

Teori- og forskningsevaluering av forskningsarbeidet:

**Evans, J. og Green, R. (2003),
"Why Did British Electricity Prices Fall after 1998?",
Cambridge Working Papers in Economics CWPE 0326**

Av Kristin Linnerud

<p>TITTEL Teori- og forskningsevaluering av forskningsarbeidet:</p> <p>Evans, J. og Green, R. (2003), "Why Did British Electricity Prices Fall after 1998?", Cambridge Working Papers in Economics CWPE 0326.</p>	<p>NOTATNR. 8/04</p>	<p>DATO 01.09.2004</p>
<p>PROSJEKTTITTEL Teori- og forskningsevaluering av utvalgt forskningsarbeid.</p>	<p>TILGJENGE åpen</p>	<p>TAL SIDER 22</p>
<p>FORFATTAR Kristin Linnerud</p>	<p>PROSJEKTTLEIAR/-ANSVARLEG Kristin Linnerud</p>	
<p>OPPDRAKSGJEVAR Obligatorisk innlevering ved doktorgradskurs i vitenskapsteori ved NHH.</p>	<p>EMNEORD Vitenskapsteori, Karl Popper, Falsifikasjon</p>	
<p>SAMANDRAG</p> <p>I notatet gir jeg en kort presentasjon av Evans og Green sin artikkel om det britiske kraftmarkedet. Deretter gjennomfører jeg en teori- og forskningsevaluering av deres arbeid med utgangspunkt i Karl Popper sitt kriterium om falsifiserbarhet. På bakgrunn av evalueringen presenterer jeg til slutt noen forslag til forbedringer i det videre forskningsarbeid med problemstillingen.</p>		
<p>SUMMARY</p> <p>I start with a brief presentation of the main issues in Evans and Green's article about the british power market. I continue with an evaluation of their research based on Karl Popper's criteria; falsification. At the end I present some suggestions to the organization of future research in this field.</p>		
<p>PRIS</p>	<p>ISSN 0806- 1696</p>	<p>ANSVARLEG SIGNATUR</p>

Forord

Dette notatet er et godkjent arbeidskrav ved doktorgradskurset ”Teori- og forskningsevaluering” ved Norges Handelshøyskole sommeren 2004. Kurset inngår i mitt doktorgradsarbeid med tittelen ”Innenlands bruk av naturgass – reguleringsregimer for infrastruktur under oppbygging”.

Målgruppa for notatet kan være studenter og kollegaer som arbeider med vitenskapsteoretiske problemstillinger, f.eks. ved utforming av forskningsdesign i prosjektoppgaver eller forskningsarbeid.

Innhold

1	Innledning.....	1
2	Presentasjon av artikkelen	2
2.1	Sammendrag	2
2.2	Bakgrunn	2
2.3	Resultater fra tidligere forskningsarbeid	3
2.4	Modell og data	4
2.5	Resultater	5
2.6	Konklusjon.....	6
3	Presentasjon av og begrunnelse for valg av vurderingskriterier.....	7
3.1	Mitt vitenskapssyn	7
3.2	Artikkelen og de fire forskningsdomenene	8
3.3	Mitt valg av evalueringskriterium	9
3.4	Falsifiserbarhet	10
4	Evaluering.....	11
4.1	Teoriens utsagn.....	11
4.2	Falsifiserbarhet og tautologiske utsagn	12
4.3	Falsifiserbarhet og verifiserbarhet	12
4.4	Falsifiserbarhet og utsagnsinhold	14
4.5	Falsifiserbarhet og utsagnets teoretiske og empiriske fundament	15
4.6	Falsifiserbarhet og utforming av empiriske tester	17
4.7	Falsifiserbarhet og vilje til å erkjenne falsifikasjon.....	18
4.8	Noen refleksjoner rundt bruken av falsifiserbarhetskriteriet.....	20
5	Konklusjon.....	22

1 Innledning

Joanne Evans og Richard Green har levert et innsiktsfullt og nyttig bidrag til debatten om hva som forårsaket prisfallet i engrosmarkedet for kraft i England og Wales i forrige tiår.

På bakgrunn av sine studier trekker de en foreløpig konklusjon om at innføringen av et nytt auksjonsprinsipp i engrosmarkedet for kraft synes å ha vært viktigste drivkraften bak prisfallet.

Jeg har foretatt en teori- og forskningsevaluering av artikkelen med utgangspunkt i Karl Popper sitt kriterie om falsifiserbarhet. Basert på en grundig gjennomgang av artikkelen har jeg følgende forslag til forbedringer i det videre arbeid med problemstillingen:

1. Det teoretiske grunnlaget for de utsagn som testes bør klarlegges bedre. Spesielt gjelder dette antakelsen om at et diskriminerende prissystem som NETA vil lede til et lavere prisnivå enn ett-prissystem som the Pool.
2. Antakelsen om et stilltiende prissamarbeid forut for innføringen av NETA bør testes gjennom empiriske studier.
3. Man bør unngå bruk av visuell inspeksjon av dataene for å formulere testbare hypoteser.
4. Gyldighetsområdet til teorien som testes bør gjøres mer generell.

Forfatterens forklaring av prisfallet i engrosmarkedet for kraft har både teoriutledede elementer og elementer som er improvisert og tilpasset situasjonen. Selv om arbeidet slik sett ikke tilfredsstillende kravet til testbarhet og falsifiserbarhet, kan den gi grunnlag for å generere nye hypoteser som kan være falsifiserbare. Jeg vil derfor advare mot en for streng bruk av kriteriet i denne sammenhengen. Artikkelen kan, mener jeg, heller vurderes som et kreativt og tidlig innspill i en prosess som kan føre til mer generelle analyser av de samfunnsmessige effektene av bruken av ulike typer markedsklaringsprinsipp.

2 Presentasjon av artikkelen

2.1 Sammendrag

De britiske myndighetene iverksatte en rekke tiltak mot slutten av forrige tiår for å redusere strømprisene i England og Wales. Konsentrasjonen av kraftprodusenter ble redusert og nye regler for markedsklarering ble tatt i bruk. Samtidig økte produksjonskapasiteten.

Evans og Green har foretatt økonometriske analyser av utviklingen i strømprisene i perioden april 1996 til september 2002. Målet har vært å finne hvilke av de tre faktorene nevnt over som bidro mest til prisfallet i perioden. Spesielt ønsket de å kaste lys over effekten av innføringen av New Electricity Trading Arrangement (NETA).

De finner støtte for to motstridende hypoteser. Fra et statisk synspunkt ser det ut til at økt konkurranse og økt kapasitetsmargin er hovedårsakene til prisfallet. Dersom produsentene hadde samarbeidet om å holde prisene oppe under de gamle reglene for markedsklarering, kan imidlertid utsikter til innføringen av NETA ha endret deres markedsatferd *noen måneder før* reglene trådte i kraft. Når Evans og Green justerer den økonometriske modellen i tråd med dette spillteoretiske resonnementet, finner de støtte for at introduksjonen av NETA var den viktigste årsaken til prisfallet.

2.2 Bakgrunn

I *avsnitt 2* presenteres utviklingen i kraftsektoren fra restruktureringen av kraftsektoren i 1990 til introduksjonen av NETA i 2001. Beskrivelsen er detaljert og relativt teknisk. De viktigste utviklingstrekkene er:

Mars, 1990: Det statlige selskapet for kraft blir splittet opp i ett nasjonalt nettselskap, NGC, og flere produksjonsselskap som delprivatiseres. Samtidig etableres et spot marked som skal sikre balanse mellom etterspørsel og tilbud i engrosmarkedet for kraft; "the Pool".

Kraftprodusentene leverer daglig inn bud på salg av kraft for hver halvtime i døgnet.

Budene rangeres, og systemprisen (SMP) settes lik gjennomsnittsbudet til den marginale produsenten. SMP danner utgangspunkt for kjøps- og salgspriser i engrosmarkedet for kraft.

1994-1998: Prisnivået ligger over variable kostnader for kraftproduksjon. Dette fører til nyinvesteringer men og til kritikk av kraftprodusentene for å utnytte sin markedsrett. De store kraftprodusentene blir nødt til å selge/lease ut deler av sin produksjonskapasitet. Systemprisregelen i "the Pool" blir kritisert. Det hevdes at produsentene kan sikre sine leveranser ved å levere lave bud på deler av sin produksjonskapasitet, mens et lite antall høye bud er med å sette prisnivået for hele markedet.

Mars, 2001: Den nye markesmekanismen; NETA, blir etablert. Mer enn 95% av handelen i kraftmarkedet skjer nå gjennom bilaterale avtaler. I tillegg er det etablert en sentral mekanisme som skal sikre stabilitet markedet de siste timene før faktisk levering. Aktørene i markedet informerer NGC om sin planlagte fysiske posisjon en time før faktisk levering. Produsenter og leverandører leverer salgs- og kjøpsbud for å justere disse posisjonene. NGC balanserer markedet ved å akseptere noen av disse. De aktørene som er i ubalanse ved levering må kjøpe kraft til "System Buy Price" (SBP) eller selge kraft til "System Sell Price" (SSP). Siden SBP normalt er høyere enn SSP vil selskapene forsøke å unngå å være i ubalanse ved levering.

2.3 Resultater fra tidligere forskningsarbeid

Avsnitt 3 viser resultatene av en rekke forskningsarbeid mht. utnyttelse av markedsrett i kraftbransjen. Framstillingen er kortfattet, og fokuserer på resultatene framfor den underliggende teorien. Noen viktige resultater er:

Utnyttelse av markedsrett kan føre til et prisnivå over marginalkostnadene til produsentene. Evans og Green viser til flere empiriske studier der prismarginen øker med økt konsentrasjon av produsenter og redusert kapasitet i forhold til etterspørsel. Teoretiske studier viser hvordan varierende grad av prissikring kan føre til variasjoner i prismarginen.

Teoretiske studier av effekten på prismarginen av ulike markedsregler gir ingen entydig konklusjon. Fabra et al (2002) konkluderer med at pris auksjoner med én pris ("the Pool") er (svakt) mer effektiv, mens en auksjon med diskriminerende priser (NETA) gir et noe høyere konsumentoverskudd. På bakgrunn av dette argumenterer Evans og Green for at innføringen av NETA kan ha bidratt til redusert prismargin.

2.4 Modell og data

Avsnitt 4 viser oppbyggingen av regresjonsmodellen og valg av data. Framstillingen er detaljert og grundig mht. valg av data og estimering av variable som inngår i modellen.

Regresjonsmodellen forsøker å forklare prismarginen i engrosmarkedet med grad av konsentrasjon i markedet, kapasitetsmargin og innføringen av nye markedsregler:

$$SMPLerner = \alpha + \beta_1 Herfindale + \beta_2 Herfindale^2 + \beta_3 Demand / Capacity + \beta_4 NETA$$

$$PSPLerner = \alpha + \beta_1 Herfindale + \beta_2 Herfindale^2 + \beta_3 Demand / Capacity + \beta_4 NETA$$

Variablene i modellen er estimert som følger:

Den avhengige variabelen:

Evans og Green bruker en Lerner indeks som avhengig variabel; prismargin delt på pris. Estimeringen av denne reiser flere problemstillinger. Evans og Green tar utgangspunkt i systemprisen og salgsprisen i engrosmarkedet under "the Pool", henholdsvis SMP og PSP. Evans og Green gjør nøye rede for hvordan de justerer prisdata fra UK Power Exchange for å komme fram til sammenlignbare data etter innføringen av NETA. Videre må de finne et estimat på marginalkostnaden ved produksjon av kraft. I dette regnestykket inkluderer de bl.a. forskjeller i kostnadsstruktur og effektivitet til ulike typer kraftverk.

De uavhengige variablene:

Herfindahlindeksen skal gi et mål på graden av konsentrasjon i markedet. Evans og Green tar utgangspunkt i månedlige registreringer av produsentenes produksjonskapasitet og kalkulerer indeksen som summen av kvadrerte andeler av kapasiteten. De inkluderer også en kvadrert Herfindahlindeks under henvisning til en studie som viser en ikke-lineær sammenheng mellom tilbudsfunksjonen og antall bedrifter. Grad av ledig kapasitet i markedet er målt som gjennomsnittlig etterspørsel over gjennomsnittlig registrert kapasitet. Evans og Green har inkludert en dummy for å vurdere effekten av innføring av NETA.

Evans og Green har tatt utgangspunkt i månedlige gjennomsnittsdata for perioden April 1996 til September 2002. De viser til intervensjoner fra regulerende myndigheter og andre hendelser i markedet, som gjør første halvdel av nittitallet lite egnet for regresjonsanalyse.

2.5 Resultater

Avsnitt 5 viser resultatene av den statiske regresjonsmodellen. Evans og Green sin første hypotese er at introduksjonen av NETA kan ha ført til en nedgang i prismarginen på tidspunktet for endringene i markedsreglene.

Forklaringskraften til modellene er på rundt 60%. Ved 5% signifikansnivå har både enkel og kvadrert Herfindahlindeks en signifikant effekt på Lerner indeksene, mens kapasitetsmarginen slik den er målt her kun har en signifikant effekt på Lerner indeksen for SMP.

Dummyvariabelen for NETA er ikke signifikant.

Resultatene samsvarer godt med en empirisk analyse utført av Bower (2002). På bakgrunn av en litt mer detaljert modell, finner Bower at effekten av konsentrasjon og kapasitetsmargin på prisnivået er signifikant, mens effekten av innføringen av NETA ikke er entydig.

Avsnitt 6 viser resultatene av den dynamiske regresjonsmodellen. Evans og Green sin andre hypotese er at utsikter til en endring i markedsreglene fører til en nedgang i prismarginen 6 måneder før innføringen av NETA.

En visuell inspeksjon av Lerner indeksene viser et markert fall i prismarginene 6 måneder før innføringen av NETA. Sweeting (2001) sin empiriske studie av kraftpriser i England og Wales indikerer et stilltiende samarbeid mellom kraftprodusentene mot slutten av 90-tallet. Stilltiende samarbeid vil normalt tilsi at man oppgir gevinster i dag fra en mer konkurrerende strategi mot utsikter til høyere overskudd over tid. Dersom aktørene ikke tror det vil være mulig å opprettholde en samarbeidsstrategi under NETA, vil de etter som tidspunktet for regelendringen nærmer seg endre strategi. Perioden med mergevinst fra samarbeid blir stadig kortere, og vil på et tidspunkt forut for regelendringen ikke dekke det kortsiktige tapet fra samarbeidsstrategien.

Evans og Green utvider regresjonsmodellene med en dummy for 6 måneders perioden forut for innføringen av NETA. Nye estimeringer viser at innføringen av NETA og endringer i kapasitetsmargin har hatt en signifikant effekt på prismarginen i perioden. Endringer i industrikonsentrasjon har nå ikke lenger en signifikant effekt.

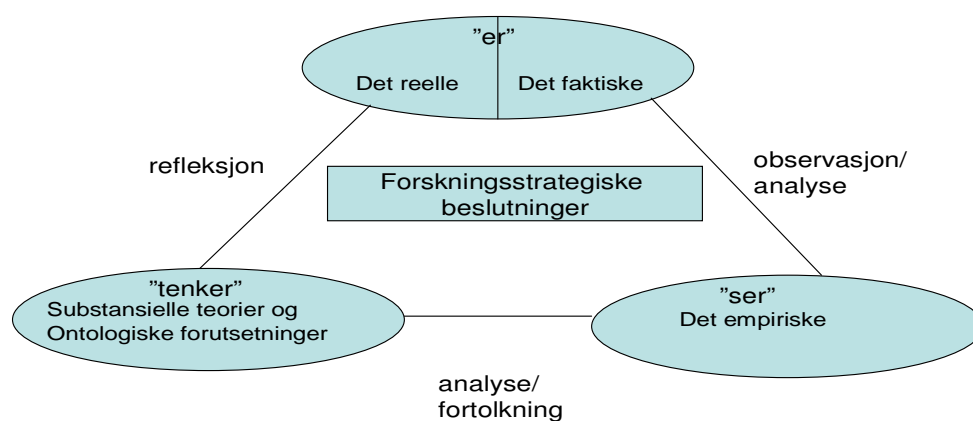
2.6 Konklusjon

Avsnitt 7 konkluderer artikkelen. Evans og Green avslutter artikkelen med at de er tilbøyelig til å støtte den andre hypotesen. De argumenterer med at denne regresjonen gir en bedre tilpasning til datane og viser til Sweeting sin empiriske studie som underbygger antakelsen om stilltiende prissamarbeid forut for NETA.

3 Presentasjon av og begrunnelse for valg av vurderingskriterier

3.1 Mitt vitenskapssyn

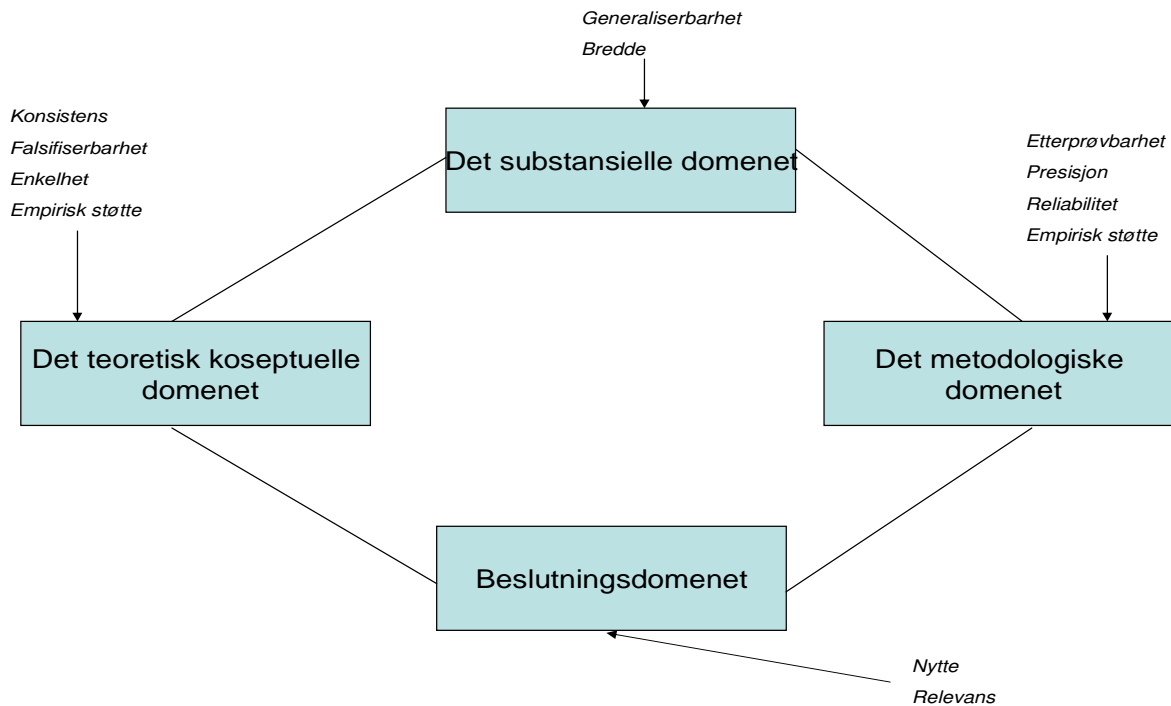
Min vurdering av artikkelen til Evans og Green vil avspeile mine verdier og grunnleggende antakelser om hvordan ting henger sammen. Jeg kan gi min tilslutning til et vitenskapssyn som *kritisk realisme* der virkeligheten har en eksistens utenfor forskeren selv, men der denne virkeligheten ikke kan observeres direkte men kan reflekteres i forskerens data og iakttagelser. Dette betyr at jeg vektlegger empirisk orientert forskning. Samtidig erkjenner jeg at observasjoner er teoriladet, slik Popper har påpekt, og at teorien styrer hvilke data vi velger og hvilke metoder vi bruker for å analysere disse. Figuren under viser hvordan et kritisk-realistisk syn vil påvirke forskeren mht hvordan hun oppfatter virkeligheten og de valg hun gjør mht teori og metode. Min vektlegging av empirisk forskning utelukker imidlertid ikke at jeg samtidig mener at deler av forskningen bør være normativ og bestå av bidrag som kan karakteriseres som rasjonalistiske. Her vil logiske resonnementer snarere enn empiri brukes for å evaluere en teoris holdbarhet. Mitt forskningssyn har mye til felles med både rasjonalisme, empirisme og idealisme, men ligger trolig noe nærmere de to førstnevnte. Her er jeg nok farget av etablerte tenkemåter innenfor mitt fagfelt; samfunnsøkonomi.



Figur 1: Et kritisk-realistisk syn

3.2 Artikkelen og de fire forskningsdomenene

De kriterier jeg legger til grunn for evalueringen av artikkelen til Evans og Green må være forenlig med mitt vitenskapsteoretiske utgangspunkt; kritisk realisme. Samtidig bør mitt valg av kriterier reflektere de forskningsdomenene som artikkelforfatterene har valgt å fokusere.



Figur 2: Ulike kriteriers relevans for ulike domener

Min vurdering er at Evans og Green sitt forskningsarbeid har hovedfokus på det *metodologiske domenet* og i noe mindre grad på *det teoretisk koseptuelle domenet*.

Utpøving av etablerte spillteoretiske resonnementer på nye problemstillinger leder Evans og Green fram til nye konklusjoner på effekten av innføring av NETA. Det teoretisk koseptuelle domenet er derfor viktig for forskningsarbeidet. Samtidig er denne delen av artikkelen relativt kortfattet og orientert mot resultater av tidligere forskningsarbeid. Derimot er en vesentlig del av artikkelen viet valg av observasjons- og analysemetode. Valg mht. hvilke data som samles inn og hvordan de blir brukt i regresjonsmodellene blir grundig gjort rede for. Resultatene fra estimering av en rekke regresjonsmodeller blir presentert i sammen med Evans og Green sine tolkninger.

De to andre domenene; det substansielle domenet og beslutningsdomenet, er og sentrale i Evans og Green sin artikkel. Deres forskningsbidrag er et eksempel på anvendt forskning som er knyttet opp mot en aktuell problemstilling. Arbeidet har en klar nytteverdi da det kan brukes til å belyse de samfunnsøkonomiske kostnadene og inntektene ved et eksisterende regelverk. På den annen side har forfatterne i liten grad problematisert og drøftet i hvilken grad deres arbeid bør føre til revisjoner av marknadsreglene. Forfatterne er og svært tydelige mht. beskrive og avgrense hvilken del av virkeligheten de ønsker å studere; effekten på kraftprisene i England og Wales av innføringen av NETA i mars 2001. På den annen side har forfatterne i liten grad problematisert og drøftet i hvilken grad deres arbeid har overføringsverdi til andre situasjoner eller om det kun gjelder dette isolerte tilfellet.

3.3 Mitt valg av evalueringskriterium

Jeg har valgt å fokusere på kriteriet *falsifiserbarhet* i min vurdering av artikkelen. Kravet om falsifiserbarhet reflekterer en empirisk vitenskaplig forankring som bl.a. er forenlig med kritisk realisme.

Kravet om falsifiserbarhet retter seg mot de valg som gjøres innenfor det teoretisk-konseptuelle domenet, dvs. hvordan de teoretiske relasjonene er formulert. Samtidig må man ved vurdering av falsifiserbarhet også stille spørsmål ved hvordan empiriske tester er utformet og hvordan de empiriske resultatene er drøftet. Det betyr at kriteriet også er sentralt innenfor det empirisk-operasjonelle plan. Kriteriet er m.a.o. sentralt innenfor de to forskningsdomenene som jeg oppfatter har det sterkeste fokuset i artikkelen.

Til syvende og sist henger mitt valg av kriterie også sammen med et spørsmål jeg stilte meg selv ved første gjennomlesning av artikkelen til Evans og Green; har forskerne ved utforming av hypoteser og testing av disse stilt seg selv de nødvendige kritiske spørsmål for å sikre erkjennelsesmessig holdbarhet?

Jeg kunne valgt å også ta med kriteriet om empirisk støtte ved vurdering av artikkelen. Siden det er lettere å oppnå empirisk støtte for teorier som har lav falsifiserbarhet, er det fornuftig å se disse kriteriene i sammenheng når vi vurderer om teorien har erkjennelsesmessig holdbarhet. Jeg har likevel valgt å ha hovedfokus på falsifiserbarhetskriteriet, men vil underveis trekke inn noen synspunkt på graden av empirisk støtte i artikkelen.

3.4 Falsifiserbarhet

For å vurdere en teori sin holdbarhet ønsker vi å teste hvorvidt prediksjonene holder empirisk. Vi tester en antakelse om prosesser (i det reelle) gjennom de antatte effekter disse har (i det faktiske) ved å sammenholde antakelsene med det vi faktiske observerer (ser).

Karl Popper introduserte falsifiserbarhet som et demarkasjonskriterie for vitenskapen i sin berømte bok; *The Logic of Scientific Discovery* (Popper, K. (1959)). Her argumenterer han for at bare de teoriene som er testbare og falsifiserbare på basis av observasjoner eller eksperiment kan kalles vitenskaplige.

Karl Poppers krav om *falsifiserbarhet* kan sammenlignes med positivistenes krav om *verifiserbarhet*. Kriteriene har det til felles at holdbarheten av slutninger fra det observerte til det ukjente *ikke kan bevises*. For positivistene er konsekvensen at dette at de begrenser seg til å avdekke funksjonsformer som beskriver observerte sammenhenger. Popper mener imidlertid at vitenskapens oppgave er å finne fram til generell kunnskap som går utover det observerte. Poppers utgangspunkt er ”Det som ikke gjelder spesielt, kan heller ikke gjelde generelt, mens det som gjelder spesielt ikke nødvendigvis gjelder generelt.” Et minimumskrav til forskningen ut fra falsifikasjonsprinsippet er at forskningen drives av antagelser som settes på prøve gjennom empirisk testing. Et holdbart vitenskaplig utsagn er ikke nødvendigvis sant, men er holdbart i den forstand at det kan vise seg å ikke være sant.

Basert på følgende oppsett hentet fra Troye, S. V. (1994) vil jeg i det følgende vurdere i hvilken grad falsifiserbarhetskritet er tilfredsstillt i artikkelen til Evans og Green.

DET TEORETISK-KONSEPTUELLE PLAN	<ol style="list-style-type: none">1) Er teoriens utsagn tautologisk?2) Er de falsifiserbare eller kun bekreftbare?3) Er de ”selvfølgelige”?4) Er påstandene framsatt som følge av en kritisk gjennomgang av eksisterende litteratur der argumenter pro og kontra tillegges rimelig vekt?
DET EMPIRISK-OPERASJONELLE PLAN	<ol style="list-style-type: none">5) Er de empiriske testene utformet på en slik måte at de vil gi en kritisk test, og at de i størst mulig grad gir grunn til å utelukke alternative forklaringer?6) Er resultatene av de empiriske testene drøftet kritisk?

4 Evaluering

4.1 Teoriens utsagn

Innledningsvis vil jeg forsøke å få klart fram hvilke teoretiske utsagn som Evans og Green vil teste i sin artikkel. Dette er ikke uten videre klart, da verken teorien eller hypotesene er eksplisitt formulert.

Evans og Green starter sin artikkel med å slå fast at det er en sammenheng mellom høye priser i kraftmarkedet og tre variable:

”It is a truth universally acknowledged, that an electricity market with high prices is in want of a remedy. The causes of high prices include insufficient generation capacity, market power stemming from excess concentration among generators, and inappropriate market rules.”

Målsettingen er m.a.o. ikke å teste hvorvidt redusert konsentrasjon, økt kapasitet og et bedre fungerende markedsregler vil gi reduserte kraftpriser. Målet med artikkelen er:

”This paper is an attempt to assess which of those remedies had the greatest impact on the level of electricity prices. In particular, we hope to shed light on the extent to which the adoption of the New Electricity Trading Arrangement (NETA) directly caused the drop in wholesale prices which occurred at about the same time.”

Min tolkning er at Evans og Green primært ønsker å teste hvorvidt NETA innebar en forbedring av markedets funksjonsmåte. Dette gjør de ved å teste hvorvidt innføringen av NETA ga et fall i prismarginen. De er åpne for at tidspunktet for endringen i aktørenes markedsatferd ikke nødvendigvis er sammenfallende med innføringen av NETA, men kan ha skjedd en tid før regelendringen. De tester derfor to alternative hypoteser. Følgende to sitat er det nærmeste jeg har kommet de to utsagnene de ønsker å teste:

”Our first hypothesis was that the introduction of NETA might have caused a change in the level of the Lerner index, at the time that the rules changed.”

“If the generators believed that they would not be able to maintain a collusive equilibrium after the introduction of NETA, then they might have had to abandon any tacitly collusive

practices once NETA drew near. The extra profit that they could have gained from only a few more months of tacit collusion was no longer enough to outweigh the immediate profits from more competitive behaviour. We therefore ran a new set of regressions, with a separate dummy for the “pre-NETA” period of October 2000 to March 2001.”

Legg merke til at vi gjennom empiriske tester kun kan finne støtte for en hypotese og ikke den underliggende teorien. Forskjellen er at teorien inneholder en forklaring på hvorfor en endring i X fører til en endring i Y. I sitat nummer to over har forfatterene inkludert en slik forklaring. Det betyr at vi ikke kan finne empirisk støtte for dette utsagnet slik det er formulert.

4.2 Falsifiserbarhet og tautologiske utsagn

Evans og Green slår innledningsvis fast at det er en sammenheng mellom ”innappropriate market rules” og høye prismarginer. Dette er et eksempel på et tautologisk utsagn. Begrepet ”inappropriate market rules” er ikke definert av Evans og Green, men jeg tolker det slik at det impliserer markedssituasjoner som avviker fra perfekt konkurranse. En forbedring av marknadsreglene vil derfor pr. definisjon medføre en prissetting mer i tråd med fri konkurranse der prisene er identisk med marginalkostnadene. Og, et marked med priser ned mot marginalkostnadene, vil normalt tyde på godt fungerende marknadsregler. Slike sirkelresonnement er ikke testbare. Evans og Green velger imidlertid å teste hvorvidt innføringen av NETA ga reduserte prismarginer. De tester dermed en relasjon som ikke er tautologisk, og tilfredsstillter derfor dette kravet til falsifiserbarhet.

4.3 Falsifiserbarhet og verifiserbarhet

Min tolkning av utsagnene er at de er både *verifiserbare* og *falsifiserbare*. Dette svekker utsagnenes falsifiserbarhet. Et viktigere brudd på kravet om falsifiserbarhet er imidlertid at utsagnene synes å forklare *en enkeltstående situasjon*, og at gyldighetsområdet til utsagnene er avgrenset til den aktuelle situasjonen som oppsto i kraftmarkedet i England og Wales rundt 2000.

Vi kan verifisere det første utsagnet for de aktuelle observasjonene dersom vi for eksempel observerer et fall i prismarginene ved innføringen av NETA når vi korrigerer for andre faktorer som kan påvirke marginen. Det betyr imidlertid ikke at vi har funnet bevis for en kausal

sammenheng mellom innføring av NETA og reduserte prismarginer som også ville funnet hadde NETA blitt introdusert fire år senere. Vi kan også falsifisere utsagnene. Dersom vi, alt annet like, hadde observert en økning i prismarginene 6 måneder forut for innføringen av NETA, ville vi falsifisert utsagn 2.

Når utsagnene er både verifiserbare og falsifiserbare, får vi en lavere grad av falsifiserbarhet enn hvis de er kun falsifiserbare. Et eksempel på et utsagn som er kun falsifiserbart er:

”Dersom vi i et engrosmarked for kraft erstatter et ettprissystem (som the Pool) med et flerprissystem (som NETA) vil prismarginene bli redusert eller forbli uendret alt annet like.”

Utsagnet kan falsifiseres, og er da usant. Mangel på falsifisering kan imidlertid ikke brukes som bevis for at denne typen markedsregler har en positiv effekt på prismarginen jamført med et ettprissystem. Derimot vil man gjennom gjentatte empiriske undersøkelser på reformer av kraftmarkedet i andre land kunne samle støtte for utsagnet.

Evans og Green har i sin artikkel formulert teoretisk relasjoner mellom innføring av NETA og endring i Lerner indeksene som jeg oppfatter som relativt situasjonsbestemte. Blant annet er det i sitatet over som jeg tolker som hypotese 2, lagt inn en antakelse om at det fant sted et stilltiende prissamarbeid forut for innføringen av NETA. Denne antakelsen er basert på ett empirisk arbeid:

” A possible explanation for this drop in the Lerner indices may come from a more “dynamic” view of the electricity market. Recall that Sweeney (2001) had found evidence of behaviour consistent with tacit collusion during the late 1990s.”

Antakelsen om prissamarbeid er ikke testet av Evans og Green og heller ikke inkludert som en variabel i modellen. Derfor blir nyttverdien av resultatene av testen avgrenset til først å fremst å forklare det som skjedde med prisnivået i det kraftmarkedet for England og Wales rundt år 2000. Forfatterne foreslår selv en forbedring av sitt arbeid ved å:

”An examination of the bid data around this time would show which companies changed their strategies, and might allow us to better discriminate between explanations for the fall in prices”.

Avslutningsvis gir forfatterne klart uttrykk for at fallet i prismarginene som følge av innføringen av NETA var situasjonsbestemt, og derfor ikke vil være reproduerbare i en annen kontekst:

”A number of academic economists (including one of us) had suggested that NETA would not necessarily be less vulnerable to high prices than the Pool-based system. If our tentative explanation for the fall in margins that preceded NETA is indeed correct, it is perhaps fortunate for consumers that the companies appear not to have listened to them.”

M.a.o. med andre aktører og andre tolkninger, kunne prissamarbeidet og de høye prismarginene kanskje ha fortsatt etter innføringen av NETA.

4.4 Falsifiserbarhet og utsagnsinnhold

Innholdet i utsagnene som testes bør ikke være trivielt ut fra eksisterende viten, men bør være slik at det knytter seg en viss usikkerhet til om utsagnet empirisk vil holde mål. Jeg forstår dette kravet dithen at verdien av empirisk støtte vil øke med hypotesens dristighet.

Hypotese nummer to som Evans og Green tester, om et prisfall forut for innføringen av NETA, kan synes dristig. Den er i hvertfall ikke trivell da verken tidligere empiriske undersøkelser eller teori skulle tilsi at en slik sammenheng opplagt eksisterer.

På den annen side er hypotese nummer to formulert etter en inspeksjon av datane:

”However, a visual inspection of our data (figure 2) reveals an obvious fall in the Lerner indices in October 2000, six months before the introduction of NETA”.

Denne observasjonen samt et spillteoretisk resonnement og Sweeting sine funn som tyder på prissamarbeid mot slutten av nittitallet, er bakgrunnen for formuleringen av hypotese nummer to. Sannsynligheten for at utsagn nummer to vil bli falsifisert ved empirisk testing er dermed sterkt redusert.

4.5 Falsifiserbarhet og utsagnets teoretiske og empiriske fundament

Evans og Green sin artikkel inneholder en rekke forutsetninger og antakelser som er nødvendige for å kunne forklare det som er teoriens sentrale fenomen, uten at disse antakelsene i seg selv er direkte testbare og falsifiserbare. Milton Friedman er blant en rekke økonomer som mener at det i slike situasjoner kan være tilstrekkelig å teste om prediksjonene holder og la dette være et kriterium på hvorvidt teorien i sin helhet holder. Dette vil igjen avhenge av at disse antakelsene hviler på et solid teoretisk og empirisk fundament. Jeg vil i det følgende se litt nærmere på de forutsetningene som Evans og Green sine to utsagn hviler på.

Evans og Green slår innledningsvis fast at prismarginen vil øke med lav kapasitetsmargin, høy konsentrasjon i industrien og "inappropriate market rules". De finner belegg for effekten av de to første variablene i en rekke teoretiske og empiriske studier. Eksempler på referanser til empiriske studier i teksten er:

" Wolfram (1998,1999) studied bids and prices in the England and Wales market. She found that the mark-ups between price and marginal cost were higher when demand was above the median level, and that generators tended to submit higher bids, relative to marginal costs, the greater the level of capacity that was infra-marginal to the plant. Sweeting (2001) finds that mark-ups in a given quarter are generally higher when there is little spare capacity. He finds that mark-ups vary over time, however, and relates this to changes in behaviour from not exploiting their (considerable) market power in the mid-1990s to exploiting the (much lower) degree of market power that remained to them, and possibly even colluding by the end of his sample period in 2000"

Jeg finner at sammenhengen mellom kapasitet og konsentrasjon på den ene siden og prismarginene på den andre siden er godt gjort rede for i Evans og Green sitt arbeid. Jeg har imidlertid mer problem med den vekt de tillegger Sweeting sin studie når de i sin hypotese 2 tar som gitt at produsentene i kraftmarkedet deltok i prissamarbeid forut for innføringen av NETA. Resultatene av Sweeting sin studie er ikke gjort rede for. Datagrunnlaget for Sweeting sin studie går kun fram til september 2000. Ytterligere dokumentasjon på et prissamarbeid i den aktuelle situasjonen finnes ikke utover henvisning til rykter:

”Some anecdotes within the industry are consistent with this view¹, although others suggest that the generators chose a lower price strategy because they were no longer selling plant.”

Dersom vi ikke tar et slikt prissamarbeid for gitt, bør man ved test av utsagn 2 ideelt sett få fram om følgende sett av påstander gjelder samtidig:

- At produsentene inngikk i stilltiende prissamarbeid forut for innføringen av NETA
- At produsentene antok at denne form for prissamarbeid ikke ville være mulig etter innføringen av NETA.
- At produsentene ut fra en nytte-kostnadsanalyse valgte å gå ut av prissamarbeidet 6 måneder før NETA.
- At denne endringen i atferd, isolert sett, bidro til å redusere prismarginene.

Resultatet av regresjonene som er gjennomført i denne artikkelen kan ikke brukes til å vurdere disse påstandene, og den empiriske støtten som teorien deres oppnår må vurderes i dette lys.

Til slutt tar forfatterne for seg en rekke teoretiske studier av effekten av ulike markedsregler:

”Bower and Bunn (2000) used a simulation modell to predict that moving from a daily auction with a uniform price to hourly auctions with discriminatory pricing (i.e. , from a simplification of the Pool system to NETA) would lead to higher prices. Fabra (2002), however, shows in a theoretical model that a discriminatory auction is less vulnerable to the exercise of market power than a uniform-price auction is. The uniform-price auction allows generators to receive a high price while still submitting low bids, and therefore minimising the pay-off from deviating to a more competitive strategy. Fabra et al (2002) compare several auction formats and find that their welfare ranking is ambiguous – uniform price auctions are (weakly) more efficient, but a discriminatory auction is (weakly) better for consumer surplus. In other words, the switch to NETA might be expected to reduce prices.

Evans og Green er nøye med å få fram at resultatene av disse studiene ikke er entydige. Jeg synes likevel denne gjennomgangen, som jeg har gjengitt i sin helhet, er for lite grundig. Jeg

¹ Dvs. gå fra en samarbeidsstrategi med høy pris til en konkurrerende strategi med lav pris i månedene forut for innføringen av NETA.

savner en god redegjørelse for hvorfor NETA vil føre til redusert prismargin. Dette er viktig ettersom ulike økonomer trekker ulike konklusjoner basert på teoretiske studier.

Kort oppsummert mener jeg kravet til falsifiserbarhet kan møtes på en bedre måte ved:

- Å gi en teoretisk redegjørelse for hvorfor et diskriminerende prissystem som NETA er forventet å gi reduserte prismarginer.
- Å framskaffe bedre empirisk grunnlag for hvorvidt aktørene inngikk i kartellignende prissamarbeid forut for innføringen av NETA.

4.6 Falsifiserbarhet og utforming av empiriske tester

De empiriske testene består av estimering av regresjoner av typen:

$$SMPLerner = \alpha + \beta_1 Herfindale + \beta_2 Herfindale^2 + \beta_3 Demand / Capacity + \beta_4 NETA$$

$$PSPLerner = \alpha + \beta_1 Herfindale + \beta_2 Herfindale^2 + \beta_3 Demand / Capacity + \beta_4 NETA$$

Evans og Green har med ulike varianter av disse regresjonsmodellene. Den viktigste modifikasjonen er at de ved testing av utsagn nummer to utvider modellene med en dummyvariabel for 6 måneders perioden forut for innføringen av NETA.

Et viktig spørsmål er hvorvidt det er et en-til-en forhold mellom teoriens innhold og logiske struktur og den måten en teori testes på. Jeg har litt problem med å gjøre denne vurderingen av Evans og Green sitt arbeid, i og med at teoriens innhold ikke er eksplisitt formulert i form av hypoteser. Noe av problemet med artikkelen, slik jeg ser det, er at forfatterene ikke klart skiller mellom de postulerte sammenhengene i ”det reelle”, det de forsøker å måle i ”det faktiske” og de dataene som ligger til grunn for de empiriske datane. Likevel vil jeg konkludere at det er godt samsvar mellom det forfatterne oppgir som artikkelens mål; jfr. andre sitat i dette hovedkapittelet, og regresjonsoppsettet over.

Et annet viktig spørsmål ved vurdering av falsifiserbarhet er i hvilken grad man har isolert effekten av variable som er teorien uvedkommende. Innenfor regresjonsanalyse kan dette oppnås gjennom styrte eksperiment, eller ved å inkludere flest mulig variable som kan tenkes

å ha innvirkning på den forklarte variabelen. Den siste teknikken er vanlig innenfor samfunnsfagene fordi vi sjelden har anledning til å iscenesette eksperiment.

Forfatterne nevner selv flere variable som kan tenkes å påvirke prismarginene, men som ikke er inkludert i modellen:

”An alternative explanation for changes in generators’ behaviour over time, to which both Sweeting and Wolfram allude, is a change in their contractual position. [] Wolak(2000) uses very detailed data from the Australian electricity market to illustrate how these relationships held in practise.”

“National Power and PowerGen would have had a strong incentive to keep prices high while they were attempting to sell power stations, but that this incentive would have fallen as soon as the sales were completed, in October 2000.”

“An alternative version of the dynamic story is that generators faced with an imminent rise in competition would be able to continue colluding, but not to the same extent.”

Felles for disse tre forholdene, er at de er vanskelig eller umulig å observere. Det kunne imidlertid være verdt et forsøk å estimere en variant av modellene over der man hadde forsøkt å estimere graden av prissikring i markedet.

Et tredje viktig spørsmål ved vurdering av falsifiserbarhet er hvorvidt forfatterne har foretatt et utvalg av objekter som har variasjon mht. det man ønsker å forklare. Siden Evans og Green kun analyserer en enkelthendelse, innfrir de ikke dette kravet. Nye empiriske studier kan imidlertid supplere studien til Evans og Green ved å få fram hva som kjennetegner markedsreformer i kraftmarkedet som har blitt etterfulgt av økning i prismarginen for eksempel .

4.7 Falsifiserbarhet og vilje til å erkjenne falsifikasjon

Evans og Green estimerer varianter av regresjonsmodellene over. Resultatene er kortfattet:

Utsagn I, statistisk modell:

Modellen forklarer mellom 58 til 65% av variasjonene i prismarginene. Innføringen av NETA er ikke signifikant (p-verdier på mellom 0,33 og 0,50). De andre forklaringsvariablene er i hovedsak signifikante ved 5% signifikansnivå.

Utsagn II, dynamisk modell:

Her er det to hovedvarianter av regresjonsmodellene. Den første har en egen dummyvariabel for de 6 månedene forut for innføringen av NETA. I den andre er eksisterende dummyvariabel for NETA utvidet med 6 ekstra måneder.

Forklaringskraften er nå svakt forbedret og ligger mellom 63-68% for begge modellene. Begge modellene gir som resultat at Herfindahl indeksene sin effekt på prismarginen ikke lenger er signifikant, dvs. at graden av konsentrasjon i næringen ikke har en isolert effekt på prismarginene. Derimot har kapasitetsmarginen en klart signifikant effekt på prismarginen.

Der det er lagt inn en egen dummy for perioden forut for innføringen av NETA, er det kun denne dummyvariabelen som er signifikant – og ikke dummyvariabelen for perioden etter innføringen av NETA. Dette vurderer forfatterne som et merkelig resultat, og de argumenterer som følger for å innføre en felles dummyvariabel for hele perioden:

”The dummy for the six month period preceding NETA is larger and more significant than the NETA dummy, which we find slightly worrying. Admittedly, the autumn of 2000 saw an increase in marginal costs, which might not have been fully passed through to prices. In general, though, there was no reason to expect that anticipation of NETA would actually drive mark-ups below the level that would be sustained under NETA itself. F-tests revealed that the coefficients on the two dummy variables were not significantly different from each other. We therefore consolidated the two dummy-variables into a single variable, “para-NETA”, equal to one from October 2000 onwards.”

Det teoretiske argumentasjonen for å bruke en gjennomgående dummyvariabel virker logisk og i henhold til teorien.

I vurderingen av de to ulike hypotesene støtter Evans og Green seg dels på en ren statistisk vurdering av resultatene:

The standard statistical criteria (adjusted R-squared, Akaike (1974) Information criterion and Schwarz (1978) Criterion) all favour the regression from table 3².

Forskjellene i de statistiske resultatene er imidlertid helt marginale, og de faller tilbake til argumenter for et prissamarbeid for å argumentere for sin endelige konklusjon:

"We are currently inclined to support this second hypothesis. The regression are a better fit to the data, and Sweeting (2001) gives convincing evidence of tacit collusion for much of the critical period, although his work stops tantalising short, in November 2000."

Personlig synes jeg det er vanskelig ut fra utfallet av regresjonsanalysen og trekke en entydig konklusjon. Forklaringskraften er rimelig lik. Den andre hypotesen er modellert etter å ha "tyvtittet" på fasiten. Videre mener jeg forfatterne i større grad burde stilt spørsmålstegn ved hvorfor graden av konsentrasjon i næringen ikke lenger synes å ha en signifikant effekt på prismarginen.

4.8 Noen refleksjoner rundt bruken av falsifiserbarhetskriteriet

Empiriske studier gjennomført av bl.a. Wason (1960) viser at falsifikasjonskriteriet ofte ikke etterleves praksis. De empiriske studiene til bl.a. Wason (1960) og Mynatt, Doherty og Tweney (1977) illustrerer at forskningen kan ha karakter av å være en adaptiv læringsprosess der "context of discovery" og "context of justification" dels glir over i hverandre.

Vurdert ut fra dette kan Evans og Green sin artikkel vurderes som et tidlig bidrag der fokus primært er på "context of discovery". Ved å introdusere en spillteoretisk tankegang, stiller forfatterne seg åpne overfor nye måter å tolke aktørenes atferd på. De er i tillegg relativt forsiktige med å trekke bastante konklusjoner. Kanskje er det viktigste resultatet av deres arbeid at estimeringen av en dynamisk modell gir resultater som er forenlig med en form for stilltiende avtaler om prissamarbeid forut for NETA. Neste skritt i forskningen rundt NETA,

² Tabell 3 inneholder estimeringen av regresjonsmodellene for testing av utsagn II der det er brukt en gjennomgående dummyvariabel for hele perioden.

kan da være grundigere studier av hvorvidt et slikt samarbeid faktisk fant sted. Dette kan være et argument mot en for streng praktisering av falsifiseringskriteriet, spesielt i en tidlig fase av et forskningsprogram.

Videre, argumenterer mange forskere for en mer moderat praktisering av falsifiserbarhetskriteriet enn det som er gjengitt innledningsvis. De tar følgende sitat fra Popper (1959) til inntekt for at falsifikasjonsprinsippet ikke skal tas helt bokstavelig:

”In point of fact, no conclusive disproof of a theory can ever be produced, for it is always possible to say that the experimental results are not reliable, or that the discrepancies which are asserted to exist between the experimental results and the theory are only apparent and that they will disappear with the advancement of our understanding.”

Basert på dette kan f.eks Evans og Green sitt funn at Herfindal indeksene ikke har signifikant effekt på prismarginene i testingen av den dynamiske modellen brukes til å forkaste hypotesen om at økt konsentrasjon av produsenter i en næring gir økt prismargin. Evans og Green trekker da heller ikke en slik slutning. De velger belyse hvilke av tre faktorer som var mest sentral når det gjaldt å forklare prisutviklingen i en gitt situasjon.

5 Konklusjon

Jeg har foretatt en teori- og forskningsevaluering av artikkelen med utgangspunkt i Karl Popper sitt kriterie om falsifiserbarhet. Basert på en grundig gjennomgang av artikkelen har jeg følgende innvendinger til studien:

- Utsagnene som testes er både *verifiserbare* og *falsifiserbare*. Dette svekker utsagnenes falsifiserbarhet. Et viktigere brudd på kravet om falsifiserbarhet er imidlertid at utsagnene synes å forklare *en enkeltstående situasjon*, og at gyldighetsområdet til utsagnene er avgrenset til den aktuelle situasjonen som oppsto i kraftmarkedet i England og Wales rundt 2000.
- Hypotese nummer to er formulert etter å inspisert datagrunnlaget. Dette gjør det vanskeligere å falsifisere utsagnene og svekker tiltroen til resultatene. Videre er hypotesen basert på en rekke antakelser om aktørenes atferd i markedet som ikke er tilstrekkelig dokumentert i empiriske studier. Jeg savner også en god teoretisk redegjørelse for hvorfor NETA vil føre til redusert prismargin. Dette er viktig ettersom ulike økonomer trekker ulike konklusjoner basert på teoretiske studier.
- Det er vanskelig ut fra utfallet av regresjonsanalysen og trekke en entydig konklusjon. Forklaringskraften er rimelig lik. Den andre hypotesen er modellert etter å ha ”tyvtittet” på fasiten. Videre mener jeg forfatterne i større grad burde stilt spørsmålstegn ved hvorfor graden av konsentrasjon i næringen ikke lenger synes å ha en signifikant effekt på prismarginen.

Avslutningsvis vil jeg likevel påpeke at artikkelen kan betraktes som et kreativt og innsiktsfullt innspill i en prosess som kan føre til mer generelle analyser av de samfunnsmessige effektene av å bruke auksjonsprinsipp med diskriminende priser kontra auksjonsprinsipp med ettprissystem. Slik kan den danne grunnlag for å formulere hypoteser som er falsifiserbare.

Litteraturliste

Evans, J. og Green R. (2003): "Why Did British Electricity Prices Fall after 1998?", *Cambridge Working Papers in Economics SWPE 0326, CMI Working Paper 26*.

Troye, S.V. (1994) *Teori og forskningsevaluering*. TANO.

Popper, K. (1959) *The Logic of Scientific Discovery*. Hutchinson of London.

Popper, K. (1968) *Conjectures and Refutations*. Harper and Row.

Popper, K. (1972) *Objective Knowledge*. Oxford University Press.

Wason, P.C. (1960): "On the Failure to Eliminate Hypothesis in a Conceptual Task", *Quarterly Journal of Experimental Psychology*.

Mynatt, C. R., Doherty, M.E. og Tweney, R. D. (1977): "Confirmation Bias in a Simulated Research Environment: An experimental Study of Scientific Inference", *Quarterly Journal of Experimental Psychology*.