

## Boligmarkedet i Kristiansandsregionen

Hvilken effekt har avstanden fra sentrum, målt i reisetid i rushtrafikk, på boligprisene?

**Lars Gunnar Abusdal**

**Veileder**

Arngrim Hunnes

*Masteroppgaven er gjennomført som ledd i utdanningen ved Universitetet i Agder og er godkjent som del av denne utdanningen. Denne godkjenningen innebærer ikke at universitetet inntår for de metoder som er anvendt og de konklusjoner som er trukket.*

Universitetet i Agder, 2013

Fakultet for Økonomi og samfunnsvitenskap

Institutt for økonomi/Handelshøyskolen i Kristiansand

# Forord

Denne oppgaven er en obligatorisk del av masterstudiet i økonomi og administrasjon ved Universitet i Agder. Oppgaven skrives som en avsluttende del av det femårige studieforløpet.

Bakgrunnen for valget om å skrive oppgaven innenfor fagområdet eiendomsøkonomi, er at det ble fattet stor interesse for dette fagfeltet da jeg ble introdusert for det gjennom valgfag høsten 2012. Samtidig markerer også masteroppgaven slutten på studietiden, og for mange studenter, undertegnede inkludert, er det nå tid for å bevege seg inn på både arbeids- og boligmarkedet. Av denne årsak er studier av boligmarkedet i Kristiansandsregionen et høyaktuelt tema.

Jeg vil takke min veileder, Arngrim Hunnes, for god veiledning og konstruktive tilbakemeldinger underveis i skriveprosessen. Samtidig vil jeg også takke Monika og Kasper for deres tålmodighet og støtte, som har gjort det mulig å fullføre denne oppgaven i en hektisk periode av våre liv.

Lars Gunnar Abusdal

Kristiansand 1. juni 2013

# Sammendrag

Etablert teori innenfor eiendomsøkonomi tilsier at verdien til en ellers lik bolig vil falle dersom avstanden til sentrum øker. Tidligere masteroppgaver har analysert boligmarkedet i Kristiansand, og funnet en slik sammenheng mellom boligprisene og antall kilometer kjøredistanse mellom boligene og Kristiansand sentrum. Denne oppgaven anerkjenner imidlertid at det finnes flere måter å beregne sentrumsavstand på. Veinettets utforming og standard i Kristiansand medfører at det tar relativt mye lengre tid å bevege seg langs veinettet i rushtrafikk på vestsiden av byen enn på østsiden. Denne oppgaven undersøker derfor i hvilken grad en boligs avstand fra sentrum, målt i reisetid i rushtrafikk, slår inn på boligprisene. Modeller med avstanden målt i reisetid blir også sammenlignet med modeller som beregner avstanden på den tradisjonelle måten.

En modell med lineær funksjonsform estimerer at verdien til en ellers lik bolig faller med 45 350 kroner dersom sentrumsavstanden økes med ett minutt. Mens en modell med kvadratisk funksjonsform anslår at effekten av økt sentrumsavstand er større nærme sentrum enn den er ute i periferien. Denne modellen estimerer den negative priseffekten av økt sentrumsavstand til å være 70 820 kroner per minutt for boliger i sentrum, men at den negative effekten av økt sentrumsavstand avtar med 1219 kroner for hvert minutt man beveger seg bort fra sentrum. For modellene med denne funksjonsformen indikerer analysen at avstanden målt i reisetid fremfor kilometer gir en mer treffsikker modell. Vår modell med dobbeltlogaritmisk funksjonsform anslår at 1 % økning i sentrumsavstand målt i reisetid, medfører en reduksjon i boligens verdi på 0,188 %. For modeller med denne funksjonsformen er det imidlertid liten forskjell på om avstanden måles i kilometer eller reisetid.

# Innhold

Forord .....	i
Sammendrag .....	ii
Oversikt over figurer .....	v
Oversikt over tabeller .....	vi
Oversikt over vedlegg .....	vii
1. Innledning .....	1
1.1 Bakgrunn .....	1
1.2 Problemstilling .....	1
1.3 Oppgavens oppbygging .....	2
2. Oversikt over Kristiansandsregionen .....	4
2.1 Kristiansand .....	4
2.2 Kristiansandsregionen .....	6
3. Teori .....	8
3.1 Boligmarkedet .....	8
3.2 Alonso-Mills-Muth-modellen .....	9
3.2.1 Om modellen og dens forutsetninger .....	9
3.2.2 Husleiegradienten .....	10
3.2.3 Befolkningsvekst og boligpriser i Alonso-Mills-Muth-modellen .....	12
3.3 Den hedonistiske metoden .....	15
3.3.1 Hedonistiske prisfunksjoner .....	15
3.3.2 Likevekt på etterspørselssiden av markedet .....	16
3.3.3 Likevekt på tilbudssiden av markedet .....	19
3.3.4 Markedslikevekt .....	23
4. Metode .....	24
4.1 Korrelasjon mellom variablene .....	24
4.2 Lineær regresjon .....	26
4.2.1 Generelt om lineær regresjon .....	26
4.2.2 Minste kvadraters metode .....	27
4.2.3 De klassiske forutsetningene .....	30
4.2.4 Modellens forklaringskraft, $R^2$ og $\bar{R}^2$ .....	31
4.2.5 Ulike funksjonsformer .....	33

5. Datainnsamling .....	35
5.1 Innsamling av boligdata .....	35
5.2 Reisetid i rushtrafikk .....	38
5.2.1 Reisetidsregistreringer .....	38
5.2.2 Kjøretid mellom boligene og Kristiansand sentrum .....	41
5.3 Presentasjon av datamateriale .....	42
6. Analyse .....	45
6.1 Korrelasjon mellom kvadratmeterpris og avstand fra sentrum .....	45
6.2 Lineær regresjonsanalyse .....	48
6.2.1 Lineær funksjonsform med avstand målt i kilometer .....	48
6.2.2 Lineær funksjonsform med avstand målt i reisetid .....	51
6.2.3 Dobbellogaritmisk funksjonsform med avstand målt i kilometer .....	53
6.2.4 Dobbellogaritmisk funksjonsform med avstand målt i reisetid .....	55
6.2.5 Kvadratisk funksjonsform med avstand målt i antall kilometer .....	57
6.2.6 Kvadratisk funksjonsform med avstand målt i reisetid .....	61
6.3 Diskusjon .....	65
7. Konklusjoner .....	73
Litteraturliste .....	75
Vedlegg 1: Testing av VIF .....	76

# Oversikt over figurer

Figur 1: Kart over Kristiansand .....	4
Figur 2: Kart over Kristiansandsregionen.....	6
Figur 3: Prosentvis andel av sysselsatte som pendler til Kristiansand i 2010.....	7
Figur 4: Husleiegradienten i Alonso-Mills-Muth-modellen .....	12
Figur 5: Befolkningsvekst og husleie i Alonso-Muth-Mills-modellen.....	13
Figur 6: Den hedonistiske metoden, husholdningenes budfunksjoner .....	18
Figur 7: Den hedonistiske metoden, tilbydernes offerfunksjon .....	22
Figur 8: Markedslikevekt ved den hedonistiske metoden .....	23
Figur 9: Spredningsplott, eksempler på korrelasjon mellom to variabler.....	24
Figur 10: Kjøreruter reisetidsregistrering .....	39
Figur 11: Histogram, observasjoner fordelt etter sentrumsavstand. ....	43
Figur 12: Plott, kvadratmeterpris mot avstand fra sentrum .....	46
Figur 13: Regresjon, lineær funksjonsform med avstand målt i kilometer. ....	48
Figur 14: Residualplott, lineær funksjonsform med avstand målt i kilometer. ....	49
Figur 15: Regresjon, lineær funksjonsform med avstand målt i reisetid. ....	51
Figur 16: Residualplott, lineær funksjonsform med avstand målt i reisetid. ....	52
Figur 17: Regresjon, dobbeltlogaritmisk funksjonsform med avstand målt i kilometer. ....	53
Figur 18: Residualplott, dobbeltlogaritmisk modell med avstand målt i kilometer. ....	54
Figur 19: Regresjon, dobbeltlogaritmisk funksjonsform med avstand målt i reisetid. ....	55
Figur 20: Residualplott, dobbeltlogaritmisk funksjonsform med avstand målt reisetid. ....	56
Figur 21: Regresjon, kvadratisk funksjonsform med avstand målt i kilometer. ....	58
Figur 22: Kvadratisk funksjonsform, total priseffekt av sentrumsavstand målt i kilometer ....	59
Figur 23: Kvadratisk funksjonsform, effekt på estimert pris av en marginal økning i avstand målt i kilometer .....	59
Figur 24: Residualplott, kvadratisk modell med avstand målt i kilometer.....	60
Figur 25: Regresjon, kvadratisk funksjonsform med avstand målt i reisetid .....	61
Figur 26: Kvadratisk funksjonsform, total priseffekt av sentrumsavstand målt i reisetid. ....	62
Figur 27: Kvadratisk funksjonsform, effekt på estimert pris av en marginal økning i avstand målt i reisetid. ....	62
Figur 28: Residualplott, kvadratisk modell med avstand målt i reisetid. ....	63
Figur 29: Effekt av lineær approksimasjon av logaritmisk kurve når avstand øker. ....	67

# Oversikt over tabeller

Tabell 1: Oversikt over variablene i oppgaven.....	36
Tabell 2: Forsinkelser på reiserutene, antall minutt per kjørte km. ....	39
Tabell 3: Forsinkelser i total kjøretid om morgenen på rutene i reisetidsregistreringen .....	40
Tabell 4: Forsinkelser i total kjøretid om ettermiddagen på rutene i reisetidsregistreringen.....	40
Tabell 5: Fordeling av omsatte boliger på kommuner .....	42
Tabell 6: Oppsummeringstabell over undersøkte boligkarakteristikker.....	43
Tabell 7: Korrelasjonsmatrise over variablene i modellen .....	45
Tabell 8: Lineær korrelasjon mellom kvadratmeterpris og avstand fra sentrum.....	46
Tabell 9: Tolkning av de estimerte $\beta$ -koeffisientene, dobbeltlogaritmisk funksjonsform med sentrumsavstand målt i kilometer. ....	53
Tabell 10: Tolkning av de estimerte $\beta$ -koeffisientene, dobbeltlogaritmisk funksjonsform med sentrumsavstand målt i reisetid i rushtrafikk. ....	55
Tabell 11: Resultater fra lineær regresjonsanalyse med lineær funksjonsform, sammenlignet med Alfheim (2005).....	66
Tabell 12: Resultater fra lineær regresjonsanalyse med dobbeltlogaritmisk funksjonsform, sammenlignet med Alfheim (2005).....	69

# Oversikt over vedlegg

Vedlegg 1: Test av Variance Inflating Factors (VIF).....	76
--	----



# 1. Innledning

## 1.1 Bakgrunn

Boligmarkedet er et svært interessant tema fordi det på en eller annen måte angår alle som har et sted å bo, og det vil jo forhåpentlig vis bety de aller fleste av oss. For mange oppleves summene det omsettes for som svært høye, og for de fleste boligeiere er investeringen i bolig kanskje den største investeringen de gjør i løpet av livet. Informasjon og kunnskap om hvordan boligmarkedet fungerer er derfor av stor interesse for veldig mange. Dette gjelder kanskje særlig unge voksne som gjerne har begrenset erfaring med boligmarkedet fra før, som skal ut og kjøpe seg sin første bolig.

Det er tidligere skrevet masteroppgaver om boligprisene i Kristiansand, og felles for flere av disse er at avstand til sentrum blir brukt som en viktig variabel ved estimering av en hedonistisk prisfunksjon. Professor Theis Theisen og dosent Karl Robertsen trekker særlig frem Alfheim (2005) i deres essay om boligprisene i Kristiansand (Robertsen & Theisen, 2010). Her blir avstanden målt i antall kilometer kjøredistanse mellom boligen og Kristiansand sentrum, og det blir funnet at boligprisene faller med et bestemt kronebeløp per kilometer, eller en bestemt prosentandel av boligprisen når man beveger seg bort fra sentrum. Dette er absolutt ikke urimelig, og helt i tråd med allment akseptert teori fra eiendomsøkonomi. Men er antall meter den eneste måten å uttrykke avstand på? Er det rimelig å anta at en bolig som ligger 15 kilometer strak motorvei fra sentrum verdsettes lavere enn en lignende bolig som ligger bare 10 kilometer unna sentrum, men hvor store deler av distansen må kjøres på en kronglete traktorvei?

## 1.2 Problemstilling

Kristiansand sentrum er arbeidssted for en stor del av byens befolkning, samtidig som en betydelig andel av de nærliggende kommunenes innbyggere også har sin arbeidsplass i Kristiansand. Dette medfører en relativt stor trafikkbelastning på innfartsårene til og

utfartsårene fra Kristiansand sentrum i rushtrafikken. Øst for byen er E18 bygd ut med fire kjørefelt, men på vestsiden består veinettet stort sett av tofeltsveier, og det kan oppstå relativt store kødannelser både i morgenrushet og i ettermiddagsrushet.

Denne oppgaven skal forsøke å belyse hvorvidt disse rushtrafikkproblemene blir hensyntatt av boligmarkedet. Kjent teori fra eiendomsøkonomi tilsier at boligprisene synker jo lenger boligen er fra sentrum, og det er tidligere gjort studier på dette området og funnet en slik sammenheng. Denne oppgaven vil i midlertid vektlegge avstanden mellom boligene og sentrum i kjøretid i rushtrafikk vel så mye som den nøyaktige avstanden i meter, og vil undersøke i hvilken grad avstand målt på denne måten slår inn på boligprisene. Oppgavens konkrete problemstilling er som følger:

**Hvilken effekt har avstanden fra sentrum, målt i reisetid i rushtrafikk, på boligprisene?**

### 1.3 Oppgavens oppbygging

Oppgaven starter med et kort innledende kapittel om bakgrunn, problemstilling og oppgavens oppbygging, mens kapittel 2 vil inneholde en oversikt over de geografiske områdene som blir nevnt i oppgaven samt en forklaring på hvorfor nettopp disse områdene er interessante i sammenheng med denne oppgaven.

Kapittel 3 vil dreie seg om to forskjellige teoretiske modeller som sier noe om prissettingen av boliger. Det vil først bli utledet en forenklet versjon av Alonso-Muth-Mills-modellen, som argumenterer for at eiendomsprisen er en funksjon av avstanden til sentrum. Det vil så vises ved en gjennomgang av den hedonistiske metoden, at prisen på en boligeiendom er en sammensatt funksjon av en rekke attributter ved boligen.

Kapittel 4 gir en beskrivelse av metodene som er brukt for å undersøke overnevnte problemstilling. Dette dreier seg hovedsakelig om en gjennomgang av regresjonsanalyse, samt en beskrivelse av de klassiske forutsetningene som kreves for at minste kvadraters metode skal være den beste lineære estimeringsmetoden.

Kapittel 5 presenter datamaterialet som er brukt i undersøkelsen. Dette kapittelet inneholder også en beskrivelse av hvordan reisetid i rushtrafikk er beregnet.

Selve analysen av boligmarkedet i Kristiansandsregionen blir gjennomgått i kapittel 6. Her blir det blant annet presentert regresjonsmodeller med flere ulike underliggende funksjonsformer, med sentrumsavstand målt både i kilometer og i tid. Disse blir så sammenlignet med hverandre og med modellen til Alfheim (2005).

Oppgavens konklusjoner presenteres så til slutt i kapittel 7.

## 2. Oversikt over Kristiansandsregionen

### 2.1 Kristiansand

Kristiansand er Norges sjette største kommune, og hadde i begynnelsen av fjerde kvartal 2012 i overkant av 84000 innbyggere (Statistisk sentralbyrå, 2012). Byen Kristiansand er lokalisert på Norges sydspiss og sentrum i byen blir gjerne kalt for Kvadraturen, noe som har sammenheng med bykjernens utforming. Sentrum er imidlertid relativt lite, og er avgrenset av havet på sørsiden og av Baneheia og bymarka på nordsiden. Størstedelen av befolkningen bor derfor i boligområder utenfor selve bykjernen, enten på østsiden eller vestsiden av byen.



Figur 1: Kart over Kristiansand

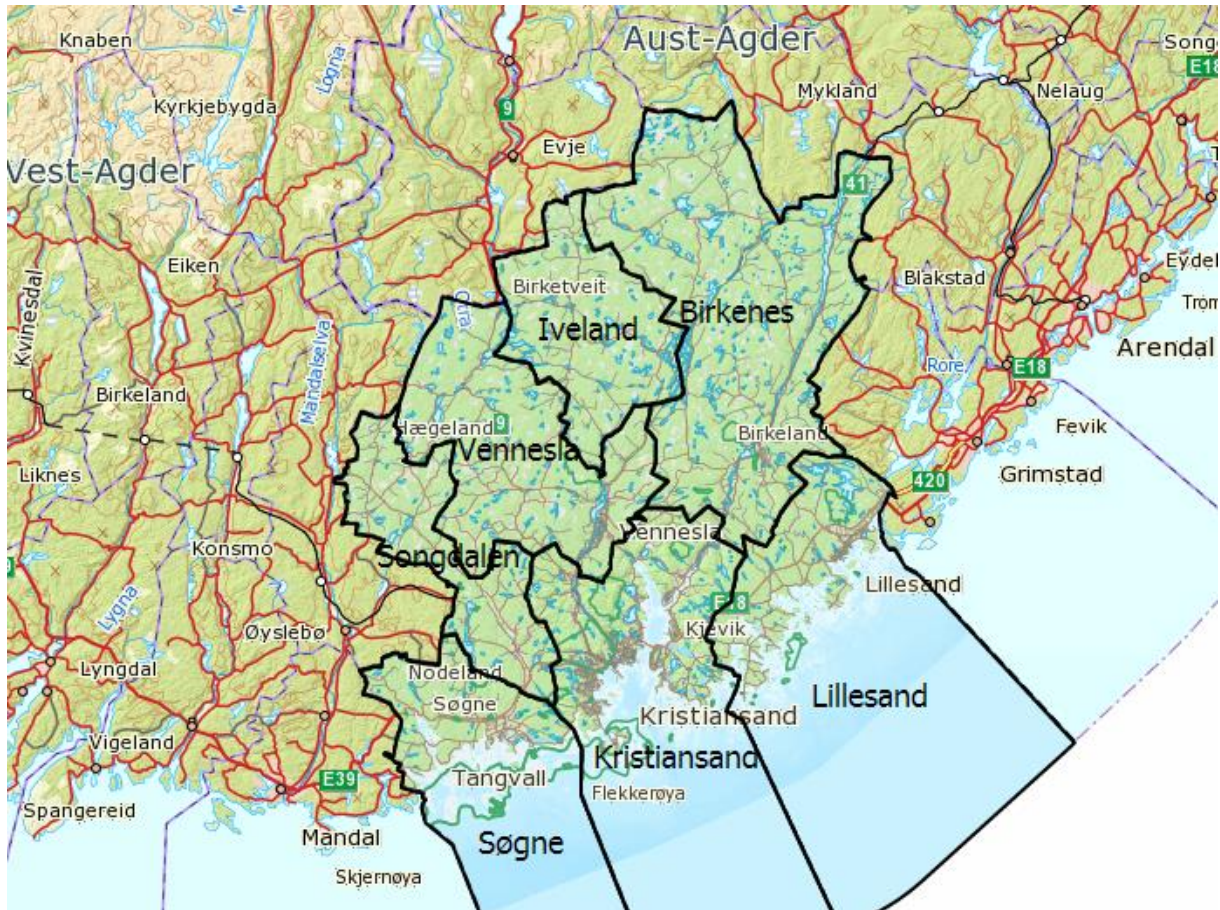
Arbeidsplassene i Kristiansand er til en viss grad er spredd utover kommunens areal. Det finnes industriområder både øst og vest for sentrum, samtidig som Sørlandssenteret har tatt over en betydelig andel av handelen som tidligere var konsentrert til sentrum. Det er likevel slik at Kvadraturen og Eg samlet fortsatt har regionens desidert største forekomst av

arbeidsplasser. Bydelen Eg ligger like nord for Kvadraturen, og er inneklemt mellom elven Otra på østsiden og bymarka på nord og vestsiden. Av denne årsak må all trafikk til denne bydelen nesten helt inn til Kvadraturen før den kan dreie nordover til Eg. Det er derfor naturlig i denne sammenheng å se på de to områdene samlet. Det er for øvrig på Eg at regionens sykehus er lokalisert, og området har til tross for en begrenset størrelse relativt mange arbeidsplasser.

I og med at en så stor del byens befolkning bor i boligområder utenfor Kristiansand sentrum, er det derfor slik at mange må pendle til jobb i Kvadraturen (Knutepunkt Sørlandet, 2013). Dette er med på å skape et press på veinettet inn til og ut fra Kristiansand sentrum i rushtrafikken. Noen arbeider også på en annen kant av byen, og må da gjennom sentrum for å komme på jobb. Det finnes ikke noe reelt alternativ til E18/E39, som er hovedåren inn og ut av sentrum, når man skal kjøre mellom øst og vest i Kristiansand.

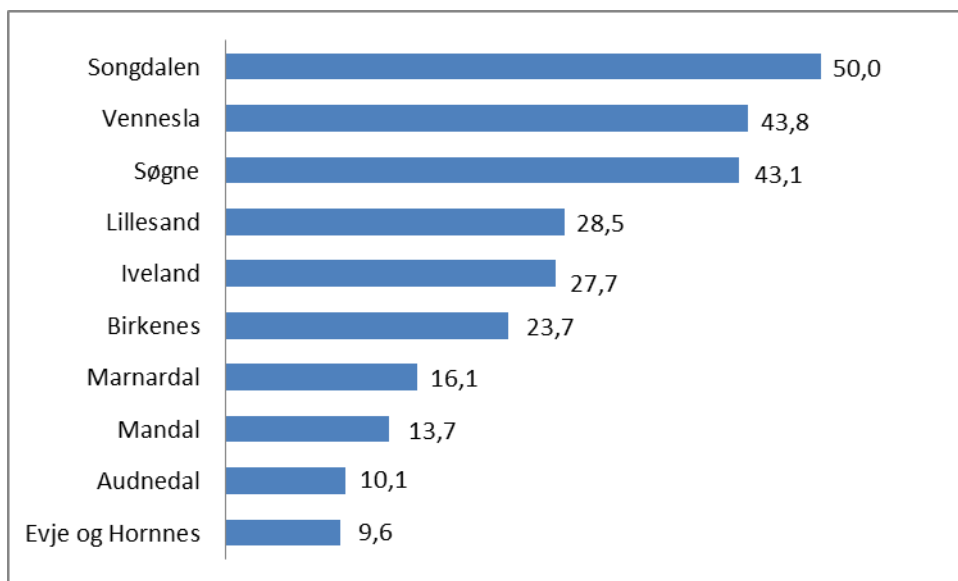
## 2.2 Kristiansandsregionen

Benevnelsen Kristiansandsregionen er ikke offisiell, men brukes gjerne om kommunene Kristiansand, Søgne, Songdalen og Vennesla i Vest-Agder fylke, og kommunene Iveland, Birkenes og Lillesand i Aust-Agder fylke.



Figur 2: Kart over Kristiansandsregionen

For denne regionen utpreger byen Kristiansand seg som et regionalt bysenter. Men til tross for en betydelig vekst i nye arbeidsplasser i Kristiansand har det vært en liten netto tilflytting fra andre deler av landet. Dette indikerer at en del av arbeidstakerne som er sysselsatt i Kristiansand er bosatt i kommuner utenfor Kristiansand, og pendler til Kristiansand for å gå på arbeid (Rådmannen i Kristiansand, 2011). Dette er særlig tilfellet for Søgne, Songdalen og Vennesla, men som vi ser av figur 3 på neste side, har også kommunene Iveland, Birkenes og Lillesand en betydelig andel av befolkningen sysselsatt i Kristiansand.



Figur 3: Prosentvis andel av sysselsatte som pendler til Kristiansand i 2010.  
Tall hentet fra Knutepunkt Sørlandet (2012)

# 3. Teori

## 3.1 Boligmarkedet

Boligmarkedet skiller seg fra de fleste andre markeder ved at produktene som blir solgt er fullstendig heterogene goder. Dette i motsetning til for eksempel markedet for hvete eller olje, som gjerne blir solgt i sekker eller fat og hvor egenskapene til et fat olje ikke skiller seg nevneverdig fra et annet fat olje. For boliger er dette annerledes i og med at hver bolig er unik.

Ved kjøp av bolig kjøper man egentlig en pakke av attributter som er av verdi for boligeieren. Noen attributter kan være like for enkelte boliger, slik som for eksempel antall kvadratmeter boareal og planløsning for to «like» leiligheter i et borettslag. Men selv i dette tilfellet vil ikke boligene være helt like. Boligene kan ikke oppta samme plass, og det vil derfor være forskjell i attributter som utsikt, solforhold eller støyskjerming mot vei og lignende. Nettopp dette at to ellers like boliger alltid vil være forskjellige på grunn av boligens lokasjon, er det som danner grunnlaget for denne oppgaven.

Tilbudet av land innenfor et gitt område vil alltid være fast, og er derfor relativt uelastisk med hensyn på pris. Etterspørselen etter land vil derimot være svært sensitiv med hensyn på pris, noe som skyldes at det som regel finnes mange konkurrerende tomter. Det er dermed etterspørselen som bestemmer prisen, og denne prisen varierer med beliggenhet. Denne tilnærmingen til tomtepriser tilsier at land må prises på en slik måte at den som benytter seg av ei tomt må belastes den relative verdien av de fordelene som eksisterer i forbindelse med nettopp denne tomten i forhold til andre tomter (DiPasquale & Wheaton, 1996).



## 3.2 Alonso-Mills-Muth-modellen

### 3.2.1 Om modellen og dens forutsetninger

Denne modellen bygger på arbeidet til William Alonso, Edwin Mills og Richard Muth og er presentert av Denise DiPasquale og William C. Wheaton (DiPasquale & Wheaton, 1996). Modellen tar utgangspunkt i kanskje den mest grunnleggende egenskapen til boligmarkedet, nemlig at boliger er dyre på attraktive lokasjoner, og relativt sett billigere på mindre attraktive lokasjoner. For å belyse dette på en enkel og ukomplisert måte, beskriver Alonso-Mills-Muth-modellen situasjonen i en by som er preget av mange forutsetninger og begrensninger:

- Byen er monosentrisk. Det er kun ett bysenter, og det er der alle jobbene er.
- Byen har en gitt bygningsstruktur, man kan ikke substituere land med høyere bygninger.
- Folk pendler til jobb i sentrum langs en rett linje fra deres hjem. Pendlingskostnaden er  $k$  kroner per km per år. Avstanden mellom hjemmet og sentrum er representert ved hjelp av variabelen  $d$ .
- Husholdningene er identiske, og antall arbeidere som pendler til og fra sentrum er det samme i alle husholdningene. Husholdningenes inntekt  $y$  er også lik for alle, og denne brukes til husleie, transport og alle andre varer og tjenester. Andre varer og tjenester er samlet og representert ved hjelp av  $x$ .
- Boligene er helt like overalt. Tomtene som husene står på er også like, og det eneste som skiller dem fra hverandre er avstanden fra sentrum. Årlig husleie vil derfor være lik  $R(d)$ , som varierer avhengig av pendlingsdistansen.
- Husleietjenester produseres ved å kombinere tomteareal  $q$  per hus og kostnadene vedrørende det å bygge og vedlikeholde boligen, her representert ved  $c$ .
- Husene blir bebodd av den husholdningen som er villig til å betale mest husleie, og tomteareal allokeres til høyeste pris.

### 3.2.2 Husleiegradienten

Siden både husene og boligtettheten i denne noe spesielle byen er identisk overalt, er den eneste mulige variasjonen mellom husholdninger beløpet de har igjen til å bruke på andre varer og tjenester etter at husleie og pendlingskostnader er betalt. Det er imidlertid slik at husholdningene er identiske, og de har dermed like preferanser når det gjelder hvor stor andel av deres disponible inntekt de ønsker å bruke på dette. Husholdningenes forbruk av andre varer og tjenester,  $x$ , må derfor være konstant over hele byen, og dette forbruket vil ligge på et eller annet nivå som settes lik  $x^0$ . Dersom det da er slik at husleia ikke faller tilstrekkelig for boliger når man beveger seg bort fra sentrum, vil beboerne i utkanten av byen ha intensiver til å flytte nærmere sentrum for å øke andelen av deres disponible inntekt som kan brukes på annet konsum. Dette vil føre til høyere etterspørsel nærme sentrum, og husleia vil øke for sentrumsnære boliger. Dersom det derimot er slik at den årlige husleia blir redusert mye mer enn de eventuelle økte pendlingsutgiftene man pådrar seg ved å flytte bort fra sentrum, vil dette skape insentiver til å flytte bort fra sentrum. Vi ser dermed at under forutsetningene som er gitt, må det være slik at når boligmarkedet er i likevekt så vil husleia reduseres med nøyaktig det samme beløpet som transportkostnadene øker med når man bosetter seg lenger bort fra sentrum (DiPasquale & Wheaton, 1996).

I og med at disponibel inntekt kun brukes til husleie, transport og annet konsum, og at både disponibel inntekt og annet konsum er likt overalt, må vi ha at husleie er lik disponibel inntekt minus transportkostnader og annet konsum. Dette vises i likning 3.1.

$$R(d) = y - kd - x^0 \quad (3.1)$$

I sentrum er pendlingsdistansen lik 0, og leddet  $-kd$  forsvinner. For sentrum har vi derfor at  $R(0) = y - x^0$ . Dersom vi beveger oss ut fra sentrum, vil husleia reduseres med nøyaktig det samme beløpet som pendlingskostnadene øker med. Husleia vil fortsette å reduseres jo lengre man beveger seg ut fra sentrum. Dette vedvarer helt til man når punktet  $b$ , som representerer bygrensa, her er husleieprisene de laveste i byen. Utenfor bygrensa er jordbruk den eneste alternative arealbruken. Siden tomteareal allokeres til den bruken som gir størst avkastning, er det avkastningen på jordbruk sammenlignet med husleieprisene som bestemmer hvor bygrensa går. Siden husleieprisene synker når man beveger seg bort fra

sentrum, vil man før eller siden komme til et punkt hvor avkastningen på boligbygging vil være lavere enn avkastningen ville vært dersom man drev med jordbruk i stedet. Dette punktet vil være bygrensa (DiPasquale & Wheaton, 1996).

Ved bygrensa vil landeierne ha like stor avkastning om de driver jordbruk eller om de bygger ut en tomt til boligformål. Årlig avkastning på jordbruk er gitt ved  $r^a$  per mål, og tomtestørrelsen er  $q$ . Vi kan dermed finne at den årlige arealleia blir lik ( $r^a q$ ). Videre må det også bygges en eller flere bygninger på tomten for at den kan brukes til boligformål. En av forutsetningene for modellen var at alle boligene i byen er like, og at den årlige kostnaden forbundet med bygging, og vedlikehold av en bolig er lik  $c$ . Den årlige kostnaden forbundet med byggingen av en bolig er annuiteten av byggekostnadene, og denne kostnaden må også dekkes inn av husleia. Denne vil dermed på bygrensa bestå av arealleie pluss bygningsleie, altså ( $r^a q + c$ ) (DiPasquale & Wheaton, 1996).

Vi kan nå sammenligne dette uttrykket for husleia ved bygrensa med likning 3.1, og se at:

$$R(b) = y - kb - x^0 = (r^a q + c)$$

Vi kan videre bruke dette, og enkelt finne at annet konsum da må være:

$$x^0 = y - kb - (r^a q + c) \quad (3.2)$$

I og med at alle husholdninger er like, vil dette uttrykket for annet konsum gjelde for alle husholdningene i byen. Vi kan dermed sette likning 3.2 inn for  $x^0$  i likning 3.1 som var det opprinnelige uttrykket for husleia:

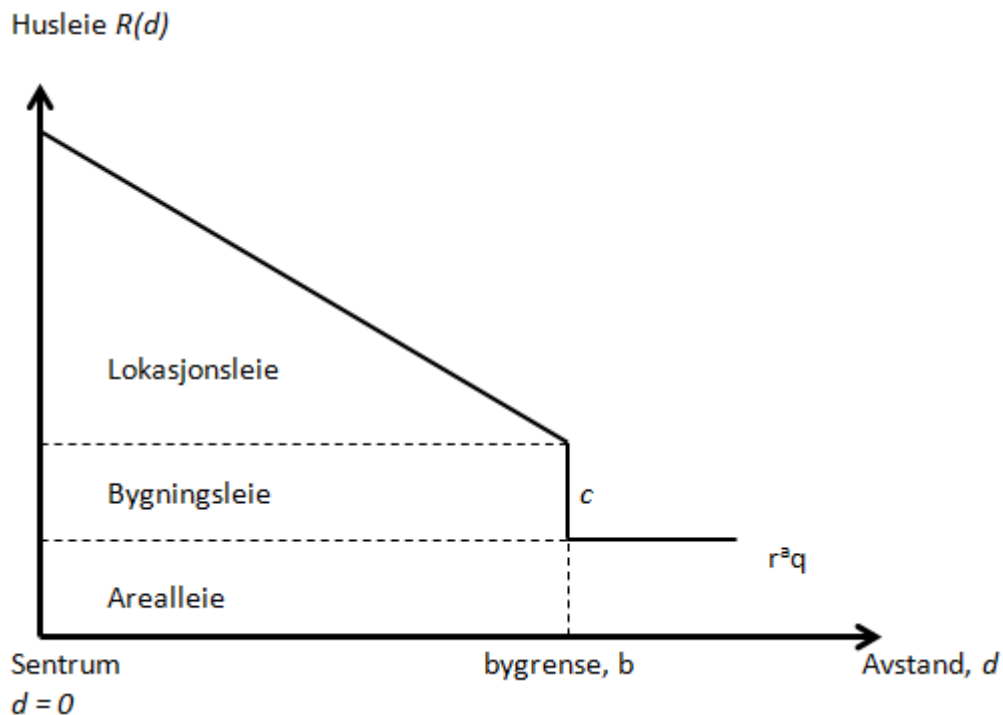
$$R(d) = y - kd - (y - kb - (r^a q + c))$$

⇓

$$R(d) = (r^a q + c) + k(b - d) \quad (3.3)$$

Vi ser her at husleia består av tre komponenter, som blir illustrert i figur 4. Arealleie  $r^a q$ , bygningsleie  $c$ , og det siste leddet,  $k(b - d)$ , er lokasjonsleie. Etersom boligene er like, vil både arealleie og bygningsleie være likt over hele byen, men vi ser at lokasjonsleien vil variere etter hvor boligen er lokalisert i byen. Dersom boligen befinner seg på bygrensa er  $d = b$ , og det siste leddet forsvinner. Husleia blir dermed lik ( $r^a q + c$ ) på bygrensa. Dersom

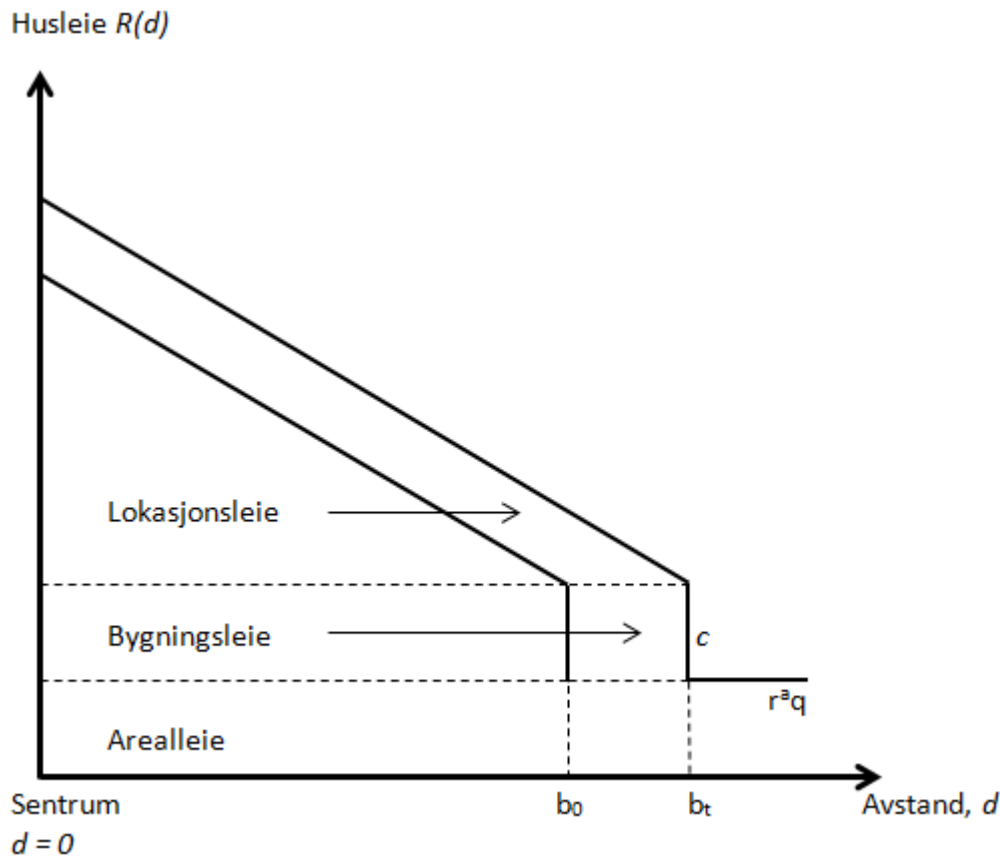
boligen imidlertid ligger helt i sentrum, vil  $d = 0$ , og husleia blir dermed lik  $(r^a q + c) + kb$ . Ved å derivere likning 3.3 med hensyn på  $d$ , får vi at  $\frac{\partial R(d)}{\partial d} = -k$ . Vi ser altså at i denne modellen synker lokasjonsleien med nøyaktig like mye som transportkostnadene øker med når man beveger seg bort fra sentrum.



Figur 4: Husleiegradienten i Alonso-Mills-Muth-modellen (DiPasquale & Wheaton, 1996)

### 3.2.3 Befolkningsvekst og boligpriser i Alonso-Mills-Muth-modellen

Dersom byen i modellen vår er gjenstand for befolkningsvekst, vil bygrensa måtte flytte seg utover for å få plass til nye boliger. Når disse nye boligene bygges utenfor det som var den gamle bygrensa, fører dette til at samtlige allerede eksisterende boliger i byen vil få en økt lokasjonsleie. Boliger som ligger på den gamle bygrensa, og som dermed ikke hadde noen lokasjonsleie før, ligger nå nærmere sentrum enn de nye boligene, og man må derfor betale en økt husleie på grunn av de sparte transportkostnadene man har i forhold til dem som ligger på den nye bygrensa. Husleia vil til en hver tid være  $(r^a q + c)$  på bygrensa, og som det illustreres i figur 5 vil husleia øke for alle boliger som er lokalisert nærmere sentrum enn den nye bygrensa.



Figur 5: Befolkningsvekst og husleie i Alonso-Muth-Mills-modellen (DiPasquale & Wheaton, 1996)

På bakgrunn av dette ser vi at for en gitt bolig i en by som vokser er ikke husleia konstant fra år til år, men den vil i stedet stige ettersom byen vokser. Uttrykket for husleia som er funnet i likning 3.3 kan fortsatt brukes, men det må presiseres hvilket tidspunkt man beregner husleia for. Husleia på tidspunkt  $t$  er gitt ved:

$$R_t(d) = (r^a q + c) + k(b_t - d), \quad d \leq b_t, \text{ for alle } t \quad (3.4)$$

Dette kan vi videre bruke til å finne eiendomsprisen på tidspunkt  $t$  for denne boligen. Eiendomsprisen på tidspunkt  $t$  er gitt som nåverdien av alle fremtidige husleiestrømmer som er definert i likning 3.4. Ved å bruke diskonteringsatts  $i$ , og forventet årlig vekst  $g$ , kan vi nå utlede likning 3.5 som gir oss eiendomsprisen for en gitt bolig (DiPasquale & Wheaton, 1996).

$$P_t(d) = NV_{t \rightarrow \infty} \{R_t(d)\} = \frac{r^a q}{i} + \frac{c}{i} + \frac{k(b_t - d)}{i} + \frac{k b_t g}{i(i - g)} \quad d < b_t, i > g \quad (3.5)$$

De to første leddene i uttrykket for eiendomsprisen er selvforklarende. Dette er nåverdien av fremtidig arealleie og bygningsleie. Det tredje leddet er nåverdien av den lokasjonsleia som er gjeldene på tidspunkt  $t$ , mens det siste leddet er nåverdien av forventet vekst i lokasjonsleie. Det siste leddet er funnet ved å ta nåverdien av fremtidig lokasjonsleie i sentrum inkludert vekst, minus nåverdien av fremtidig lokasjonsleie i sentrum uten vekst.

$$\frac{kb_t}{(i-g)} - \frac{kb_t}{i} = \frac{kb_t g}{i(i-g)}$$

Alternativt kan likning 3.5 omskrives til et litt enklere uttrykk ved å slå sammen de to siste leddene, og med litt bearbeiding ender vi da opp med følgende uttrykk for eiendomsprisen for en gitt eiendom på tidspunkt  $t$ :

$$P_t(d) = \frac{r^a q}{i} + \frac{c}{i} + \frac{kb_t}{(i-g)} - \frac{kd}{i} \quad (3.6)$$

Vi ser her at denne modellen beskriver eiendomsprisene som en funksjon av avstanden til sentrum. Versjonen av Alonso-Muth-Mills-modellen som er presentert her er svært forenklet, men det viser seg imidlertid at hovedresultatene er robuste (Robertsen & Theisen, 2010). En mer generell variant av modellen ble presentert av Jan K. Brueckner i 1987. Her er forutsetningene ikke like strenge, og han finner blant annet at tomtene blir større når man beveger seg bort fra sentrum og at husleiegradienten blir ikke-lineær (Brueckner, 1987).

## 3.3 Den hedonistiske metoden

### 3.3.1 Hedonistiske prisfunksjoner

Den forenklete Alonso-Muth-Mills-modellen som ble gjennomgått i forrige delkapittel belyser hvordan to ellers like boliger kan variere stort i pris avhengig av deres lokasjon. Den tar imidlertid ikke hensyn til at det finnes en rekke andre attributter ved en bolig som er med på å gi en boligeiendom sin verdi. Dette bringer oss videre til den hedonistiske metoden.

Betegnelsen hedonisme kommer av det greske ordet hedone, som betyr lyst eller glede. I økonomifaget forbindes den hedonistiske metoden med heterogene goder, og ordet hedonisme brukes fordi man tar utgangspunkt i at de heterogene godene er karakterisert ved ulike egenskaper eller attributter, og at det er disse attributtene som gir glede eller nytte (Osland, 2001).

De første hedonistiske analysene bygger implisitt på Kelvin J. Lancasters teorier angående konsumentenes tilpasning. Han hevdet blant annet at en vare i seg selv ikke gir noen nytte til konsumenten, men at en vare har forskjellige egenskaper og at det nettopp er disse ulike egenskapene som gir konsumenten nytte (Lancaster, 1966). Disse ideene ble så videreutviklet av blant andre Sherwin Rosen til et mer fullstendig rammeverk (Rosen, 1974). I Rosens artikkel får man en teoretisk forklaring på sammenhengen mellom den hedonistiske prisfunksjonen og tilpasningen til enkeltaktørene på begge sider av markedet (Osland, 2001).

Rosens modell er statisk og tar utgangspunkt i at et gode kan betraktes som en vektor bestående av  $n$  objektivt målte attributter:  $Z = (Z_1, \dots, Z_n)$ . Dette vil si at for en bolig, betaler man egentlig for en «pakke» av attributter. Eksempler på dette kan være størrelse, antall bad, garasje, sentrumsnærhet osv. Totalprisen for godet blir dermed en funksjon av mengden attributter og deres implisitte pris, slik at den hedonistiske prisfunksjonen kan defineres som  $P(Z) = P(z_1, z_2, \dots, z_n)$  (Osland, 2001).

Som med de fleste modeller, er det også her gjort noen nødvendige forutsetninger. En forutsetning for modellen er at det finnes tilstrekkelig antall boliger på markedet slik at valgene mellom ulike attributtvektorer er kontinuerlige. Det forutsettes også at det finnes mange små markedsaktører som enkeltvis ikke har noen innflytelse på markedsforhold og

priser. Det antas videre at transaksjons -og flyttekostnader antas neglisjerbare, og at aktørene har full informasjon om priser og attributter for alle boligene. Gitt disse forutsetningene kan den hedonistiske teorien forklare hvordan den hedonistiske prisfunksjonen er et resultat av samspillet mellom tilbyderne og etterspørrene i markedet for et heterogent gode. Dette samspillet utpreger seg ved at den hedonistiske prisfunksjonen er en omhylling av etterspørrenes budfunksjoner og tilbydernes offerfunksjoner for hvert enkel attributt (Osland, 2001).

### *3.3.2 Likevekt på etterspørselssiden av markedet*

Husholdningene som etterspør det heterogene godet, i vårt tilfelle en boligeiendom, vil tilpasse seg slik at nytten maksimeres. De er imidlertid bundet av en ikke-lineær budsjettrestriksjon, og må gjøre avveininger om hvor mye nytte de får fra ulike boligattributter i forhold til andre konsumvarer. Vi lar følgende nyttefunksjon beskrive en vilkårlig husholdnings nytte:

$$U_j = U(X, Z, \alpha_j) \quad (3.7)$$

Hvor  $X$  er en vektor av alle andre konsumvarer enn boligen, mens  $\alpha_j$  representerer en vektor av parametere som karakteriserer preferansene til husholdning  $j$ . Nyttefunksjonen antas å være strengt konkav. Dersom vi nå setter prisen på  $X$  lik 1, kan budsjettrestriksjonen beskrives på følgende måte:

$$Y_j = X + P(Z) \quad (3.8)$$

$Y_j$  angir inntekt målt i enheter av  $X$  for husholdning  $j$ , og det forutsettes at hver husholdning kun kjøper en bolig og at den er et konsumgode. I optimum til dette nyttemaksimeringsproblemet vil det være slik at den marginale substitusjonsrate mellom  $Z_i$  og  $X$  vil være lik den partiellderiverte av prisfunksjonen med hensyn til de forskjellige attributtene:

$$\frac{\frac{\partial U_j}{\partial Z_i}}{\frac{\partial U_j}{\partial X}} = \frac{\partial P}{\partial Z_i} \quad (3.9)$$



Høyre side i likning 3.9 svarer til marginale implisitte priser eller hedonistiske priser for attributt  $i$ . Den angir helning til prisfunksjonen i punkter for optimal mengde av  $Z_i$  (Osland, 2001).

Sentralt i forklaringen av markedslikevekt på etterspørselsiden står etterspørres budfunksjon. En utledning av denne tar utgangspunkt i de optimale verdiene for boligvektoren  $Z^*$  og andre goder  $X^*$ , slik at  $X^* = Y_j - P(Z^*)$ . Vi kan nå sette dette inn i nyttefunksjonen (3.7) som da gir oss:

$$U_j = U(Z^*, Y_j - P(Z^*), \alpha_j) = U_j^*$$

Dersom vi nå lar nyttenivået være konstant lik  $U^*$  og antar at inntekten er gitt, kan vi ved å forutsette at den maksimale betalingsvilligheten  $\theta$  er lik den prisen man faktisk betaler  $P(Z^*)$  finne følgende uttrykk for nyttefunksjonen:

$$U_j = U(Z^*, Y_j - P(Z^*), \alpha_j) = U_j^* = U(Z, Y_j - \theta_j, \alpha_j) \quad (3.10)$$

Vi ser her at dette uttrykket (3.10) implisitt definerer en relasjon for maksimal betalingsvillighet ved andre kombinasjoner av boligattributter enn den optimale, på samme tid som husholdningen oppfatter det som om disse ulike sammensetningene gir nøyaktig like mye nytte. Dette indikerer at for andre attributtsammensetninger enn den optimale, vil husholdningen beregne en subjektiv pris som gjør at de vil bli på det samme optimale nyttenivået. Vi ser dermed at budfunksjonen er en indifferenskurve som vil variere med valgt inntekt og nyttenivå, og kan uttrykkes slik:

$$\theta_j = \theta(Z, Y_j, U_j, \alpha_j) \quad (3.11)$$

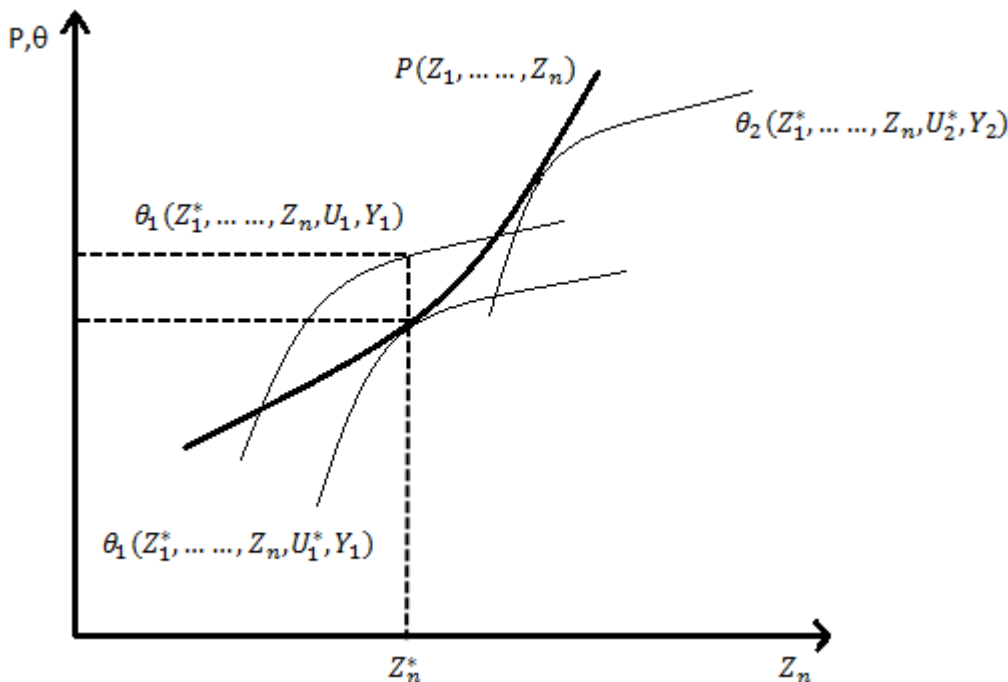
Implisitt derivasjon av (3.11) gir oss:

$$\frac{\partial \theta_j}{\partial Z_i} = \frac{\frac{\partial U_j}{\partial Z_i}}{\frac{\partial U_j}{\partial X}} > 0 \quad (3.12)$$

Den partiellderiverte av budfunksjonen med hensyn på et attributt, tolkes som maksimal betalingsvillighet for en partiell økning i dette attributtet. Siden vi har at nyttefunksjonen er strengt konkav, vil det også være slik at  $\frac{\partial^2 \theta_j}{\partial Z_i^2} < 0$ . Dette betyr at betalingsvilligheten er

positiv men avtakende for partielle økninger i attributter (Osland, 2001).

Det kan vises grafisk hvordan budfunksjonen gir et sett av indifferenskurver for hvert nyttenivå. Figur 6 nedenfor, illustrer en situasjon hvor husholdningen er optimalt tilpasset i alle attributter bortsett fra  $Z_n$ . Langs den vertikale akse måles kroner, og  $\theta_1$  er indifferenskurver for en valgt husholdning. Vi har nå at gitt den hedonistiske prisfunksjonen  $P(Z)$ , vil det å maksimere nytten være det samme som å finne den sammensetningen av attributter som gjør at husholdningen kommer på den laveste budkurven som er mulig. Vi ser at dette punktet finnes akkurat der hvor den hedonistiske prisfunksjonen tangerer husholdningens budkurve. Siden nyttefunksjonen  $U$  inneholder preferanseparameteren  $\alpha$ , vil hver husholdning ha forskjellig budkurve. I figur 6 representerer  $\theta_2$  budkurven til en husholdning som relativt til den første husholdningen har preferanser for større mengde av attributt  $Z_n$ . Denne husholdningen tilpasser seg derfor lengre oppe langs prisfunksjonen.



Figur 6: Den hedonistiske metoden, husholdningenes budfunksjoner

Som illustrert i figur 6, ser vi at de ulike husholdningene tilpasser seg i punktet hvor deres budkurve tangerer den hedonistiske prisfunksjonen. Det vil si at i dette punktet er helningen til de to kurvene helt lik, og dette kan tolkes som at den marginale betalingsvilligheten for

den siste enheten er nøyaktig lik den implisitte prisen for dette attributtet. Dette kan også vises ved å kombinere likningene (3.9) og (3.12), noe som gir oss likevektsbetingelsen på etterspørselssiden:

$$\frac{\partial \theta_j}{\partial Z_n} = \frac{\frac{\partial U_j}{\partial Z_n}}{\frac{\partial U_j}{\partial X}} = \frac{\partial P}{\partial Z_n} \quad j = 1, 2, \dots, m \quad (3.13)$$

I tillegg til tangeringsbetingelsen krever likevekt at  $\theta_j(Z^*, U_j^*, Y_j, \alpha_j) = P(Z)$ . Dette kreves fordi for en hver indifferenskurve  $\theta_j$  vil man kunne finne et punkt som har samme helning som  $P(Z)$ , men det er kun en av disse kurvene som tangerer  $P(Z)$  i dette punktet, nemlig den hvor  $\theta_j(Z^*, U_j^*, Y_j, \alpha_j) = P(Z)$ . Dette kan også forklares intuitivt ved at  $P(Z)$  er det minste beløpet man må betale på markedet for en bolig med attributtvektor  $Z$ , mens det maksimale en husholdning er villig til å betale er gitt ved budfunksjonen  $\theta_j(Z, U_j, Y_j, \alpha_j)$ . Det er dermed en forutsetning for nyttemaksimum at betalingsvilligheten er lik det minste beløpet man er nødt til å betale for en bolig med den optimale sammensetningen av attributter. Enhver annen tilpasning på den lavest oppnåelige budkurven enn tangeringspunktet vil ikke bli akseptert. Dette fordi det vil finnes andre husholdninger med en annen preferansestruktur som er har høyere betalingsvillighet for denne boligtypen. Den hedonistiske prisfunksjonen  $P(Z)$  er på denne måten en omhylling av alle husholdningers budfunksjoner (Osland, 2001).

### 3.3.3 Likevekt på tilbudssiden av markedet

På tilbudssiden av markedet antas det at det finnes mange små bedrifter som produserer boliger, og som tilpasser antall enheter og sammensetning av attributter slik at profitten maksimeres. Det antas videre at bedriftene spesialisere seg, slik at hver bedrift produserer en gitt boligtype med en spesiell attributtsammensetning. Dersom  $\pi$  angir profitten til en bedrift, kan profittfunksjonen til hver enkelt bedrift skrives som:

$$\pi = M \times P(Z) - C(M, Z, \beta) \quad (3.14)$$

Her ser vi at den første delen av profittfunksjonen er en inntektsfunksjon, hvor  $M$ , som representerer antall boliger en bedrift produserer med en gitt attributtvektor  $Z$ , blir

multiplisert med den hedonistiske prisfunksjonen  $P(Z)$ . Prisfunksjonen oppfattes av bedriftene som gitt, og blir derfor ikke påvirket av bedriftenes produksjonsbeslutninger. For å finne profitten blir det så trukket fra kostnader  $C(M, Z, \beta)$  fra inntektene.

Grensekostnadene i produksjonen av attributter er positive og ikke avtagende, noe som betyr at kostnadsfunksjonen  $C$  er en stigende og konveks funksjon av antall boliger  $M$ .  $\beta$  representerer en vektor av skiftparametre som for eksempel faktorpriser eller produksjonsteknologi for den enkelte bedrift, dette sørger for kontinuerlig variasjon i attributter siden markedet består av mange små bedrifter.

Det antas at tilbudet av nye boliger er identisk med produksjonen av nye boliger.

Modifisering og salg av gamle boliger kan imidlertid inkluderes i modellen ved å anta selgere som tilbyr boliger slik at  $M = 1$ , og gitt en kostnadsfunksjon for blant annet modifisering av bolig  $C(\cdot)$  (Hite, 1998).

Vi har følgende førsteordensbetingelser for maksimal fortjeneste:

$$\frac{\partial P}{\partial Z_i} = \frac{\partial C}{\partial Z_i} M \quad (i = 1, \dots, n) \quad (3.15)$$

$$P(Z) = \frac{\partial C}{\partial M} \quad (3.16)$$

Likning 3.15 viser oss at hver tilbyder av bolig bør velge den sammensetningen av attributter som gjør at den implisitte prisen for et gitt boligattributt er lik grensekostnad per bolig ved en partiell økning i attributter. Likning 3.16 sier oss at hver tilbyder av bolig bør produsere et antall boliger slik at grenseinntekt, som er gitt ved prisen på boligen, er lik grensekostnader i produksjonen av boliger (Osland, 2001).

Andreordensbetingelsene er imidlertid ikke helt sikret gjennom antakelsen at kostnadsfunksjonen er konveks. Siden den hedonistiske prisfunksjonen også kan være en

konveks funksjon, må vi derfor forutsette at  $\frac{\partial^2 C}{\partial Z_i^2} > \frac{\partial^2 P}{\partial Z_i^2}$  i det relevante området for

maksimum. Dette vil si at dersom begge funksjonene er konvekse, må kostnadsfunksjonen krumme mer enn prisfunksjonen.

Tilsvarende til budfunksjonen på etterspørselssiden, har vi på tilbudssiden en offerfunksjon  $\Phi = \Phi(Z, \pi, \beta)$ . Denne offerfunksjonen beskriver det minste beløpet produsenter er villige til å akseptere for å kunne tilby boliger med ulike attributter, gitt det optimale antallet boliger som produseres og et konstant profittnivå. For å utlede offerfunksjonen tar vi utgangspunkt i likning 3.14 og setter inn de optimale verdiene  $Z^*$ ,  $M^*$  og  $\pi^*$ , noe som gir oss profittfunksjonen:

$$\pi^* = M^* \times P(Z^*) - C(M^*, Z^*, \beta) \quad (3.17)$$

Vi lar nå profitten være konstant lik  $\pi^*$ . Dersom vi nå antar at prisen som blir betalt for en bolig er lik den laveste prisen en produsent kan tilby, kan profittfunksjonen skrives:

$$\pi^* = M^* \times \Phi(Z^*, \pi^*, \beta) - C(M^*, Z^*, \beta) \quad (3.18)$$

Derivasjon av likning 3.18 med hensyn til  $M$  og  $Z_i$  gir oss førsteordensbetingelsene:

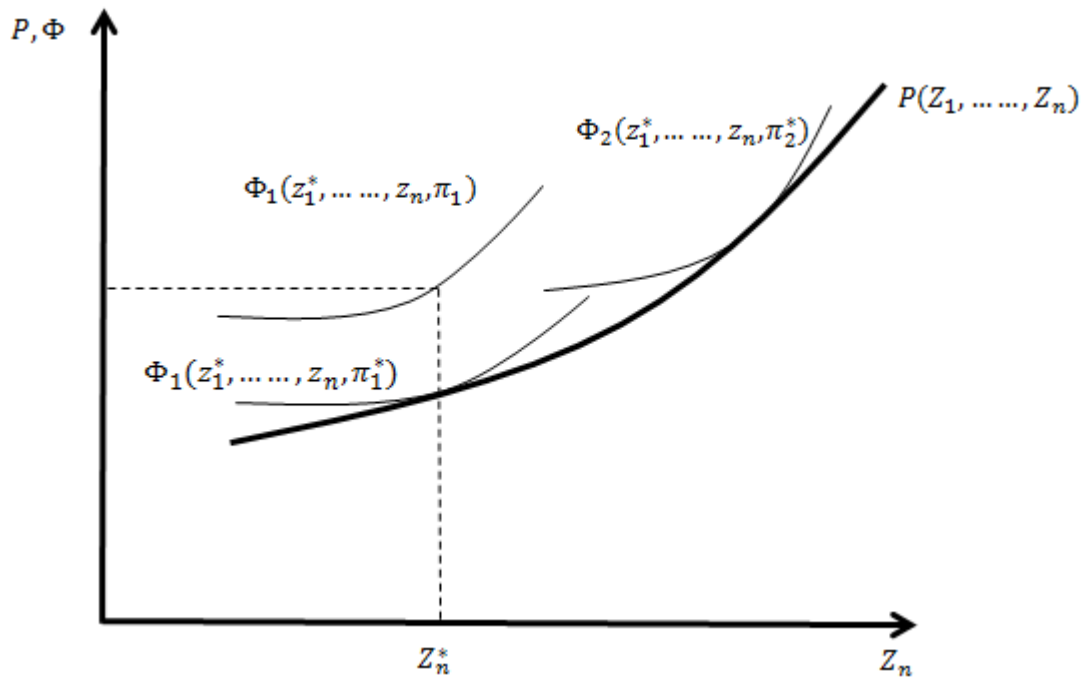
$$\Phi(Z^*, \pi^*, \beta) = \frac{\partial C}{\partial M} \quad (3.19)$$

$$\frac{\partial \Phi}{\partial Z_i} = \frac{\frac{\partial C}{\partial Z_i}}{M} \quad i = 1, \dots, n \quad (3.20)$$

Ved å løse likning 3.19 med hensyn til  $M$  og sette uttrykket inn i likning 3.18, elimineres  $M$ . Profittfunksjonen definerer dermed implisitt en relasjon mellom offerpriser og boligattributter (Osland, 2001):

$$\Phi = \Phi(Z, \pi^*, \beta) \quad (3.21)$$

Offerfunksjonen kan presenteres grafisk. I figur 7 ser vi at den presenteres som isoprofitkurver hvor det antas at tilbyderen er optimalt tilpasset i alle attributter unntatt  $Z_n$ . Kurvene er konvekse og profitten stiger når man beveger seg oppover i diagrammet, for å maksimere profitten ønsker derfor tilbydere å befinne seg på den isoprofitkurven som ligger høyest mulig i diagrammet. Isoprofitkurvene til forskjellige tilbydere av boliger er ulike på grunn av skiftparameteren  $\beta$ , og de forskjellige tilbyderne vil derfor tilpasse seg på ulike steder langs prisfunksjonen.



Figur 7: Den hedonistiske metoden, tilbydernes offerfunksjon

Vi ser av figur 7 at for å maksimere profitten bør tilbyderen med offerfunksjon  $\Phi_1$  tilpasse seg slik at hans isoprofittkurve tangerer den eksogent gitte prisfunksjonen, og produsere boliger med mengde  $Z_n^*$  av attributtet  $Z_n$ . Dette illustrerer hvordan likevekten på tilbudssiden finnes ved å kombinere likningene 3.15 og 3.20, slik at offerkurvene for hver tilbyder tangerer den eksogent gitte prisfunksjonen:

$$\frac{\partial \Phi}{\partial Z_n} = \frac{\frac{\partial C}{\partial Z_n}}{M} = \frac{\partial P}{\partial Z_n} \quad (3.22)$$

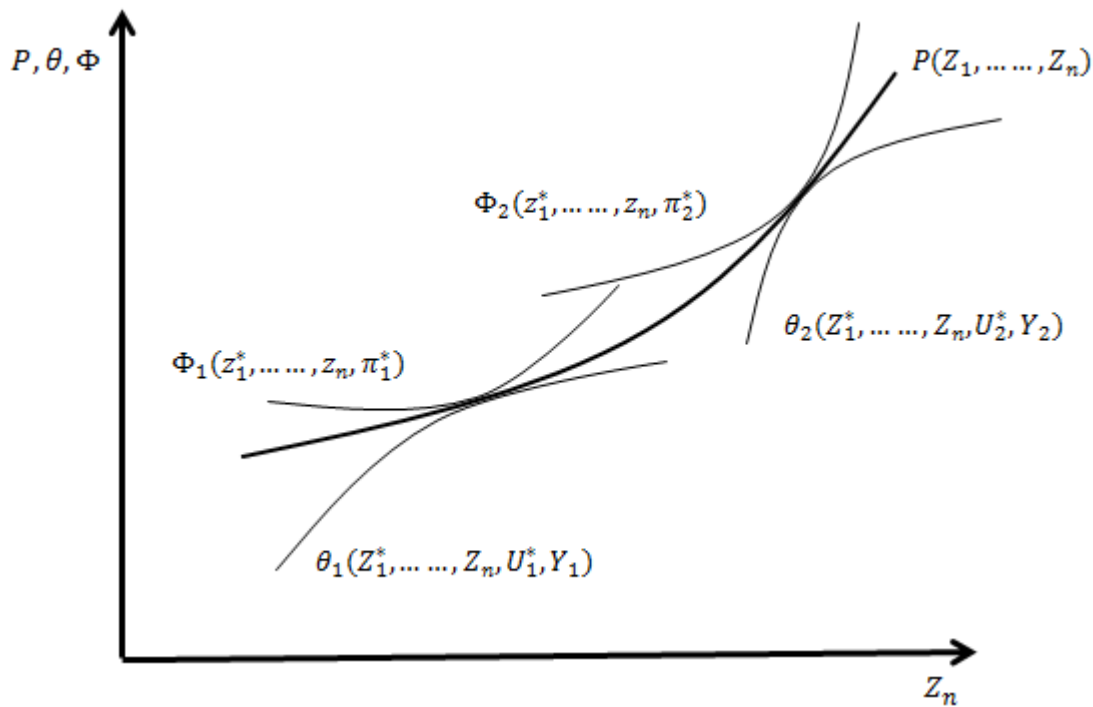
Likning 3.22 sier egentlig kun at i likevekt må offerfunksjonen og prisfunksjonen ha samme helning. Denne helningen vil man imidlertid finne på uendelig mange forskjellige isoprofittkurver for en tilbyder, det kreves derfor også i tillegg at  $\Phi(Z^*, \pi^*, \beta) = P(Z^*)$ . Vi ser dermed at i likevekt er offerprisen lik den eksogent gitte prisfunksjonen.

### 3.3.4 Markedslikevekt

Markedslikevekt oppnås ved at husholdningenes budfunksjon og tilbydernes offerfunksjon tangerer hverandre:

$$\frac{\partial \theta}{\partial Z_i} = \frac{\partial P}{\partial Z_i} = \frac{\frac{\partial C}{\partial Z_i}}{M} = \frac{\partial \Phi}{\partial Z_i} \quad (3.23)$$

Den hedonistiske prisfunksjonen er på denne måten en omhylling av både konsumentenes budfunksjoner og tilbydernes offerfunksjoner (Osland, 2001).



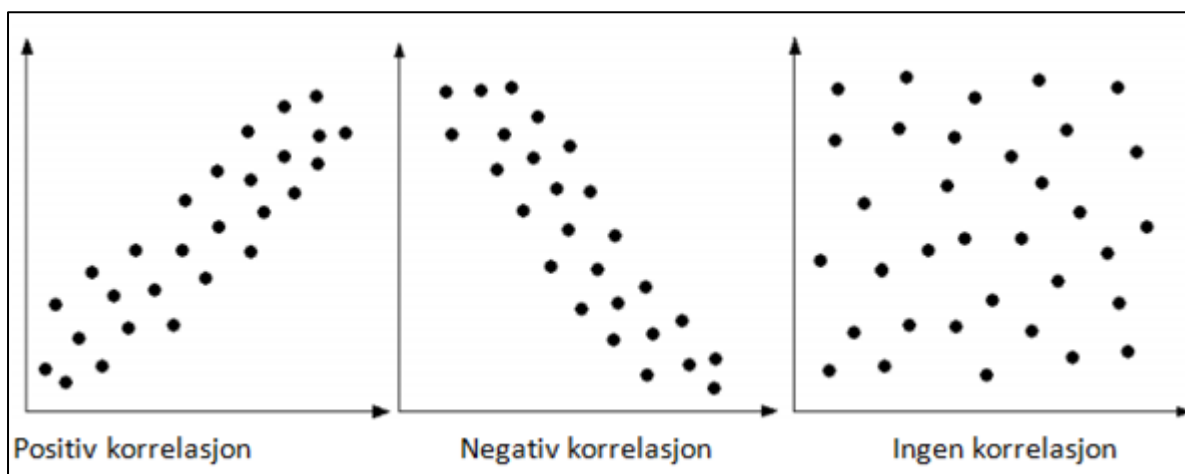
Figur 8: Markedslikevekt ved den hedonistiske metoden

Figuren over illustrerer hvordan en rekke punkter med likevekt mellom tilbydere og etterspørrere danner den hedonistiske prisfunksjonen.

## 4. Metode

### 4.1 Korrelasjon mellom variablene

Når man ønsker å undersøke om det er en sammenheng mellom variasjoner i to variabler, kan et enkelt første skritt være å lage et spredningsplott med de to variablene. Et slikt plott vil ofte raskt og enkelt kunne gi indikasjoner på om det er noen sammenheng, og hvilken form denne sammenhengen har.



Figur 9: Spredningsplott, eksempler på korrelasjon mellom to variabler

Figur 9 viser forskjellige typer korrelasjon mellom to variabler. De positivt korrelerte variablene vil ofte variere på samme måte, slik at en variasjon i den ene variabelen ofte medfører en variasjon i den andre variabelen i samme retning. Negativt korrelerte variabler varierer også samtidig, men i da i motsatt retning. Dersom den ene variabelen stiger, vil dette gjerne sammenfalle med en reduksjon i den andre variabelen. Vi har også tilfeller med ingen korrelasjon, hvor variasjonene i en variabel er fullstendig upåvirket av variasjonen i den andre.



Et mer deskriptivt mål på den lineære sammenhengen mellom to variabler er korrelasjonskoeffisienten  $r$ , som er definert på følgende måte:

Hvis vi lar  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$  og  $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ , da er

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

Korrelasjonskoeffisienten vil variere fra -1 til 1, og er et mål på styrken og retningen på den lineære sammenhengen mellom to variabler. Jo nærmere absoluttverdien til  $r$  er lik 1, jo sterkere er den lineære korrelasjonen mellom de to variablene.

- Hvis to variabler er perfekt positivt korrelert er  $r = 1$
- Hvis to variabler er perfekt negativt korrelert er  $r = -1$
- Hvis to variabler er fullstendig ukorrelert er  $r = 0$

Dersom to variabler imidlertid korrelerer på en ikke-lineær måte, vil dette ikke nødvendigvis bli fanget opp av korrelasjonskoeffisienten  $r$ .

## 4.2 Lineær regresjon

### 4.2.1 Generelt om lineær regresjon

En regresjonsanalyse er en statistisk teknikk som forsøker å forklare variasjoner i en variabel, gjennom en funksjon av bevegelser i et sett av andre variabler. Den avhengige variabelen kan for eksempel være prisen på en bolig, som kan estimeres ved hjelp av forskjellige boligattributter, som da er de uavhengige variablene.

Den enkleste formen for lineær regresjon er  $Y = \beta_0 + \beta_1 X$ . Her representerer  $\beta_0$  konstantleddet, som indikerer verdien til  $Y$  dersom verdien til den uavhengige variabelen  $X$  er lik 0.  $\beta_1$  er helningskoeffisienten, den indikerer hvor mye  $Y$  vil endres dersom den uavhengige variabelen  $X$  økes med en enhet. I tillegg til variasjonen i  $Y$  som er forårsaket av den uavhengige variabelen  $X$ , finnes det nesten alltid ytterligere variasjon som er forårsaket av andre årsaker, og som rett og slett ikke kan forklares av modellen. Dette kan stamme fra viktige variabler som ikke er tatt med i modellen, målefeil, feil funksjonsform eller kan stamme fra fullstendig vilkårlige og uforutsigbare hendelser. For å ta hensyn til denne variasjonen som ikke kan forklares av de uavhengige variablene, blir et stokastisk feilledd,  $\varepsilon$ , lagt til regresjonslikningen som da blir seende slik ut:  $Y = \beta_0 + \beta_1 X + \varepsilon$

Det er imidlertid slik at en regresjon som skal beskrive boligpriser sannsynligvis trenger mer enn bare én uavhengig variabel, noe som bringer oss videre til en multippel lineær regresjonsmodell. En slik regresjonsmodell med for eksempel tre uavhengige variabler, vil se slik ut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Betydningen av  $\beta_0$  er den samme som før, altså verdien til  $Y$  dersom alle de uavhengige variablene er lik 0. Betydningen til koeffisienten  $\beta_1$  er den effekten en enhets økning i  $X_1$  har på  $Y$  når man holder de andre uavhengige variablene konstant. På samme måte angir  $\beta_2$  den effekten en marginal økning i  $X_2$  har på  $Y$  når man holder  $X_1$  og  $X_3$  konstant, og så videre.

Ved å inkludere en spesifikk referanse til hver av observasjonene av de forskjellige variablene, kan man skrive den generelle lineære multiple regresjonsmodellen med  $N$  observasjoner og  $k$  uavhengige variabler på denne måten:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (i = 1, 2, \dots, N) \quad (4.1)$$

Et viktig poeng er derimot at denne regresjonslikningen er teoretisk, og umulig å finne i virkeligheten siden det i de fleste tilfeller vil være umulig å både finne og måle riktig absolutt alle variabler som har en effekt på den avhengige variabelen  $Y$ . Den virkelige regresjonslikningen er altså ukjent, og de virkelige regresjonskoeffisientene er umulige å observere. Regresjonslikningen kan likevel estimeres ved hjelp av observasjoner av virkelige verdier til de forskjellige variablene. Den estimerte verdien til  $Y$  er gitt betegnelsen  $\hat{Y}$ , og de estimerte regresjonskoeffisientene er betegnet  $\hat{\beta}$ . Den estimerte regresjonslikningen ser følgende ut på denne måten:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki} \quad (i = 1, 2, \dots, N)$$

$\hat{Y}_i$  er dermed den estimerte verdien til  $Y_i$ , og representerer verdien til  $Y$  for observasjon nummer  $i$ , beregnet med den estimerte regresjonslikningen. Jo nærmere  $\hat{Y}_i$  er de observerte  $Y_i$  i datautvalget, jo bedre passer den estimerte likningen. Forskjellen mellom den estimerte verdien til den avhengige variabelen og den observerte verdien, kalles residualen ( $e_i$ ).

$$e_i = Y_i - \hat{Y}_i \quad (4.2)$$

De fleste regresjonsteknikker forsøker å bestemme  $\hat{\beta}$ -verdier slik at residualene blir minst mulig. Jo mindre residualer, jo bedre passer den estimerte modellen. En av de vanligste metodene som brukes til dette er minste kvadraters metode.

#### 4.2.2 Minste kvadraters metode

Før minste kvadraters metode gjennomgås kan det være hensiktsmessig å presentere likning 4.1 i matriseformat

$$Y_i = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \dots & X_{1,k} \\ 1 & X_{2,1} & X_{2,2} & \dots & X_{2,k} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{N,1} & X_{N,2} & \dots & X_{N,k} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

Eller enda mer kompakt:

$$Y_{(N,1)} = X_{(N,k+1)}\beta_{(k+1,1)} + \varepsilon_{(N,1)}$$

Videre i dette avsnittet om minste kvadraters metode er det, dersom ikke annet er opplyst, matriser som blir behandlet. Matrisenotasjonen utelates derfor, og det antas at det er underforstått at når for eksempel  $\beta$  omtales, er det snakk om vektoren  $\beta_{(k+1,1)}$ .

Prinsippet bak minste kvadraters metode er egentlig relativt enkelt. Man forsøker å estimere regresjonskoeffisientene slik at avvikene mellom de estimerte  $Y$  verdiene og de observerte  $Y$  verdiene blir minst mulig totalt sett. For at et positivt avvik og et negativt avvik ikke skal utligne hverandre, blir avvikene kvadrert før de plusses sammen. Disse avvikene er de samme som residualene presentert i likning 4.2. Vi ser nå at totalsummen av de kvadrerte residualene avhenger av de estimerte regresjonskoeffisientene  $\hat{\beta}$ , og at summen kan beskrives ved funksjonen:

$$L(\hat{\beta}) = e'e = (e_1 \quad e_2 \quad \dots \quad e_N) \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{pmatrix} = e_1^2 + e_2^2 + \dots + e_N^2 = \sum_{i=1}^N e_i^2 \quad (4.3)$$

Minste kvadraters metode går altså ut på å minimere  $L(\hat{\beta})$  med hensyn til  $\hat{\beta}$ . Løsningen på dette minimeringsproblemet kalles ordinary least squares (OLS) estimatoren til regresjonsparameteren  $\beta$ .

For å løse dette problemet endrer vi litt på likning 4.3. Siden vi har at  $e = Y - \hat{Y}$  og at  $\hat{Y} = X\hat{\beta}$ , kan likning 4.3 også skrives på denne måten:

$$L(\hat{\beta}) = e'e = (Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

Vi ser nå at både  $\hat{\beta}'X'Y$  og  $Y'X\hat{\beta}$  har dimensjonene  $1 \times 1$ , så det er dermed slik at  $\hat{\beta}'X'Y = Y'X\hat{\beta}$ . Likningen over kan da skrives:

$$L(\hat{\beta}) = e'e = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

Førsteordensbetingelsen for minimum av  $L(\hat{\beta})$  gir oss:

$$\frac{\partial L}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0 \quad (4.4)$$

Vi ordner litt i likning 4.4, og ser at:

$$\begin{aligned} 2X'Y &= 2X'X\hat{\beta} \\ X'Y &= X'X\hat{\beta} \\ \Downarrow \\ \hat{\beta} &= (X'X)^{-1}X'Y \end{aligned}$$

Vi ser altså at vektoren med OLS-koeffisientestimatene for en regresjonsmodell med  $k$  uavhengige variabler er gitt ved:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = (X'X)^{-1}X'Y \quad (4.5)$$

Som vi ser, er det mulig ved hjelp av likning 4.5 å estimere OLS-koeffisientene for hånd. Dette vil imidlertid være svært tidkrevende selv for en relativt liten regresjonsmodell, heldigvis kan en datamaskin gjøre dette arbeidet for oss på en brøkdel av et sekund. I denne oppgaven er det brukt programvaren Stata.

Variansen til de estimerte koeffisientene er gitt ved:

$$\text{var}(\hat{\beta}) = s^2(X'X)^{-1} \quad \text{hvor} \quad s^2 = \frac{e'e}{N - (k + 1)}$$

Fullstendig utledning av varians er ikke gitt her, men kan finnes i de fleste lærebøker om økonometri, som for eksempel i boken til Brooks (2008).

### 4.2.3 De klassiske forutsetningene

De klassiske forutsetningene referer til et sett av grunnleggende forutsetninger som må holde for at minste kvadraters metode skal være den beste tilgjengelige estimatoren for regresjonsmodeller (Studenmund, 2006).

1.  $E(\varepsilon_i) = 0$  Gjennomsnittlig er feilleddene lik 0
2.  $var(\varepsilon_i) = \sigma^2 < \infty$  Feilleddenes varians er konstant og endelig for alle verdier til  $X_i$
3.  $cov(\varepsilon_i, \varepsilon_j) = 0$  Feilleddene er lineært uavhengig av hverandre
4.  $cov(\varepsilon_i, X_i) = 0$  Det er ingen sammenheng mellom feilleddene og den korresponderende  $X$  variansen
5.  $\varepsilon_i \sim N(0, \sigma^2)$  Feilleddene er normalfordelte

Dersom forutsetning 1-4 holder, har OLS-estimatoren en rekke forutsetninger som beskrives ved akronymet BLUE. Dette står for Best Linear Unbiased Estimator. Best betyr i denne sammenheng at OLS-estimatoren  $\hat{\beta}$  har minimum varians. Lineær fordi regresjonsmodellen må være lineær i koeffisientene. Unbiased betyr at estimatoren er forventningsrett, altså at  $E(\hat{\beta}) = \beta$ . Forutsetning nr. 5, om normalfordelte feilledd, er ikke nødvendig for OLS-estimering. Den er imidlertid viktig for å ta riktige slutninger angående estimatene i forbindelse med hypotesetesting.

En implisitt forutsetning ved utledningen av vektoren med OLS-estimatene for regresjonskoeffisientene,  $\hat{\beta} = (X'X)^{-1}X'Y$ , er at to eller flere uavhengige variabler ikke er perfekt korrelert med hverandre. Dette fenomenet kalles perfekt multikollinearitet, og gjør det umulig å estimere alle koeffisientene i modellen. Dersom to uavhengige variabler er perfekt korrelert med hverandre, inneholder de bare nok informasjon til å estimere en parameter, ikke to. Teknisk sett oppstår problemet ved at matrisen  $(X'X)$  ikke er mulig å invertere når to av kolonnene er lineært avhengige av hverandre. Vi har da at determinanten,  $det(X'X) = 0$ , og dermed vil ikke  $(X'X)^{-1}$  eksistere og  $\hat{\beta} = (X'X)^{-1}X'Y$  vil dermed være umulig å kalkulere.

Perfekt multikollinearitet oppstår vanligvis kun når samme variabel ved et uhell er tatt med to ganger i samme regresjon, eller at verdien til en variabel alltid er i et konstant forhold til verdien til en annen variabel. Dette er som oftest relativt enkelt å unngå, og i praksis er det

heller situasjoner hvor vi har nesten perfekt multikollinearitet som man skal være oppmerksom på. Dersom det er ingen korrelasjon mellom de uavhengige variablene, sier man at de er ortogonale til hverandre, men i praksis vil korrelasjonen mellom forklaringsvariablene nesten alltid være forskjellig fra null.

Dette tyder på at det vil alltid være en viss grad av multikollinearitet til stede, men en liten grad av korrelasjon mellom forklaringsvariablene vil imidlertid ikke føre til noe særlig tap av presisjon. Problemer kan imidlertid oppstå dersom graden av multikollinearitet er stor, altså at de uavhengige variablene i stor grad er korrelert med hverandre. Man kan da ende opp med en modell som tilsynelatende har god forklaringskraft, men hvor de individuelle variablene ikke er signifikante. Dette skyldes at ved nært avhengige forklaringsvariabler kan det være vanskelig å observere det individuelle bidraget hver variabel gir til den samlede regresjonen. Konfidensintervallene for parameterne vil derfor bli veldig vide, og tester av signifikansnivå vil dermed kunne gi tvilsomme konklusjoner. Et annet problem med stor grad av multikollinearitet er at regresjonen blir veldig sensitiv med hensyn til små endringer i spesifisering, slik at dersom man fjerner eller legger til en forklaringsvariabel, kan dette føre til store endringer i koeffisientestimer eller signifikansnivå til andre uavhengige variabler (Brooks, 2008).

#### 4.2.4 Modellens forklaringskraft, $R^2$ og $\bar{R}^2$

Den totale variasjonen fra gjennomsnittet av de observerte avhengige variablene, kalles for Sum of Squares Total (SST). Denne totalsummen kan deles opp i to deler. Sum of Squares Regression (SSR), og Sum of Squares Error (SSE). SSR er den delen av variansen som blir forklart av den estimerte regresjonsmodellen, mens SSE er den delen av variansen som ligger i residualene og som modellen dermed ikke har klart å forklare. Vi har altså at  $SST = SSR + SSE$ , denne sammenhengen vises også i likning 4.6.

$$\sum_{i=1}^N (Y_i - \bar{Y})^2 = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^N e_i^2 \quad (4.6)$$

Som mål på regresjonsmodellens forklaringskraft, som også ofte kalles goodness of fit, brukes gjerne koeffisienten  $R^2$ .

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

Vi ser at  $R^2$  beskriver hvor stor andel av den totale variansen som blir forklart av modellen. Siden SST, SSR og SSE alle er ikke-negative størrelser, og siden  $SSR \leq SST$ , må  $R^2$  ligge i intervallet

$$0 \leq R^2 \leq 1$$

Dersom  $R^2$  har en verdi i nærheten av 1, tyder dette på at vi har en regresjonsmodell med svært god forklaringskraft, mens en verdi i nærheten av 0 tyder på at vi har en regresjonsmodell som ikke klarer å estimere verdien til  $Y_i$  noe særlig bedre enn et enkelt gjennomsnitt av alle  $Y$  i utvalget ville gjort.

Et problem med bruken av  $R^2$  som mål på hvor godt modellen passer, er at  $R^2$  aldri blir redusert når nye variabler blir lagt til regresjonsmodellen.  $R^2$  forblir uendret dersom den estimerte  $\beta$ -verdien til den nye variabelen er nøyaktig lik 0, men dette er imidlertid svært sjeldent i praksis. De estimerte  $\beta$ -verdiene er nesten alltid forskjellig fra null, selv om de nødvendigvis ikke er signifikante. Konsekvensen av dette er at  $R^2$  i praksis alltid vil stige dersom nye variabler legges til modellen. Man kan da følgelig ende opp med en modell med en enorm mengde uavhengige variabler, som ikke nødvendigvis har noen sammenheng med den avhengige variabelen, dersom man utelukkende baserer seg på  $R^2$  som mål på hvor godt modellen passer.

For å komme rundt dette problemet med  $R^2$ , blir det ofte gjort en modifisering av  $R^2$  som tar høyde for tapet av frihetsgrader som oppstår når man inkluderer en ekstra variabel i regresjonsmodellen. Dette er kjent som justert  $R^2$ , og er definert som følger:

$$\bar{R}^2 = 1 - \left[ \frac{N-1}{N-k-1} (1 - R^2) \right]$$

Vi ser her at dersom en variabel blir lagt til regresjonen uten at den bidrar til en tilstrekkelig økning i  $R^2$ , vil dette fanges opp av  $\bar{R}^2$  som dermed blir redusert. På denne måten kan  $\bar{R}^2$  bli



brukt som et verktøy i beslutningsprosessen vedrørende om en variabel skal inkluderes i regresjonen eller ikke. En hovedregel kan være å inkludere variabelen dersom  $\bar{R}^2$  øker, mens man bør utelate den dersom  $\bar{R}^2$  synker. Man bør likevel ikke ukritisk basere seg på at å maksimere  $\bar{R}^2$  gir den beste modellen. I hvilken grad de estimerte koeffisientene stemmer med etablert teori og forventningene man hadde til dem før regresjonen ble gjennomført, er vel så viktig som  $\bar{R}^2$  når man skal vurdere hvor godt modellen passer.

#### 4.2.5 Ulike funksjonsformer

De estimerte  $\beta$ -koeffisientene i den lineære modellen som ble presentert i avsnitt 4.2.1, viser en endring i estimert salgpris ved en marginal endring i deres respektive forklaringsvariabler som er like stor uavhengig av hvilken verdi variabelen har. Altså at en økning i for eksempel 20 m<sup>2</sup> BOA for en ellers lik bolig har like stor verdi uavhengig av hvor mange m<sup>2</sup> BOA man har fra før. Dette kan synes litt urimelig. Fra gjennomgangen av den hedonistiske metoden har vi at betalingsvilligheten er positiv men avtagende for partielle økninger i attributter. Det høres også fornuftig ut at en økning i BOA fra 20 m<sup>2</sup> til 40 m<sup>2</sup> har større verdi for en boligeier enn en økning fra 250 m<sup>2</sup> til 270 m<sup>2</sup> dersom alt annet holdes likt. For å ta hensyn til dette kan man ta i bruk en logaritmisk modell, som for eksempel kan være på denne formen:

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} e^{\beta_4 X_4 + \beta_5 X_5 + \varepsilon} \quad (4.7)$$

Modellen i likning 4.7 kunne selvfølgelig hatt flere variabler, dette er kun et eksempel med fem uavhengige variabler hvorav  $X_4$  og  $X_5$  er dummyvariabler. Vi ser for øvrig at modellen beskrevet i likning 4.7 ikke er lineær, men dersom vi nå tar den naturlige logaritmen på begge sider av likhetstegnet ender vi opp med følgende lineære modell:

$$\ln Y = \ln \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (4.8)$$

En hendig egenskap ved dobbeltlogaritmiske regresjonsmodeller, er at de individuelle regresjonskoeffisientene kan tolkes som elastisiteter (Studenmund, 2006). Dersom for

eksempel  $X_2$  øker med 1 % mens de andre forklaringsvariablene holdes konstant, vil dette medføre en endring i  $Y$  på  $\beta_2$  %.

Før man kan gjennomføre en regresjon som har en slik dobbeltlogaritmisk funksjonsform, kan det imidlertid være nødvendig å gjøre visse modifiseringer av noen av variablene. Siden  $\ln(X)$  ikke er definert for  $X = 0$ , blir det problematisk å ta logaritmen til variabler hvor verdien 0 er observert. I denne oppgaven er dette løst ved at verdiene til de variablene som har observasjoner av verdien 0, får lagt til verdien 1. På denne måten blir for eksempel den nye variabelen  $\ln alder$ , som er logaritmen til boligens alder, lik  $\ln(alder + 1)$ . Dummyvariablene blir i teorien gjort om fra å ha verdiene 0 og 1, til å ha verdiene 1 og  $e$ . I praksis betyr dette at vi ikke trenger å modifisere dummyvariablene i datasettet da  $\ln(1) = 0$  og  $\ln(e) = 1$ .

En annen funksjonsform som tar hensyn til at betalingsvilligheten er marginalt avtagende for partielle økninger i attributtet, er en polynomisk funksjonsform av andre grad. En slik funksjonsform kalles gjerne kvadratisk, og uttrykker den avhengige variabelen  $Y$ , som en funksjon av uavhengige variabler hvor minst en av de uavhengige variablene er kvadrert (Studenmund, 2006). Et eksempel på en slik funksjonsform vises i likning 4.9.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_2)^2 + \varepsilon \quad (4.9)$$

En slik funksjon kan produsere kurver for hver av variablene hvor helningen forandrer seg ettersom verdien til den uavhengige variabelen endrer seg. For eksempel er helningen til  $Y$  med hensyn til  $X_2$  i likning 4.9 lik:

$$\frac{\partial Y}{\partial X_2} = \beta_2 + 2\beta_3 X_2 \quad (4.10)$$

Vi kan her se at helningen avhenger av nivået på  $X_2$ . For små verdier av  $X_2$  vil  $\beta_2$  dominere, mens ettersom verdien til  $X_2$  blir større vil  $\beta_3$  overta dominansen.

# 5. Datainnsamling

## 5.1 Innsamling av boligdata

Boligdata i denne analysen er hentet fra databasen til Eiendomsverdi. Dette er et selskap som overvåker og registrerer aktivitet og utvikling i de norske eiendomsmarkedene, og deres databaser inneholder informasjon om alle landets eiendommer samt deres omsetningspriser i nyere tid (Eiendomsverdi, 2013).

Fra databasen til Eiendomsverdi er det hentet ut informasjon om samtlige boligsalg i Kristiansandsregionen i 2010. Dette omsetningsåret er valgt for at omsetningsprisene tidsmessig skal samsvare i høyest mulig grad med rapporten fra Statens veivesen som er brukt til å beregne reisetid i rushtrafikk til og fra Kristiansand sentrum. Relevant data fra denne rapporten blir presentert senere i dette kapitlet.

For noen få av boligene som ble omsatt i 2010, var det imidlertid ikke mulig å finne andre opplysninger enn selve salgssummen. Det vil si at det ikke fantes opplysninger om boareal eller andre boligattributter, og det var heller ikke mulig å finne noe salgsprospekt for boligen. Det kan være flere årsaker til at det mangler informasjon angående disse boligene, men en av disse årsakene kan være at dette ikke har vært et reelt markedssalg. Av denne årsak er derfor boliger som ble omsatt uten at det er mulig å finne denne type informasjon strøket fra listen, og vi sitter igjen med 2526 solgte boliger i Kristiansandsregionen i 2010. Fra denne populasjonen er det trukket ut 253 tilfeldige boliger som er undersøkt videre med hensyn på boligkarakteristikkene som er presentert i tabell 1 på neste side.

Tabell 1: Oversikt over variablene i oppgaven

Variabelnavn	Forklaring	Forventning til $\hat{\beta}$
<i>pris</i>	Boligens salgspris i 1000 kroner	
<i>boa</i>	Boligens boareal	Positiv
<i>alder</i>	Boligens alder	Negativ
<i>kmavstand</i>	Boligens avstand fra sentrum i kilometer snitt tur/retur	Negativ
<i>reisetid</i>	Boligens avstand fra sentrum i reisetid i rushtrafikk snitt tur/retur	Negativ
<i>sjø</i>	Dummyvariabel som indikerer at boligen er en enebolig med egen strandlinje	Positiv

Omsetningsprisen til boligene er hentet fra Eiendomsverdi, men prisen som blir oppgitt der er ikke inkludert boligens andel av en eventuell fellesgjeld i borettslag eller annen type sameie. Det er rimelig å anta at dersom man ved kjøp av bolig også overtar en betydelig andel fellesgjeld, vil dette slå kraftig inn på hvor mye man er villig til å betale for boligen. Dette er undersøkt av Robertsen og Theisen som finner at ved overtagelse av fellesgjeld i borettslag vil gjeldsbeløpet slå nesten fullt ut i prisreduksjon for boligen (Robertsen & Theisen, 2009). De finner at man får en slags rabatt på 9,3 % grunnet mer gunstige rentebetingelser på store felleslån enn hva hver enkelt boligeier kan få på sine boliglån. En andel på 100 000 kr av en stor fellesgjeld vil altså i snitt føre til en redusert omsetningspris på 90 700 kr. Den avhengige variabelen *pris* som er funnet for hver bolig, er derfor salgssummen pluss 90,7 % av en eventuell andel av fellesgjeld.

Boareal og alder er også hentet fra Eiendomsverdi. Boligens alder er funnet ved å trekke boligens byggeår fra omsetningsåret 2010. Hvorvidt en bolig er en enebolig med egen strandlinje, er undersøkt ved å studere boligenes salgsprospekt på finn.no. Avstanden fra sentrum i kilometer er beregnet med ruteplanleggeren Visveg (Visveg, 2013). Distansen er blitt målt begge veier mellom boligen og sentrum, og avstanden som er oppgitt her er gjennomsnittet av dette. Metoden for å beregne avstanden fra sentrum i tid blir grundig gjennomgått i neste delkapittel, som omhandler reisetid i rushtrafikk.

Variabelen *kmavstand* forteller hvor mange km boligen ligger fra Kristiansand sentrum dersom man kjører den mest hensiktsmessige ruten mellom de to punktene. Punktet som representerer sentrum er valgt til krysset mellom Festningsgata og Dronningens gate. Dette

punktet er satt som endepunkt for de fleste av rutene i reistidsundersøkelsen til Statens veivesen, samtidig som det også er relativt nærme Kvadraturens geografiske sentrum. Veiene antas også svært trafikkerte i rushtrafikk da tilkoplingen mellom sentrum og E18 i Baneheitunnelen går via Festningsgata, og all trafikk til og fra Lund over Lundsbroa går via Dronnings gate.

Selv om boligene som er omsatt i Kristiansand sentrum nødvendigvis må befinne seg en viss avstand fra punktet som definerer sentrum, vil denne avstanden aldri være særlig stor. Det ville også vært urimelig å anta en negativ innvirkning på boligprisen inne i Kvadraturen basert på avstand fra dette punktet. På dette grunnlag er det valgt å behandle alle boliger inne i Kvadraturen likt, dette er gjort ved å sette *kmavstand* og *reisetid* lik 1 for samtlige boliger i Kvadraturen.

Det er blitt vurdert å undersøke boligene for flere enn de overnevnte boligkarakteristikker, men dette er ikke funnet hensiktsmessig. Med dette menes det absolutt ikke at egenskaper som nyoppusset bad og kjøkken eller en fantastisk sjøutsikt er irrelevant for en boligkjøper, men boligattributter av denne typen er svært vanskelig å vurdere objektivt og er derfor utelatt.

Noen vil kanskje også savne tomtestørrelse som forklaringsvariabel. Den individuelle tomtestørrelsen for boligtyper som har en andel i en større felles tomt kan være problematisk å beregne. Denne boligtypen utgjør en betydelig andel av de omsatte boligene i Kristiansandsregionen i 2010. Det er mulig å forsøke å estimere disse tomtene, men det er usikkert hvor godt dette estimatet egentlig blir. Det er også usikkert om en tiendedel av en felles eiet tomt på 4000 kvadratmeter er verdt det samme som en tomt på 400 kvadratmeter som man eier alene.

## 5.2 Reisetid i rushtrafikk

### 5.2.1 Reisetidsregistreringer

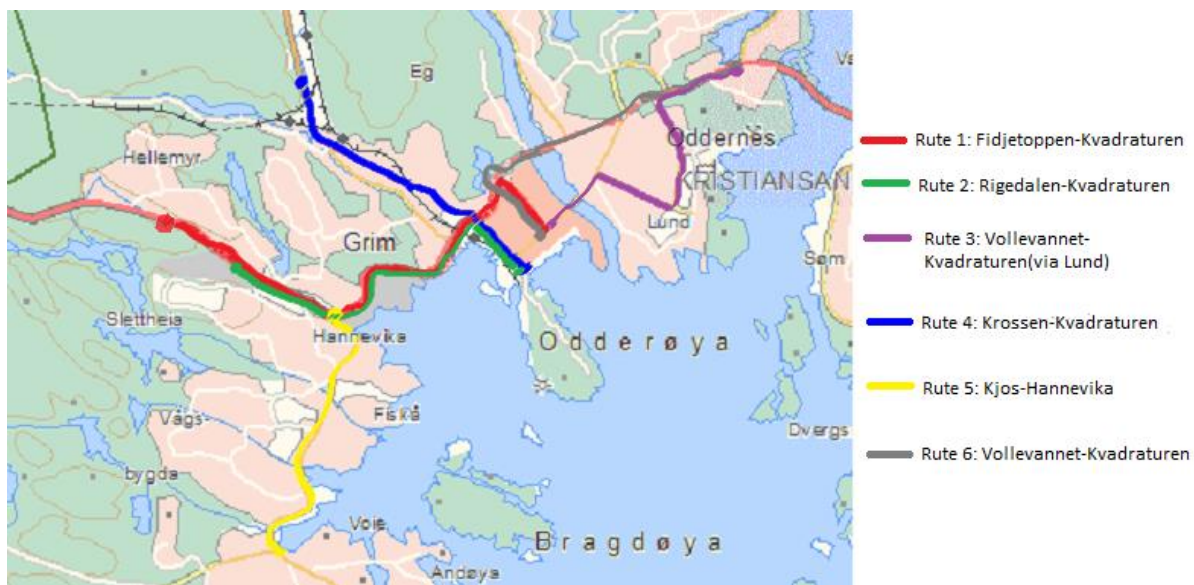
For å fastslå reisetiden folk bruker for å komme seg til og fra sentrum i rushtrafikken, er det tatt utgangspunkt i Samferdselspakke fase 2 for Kristiansandsregionen KVU, utgitt av Statens vegvesen i 2010 (Statens vegvesen, 2010). Denne rapporten presenterer blant annet resultater fra reisetidsregistreringer langs seks ruter i Kristiansand. Disse registreringene er i midlertid gjort allerede høsten 2006, men det vurderes i rapporten at disse tallene fortsatt kan brukes i 2010 (Statens vegvesen, 2010).

Veinettet har ikke hatt noen større endringer i tiden siden reisetidsregistreringene, men arbeidet med en ny vei til Vågsbygd er byggefasen. Denne nye firefeltsveien på fv. 456 vil, når den er ferdig, utvilsomt korte ned på reisetiden mellom Vågsbygd og Kristiansand sentrum i rushtrafikken. Dette prosjektet ferdigstilles i midlertid ikke før i 2014, og basert på Grimstvedt og Eggum Adolfsens konklusjoner vedrørende effekten av nye E18 på boligprisene i Lillesand, vil ikke en slik veibyggning ha signifikant effekt på boligprisene før tidspunktet når veien faktisk åpner (Grimstvedt & Eggum Adolfsen, 2011). Det antas derfor at boligprisene i 2010 var upåvirket av dette.

Rutene i reisetidsregistreringen dekker alle hovedveiene inn mot Kristiansand sentrum, og det antas at det er her veinettet blir hardest belastet i rushtiden. Det forutsettes videre at trafikken flyter tilnærmet normalt i veinettet utenfor disse rutene.

Reisetidsregistreringene som ble gjort i 2006 gjaldt følgende 6 ruter:

- Rute 1: E39 Fidjetoppen – Kvadraturen (via Vesterveibroa)
- Rute 2: E39: Rigidalen – Kvadraturen (via Gartnerløkka)
- Rute 3: E18 Øst: Vollevannet – Kvadraturen (via Lund)
- Rute 4: Rv.9 Krossen – Kvadraturen (kun morgenrush)
- Rute 5: Fv. 456 Kjos – Hannevika (ved isfabrikken)
- Rute 6: E18 øst: Vollevannet – Kvadraturen (via Banneheia tunnel)



Figur 10: Kjøreruter reisetidsregistrering (Statens vegvesen, 2010)

For hver av de overnevnte kjøreruter er det i undersøkelsen fra Statens vegvesen beregnet forsinkelser. Dette er gjort ved at hver rute er kjørt ti ganger i rushtiden, og deretter er gjennomsnittstiden sammenlignet med en normaltids. Denne normaltiden er fremkommeligheten på ruten med angitt fartsgrense relativt uhindret, det er likevel tatt hensyn til forsinkelser forårsaket av signalanlegg og lignende. Det presenteres så en oversikt over antall minutter forsinkelse per kjørte kilometer i henholdsvis morgenrush og ettermiddagsrush på de aktuelle kjørerutene i forhold til normaltiden (Statens vegvesen, 2010).

Tabell 2: Forsinkelser på reiserutene, antall minutt per kjørte km.  
Tall hentet fra Samferdselspakke fase 2 for Kristiansandsregionen KVVU(Statens vegvesen, 2010)

	Rute 1	Rute 2	Rute 3	Rute 4	Rute 5	Rute 6
Morgen	0,91	1,29	0,88	0,83	2,08	0,75
Ettermiddag	0,94	0,75	0,99	Ingen reg.	1,38	0,53

For rute 4 ser vi at det ikke er gjort noen registrering angående forsinkelser i ettermiddagsrushet, og rapporten fra Statens vegvesen nevner dessverre ingenting om årsakene til dette. Man kan imidlertid se på registreringene fra de andre rutene og observere

at det stort sett er mindre forsinkelser om ettermiddagen enn om morgenen. Ved å beregne gjennomsnittsforsinkelsen på de fem andre rutene om morgen og ettermiddag, kan man sammenligne disse og se at forsinkelsene om ettermiddagen er gjennomsnittlig 77,7 % av forsinkelsen om morgenen. Dette blir så videre brukt til å estimere forsinkelser i ettermiddagsrushet på rute 4, som da blir estimert til  $0,83 \text{ minutt} * 0,777 = 0,64 \text{ minutt per km}$ .

Hver av de seks rutenes distanse blir beregnet ved hjelp av ruteplanleggeren Visveg. Dette er en ruteplanlegger som inneholder alle offentlige og private veier fra Kartverkets database med lengde over femti meter som er kjørbare med personbil. Nøyaktigheten på veinettet er ti meter eller bedre, og veiviseren tar hensyn til fartsgrenser og andre kjørerestriksjoner (Visveg, 2013). Vi ender da opp med følgende tabeller for kjøretid i rushtrafikk om morgenen og om ettermiddagen på rutene:

**Tabell 3: Forsinkelser i total kjøretid om morgenen på rutene i reisetidsregistreringen. Kjøretiden er avrundet til nærmeste minutt.**

	Rute 1	Rute 2	Rute 3	Rute 4	Rute 5	Rute 6
Distanse i km beregnet med Visveg	6,00	4,60	4,30	3,20	3,30	4,10
Forsinkelse i minutt pr km	0,91	1,29	0,88	0,83	2,08	0,75
Forsinkelse på ruten i hele minutter	5	6	4	3	7	3

**Tabell 4: Forsinkelser i total kjøretid om ettermiddagen på rutene i reisetidsregistreringen. Kjøretiden er avrundet til nærmeste minutt.**

	Rute 1	Rute 2	Rute 3	Rute 4	Rute 5	Rute 6
Distanse i km beregnet med Visveg	5,9	4,4	4	3,2	3,3	3,9
Forsinkelse i minutt pr km	0,94	0,75	0,99	0,64	1,38	0,53
Forsinkelse på ruten i hele minutter	6	3	4	2	5	2



### 5.2.2 Kjøretid mellom boligene og Kristiansand sentrum

For å beregne kjøretiden mellom Kristiansand sentrum og de aktuelle boligene i denne undersøkelsen, er det tatt utgangspunkt i kjørerutene som ble presentert i figur 10. For hver av boligene er det beregnet kjøretid ved hjelp av Visveg mellom boligen og Kristiansand sentrum via den nærmeste av de overnevnte kjøreruter. I tilfellet for rute 3 og rute 6, hvor startpunkt og endepunkt er det samme for de to rutene, forutsettes det at bilisten velger den ruten som går fortest. Det vil si at det forutsettes at bilister som kommer østfra velger rute 6. For samtlige boliger er det beregnet reisetid inn til sentrum om morgenen, og tilbake til boligen igjen om ettermiddagen. Variabelen *reisetid* som benyttes i analysen er gjennomsnittet av disse reisetidene for hver bolig.

Når det gjelder rute 5, ser vi av figur 10 at denne ruten går fra Kjos til Hannevika. Det vil si at denne ruten ikke har noe endepunkt i sentrum. Dette blir løst ved at det forutsettes at pendlere fra sørvest kjører mellom sentrum og Hannevika med samme gjennomsnittsfart som er blitt registrert på rute 2, som er ruten mellom sentrum og Rigidalen. Denne ruten går fra sentrum til Hannevika hvor den i stedet for å gå sørvest mot Vågsbygd, dreier mot nordvest og fortsetter mot Rigidalen. Gjennomsnittsforsinkelsen på denne strekningen i rushtrafikk er 1,29 min per km om morgenen, og 0,75 min per km om ettermiddagen. Distansen mellom Hannevika og sentrum er i følge Visveg 3,6 km, og estimatet av forsinkelsen i rushtrafikk mellom Hannevika og sentrum blir da  $3,6 \cdot 1,29 \text{ min} = 4,64 \text{ min}$  om morgenen, og  $3,6 \cdot 0,75 \text{ min} = 2,70 \text{ min}$  om ettermiddagen. Ved å legge disse estimatene sammen med beregnet forsinkelser på rute 5, vil vi finne at total forsinkelse mellom Kjos og Kristiansand sentrum i rushtrafikk kan estimeres til henholdsvis  $4,64 \text{ min} + 6,86 \text{ min} \approx 12$  minutter om morgenen, og  $2,70 \text{ min} + 4,55 \text{ min} \approx 7$  minutter om ettermiddagen.

## 5.3 Presentasjon av datamateriale

Populasjonen i datasettet som blir brukt i denne oppgavene er samtlige reelle markedsolgte boliger i Kristiansandsregionen i 2010. Tabell 5 viser hvordan de omsatte boligene fordeler seg på de forskjellige kommunene i regionen.

Tabell 5: Fordeling av omsatte boliger på kommuner

	Populasjon		Tilfeldig utvalg	
	Antall	Prosent	Antall	Prosent
Kristiansand	1872	74,11 %	187	73,91 %
Søgne	186	7,36 %	18	7,11 %
Songdalen	101	4,00 %	11	4,35 %
Vennesla	170	6,73 %	20	7,91 %
Iveland	11	0,44 %	1	0,40 %
Birkenes	51	2,02 %	4	1,58 %
Lillesand	135	5,34 %	12	4,74 %
Totalt	2526	100 %	253	100 %

Som vi kan se av tabellen over, er det tilfeldig trukket ut rett over 10 % av populasjonen.

Dette er gjort ved at hver observasjon i populasjonen er gitt et tilfeldig tall mellom 0 og 1 ved bruk av funksjonen *Tilfeldig()* i Excel. Observasjonene er deretter sortert med hensyn på dette tallet, og det tilfeldige utvalget består av de 253 observasjonene som da ble liggende øverst i regnearket.

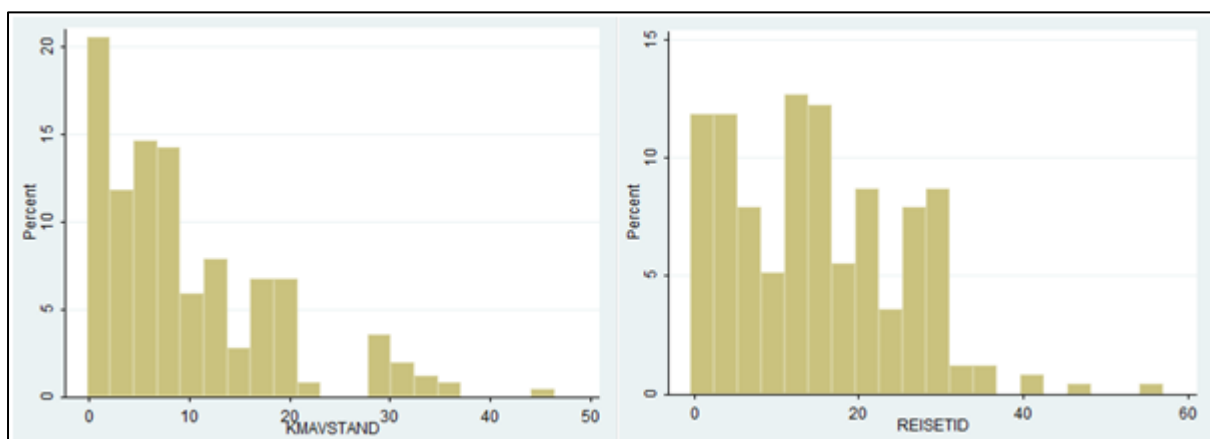
Det tilfeldige utvalget representerer den fulle populasjonen godt med hensyn til fordeling på kommuner. I denne oppgaven blir imidlertid arbeidsmarkedet i Kristiansandsregionen vurdert som ett felles arbeidsmarked, med hovedtyngde i Kristiansand sentrum. Det vil derfor ikke bli undersøkt om det er noen forskjeller i boligpris i de forskjellige kommunene uavhengig av distanse fra Kristiansand sentrum.

Samtlige boliger i det tilfeldige utvalget er undersøkt med hensyn til karakteristikkene som ble presentert i kapittel 5.1. Funnene kan oppsummeres i følgende tabell:

Tabell 6: Oppsummeringstabell over undersøkte boligkarakteristikk

variable	Obs	Mean	Std. Dev.	Min	Max
pris	253	2507.14	1382.608	850	13235
boa	253	108.0988	52.53776	19	345
alder	253	32.92885	28.04791	0	223
reisetid	253	14.98814	10.29553	1	54.5
kmapstand	253	9.834387	8.799466	1	44.3
sjø	253	.0079051	.0887343	0	1

Oppsummeringstabellen over viser blant annet gjennomsnittsverdiene til de observerte boligkarakteristikkene. Vi ser at gjennomsnittverdien til både reisetid og avstand i km ligger mye nærmere minimumsverdien enn maksimumsverdien. Dette kan tyde på at vi har relativt sett flere observasjoner av solgte boliger nærme sentrum enn i utkanten av regionen. Dette kan også illustreres ved å lage et histogram hvor høyden på søylene symboliserer andel av de totale observasjonene, mens den vertikale aksene beskriver avstand fra sentrum.



Figur 11: Histogram, observasjoner fordelt etter sentrumsavstand

Denne figuren illustrerer også godt at det er en forskjell i de to måtene å beregne avstand fra sentrum på. Hver søyle representerer et 5 % intervall av den observerte maksimumsverdien, slik at søylen helt til venstre indikerer antall observasjoner av boliger som er lokalisert mindre enn 5 % av den høyeste observerte avstanden fra sentrum. Den neste søylen indikerer antall observasjoner som ligger mellom 5 % og 10 % av maksverdien, og så videre.

Til venstre i figur 11 ser vi at når avstanden fra sentrum beregnes i kilometer, vil andelen av observasjonene synke drastisk når vi beveger oss lenger bort fra sentrum. Til høyre i figuren ser vi derimot at når avstanden beregnes i reisetid i rushtrafikk, vil de samme boligene være relativt jevnt fordelt hvis vi ser bort fra de mest ekstreme observasjonene. Årsaken til dette antas å være reisetidsforsinkelser som kun oppstår nærme sentrum på grunn av rushtrafikkproblemer. Dette slår relativt sett mye hardere ut på reisetiden til folk som bor nærme sentrum, enn på reisetiden til folk som bor lenger ute i regionen.

Noe annet vi også kan observere av oppsummeringstabellen, er at gjennomsnittsverdien til dummyvariabelen *sjø* er veldig nærme null. Dette kommer av at det i det tilfeldige utvalget kun er to observasjoner av eneboliger med egen strandlinje. Det er vurdert å fjerne variabelen fra regresjonsmodellene, men disse to boligene ligger i en helt egen prisklasse og dette antas å hovedsakelig skyldes nettopp denne egenskapen. Ved å fjerne variabelen vil disse to boligene få en uforklarlig høy observert salgspris, noe som vil føre til en betydelig reduksjon av modellenes forklaringskraft, samtidig som også de estimerte koeffisientene til de resterende variablene kan bli påvirket.

Det er også vurdert å fjerne disse to observasjonene fra hele datasettet, men dette er heller ikke funnet hensiktsmessig. Boliger av denne typen er absolutt en del av boligmarkedet i Kristiansandsregionen, og det er ikke funnet noen tungtveiende argumenter som kan rettferdiggjøre at disse to boligene skal fjernes fra datasettet. Variabelen beholdes derfor i modellene, men ved tolkning av den estimerte koeffisienten til *sjø* bør man ha i bakhodet at dette estimatet baserer seg på kun to observasjoner og den nøyaktige verdien til estimatet bør derfor ikke ilegges for stor vekt.

## 6. Analyse

### 6.1 Korrelasjon mellom kvadratmeterpris og avstand fra sentrum

Vår analyse begynner enkelt og greit med å presentere en korrelasjonsmatrise for de involverte variabler.

Tabell 7: Korrelasjonsmatrise over variablene i modellen

	pris	boa	alder	reisetid	kmavst~d	sjø
pris	1.0000					
boa	0.6468	1.0000				
alder	0.0153	0.0281	1.0000			
reisetid	-0.1271	0.2878	-0.3208	1.0000		
kmavstand	-0.1041	0.2504	-0.2808	0.9269	1.0000	
sjø	0.5732	0.2058	-0.0412	-0.0216	-0.0502	1.0000

Korrelasjonsmatrisen viser en positiv lineær korrelasjon mellom prisene til boligene i utvalget og deres respektive boareal, noe som er helt i tråd med hva man forventet på forhånd. Noe uventet derimot, viser korrelasjonsmatrisen ingen lineær sammenheng mellom boligprisene og boligenes alder. Boligprisene har en svak negativ lineær korrelasjon med avstanden fra sentrum, målt både i tid og i kilometer. Mens til sist ser vi at det er relativt kraftig positiv lineær korrelasjon mellom prisen og hvorvidt boligen er en enebolig med egen strandlinje.

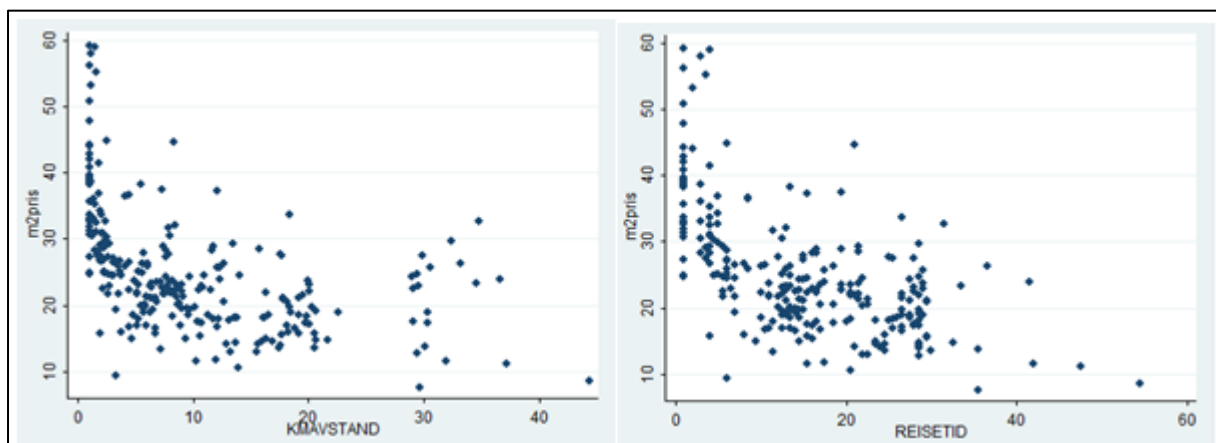
Dersom vi nå går videre og ser på den lineære korrelasjonen mellom de forskjellige uavhengige variablene, ser vi blant annet at boareal har en moderat positiv korrelasjon med avstanden fra sentrum. Dette gjelder når avstanden er målt både i tid og i kilometer, og kan tyde på at boligene blir større ettersom man beveger seg bort fra sentrum. Dette er interessant med tanke på den lineære korrelasjonen mellom observerte priser og boareal, og den lineære korrelasjonen mellom priser og avstand fra sentrum. Siden boligene blir større når man beveger seg bort fra sentrum, vil dette isolert sett medføre høyere priser når man beveger seg bort fra sentrum. I følge teori presentert tidligere i oppgaven skal imidlertid prisene falle ettersom avstanden til sentrum blir større, men det kan altså se ut som om

dette til en viss grad blir motvirket av høyere priser på grunn av en økning i boareal. For å forsøke å isolere effekten av avstanden til sentrum, kan det derfor være hensiktsmessig å dele boligenes pris på antall kvadratmeter boareal. På denne måten kan vi undersøke om det er noen sammenheng mellom prisen man betaler per kvadratmeter boareal og avstand mellom boligen og sentrum.

Tabell 8: Lineær korrelasjon mellom kvadratmeterpris og avstand fra sentrum

	m2pris	kmavst~d	reisetid
m2pris	1.0000		
kmavstand	-0.4870	1.0000	
reisetid	-0.5951	0.9269	1.0000

Vi ser her at vi får frem en stor grad av negativ lineær korrelasjon mellom kvadratmeterpris og sentrumsavstand målt i kilometer, og en enda større grad av lineær korrelasjon når denne avstanden måles i reisetid. Hvordan kvadratmeterprisen faller når man beveger seg ut fra sentrum, kan illustreres ved å plote de observerte kvadratmeterprisene mot avstanden fra sentrum.



Figur 12: Plott, kvadratmeterpris mot avstand fra sentrum

Ved å studere plottene i figur 12 ser vi en tydelig trend ved at kvadratmeterprisen er lavere jo lengre bort fra sentrum vi befinner oss. Plottene viser også noe mer, som korrelasjonsmatrisen ikke kunne vise oss. Nemlig at korrelasjonen mellom kvadratmeterpris

og sentrumsavstand, da særlig målt i kilometer, følger en mer logaritmisk trend. Dette indikerer at kvadratmeterprisen faller veldig raskt i starten når man beveger seg ut fra sentrum målt i kilometer, for så å falle mindre og mindre ettersom man kommer lenger bort fra sentrum. Befinner man seg langt borte fra sentrum vil altså en marginal endring i avstand ha svært liten effekt på salgsprisen.

Det er nå vist hvordan den lineære korrelasjonskoeffisienten  $r$  ikke nødvendigvis gir et fullstendig bilde av den reelle korrelasjonen mellom den avhengige og de uavhengige variablene når det eksisterer lineær korrelasjon mellom to eller flere av de uavhengige variablene. Vi ser også i korrelasjonsmatrisen i tabell 7 at det blant annet er negativ korrelasjon mellom avstand fra sentrum og alder, noe som sannsynligvis er hovedårsaken til at man ikke finner noen lineær korrelasjon mellom pris og alder.

## 6.2 Lineær regresjonsanalyse

### 6.2.1 Lineær funksjonsform med avstand målt i kilometer

Vår regresjonsanalyse begynner med å presentere en modell med en lineær underliggende funksjonsform hvor sentrumsavstanden er målt i antall kilometer fra sentrum.

Source	SS	df	MS			
Model	325298194	4	81324548.5	Number of obs =	253	
Residual	156426551	248	630752.221	F( 4, 248) =	128.93	
				Prob > F =	0.0000	
				R-squared =	0.6753	
				Adj R-squared =	0.6700	
				Root MSE =	794.2	
Total	481724745	252	1911606.13			

pris	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
boa	16.41079	1.01734	16.13	0.000	14.40706	18.41451
alder	-2.756368	1.87566	-1.47	0.143	-6.450623	.9378867
kmavstand	-39.98012	6.201665	-6.45	0.000	-52.19477	-27.76547
sjø	6696.298	581.5695	11.51	0.000	5550.852	7841.743
_cons	1164.163	137.0273	8.50	0.000	894.2768	1434.048

Figur 13: Regresjon, lineær funksjonsform med avstand målt i kilometer

Regresjonen viser at selv ved bruk av kun fire uavhengige variabler, kan modellen forklare over 67 % av variasjonen i omsetningspris til boligene i datasettet. Dette må i aller høyeste grad sies å være akseptabelt. De estimerte koeffisientene har alle forventet fortegn og fornuftige størrelser. Vi ser imidlertid at den estimerte koeffisienten til variabelen *alder* ikke er signifikant på 95 % nivå. Dette betyr at vi på statistisk grunnlag ikke kan være minst 95 % sikker på at den virkelige koeffisienten er forskjellig fra null. Det er likevel slik at all fornuft tilsier at en boligs verdi faller med økt alder, så dette får ingen konsekvenser for vår modell.

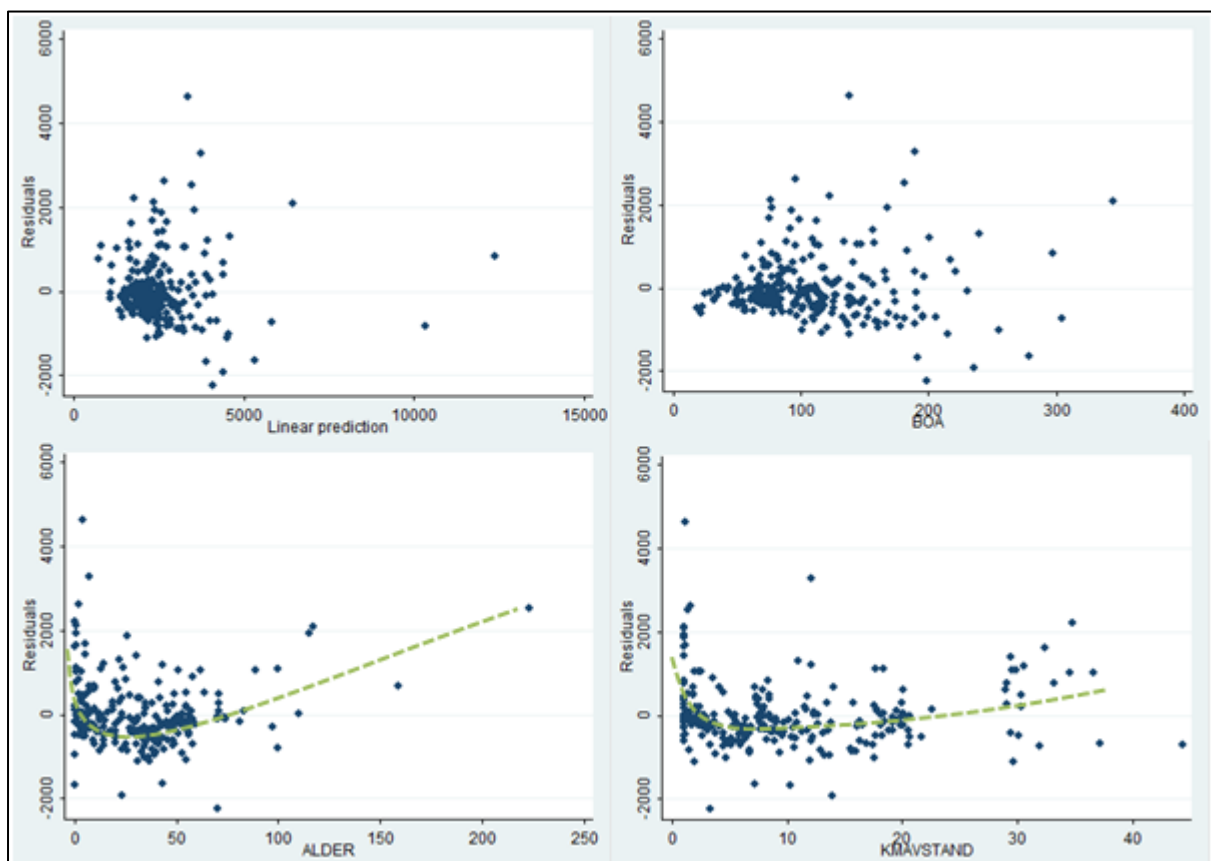
Den estimerte regresjonsmodellen har et konstantledd på 1164,163. Dette indikerer den estimerte prisen i 1000 kr til en hypotetisk bolig som ligger nøyaktig i sentrum, er helt ny, har null boareal og som heller ikke er en enebolig med egen strandlinje. Merk at dette er en fullstendig teoretisk bolig som er vanskelig å se for seg i praksis. Vi har heller ikke noen observasjoner av boliger av denne typen, og den ligger derfor utenfor modellens relevante



område. Denne hypotetiske boligen er likevel interessant, fordi den danner grunnlaget for estimeringen av boligene i det relevante området.

Denne modellen estimerer en boligs verdi i tusen kroner, ved å ta utgangspunkt i konstantleddet for så å legge til  $16,41 \times$  boligens boareal, trekke fra  $2,76 \times$  boligens alder i år, trekke fra  $39,98 \times$  boligens avstand fra sentrum i kilometer og legge til  $6696,30$  dersom boligen er en enebolig med egen strandlinje. På denne måten ser vi at modellen estimerer at en ellers lik bolig faller i verdi med  $39\,980$  kroner for hver kilometer den ligger utenfor Kvadraturen.

De klassiske forutsetningene som skal til for at estimering med minste kvadraters metode skal være BLUE, blir ikke fullstendig gjennomgått her. Vi nøyer oss med å presentere plott av residualene mot den lineære prediksjonen og mot de uavhengige variablene. Det blir ikke plottet mot variabelen *sjø*, da vi kun har to observerte boliger med denne egenskapen.



Figur 14: Residualplott, lineær funksjonsform med avstand målt i kilometer

Residualplottene i figur 14 avslører at variasjonen i residualene ser ut til å øke ettersom den estimerte prisen blir høyere. Det ser også ut som om variasjonen i residualene øker med økt boareal. Når vi ser på plottet nederst til venstre i figuren, kan det se ut som om residualene varierer omkring en slags liggende J-kurve, illustrert ved den stiplede linjen. Dette indikerer at modellen underestimerer nye boliger og svært gamle boliger, mens boligene i mellom til en viss grad blir overestimert. Det må understrekes at dette ikke er noen tydelig trend, men residualplottet synes å vise svake indikasjoner på dette mønsteret. Man kan også ane en svak kontur av en liggende J-kurve i plottet nederst til høyre, her kanskje enda svakere enn i residualplottet for alder. Dette gir også noen små indikasjoner på at modellen underestimerer sentrumsnære boliger og boliger som ligger svært langt fra sentrum, mens boligene i mellom til en viss grad blir overestimert.

Det ble i forrige delkapittel, angående korrelasjon, pekt på at det er en viss grad av lineær korrelasjon mellom noen av de uavhengige variablene. Denne korrelasjonen er imidlertid ikke av et slikt omfang at den skulle tilsa at vår modell har noen alvorlige problemer med hensyn til multikollinearitet. For sikkerhets skyld er det gjennomført VIF-test (Variance Inflating Factor-test) av alle regresjonsmodellene i denne oppgaven som har lineær eller dobbeltlogaritmisk funksjonsform, og ingen tyder på noen større grad av multikollinearitet. Fremgangsmåte og resultater av VIF-testene presenteres imidlertid ikke her, men kan finnes i vedlegg 1 bakerst i oppgaven.

## 6.2.2 Lineær funksjonsform med avstand målt i reisetid

Analysen fortsetter nå med en presentasjon av en nesten identisk regresjonsmodell. Vi har den samme lineære underliggende funksjonsformen, og vi har de samme uavhengige variablene med unntak av variabelen som beskriver avstand fra sentrum. Vi har nå byttet ut avstanden målt i kilometer med avstanden målt i reisetid i rushtrafikk. Figur 15 oppsummerer denne nye estimerte regresjonsmodellen.

Source	SS	df	MS			
Model	<b>342989495</b>	<b>4</b>	<b>85747373.8</b>	Number of obs =	<b>253</b>	
Residual	<b>138735250</b>	<b>248</b>	<b>559416.329</b>	F( 4, 248) =	<b>153.28</b>	
				Prob > F =	<b>0.0000</b>	
				R-squared =	<b>0.7120</b>	
				Adj R-squared =	<b>0.7074</b>	
Total	<b>481724745</b>	<b>252</b>	<b>1911606.13</b>	Root MSE =	<b>747.94</b>	

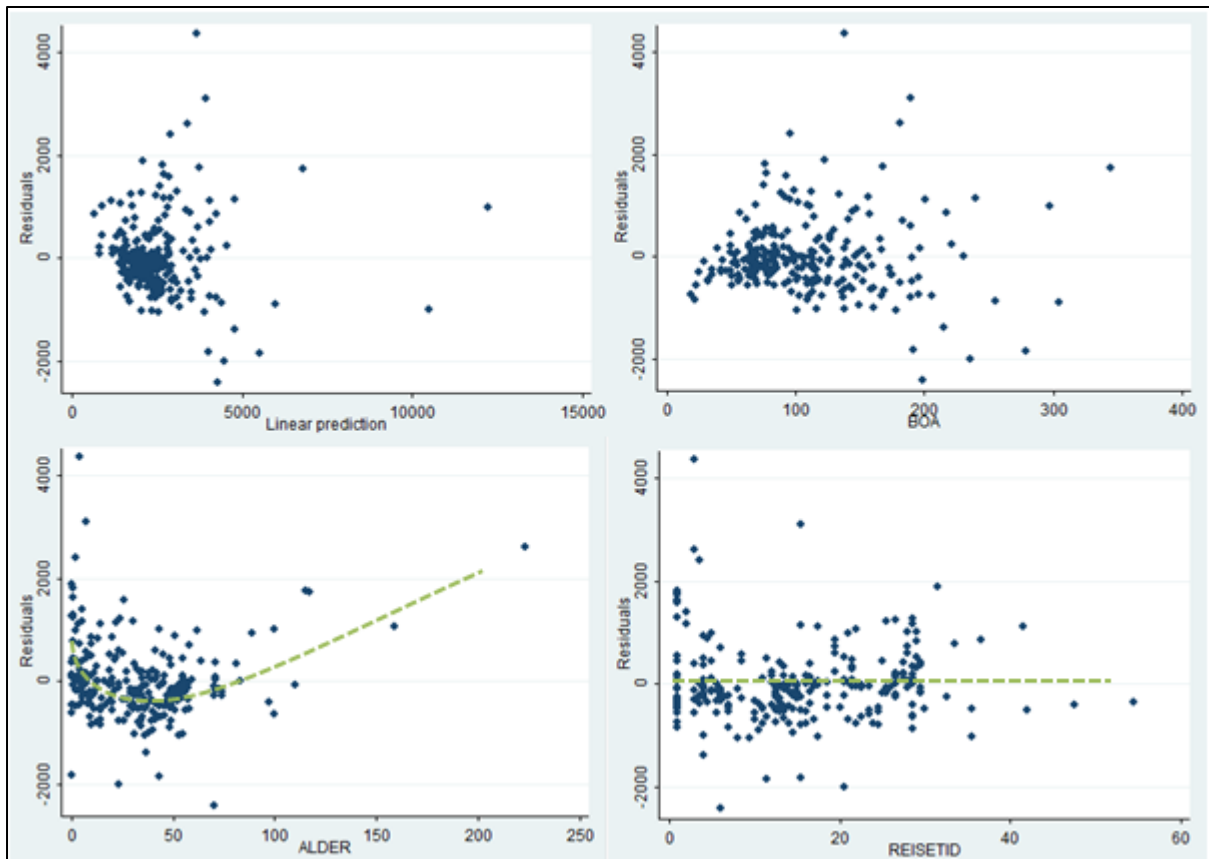
  

pris	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
boa	<b>17.33809</b>	<b>.9709546</b>	<b>17.86</b>	<b>0.000</b>	<b>15.42572</b>	<b>19.25046</b>
alder	<b>-4.630189</b>	<b>1.795596</b>	<b>-2.58</b>	<b>0.010</b>	<b>-8.166751</b>	<b>-1.093626</b>
reisetid	<b>-45.35033</b>	<b>5.119062</b>	<b>-8.86</b>	<b>0.000</b>	<b>-55.43271</b>	<b>-35.26795</b>
sjø	<b>6643.985</b>	<b>546.5135</b>	<b>12.16</b>	<b>0.000</b>	<b>5567.585</b>	<b>7720.385</b>
_cons	<b>1412.576</b>	<b>136.3018</b>	<b>10.36</b>	<b>0.000</b>	<b>1144.119</b>	<b>1681.032</b>

Figur 15: Regresjon, lineær funksjonsform med avstand målt i reisetid

Vi ser at forklaringskraften til denne relativt enkle modellen er på over 71 %. De estimerte koeffisientene har forventet fortegn, og fornuftige størrelser. Konstantleddet er nå litt høyere enn det var i den forrige modellen, mens den estimerte prisen nå reduseres litt mer ved en økning i boligens alder. Koeffisientene til *boareal* og *sjø* er tilnærmet uendret fra den forrige modellen. Denne modellen, med underliggende lineær funksjonsform og sentrumsavstanden målt i reisetid, estimerer at prisen til en bolig vil reduseres med 45 350 kr for hvert minutt boligen ligger utenfor sentrum i reisetid i rushtrafikk.

Det vil for denne modellen, i likhet med den foregående, bli studert plott av residualene mot den lineære prediksjonen og de uavhengige variablene, med unntak av variabelen *sjø*.



Figur 16: Residualplott, lineær funksjonsform med avstand målt i reisetid

Residualplottene viser nesten det samme som plottene i den foregående modellen. Vi ser at variansen til residualene ser ut til å øke ettersom boligene får høyere estimert salgsverdi. Dette stemmer overens med det vi kan observere i de andre residualplottene, nemlig at residualenes varians ser ut til å øke ettersom boareal øker, og synke med økt alder og økt reisetid. Dette vil si at variansen i residualene er størst for relativt nye boliger, boliger som har stort boareal og for boliger som er sentrumsnære. Alle disse egenskapene karakteriserer nettopp boliger som har høy estimert salgspris.

Som vi ser er det også for disse residualplottene tegnet inn en stiplet kurve i de to nederste plottene, på samme måte som for foregående modell. Vi legger imidlertid merke til at residualene plottet mot reisetid ikke ser ut til å følge den samme liggende J-konturen som de gjorde i foregående modell, når avstanden ble målt i kilometer. Nå ser de mer ut til å variere omkring en mer rett linje.

### 6.2.3 Dobbellogaritmisk funksjonsform med avstand målt i kilometer

Denne delen av analysen tar for seg en regresjonsmodell med de samme variablene som i de foregående modellene, men hvor den underliggende funksjonformen er dobbellogaritmisk. Modellen med sentrumsavstand målt i kilometer oppsummeres her i figur 17.

Source	SS	df	MS			
Model	32.4745294	4	8.11863236	Number of obs =	253	
Residual	11.2917477	248	.045531241	F( 4, 248) =	178.31	
				Prob > F =	0.0000	
				R-squared =	0.7420	
				Adj R-squared =	0.7378	
				Root MSE =	.21338	
Total	43.7662771	252	.173675703			

lnpris	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnboa	.7339773	.0309168	23.74	0.000	.6730842	.7948703
lnalder	-.08227	.012188	-6.75	0.000	-.1062751	-.0582648
lnkmavstand	-.1934548	.0145701	-13.28	0.000	-.2221517	-.1647579
sjø	.838642	.1548091	5.42	0.000	.5337338	1.14355
_cons	4.97165	.1381412	35.99	0.000	4.699571	5.24373

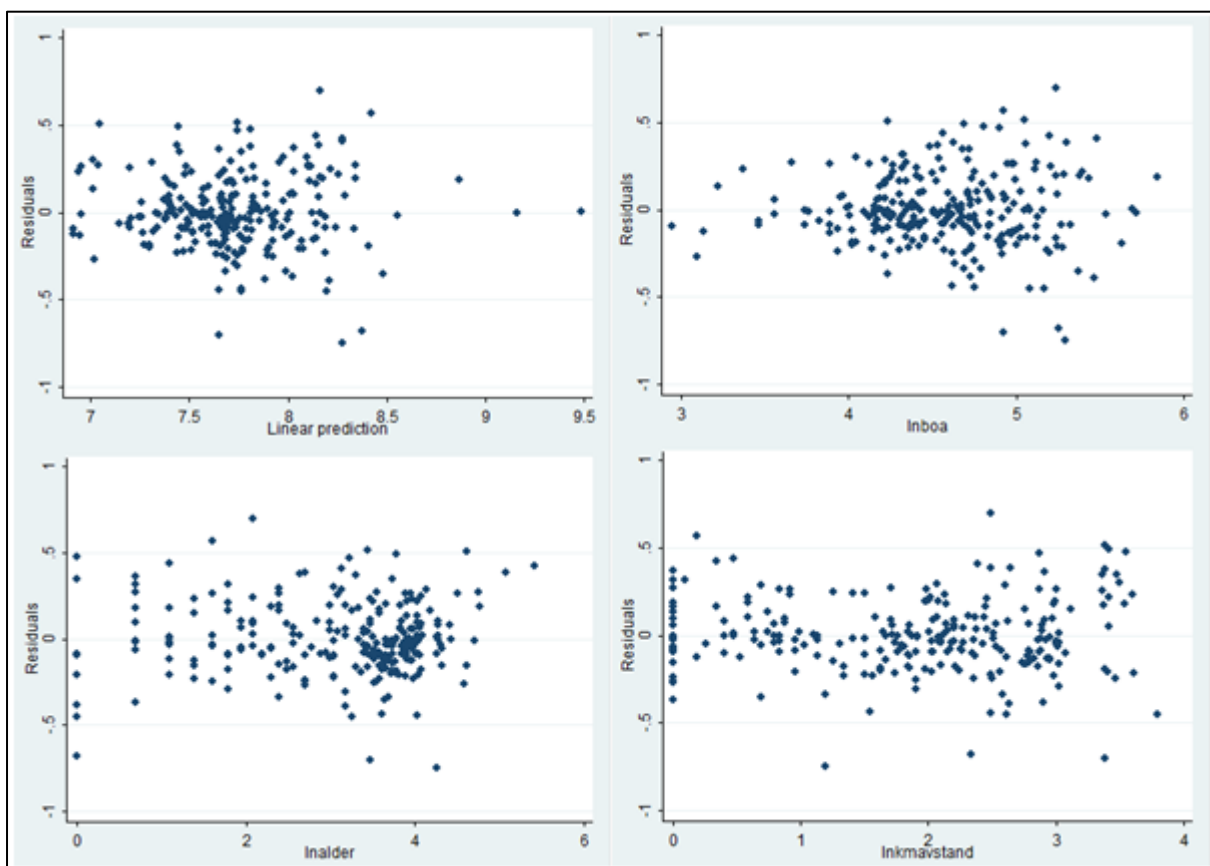
Figur 17: Regresjon, dobbellogaritmisk funksjonsform med avstand målt i kilometer

Vi ser at denne modellen, med en  $R^2$  på 0,742, forklarer variasjonene i salgspris godt. De estimerte koeffisientene har forventet fortegn, og størrelsene ser fornuftige ut. I en modell av denne typen, med dobbellogaritmisk funksjonsform, kan de estimerte koeffisientene tolkes som elastisiteter. Tolkningen av de estimerte  $\beta$ -koeffisientene til de uavhengige variablene oppsummeres i tabellen under.

Tabell 9: Tolkning av de estimerte  $\beta$ -koeffisientene, dobbellogaritmisk funksjonsform med sentrumsavstand målt i kilometer

Variabel	$\hat{\beta}$	Tolkning
<i>lnboa</i>	0,7340	Den estimerte salgsprisen øker med 0,734 % dersom boareal økes med 1 % og alt annet holdes konstant
<i>lnalder</i>	-0,0823	Den estimerte salgsprisen reduseres med 0,0823 % dersom alder økes med 1 % og alt annet holdes konstant
<i>lnkmavstand</i>	-0,1935	Den estimerte salgsprisen reduseres med 0,1935 % dersom avstanden i kilometer økes med 1 % og alt annet holdes konstant
<i>sjø</i>	0,8364	Den estimerte salgsprisen øker med 83,64 % dersom den estimerte boligen er en enebolig med egen strandlinje

Også for denne modellen er det plottet residualer mot den lineære prediksjonen og de uavhengige variablene. Av figuren under ser vi at variansen til residualene ser ut til å være mye mindre varierende enn i de to foregående modellene. Residualene plottet mot *lnboa* viser en svak økning i varians ettersom *lnboa* øker, noe som til en viss grad gjenspeiler seg i plottet med residualene plottet mot den lineære prediksjonen ettersom stort boareal gjerne sammenfaller med høy estimert pris. De øvrige residualplottene viser ingen tydelige tegn til avvik fra konstant varians. Man kan heller ikke finne noen tydelige tegn til andre avvik fra tilfeldig spredning i noen av residualplottene.



Figur 18: Residualplott, dobbeltlogaritmisk modell med avstand målt i kilometer

## 6.2.4 Dobbellogaritmisk funksjonsform med avstand målt i reisetid

Dette er i utgangspunktet nesten samme modell som ble beskrevet i forrige delkappitel. Vi har samme funksjonsform og de samme variablene, bortsett fra at avstandsvariabelen nå er målt i reisetid i rushtrafikk i stedet for i antall kilometer kjøredistanse. Modellen oppsummeres i figuren under.

Source	SS	df	MS			
Model	32.3684008	4	8.09210019	Number of obs =	253	
Residual	11.3978764	248	.045959179	F( 4, 248) =	176.07	
Total	43.7662771	252	.173675703	Prob > F =	0.0000	
				R-squared =	0.7396	
				Adj R-squared =	0.7354	
				Root MSE =	.21438	

lnpris	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnboa	.7307186	.030995	23.58	0.000	.6696716	.7917655
lnalder	-.0718155	.0120659	-5.95	0.000	-.0955801	-.0480509
lnreisetid	-.1883035	.0143438	-13.13	0.000	-.2165546	-.1600524
sjø	.9381384	.1547018	6.06	0.000	.6334414	1.242835
_cons	5.038067	.138094	36.48	0.000	4.76608	5.310054

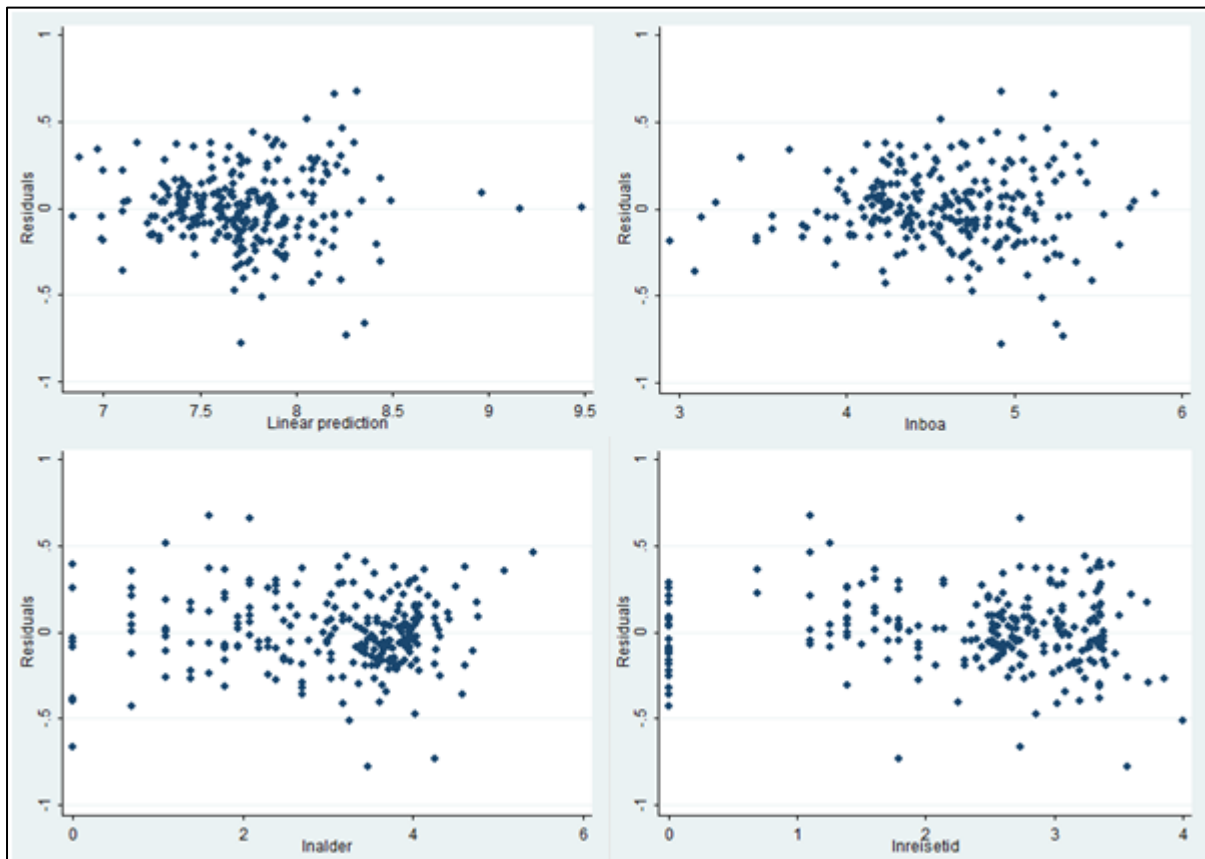
Figur 19: Regresjon, dobbellogaritmisk funksjonsform med avstand målt i reisetid

Vi ser av figuren over at denne modellen, i likhet med den foregående, forklarer ca. 74 % av variasjonen i salgspris. De estimerte koeffisientene har alle forventet fortegn og har også verdier i områdene som man skulle forvente. På samme måte som ved foregående modell, blir tolkningen av de estimerte koeffisientene til de uavhengige variablene presentert i en egen tabell.

Tabell 10: Tolkning av de estimerte  $\beta$ -koeffisientene, dobbellogaritmisk funksjonsform med sentrumsavstand målt i reisetid i rushtrafikk

Variabel	$\hat{\beta}$	Tolkning
<i>lnboa</i>	0,7307	Den estimerte salgsprisen øker med 0,7307 % dersom boareal økes med 1 % og alt annet holdes konstant
<i>lnalder</i>	-0,0718	Den estimerte salgsprisen reduseres med 0,0718 % dersom alder økes med 1 % og alt annet holdes konstant
<i>lnreisetid</i>	-0,1883	Den estimerte salgsprisen reduseres med 0,1883 % dersom reisetiden i minutter økes med 1 % og alt annet holdes konstant
<i>sjø</i>	0,9381	Den estimerte salgsprisen øker med 93,81 % dersom den estimerte boligen er en enebolig med egen strandlinje

Tilsvarende som for de andre modellene, er det også her plottet residualer mot den lineære prediksjonen og de uavhengige variablene, med unntak av variabelen *sjø*. Vi har følgende figur som viser de forskjellige residualplottene:



Figur 20: Residualplott, dobbeltlogaritmisk funksjonsform med avstand målt reisetid

Vi ser at også her har variansen til residualene en liten tendens til øke for boliger med større boareal, men alt i alt er variansen i residualene mye mer stabil enn i de lineære modellene. Vi ser heller ikke her noen tydelige tegn til avvik fra tilfeldig spredning av residualene. Det kan imidlertid se ut som boligene som ligger rett utenfor Kvadraturen i reisetid, har en tendens til å bli underestimert. På grunn av både rushtidsforsinkelser og måten funksjonen  $\ln(X)$  er definert, har vi imidlertid svært få observasjoner av  $\ln(\text{reisetid})$  i dette området.



### 6.2.5 Kvadratisk funksjonsform med avstand målt i antall kilometer

Basert på våre erfaringer med de to foregående funksjonsformene og korrelasjonsanalysen, kan det nå være interessant å forsøke en modell med kvadratisk funksjonsform. Korrelasjonsanalysen indikerte at den negative effekten av økt sentrumsavstand ikke fulgte en rett linje, men at den heller var konveks. Residualplottetene i regresjonsanalysen med lineær underliggende funksjonsform viste også tegn til avvik fra tilfeldig spredning av residualene plottet mot sentrumsavstand, noe som styrker mistanken om at den reelle sammenhengen ikke er lineær. Det er derfor valgt å kvadrere avstandsvariabelen for å prøve å fange opp konveksiteten som korrelasjonsanalysen antyder. I modellene med lineær funksjonsform viste også residualplottene for boligens alder avvik fra tilfeldig spredning, noe som antyder at heller ikke her er den reelle sammenhengen lineær. Det antas at den virkelige effekten av økt alder på boligprisen er negativ, men avtagende ved en marginal økning i alder. Dette betyr altså at man antar at en økning i alder har større negativ effekt på prisen til relativt nye boliger enn det har på gamle boliger. Av denne årsak er også variabelen *alder* kvadrert.

Når det gjelder variabelen *boa*, viste residualplottene til modellene med lineær funksjonsform noen små tegn til avvik fra tilfeldig spredning. Det var imidlertid svært vanskelig å se noen tydelig trend siden variansen økte relativt mye med økt boareal. I teorien som ble gjennomgått i forbindelse med den hedonistiske metoden ble det for øvrig antatt at priseffekten av en økning i boligattributt av denne typen er avtagende for en marginal økning i attributtet. Det synes også å virke fornuftig at verdien til en ekstra kvadratmeter boareal er større dersom man har veldig få kvadratmeter boareal fra før enn dersom man allerede har svært stor bolig. Det antas derfor at kurven som beskriver priseffekten av en boligs boareal er stigende og konkav. Det er derfor også valgt å kvadrere variabelen *boa*.

Figuren på neste side oppsummerer regresjonen av denne modellen med kvadratisk funksjonsform og som har sentrumsavstanden målt i antall kilometer. Variablene som har et 2-tall på slutten av variabelnavnet indikerer at dette er en uavhengig variabel som er kvadrert, slik at for eksempel *kmavstand2* er den opprinnelige uavhengige variabelen *kmavstand* opphøyd i andre potens.

Source	SS	df	MS			
Model	359642057	7	51377436.8	Number of obs =	253	
Residual	122082687	245	498296.683	F( 7, 245) =	103.11	
				Prob > F =	0.0000	
				R-squared =	0.7466	
				Adj R-squared =	0.7393	
				Root MSE =	705.9	
Total	481724745	252	1911606.13			

pris	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
boa	18.00863	3.185428	5.65	0.000	11.73431	24.28294
boa2	-.0044925	.0105716	-0.42	0.671	-.0253153	.0163303
alder	-22.36557	3.18923	-7.01	0.000	-28.64738	-16.08376
alder2	.143961	.0216864	6.64	0.000	.1012455	.1866766
kmavstand	-122.7198	16.50195	-7.44	0.000	-155.2236	-90.21601
kmavstand2	2.432739	.4709956	5.17	0.000	1.505021	3.360456
sjø	6532.595	523.5369	12.48	0.000	5501.387	7563.802
_cons	1825.173	219.2696	8.32	0.000	1393.279	2257.067

Figur 21: Regresjon, kvadratisk funksjonsform med avstand målt i kilometer

Modellen, med en  $R^2$  på 0,7466, forklarer variasjonene i salgspris svært godt. Vi ser imidlertid at den estimerte koeffisienten til *boa2* er svært lav og at den heller ikke er signifikant. Det er vurdert å fjerne det kvadrerte leddet til *boa*, men det er besluttet å beholde modellen slik som den er beskrevet i figuren over. Det at den estimerte koeffisienten ikke er signifikant, betyr at vi basert på statistikk ikke kan være sikre på at den virkelige koeffisienten er forskjellig fra null. Vi har imidlertid både teori og fornuftige forventninger som tilsier at den virkelige koeffisienten ikke er lik null. Fortegnet til den estimerte koeffisienten, selv om den ikke er signifikant, stemmer også overens med både teorien og forventningene vi hadde før regresjonen ble gjennomført.

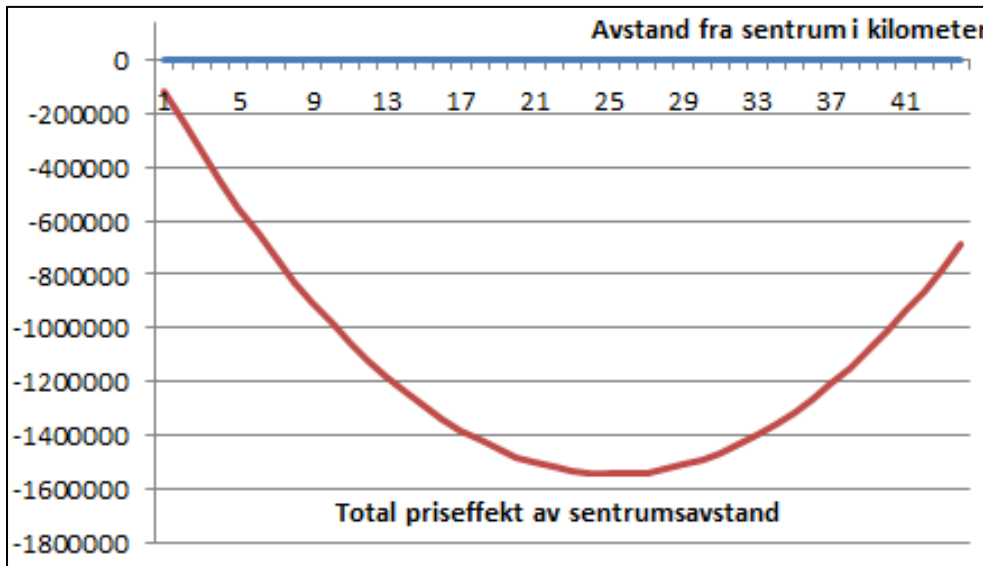
Vi skal videre nøye oss med å tolke koeffisientene til variablene *kmavstand* og *kmavstand2*. Tolkning av de resterende estimerte regresjonskoeffisientene ville for øvrig ha fulgt samme fremgangsmåte.

Som vi ser av figuren over, denne modellen estimerer følgende sammenheng mellom prisen til en bolig og dens avstand fra sentrum målt i kilometer:

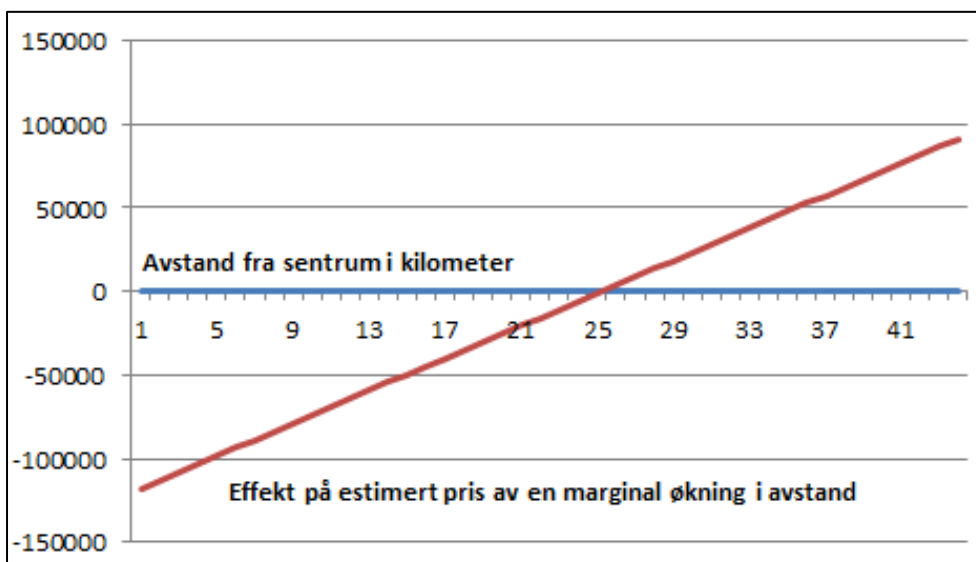
$$\widehat{Pris}_{avstand} = -122720 \times kmavstand + 2433 \times (kmavstand)^2 \quad (6.1)$$

En marginal økning i sentrumsavstand har dermed følgende effekt på den estimerte boligprisen:

$$\frac{\partial \widehat{Pris}}{\partial kmavstand} = -122720 + 4865 \times kmavstand \quad (6.2)$$



Figur 22: Kvadratisk funksjonsform, total priseffekt av sentrumsavstand målt i kilometer



Figur 23: Kvadratisk funksjonsform, effekt på estimert pris av en marginal økning i avstand målt i kilometer

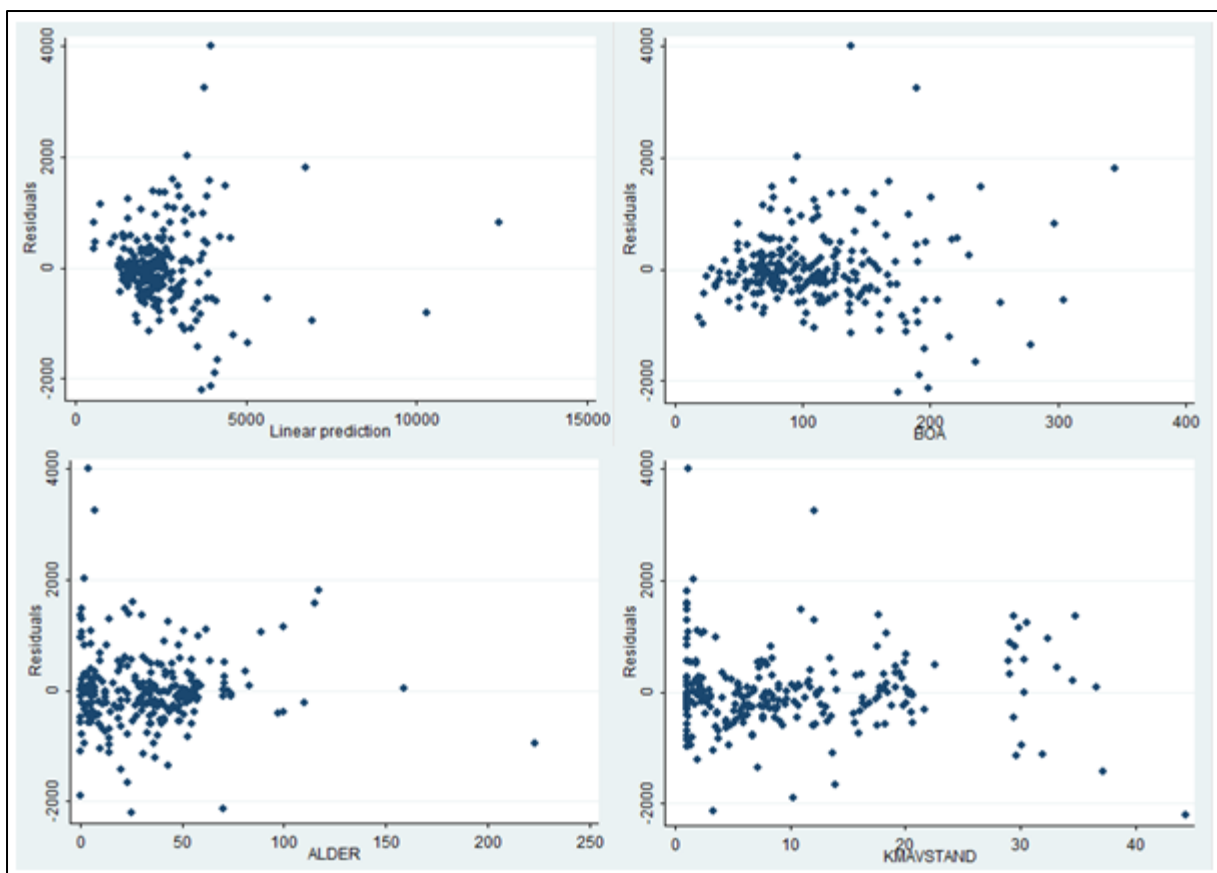
Av figurene over ser vi at denne modellen anslår at prisen til en bolig faller med ca. 120 000 kroner dersom vi befinner oss i sentrum og beveger oss en kilometer utover mens alt ellers holdes konstant. Prisen vil fortsette å reduseres ettersom vi flytter oss lenger og lenger bort

fra sentrum, men reduksjonen vil avta med 4865 kroner for hver kilometer. På denne måten ser vi at omkring 25 kilometer fra sentrum vil den estimerte effekten av økt sentrumsavstand ikke lenger være negativ. Det nøyaktige punktet kan finnes ved å sette likning 6.2 lik null, noe som gir oss at:

$$4865 \times kmavstand = 122720 \Rightarrow kmavstand = 25,23$$

Dette innebærer at vår modell med denne funksjonsformen estimerer at for boliger hvor sentrumsavstanden er over 25,23 kilometer, vil prisen øke dersom vi flytter oss enda lenger bort fra sentrum og alt annet holdes likt. Hvor rimelig dette egentlig er, blir behandlet i kapittel 6.3, som oppsummerer og diskuterer samtlige modeller i oppgaven.

Også for denne modellen blir det presentert residualplott. Figuren under viser at i likhet med modellene med lineær funksjonsform, har vi også her problemer med hensyn til forutsetningen om konstant varians. Særlig residualene plottet mot boareal ser ut til å ha en stor økning i varians ettersom mengden boareal blir større.



Figur 24: Residualplott, kvadratisk modell med avstand målt i kilometer

## 6.2.6 Kvadratisk funksjonsform med avstand målt i reisetid

Den siste modellen som blir presentert i denne analysen, har en kvadratisk funksjonsform og benytter reisetiden i rushtrafikk som mål på sentrumsavstand. Figuren under oppsummerer regresjonen av denne modellen.

Source	SS	df	MS			
Model	366556042	7	52365148.9	Number of obs = 253		
Residual	115168703	245	470076.337	F( 7, 245) = 111.40		
				Prob > F = 0.0000		
				R-squared = 0.7609		
				Adj R-squared = 0.7541		
				Root MSE = 685.62		
Total	481724745	252	1911606.13			

pris	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
boa	18.5354	3.089104	6.00	0.000	12.45081	24.61999
boa2	-.0054407	.0102568	-0.53	0.596	-.0256435	.0147621
alder	-22.36711	3.084697	-7.25	0.000	-28.44302	-16.2912
alder2	.1412028	.0210559	6.71	0.000	.0997292	.1826765
reisetid	-70.82024	12.59579	-5.62	0.000	-95.63009	-46.01039
reisetid2	.6093133	.3159734	1.93	0.055	-.0130575	1.231684
sjø	6662.123	507.299	13.13	0.000	5662.899	7661.347
_cons	1862.383	215.4412	8.64	0.000	1438.029	2286.736

Figur 25: Regresjon, kvadratisk funksjonsform med avstand målt i reisetid

Vi ser at denne modellen kan forklare over 76 prosent av variasjonen i de observerte salgspriene. Den estimerte koeffisienten til *boa2* er ikke signifikant, men det er besluttet å beholde leddet i regresjonslikningen av samme årsak som ble beskrevet i foregående modell. For øvrig kan vi også observere at den estimerte koeffisienten til *reisetid2* ikke er signifikant på 95 % nivå. Den er imidlertid signifikant på 90 % nivå, har forventet fortegn og all fornuft tilsier at den bør beholdes i regresjonslikningen. Vi skal som ved forrige modell, også her nøye oss med kun å tolke de estimerte koeffisientene til avstandsvariabelen.

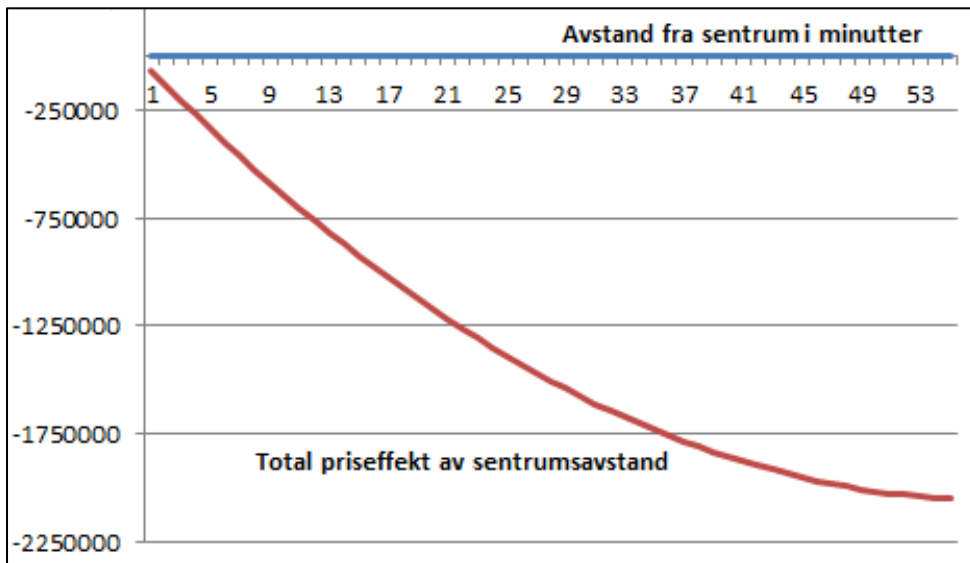
Denne modellen estimerer at en boligs avstand fra sentrum har følgende effekt på boligprisen:

$$\widehat{Pris}_{avstand} = -70820 \times reisetid + 609 \times (reisetid)^2 \quad (6.3)$$

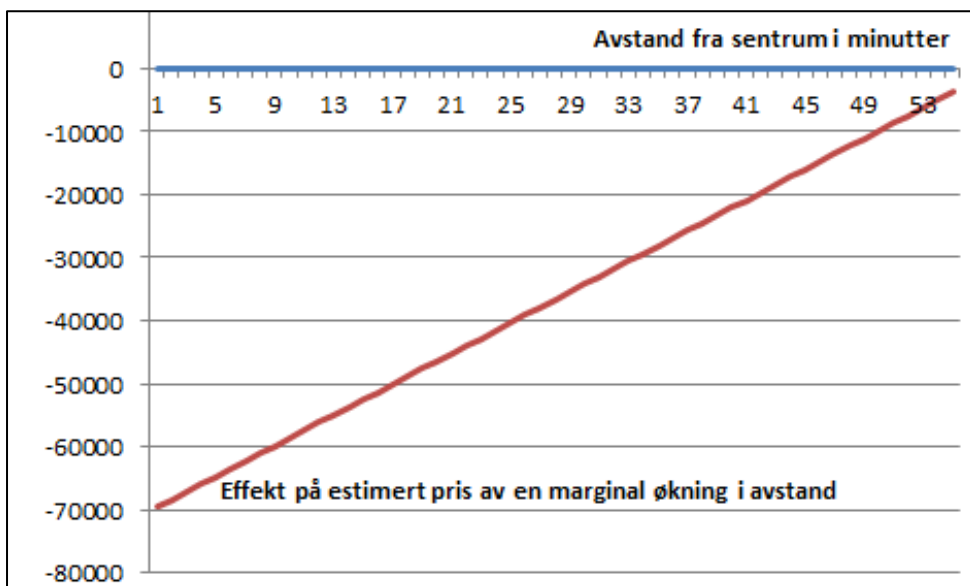
En marginal økning i sentrumsavstand har dermed følgende effekt på den estimerte boligprisen:

$$\frac{\partial \widehat{Pris}}{\partial reisetid} = -70820 + 1219 \times reisetid \quad (6.4)$$

Denne sammenhengen illustreres i figurene under.



Figur 26: Kvadratisk funksjonsform, total priseffekt av sentrumsavstand målt i reisetid

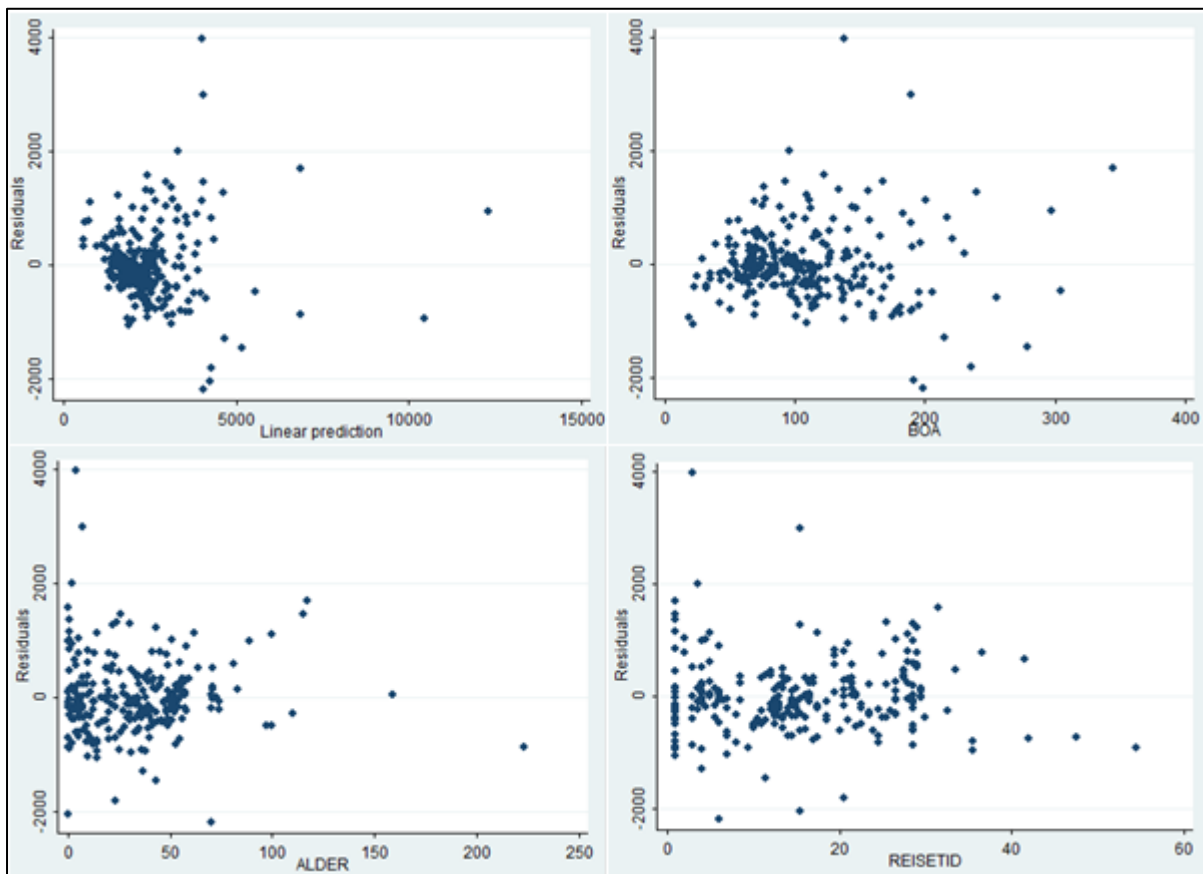


Figur 27: Kvadratisk funksjonsform, effekt på estimert pris av en marginal økning i avstand målt i reisetid

Vi ser av figur 27 at vår modell estimerer effekten av en ett minutt økning i reisetid til å være ca. 70 000 kroner for en bolig i sentrum, dersom alt annet holdes konstant. Den estimerte prisen vil falle ytterligere jo lenger bort fra sentrum vi beveger oss, men reduksjonen vil bli mindre og mindre helt til vi når punktet hvor  $\frac{\partial \widehat{pris}}{\partial reisetid} = 0$ . Dette punktet finner vi ved å sette likning 6.4 lik null, og løse likningen med hensyn på *reisetid*:

$$1219 \times reisetid = 70820 \quad \Rightarrow \quad reisetid = 58,1 \text{ minutter}$$

I og med at vår høyeste observerte avstand fra sentrum er på 54,5 minutter, ser vi at den estimerte effekten av økt sentrumsavstand på boligprisen er negativ men marginalt avtagende i hele det relevante området. Dette kan også observeres i figur 26 på forrige side.



Figur 28: Residualplott, kvadratisk funksjonsform med avstand målt i reisetid

Residualplottene for denne modellen viser at også her er det forekomst av heteroskedastisitet. Variansen til residualene ser ut til øke med økt med økt boareal, men

synke med økt alder og økt reisetid. Dette medfører at variansen til residualene er høyere for boliger med høy estimert salgpris, enn til boliger med lav estimert salgpris. Vi ser imidlertid ikke at noen spesielle boligtyper systematisk blir overestimert eller underestimert.



## 6.3 Diskusjon

Analysedelen i denne oppgaven startet med en korrelasjonsanalyse hvor den lineære korrelasjonen mellom de ulike variablene i datasettet ble undersøkt. Denne korrelasjonsanalysen synes tydelig å vise at det eksisterer en negativ korrelasjon mellom verdien til en bolig og dens avstand fra sentrum. Dette gjelder både om avstandene måles i kilometer eller om den måles i antall minutter reisetid i rushtrafikk.

For å prøve å tallfeste hvor stor negativ innvirkning en boligs avstand fra sentrum har på boligprisen, ble det gjennomført lineære regresjonsanalyser. Dette er tidligere blitt gjort for boligmarkedet i Kristiansand i flere masteroppgaver. Dosent Karl Robertsen og professor Theis Theisen ved Universitet i Agder, trekker særlig frem resultatene til Alfheim (2005) i et essay om boligmarkedet i Kristiansand (Robertsen & Theisen, 2010).

Alfheim benytter seg av regresjonsmodeller med både en lineær og en dobbeltlogaritmisk funksjonsform, og hvor variabelen som beskriver sentrumsavstand er målt i antall kilometer kjøredistanse mellom boligen og Kristiansand sentrum. Det er i denne oppgaven gjennomført regresjonsanalyser av boligmarkedet i Kristiansandsregionen med sentrumsavstanden målt både i antall kilometer, for å se i hvilken grad dette sammenfaller med Alfheims resultater, og i reisetid i rushtrafikk. Vi legger imidlertid merke til at Alfheim kun har sett på boliger i Kristiansand kommune, mens det i denne oppgaven er argumentert for at vi har et felles arbeids- og boligmarked i hele Kristiansandsregionen. Det er i denne oppgaven derfor også hensyntatt boligsalg i de omkringliggende kommunene.

Det er også slik at ikke alle de estimerte koeffisientene fra modellene i denne oppgaven kan sammenlignes direkte mot koeffisientene i Alfheims modeller. Alfheim har riktig nok benyttet seg av de samme forklaringsvariablene som er brukt i denne oppgaven med unntak av variabelen *sjø*, men har i tillegg benyttet seg av en rekke andre forklaringsvariabler som det antas at til en viss grad korrelerer med disse og dermed påvirker de estimerte verdiene. Det er likevel svært interessant å se på hvordan estimatene for effekt av sentrumsavstand i denne oppgaven, hvor modellene har relativt få forklaringsvariabler, sammenfaller med resultatene til Alfheim, som benytter modeller med langt flere forklaringsvariabler. Det er også interessant å se på i hvilken grad modellene med kun fire uavhengige variabler klarer å

forklare variasjonene i salgpris sett i forhold til Alfheims modell, som har fem ganger så mange forklaringsvariabler.

Vi starter med å sammenlikne modellene med en lineær underliggende funksjonsform. Her presenterer Alfheim først en full modell, som han så reduserer ned til en modell med 20 forklaringsvariabler. Tabell 11 oppsummerer de mest interessante parameterne i Alfheims modell sammenlignet med tilsvarende parametere fra modellene i denne oppgaven.

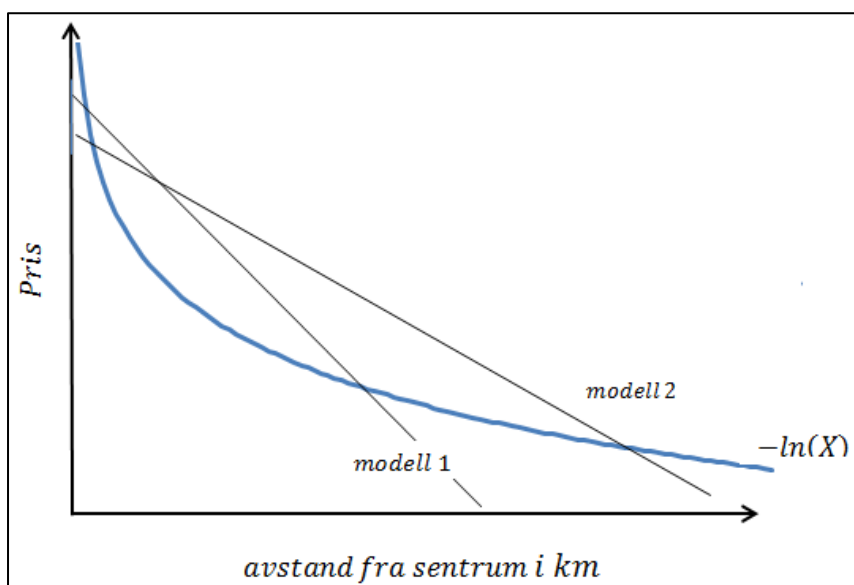
Tabell 11: Resultater fra lineær regresjonsanalyse med lineær funksjonsform, sammenlignet med Alfheim (2005)

	Alfheim(2005)	Vår modell, med avstand målt i antall kilometer	Vår modell, med avstand målt i reisetid i rushtrafikk
Antall forklaringsvariabler	20	4	4
Forklaringskraft, $R^2$	70,3 %	67,5 %	71,2 %
Effekt av avstand fra sentrum på estimert salgpris	-53 631 kr per km	-39 980 kr per km	-45 350 kr per minutt
Gjennomsnittlig observert avstand fra sentrum	5,4 km	9,8 km	15,0 minutter
Høyeste observerte avstand fra sentrum	19,4 km	44,3 km	54,5 minutter

Det første vi observerer ved å studere tabell 11, er at selv med kun fire forklaringsvariabler klarer de lineære modellene i denne oppgaven å forklare variasjonene i salgpris i tilnærmet samme grad som modellen til Alfheim. For modellen som benytter sentrumsavstand målt i reisetid i rushtrafikk i stedet for antall kilometer, ser vi til og med at  $R^2$  faktisk er litt høyere enn i hans modell.

Vi ser videre at den negative effekten av økt sentrumsavstand er noe høyere i Alfheims modell enn i vår modell med avstand målt kilometer. Forskjellen er ikke veldig stor, men den er verdt å merke seg. Alfheims modell er også fra 2005, og man må anta en viss stigning i det generelle prisnivået mellom 2005 og 2010, så forskjellen er enda litt større dersom man betrakter beløpet i 2010-kroner.

At det er en slik forskjell i den estimerte effekten av økt sentrumsavstand målt i antall kilometer, er likevel ikke spesielt uventet. Vår korrelasjonsanalyse viste at det er negativ korrelasjon mellom prisen til en bolig og dens avstand fra sentrum, men den antydte også at denne sammenhengen ikke var lineær. Særlig når avstanden ble målt i antall kilometer kunne det se ut som om sammenhengen fulgte en mer logaritmisk kurve, ved at den observerte kvadratmeterprisen falt drastisk i starten når man beveget seg ut fra sentrum, for så å falle mindre og mindre. Nettopp en slik negativ logaritmisk kurve er illustrert nedenfor i figur 29, hvor det også er tegnet inn to lineære approksimasjoner av den samme kurven.



Figur 29: Effekt av lineær approksimasjon av logaritmisk kurve når avstand øker

Modell 1 i figuren over illustrerer den lineære approksimasjonen til den logaritmiske kurven dersom et datasett med observasjoner relativt nærmere sentrum er benyttet. Modell 2 beskriver en lineær approksimasjon til den samme logaritmiske kurven, men her har vi flere observasjoner relativt langt borte fra sentrum. Vi ser at helningen til den lineære approksimasjonen i modell 2 er mindre enn i modell 1, selv om det er den samme kurven de forsøker å tilnærme.

Dersom vi nå går tilbake til tabell 11, som sammenligner modellen til Alfheim og modellene som er presentert i denne oppgaven, ser vi at i Alfheims datasett ligger boligene gjennomsnittlig mye nærmere sentrum enn i datasettet som er benyttet i denne oppgaven.

Dette skyldes etter all sannsynlighet at Alfheim kun har sett på boligsalg i Kristiansand kommune, mens det i denne oppgaven også er sett på salg i de omkringliggende kommunene. Dette kan forklare hvorfor den estimerte negative effekten av økt sentrumsavstand er noe mindre i vår lineære modell enn i Alfheims.

Det at effekten av boligens sentrumsavstand på boligprisen ser ut til å følge en negativ logaritmisk kurve, gjør at den lineære approksimasjonen ikke blir helt nøyaktig. Man ser typisk at prisen til sentrumsnære boliger og boliger svært langt fra sentrum blir underestimert, mens boligene i mellom blir overestimert. Residualplottet til den lineære modellen med avstand målt i kilometer viste også tegn til nettopp dette. Det blir imidlertid ikke funnet like tydelige tegn til dette i residualplottet til modellen som bruker reisetid som mål på sentrumsavstand. Årsaken til dette kan være, som også korrelasjonsanalysen viste, at korrelasjonen mellom boligpris og avstand målt i reisetid følger en nærmere lineær linje enn avstanden målt i kilometer. Dette antas også delvis å være årsaken til at den lineære modellen som bruker avstand målt i reisetid har noe større forklaringskraft enn modellen som måler avstanden i kilometer. En annen faktor som antas å bidra til større forklaringskraft når avstanden måles i reisetid, er at denne modellen tar hensyn til at veistandarden er relativt dårlig vest for kvadraturen, og at en kilometer dermed tar lengre tid å kjøre i rushtrafikk vest for sentrum enn på østsiden og nordsiden av byen.

Residualplottene til begge modellene våre med lineær funksjonsform viste tegn til avvik fra tilfeldig spredning. Det antas at de virkelige funksjonsformene til variablene *alder* og *boareal* er henholdsvis avtagende og konveks og økende og konkav. Basert på dette var det nærliggende å tro at modeller med kvadratisk eller dobbeltlogaritmisk funksjonsform ville passe bedre til å modellere boligmarkedet i Kristiansandsregionen, noe også vår analyse synes å avsløre.

Siden også Alfheim presenterer en dobbeltlogaritmisk regresjonsmodell i sin oppgave, ble det valgt å først forsøke modeller med denne funksjonsformen. Residualene i de to modellene med dobbeltlogaritmisk funksjonsform ble på lik linje med residualene i de to modellene med lineær funksjonsform, plottet mot den lineære prediksjonen og hver av de uavhengige variablene med unntak av *sjø*. Plottene viste ingen tydelige tegn til avvik fra tilfeldig spredning, og problemene vi hadde med heteroskedastisitet i de to første modellene

var nå lindret betraktelig. Tabell 12 sammenligner utvalgte parametere fra disse to modellene med den dobbeltlogaritmiske modellen til Alfheim (2005).

Tabell 12: Resultater fra lineær regresjonsanalyse med dobbeltlogaritmisk funksjonsform, sammenlignet med Alfheim (2005)

	Alfheim(2005)	Vår modell, med avstand målt i antall kilometer	Vår modell, med avstand målt i reisetid i rushtrafikk
Antall forklaringsvariabler	20	4	4
Forklaringskraft, $R^2$	77,7 %	74,2 %	74,0 %
Endring i estimert salgspris ved 1 % økning i avstand	-0,129 %	-0,193 %	-0,188 %

Vi ser at jevnt over forklarer alle de tre modellene variasjonene i salgspris godt. Modellene som er presentert i denne oppgaven ser for øvrig ut til å estimere en litt kraftigere negativ effekt på boligprisen ved økt sentrumsavstand, enn modellen til Alfheim. Det er ikke funnet noen åpenbare årsaker til dette, men modellen til Alfheim benytter seg av dummyvariabler for enkelte områder i Kristiansand og de fleste av disse har negative estimerte regresjonskoeffisienter. Modellene som er presentert i denne oppgaven har ikke slike områdedummyer, og det kan spekuleres i om at dette kan være årsaken til at den estimerte effekten av sentrumsavstand i vår modell avviker en del fra Alfheims estimat.

Når det gjelder de to dobbeltlogaritmiske modellene våre, er det lite som skiller dem. Både forklaringskraft og de estimerte koeffisientene er relativt like. Mens det for de to modellene med lineær funksjonsform viste seg at å beregne avstanden i reisetid i rushtrafikk fremfor i antall kilometer hadde visse fortrinn, kan man ikke si at man finner dette for de dobbeltlogaritmiske modellene.

Den dobbeltlogaritmiske funksjonsformen ser for øvrig ut til å passe bedre enn den lineære formen til å estimere boligprisene i Kristiansandsregionen. Det antas at årsaken til dette er at effekten på den implisitte prisen til de fleste boligattributtene er marginalt avtagende ved en økning i attributtet, og at de dermed estimeres bedre med en logaritmisk kurve enn med en

lineær. Hva man så velger å måle avstanden i, om det er antall kilometer eller reisetid i rushtrafikk, spiller ikke en vesentlig rolle. I denne analysen ble resultatene relativt like, og man finner ingen overbevisende argumenter for å foretrekke det ene foran det andre. Et lite moment som likevel kan trekke i retning av å velge avstanden målt i kilometer, er at folk som bor like utenfor kvadraturen i mange tilfeller vil velge å spasere eller sykle til jobben i den verste rushtiden. For disse boligene vil distansen målt i kilometer være et bedre mål på sentrumsavstanden enn reisetid i rushtrafikk. For de aller fleste boliger er det også enklere å måle avstanden fra sentrum i kilometer relativt nøyaktig, enn å lage et godt estimat på reisetid i rushtrafikk.

Selv om vi har indikasjoner på at en dobbeltlogaritmisk funksjonsform passer bedre til å beskrive boligmarkedet i Kristiansandsregionen enn en lineær, betyr dette absolutt ikke at modellen med lineær funksjonsform er verdiløs. Modellen forklarer variasjonene i salgspris svært godt, og de estimerte koeffisientene er for de fleste enklere å forholde seg til enn de estimerte koeffisientene i modellen med dobbeltlogaritmisk funksjonsform. Skal modellene brukes til videre analyser, som for eksempel hypotesetesting, bør imidlertid den dobbeltlogaritmiske funksjonsformen benyttes, da den lineære funksjonsformen viste tydelige tegn til heteroskedastisitet.

Dersom man velger å bruke modellen med lineær funksjonsform, gir sentrumsavstanden målt i reisetid i rushtrafikk en bedre modell enn sentrumsavstanden målt i kilometer. Dette antakeligvis fordi effekten av sentrumsdistanse målt i kilometer gir en mer logaritmisk kurve som medfører en større unøyaktighet i den lineære approksimeringen. Alfheim (2005) hevder at denne logaritmiske kurven kommer av at effekten av en økning i reisetid på ett minutt er mye større dersom den totale reisetiden er relativt kort, enn dersom den er lang. Han skriver videre at dersom man først har kjørt 20 minutter, spiller det liten rolle om man må kjøre 5 minutter ekstra. Dette synes ikke å fullt ut stemme overens med resultatene i denne oppgaven.

Analysen i denne oppgaven, samt kapitlet som omhandler reisetid, indikerer at i hvert fall deler av det logaritmiske utseende til denne kurven skyldes at man bruker mye lengre tid på å kjøre en kilometer nærme sentrum, enn en kilometer lengre borte. Av denne årsak får en kilometers økning i avstand en større effekt på boligprisen nærme sentrum enn den får ute i

periferien. Det synes urimelig at effekten på boligprisen av et minutts økning i avstand skal variere i like stor grad avhengig av hvor man befinner seg i forhold til sentrum. Vår korrelasjonsanalyse indikerte at kvadratmeterprisen plottet mot avstand målt i reisetid fulgte en nærmere rett linje, og at en slik stor variasjon i verdien til ett minutt heller ikke er tilfellet.

Våre modeller med kvadratisk funksjonsform forsterker også dette inntrykket. Vår analyse viser at kurven til den estimerte totale priseffekten av en boligs sentrumsavstand, krummer langt mindre når avstanden måles i reisetid enn når den måles i antall kilometer. De to modellene med kvadratisk funksjonsform er for øvrig de to modellene i analysen som har høyest  $R^2$ , og er dermed de to modellene med høyest forklaringskraft av de seks modellene som er presentert i denne oppgaven. Det er imidlertid slik at alle modellene våre forklarer variasjonene i salgspris godt, og vi bør ikke utelukkende basere oss på at å maksimere  $R^2$  gir den beste modellen. Analysen viste for eksempel at når avstanden ble målt i kilometer, estimerte vår modell med kvadratisk funksjonsform at dersom boligen er lokalisert mer enn 25 kilometer fra sentrum, vil en marginal økning i sentrumsavstand ha en positiv effekt på salgsprisen. Dette strider både mot teorien presentert i kapittel 3 og mot generell sunn fornuft. Dette kan tyde på at denne funksjonsformen ikke passer så godt til å estimere boligprisene i Kristiansandsregionen når sentrumsavstanden måles i antall kilometer kjøredistanse og det er prisen til en bolig som ligger langt fra sentrum som skal estimeres. Vi ser også av residualplottene til denne modellen at boligen som ligger lengst bort fra sentrum har det største negative residualet av samtlige observasjoner, noe som indikerer at modellen har kraftig overestimert verdien til denne boligen.

Dersom vi bytter ut sentrumsavstanden målt i antall kilometer kjøredistanse med antall minutter reisetid i rushtrafikk, ser vi imidlertid at en kvadratisk funksjonsform nå passer mye bedre. Denne modellen har høyest forklaringskraft av samtlige, og kurven som beskriver den totale effekten av en boligs avstand fra sentrum på den estimerte prisen, er svakt konveks og fallende i hele det relevante området. Som for modellene med lineær funksjonsform, har vi også her visse problemer med heteroskedastisitet. Dette påvirker imidlertid kun påliteligheten til de estimerte standardavvikene, ikke de estimerte regresjonskoeffisientene. Residualplottene viser for øvrig ingen tegn til at denne modellen overestimerer eller

underestimerer noen spesielle boligtyper systematisk, slik tilfellet var for modellene med lineær funksjonsform.



## 7. Konklusjoner

Hovedformålet med denne oppgaven var å belyse i hvilken grad boligenes avstand fra sentrum, målt i reisetid i rushtrafikk, slår inn på boligprisene i Kristiansandsregionen, for så å sammenligne dette med mer tradisjonelle modeller hvor avstanden er målt i antall kilometer kjørevei. Analysen antyder en relativt sterk negativ korrelasjon mellom observert kvadratmeterpris og avstanden til sentrum, målt både i minutter reisetid og i kilometer kjørevei. Analysen indikerer for øvrig at avstanden målt i reisetid har høyere lineær korrelasjon med kvadratmeterprisen enn avstanden målt i kilometer.

Det er utført lineære regresjonsanalyser med både lineære, dobbeltlogaritmiske og kvadratiske underliggende funksjonsformer. Regresjonsmodellene i denne oppgaven bruker kun fire uavhengige variabler, men viser seg å kunne forklare variasjonene i salgspris like godt som mye større modeller som er brukt i tidligere oppgaver av denne typen. Modellen med lineær funksjonsform anslår at verdien til en ellers lik bolig synker med 45 350 kroner dersom sentrumsavstanden økes med ett minutt. Analysen viser også at for en modell med denne type funksjonsform gir sentrumsavstanden målt i reisetid en modell med større forklaringskraft enn dersom avstanden blir målt i antall kilometer kjørevei.

Det viser seg for øvrig at modellene med dobbeltlogaritmiske eller kvadratiske funksjonsformer, passer bedre til å estimere boligpriser i Kristiansandsregionen. Den dobbeltlogaritmiske modellen estimerer, med reisetid i rushtrafikk som mål på sentrumsavstand, at verdien til en ellers lik bolig vil reduseres med 0,188 % når avstanden til sentrum økes med 1 %. Ved bruk av denne funksjonsformen er det relativt liten forskjell på om avstanden måles i minutter reisetid eller kilometer kjøretid. Dette gjelder både de estimerte verdiene og modellens forklaringskraft.

Modellene med kvadratisk funksjonsform har for øvrig høyest forklaringskraft av samtlige modeller, og for denne funksjonsformen viser analysen at sentrumsavstanden målt i reisetid i rushtrafikk er å foretrekke fremfor avstanden målt i antall kilometer kjørevei. Vår modell estimerer at for en bolig i sentrum vil en ett minuts økning i sentrumsavstand, alt annet konstant, redusere boligprisen med 70 820 kroner. Etter hvert som vi beveger oss ut fra

sentrum vil den estimerte prisen fortsette å falle, men størelsen på reduksjonen per minutt vil avta med 1219 kroner for hvert minutt sentrumsavstanden øker.

# Litteraturliste

- Alfheim, T. (2005). *Boligmarkedet i Kristiansand - Hvilken betydning har avstand til sentrum for boligprisene?* Siviløkonomoppgave, Høgskolen i Agder, Kristiansand.
- Brooks, C. (2008). *Introductory econometrics for finance*. New York: Cambridge University Press.
- Brueckner, J. K. (1987). The structure of urban equilibria: A unified treatment of the Muth-Mills model. I E. S. Mills (Red.), *Handbook of regional and urban economics* (Vol. II, s. 821-845): Elsevier Science Publishers B.V.
- DiPasquale, D., & Wheaton, W. C. (1996). *Urban economics and real estate markets*. Englewood Cliffs, N.J.: Prentice Hall.
- Eiendomsverdi. (2013). Eiendomsverdi. Hentet 07.05, 2013, fra <http://eiendomsverdi.no>
- Grimstvedt, A., & Eggum Adolfsen, T. (2011). *Hvordan har utbyggingen av ny motorvei i Agder påvirket boligprisene i Lillesand?* Masteroppgave, Grimstvedt & Eggum Adolfsen, Kristiansand.
- Hite, D. (1998). Information and Bargaining in Markets for Environmental Quality. *Land Economics*, 74(3), 303-316.
- Knutepunkt Sørlandet. (2012). Pendling på Agder. fra <http://www.knutepunktsorlandet.no/artikkel.aspx?AId=545&back=1&MId1=30&MId2=185>
- Knutepunkt Sørlandet. (2013). Slik pendler arbeidstakere i regionen. fra <http://www.knutepunktsorlandet.no/artikkel.aspx?MId1=27&AId=1117&Back=1>
- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *The Journal of Political Economy*, 74(2), 132-157.
- Osland, L. (2001). Den hedonistiske metoden og estimering av attributtpriser. *Norsk Økonomisk Tidsskrift*, 115(1), 1-22.
- Robertsen, K., & Theisen, T. (2009). The impact of financial arrangements and institutional form on housing prices. *The Journal of Real Estate Finance and Economics*, Publisert online 11.november. doi: 10.1007/s11146-009-9213-z
- Robertsen, K., & Theisen, T. (2010). Boligmarkedet i Kristiansand *Økonomi og tid : 18 essays i Pufendorf-tradisjon : festskrift til professor Arild Sæther på 70-årsdagen, 8. august 2010* (s. s. 243-260). Bergen: Fagbokforl.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34-55.
- Rådmannen i Kristiansand. (2011). *Styrke i muligheter: Kommuneplan 2011-2021, samfunnsdelen*. Kristiansand: Kristiansand kommune.
- Statens vegvesen. (2010). Samferdselspakke fase 2 for Kristiansandsregionen KVVU *Trafikkundersøkelse og trafikkdata - Kristiansand* Hentet fra <http://www.vegvesen.no/attachment/154734/binary/264433>
- Statistisk sentralbyrå. (2012). Tabell 01222: Befolkningsutvikling *Statistikkbanken*. Oslo: Statistisk sentralbyrå.
- Studenmund, A. H. (2006). *Using econometrics: a practical guide*. Boston, Mass.: Pearson Education.
- Visveg. (2013). Ruteplanlegger. Hentet 08.05, 2013, fra <http://visveg.vegvesen.no>

## Vedlegg 1: Testing av VIF

Test av VIF (Variance Inflating Factors) er en metode for å undersøke alvorlighetsgraden av multikollinearitet. Denne undersøker hvorvidt en gitt uavhengig variabel kan bli forklart av alle de andre uavhengige variablene i regresjonslikningen. Dersom man for eksempel ønsker å sjekke en regresjonslikning med  $k$  uavhengige variabler for multikollinearitet ved hjelp av en VIF-test, tar man utgangspunkt den opprinnelige regresjonslikningen:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Man gjennomfører så en regresjon som har  $X_i$  som den avhengige variabelen som skal bli forklart av de resterende forklaringsvariablene. Dersom for eksempel  $i = 1$ , har vi at:

$$X_1 = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \alpha_k X_k + v \quad (8.1)$$

Hvor  $v$  er et klassisk stokastisk feilledd. Vi ser at vi vil ha  $k$  slike regresjoner, en for hver av de uavhengige variablene i den opprinnelige regresjonslikningen.

VIF for  $\hat{\beta}_i$  er nå gitt ved følgende uttrykk:

$$VIF(\hat{\beta}_i) = \frac{1}{(1 - R_i^2)}$$

Hvor  $R_i^2$  er forklaringskraften til regresjonene av typen i likning 8.1. På denne måten kan man finne VIF for hver av de uavhengige variablene i den opprinnelige regresjonslikningen. Det finnes ingen tabell over kritiske verdier til VIF, men en tommelfingerregel er at dersom  $VIF(\hat{\beta}_i) > 5$ , er det betydelig multikollinearitet til stede. Denne grensen kan økes litt dersom vi har mange variabler i modellen (Studenmund, 2006).

Figurene på neste side viser resultatene av VIF-testene av våre modeller med lineær- og dobbeltlogaritmisk funksjonsform. Da det åpenbart er stor korrelasjon mellom  $X$  og  $X^2$  er det ikke foretatt VIF-test av modellene med kvadratisk funksjonsform.

## Lineær funksjonsform

Avstand målt i kilometer

Avstand målt i reisetid

variable	VIF	1/VIF	variable	VIF	1/VIF
kmavstand	<b>1.19</b>	<b>0.840483</b>	reisetid	<b>1.25</b>	<b>0.799202</b>
boa	<b>1.14</b>	<b>0.876159</b>	boa	<b>1.17</b>	<b>0.853088</b>
alder	<b>1.11</b>	<b>0.904375</b>	alder	<b>1.14</b>	<b>0.875218</b>
sjø	<b>1.06</b>	<b>0.939880</b>	sjø	<b>1.06</b>	<b>0.943952</b>
Mean VIF	<b>1.13</b>		Mean VIF	<b>1.16</b>	

## Dobbeltlogaritmisk funksjonsform

Avstand målt i kilometer

Avstand målt i reisetid

variable	VIF	1/VIF	variable	VIF	1/VIF
lnkmavstand	<b>1.30</b>	<b>0.769252</b>	lnreisetid	<b>1.26</b>	<b>0.795883</b>
lnboa	<b>1.24</b>	<b>0.806611</b>	lnboa	<b>1.23</b>	<b>0.810093</b>
lnalder	<b>1.08</b>	<b>0.925091</b>	lnalder	<b>1.05</b>	<b>0.952786</b>
lnsjø	<b>1.04</b>	<b>0.957488</b>	lnsjø	<b>1.03</b>	<b>0.967828</b>
Mean VIF	<b>1.17</b>		Mean VIF	<b>1.14</b>	