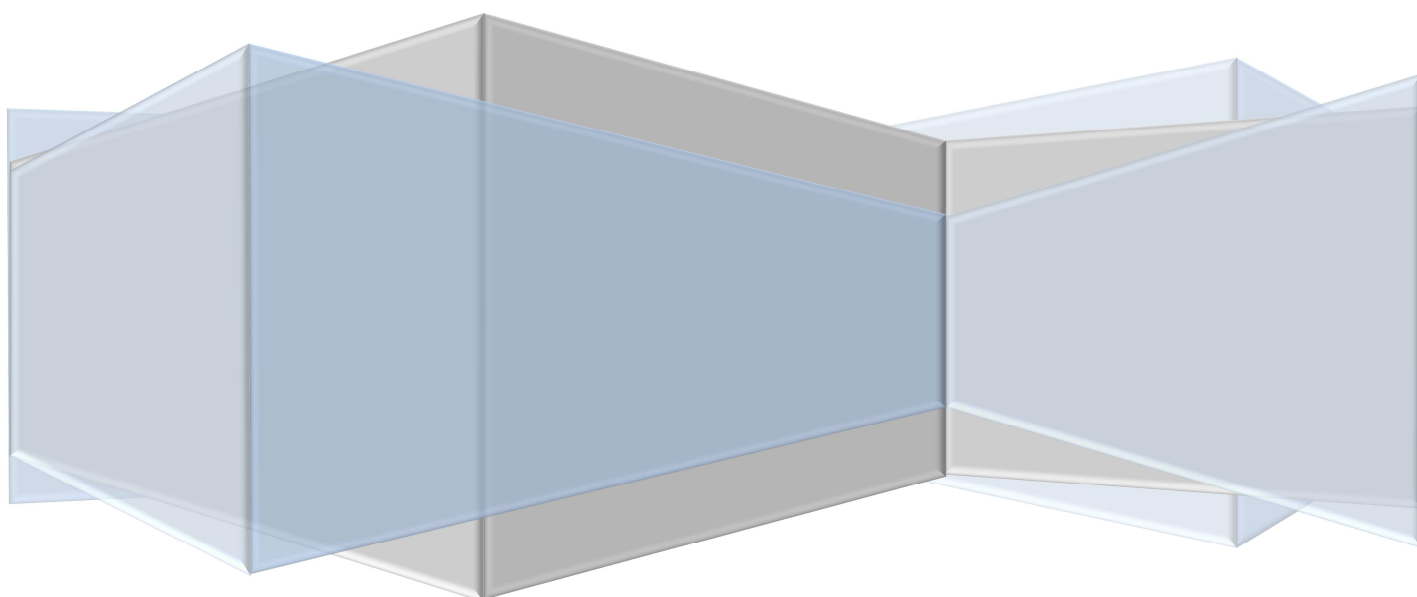


Sekretariatet for Energitilsynet

Baggrundsnotat: Model- teknisk materiale

Store forskelle i varmepriserne – hvorfor?

Center for Varme



Tekniske bilag

I dette baggrundsnotat gennemgås de økonomiske forhold omkring validiteten af den statistiske model. Der vil være fokus på de tekniske forhold omkring data og regressionsmodellen, som forklares og testes for de mest almindelige fejltyper og misspecifikationer. Regressionsmodellens robusthed undersøges ved forskellige følsomhedsanalyser og ved at gennemføre tilsvarende analyser på reducerede datagrundlag, for at se om de samme tendenser gør sig gældende, eller om resultaterne skyldes enkelte ekstreme observationer.

Indledningsvist argumenteres der for, hvorfor det er nødvendigt at benytte en statistisk model til at analysere forskellene i varmepriserne. Herefter diskuteres de generelle styrker og begrænsninger ved at benytte regressionsmodeller i analysesammenhæng. Dernæst vises to eksempler på, hvordan de modelforudsagte varmepriser beregnes, og hvordan modellen kan forudsige de meget forskellige varmepriser. Herefter beskrives datagrundlaget for analysen. Der præsenteres overvejelser om signifikansen af modellens parametre, hvorvidt heteroskedasticitet påvirker resultaterne og om graden af multikollinearitet vurderes at være problematisk. Endelig argumenteres for brugen af dummy-kategorier for fjernvarmeforsyningernes bymæssighed pga. manglende data for forbrugertæthed og fordi det tilgængelige data viser at forbrugertæthed ikke er statistisk signifikant.

Bilagene er tiltænkt den teknisk interesserede læser, med en grundlæggende forståelse af regressionsanalyser og statistiske test.

Hvorfor benytte en statistisk model?

I Baggrundsnotatet "Beskrivende statistik" til analysen: "Store forskelle i varmepriserne – hvorfor?" vises det at der er store forskelle på de gennemsnitlige fjernvarmepriser mellem forskellige grupper af fjernvarmeforsyninger. For eksempel fremstår kommunale fjernvarmeforsyninger billigere end forbrugerejede forsyninger og forsyninger med andre ejerskabstyper. Barmarksværkerne findes i gennemsnit at være betydeligt dyrere end fjernvarmeforsyninger, der ligger i områder med en anden bymæssighed. Samtidig er der stor variation i varmepriser blandt gruppen af kommunale fjernvarmeforsyninger og gruppen af barmarksværker.

Analysen i baggrundsnotatet "Beskrivende statistik" til analysen: "Store forskelle i varmepriserne – hvorfor?", der undersøger forskelle i gennemsnitspriserne og prisvariationerne i undergrupper af fjernvarmeforsyninger kan imidlertid ikke besvare spørgsmål, som for eksempel om de lave priser blandt kommunale fjernvarmeforsyninger i virkeligheden skyldes at kommunale forsyninger typisk er større end forbrugerejede forsyninger, og dermed kan opnå stordriftsfordele og lave priser? Om barmarksværkerne fremstår dyre fordi de typisk er meget små og fyrer med naturgas? Eller om det skyldes barmarksværkernes bymæssighed - altså at der er meget langt mellem aftagerne af fjernvarmen? De store prisforskelle inden for hver undergruppe af fjernvarmeforsyninger, indikerer at der er flere forhold, der påvirker fjernvarmeforsyningerne samtidigt, og derfor giver stor spredning i priserne.

Spørgsmålene kan ikke besvares ved at sammenligne simple eller vægtede gennemsnitspriser, som det gøres i baggrundsnotatet "Beskrivende statistik" til analysen: "Store forskelle i varmepriserne – hvorfor?" og i statistik fra Dansk Fjernvarme (i udgivelserne Årsstatistik og Benchmarking 2011). For at kunne afgøre hvor meget hver variabel betyder for fjernvarmepriserne er det nødvendigt at isolere effekten af de variable, der menes at påvirke fjernvarmepriserne. Statistiske modeller kan isolere effekten af hver

variabel, hvilket er baggrunden for at benytte en statistisk model i analysen "Store forskelle i fjernvarmepriserne – hvorfor?".

Eksempel: Forskellen på en statistisk model og et simpelt gennemsnit

Betydningen af at isolere effekten af en variabels påvirkning varierer alt efter, hvor meget variablene 'bevæger sig i samme retning' (er korrelerede). Hvis der er en høj grad af korrelation mellem variablene, vil effekten af hver variabel fejlvurderes, hvis der benyttes sammenligning af simple gennemsnit. Effekten af hver variabel for sig vil enten overvurderes eller undervurderes.

I dette eksempel vil vi fokusere på forskellen mellem resultaterne fra en statistisk model sammenlignet med, hvis man kigger på forskelle i gennemsnit mellem forskellige kategorier, som f.eks. naturgasfyrede forsyninger i forhold til kulfyrede forsyninger. Grunden til at den statistiske models resultater sammenholdes med de gennemsnitlige forskelle mellem forskellige kategorier er at hver forsyning vægtes ligeligt i det simple gennemsnit, hvilket vil sige at informationerne om sammenhængen mellem f.eks. brændselskøb og fjernvarmepris behandles ligeligt på tværs af fjernvarmeforsyninger.

Hvis der alternativt var benyttet vægtede gennemsnitspriser ville informationerne om de forhold, der påvirker de store fjernvarmeforsyningers priser, dominere billedet så det næsten udelukkende er en lille gruppe af meget store fjernvarmeforsyninger, der undersøges.

I Tabel 1 nedenfor viser kolonne 2 de uvægtede gennemsnitspriser for forsyninger med forskellige typer af ejerskab. I 3. kolonne fremgår forskellene i gennemsnitspris mellem den pågældende ejerskabstype og de forbrugerejede forsyninger, som er referencen for ejerskab. I 4. kolonne fremgår de forskelle i fjernvarmepriser, som er beregnet ved den statistiske model (regressionsanalyse).

Tabel 1: Hvorfor benytte en statistisk model?

	Uvægtet gns. pris	Forskel til reference	Statistisk model
Ejerskab			
Forbrugerejet (Reference)	16.398		
Kommunalt	14.569	-1.829	1.167
Kommercielt	21.505	5.108	5.719

Hvis der ses på forskellen i gennemsnitspriser, i 3. kolonne, fremgår det at kommunalt ejede forsyninger i gennemsnit er godt 1.800 kr. billigere end forbrugerejede forsyninger, og kommercielt ejede forsyninger i gennemsnit er godt 5.100 kr. dyrere end forbrugerejede forsyninger.

Sammenlignes forskellene i gennemsnitspriser med de forskelle der er beregnet i den statistiske model (kolonne 4), er betydningen af at være kommunalt nu ikke længere, at kommunale forsyninger har lavere priser, men derimod at de er små 1.200 kr. dyrere end forbrugerejede forsyninger. Forskellen skyldes at den statistiske model tager højde for betydningen af forsyningernes størrelse, primære brændselstype, og bymæssighed. Når der er taget højde for disse variable er forbrugerejede forsyninger faktisk billigere end kommunale forsyninger.

Betydningen af at være kommercielt ejet ændres stort set ikke, når der tages højde for variablene størrelse, primært brændsel og bymæssighed. De kommercielle forsyninger er betydelig dyrere end forbrugerejede forsyninger, uanset hvilken metode der benyttes.

Hvis der benyttes en statistisk model frem for at sammenligne gennemsnitlige varmepriser for forskellige grupper af fjernvarmeforsyninger, nås der frem til vidt forskellige konklusioner på betydningen af at en forsyning er kommunalt ejet. Det illustrerer, hvorfor det er vigtigt at isolere effekten af de forhold, der påvirker varmepriserne, inden det konkluderes hvilken betydning et givent forhold har for varmeprisen.

Den statistiske model, der benyttes i analysen af forskelle i varmepriser er en såkaldt regressionsmodel.

Regressionsanalysens styrker og begrænsninger

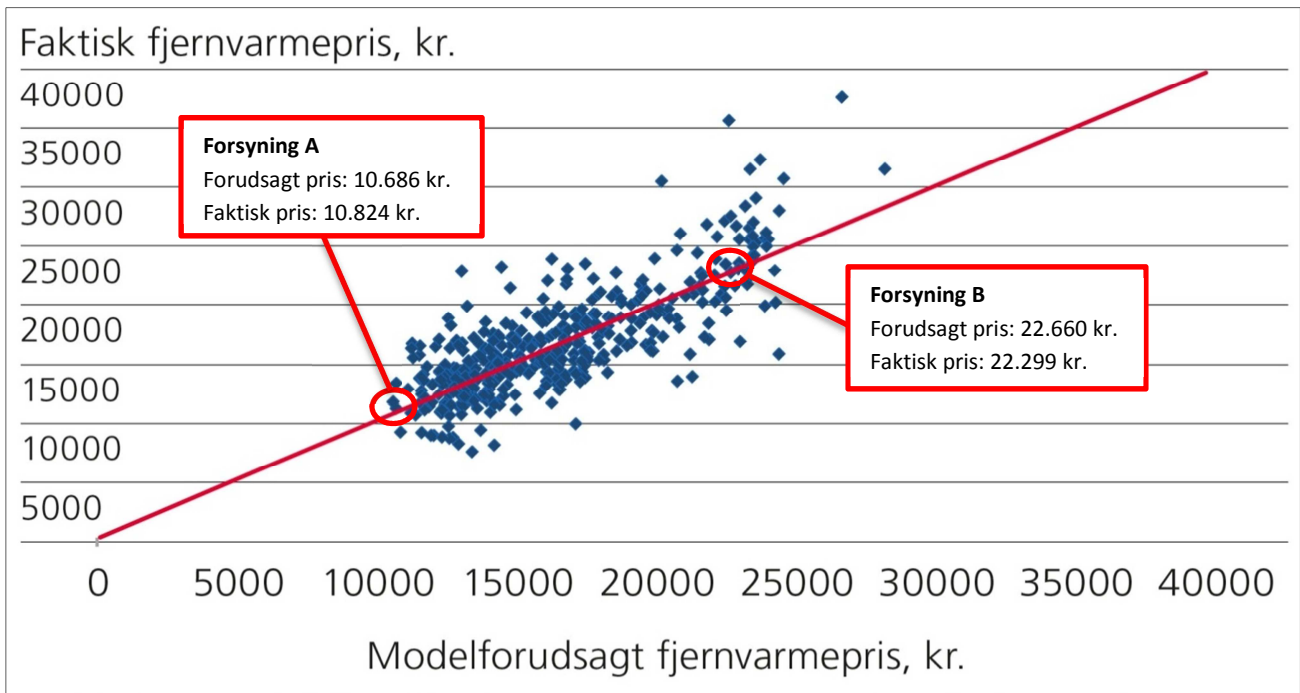
Analysen af forskelle i varmepriser er baseret på en såkaldt regressionsmodel. Analysen af varmepriser baseret på regressionsmodeller har flere fordele i forhold til analyser baseret på simple gennemsnit. Regressionsmodeller er et meget stærkt, anerkendt og anvendt analyseredskab til at belyse komplekse økonomiske sammenhænge. Udover at regressionsmodeller kan isolere effekten af de variable og kategorier, der indgår i modellen, har metoden også den store fordel, at regressionsmodeller er fleksible og derfor nemt kan udvides med nye variable. Efterhånden som datagrundlaget forbedres eller nye variable identificeres, kan regressionsmodellen bruges til at teste om nye variable skal inkluderes i modellen og hvordan det påvirker resultaterne.

Selv om den regressionsmodel, der anvendes i analysen af fjernvarmepriserne, kan forklare en stor del (ca. 2/3) af prisforskellene er modellen ikke perfekt eller endegyldig. Der er fortsat flere variable, der har betydning for fjernvarmepriserne, som ikke er inkluderet i analysen på grund af manglende eller dårlige data. Derfor skal resultaterne af analysen også tolkes med forbehold for, at fjernvarmeområdet fortsat ikke er gennemanalyseret og at kendte såvel som nye variable muligvis kan forklare endnu mere af prisforskellene mellem fjernvarmeforsyninger i Danmark.

Beregning af den model-forudsagte fjernvarmepris

I dette afsnit vises det, hvordan den af modellen forudsagte fjernvarmepris beregnes. I analysen tages der udgangspunkt i en reference-forsyning, ud fra hvilken de forskellige variables betydning beregnes som forskellen til referencen. I figur 1 nedenfor, er der udvalgt to konkrete forsyninger, kaldet forsyning A og forsyning B.

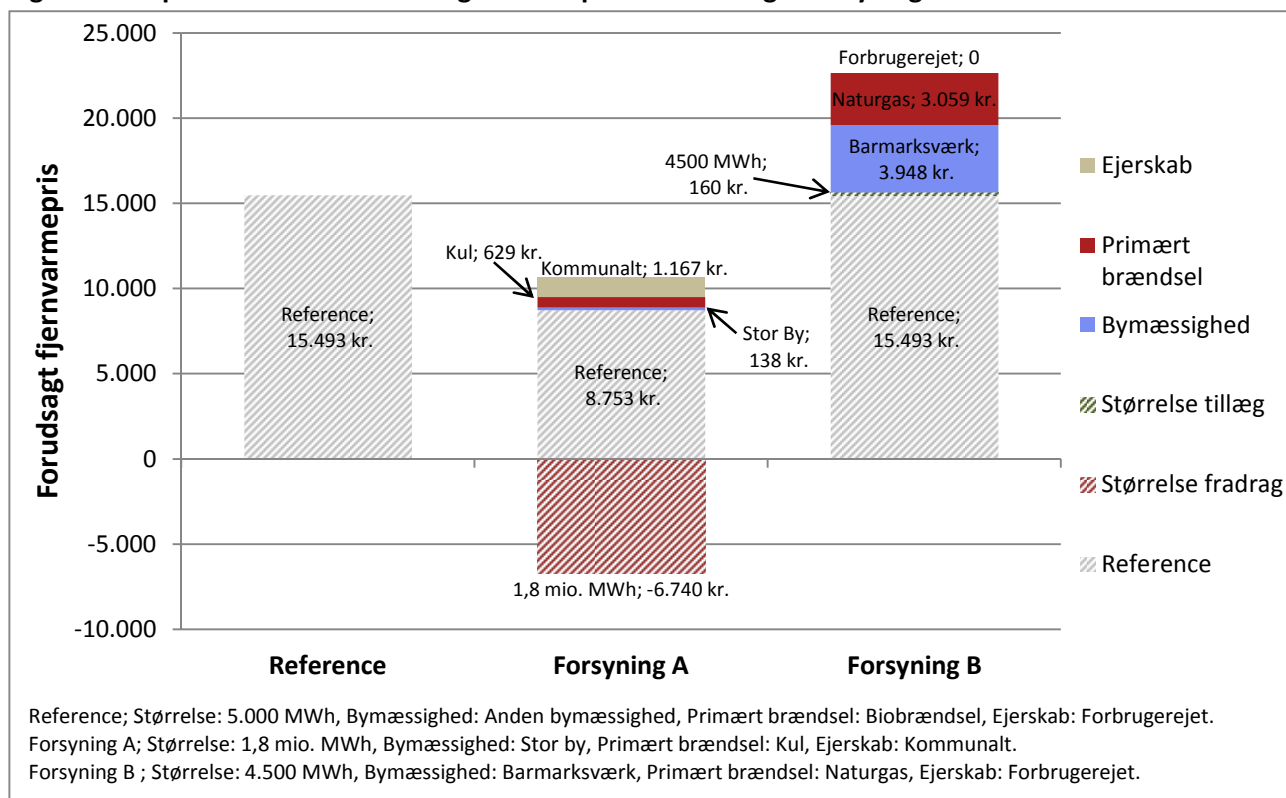
Figur 1: Faktiske priser sammenholdt med de forudsagte priser i analysemodellen



De to forsyninger er udvalgt, fordi modellens forudsagte pris næsten er identisk med den faktiske pris, og fordi de har meget forskellige priser og karakteristika, og derfor egner sig godt til at illustrere, hvordan den statistiske model kan forklare de store forskelle i varmepriser, vi ser i fjernvarmesektoren.

I det følgende vil beregningen af de to forsyningers model-forudsagte varmepriser i forhold til referenceforsyningen blive gennemgået og forklaret. Figur 2 viser dekomponerer og sammenligner den statistiske models forudsagte priser, for de to udvalgte forsyninger A og B.

Figur 2: Komponenter af den forudsagte varmepris for 2 udvalgte forsyninger



Reference-forsyningen

Reference-forsyningen er en forsyning, med et budgetteret årssalg på 5.000 MWh, som hverken kategoriseres som et barmarksværk eller ligger i en stor by. Reference-forsyningen sælger varme, der primært er produceret på biobrændsel, og forsyningen er ejet af forbrugerne. Den statistiske model forudsiger, at en forsyning med disse karakteristika har en varmepris på ca. 15.500 kr. pr. år for at opvarme et enfamiliehus på 130 kvm.

Størrelse

En fjernvarmeforsynings størrelse, målt efter budgetteret årssalg, er en af de variable, som kan forklare en del af forskellene i fjernvarmepriser. Ifølge den statistiske model er fjernvarmeforsyninger med et stort årssalg billigere end reference-forsyningen, som har et budgetteret årssalg på 5.000 MWh.

Fjernvarmeforsyninger, der budgetterer med et årssalg på under 5.000 MWh, vil have et tillæg til varmeprisen i forhold til reference-forsyningen, fordi det ikke kan opnå de samme stordriftsfordele.

Forsyning A (stor forsyning) har et fradrag på 6.740 kr. i forhold til reference-forsyningen, på grund af dets størrelse (budgetteret årssalg på ca. 1,8 mio. MWh). Fradraget fremgår som en negativ værdi i figur 2, og er markeret med en skraveret rød farve.

Forsyning B (lille forsyning) har et tillæg på 160 kr., fordi det er marginalt mindre end reference-forsyningen (budgetteret årssalg på ca. 4.500 MWh). Tillægget er markeret med skraveret grøn farve i figuren.

Bymæssighed

Reference-forsyningen har 'anden bymæssighed'. Derfor er tillæggene for 'stor by' og 'barmarksværk'

beregnet i forhold til referencen, 'anden bymæssighed'. Tillæg for fjernvarmeforsyningers bymæssighed vises med blå farve i figur 2 ovenfor.

Der lægges 138 kr. til forsyning A's forudsagte varmepris, fordi det ligger i en stor by.

Der lægges 3.948 kr. til Forsyning B's forudsagte pris, fordi det er et barmarksværk.

Primært brændsel

Reference-forsyningens varme er primært produceret på biobrændsel. Tillæggene for at benytte andre brændselstyper end biobrændsel er derfor beregnet i forhold til en forsyning, hvor varmen produceres på biobrændsel. Tillæg på baggrund af forsyningernes primære brændselskilde er markeret med rød farve i figur 2 ovenfor.

Der lægges 629 kr. til forsyning A's forudsagte pris, fordi det fyrer med kul frem for biobrændsel.

Der lægges 3.059 kr. til forsyning B's forudsagte pris, fordi det fyrer med naturgas frem for biobrændsel.

Ejerskab

Referencen-forsyningen er forbrugerejet, og derfor er tillæggene til de forudsagte varmepriser beregnet, som forskellen til en tilsvarende forbrugerejet forsyning. Tillæg til varmeprisen, som følge af forsyningernes ejerskabstype fremgår med beige farve i figur 2 ovenfor.

Der lægges 1.167 kr. til forsyning A's pris, fordi det er kommunalt ejet.

Der lægges ikke noget til forsyning B's pris, fordi det er forbrugerejet og derfor på dette punkt er ens med referenceforsyningen.

Samlede forudsagte varmepriser

I tabellen nedenfor opsummerer de samlede model-forudsagte fjernvarmepriser, og sammenholder dem med de to forsyningers faktiske varmepriser.

Tablet 2: Model-forudsagte varmepriser

	Reference-forsyning	Størrelse	Bymæssighed	Primært brændsel	Ejerskab	Samlet forudsagt pris	Faktisk pris
Forsyning A	15.493	-6.740	138	629	1.167	10.687	10.824
Forsyning B	15.493	160	3.948	3.059	0	22.660	22.299

Eksemplet viser, hvordan de to forsyningers varmepriser forudsiges i den statistiske model og illustrerer samtidig, at den statistiske model er fleksibel og kan forudsige både relativt lave priser, som mange forsyninger i Danmark har, og samtidig kan forudsige de meget høje priser, som også findes.

Datagrundlag

Analysens primære datakilde er prisanmeldelser til Energitilsynet fra 2011/2012, der var gældende i august 2012. Fra anmeldelserne er hentet oplysninger om:

- Fjernvarmepriser
- Budgetteret årssalg

- Primært brændsel

Variablen 'bymæssighed' er lavet på baggrund af Sekretariatets inddeling af fjernvarmeforsyningerne, på baggrund af om de ligger i en af Danmarks 20 største byer, eller i hovedstadsområdet, samt Energistyrelsens liste over barmarksværker i Danmark.

Sekretariatet har inddelt forsyningerne efter ejerskabstype. Selskaberne er inddelt i grupperne kommunalt, forbrugerejet, kommercielt og boligforening, på baggrund af hvem der har bestemmende indflydelse i forsyningerne.

Signifikans af modellens parametre

I tabellen nedenfor vises resultaterne af regressionsanalysen fra artiklen "Store forskelle i varmepriserne – hvorfor?" med flere tekniske detaljer. Kolonnen koefficienter indeholder de estimerede koefficienter for modellen. Grunden til at budgetteret årssalg ikke fremgår som forskelle mellem en referencegruppe er, at variabelen statistisk set er inkluderet som en såkaldt kontinuert variabel, hvorimod de øvrige variable er dummy variable (binære variable). Det budgetterede årssalg indgår på følgende vis:

$$Påvirkning_{størrelse} = 36.091 \times Forventet \text{ årssalg}^{-0,11}$$

Koefficienterne på 36.091 og -0,11 er estimeret.

Denne specifikation er valgt, fordi det er den specifikation, der findes bedst til at beskrive sammenhængen mellem budgetteret årssalg og forsyningernes fjernvarmepriser. Hertil kommer, at den funktionelle form i øvrigt er anvendelig til at beskrive forekomsten af eventuelle skala/stordriftsfordele. Modellen har en justeret forklaringsgrad på 62 pct., hvilket er relativt godt, for denne type statistiske modeller.

Tablet 3: Resultater af regressionsanalysen

	Koefficienter	Nedre 95%	Øvre 95%	Standardfejl	t-stat	P-værdi	Antal observationer
Konstant	1.351	-1.167	3.870	1.281	1,05	0,29	
Ejerskab (Reference: Forbrugerejet)							354
Kommunalt	1.167	249	2.084	467	2,50	0,01	53
Kommercielt	5.719	4.265	7.173	740	7,73	0,00	16
Boligforening	1.130	-1.226	3.486	1.199	0,94	0,35	6
Andet ejerskab	4.161	233	8.089	1.998	2,08	0,04	2
Brændselstype (Reference: Biobrændsel)							144
Kul	629	-830	2.087	742	0,85	0,40	19
Affald	607	-764	1.977	697	0,87	0,38	20
Naturgas	3.059	2.433	3.685	318	9,61	0,00	181
Andet brændsel	962	83	1.842	448	2,15	0,03	67
Bymæssighed (Reference: Anden bymæssighed)							303
Barmarksværk	3.948	3.161	4.734	400	9,86	0,00	87
Stor by	138	-972	1.247	564	0,24	0,81	41
Størrelse							
Forventet årssalg	36.091	28.836	43.345	3.691	9,78	0,00	431

Kilde: Regressionen bygger på 431 observationer fra danske fjernvarmeforsyninger i 2011/12. Regressionsmodellen er nærmere forklaret i et uddybende baggrundsnotat på Energitilsynets hjemmeside.

I kolonnerne "Nedre 95%" og "Øvre 95%" fremgår den øvre og nedre grænse for konfidensintervallet omkring den estimerede koefficient. Spændet mellem den øvre og den nedre grænse skal forstås som et

spænd for den statistiske usikkerhed omkring de estimerede koefficienter. Med andre ord er effekten af at være kommunalt ejet estimeret til 1.167 kr., men estimatet er usikkert. Den øvre og nedre grænse angiver at koefficienten for "kommunalt" med 95 % sandsynlighed ligger mellem 249 kr. og 2084 kr.

Kolonnerne "standardfejl", "t-stat" og "P-værdi" fortæller, hvor statistisk signifikant hver kategori er. En p-værdi på under 0,05 betragtes sædvanligvis som statistisk signifikant.

I kolonnen "observationer" fremgår, hvor mange fjernvarmeforsyninger, der findes i hver kategori. Jo flere observationer, der er i en kategori, desto bedre er grundlaget for at fastslå om en forskel er statistisk signifikant. Hvis sammenhængen er stærk, kan kategorier med meget få observationer være signifikante.

Gruppen 'boligforening' har kun 6 observationer, hvilket skyldes at gruppen kun omfatter boligforeninger, med salg til husholdninger, der ikke er medlemmer af boligforeningen. Dvs. husstande udenfor boligforeningens område. Det har ikke betydning for de i analysen estimerede koefficienter, eller modellens forklaringsgrad, hvis boligforeninger fjernes fra analysen.

Kommercielt drevne fjernvarmeforsyninger er signifikant dyrere end andre forsyninger selv om forskellen er estimeret på grundlag af blot 16 observationer. Til trods for det relativt spinkle datagrundlag er effekten af at være kommercielt ifølge regressionsmodellen med 95 % sikkerhed mellem 4.265 og 7.173, og er altså både betydelig og signifikant på et 1 % -niveau.

Referencens betydning for kategoriernes signifikans

Det er ikke alle parametrene i modellen, der er signifikante. Det skyldes imidlertid den måde modellen er specificeret med dummy-kategorier. Når der anvendes referencer i modellen, vil koefficienterne være et estimat på forskellen i kr. mellem referencen og den givne variabel. Det vil sige at koefficienten for "kommunalt" på 1.167 angiver den estimerede prisforskel mellem en kommunal forsyning og en forbrugerejet forsyning (referencen). Benyttes en anden reference. som f.eks. "andet ejerskab", er det ikke sikkert at kategorien "kommunalt" ville være signifikant, fordi forskellen til referencen vil være en anden.

Man kunne altså stille spørgsmål ved, om modellen og analysens resultater er følsomme overfor valget af referencer. For at illustrere hvordan koefficienterne og signifikansen af parametrene ændrer sig, når referencen ændres, sammenlignes resultaterne af en regression, hvor naturgas benyttes som reference for primært brændsel med den regression, der benyttes i analysen, hvor referencen er biobrændsel.

Table 4: Reference significance for categories.

Regression statistics				
	Reference: Biomass		Reference: Natural Gas	
Multipel R	0,79		0,79	
R-squared	0,63		0,63	
Adjusted R-squared	0,62		0,62	
Standard error	2.803		2.803	
Observations	431		431	
Coefficients and P-values				
	Coefficient	P-value	Coefficient	P-value
Intercept	1.351	0,29	4.411	0,00
Ownership				
Municipal	1.167	0,01	1.167	0,01
Commercial	5.719	0,00	5.719	0,00
Apartment association	1.130	0,35	1.130	0,35
Other ownership	4.161	0,04	4.161	0,04
Primary fuel				
Coal	629	0,40	-2.431	0,00
Waste	607	0,38	-2.453	0,00
Biomass	3.059	0,00	-3.059	0,00
Other fuel	962	0,03	-2.097	0,00
Urbanization				
Rural area	3.948	0,00	3.948	0,00
Urbanization	138	0,81	138	0,81
Size				
Expected sales	36.091	0,00	36.091	0,00

From the table it appears that the explanatory power and standard error are identical, just as the coefficients for the variables ownership, urbanization and size are also unchanged. The constant is affected by the change in reference. The difference in constants is precisely the estimated difference between using natural gas instead of biomass (3.059).

The coefficients and p-values for primary fuel change significantly when the fuel reference is changed. When natural gas is used as reference, all categories in the variable primary fuel are significant – also those that were not significant when biomass was used as reference. The results underline that the choice of reference does not have any effect on the model's validity and explanatory power, and therefore does not affect the analysis's conclusions.

Econometric theory suggests that variables and categories that are not significant should be removed or combined. When biomass is used as reference for fuel type, it would mean that the categories "coal" and "natural gas" should be combined, as the cost difference between using these fuels is not significant compared to using biomass. The categories are, however, not combined in this analysis, as the data base is large enough to estimate the parameters. The difference between R^2 and adjusted R^2 shows that it is not too problematic to include several explanatory categories. The difference is not too large.

hvorfor der ikke er et pres for at slå kategorier sammen. Der er desuden tekniske forskelle på forsyninger med forskelligt brændsel, hvilket taler imod at slå biobrændsel sammen med affald og kul selvom analysen viser, at der ikke er signifikant forskel på prispåvirkningerne.

Heteroskedasticitet

Heteroskedasticitet, eller uens varians i fejledene, er et velkendt fænomen i regressionsanalyser. Tilstedeværelsen af heteroskedasticitet betyder sædvanligvis ikke, at de estimerede koefficienter er misvisende (biased), men påvirker signifikansniveauet af modellens parametre. Det vil sige, at koefficienterne fortsat er pålidelige, men at tilstedeværelsen af heteroskedasticitet nødvendiggør en justering af de t-værdier, der benyttes til at fastslå om en variabel er signifikant.

I tabel 5 nedenfor vises resultatet af to statistiske test for heteroskedasticitet, som begge peger på, at der er heteroskedasticitet til stede i regressionsmodellen.

Tabel 5: Heteroskedasticitet i regressionsmodellen.

Test for Heteroskedasticitet			
	F-test	Test-værdi	P-værdi
Hetero test:	F(12,418)	56,9	[0.0000]**
Hetero-X test:	F(12,418)	56,9	[0.0000]**

For at kunne konkludere, om tilstedeværelsen af heteroskedasticitet påvirker analysens konklusioner, er såkaldte heteroskedasticitetskonsistente standard errors¹ (HCSE) sammenlignet med ukorrigerede standard errors, som ikke tager højde for heteroskedasticitet. Sammenligningen fremgår af tabel 6.

Tabel 6: Heteroskedasticitets betydning for modellens resultater.

Heteroskedasticitet							
	Koefficienter	SE	t-SE	t-prob	HCSE	t-HCSE	t-HCSE prob
Konstant	1.351	1.281	1,05	0,29	1.780	0,76	0,45
Ejerskab							
Kommunalt	1.167	467	2,50	0,01	512	2,28	0,02
Kommercielt	5.719	740	7,73	0,00	1.380	4,14	0,00
Boligforening	1.130	1.199	0,94	0,35	639	1,77	0,08
Andet ejerskab	4.161	1.998	2,08	0,04	2.481	1,68	0,09
Primært brændsel							
Kul	629	742	0,85	0,40	736	0,85	0,39
Affald	607	697	0,87	0,38	823	0,74	0,46
Naturgas	3.059	318	9,61	0,00	303	10,08	0,00
Andet brændsel	962	448	2,15	0,03	440	2,19	0,03
Bymæssighed							
Barmarksværk	3.948	400	9,86	0,00	473	8,35	0,00
Stor by	138	564	0,24	0,81	669	0,21	0,84
Størrelse							
Forventet årssalg	36.091	3.691	9,78	0,00	5.209	6,93	0,00

¹ Der er benyttet White's standard errors

Hvis værdien i kolonnen "t-prob" er betydelig forskellig fra værdien i kolonnen "t-HCSE prob", skyldes det at heteroskedasticitet påvirker pålideligheden af de ukorrigerede standard errors. Det fremgår af tabellen at p-værdien for "boligforening" falder fra 0,35 til 0,08, når der tages højde for heteroskedasticitet. kategorien bliver altså mere statistisk signifikant, men er stadig ikke inden for grænsen på 0,05, og betragtes derfor fortsat ikke som statistisk signifikant. Kategorien "andet ejerskab" ændres fra at være signifikant (0,04), til at være svagt insignifikant (0,09), men det har ingen betydning for analysens resultater, da denne rest-gruppe bestående af to værker ikke indgår i selve analysens resultater.

På baggrund af ovenstående overvejelser vurderes det at tilstedeværelsen af heteroskedasticitet i modellen ikke påvirker analysens konklusioner.

Multikollinearitet

Ud fra et almindeligt kendskab til fjernvarmebranchen kunne man forvente at flere af variablene i modellen er højt korrelerede, hvilket kunne give en forventning om en betydelig grad af multikollinearitet.

Multikollinearitet opstår, hvis to variable i høj grad bevæger sig i samme retning, hvilket bevirker at statistiske modeller ikke kan separere betydning af den ene forklarende variabel fra den anden.

Hvis den statistiske model ikke klart kan identificere hvilken af to forklarende variable, der påvirker den forklarede variabel, vil selv små ændringer i modellen eller én af de to variable kunne medføre store ændringer i de estimerede koefficienter. Det betyder at højt korrelerede variable i modellen vil have meget høje standardafvigelse, og at det som resultat heraf vil være svært at afgøre om en variabel er signifikant. En høj grad af multikollinearitet kan medføre at de estimerede koefficienter er misvisende (biased).

Sædvanligvis betragtes en korrelationskoefficient med en numerisk værdi på 0,8 eller højere som problematisk i forbindelse med regressionsanalyse. Korrelationskoefficienterne får den statistiske models forklarende variable og kategorier fremgår af tabel 7 nedenfor.

Tabel 7: Korrelationsmatrice

	Forbrugerejet	Kommunalt	Kommercielt	Boligforening	Andet ejerskab	Biobrændsel	Kul	Affald	Naturgas	Andet brændsel	Stor by	Anden bymæssighed	Barmarksværk	Årssalg
Forbrugerejet		-0,80	-0,42	-0,25	-0,15	0,11	-0,25	-0,13	0,08	-0,03	-0,34	0,17	0,07	0,28
Kommunalt			-0,07	-0,04	-0,03	-0,07	0,19	0,19	-0,12	0,03	0,36	-0,14	-0,11	-0,36
Kommercielt				-0,02	-0,01	-0,04	0,20	-0,04	0,01	-0,05	-0,02	-0,07	0,07	0,03
Boligforening					-0,01	-0,08	-0,03	-0,03	0,02	0,11	0,16	-0,06	-0,05	0,07
Andet ejerskab						-0,05	-0,01	-0,02	0,08	-0,03	-0,02	-0,03	0,04	-0,03
Biobrændsel							-0,15	-0,16	-0,60	-0,30	-0,20	-0,05	0,17	0,07
Kul								-0,05	-0,18	-0,09	0,12	-0,11	0,02	-0,28
Affald									-0,19	-0,09	0,15	-0,11	0,00	-0,20
Naturgas										-0,37	-0,18	0,20	-0,06	0,26
Andet brændsel											0,34	-0,09	-0,14	-0,17
Stor by												-0,16	-0,50	-0,40
Barmarksværk													-0,77	0,51
Anden bymæssighed														-0,19
Årssalg														

Generelt er korrelationen mellem de forklarende kategorier lav, men korrelationen mellem at være forbrugerejet og kommunal, samt at være barmarksværk og have anden bymæssighed ligger tæt på den kritiske værdi på 0,8. Det er ikke overraskende at netop disse kategorier er højt korrelerede.

”Forbrugerejet” og ”kommunalt” er begge såkaldte dummy-kategorier i variabelen ’ejerskab’, som opdeler datasættet i fire grupper. Derfor er det uundgåeligt at disse dummy-kategorier vil være korreleret med hinanden. Havde man opdelt fjernvarmeforsyningerne efter om de var kraftvarmeverker (el- og varmeproduktion) eller varmeverker (kun varmeproduktion), ville korrelationen mellem de to kategorier være perfekt negativ (-1). Opdeling af datasættet med dummy-kategorier er en almindelig anerkendt fremgangsmåde inden for økonometri, og derfor er høj korrelation mellem kategorier inden for samme variabel, som f.eks. bymæssighed eller ejerskab ikke problematisk for modellens resultater.

I tabel 8 nedenfor er korrelationer mellem kategorier inden for samme variabel fjernet, for at fremhæve de korrelationskoefficienter, der er relevante for vurderingen af graden af multikollinearitet i modellen.

Tabel 8: Korrelationsmatrice (uden intra-variabel korrelation)

	Forbrugerejet	Kommunalt	Kommercielt	Boligforening	Andet ejerskab	Biobrændsel	Kul	Affald	Naturgas	Andet brændsel	Stor by	Barmarksværk	Anden bymæssighed	Årssalg
Forbrugerejet														
Kommunalt														
Kommercielt														
Boligforening														
Andet ejerskab														
Biobrændsel														
Kul														
Affald														
Naturgas														
Andet brændsel														
Stor by														
Barmarksværk														
Anden bymæssighed														
Årssalg														

Den største absolutte korrelationskoefficient er på 0,51, hvilket vurderes ikke at give anledning til en problematisk grad af multikollinearitet i modellen.

Hovedparten af barmarksværkerne anvender naturgas som deres primære brændsel. Derfor kan det umiddelbart undre at korrelationen mellem kategorierne barmarksværk og naturgas ikke er højere end 0,2. Årsagen til den relativt lave korrelation mellem kategorierne er at "kun" 55 af 87 barmarksværker er naturgasfyrede, samtidig er hen ved 180 af 431 fjernvarmeforsyninger naturgasfyrede. På baggrund af de ca. 125 naturgasfyrede fjernvarmeforsyninger, som ikke er barmarksværker, og de 32 barmarksværker, der ikke er naturgasfyrede, er der statistisk grundlag for at isolere effekten af at anvende naturgas fra effekten på prisen af at være et barmarksværk.

På samme vis kunne der argumenteres for, at barmarksværker generelt er små, og at det derfor ikke vil være muligt at isolere effekten af at være lille, fra effekten af at være et barmarksværk med lav forbrugertæthed. Der er imidlertid 239 små fjernvarmeforsyninger med et budgetteret årssalg på under 20.000 MWh i datasættet, hvoraf kun de 87 er barmarksværker. Derfor er der tilstrækkelig mange små fjernvarmeforsyninger, som ikke er barmarksværker til, at effekten af at være en lille fjernvarmeforsyning kan estimeres uden at det giver betydelige problemer med multikollinearitet.

På baggrund af ovenstående overvejelser, og de relativt lave standardafvigelser i modellen, anses graden af multikollinearitet ikke som problematisk for resultaterne i den statistiske model.

Forbrugertæthed

Barmarksværk er i analysen anvendt som betegnelse for fjernvarmeforsyninger, med langt mellem forbrugerne. Kategorien 'stor by' er anvendt som betegnelse for områder med meget høj forbrugertæthed. Der kunne argumenteres for, at der i stedet burde være anvendt et mål som forbrugertæthed frem for

inddelingen i 'barmarksværk', 'stor by' og 'anden bymæssighed', men data for forbrugertæthed er endnu ikke af samme kvalitet, som for f.eks. budgetteret årssalg og primær brændselstype og det har derfor ikke været muligt at inkludere variabelen på nuværende tidspunkt.

For at teste om fjernvarmeforsyningernes forbrugertæthed kan forklare prisforskellene mellem fjernvarmeforsyninger, undersøges variabelens evne til at forklare fjernvarmeprisen for en delgruppe af forsyninger på 253, hvor data for forbrugertætheden er tilgængelig om end data ikke er grundigt valideret.

I undersøgelsen anvendes det samme mål for forbrugertæthed, som Dansk Fjernvarme benytter i deres publikation: "Udviklingstendenser i dansk fjernvarme" (Årsstatistik 2011). Målet er defineret på følgende måde:

$$\text{Forbrugertæthed} = \frac{\text{leveret MWh varme}}{\text{meter hovedledning} + \text{meter stikledning}}$$

Denne definition på forbrugertæthed er blevet undersøgt i en model med data for 253 forsyninger, hvor variabelen bymæssighed ikke indgår, kaldet "Regression med forbrugertæthed" (midterste kolonne). Derudover er variabelens betydning også undersøgt i en model, hvor både forbrugertæthed og bymæssighed er inkluderet, kaldet "Regression med alle variable", i kolonnen yderst til højre.

I den første kolonne af tabel 9, indgår forbrugertætheden ikke. Denne kolonne er tilføjet, dels som sammenligningsgrundlag for modelspecifikationerne med forbrugertæthed, dels for at vise, at analysens resultater ikke påvirkes betydeligt af at 178 observationer ekskluderes.

Tabel 9: Resultater af regressioner med forbrugertæthed

	Standard regression		Regression med forbrugertæthed		Regression alle variable.	
R ²	0,60		0,53		0,60	
Justeret R ²	0,58		0,51		0,59	
Observationer	253		253		253	
	Koefficienter	P-værdi	Koefficienter	P-værdi	Koefficienter	P-værdi
Konstant	786	0,66	-4.556	0,00	916	0,60
Ejerskab (Reference: Forbrugerejet)						
Kommunalt	1.394	0,01	1.546	0,01	1.344	0,01
Kommercielt	2.841	0,01	3.092	0,02	3.775	0,00
Boligforening	1.584	0,29	-5	1,00	1.548	0,30
Andet ejerskab						
Brændselstype (Reference: Biobrændsel)						
Kul	980	0,31	2.290	0,03	1.371	0,17
Affald	1.489	0,05	2.256	0,01	1.662	0,03
Naturgas	2.692	0,00	2.916	0,00	2.714	0,00
Andet brændsel	1.052	0,04	1.463	0,01	1.122	0,03
Bymæssighed (Reference: Anden bymæssighed)						
Barmarksværk	3.607	0,00			3.576	0,00
Stor by	-283	0,68			19	0,98
Størrelse						
Forventet årsalg	38.088	0,00	55.491	0,00	38.004	0,00
Forbrugertæthed						
MWh leveret pr. meter			-214	0,12	-218	0,10

Regressionerne har en forklaringsgrad på mellem 53 og 60 %, hvilket er en smule lavere end i analysen med 431 observationer (63 %).

Det er bemærkelsesværdigt at målet for forbrugertæthed ikke har signifikant forklaringsevne (P-værdi: 0,12), og at koefficienten for barmarksværk ikke ændres betydeligt ved inklusion af forbrugertæthed (fra 3.607 til 3.576). Generelt ændres koefficienterne for bymæssighed ikke ved inklusion af forbrugertæthed. Resultaterne er overraskende, men det kan ikke på baggrund af undersøgelsen konkluderes at forbrugertæthed ikke er korreleret med variabelen 'bymæssighed', da data ikke er tilstrækkelig valideret. Af samme årsag kan det ikke konkluderes at forbrugertæthed ikke har betydning for fjernvarmeprisen, selvom resultaterne peger på det. Det vil være nødvendigt med betydeligt datavalidering og test af andre specifikationer for forbrugertæthed, før der kan drages endelige konklusioner om variabelens betydning.

Resultaterne af regressionen med 253 observationer stemmer generelt overens med de resultater, der fandtes i analysen med 431 observationer. Det underbygger at resultaterne af analysen er robuste overfor ændringer i datagrundlaget. Koefficienten for "kommercielt" er faldet betydeligt, men er stadig væsentlig højere end de øvrige ejerskabsformer. Det kan tilskrives at seks af de dyreste kommercielle forsyninger ikke indgår blandt de 253 observationer.