasientgrunnlaget. Der CCR-modellen identifiserer de(n) mest effektive DMUen(e) og former fronten basert på det konstante skalautbyttet, vil BCC-modellen legge fronten nærmere observasjonene siden det tas hensyn til ulikheter i operasjonell størrelse. Hvis det eksisterer skalaforskjeller i datamaterialet vil en BCC-modell forverre den rapporterte effektiviteten via en skaladel som ikke blir eksplisitt målt. Omhyllingsvarianten av den inputminimerende BCC-modellen (3) er presentert under.

$$\begin{aligned} & \min \theta_o \\ & \text{når} \\ (3) \quad & \sum_{j=1}^N \lambda_j x_{ij} \leq \theta_o x_{io} \quad (i = 1, \dots, M) \\ & \sum_{j=1}^N \lambda_j y_{rj} \geq y_{ro} \quad (r = 1, \dots, S) \\ & \sum_{j=1}^N \lambda_j = 1 \\ & \lambda_j \geq 0 \quad (j = 1, \dots, N) \end{aligned}$$

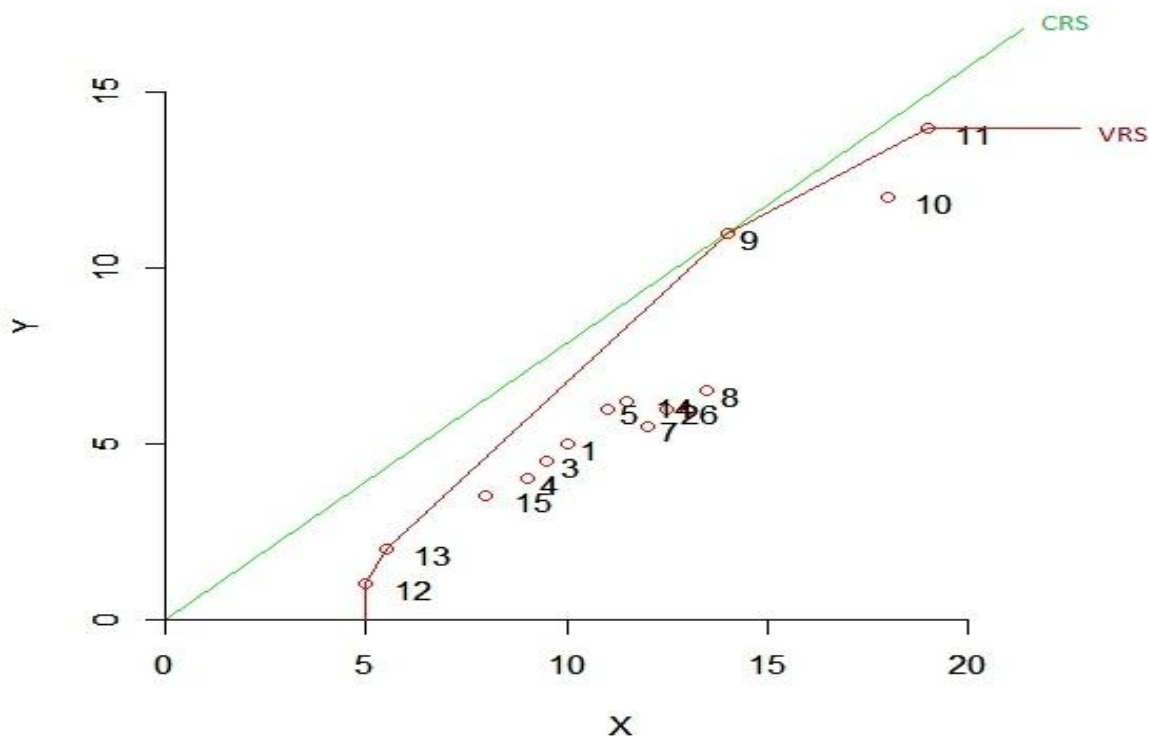
Modellen er nesten identisk med CCR-modellen, bortsett en ny konveksitetsbeskranking gitt ved $\sum_{j=1}^N \lambda_j = 1$. Denne beskranking legges effektivitetsfronten nærmere observasjonene, siden analyseobjektene kun vil benchmarkes mot DMUer av lik størrelse. Dette vil gi den enkelte DMU i BCC-modellen en effektivitetsscore som er lik eller større enn CCR-scoren.

Ved å måle effektiviteten både i CCR- og BCC-modellen kan man angi hva som er ren teknisk ineffektivitet, og hva som er skalaineffektivitet. Hvis det er forskjell i den tekniske effektiviteten mellom CCR og BCC betyr det at det eksisterer skalaineffektivitet.

Skalaineffektiviteten måles ved

$$\text{Skalaeffektivitet (SE)} = \frac{TE_{CCR}}{TE_{BCC}}$$

CCR-modellen vil altså måle både teknisk- og skalaeffektivitet, men man må beregne BCC-modellen for å skille ut hvor stor andel av den totale ineffektiviteten som skyldes feil skala.



Figur 2: Eksempel på en effektivitetsfront

I figur 2 ser vi et eksempel på to effektivitetsfronter. En for konstant skalautbytte (CRS) og en for variabelt skalautbytte (VRS). Datasettet til eksempelet ligger i vedlegg 1. Den rette linjen i figuren representerer CRS-fronten. Denne linjen går fra origo og gjennom det punktet (DMU 9) med best observert output/input-ratio. Den konkave linjen viser fronten når man tillater variabelt skalautbytte. Som vi ser er VRS-fronten tettere omhyllt observasjonene, og vil dermed gi en høyere gjennomsnittlig effektivitet for utvalget. DMU 11, 12 og 13 går fra å være et stykke fra den effektive fronten til å ligge på fronten. De har nå blitt effektive i forhold til skalaen de opererer på. Som vi ser av helningen på VRS-fronten er den høyere enn CRS-linjen til venstre for DMU 9, mens den er slakkere til høyre for dette punktet. Det vil si at for DMUene på venstresiden av DMU 9 vil en relativ økning i input gi en høyere relativ økning i output. Dette området kalles *økende skalautbytte (IRS)*. Det motsatte vil være tilfelle for DMUene på høyresiden, der en relativ økning i input vil gi en relativt lavere økning i output. Dette området kalles *avtagende skalautbytte (DRS)*.

4.4 Outliers

Et problem når det kommer til estimering av en effektiv front, er at det kan finnes bedrifter eller DMUer som ikke er sammenlignbare med de andre DMUene i datasettet, og dermed spiller en stor rolle for hvordan fronten blir seende ut. Disse kalles heretter outliers. Bogetoft og Otto (2010) peker på tre årsaker til at det finnes outliers. Feil i data kan forekomme, observasjonen kan være atypisk, eller så kan DMUen prestere såpass mye bedre eller dårligere enn de andre at den rett og slett ikke lengre er sammenlignbar med de andre DMUene i datasettet.

Det finnes flere måter å søke etter disse DMUene på. I parametriske modeller som OLS vil man enkelt kunne finne outliers ved å undersøke residualene til hver enkelt observasjon, og legge merke til store avvik. Hvis man bruker DEA som analysemetode finnes det ikke residualer, slik at man må benytte seg av alternative metoder for å finne outliers.

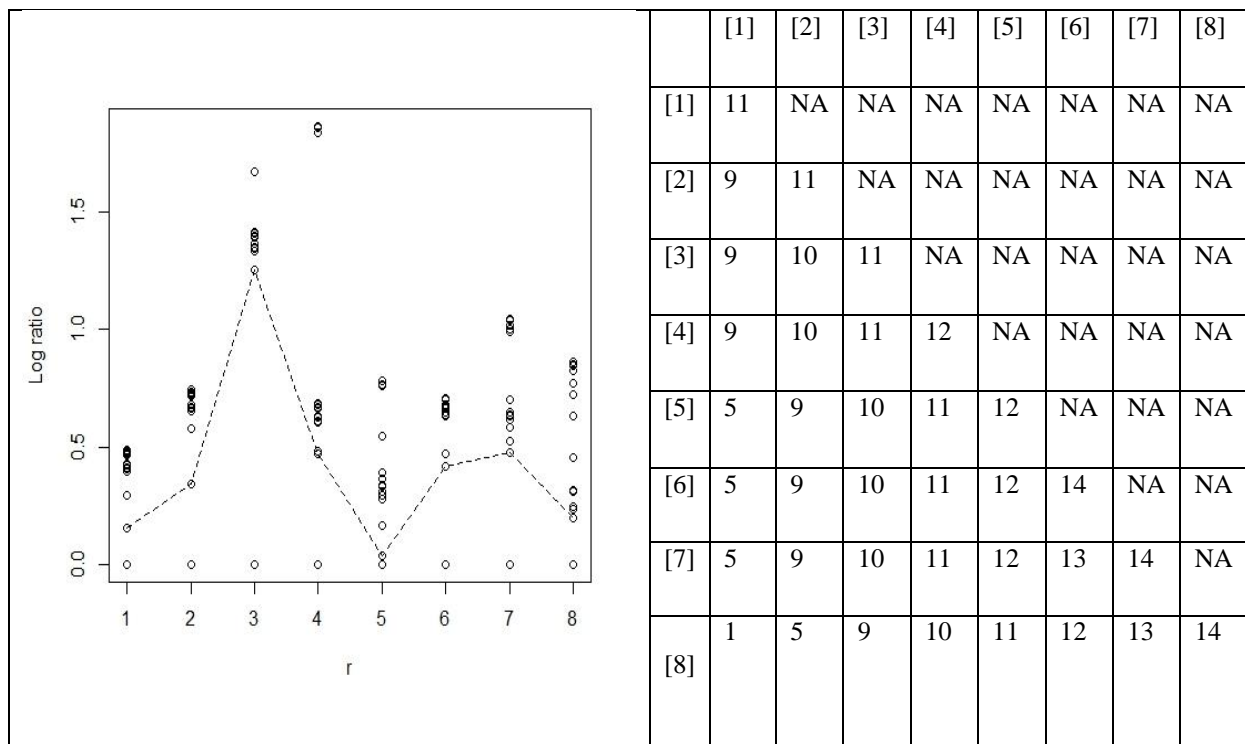
Deterministiske modeller som DEA er sensitiv ovenfor ekstreme observasjoner siden de tolker distansen til fronten som ineffektivitet.

4.4.1 Wilson-rutinen

Wilson(1993) utarbeidet en metode for å finne outliers i ikke-parametriske frontmodeller. Arbeidet er basert på (Andrews & Pregibon, 1978), som er en metode for å finne outliers når man kun har en output.

Wilson-rutinen minimerer $R^{(i)}$ for alle i , som er observasjoner vi velger å fjerne fra settet. $R^{(i)}$ er en logaritmeratio mellom de DMUene man fjerner fra settet og den DMUen i datasettet som ligger nærmest de man fjerner. Framgangsmåten gir oss både en grafisk framstilling og en tabell med resultater.

Grafisk vil det være enkelt å identifisere eventuelle DMUer som skiller seg fra de andre.



Figur 3: Outlieranalyse med medfølgende analyseobjekter som fjernes i hvert steg

I figur 3 har vi laget et eksempel på en outlieranalyse basert på tallmaterialet i vedlegg 1. Figuren viser at log-ratioen har en klar topp, og observasjonene som tilhører toppen kan sees

på som medlemmer av en outliergruppe. Det vi kan tolke av en slik tabell er at det potensielt finnes en gruppe outliers i dette datasettet, og de forekommer rundt $R^{(3)}$. Potensielt tre outliers finnes i eksempeldataen. Vi observerer at DMU 11,10 og 9 inngår i den første toppen i figur 3. Ved å studere figur 2 (i avsnitt 4.3.3) ser vi også at disse er samlet noe unna resten av observasjonene.

Analysere man data med denne metoden kan også ineffektive DMUer få betegnelsen outliers. De outlierne som ikke er effektive vil ikke påvirke hvordan effektivitetsfronten dannes, siden de ikke påvirker fronten direkte. Problemene oppstår først når effektive DMUer framstår som outliers, da disse er med på å konstruere effektivitetsfronten. Ineffektive outlierne kan skape problemer når man skal tolke resultatene og drøfte eventuelle konklusjoner, da det er noe som skiller dem fra de andre observasjonene i datasettet. Wilson-outliers trenger ikke å skape problemer på samme måte som supereffektive outliers. At en DMU eller en gruppe DMUer framstår som outlierne kan det rett og slett bety at de har en annen miks av input eller output enn de andre DMUene. Dette framgår også av grafen (figur 2), der grupper som har tilnærmet lik ressursbruk eller produksjon havner i samme gruppe av outliers (figur 3).

4.4.2 Supereffektivitet

I DEA-metoden vil alle effektive DMUer få effektivitetstall 1. I mange undersøkelser vil flere DMUer framstå som effektive, men resultatene sier ikke noe om rangeringen mellom disse eller hvem som er mest effektiv. Supereffektivitet forsøker å rangere disse og tillegger DMUene som er mer effektive enn andre en effektivitetsscore høyere enn 1. Hvis enkelte DMUer får en betydelig høyere effektivitetsscore enn de andre kan dette tyde på at den er en outlier. Adler, Friedman, og Sinuany-Stern (2002) peker på to mangler/problemer med supereffektivitet slik Andersen og Petersen (1993) definerte det. Supereffektivitet fungerer dårlig som rangeringsmetode fordi hver enhet er vurdert med forskjellige vekter. Metoden kan også gi ”spesialiserte” enheter urimelig høy supereffektivitet. Supereffektivitet måles ved (4) som er en omspesifisering av omhyllingsmodellen (2).

$$\begin{aligned}
& \min \theta_o \\
& \text{når} \\
(4) \quad & \sum_{j=1}^N \lambda_j x_{ij} \leq \theta_o x_{io} \quad (i = 1, \dots, M) \\
& \sum_{j=1}^N \lambda_j y_{rj} \geq y_{ro} \quad (r = 1, \dots, S) \\
& \lambda_j \geq 0 \quad (j \neq O)
\end{aligned}$$

Supereffektivitetsproblemet er nesten identisk med omhyllingsformuleringen av CCR-modellen (2). Forskjellen ligger i at modellen betinger at j ikke er lik o når modellen leter etter den observerte DMUs referanse- λ . I praksis betyr det at ingen DMU kan ha seg selv som referanse. Framgangsmåten er at man formulerer et minimeringsproblem der DMUen som analyseres ikke lengre har en teknologi som begrenser referansesettet. Effektive DMUer vil ikke kunne ha seg selv som referanse, slik at disse må finne en tidligere ineffektiv DMU eller mindre supereffektive DMUer som referanse. Utfallet av dette grepet er at effektivitetsscorene nå kan bli høyere enn 1, og noen av DMUene vil potensielt kunne ligge utenfor det opprinnelige effektivitetsområdet.

4.5 Stokastisk frontanalyse

Stokastisk frontanalyse er, i motsetning til DEA, en parametrisk metode. Det vil si at man stokastisk estimerer effektivitetsfronten basert på en gitt produktfunksjon. Beregningen av produktfunksjonen gir oss estimerte parametere eller koeffisienter som sier noe om hvordan inputene påvirker produksjonen. Der man i DEA ikke trenger en *a priori* antagelse om funksjonell form, behøver man dette ved bruk av SFA. Modellen er basert på regresjon med sammensatt feilledd. Feilleddet består av to komponenter: et tilfeldig og et systematisk, der det systematiske leddet er et mål på teknisk ineffektivitet. Modellen ble først presentert i Aigner, Lovell, og Schmidt (1977), og vi benytter oss av en modellspesifikasjon presentert av Battese og Coelli (1992). Modellen (5) baserer seg på paneldata, altså observasjoner av en gruppe DMUer over flere tidsperioder. Dataene våre er et tverrsnitt, dermed estimeres modellen over én tidsperiode, slik at $t=1$ og vil falle ut av modellen.

$$(5) \quad y_i = x_i\beta + (v_i + u_i) \quad i = 1 \dots, N$$

Der y_i er den naturlige logaritmen av produksjonen til den i 'ende DMU. x_i er en vektor av inputene til den i 'ende DMU (disse er også i naturlige logaritmer). β er en vektor bestående av ukjente parametere som skal estimeres, mens v_i er et tilfeldig feilledd (med forventet verdi null) og u_i er det systematiske feilleddet som måler teknisk ineffektivitet.

Beregning av denne modellen, der både input og output er i naturlige logaritmer, gjøres med en tilhørende Cobb-Douglas produktfunksjon. Det systematiske feilleddet u_{it} er ikke-positiv, og bestemmer om en DMU ligger på effektivitetsfronten eller under. En produsent kan påvirke dette feilleddet gjennom arbeidsinnsats, andel av defekte produserte enheter, overforbruk av innsatsfaktorer m.m. Det tilfeldige feilleddet v_{it} gjør derimot effektivitetsfronten i stand til å endres uavhengig på tvers av de ulike DMUene. Dette gir et stokastisk avvik. Årsakene til endringene her er utenfor den enkelte DMUs kontroll, slik som flaks og vær (Aigner et al., 1977). Man antar ulike fordelinger av de to feilleddene. Det systematiske leddet er antatt fordelt halv-normalt (effektivitetene ligger enten på fronten eller under), mens det stokastiske leddet er normalfordelt. Med variansen til de to leddene kan man konstruere (6)

$$(6) \quad \lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}$$

Jo høyere variansen til det systematiske feilleddet er i forhold til det stokastiske, desto høyere er λ (Bogetoft & Otto, 2010). Hvis $\lambda = 0$ finnes det ingen teknisk ineffektivitet i det analyserte datasettet. Man kan alternativt se andelen av variansen i modellen som skyldes variansen i teknisk ineffektivitet ved (7)

$$(7) \quad \text{andel T. ineff} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}$$

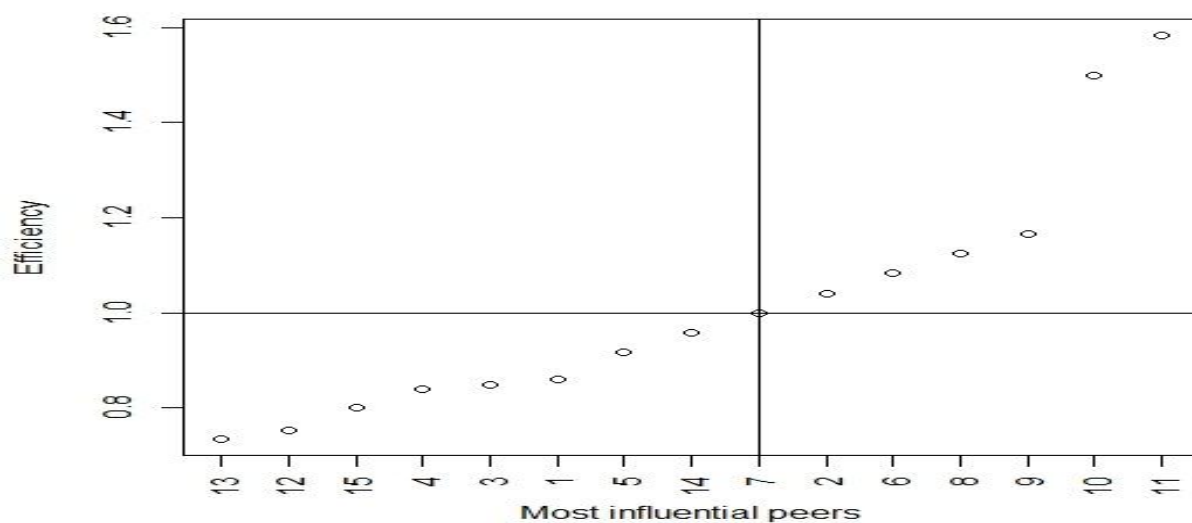
Ved beregningen av SFA får man presentert den totale summen av feilleddene u_{it} og v_{it} . Selv om beregningen gir et mål på individuell ineffektivitet for hver enkelt DMU vil man ikke kunne skille ut den individuelle effekten av de to residualene for den enkelte DMU.

4.6 Sensitivitet og robustheten til effektivitetsestimaterne

Det faktum at DEA er en ikke-parametrisk metode impliserer at den/de effektive DMUene har stor påvirkning på de ineffektive DMUenes effektivitetsestimater. En måte å teste sensitiviteten eller robustheten til effektivitetsestimaterne er ved å bruke en såkalt effektivitetsstige basert på Edvardsen (2004). Metoden baserer seg på at man velger en DMU man ønsker å undersøke robustheten til. Videre fjerner man sekvensielt analyseobjektets mest innflytelsesrike benchmark (DMUen som bidrar med høyest λ) helt til analyseobjektet selv ligger på effektivitetsfronten. Poenget er at man underveis ser på endringene i effektivitetsestimaterne som er forårsaket av at man fjerner den mest innflytelsesrike referanseenheten. Hvis analyseobjektet får en dramatisk effektivitetsøkning som et resultat av at man fjerner en referanseenhet kan dette være et tegn på at effektivitetsestimateret er sensitivt i den forstand at feilmålinger hos referanseenhetene får stor påvirkning på modellen og resultatene. Som Edvardsen selv poengterer;

All in all, the Efficiency stepladder measure is simple and crude, but it is shown that it can provide useful information for practitioners about the robustness of the efficiency scores in DEA”(Edvardsen, 2004, p. 59)

I steg 1 estimerer man effektiviteten til DMUen som undersøkes, og noterer seg den mest innflytelsesrike referanseenheten til DMUen. I steg 2, fjerner man referanseenheten fra dataen, og estimerer på nytt effektiviteten til DMUen. Gjør man dette tilstrekkelige antall ganger vil til slutt DMUen man ønsker å undersøke bli effektiv og selv ligge på effektivitetsfronten. Metoden baserer seg tungt på (Sinuany-Stern, Mehrez, & Barboy, 1994) der de måler hvor mange akademiske avdelinger man må fjerne før de ineffektive blir effektive. Effektivitetsstigen rapporterer i tillegg effektivitetsendringen for hver DMU man fjerner. Under følger et lite eksempel basert på samme eksempeldata brukt tidligere i teoridelen (Vedlegg 1). DMU 7 er analyseobjektet.



Figur 4: Eksempel på effektivitetsstige

Figur 4 viser en stigeestimering gjennomført i R via tillegget Benchmarking. Vi ser at effektivitetsestimaterne til DMU7 er meget robuste i forhold til endringer i effektivitetsfronten. Det tar 8 steg før DMU7 blir effektiv, og stigningen virker lineær, og dette tyder på at effektiviteten til DMU7 endrer seg like mye i steg 1 som i steg 7. Stegene etter at DMU7 har blitt effektiv estimerer supereffektiviteten til analyseobjektet.

4.7 DEA og hypotesetester

DEA er i utgangspunktet ikke en statistisk metode, men det faktum at datasettet kan sees på som et utvalg, gjør at man også kan gjøre antakelser om den ukjente underliggende datagenereringsprosessen (DGP). Ved hjelp av denne antakelsen kan man utvikle statistiske diagnoseverktøy også for DEA (Banker, 1993). I følge dette arbeidet kan man utarbeide statistiske mål som gjør at effektivitetsestimatoren er statistisk konsistent, og videre utvide denne estimatoren slik at man kan gjøre hypotesetester basert på dette statistiske grunnlaget.

For å kunne gjøre tester knyttet til estimatoren θ_j må vi gjøre enda en antakelse knyttet til fordelinga. Banker peker på to potensielle fordelinger estimatoren kan ha.. Banker (1993) utviklet ulike tester for egenskapene til estimatoren basert både på antakelse om eksponentialfordelt og halv-normalfordelt θ . Testen kan både brukes til å undersøke om det

finnes skalaforskjeller i utvalget og teste for signifikante forskjeller mellom effektivitene under ulike modellformuleringer.

For eksponentialfordeling:

$$(8) T_{EK} = \frac{\sum_{j=1}^N (1 - \theta_j^1)}{\sum_{j=1}^N (1 - \theta_j^2)}$$

Sammenlignet mot en kritisk verdi fra en F-test med $(2N, 2N)$ frihetsgrader

For halv-normalfordeling:

$$(9) T_{HN} = \frac{\sum_{j=1}^N (1 - \theta_j^1)^2}{\sum_{j=1}^N (1 - \theta_j^2)^2}$$

Sammenlignet mot en kritisk verdi fra F-test med (N, N) frihetsgrader.

θ_j^1 er effektiviteten til hver enkelt DMU i originalmodellen, og θ_j^2 er effektiviteten når modellen er spesifisert etter alternativet. Nullhypotesen for begge testene er at modellene er like. Vi avviser nullhypotesen hvis $F_{(N,N)} > T_{HN}$ eller $F_{(2N,2N)} > T_{EK}$.

En måte å se på eksogene variablers påvirkning på effektivitet kan være en form for to-steps analyse. Først estimerer vi en DEA-modell med de opprinnelige modellvariablene, for så i neste steg se på hvorvidt et sett eksogene variabler kan hjelpe til å forklare effektivitetsforskjellene. Banker og Natarajan (2008) påpeker at det er essensielt at korrelasjonen mellom input i DEA-modellen og de kontekstuelle variablene ikke er for høy, mens høy korrelasjon mellom de kontekstuelle variablene ikke er problematisk.

Den vanligste og enkleste måten å gjøre dette på er ved hjelp av ordinær regresjon eller OLS (ordinary least square). Banker og Natarajan (2008) argumenterer for at OLS ikke står tilbake for blant annet TOBIT når det kommer til å teste slike hypoteser. Resultatet av regresjonen forteller oss hvor mye av variasjonen i den avhengige variabelen som kan tilegnes variasjon i de(n) uavhengige variabelen(e), samtidig som vi får verdier på om koeffisientene til de uavhengige variablene har en signifikant påvirkning på den avhengige variabelen.

5 Datasett - input, output og andre variabler

For å kunne gjøre en tilfredsstillende effektivitetsanalyse av en produksjonsprosess kreves det data som gjenspeiler hele produksjonsprosessen. Godt dokumenterte observasjoner av hvilke ressurser som brukes som innsatsfaktorer er viktig for å kunne rangere DMUene. Man kan bruke vidt forskjellige framgangsmåter for å produsere samme produkt(er). Datasettet man skal bruke når man skal gjøre en effektivitetsanalyse ved hjelp av DEA, må inneholde mest mulig informasjon om veien fra innsatsfaktorer til sluttprodukt. Typiske eksempler på innsatsfaktorer kan være lønnskostnader, materialkostnader og kapitalkostnader. Man kan også ha datasett som er målt i enheter. Da kan innsatsfaktorer være årsverk eller materialforbruk målt i meter. En naturlig tilleggsinformasjon til slike enhetsmål vil være priser tilhørende innsatsfaktorene, slik at analysen kan utvides til også å se på allokeringseffektivitet og kostnadseffektivitet. Typiske sluttprodukter kan være inntekter eller produserte enheter. Kravet for å kunne sammenligne ulike DMUer er at de har noen lunde samme produksjonsprosess, og at de produserer samme sluttprodukt(er).

I *Pitfalls and Protocols in DEA* av Dyson et al. (2001) foreslår de at følgende kriterier må oppfylles i valg av datasettet:

- Dekker hele spekteret av innsatsfaktorer
- Dekker hele aktivitetsnivået og alle ytelsesmål
- Alle faktorer er lik for alle enheter
- Variasjon i omgivelsene er oppdaget og korrigert

I valg av antall input og output er det flere forhold å ta hensyn til. Først og fremst må de overnevnte kriteriene dekkers. Problemet blir å holde antallet variabler tilstrekkelig lavt for å oppnå best mulig diskriminering i effektivitetsvurderingen. Forfatterne legger frem et eksempel med fire inputs og fire outputs. Siden DEA er konstruert slik at den velger vektene selv, vil flere faktorer føre til lavere diskriminasjon. Med 16 ulike kombinasjoner av input og output vil man kunne få 16 effektive enheter. Dette kommer av at dersom en enhet har den beste sammensetningen av en unik input/output-konfigurasjon, vil den komme ut som effektiv. Det trenger ikke være slik at det er en effektiv enhet for hver input/output-konfigurasjon, men for å diskriminere bør man ha et antall analyseobjekter som er større enn $2m*s$, der $m*s$ er produktet av $input*output$. Som en remedie for å oppnå diskriminasjon kan alle inputs som er mulig å prise samles. Slik vil innsatsfaktorene kunne samles under en

kostnadsfane. På outputsiden kan de outputer som ikke er direkte relatert til organisasjonens mål fjernes. En annen vinkling er å begrense vektene på input og/eller output slik at det skapes en rangering basert på viktigheten til den enkelte faktor.

5.1 Datasettet

Analyseobjektene i denne oppgaven er sengeposter ved UNN. Data vedrørende sengepostene finnes ikke i offentlige databaser som StatRes eller hos SSB. Temaet i oppgaven ble foreslått av UNN selv, og datasettet er hentet fra UNNs databaser, ved hjelp av ansatte ved økonomi og analyseavdelingen. Datasettet inneholdt i utgangspunktet tall knyttet til 31 sengeposter ved UNN, fordelt rundt i hele helseforetaket. Dataene gjelder for 2010 og 2011. Datasettet inneholdt også tall for rehabiliterings- og palliative poster. Dette er poster vi ikke har tatt med i analysen. Vi måtte også fjerne sengeposter som fungerer som akuttposter eller intensivposter, fordi aktiviteten ikke måles her, men fordeles på andre kostnadssteder samtidig som denne aktiviteten ikke er sammenlignbar med de andre postene. Etter denne avgrensingen stod vi igjen med 22 sengeposter fordelt på sykehusene i Tromsø, Harstad og Narvik. Den siste justeringen vi måtte gjøre var å slå sammen de to medisinske sengepostene ved sykehuset i Harstad. Grunnen til denne sammenslåingen var at store deler av kostnadene for medisinsk post B var ført på medisinsk post A, og sykehuset hevder selv at disse to sammen kan sees på som én sengepost. Utgangspunktet for analysene våre er data for 21 sengeposter i perioden 2010-2011. Det er 7 kirurgiske sengeposter, 12 medisinske sengeposter og 2 som både håndterer medisinske og kirurgiske tilfeller.

Datasettet inneholder detaljert informasjon om lønnskostnader, andre driftskostnader, DRG-produksjon (både på sengepostnivå og for hver enkelte diagnosegruppe), informasjon om senger og ulik informasjon om årsverkene som er skapt ved sengepostene. Fra Helsedirektoratet har vi fått detaljert informasjon om kostnadsvektene tilhørende alle DRGene (Helsedirektoratet, 2011c, 2011d). UNN har også bidratt med en form for oversikt over hvilke hoveddiagnosegrupper hver enkelt sengepost operer innenfor og andelen hver enkelt hoveddiagnosegruppe står for av aktiviteten ved sengeposten.

5.1.1 Valg av variabler

I spesifiseringen av hvilke indikatorvariabler modellen bør inneholde, har vi valgt å teste ulike former av omhyllingsmodellen (2) mot hverandre ved hjelp av Banker-testene beskrevet i kapittel 4.7.

F1: $y = \text{Justerte DRG-poeng}$; $x = \text{Totalkostnader}$.

F2: $y = \text{Justerte DRG-poeng}$; $x1 = \text{Lønnskostnader}$, $x2 = \text{Andre driftskostnader}$, $x3 = \text{Sengedøgn}$.

Denne testen finner ingen signifikant forskjell på en modellform der vi kun benytter oss av *totalkostnader* som input mot en modell med inputene *lønnskostnader*, *andre driftskostnader* og *sengedøgn*. Dette gjelder for begge årene og for VRS- og CRS-modellen. For komplett beregning se vedlegg 2. Resultatene betyr at fordelingen av ineffektiviteten i *F1* og *F2* ikke er statistisk forskjellig.

Det vi derimot observerer er at sensitiviteten (jf. kapittel 4.6) til effektivitetsmålene øker betraktelig i førstnevnte modell (*F1*), noe som tilsier at man bør benytte seg av den siste (*F2*). Ved å fjerne den mest influerende DMUen fra referansesettet til den enkelte DMU ser man at effektivitetsscoren jevnt over øker betraktelig mer i *F1* enn *F2*, noe som tilsier at settet er sårbart for målefeil. Dette er basert på metoden beskrevet i kapittel 4.6. En annen årsak til at vi har valgt den disaggregerte *F2* framfor *F1* er at vi ønsker opplysninger om input på kostnadstypenivå.

Vi har også testet to ulike varianter av *F2*, der den ene har *justert DRG* som output (den originale *F2*), og i *F2'* har vi med en ekstra output i form av *antall utskrevne pasienter*. Se vedlegg 3 for R-koder og utregning.

F2: $y = \text{Justerte DRG-poeng}$; $x1 = \text{Lønnskostnader}$, $x2 = \text{Andre driftskostnader}$, $x3 = \text{Sengedøgn}$.

F2': $y1 = \text{Justerte DRG-poeng}$, $y2 = \text{Utskrevne pasienter}$; $x1 = \text{Lønnskostnader}$, $x2 = \text{Andre driftskostnader}$, $x3 = \text{Sengedøgn}$.

Her finner vi heller ingen signifikant forskjell. I modell *F2'* er gjennomsnittseffektiviteten jevnt over høyere enn i *F2*, noe vi finner naturlig på grunn av flere variabler. DRG-produksjonen i seg selv skal være et konsistent mål på hvordan ressursene utnyttes (Biørn et

al., 2003) og måle den korrekte aktiviteten. Tolkningen av dette resultatet er at de justerte DRG-poengene alene er et tilstrekkelig aktivitetsmål.

Som en supplerende robusthetsanalyse bør man estimere en SFA-modell som er så nær DEA-modellen som mulig (Kittelsen et al., 2009). Ved ulike estimeringer av SFA-modellen forandres signifikansnivået til de uavhengige variablene avhengig av hvilke variabler vi inkluderer. Beholder vi de samme variablene som i $F2$ får vi insignifikante verdier på koeffisientene til *lønn* og *andre driftskostnader*, mens *senger* beholder et signifikant påvirkningsnivå. En av antagelsene ved SFA er at man kjenner den funksjonelle formen til produktfunksjonen. Med tanke på at produksjonsprosessen ved sengepostene ikke er kjent, vil det være logisk å benytte seg av samme antagelser som i DEA-modellen.

5.2 Input

I arbeidet med å identifisere de relevante innsatsfaktorene for produksjon av DRG-poeng ved sengepostene, har vi hatt vanskeligheter med å finne tydelige framgangsmåter i den relevante litteraturen. Innsatsfaktorene våre virker å være godt dokumenterte som passende input i DEA-modeller, spesielt når det kommer til produksjon av internasjonale og testede output som *dagpasienter*, *casemix justerte utskrivelser* og *casemix justerte innleggelser*. Kittelsen et al. (2009) bruker DRG-poeng som output, og har valgt en tilnærmet lik inputgruppe som det vi tenker å bruke. Det som skiller deres innsatsfaktorer fra våre er at vi også har valgt å inkludere en enhetsstørrelse som kalles *sengekapasitet* i tillegg til driftskostnader, som vi har delt inn i en del lønn og en del andre driftskostnader. Vi hadde også tilgang på data vedrørende årsverk/månedsværk, men vi valgte å bruke de reelle lønnsdataene framfor planlagte årsverk. En av hovedårsakene til dette er at hvis vi kun ser på årsverk vil effektene av sykefravær (kortsikt), innleie og dyrere arbeidskraft falle bort i modellen. Enhetsmålinger på andre driftskostnader ville ført til et meget høyt antall innsatsfaktorer innenfor denne kategorien. Dermed vil en fellesbenevnelse i form av aggregerte kostnader innenfor hver kategori fungere vel så bra, selv om vi mister den potensielle informasjonen modellen kunne gitt oss om disse elementene på enhetsnivå. De tre valgte inputene er alle tidligere benyttet i DEA-modeller innenfor helsevesenet (O'Neill, Rauner, Heidenberger, & Kraus, 2008).

Inputmålet lønn inneholder alle lønnskostnadene knyttet til sengeposten. For noen av sengepostene er det 30 forskjellige poster med ulike lønnskostnader og refusjoner, men felles for alle er at arbeidsgiveravgift, pensjonskostnader, overtid, sykepenger og innleie av ekstern arbeidskraft er inkludert i beregningen. Arbeidsgiveravgiften i Tromsø og Harstad/Narvik har en differanse på 2 prosent. Vi har valgt å ikke gjøre korrigeringer med tanke på dette, da det får minimale til ingen effekt på estimering. Årlig utgjør dette maksimalt en økning på 0,001 prosent i totale kostnader for de berørte sengepostene. Det årlige budsjettarbeidet setter rammene for hvor mange årsverk hver sengepost kan bruke, og hvor mye penger hver sengepost kan bruke på lønn. I datasettet finnes det også budsjettall for hver sengepost på årlig basis med hensyn på lønn.

Andre driftskostnader er også tall som sykehuset har knyttet til sengepostene. Denne posten inneholder alt fra medikamenter, medisinsk utstyr og vanlige forbruksvarer. Vi har riktignok også her sett oss nødt til å fjerne enkelte store kostnader som blir kostnadsført på sengepostene. Årsaken til dette er at enkelte medikamenter som TNF-hemmere⁴, ikke fører til DRG-produksjon for sengepostene, og tilstedeværelsen av disse ville ført til at andre driftskostnader, da spesielt ved de medisinske sengepostene, hadde blitt urealistisk høye. Datasettet inneholder også her budsjetterte årlige kostnader fordelt på sengepostene.

Den siste inputen, sengekapasitet er også knyttet til hver enkelt sengepost. Alle sengepostene rapporterer en normert mengde senger de har tilgjengelig. Vi har valgt å bruke gjennomsnittlig antall senger som mål på kapasitet. Disse reflekterer hvor mange senger sengeposten i gjennomsnitt har klargjort hver dag. Vi multipliserer dette tallet med 365 slik at vi får et mål på den årlige kapasiteten. I gjennomgangen av ulike DEA-studier innenfor helsevesenet presiserer også O'Neill et al. (2008) at senger er et godt mål på størrelse. Sherman (1984) benytter seg av tilgjengelige sengedøgn som en fast innsatsfaktor. Mulige årsaker til at normerte senger og gjennomsnittlig antall senger i bruk ikke er helt lik kan skyldes at det i perioder av året planlegges lavere kapasitet på grunn av ferieavvikling.

⁴ Biologiske betennelsesdempende medikamenter som blir brukt på pasienter med revmatiske sykdommer og på pasienter med hud- og tarmsykdommer

5.2.1 Faste innsatsfaktorer

Sykehus og sengeposter har store faste kostnader/innsatsfaktorer, og på kort sikt kan man si at majoriteten av sykehuskostnadene er faste (Roberts et al., 1999). I den aktuelle studien ble hele 84 (tidligere undersøkelser sier mellom 58 og 75) prosent av sykehuskostnadene definert som faste, og dette inneholder kapitalkostnader, lønnskostnader, kostnader knyttet til bygg og anlegg og kostnader knyttet til tilstedeværelsen av støttefunksjoner. I vårt datasett har vi en input (lønn) som ifølge Roberts et al. (1999) tilhører kategorien faste kostnader, og vi har også en fast input i form av sengekapasitet. Andre driftskostnader er lettere å knytte til pasienter, og kan dermed sees på som en ren variabel kostnad, enda denne posten også vil inneholde elementer av faste kostnader på kort sikt. Med tanke på at sykehuset har en viss fleksibilitet til å endre antall klargjorte senger og bemanning, er det vanskelig å behandle disse målene som rene faste innsatsfaktor. Vi har derfor valgt å ikke behandle disse som faste jamfør Zhu (2004), der det beskrives en modell for å håndtere faste/ukontrollerbare variabler i en DEA setting.

5.3 Output

Valg av output i tidligere DEA-analyser av helsesektoren varierer i forhold til hva slags sykehus man analyserer og i hvilken geografisk region analysen er foretatt. Amerikanske studier har ofte basert seg på *case-mix adjusted discharges/visits* (O'Neill et al., 2008), det vil si innleggelser/utskrivelser justert etter oppholdstype. Nordiske studier benytter seg gjerne av sykehusenes DRG-produksjon. Vi har fått tak i et utdrag fra en rapport den norske arbeidsgiverforeningen Spekter arbeidet med i (Spekter, 2010). I denne rapporten gjennomfører de benchmarking på tvers av norske sykehus. De presenterer en liste over aktuelle aktivitetsmål man kan benytte seg av. På avdelingsnivå er pasientopphold presisert som et alt for grovt mål siden de vil variere i omfang og ha ulike ressursbehov. Pasientdøgn er også en vanskelig indikator siden den kan være høy hvis en pasient ligger lenge, eller mange pasienter ligger i kort tid. Dette kan allikevel korrigeres ved å måle gjennomsnittsdøgn per pasient. Slik vi ser det vil det uansett være vanskelig å måle på tvers av diagnosegrupper siden det er forskjellig rekonvalesenstid for ulike behandlinger. DRG-poeng er presentert som et gunstig uttrykk for produksjon på grunn av at vektingen gjør ulike diagnoser sammenlignbare.

Vi har valgt å måle den totale aktiviteten ved sengepostene i form av produksjon av DRG-poeng. Det vil med andre ord bety at sengepostene bruker arbeidskraft, driftsmateriell og kapasitet (senger) for å lage DRG-poeng. DRG-poengene er rapportert etter 2011-logikk. Det betyr at prosedyrene fra 2010 er rapportert med vektene og rapporteringssystemet fra 2011. Dette gjør at poengene fra de to periodene lett kan sammenlignes. Som nevnt innledningsvis vil produksjonen av DRG-poeng foregå i flere avdelinger. Den rapporterte DRG-produksjonen som ender opp på sengeposten, har tilhørende kostnader som hører hjemme på operasjonsstuen, laboratorier og lignende. Som Dyson et al. (2001) poengterer, er det viktig at man ivaretar homogenitet i inputsettet. Vi har dermed justert DRG-poengene slik at de kan benyttes i en sammenligning på tvers av sengepostene.

Justering av DRG-poengene

DRG-poengene består av ni ulike elementer (For eksempel operasjonskostnad, pleiekostnad og laboratorium). Disse elementene har ulik vekt avhengig av hvilken DRG det er snakk om. Siden vi ser på sengepostene hos UNN (som primært driver med pleie og rekonvalens) har vi valgt å beholde hver enkelte DRGs grunn- og pleieelementet. Grunnkostnaden utgjør restkostnadene man ikke kan plassere på et bestemt kostnadselement (Helsedirektoratet, 2011d). Siden disse utgjør kostnader knyttet til pasientens opphold har vi valgt å beholde dem i korrigeringen av sengepostens DRG-produksjon. Pleieelementet er selvsagt inkludert siden sengepostene er et pleiested.

UNN har gitt oss en oversikt over all DRG-produksjon i årene 2010 og 2011 basert på 2011 logikk. Videre har vi sortert ut de ulike sengepostene, slik at vi i tur og orden har kunnet sammenstilt deres produksjon innenfor hver enkelte DRG mot et regneark fra Helsedirektoratet som inneholder informasjon om hvordan vektene til hver DRG er bygd opp.

Tabell 5 viser et utdrag fra korrigeringen av sengeposten *Store barn*.

Tabell 5: Eksempel på justering av DRG-poeng

DRG		Grunnkostnad	Pleiekostnad	DRG	P2011	P2010	Jus2011	Jus2010
19	Sykdm. i hjernenerv. og per. nerver u/bk	0,67	0,06	19	1,40	2,80	1,02	2,03
20	Infks. i nervesys ekskl viral meningitt	0,70	0,14	20	27,95	22,40	23,49	18,83
21	Viral meningitt	0,78	0,11	21	0,00	4,06	0,00	3,61
22	Hypertensiv encefalopati	0,00	0,00	22	0,00	0,00	0,00	0,00
23	Ikke-traumatisk koma og stupor	0,39	0,08	23	0,00	0,00	0,00	0,00
24	Kramper og hodepine > 17 år m/bk	0,69	0,12	24	0,00	0,00	0,00	0,00
25	Kramper og hodepine > 17 år u/bk	0,74	0,08	25	0,00	0,00	0,00	0,00
26	Kramper og hodepine 0-17 år	0,74	0,11	26	49,56	30,37	41,83	25,63
27	Alvorlig traumatisk hjerneskade	0,59	0,10	27	4,03	5,50	2,78	3,79

Vi har lagt all DRG-produksjon for årene 2010 og 2011 inn i et eget ark for den aktuelle sengeposten, og så koblet denne mot det overordnede regnearket med all vektinformasjonen. Videre har vi konstruert en spørring som sorterer produksjonen basert på identifikatoren *DRG*. Denne søker etter likhet i DRG-nummeret, og vil sortere avdelingens DRG-poeng ved siden av det korrigerende DRG-nummeret i vektarket. Kolonnene *Jus(tert)2011* og *Jus2010* består av verdien i den tilhørende poeng-cellen (*P2010-11*) multiplisert med *Grunnvekt+Pleievekt*. Til slutt summeres disse to kolonnene for å få sengepostens totale justerte DRG-produksjon for de to periodene. Nedenfor i tabell 6 følger en oversikt over hvor stor andel av DRG-produksjonen som hører hjemme på sengepostene, samt en liten deskriptiv statistikk.

Tabell 6: Deskriptiv statistikk av sengepostenes pleie -og grunnandel

	Pleieandel		Grunnandel	
	2010	2011	2010	2011
Gastrokir. post Tromsø	0,089	0,098	0,424	0,455
Ort/plastikkir post Tromsø	0,069	0,068	0,369	0,365
Kreft post Tromsø	0,122	0,111	0,627	0,590
Uro/Gyn post Tromsø	0,086	0,088	0,477	0,475
Hjertemed. post Tromsø	0,067	0,068	0,338	0,333
Kirurgisk post Harstad	0,085	0,086	0,443	0,442
Hjerte/lu/karkir post Tromsø	0,110	0,106	0,440	0,433
Hud/ØNH/Revma post Tromsø	0,092	0,090	0,568	0,538
Nevrologisk post Tromsø	0,058	0,058	0,377	0,363
Kirurgisk post Narvik	0,087	0,089	0,487	0,492
Øye/Nevkir. post Tromsø	0,038	0,039	0,208	0,208
Medisinsk post Narvik	0,120	0,120	0,573	0,588
Lungemed. post Tromsø	0,134	0,136	0,605	0,630
Føde/barsel/photel post Tromsø	0,225	0,224	0,642	0,649
Med Harstad	0,117	0,112	0,592	0,564
Føde/gyn post Harstad	0,176	0,174	0,594	0,590
Infeksjonsmed post Tromsø	0,105	0,106	0,559	0,551
Hematol/Endo post Tromsø	0,104	0,099	0,598	0,564
Føde post Narvik	0,182	0,182	0,619	0,626
Gastro/nyre/gerat post Tromsø	0,095	0,101	0,551	0,562
Store barn	0,097	0,094	0,503	0,484
Gjennomsnitt	0,108	0,107	0,505	0,500
Maks	0,225	0,224	0,642	0,649
Min	0,038	0,039	0,208	0,208

Tabellen over gir et bilde på hvor stor pleieandelen er på de enkelte sengepostene. Det kan merkes at gjennomsnittene for pleie- og grunnandelen er lik landsgjennomsnittene presentert i Helsedirektoratets vektarbeid.

Tabell 7: Deskriptiv statistikk input og output 2010

2010				
	Lønnskostnader (i NOK)	Andre driftskostnader (i NOK)	Sengedøgn (klargjort)	DRG-Justert
Gjennomsnitt	22 355 894	3 019 986	7594	1117
Standardavvik	7 122 143	1 948 637	2577	533
Min	9 424 657	552 791	2555	350
Max	35 763 750	7 013 802	11680	2270

Tabell 8: Deskriptiv statistikk input og output 2011

2011				
	Lønnskostnader (i NOK)	Andre driftskostnader (i NOK)	Sengedøgn (klargjort)	DRG-Justert
Gjennomsnitt	23 410 869	3 451 961	7634	1118
Standardavvik	7 506 791	2 394 912	2656	517
Min	9 591 960	607 530	2555	326
Max	38 145 354	9 685 053	11607	2114

5.4 Kontrollvariabler

For å kunne svare på de postulerte hypotesene i problemstillingen, har vi valgt å inkludere en rekke kontrollvariabler.. Alle variabler kommer fra det opprinnelige datasettet fra UNN. De valgte kontrollvariablene er som følger

- Andel av totale utskrevne pasienter
- Kompleksitet
- Art (medisinsk/kirurgisk)
- Beleggsgrad

Beleggsgrad vil være et godt mål på hvor ofte sengepostene har ledig kapasitet. Vi antar at sengepostene med høyt belegg har muligheter til å holde et høyere aktivitetsnivå enn de med

lav beleggsgrad. Gitt at innsatsfaktorer som lønn og kapasitetskostnader er faste på kort sikt, kan det være et positivt forhold mellom effektivitet og beleggsgrad. Sengepostene med høyt belegg har i utgangspunktet bedre forutsetninger for å prestere mer effektivt enn de med lavt belegg.

Siden UNN (som alle sykehus) er nødt til å opprettholde en planlagt bemanning i tillegg til å ikke kunne kontrollere pasientmengden fullstendig, har vi valgt å måle effektiviteten mot hver sengeposts *andel av totalt utskrevne pasienter*. Selv om DRG-systemet fungerer, vil det kunne antas at en større andel av pasientene vil gi høyere effektivitet (med mindre planlagt aktivitet er lik faktisk aktivitet).

Kontrollvariabelen *Kompleksitet* kan gi svar på om kompleksiteten hos sengepostenes pasienter har en innvirkning på effektiviteten. Det justerte DRG-målet inneholder kun produksjon som kan tilegnes sengeposten. I et tilfelle der man har to relativt like sengeposter sammenlignet i størrelse, kostnadsnivå, og betjente pasienter, vil den mest komplekse ha en høyere ujustert DRG-produksjon. I tilfelle justeringsvekten for disse postene er noenlunde likt, vil den komplekse oppnå høyere produksjon i forhold til innsatsfaktorene. Denne kontrollvariabelen kan gi et svar på om DRG-systemet klarer å homogenisere outputen tilstrekkelig.

I en rapport for Beregningsutvalget for spesialhelsetjenesten poengterer Holmeide et al. (2008) at det er en underdekning for medisinske DRGer samtidig som kirurgiske DRGer opplever en overdekning. De faktiske kostnadene samsvarer altså ikke med refusjonen. For å teste om disse eventuelle forskjellene kommer frem i effektivitetstallene har vi inkludert en dummyvariabel *Art* for å kontrollere for forskjeller.

Tabell 9: Korrelasjon mellom input, output og kontrollvariabler 2011

	<i>Lønnskostnad</i>	<i>Andre driftskost</i>	<i>Sengekapasitet</i>	<i>justert DRG</i>	<i>andel med/kir</i>	<i>komp</i>	<i>belegg</i>
Lønnskostnad	1,00						
Andre driftskostnader	0,31	1,00					
Sengekapasitet	0,81	0,37	1,00				
justert DRG	0,65	0,20	0,80	1,00			
andel med/kir	0,66	-0,13	0,66	0,69	1,00		
komp	0,09	-0,23	0,28	0,39	0,13	1,00	
belegg	0,17	0,03	0,22	0,55	0,25	0,53	1,00
	0,29	0,05	0,13	0,20	0,39	-0,13	0,50

5.5 Valg av modell

Det ligger flere betraktninger i grunn for valg av modell. For det første er det viktig å finne robuste indikatorer på input og output. For det andre er det ønskelig å velge en modell som er konsistent med tidligere effektivitetsanalyser innenfor helsevesenet. I diskusjonen over har vi rasjonalisert for våre valg av input- og outputvariabler, mens vi i dette avsnittet drøfter valg av modell.

I Chilingirian og Sherman (2011) legger forfatterne frem ulike modeller basert på hva man analyserer. På grunn av manglende modeller i forhold til en intern analyse, har vi valgt å ta utgangspunkt i *Urban Hospital Model* (Ozcan, Luke, & Haksever, 1992).

Tabell 10: Urban Hospital Model

4 INPUTS	3 OUTPUTS
Number of Diagnostic & Specialty Services	Case Mix Adjusted Discharges
FTE Labor	Quantity of Outpatient Visits
Supplies & Materials	FTE Trainees
Expenditures	
Quantity of Beds	

Den skisserte modellen passer bra, siden UNN besitter karakteristikene til et urbant sykehus (sentralsykehus) og opprettholder et fullt tjenestetilbud. Vi er nødt til å endre på deler av modellen grunnet vår interne analyse av sengepostene. På inputsiden gjør vi én endring. Siden vi kun analyserer sengepostene har vi fjernet støttefunksjonene fra inputsiden. Deres bidrag til DRG-produksjonen er dessuten korrigert bort ifra datasettet. Vi beholder personellinnsats, ikke i form av årsverk, men som totale lønnskostnader. Andre driftskostnader står som de er, og antall senger er multiplisert med 365 for å få tilgjengelig kapasitet. På outputsiden har vi valgt å erstatte Case-mix justerte utskrivelser med DRG-produksjon. Dette er konsistent med (Biørn et al., 2003) og (Kittelsen et al., 2009). Outpatient(poliklinikk)besøk fjernes siden det kun er dagkonsultasjoner/dagopphold. Antall lærlinger har vi ikke tall på, og vil dermed ikke inkluderes. Vi står igjen med følgende modell:

Tabell 11: Vår modell

3 INPUT	1 OUTPUT
Lønnskostnader	Justert DRG-Produksjon
Andre driftskostnader	
Sengedøgn	

I resultatdelen av oppgaven presenterer vi resultatet av modellen ved å gjennomføre en DEA-analyse i omhyllingsform (2). Vi har valgt å gjøre analysen i inputretningen (minimere input) fordi det er stor usikkerhet knyttet til pasientstrømmen, og de variablene UNN kan kontrollere ligger på inputsiden. En outputmaksimerende modell ville antatt at hver sengepost drev kostnads- og kapasitetsoptimalt, noe vi ikke kan anta - gitt UNNs høye gjennomsnittskostnader i forhold til resten av landet. Et annet moment er at utnyttelsen av sengekapasiteten allerede er høy, så det vil være vanskelig å generere flere DRG-poeng gjennom økt pasientbehandling (med mindre den gjennomsnittlige liggetiden er for høy per dags dato). Resultatene presenteres både med antagelsene om konstant og variabelt skalautbytte. Resultatene fra Banker-testene viser at det er signifikante forskjeller mellom CRS og VRS-effektivitetene i våre data (Vedlegg 4).

Som et tillegg til DEA-analysen kommer vi også til å presentere resultatene fra en stokastisk frontanalyse som en robusthetstest. Modellen vi benytter oss av beholder de samme målene, i logaritmisk form, på input og output som i tabell 11.

$$(10) \ln(jDRG) = \beta_1 + \beta_2 \ln(LK_i) + \beta_2 \ln(ADK_i) + \beta_3 \ln(SK_i) + (v_i + u_i)$$

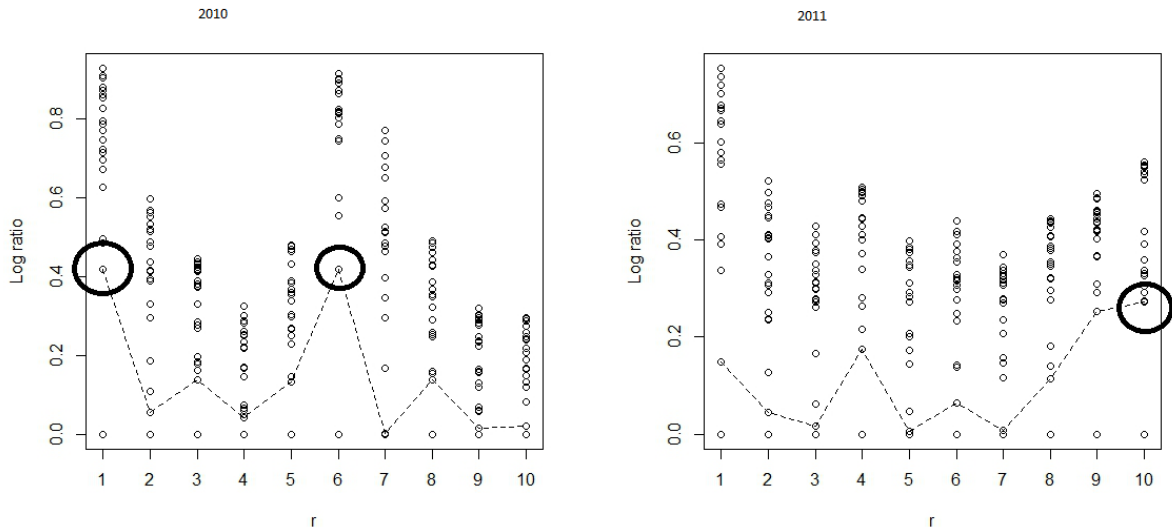
Ved estimeringen av modellen finner vi at variansen til det systematiske feilleddet v_i kollapser. En antagelse ved ordinær regresjon er at feilleddets varians er konstant for alle observasjoner. Viser det seg at variansen er positivt korrelert med en av datasettets forklarende variabler vil man ha et heteroskedastisk feilledd. Et vanlig tilfelle er at det tilfeldige feilleddet v_i , det bedriftsspesifikke feilleddet u_i – eller begge samtidig – varierer i samsvar med størrelsen på den enkelte DMU (Kumbhakar & Lovell, 2003, pp. 116-118). Vi har derfor estimert (8) på ny, der det stokastiske feilleddet varierer i samsvar med størrelsesvariabelen *sengekapasitet*.

$$(11) \ln(jDRG) = \beta_1 + \beta_2 \ln(LK_i) + \beta_2 \ln(ADK_i) + (v_i + u_i)$$

Forskjellen her er at vi har fjernet *sengekapasitet* eksplisitt fra modellen. Nå som modellen tillater heteroskedastisitet har vi pålagt det stokastiske feilleddet å variere som en funksjon av størrelsesvariabelen z_i (*sengekapasitet*) gjennom $\sigma_{v_i}^2 = g_1(z_i; \delta_i)$ slik det er beskrevet i (Kumbhakar & Lovell, 2003). I vedlegg 6 ligger statakodene for estimeringen, og beviset på heteroskedastisitet.

5.6 Outliers

Vi har valgt å anvende metoden beskrevet i kapittel 4.4.1 på våre data for å se om det er potensielle observasjoner som kan karakteriseres som outliers i forhold til resten av datasettet.



Figur 5: Outlieranalyse 2010 og 2011

Tabell 12: Oversikt over hvilke enheter som fjernes i hvert steg 2010

i	R(i)
[1,] 7	0,38
[2,] 5 7	0,20
[3,] 5 7 20	0,10
[4,] 5 7 15 20	0,06
[5,] 3 5 7 15 18	0,03
[6,] 3 5 7 14 15 18	0,01
[7,] 2 3 5 7 14 15 18	0,00
[8,] 2 3 5 7 14 15 18 20	0,00
[9,] 2 3 5 6 7 14 15 18 20	0,00
[10,] 2 3 4 5 6 7 14 15 18 20	0,00

Tabell 13: Oversikt over hvilke enheter som fjernes hvert steg 2011

i	R(i)
[1,] 7	0,45
[2,] 7 15	0,24
[3,] 3 7 15	0,13
[4,] 3 7 15 18	0,06
[5,] 3 7 14 15 18	0,03
[6,] 3 7 14 15 18 20	0,02
[7,] 3 7 14 15 18 20 21	0,01
[8,] 3 7 9 14 15 18 20 21	0,00
[9,] 3 5 7 9 14 15 18 20 21	0,00
[10,] 2 3 5 7 9 14 15 18 20 21	0,00

Figur 5 viser outlier-analysen med 2010 til venstre og 2011 til høyre. Grafisk observerer vi at det finnes ulike grupperinger/størrelser innenfor datasettet, men ingen virker ekstreme i forhold til resten. I 2010 virker DMUen i steg 1 å skille seg ut fra resten, mens steg 2 til 6 viser at det finnes 5 DMUer som er satt sammen forskjellig fra resten. I 2011 er det ingen av DMUene som skiller seg ut i de første stegene, og tolkningen av toppen mot slutten tyder på at det en plass midt i datasettet går et skille i form av størrelse eller sammensetningen av inputs og output. Resultatene tyder på at det kan eksistere forskjeller i faktorsammensetningen i datasettet, men vi velger å beholde alle de opprinnelige observasjonene videre i analysen.

Videre har vi brukt supereffektivitetstankegangen for å se om det finnes potensielle effektive DMUer som skiller seg såpass mye fra de andre at vi må justere datasettet for disse. Vi har gjort prosedyren beskrevet i kapittel 4.4.2.

Tabell 14: Supereffektivitet

crs 2010	super-effektivitet	crs 2011	supereffektivitet
Hjerte/lu/karkir post Tromsø	1,45	Hjerte/lu/karkir post Tromsø	1,46
Hjertemed. post Tromsø	1,26	Hjertemed. post Tromsø	1,32

Ingen av enhetene i analysen skiller seg ut som ekstreme i forhold til de andre, selv om 1,46 kan virke noe høyt. Supereffektivitet gjør en form for rangering av de effektive enhetene, og en høy effektivitetsscore relativt til 1 tyder på at DMUen av en eller annen grunn presterer signifikant bedre enn de andre analyseobjektene. Supereffektivitet som rangeringsmetode har sine svakheter. Adler et al. (2002) påpeker at det er en svakhet at DMUene vurderes etter sine egne vekter, pluss at man kan få en feilvurdering av de ineffektive DMUenes virkelige rangering som ikke hensyntas i metoden. Ved bruk av de opprinnelige DEA-estimatene og supereffektivitetsestimater rangerer disse målene de ineffektive sengepostene likt, etter at de effektive har fått sin egen rangering.

6 Resultat

I dette kapitlet skal vi presentere resultatene av vår undersøkelse. Selve DEA-undersøkelsen baserer seg på dataene i vedlegg 5. Vi har valgt å dele kostnader tilknyttet sengepostene inn i to grupper: lønnskostnader og andre driftskostnader. I tillegg har vi inkludert en variabel som beskriver kapasitet i form av sengepostenes tilgjengelige sengerdøgn. Som sluttprodukt eller mål på aktiviteten ved sengepostene har vi valgt en justert versjon av DRG-poengene.

Produktfunksjonen vi baserer oss på inneholder både pengestørrelser (lønnskostnader og utstyrs/medisin-kostnader) og enhetsstørrelser (sengerdøgn tilgjengelig). Vi har valgt å gjøre analysen i inputretningen, dvs. at vi ønsker å se på hvor mye hver enkelt sengepost kan minke sitt forbruk av innsatsfaktorene når man holder produksjonen av DRG-poeng konstant.

Totaleffektiviteten som fremkommer ved antakelse om konstant skalautbytte tar ikke hensyn til skala. Dermed har vi valgt å inkludere en hypotese om at skala kan være en årsak til effektivitetsforskjeller. Vi har ved å bruke Bankertestene beskrevet i kapittel 4.7 testet hvorvidt det er forskjell på CRS og VRS estimatene. Resultatene finnes i vedlegg 4. Alle teststatistikkene fra Banker-testene rapporterer en signifikant forskjell mellom CRS og VRS-estimatene på et 95 % konfidensnivå.

6.1 Sengepostenes effektivitet 2010

Ved å løse omhyllingsmodellen (2) for CRS og (3) for VRS får vi følgende resultater.

Tabell 15: Effektivitet 2010

	DEA-CRS	DEA-VRS
<i>Gastrokir. post Tromsø</i>	0,64	0,69
<i>Ort/plastikkir post Tromsø</i>	0,85	0,91
<i>Kreft post Tromsø</i>	0,57	0,67
<i>Uro/Gyn post Tromsø</i>	0,54	0,65
<i>Hjertemed. post Tromsø</i>	1,00	1,00
<i>Kirurgisk post Harstad</i>	0,61	0,70
<i>Hjerte/lu/karkir post Tromsø</i>	1,00	1,00
<i>Hud/ØNH/Revma post Tromsø</i>	0,71	0,85
<i>Nevrologisk post Tromsø</i>	0,32	0,53
<i>Kirurgisk post Narvik</i>	0,79	0,96
<i>Øye/Nevkir. post Tromsø</i>	0,52	0,72
<i>Medisinsk post Narvik</i>	0,85	1,00
<i>Lungemed. post Tromsø</i>	0,56	0,71
<i>Føde/barsel/photel post Tromsø</i>	0,74	0,76
<i>Med Harstad</i>	0,81	0,86
<i>Føde/gyn post Harstad</i>	0,83	1,00
<i>Infeksjonsmed post Tromsø</i>	0,41	0,81
<i>Hematol/Endo post Tromsø</i>	0,48	0,86
<i>Føde post Narvik</i>	0,43	1,00
<i>Gastro/nyre/gerat post Tromsø</i>	0,33	0,42
<i>Store barn</i>	0,48	0,67
<i>Gjennomsnitt</i>	0,64	0,80
<i>Standardavvik</i>	0,20	0,17
<i>Min</i>	0,32	0,42
<i>Maks</i>	1,00	1,00

Tabell 15 viser den rapporterte effektiviteten ved sengepostene i 2010. Ved konstant skalautbytte ligger *hjertemedisinsk-* og *hjerte/lunge/karkirurgisk* sengepost på effektivitetsfronten. Tar man hensyn til at sengepostene opererer under forskjellige skalaforutsetninger vil *fødepost Narvik*, *fødepost Harstad* og *medisinsk sengepost Narvik* også komme ut til den effektive fronten. Gjennomsnittseffektiviteten for CRS og VRS er henholdsvis 64- og 80 prosent. Basert på forutsetningen om konstant skalautbytte (CRS) er *hjertemedisinsk sengepost* benchmark for 11 sengeposter, og *hjerte/lunge/kar-kirurgisk*

sengepost er benchmark for 18 poster. Når vi endrer modellen til VRS får vi nye effektive enheter som også inngår i måling av optimal bruk av innsatsfaktorene.

Basert på omhyllingsmodellen (2) og målberegningene basert på de rapporterte lambdaer og referanse-DMUer kan vi se på hvor mye sengepostene i gjennomsnitt må redusere sine innsatsfaktorer for å bli effektive. Resultatene er rapportert under i tabell 16.

Tabell 16: Mål på optimal reduksjon 2010

CRS 2010			
	Lønn	Driftskostnader	Senger
Gjennomsnitt	39,9 %	48,8 %	43,8 %
Standardavvik	22,2 %	27,0 %	18,0 %
Min	0,0 %	0,0 %	0,0 %
Maks	73,0 %	90,0 %	68,1 %

VRS 2010			
	Lønn	Driftskostnader	Senger
Gjennomsnitt	22,5 %	37,5 %	23,5 %
standardavvik	18,2 %	29,1 %	16,4 %
Min	0,0 %	0,0 %	0,0 %
Maks	60,4 %	84,2 %	57,5 %

I CRS-modellen er det *Nevrologisk post Tromsø* som trenger størst reduksjon i lønn og senger for å bli effektiv, mens *Hematologisk/Endo post Tromsø* må redusere sine andre driftskostnader med hele 90 % for å bli effektiv. Dette fordi de har det tredje høyeste forbruk av *andre driftskostnader*, noe som er altfor høyt i forhold til DRG-produksjonen som er fjerde lavest i settet. For VRS er den største reduksjonen i kostnader og senger lavere enn i CRS, og det er *Gastro/nyre/geriatriisk post* som trenger størst reduksjon i lønnskostnader og senger, mens *Hematologisk post Tromsø* fortsatt trenger å redusere sine andre driftskostnader med 84,2 prosent for å bli effektive i den retningen.

6.2 Sengepostenes effektivitet 2011

Ved å løse omhyllingsmodell (2) for CRS og (3) for VRS får vi følgende resultater.

Tabell 17: Effektiviteter 2011

	DEA-CRS	DEA-VRS
<i>Gastrokir. post Tromsø</i>	0,68	0,73
<i>Ort/plastikkir post Tromsø</i>	0,89	0,93
<i>Kreft post Tromsø</i>	0,62	0,72
<i>Uro/Gyn post Tromsø</i>	0,67	0,77
<i>Hjertemed. post Tromsø</i>	1,00	1,00
<i>Kirurgisk post Harstad</i>	0,67	0,74
<i>Hjerte/lu/karkir post Tromsø</i>	1,00	1,00
<i>Hud/ØNH/Revma post Tromsø</i>	0,72	0,90
<i>Nevrologisk post Tromsø</i>	0,32	0,54
<i>Kirurgisk post Narvik</i>	0,69	0,88
<i>Øye/Nevkir. post Tromsø</i>	0,49	0,72
<i>Medisinsk post Narvik</i>	0,78	1,00
<i>Lungemed. post Tromsø</i>	0,64	0,77
<i>Føde/barsel/photel post Tromsø</i>	0,68	0,71
<i>Med Harstad</i>	0,76	0,81
<i>Føde/gyn post Harstad</i>	0,71	1,00
<i>Infeksjonsmed post Tromsø</i>	0,43	0,82
<i>Hematol/Endo post Tromsø</i>	0,52	0,89
<i>Føde post Narvik</i>	0,44	1,00
<i>Gastro/nyre/gerat post Tromsø</i>	0,35	0,44
<i>Store barn</i>	0,54	0,73
<i>Gjennomsnitt</i>	0,65	0,81
<i>Standardavvik</i>	0,19	0,15
<i>Min</i>	0,32	0,44
<i>Maks</i>	1,00	1,00

Tabell 17 viser den estimerte effektiviteten for 2011. Som i 2010 er det *hjertemedisinsk* og *hjerte/lunge/karkirurgisk sengepost* som er effektive i CRS-modellen. Løser man VRS-modellen kommer også *fødepost Harstad*, *medisinsk post Narvik* og *fødepost Narvik* ut på effektivitetsfronten, noe som også er likt for 2010. Den gjennomsnittlige effektiviteten er her henholdsvis 65 prosent for CRS og 81 prosent for VRS. Selv om det er en høyere gjennomsnittlig effektivitet enn i 2010 kan man ikke si at det har vært en forbedring. Dette kan analyseres ved hjelp av andre modeller, men går utenfor omfanget til oppgaven. For CRS-

modellen er hjerte/lunge/karkirurgisk sengepost benchmark for nitten sengeposter og hjertemedisinsk sengepost er benchmark for syv andre DMUer.

Ved å se på de rapporterte lambdaene og referanse-DMUene fra omhyllingsmodellene kan man beregne hvor mye DMUene i gjennomsnitt må redusere sine inputs for å bli effektive. Resultatene er rapportert i tabell 18.

Tabell 18: Mål på optimal reduksjon 2011

CRS 2011	Lønnskostnader	Driftskostnader	Senger
Gjennomsnitt	39,1 %	52,3 %	41,9 %
Standardavvik	21,4 %	27,9 %	17,7 %
Min	0,0 %	0,0 %	0,0 %
Maks	72,4 %	91,3 %	68,4 %

VRS 2011	Lønnskostnader	Driftskostnader	Senger
Gjennomsnitt	21,6 %	40,4 %	21,8 %
Standardavvik	17,4 %	31,6 %	16,2 %
Min	0,0 %	0,0 %	0,0 %
Maks	57,3 %	86,9 %	56,5 %

I CRS-modellen er det *Nevrologisk post Tromsø* som trenger størst reduksjon i lønn og senger for å bli effektiv, mens *Hematologisk/Endo post Tromsø* må redusere sine andre driftskostnader med hele 91 % for å bli effektiv. Årsaken til at *Nevrologisk post Tromsø* må redusere sine lønnskostnader og sengekapasitet er at den sammenlignbare sengeposten, *Hjerte/lu/karkir post Tromsø*, har en DRG-produksjon som er fire ganger så høy.

For VRS er den største reduksjonen i kostnader og senger lavere enn i CRS, og det er *Gastro/nyre/geriatrisk sengepost* som trenger størst reduksjon i lønnskostnader og senger, mens *Hematologisk/endo post* fortsatt trenger å redusere sine andre driftskostnader med 87 %.

6.4 Stokastisk frontanalyse – 2010 og 2011

Nedenfor er utskrift av resultatene av SFA-estimeringen (11) for 2010 og 2011.

Tabell 19: SFA-data 2010

Log	likelihood	=	-7,8				
Variabel:	Koeffisient	Stdfeil	z	P> z	95% konf.int		
ln lønn	1,03	0,35	2,92	0,00	0,33	1,72	
ln driftskost	0,03	0,12	0,24	0,81	-0,20	0,25	
konstant	-10,45	5,46	-1,91	0,06	-21,15	0,25	
ln sig2v:							
ln senger	-1,92	2,15	-0,89	0,37	-6,15	2,31	
konstant	13,60	18,58	0,73	0,46	-22,80	50,00	
ln sig2u:							
konstant	-1,30	0,65	-1,98	0,05	-2,58	0,01	
sigma_u	0,52	0,17			0,27	1,00	
Lambda	0,53						

Tabell 20: SFA-data 2011

Log	likelihood=	=	-5,81				
Variabel:	Koeffisient	Std.feil	z	P> z	95% konf.int		
ln lønn	1,24	0,34	3,64	0,00	0,57	1,90	
ln driftskost	-0,05	0,10	-0,55	0,58	-0,23	0,13	
konstant	-12,83	5,42	-2,37	0,018	-23,45	-2,22	
ln sig2v:							
ln senger	-2,10	2,00	-1,06	0,29	-6,00	1,80	
konstant	14,85	17,51	0,85	0,40	-19,47	49,20	
ln sig2u:							
konstant	-1,40	0,48	-2,93	0,00	-2,32	-0,46	
sigma_u	0,50	0,12			0,31	0,79	
Lambda	0,57						

Effektene av variablene er noenlunde like i begge årene. Det vi har gjort som skiller dette fra en ”rett fram” stokastisk frontanalyse, er at vi har håndtert sengepostenes størrelse ($\ln \text{ senger}$) som nettopp en størrelse og ikke som en faktor i produktfunksjonen. Vi antar at størrelsen i seg selv ikke har noen direkte effekt på produktfunksjonen; de ønsker alle å maksimere produksjonen, men størrelsen har en effekt på hvordan de opptrer (lik store og små selskaper har forskjellige opptreden, eller rike og fattige segmenter (sub-populasjoner) har forskjellige opptreden) som da gir heteroskedastisitet. Ved å modellere variansen til det stokastiske feilleddet ($\sigma^2 v$) som en funksjon av størrelse får vi mer korrekte resultater (Kumbhakar & Lovell, 2003).

$\sigma^2 v$ er variansen til det stokastiske feilleddet, og dette er nå en funksjon av støy i modellen og sengepostenes størrelse. $\sigma^2 u$ er variansen til det bedriftsspesifikke feilleddet. For 2010 er 22 prosent av variansen i modellen forklart av ren ineffektivitet (7), mens den i 2011 er steget til 25 prosent. $\sigma^2 u$ er signifikant på et 95 prosent konfidensnivå for begge årene. For begge årene er koeffisientene i $\ln \sigma^2 v$ riktignok ikke selvstendig signifikant forskjellig fra 0 basert på disse estimatene, og vi tester koeffisientene sammen ved hjelp av en Wald-test. Fra testen får vi en chi-kvadrert-verdi på 0,03 for 2010 og 0,34 for 2011. Dette er ikke nok til å avvise nullhypotesen om at koeffisientene sammen er lik null. Det må allikevel poengteres at dersom den stokastiske støyen eksisterer, slik koeffisientene er rapportert, så vil den dominere den rene tekniske ineffektiviteten, det vil si at høyest andel ineffektivitet stammer fra eksogene faktorer utenfor sengepostenes kontroll.

Tabell 21: SFA-effektiviteter

Sengepost:	sfa 2010	sfa 2011
Gastrokir. post Tromsø	0,76	0,80
Ort/plastikkir post Tromsø	0,85	0,88
Kreft post Tromsø	0,68	0,77
Uro/Gyn post Tromsø	0,66	0,81
Hjertemed. post Tromsø	0,89	0,85
Kirurgisk post Harstad	0,73	0,79
Hjerte/lu/karkir post Tromsø	0,92	0,94
Hud/ØNH/Revma post Tromsø	0,82	0,87
Nevrologisk post Tromsø	0,38	0,39
Kirurgisk post Narvik	0,78	0,70
Øye/Nevkir. post Tromsø	0,61	0,61
Medisinsk post Narvik	0,89	0,90
Lungemed. post Tromsø	0,68	0,72
Føde/barsel/photel post Tromsø	0,63	0,53
Med Harstad	0,90	0,91
Føde/gyn post Harstad	0,68	0,67
Infeksjonsmed post Tromsø	0,57	0,57
Hematol/Endo post Tromsø	0,57	0,63
Føde post Narvik	0,62	0,63
Gastro/nyre/gerat post Tromsø	0,35	0,37
Store barn	0,49	0,48
Gjennomsnitt	0,69	0,70
Standardavvik	0,16	0,17
Min	0,35	0,37
Maks	0,92	0,94

I tabell 21 vises de beregnede effektivitetene for 2010 og 2011. Analogt med DEA-estimatene har *hjerte/lunge/karkirurgisk sengepost* den høyest beregnede effektiviteten, mens det her er *gastro/nyre/geriatrisk sengepost* som har lavest rapportert effektivitet.

For å avgjøre skalaspørsmålet i SFA-undersøkelsen kan vi bruke en av egenskapene til en Cobb-Douglas produktfunksjon. Produktfunksjonen som SFA-estimatene i vår analyse er basert på er en Cobb-Douglas-funksjon. Hvis summen av koeffisientene er lik 1 er det konstant skalautbytte. Overstiger koeffisientene 1 er det tiltakende skalautbytte. Testen vi utfører for å sjekke om summen av koeffisientene til lønn og andre driftskostnader er lik 1 er en Wald-test. Vi utfører denne testen fordi den klarer å teste signifikansen til koeffisientene når vi slår dem sammen. Nullhypotesen er i vårt tilfelle

$\beta_{Lønnskostnad} + \beta_{Andre\ driftskostnader} - 1 = 0$, da vi tester om det er konstant skalautbytte.

Chi-kvadrert verdien for 2010 er 0,08 og for 2011 0,44. Vi kan ikke forkaste nulhypotesen for noen av årene. For 2010 og 2011 kan vi ikke si at vi ikke har konstant skalautbytte, men vi kan heller ikke utelukke variabelt skalautbytte.

6.5 Sensitiviteten og robustheten til effektivitetsestimaterne

Vi har utført stigemetoden (Edwardsen 2004) på CRS og VRS-estimaterne for begge årene i henhold til kapittel 4.6. Vi har sett på endringene i effektiviteten til hver enkelt DMU i deres tre første trinn. Store endringer i effektiviteten mellom hvert trinn tyder på at effektivitetsestimaterne er sensitive for feilmålinger eller bortfall av referanseenhetene. Metoden baserer seg på at man regner ut endringen til hver enkelt DMU hvis man fjerner den mest innflytelsesrike DMUen. Vi har ikke tatt hensyn til endringene som forekommer etter at DMUene har blitt effektive, det vil si supereffektivitet. Tallene som er rapportert her er gjennomsnittet av endringene til alle DMUene for de tre første stegene mot effektivitetsfronten.

Tabell 22: Sensitivitetsestimater basert på effektivitetsstiger

	Første steg	andre steg	tredje steg	Første steg	andre steg	tredje steg
CRS 2010				VRS 2010		
Gjennomsnitt	0,15	0,08	0,01		0,07	0,04
Standardavvik	0,07	0,07	0,01		0,07	0,05
Min	0,00	0,00	0,00		0,00	0,00
Maks	0,25	0,19	0,04		0,19	0,17
CRS 2011				VRS 2011		
Gjennomsnitt	0,17	0,02	0,04		0,06	0,03
Standardavvik	0,08	0,02	0,06		0,07	0,05
Min	0,00	0,00	0,00		0,00	0,00
Maks	0,29	0,07	0,22		0,22	0,15

Vi ser at CRS-estimaterne er mer sensitive enn VRS-estimaterne, og dette er et resultat av den tettere omhyllingen VRS fører til. Gjennomsnittlig effektivitetsøkning på 15 prosent i 2010 og 17 prosent i 2011 gjør at eventuelle målefeil i referanseenhetene vil få store konsekvenser for effektivitetsestimaterne våre. Den høyeste sensitiviteten i første steg er på hele 25 og 29 prosent for henholdsvis 2010 og 2011. Det er *lungemedisinsk post Tromsø* som er mest

sensitiv i det første steget i CRS-modellen, mens *infeksjonsmed post Tromsø* er mest sensitiv for effektivitetsendringer om man fjerner den mest innflytelsesrike sengeposten i VRS-modellene.

I CRS-estimatene er endringene i første steg og andre steg de høyeste, mens i VRS vil konsekvensen av å fjerne en "liten" effektiv DMU føre til at en liten DMU som i utgangspunktet er ineffektiv, blir en del av effektivitetsfronten. Dette er på grunn av omhyllingsegenskapene til VRS-modellen.

Denne metoden lar seg ikke bruke på SFA-estimatene, men siden SFA er basert på en statistisk metode kan sensitiviteten og robustheten evalueres via ulike tester knyttet til t-scorene til koeffisientene, feilledd og p-verdier. Det er stor usikkerhet knyttet til koeffisienten til *andre driftskostnader*, og effekten av denne er ikke statistisk signifikant forskjellig fra null, og den skifter fortegn mellom periodene. Usikkerheten knyttet til effekten av *andre driftskostnader* er en svakhet i modellen, og kan føre med seg potensielle problemer knyttet til tolkning og anvendelse av SFA-estimatene. Usikkerheten i SFA-estimatene kan også knyttes til verdien på σ_{2u} og σ_{2v} . σ_{2u} er signifikant forskjellig fra null på et 95 prosent konfidensnivå, så det er rimelig stor sikkerhet knyttet til at det faktisk finnes teknisk ineffektivitet i virkeligheten, mens det for σ_{2v} er knyttet større usikkerhet rundt tilstedeværelsen av varians via det stokastiske feilleddet. Hvis vi bruker koeffisientene slik de er presentert i tabell 19 og 20 vil den stokastiske variansen dominere den bedriftsspesifikke, men på grunn av usikkerhet rundt signifikansen til det stokastiske feilleddet kan det bedriftsspesifikke feilleddet stå for hele variasjonen.

6.6 Testing av hypotesene

Vi ønsker å se på hvilken effekt fire eksogene variabler har på effektivitetsestimaterne våre. Tre av variablene er tallstørrelser, og variabelen som tester for forskjell mellom medisinske og kirurgiske poster er en dummyvariabel, der 0=medisinsk og 1=kirurgisk sengepost. Som et utgangspunkt ønsker vi å estimere følgende regresjonslikning. Gitt hypotesene i kapittel 1.3, er utgangspunktet at de beregnede koeffisientene ikke er forskjellig fra null.

Vi tester dette gjennom en multippel regresjonsanalyse.

$$\text{Effektivitet} = \beta_0 + \beta_1 * \text{Andel utskrevne pasienter} + \beta_2 * D(\text{med|kir}) + \beta_3 * \text{Kompleksitet} + \beta_4 * \text{Beleggsgrad} + \varepsilon$$

6.6.1 Regresjon 2010

Resultatet av regresjonen for CRS, VRS og SFA-effektivitetene for 2010 vises henholdsvis i tabell 23,24 og 25. Tabell 26 viser korrelasjonen mellom kontrollvariablene.

Tabell 23: Regresjonsresultater CRS 2010

CRS 2010				
Justert R-kvadrat	0,54			
	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>	<i>P-verdi</i>
Skjæringspunkt	0,61	0,10	5,94	0,00
andel	4,73	1,20	3,93	0,00
med/kir	-0,18	0,09	-1,89	0,08
komp	0,32	0,09	3,42	0,00
belegg	-0,73	0,19	-3,92	0,00

Tabell 24: Regresjonsresultater VRS 2010

VRS 2010				
Justert R-kvadrat	0,13			
	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>	<i>P-verdi</i>
Skjæringspunkt	0,91	0,12	7,83	0,00
andel	1,30	1,36	0,96	0,35
med/kir	-0,17	0,10	-1,61	0,13
komp	0,23	0,10	2,14	0,05
belegg	-0,54	0,21	-2,58	0,02

Tabell 25: Regresjonsresultater SFA 2010

SFA 2010				
Justert R-kvadrat	0,41			
	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>	<i>P-verdi</i>
Skjæringspunkt	0,69	0,09	7,44	0,00
andel	2,59	1,08	2,39	0,03
med/kir	-0,11	0,08	-1,27	0,22
komp	0,26	0,08	3,08	0,01
belegg	-0,57	0,17	-3,38	0,00

Tabell 26: Korrelasjon mellom kontrollvariablene 2010

	<i>andel</i>	<i>med/kir</i>	<i>komp</i>	<i>belegg</i>
andel	1,00			
med/kir	0,12	1,00		
komp	0,29	0,56	1,00	
belegg	0,45	-0,08	0,52	1,00

Den justerte R^2 (hvor ”godt” prediksjonen treffer de faktiske observasjonene) i tabell 23 forteller oss at 54 prosent av variasjonen i effektiviteten er forklart av variasjonen i kontrollvariablene, mens tabell 24 gir en tilsvarende justert R^2 på 13 prosent. For SFA-estimatene er den justerte R^2 -verdien 41 prosent. Videre vil man ved et 95 prosent konfidensnivå fastslå at tre av kontrollvariablene har en signifikant påvirkning på effektiviteten for både CRS og SFA. For VRS-estimatene er ikke effekten av å være medisinsk eller kirurgisk post eller andel utskrevne pasienter, signifikant. Fortegnene på koeffisientene er like i alle 3 regresjonene. Økning i den totale andelen av utskrevne pasienter og kompleksitet har en positiv påvirkning på effektiviteten, mens økning i belegg har en negativ innvirkning. Når dummyvariabelen er 1 (kirurgisk sengepost) vil dette redusere effektiviteten, men denne effekten er ikke statistisk signifikant.

6.6.2 Regresjon 2011

I tabell 27,28 og 29 vises henholdsvis resultatet av regresjonen for CRS-, VRS- og SFA-effektivitetene for 2011. Tabell 30 viser korrelasjonen mellom kontrollvariablene.

Tabell 27: Regresjonsresultater CRS 2011

CRS 2012				
Justert R-kvadrat	0,46			
	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>	<i>P-verdi</i>
Skjæringspunkt	0,52	0,12	4,54	0,00
andel	4,25	1,27	3,35	0,00
med/kir	-0,10	0,09	-1,04	0,31
komp	0,25	0,09	2,80	0,01
belegg	-0,50	0,21	-2,42	0,03

Tabell 28: Regresjonsresultater VRS 2011

VRS 2011				
Justert R-kvadrat	-0,04			
	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>	<i>P-verdi</i>
Skjæringspunkt	0,90	0,13	6,76	0,00
andel	0,53	1,46	0,36	0,72
med/kir	-0,12	0,11	-1,08	0,30
komp	0,17	0,10	1,63	0,12
belegg	-0,39	0,24	-1,62	0,12

Tabell 29: Regresjonsresultater SFA 2011

SFA 2011				
Justert R-kvadrat	0,13			
	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>	<i>P-verdi</i>
Skjæringspunkt	0,66	0,13	4,96	0,00
andel	1,51	1,46	1,03	0,32
med/kir	-0,03	0,11	-0,24	0,82
komp	0,19	0,10	1,85	0,08
belegg	-0,36	0,24	-1,50	0,15

Tabell 30: Korrelasjon mellom kontrollvariablene 2011

	<i>andel</i>	<i>med/kir</i>	<i>komp</i>	<i>belegg</i>
andel	1,00			
med/kir	0,13	1,00		
komp	0,25	0,53	1,00	
belegg	0,39	-0,13	0,50	1,00

I motsetning til 2010, ser man av tabell 27 og 28 at forklaringsgraden i den justerte R^2 har falt for både CRS og VRS, der den nå er på henholdsvis 46 prosent for CRS og -4 prosent for VRS. Negativ justert R^2 tyder på regresjonen passer dataen dårlig, men man kan fremdeles finne signifikante sammenhenger. Som vi ser av tabell 29 er kun 13 prosent av den totale variasjonen i den SFA-estimerte effektiviteten forklart av variasjonen i kontrollvariablene. For CRS-estimatene har vi samme sammenhengen som i 2010, mens både VRS og SFA ikke

lenger finner noen signifikante sammenhenger mellom kontrollvariablene og effektivitet. 2011-estimatene får samme fortegn på alle koeffisientene som 2010-tallene, men vi ser færre signifikante sammenhenger.

I tabell 26 og 30, vises korrelasjonen mellom forklaringsvariablene. Den høyeste korrelasjon (mellom *kompleksitet og med/kir*) er litt over 50 prosent. En av utfordringene ved multippel regresjon er multikollinearitet. Det vil si at en eller flere variabler samvarierer, og dermed kan forårsake insignifikans i estimatene, eller forandre fortegnet på koeffisientene. Å se på den parvise korrelasjonen er ikke nok for å søke etter multikollinearitet, dermed har vi beregnet en hjelpemodell for kontrollvariablene forklart i kapittel 6.4.3 i (Hill, Griffiths, & Lim, 2008). Den høyeste R^2 verdien vi får rapportert er godt under deres tenkte cut-off-verdi på 0,8. Vi har dermed valgt å beholde OLS-estimatene.

6.7 Effektivitet og sammenhenger

Den første hypotesen vi undersøker er "*Belegg/utnyttelse av kapasitet har ingen innvirkning på effektivitet*". For *belegg* er de rapporterte sammenhengene vi fant i 6.7 negative i alle modellene, der VRS- og SFA-estimatene fra 2011 ikke kan sies å være signifikant forskjellig fra null. Denne sammenhengen kan ved første øyekast virke underlig. En rasjonell betraktning vil være å koble en økning i utnyttelse av sengene mot en økning i effektiviteten. Et poeng som kan støtte denne negative sammenhengen, er liggetid over normert tid. En sengepost med et belegg på tilnærmet hundre prosent kan ha et stort antall pasienter som ligger ut over refusjonstiden gitt av DRG-systemet. Denne kapasitetsbruken vil man dermed ikke premieres for, men heller straffes for i form av at kostnadene løper uten DRG-generering. Vi velger dermed å betrakte den første nullhypotesen som avkreftet, og fastslår at beleggsgrad har hatt en negativ innvirkning på effektiviteten.

Vår andre hypotese er "*Det eksisterer ingen effektivitetsforskjeller mellom kirurgiske og medisinske sengeposter*". Dummyvariabelen vi tester er lik én for kirurgiske sengeposter. Sammenhengen er predikert negativ for alle modeller, men vi finner ingen signifikante sammenhenger. Denne hypotesen var grunnlagt i Holmeide et al. (2008) som fastslo en underdekning av medisinske DRGer og en overdekning av kirurgiske. Man kunne derfor forventet en positiv sammenheng. En mulig forklaring på at denne er negativ kan være at de

kirurgiske sengepostene i snitt har vært utsatt for en større DRG-korreksjon enn de medisinske. Homogeniseringen i DRG-systemet skal sørge for at denne sammenhengen ikke skal eksistere. Årsaken til at den kommer frem kan være regionale forskjeller fra DRG-standarden, eller fordi DRG-systemet ikke egner seg som homogent aktivitetsmål på sengepostnivå. Dette er ikke et funn som står i motsetning til Holmeide, da vi kun har sett på sengepostene og deres tilhørende vektorer (grunnkostnad og pleiekostnad). Vi har ikke sett på alle kostnadsstedene ved et sykehus, slik at sammenhengen en fant i 2008 kan være gjeldene på sykehusnivå når hele pasientforløpet analyseres. Det må presiseres at vi ikke finner noen sammenfallende signifikans i modellene, og dermed kan vi ikke avslå den andre nullhypotesen

Den tredje hypotesen vår er *”Andel av utskrevne pasienter har ingen påvirkning på effektivitet”*. Her er det positive sammenhenger for alle modellene. Her finner vi ikke signifikante sammenhenger for VRS 2010 og 2011 og SFA 2011. En positiv sammenheng mellom totalandelen av alle utskrevne pasienter og effektivitet er logisk. Så lenge hver behandlede pasient gir DRG-refusjon, vil man få økt dekning av sine kostnader ved å øke sin pasientandel (nye pasienter). For en sengepost vil en relativ økning i pasientandelen følgelig gi andelsreduksjon for en annen. Hovedårsaken til at denne kontrollvariabelen ikke blir signifikant i VRS-modellen er at sengeposter med lav andel, hovedsakelig sengepostene i Harstad og Narvik får en effektivitetskompensasjon på grunn av størrelse. Grunnet dette velger vi å tolke sammenhengen som signifikant positiv. Vi anser dermed den tredje nullhypotesen som avkrefte og konstaterer at pasientandelen påvirker effektivitet positivt.

Den fjerde hypotesen er *”Kompleksiteten har ingen påvirkning på effektivitet”*. Etter inspirasjon fra Nyland and Bjørnenak (2000) har vi valgt å teste om kompleksitet har innvirkning på sengepostenes effektivitet. Kompleksitet er definert som hvor kompleks en pasient er i forhold til gjennomsnittspasienten. Høy kompleksitetsscore tyder på at sengeposten behandler diagnoser med høye DRG-poeng. I vårt tilfelle er det signifikante positive sammenhenger mellom kompleksitet og alle effektivitetene med unntak av VRS og SFA i 2011. Dette resultatet kan tolkes dit hen at de effektive sengepostene ved UNN har høy grad av kompleksitet, og at komplekse pasienter, slik disse her er definert, gir høyere inntekter i forhold til kostnadene enn mindre komplekse pasienter. Dette strider mot funn fra Nyland og Bjørnenak som argumenterer for at kompleksitet er en betydelig kostnadsdriver ved norske sykehus, og våre analyser kan peke mot at komplekse pasienter overfinansieres i DRG-

systemet. Funnene våre leder til at vi anser den fjerde nullhypotesen som avkreftet og konkluderer med en positiv sammenheng mellom effektivitet og kompleksitet.

6.8 Sammenlikning av analysemetodene

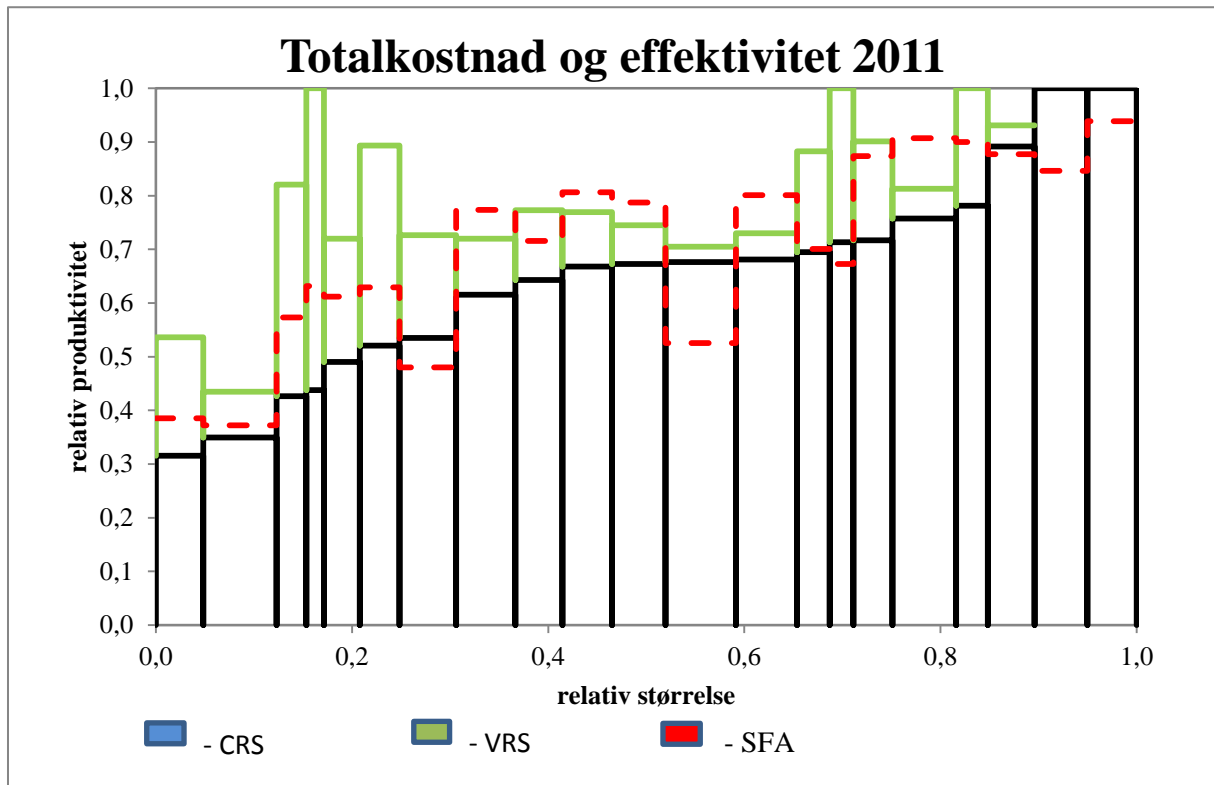
De ulike modellene i analysen gir en ulik rangering av sengepostene, samt en forskjell i gjennomsnittlig effektivitet. I avsnittene under vises en grafisk sammenstilling av de tre ulike effektivitetsberegningene i tillegg til et korrelogram med begge års estimerte effektiviteter.

Ved å se på regresjonsresultatene ser vi at modellene håndterer variablene på samme måte, dog med varierende signifikans. I SFA-modellen ser vi at andre driftskostnader vektet signifikant mindre enn lønn når det kommer til estimering av DRG-produksjon. I DEA-modellene finnes det flere måter å produsere DRG-poeng på, i og med at vi ikke har pålagt variablene noen restriksjoner.

Selv om Banker-testene (vedlegg 3) forteller at det er en signifikant forskjell mellom effektivitetssettene, er det viktig å huske på at VRS-modellen omhyller dataen tettere enn CRS.

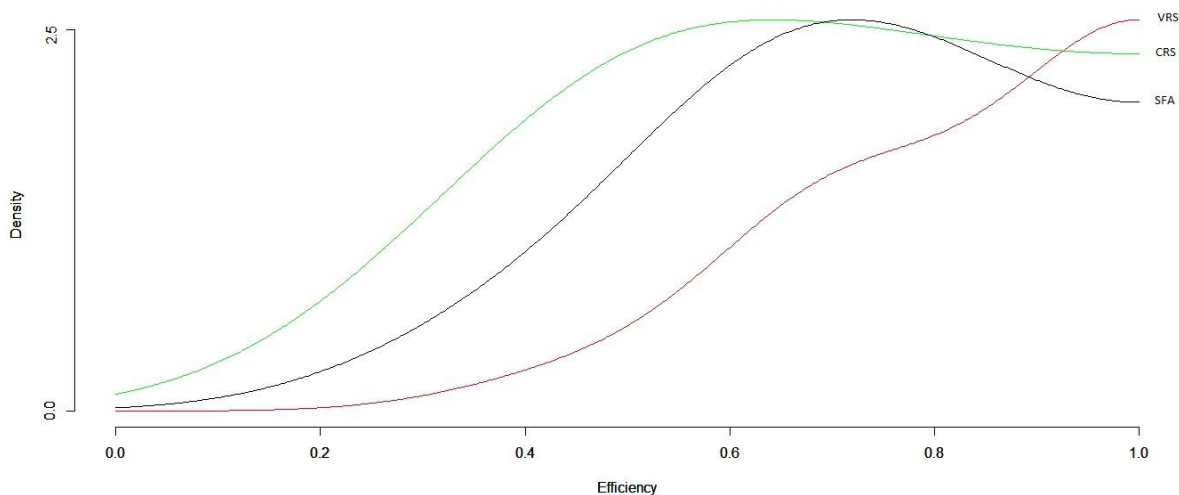
“Adopting the VRS model can, however, lead to a pitfall itself. It is clear that the VRS model will always envelop the data more closely than the CRS model, irrespective of whether variable returns to scale exist. It follows therefore that, if the VRS model is used, where there are no inherent scale effects, small and large units will tend to be over-rated in the efficiency

assessment.” (Dyson et al., 2001, p. 248)

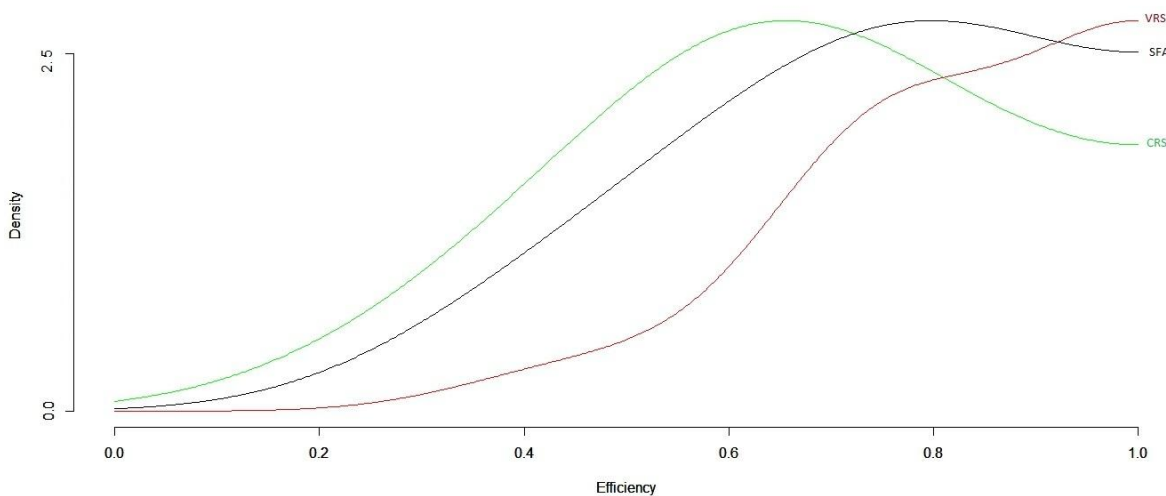


Figur 6: Salterdiagram med totalkostnad og effektiviteter

Som vi ser av figur 6 er det fortrinnsvis de små DMUene som øker mest i effektivitet ved VRS-teknologi (to ”små” blir effektive). Dette gjelder for 2011-dataene. Et element som er til støtte for at det ikke finnes skalafordeler/ulempene i produksjonsprosessen, er at spredningen mellom store og små sengeposter er god, slik at det ikke virker å være større sannsynlighet for at en stor sengepost er effektiv framfor en liten. Måten VRS omhyller estimatene på gjør at en liten og stor DMU automatisk blir effektive som et resultat av modellens utforming. Fronten trenger et naturlig start- og endepunkt, mens CRS kun tegner fronten basert på den/de ”beste” observasjonen(e). SFA-estimatene utelukker ikke CRS (vedlegg 5). Dette kan vi også se av figur 6, der relativ ulik størrelse ikke virker å ha noen sammenheng med forskjell i effektivitet. Skalaspørsmålet er ennå ikke avklart, da beregningene rundt hvilke skalaforutsetninger som regjerer ikke konkluderer i den ene eller andre retningen. Ved estimering av (2) forteller $\Sigma\lambda$ at alle ineffektive sengeposter er i et område med økende utbytte, men vi kan ikke fastslå at dette gjenspeiler virkeligheten. På sykehusnivå fant blant annet (Bjørn et al., 2003) ut at det ikke eksisterte skalaforskjeller. For SFA kan vi ikke avvise at det eksisterer konstant skalautbytte.



Figur 7: Tetthetsfunksjonen til 2010-estimatene



Figur 8: Tetthetsfunksjonen til 2011-estimatene

Tetthetsfunksjonene til effektivitetsestimaterne rapporterer hvor det er størst sannsynlighet for å finne effektivitetsestimaterne. Figur 7 og 8 er konstruert ved hjelp av R.. Toppunktet til CRS og SFA ligger ved gjennomsnittet, mens gjennomsnittet til VRS ligger i den siste bølgedalen før linja stiger igjen. SFA ligger som en mellomting mellom CRS og SFA, med høy korrelasjon med både CRS og VRS-estimatene. Linjene gir en god indikasjon på hvilket effektivitetsnivå man finner flest DMUer. Det er dette som gir den siste toppen for VRS-estimatene, siden det relativt sett er flest effektive observasjoner.

SFA-modellen er inkludert som en robusthetstest for DEA-estimatene. Vi har ikke kunnet avkrefte variabelt skalautbytte i datasettet, og heller ikke avkrefte eller bekrefte variabelt skalautbytte ved hjelp av SFA. Men som vi ser av figurene 6-8 legger SFA-estimatene seg som et snitt av den estimerte CRS- og VRS-effektiviteten. Det kan dermed tenkes at den ”ekte” effektiviteten finnes i et intervall mellom CRS og VRS. Sengepostenes prestasjoner vil påvirkes av de andre funksjonene ved sykehuset, og den stokastiske delen av ineffektiviteten vi finner i SFA-modellen kan være et mål på denne sammenhengen.

Tabell 31: Korrelasjon mellom effektivitetsestimaterne

	<i>crs2010</i>	<i>vrs2010</i>	<i>sfa2010</i>	<i>crs2011</i>	<i>vrs2011</i>	<i>sfa2011</i>
<i>crs2010</i>	1,00					
<i>vrs2010</i>	0,80	1,00				
<i>sfa2010</i>	0,89	0,75	1,00			
<i>crs2011</i>	0,95	0,71	0,91	1,00		
<i>vrs2011</i>	0,77	0,96	0,76	0,74	1,00	
<i>sfa2011</i>	0,77	0,63	0,95	0,84	0,70	1,00

Vi har ikke sett på effektivitetsendringer i perioden. Det betyr at selv om en enhet har lavere effektivitetsscore i 2011 enn i 2010, trenger ikke årsaken til dette å være effektivitetsnedgang, bare at man har endret seg relativt til fronten. Vi har også utført bankertestene (8) og (9) fra kapittel 4.7 for å se om det er noe forskjell mellom estimatene i 2010 og 2011. For alle tilfellene finner vi kun minimale, insignifikante endringer i effektivitet mellom årene.

6.8.1 Sammenligning mot tradisjonelle mål

Tradisjonelle prestasjonsmål som *resultat/budsjettavvik*, *kostnad per utskrevne pasient(KPUP)* og *kostnad per sengedøgn(KPSD)* er testet opp mot effektivitetene for å se om det eksisterer en sammenheng mellom effektivitet og tradisjonelle prestasjonsmål. Vi har sett på korrelasjonen mellom CRS-estimatene for 2011 og de tradisjonelle målene. Merk at dummyvariabelen BUD er lik én for positive avvik. Sammenhengen er presentert i tabell 32.

Tabell 32: Korrelasjon mellom CRS 2011 og tradisjonelle prestasjonsmål

	<i>CRS</i>	<i>BUD2011</i>	<i>KPUP11</i>	<i>KPSD11</i>
<i>CRS</i>	1,00			
<i>BUD2011</i>	0,24	1,00		
<i>KPUP11</i>	-0,55	-0,28	1,00	
<i>KPSD11</i>	-0,31	-0,32	0,20	1,00

Ved hjelp av regresjon fant vi ingen sammenheng mellom resultat/budsjettavvik, og observerer flere sengeposter med positive budsjettavvik som ender opp med lav effektivitet. Dette ser vi også av korrelasjonen som er lav. Det vil være en naturlig sammenheng mellom kostnad per DRG og effektivitetene, da begge inngår i modellen, derfor måler vi ikke denne sammenhengen. Vi observerer en negativ korrelasjon mellom effektivitet og kostnad per sengedøgn, men sammenstiltes de to rangeringene observerer vi at de billigste sengepostene ikke er de mest effektive i analysen. Korrelasjonen er heller ikke veldig høy. Den høyeste korrelasjonen finner vi mellom *kostnad per sengedøgn* og effektivitet. Det må poengteres at rangeringen av sengepostene ikke er helt lik ved å bruke disse metodene, men vil allikevel kunne brukes som prestasjonsmål på sengepostnivå. Høye kostnader per pasient eller sengedøgn trenger ikke å bety at sengeposten er ineffektiv eller presterer dårligere enn en billig sengepost, da komplekse pasienter naturlig vil kreve mer.

7 Oppsummering og konklusjon

7.1 Oppsummering

Universitetssykehuset i Nord-Norge ønsket å analysere ressursutnyttelsen ved sine somatiske sengeposter. Ved å benytte oss av metodene *Data Envelopment Analysis* og *stokastisk frontanalyse* har vi analysert den individuelle effektiviteten for årene 2010 og 2011. I forrige kapittel presenterte vi resultatene av selve effektivitetsanalysen. I begge årene lå den gjennomsnittlige effektiviteten ved UNNs sengeposter i intervallet 64-81 prosent for henholdsvis CRS og VRS. Ved bruk av stokastisk frontanalyse var den gjennomsnittlige effektiviteten omtrent 70 prosent i begge årene. Outlieranalysen vi foretok i 5.6 påviste ingen vesentlige outliers, noe som tilsier at analysen ikke påvirkes av ikke-tilhørende DMUer.

Vi har funnet to stabile referanseenheter som skiller seg ut i begge årene. *Hjertemedisinsk-* og *hjerte/lunge/karkirurgisk* sengepost ser ut til å forme effektivitetsfronten når man baserer seg på konstant skalautbytte(DEA), mens i SFA-estimatene er *hjerte/lunge/karkirurgisk* alene den mest effektive sengeposten. Om lag 51 prosent av ineffektiviteten i DEA-modellen kan forklares av ren teknisk ineffektivitet, mens den resterende ineffektiviteten på 49 prosent skyldes feil skala. I SFA-modellene forklares om lag 25 prosent av ineffektiviteten av rene tekniske årsaker, mens den resterende ineffektiviteten er forklart av tilfeldighet og størrelsen på sengepostene i form av senger.

DEA-estimatene avdekker store muligheter til kostnadsforbedring. Det er vanskelig å si noe om karakteristikken ved de mest effektive og minst effektive sengepostene. De to hjerterelaterte sengepostene skiller seg ut i positiv retning, mens de tre fødepostene er spredt utover effektivitetsskalaen. Blant de mest effektive sengepostene finner vi både kirurgiske og medisinske sengeposter, og innenfor denne gruppen finnes det både små og store sengeposter målt i form av kostnader og kapasitet. Blant de 10 sengepostene med lavest effektivitet finner vi kun én kirurgisk sengepost. Holmeide et al. (2008) poengterer at det er en underdekning i refusjon for de yngste og eldste pasientene ved undersøkelsesobjektene. Eksisterer denne sammenhengen ved UNN, vil *Gastro/nyre/geriatrisk* sengepost følgelig komme dårligere ut på grunn av en høy andel eldre pasienter. Dette kan også være årsaken til at sengeposten *Store barn* er å finne i det nedre effektivitetssjiktet. Underfinansiering som følge av en mangel i systemet kan skape inkommensurable målsetninger mellom klinikkledelse og styringselementet i finansieringssystemet. De beregnede optimale innsatsfaktorene vil bli noe

upresis å bruke for å finne den reelle optimale ressursmiksen for hver enkelt sengepost. Selv om aktiviteten ved sengepostene er sammenlignbar vil det på prosedyrenivå være ulikheter i ressurs sammensetningen mellom pleie og andre driftskostnader.

Geografi virker ikke å være en effektivitetsdrivende faktor. Lokalsykehusene i Harstad og Narvik har ikke samme type kompetanse som hovedsykehuset, slik at de mest komplekse og alvorlige pasientene vil bli behandlet ved sykehuset i Tromsø. Gjennomsnittskompleksiteten til sengepostene i Harstad og Narvik er 0,94 mens sengepostene i Tromsø har et gjennomsnitt på 1,48.

Ved å se på budsjettavvikene til de ulike sengepostene finner vi positive avvik både for effektive og ineffektive sengeposter. Budsjettene er utarbeidet på bakgrunn av historiske tall og prestasjoner slik at planlagt ineffektivitet i form av for ”høye” budsjetter vil kunne forekomme. Et bedre tradisjonelt prestasjonsmål vil for UNN sin del være kostnad per utskrevet pasient, som korrelerer bra med våre effektivitetsestimater. Ineffektivitet kan også komme fra svingninger i pasientgrunnet. Hvis det faktiske pasientantallet er lavere enn forventet kan det være vanskelig å omfordele ressursene (faste på kort sikt). Tallene vi har på kapasitetsutnyttelsen forteller oss derimot at denne er tett på 100 prosent for de fleste sengepostene. Det må presiseres at det høye belegget også vil stamme fra kødannelser ved sengepostenes støttefunksjoner og/eller operasjonsstuer. Dermed kan det tenkes at UNN helhetlig bør utnytte kapasiteten på en bedre måte.

Som presisert i drøftelsen av sammenhengene mellom effektiviteten og kontrollvariablene er det ulikt signifikans på variablene for de ulike modellene. Vi presenter følgende svar på hypotesene

- Vi avslår nullhypotesen og konkluderer at beleggsgard har hatt en negativ påvirkning på effektiviteten.
- Vi beholder nullhypotesen da vi ikke har tilstrekkelige bevis for at det eksisterer effektivitetsforskjell mellom kirurgiske og medisinske sengeposter
- Vi forkaster nullhypotesen og bekrefter at andel utskrevne pasienter påvirket effektiviteten positivt.
- Vi forkaster nullhypotesen og slår fast at det var en positiv sammenheng mellom effektivitet og kompleksitet.

7.2 Konklusjon

Snitteffektiviteten for de to årene ligger i intervallet 64-81 prosent, avhengig av hvilken metode man benytter. Vi har ingen sammenlignbare undersøkelser på sengepostnivå, men merker oss at snitteffektiviteten i tilsvarende interne effektivitetsanalyser for leger rapporterer en god del ineffektivitet. Analyser på sykehusnivå gir oftere en lavere ineffektivitet. Selv om DRG-systemet skal homogenisere aktiviteter, vil det være usikkert om det oppnår fullstendig homogenisering. Når internanalyser gir store svingninger, mens de overordnede sykehusanalysene reduserer intervallene for ineffektivitet, kan det virke som ineffektiviteten i helsevesenet først og fremst kommer frem mellom ulike diagnoser. Som presisert i 2.1 må man være forsiktig i tolkningen av våre resultater, men analysene antyder store rom for kostnadsforbedringer. Vi må stille oss selv de samme spørsmålene andre har gjort før oss: kan man sammenligne ulike avdelinger og prosedyrer – lar det seg gjøre på en robust måte? Fordelingsnøkklene i ISF-systemet er basert på data av virkelig ressursbruk i avdelingene knyttet til hver enkelt DRG-kode. Dette mener vi forsvarer bruken av DRG-poeng som aktivitetsmål.

Vi har ikke funnet spesielle kjennetegn ved de mest effektive sengepostene, bortsett fra at disse er de med høyest DRG-produksjon og virker å ha fagfeltet *hjerte* som fellesnevner.

Den høye korrelasjonen mellom DEA- og SFA-estimatene er betryggende. Begge metodene virker å håndtere inputene og aktivitetsmålet på en lik måte, noe som gir analysen ekstra styrke. Når det kommer til forklaringsevnen til de kontekstuelle variablene virker SFA og VRS å være mest sammenfallende. UNN kan ikke selv påvirke alle de kontekstuelle variablene som inngår i hypotesene vi har testet. Som presisert i drøftelsen av sammenhengene mellom effektiviteten og kontrollvariablene, er det ulik signifikans på variablene for de ulike modellene. Den negative sammenhengen mellom beleggsgrad og effektivitet antyder at kapasiteten kan utnyttes på en bedre måte. Dette kan gjøres ved å redusere antall utskrivningsklare pasienter for å få inn nye, behandlingsklare pasienter. Omløpshastigheten på pasientene bør økes, slik at liggetiden nærmer seg den normerte liggetiden angitt av DRG-systemet.

7.3 Veien videre

Denne analysen ser kun på sengepostdelen av et DRG-opphold. At en sengepost er mer effektiv enn en annen betyr ikke at hele aspektet ved et DRG-opphold er effektivt.

Pleieandelen og grunnandelen står i gjennomsnitt for 60 % av et DRG-poeng. For å ta analysen et steg videre kan man analysere effektivitet med de samme innsatsmålene på klinikknivå. Da får man med alle aspektene ved et DRG-opphold som operasjon, laboratorium og transplantater osv., som står for en betydelig andel av enkelte DRG-opphold. Hvis et DRG-opphold og refusjonen virkelig reflekterer sykehusets kostnader knyttet til dette oppholdt kan DRG-systemet brukes som en klassisk kostnadsstandard, der man i tillegg kan gå inn i å se på hvor kostnadsavvikene faktisk stammer fra.

Et annet aspekt ved våre funn det vil være interessant å undersøke videre er om DRG-systemet systematisk overkompenserer hjerterelaterte diagnoser, eller om de hjerterelaterte sengepostene kun opererer bedre enn de andre. Fra bevilgende myndighets side vil det også være interessant å undersøke om vektgruppene i DRG-poengene kan bedre fordeles på sykehusenes funksjoner. Dette er allerede gjort for blant annet laboratorium og operasjon. Grunnandelen står for en stor andel av alle DRG-poengene slik at en bedre fordeling av grunnandelen til mer spesifikke kategorier, vil kanskje kunne gi bedre informasjon om ressursbruken knyttet til de spesifikke sykehusfunksjonene.

8 Litteraturliste

- Adler, N., Friedman, L., & Sinuany-Stern, Z. (2002). Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research*, 140(2), 249-265.
- Aigner, D., Lovell, C., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1), 21-37.
- Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 1261-1264.
- Andrews, D. F., & Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society. Series B (Methodological)*, 85-93.
- Banker, R. D. (1993). Maximum-Likelihood, Consistency and Data Envelopment Analysis - a Statistical Foundation. *Management Science*, 39(10), 1265-1273.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 1078-1092.
- Banker, R. D., & Natarajan, R. (2008). Evaluating contextual variables affecting productivity using data envelopment analysis. *Operations Research*, 56(1), 48-58.
- Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis*, 3(1), 153-169.
- Biørn, E., Hagen, T. P., Iversen, T., & Magnussen, J. (2003). The effect of activity-based financing on hospital efficiency: a panel data analysis of DEA efficiency scores 1992–2000. *Health Care Management Science*, 6(4), 271-283.
- Bogetoft, P., & Otto, L. (2010). *Benchmarking with DEA, SFA, and R* (Vol. 157): Springer Verlag.
- Burns, L. R., Chilingerian, J. A., & Wholey, D. R. (1994). The effect of physician practice organization on efficient utilization of hospital resources. *Health Services Research*, 29(5), 583.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429-444.
- Chilingerian, J. A. (1995). Evaluating Physician Efficiency in Hospitals - a Multivariate-Analysis of Best Practices. *European Journal of Operational Research*, 80(3), 548-574.
- Chilingerian, J. A., & Sherman, H. D. (1996). Benchmarking physician practice patterns with DEA: A multi-stage approach for cost containment. *Annals of Operations Research*, 67, 83-116.
- Chilingerian, J. A., & Sherman, H. D. (1997). DEA and primary care physician report cards: Deriving preferred practice cones from managed care service concepts and operating strategies. *Annals of Operations Research*, 73, 35-66.
- Chilingerian, J. A., & Sherman, H. D. (2011). *Handbook of Data Envelopment Analysis, Health Care Applications - Chapter 16- From Hospitals to Physicians, from Productive Efficiency to Quality Frontiers* Boston, MA : Springer Science+Business Media, LLC.
- Chirikos, T. N., & Sear, A. M. (2000). Measuring hospital efficiency: A comparison of two approaches. *Health Services Research*, 34(6), 1389-1408.
- Coelli, T., Prasada Rao, D. S., O'Donnell, C. J., & Battese, G. (2005). *An introduction to efficiency and productivity analysis*: Springer Verlag.

- Cooper, W. W., Seiford, L. M., & Tone, K. (2001). Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. *Journal-operational Research Society*, 52(12), 1408-1409.
- Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2), 245-259.
- Edvardsen, D. (2004). *Four Essays on the Measurement of Productive Efficiency: Climbing the efficiency stepladder: Robustness of efficiency scores in DEA*. University of Gothenburg.
- Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society Series a-General*, 120(3), 253-290.
- Helsedirektoratet. (2010). Innsatsstyrt finansiering 2011.
- Helsedirektoratet. (2011a). DRG historie i Norge fra slutten av 70-tallet Retrieved 26.04, 2012, from <http://www.helsedirektoratet.no/finansiering/drg/drg-struktur/Sider/drg-historie-i-norge-fra-slutten-av-70-tallet.aspx>
- Helsedirektoratet. (2011b). Formål med innsatsstyrt finansiering (ISF), from <http://www.helsedirektoratet.no/finansiering/isf/om-isf-ordningen/formal/Sider/default.aspx>
- Helsedirektoratet. (2011c). Kostnadsgruppeandel 2012. www.helsedirektoratet.no.
- Helsedirektoratet. (2011d). Kostnadsvekstarbeidet 2011-2012 - DRG-systemet www.helsedirektoratet.no.
- Helsedirektoratet. (2011e). Samdata spesialisthelsetjenesten 2010.
- Helsedirektoratet. (2011f). Utvikling av DRG-systemet Retrieved 26.04, 2012, from <http://www.helsedirektoratet.no/finansiering/drg/systemutvikling/Sider/default.aspx>
- Helsedirektoratet. (2012). Nøkkeltall for helsesektoren - rapport 2011. www.helsedirektoratet.no.
- HelseNord. (2006). Strategi for Helse Nord RHF Retrieved 10.05, 2012, from <http://www.helse-nord.no/maal-og-strategier/category4520.html>
- HelseNord. (2011). Budsjett 2012- Bærekraftanalyse, from <http://www.helse-nord.no/getfile.php/RHF/Styret/Styredokumenter/2011/111021%20B%C3%A6rekraftanalyse.pdf>
- HelseNord. (2012). Årsregnskap 2011
- Hill, R. C., Griffiths, W. E., & Lim, G. C. (2008). *Principles of econometrics* (Vol. 5): Wiley.
- Hollingsworth, B., Dawson, P. J., & Maniadakis, N. (1999). Efficiency measurement of health care: a review of non - parametric methods and applications. *Health Care Management Science*, 2(3), 161-172.
- Holmeide, A. M., Seierstad, T. Ø., & Eimot, M. (2008). Sykehuskostnader og alder. *Oslo: Analysesenteret*.
- Kittelsen, S. A. C., Magnussen, J., Sarheim Anthun, K., Häkkinen, U., Linna, M., Medin, E., . . . Rehnberg, C. (2009). Hospital productivity and the Norwegian ownership reform—A Nordic comparative study. *HERO On line Working Paper Series*.
- Kumbhakar, S. C., & Lovell, C. A. K. (2003). *Stochastic frontier analysis*: Cambridge Univ Pr.
- Lyrودي, K., Glaveli, N., Koulakiotis, A., & Angelidis, D. (2006). The productive performance of public hospital clinics in Greece: a case study. *Health Services Management Research*, 19(2), 67-72.
- Magnussen, J., & Solstad, K. (1994). Case-Based Hospital Financing - the Case of Norway. *Health Policy*, 28(1), 23-36.
- Medin, E., Anthun, K. S., Hakkinen, U., Kittelsen, S. A. C., Linna, M., Magnussen, J., . . . Rehnberg, C. (2011). Cost efficiency of university hospitals in the Nordic countries: a

- cross-country analysis. *European Journal of Health Economics*, 12(6), 509-519. doi: DOI 10.1007/s10198-010-0263-1
- Nyland, K., & Bjørnenak, T. (2000). Kompleksitetens økonomi; dimensjoner, effekter og finansiering. *Universitetet i Oslo, helseøkonomisk forskningsprogram. Skriftserie 2000:2*.
- O'Neill, L., Rauner, M., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences*, 42(3), 158-189.
- Ozcan, Y. (2008). Health Care Benchmarking and Performance Evaluation: An Assessment Using Data Envelopment Healthcare Benchmarking and Performance Evaluation: Analysis (DEA)
- Ozcan, Y. A., Luke, R. D., & Haksever, C. (1992). Ownership and organizational performance: a comparison of technical efficiency across hospital types. *Medical Care*, 781-794.
- Regjeringen.no. (1997). Tabell 5.4 Kliniske avdelinger og seksjoner ved ulike sykehustyper Retrieved 27.04, 2012, from <http://www.regjeringen.no/nb/dep/hod/dok/nouer/1997/nou-1997-2/6.html?id=140695>
- Roberts, R. R., Frutos, P. W., Ciavarella, G. G., Gussow, L. M., Mensah, E. K., Kampe, L. M., . . . Rydman, R. J. (1999). Distribution of variable vs fixed costs of hospital care. *Jama-Journal of the American Medical Association*, 281(7), 644-649.
- Sherman, H. D. (1984). Hospital efficiency measurement and evaluation: empirical test of a new technique. *Medical Care*, 922-938.
- Sinuany-Stern, Z., Mehrez, A., & Barboy, A. (1994). Academic departments efficiency via DEA. *Computers & Operations Research*, 21(5), 543-556.
- Spekter. (2010). *Kunnskap om det grunnleggende – og trygghet for helse i hver krone*. Unpublished.
- StoreNorskeLeksikon. (2012). Helsevesenets historie Retrieved 31.01.2012, from http://snl.no/Helsevesenets_historie
- Wilson, P. W. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business & Economic Statistics*, 319-323.
- Zhu, J. (2004). Quantitative models for performance evaluation and benchmarking: Data envelopment analysis with spreadsheets and DEA excel solver.
- Zuckerman, S., Hadley, J., & Iezzoni, L. (1994). Measuring Hospital Efficiency with Frontier Cost-Functions. *Journal of Health Economics*, 13(3), 255-280.

Vedlegg

Vedlegg 1 Eksempeldata

dmu	x	y
1	10	5
2	12,5	6
3	9,5	4,5
4	9	4
5	11	6
6	13	6
7	12	5,5
8	13,5	6,5
9	14	11
10	18	12
11	19	14
12	5	1
13	5,5	2
14	11,5	6,2
15	8	3,5

Vedlegg 2 Spesifikasjonstest 1

Input til R med resultater:

$H_0: F_1 = F_2$ mot H_A F_1 ikke lik F_2

$F_1 = y = \text{justerte drgpoeng}$, $x_2 = \text{total kostnader}$

$F_2 = y = \text{justerte drgpoeng}$, $x = \text{lønnskostnader, andre driftskostnader, sengedøgn}$

Til R:

```
F1=eff (dea (x2, y, RTS="crs/vrs"))
```

```
F2=eff (dea (x, y, RTS="crs/vrs"))
```

2010:

<p>CRS:</p> <p>Ekspensialfordelt ineffektivitet</p> <pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre> <p>TEX</p> <p>[1] 1.26308</p> <pre>qf(.95, 2*length(F1), 2*length(F2))</pre> <p>[1] 1.670971</p> <p>Halvnormalfordelt ineffektivitet</p> <pre>THN <- sum((F1-1)*(F1- 1))/length(F1) / (sum((F2- 1)*(F2-1))/length(F2))</pre> <p>THN</p> <p>[1] 1.44</p> <pre>qf(.95, length(F1), length(F2))</pre> <p>[1] 2.084189</p>	<p>VRS:</p> <p>Ekspensialfordelt ineffektivitet</p> <pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre> <p>TEX</p> <p>[1] 1.45</p> <pre>qf(.95, 2*length(F1), 2*length(F2))</pre> <p>[1] 1.670971</p> <p>Halvnormalfordelt ineffektivitet</p> <pre>THN <- sum((F1-1)*(F1- 1))/length(F1) / (sum((F2- 1)*(F2-1))/length(F2))</pre> <p>THN</p> <p>[1] 1.75</p> <pre>qf(.95, length(F1), length(F2))</pre> <p>[1] 2.084189</p>
---	--

2011:

CRS:	VRS:
<p>Ekspensialfordelt ineffektivitet</p> <pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre> <p>TEX</p> <p>[1] 1.26</p> <pre>qf(.95, 2*length(F1), 2*length(F2))</pre> <p>[1] 1.670971</p>	<p>Ekspensialfordelt ineffektivitet</p> <pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre> <p>TEX</p> <p>[1] 1.48</p> <pre>qf(.95, 2*length(F1), 2*length(F2))</pre> <p>[1] 1.670971</p>
<p>Halvnormalfordelt ineffektivitet</p> <pre>THN <- sum((F1-1)*(F1- 1))/length(F1) / (sum((F2- 1)*(F2-1))/length(F2))</pre> <p>THN</p> <p>[1] 1.50</p> <pre>qf(.95, length(F1), length(F2))</pre> <p>[1] 2.084189</p>	<p>Halvnormalfordelt ineffektivitet</p> <pre>THN <- sum((F1-1)*(F1- 1))/length(F1) / (sum((F2- 1)*(F2-1))/length(F2))</pre> <p>THN</p> <p>[1] 1.83</p> <pre>qf(.95, length(F1), length(F2))</pre> <p>[1] 2.084189</p>

Vedlegg 3 Spesifikasjonstest 2

F1= y=justerte drgpoeng, x'= lønnskostnader, andre driftskostnader, sengedøgn

F2= y'=(justerte drgpoeng, antall utskrevne pasienter) x'=(lønnskostnader, andre driftskostnader, sengedøgn)

```
F1=eff (dea (x, y, RTS="crs"))
```

```
F2=eff (dea (x, y1, RTS="crs"))
```

2010:

CRS:	VRS:
<pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre>	<pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre>
TEX	TEX
[1] 1.25	[1] 1.36
<pre>qf(.95, 2*length(F1), 2*length(F2))</pre>	<pre>qf(.95, 2*length(F1), 2*length(F2))</pre>
[1] 1.670971	[1] 1.670971
<pre>THN <- sum((F1-1)*(F1- 1))/length(F1) / (sum((F2- 1)*(F2-1))/length(F2))</pre>	<pre>THN <- sum((F1-1)*(F1- 1))/length(F1) / (sum((F2- 1)*(F2-1))/length(F2))</pre>
THN	THN
[1] 1.45	[1] 1.39
<pre>qf(.95, length(F1), length(F2))</pre>	<pre>qf(.95, length(F1), length(F2))</pre>
[1] 2.084189	[1] 2.084189

2011:

CRS:	VRS:
<pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre>	<pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre>
TEX	TEX
[1] 1.29	[1] 1.27
<pre>qf(.95, 2*length(F1), 2*length(F2))</pre>	<pre>qf(.95, 2*length(F1), 2*length(F2))</pre>
[1] 1.670971	[1] 1.670971
<pre>THN <- sum((F1-1)*(F1-</pre>	<pre>THN <- sum((F1-1)*(F1-</pre>

<pre>1))/length(F1) / (sum((F2-1)*(F2-1))/length(F2))</pre>	<pre>1))/length(F1) / (sum((F2-1)*(F2-1))/length(F2))</pre>
THN	THN
[1] 1.52	[1] 1.43
qf(.95, length(F1), length(F2))	qf(.95, length(F1), length(F2))
[1] 2.084189	[1] 2.084189

Vedlegg 4 Skalatest

Skalatest:

```
F1=eff(dea(x,y,RTS="crs"))
```

```
F2=eff(dea(x,y,RTS="vrs"))
```

<p>2010:</p> <pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre>	<p>2011:</p> <pre>TEX <- (sum(F1-1)/length(F1)) / (sum(F2-1)/length(F2))</pre>
TEX	TEX
[1] 1.77	[1] 1.9
qf(.95, 2*length(F1), 2*length(F2))	qf(.95, 2*length(F1), 2*length(F2))
[1] 1.670971	[1] 1.670971
THN <- sum((F1-1)*(F1-1))/length(F1) / (sum((F2-1)*(F2-1))/length(F2))	THN <- sum((F1-1)*(F1-1))/length(F1) / (sum((F2-1)*(F2-1))/length(F2))
THN	THN
[1] 2.5	[1] 2.75
qf(.95, length(F1), length(F2))	qf(.95, length(F1), length(F2))
[1] 2.084189	[1] 2.084189

Vedlegg 5 Datasett 2010 og 2011

Datasett 2010				
DMU	LK	ADK	Senger	DRG
1	28871818	4324029	10257	1630
2	23292343	1650121	9965	1459
3	25103293	6917122	9928	1264
4	25235776	3960532	9125	1187
5	27616710	1728029	9855	1945
6	27344209	5398264	8724	1475
7	25835835	2679769	7957	2270
8	18497207	3231759	7483	1153
9	23354197	2005034	6132	548
10	15967746	980601	7300	868
11	17843380	1430900	6899	720
12	14636344	1621476	7300	1097
13	22822000	3117853	6643	1069
14	35487357	2029450	10585	1612
15	24008268	7013802	11680	1716
16	12502832	552791	3942	518
17	13926082	1591293	3650	430
18	15508881	5936899	3650	503
19	9424657	718793	2555	283
20	35763770	2997171	9709	855
21	26431069	3534009	6132	847

Datasett 2011				
DMU	LK	ADK	Senger	DRG
1	30103212	5139403	10220	1677
2	25066162	1829304	9891,5	1603
3	25995045	8138008	10074	1309
4	24932548	3658893	9526,5	1362
5	28916279	1565601	9855	1812
6	27780425	2952314	8796,5	1529
7	25835835	2425675	7884	2114
8	18704153	3732336	7373	1097
9	23003255	4121944	6132	519
10	17509750	1204483	7373	848
11	18924401	1634891	6971,5	725
12	16403000	1816445	7300	1049
13	23898309	3131002	6497	1120
14	38145354	2208365	10439	1508
15	26952720	9685053	11607	1669
16	13047956	607530	3905,5	502

17	14782879	2277744	3650	417
18	16151416	6724206	3650	510
19	9591960	643851	2555	265
20	38030733	4185015	10877	1020
21	27852857	4734415	5730,5	822

Vedlegg 6 Heteroskedastisitet, SFA og Wald-test

Bevis for heteroskedastisitet:

HETEROSKEDASTICITY TESTS			
	2010		
	CHI-SQUARE TEST STATISTIC	D.F.	P-VALUE
E**2 ON YHAT:	0.387	1	0.53410
E**2 ON YHAT**2:	0.345	1	0.55671
E**2 ON X X**2 (WHITE) TEST:			
KOENKER(R2):	4.072	6	0.66691
B-P-G (SSR) :	7.383	6	0.28684
HETEROSKEDASTICITY TESTS			
	2011		
	CHI-SQUARE TEST STATISTIC	D.F.	P-VALUE
E**2 ON YHAT:	0.453	1	0.50087
E**2 ON YHAT**2:	0.411	1	0.52145
E**2 ON X X**2 (WHITE) TEST:			
KOENKER(R2):	3.410	6	0.75596
B-P-G (SSR) :	5.183	6	0.52059

STATA-koder for SFA-estimering:

```
frontier DRG LØNN Andre driftskost, vhet(senger)
```

```
predict te
```

```
predict efficiency, te
```

```
list efficiency
```

WALD-TEST FOR SKALA I SFA:

2010:

```
test _b[lønn]+_b[driftskostnader]=1
```

```
chi2( 1) = 0.03
```

```
Prob > chi2 = 0.85
```

2011:

```
test _b[lønn]+_b[driftskostnader]=1
```

```
chi2( 1) = 0.34
```

```
Prob > chi2 = 0.56
```