Prisutvikling på leiligheter i Kristiansands sentrum og periferi

Bente Fiddan

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 innestår for de metoder som er anvendt og de konklusjoner som er trukket.

Universitetet i Agder, 2010
Fakultet for Økonomi og samfunnsvitenskap
Institutt for Økonomi

Forord

Denne oppgaven er skrevet som en del av masterprogrammet i Økonomi og Administrasjon ved Universitetet i Agder.

Jeg vil rette en stor takk til min veileder, Theis Theisen, ved Universitetet i Agder, for god veiledning og oppfølging gjennom hele prosessen.

Jeg vil også takke for alle som har støttet og heiet på meg, og en spesiell takk til min mann, Hallgeir.

Konsmo, 14.juni 2010

Bente Fiddan

Innholdsfortegnelse

Figurliste *Sidetall*

Tabelliste

Sammendrag

Hensikten med oppgaven var å undersøke om det er forskjell i prisutviklingen på leiligheter sentralt i Kristiansand, mot usentralt i Kristiansand. For å finne ut av dette, har vi sett på salg av leiligheter som var solgt i Kristiansand i perioden 1990-2010.

I løpet av denne perioden, og etter å ha tatt ut noen case som manglet nødvendig opplysninger, satt vi igjen med 1812 antall case. Det vil si 1291 salg som hadde skjedd sentralt, og 521 salg som var skjedd usentralt.

Med bakgrunn i teori har vi kunnet utlede en prisfunksjon til disse casene, hvor prisen er avhengig av ulike egenskaper ved leiligheten.
Ved hjelp av den økonometriske modellen, presenterte vi to ulike prisformer. En lineær og en dobbeltlogaritmisk.
Av disse to måtte vi velge den prisformen som best fanget opp variasjonen i prisen på vårt datamateriale. Den dobbeltlogaritmiske formen skilte seg ikke veldig fra den lineære formen, men det ble likevel den vi valgte på grunn av at den var greiest å gjøre estimeringer videre på.
Vi gjennomførte regresjonsanalyse på den valgte prisformen, og testet 3 hypoteser vi hadde utledet, satt opp med bakgrunn i teorien og problemstillingen vår.

Vi fant at alderen på leiligheten på salgstidspunktet hadde en negativ påvirkning på prisen.

Vi fant også at ikke alle salgsårene hadde signifikant påvirkning på prisen. På grunn av det kunne vi ikke si at påvirkningen i prisen, som kom fra salgsåret, økte i løpet av hele perioden. Dette var bare tilfelle fra år 1996.

Til slutt fant vi at prisutviklingen fra år til år, har vært forskjellig sentralt og usentralt. De to stedenes prisutvikling samvarierer ganske bra, men har ikke variert helt likt. Forskjellen har vært at det sentralt har vært en tendens til større bevegelser i begge retninger (både ved lav prosentvis vekst og høy prosentvis vekst på begge steder).

1. Innledning

Utgangspunktet for å velge å skrive en oppgave som går på prisutvikling i eiendomsmarkedet, er at det er et tema som mange er engasjert i. Blant annet på grunn av den sterke prisveksten vi har sett de siste årene.
Å investere i en bolig er noe de fleste vil oppleve i løpet av livet. Så ved å finne ut mer om hvordan eiendomsmarkedet og prisutviklingstrenden fungerer, kan man ta med seg denne kunnskapen videre i livet.
Vi ønsker å se nærmere på leiligheters prisutvikling. Nærmere bestemt om prisutviklingen sentralt i en by er forskjellig fra prisutviklingen usentralt i byen. Dette er et mindre utforsket område av eiendomsmarkedet, som kan være interessant å vite mer om. Vi ønsker å se på denne prisutviklingen i byen Kristiansand.
Dette gir oss problemstillingen vår:

"Er prisutviklingen på leilighetene i Kristiansand sentrum, den samme som for leilighetene i byens periferi?"

Oppgaven i seg selv er kanskje for snever til å fange interessen til alle og enhver, men for mennesker som bor i, eller tenker å flytte til en by, kan det være veldig interessant.
Mennesker som ønsker en leilighet, er ofte enten i startfasen eller sluttfasen av livet. På disse tidspunktene er det ofte noe lite og sentralt som er å foretrekke, fremfor en hel enebolig som ofte er lokalisert mer usentralt, i tillegg også kan kreve masse vedlikehold.
Blant annet på grunn av stort trykk i leiemarkedet for studenter i byen Kristiansand, og altfor få leiligheter tilgjengelige, så er kjøp av leilighet en mulighet som en del benytter seg av. Da er det ofte en fordel at denne leiligheten kan selges med gevinst, etter endt studietid. Dette gjelder selvfølgelig ikke bare for studenter, men for mennesker som ønsker å eie en leilighet over en kortere periode i livet. Det kan også være av interesse for eiendomsspekulanter som ønsker og gjøre penger på investering i leiligheter eller leilighetsprosjekter.
Ved å studere problemstillingen vår, kan vi finne ut om gevinstens størrelse ved salg, har betydning av om leiligheten ligger sentralt eller usentralt i byen.

Vi begynner i kapittel 2 med litt bakgrunn og statistikk av eiendomsmarkedet i Kristiansand. I neste kapittel tar vi for oss eiendomsøkonomisk teori, som gir oss en forståelse av prisdannelsen i markedet, og hva som påvirker prisen. I kapittel 4 går vi inn på teori om den økonometriske modellen. Modellen går ut på presentasjonen av to ulike prisfunksjoner, hvor en av dem potensielt er gunstigst å bruke for å fange opp variasjonene i prisene fra vårt datamateriale. I dette kapittelet går vi også inn på teori om hypoteser, og til slutt i kapittelet utledes de hypotesene vi ønsker å få svar på, knyttet til problemstillingen vår. I kapittel 5 er det en presentasjon av datamaterialet vårt, og noen innledende analyser om dette. I kapittel 6 estimerer og velger vi hvilken prisform vi tror best beskriver våre data, på grunnlag av regresjonsanalyse og kapittel 4. I kapittel 7 går vi inn på temaet om prisutvikling. I kapittel 8 tester vi hypotesene vi utledet fra kapittel 4, med den prisformen vi valgte i kapittel 6. I kapittel 9 har vi konklusjoner og kritisk vurdering av arbeidet.

2. Bakgrunn

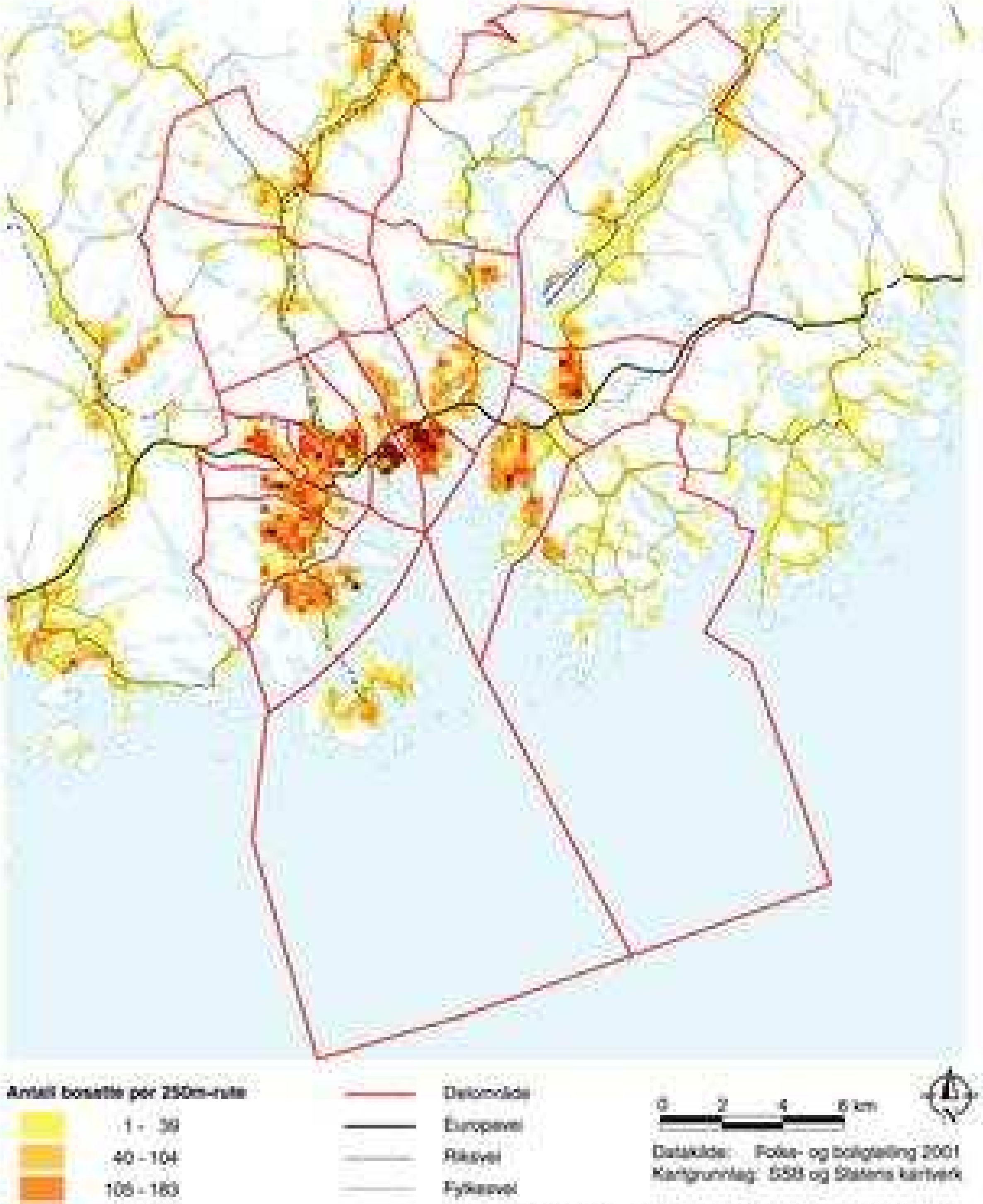
I dette kapittelet skal vi ta for oss litt om byen Kristiansand, og hovedsakelig statistikk fra boligmarkedet der.

Kristiansand kommune

Kristiansand som by ble grunnlagt i år 1641 av Christian IV. Den er 261,1 km^2 stor.
Per i dag bor det omkring 80 000 mennesker i byen. (Kilde: Kristiansand Kommunes hjemmeside.)

Bosettingsmønster

Det kan være interessant å se på bosettingsmønsteret i Kristiansand Kommune, for å få et bilde av hvor det er størst befolkningstetthet. Dette kan vi se i figur 2.1 fra 2001, nedenfor. Kristiansand Kommune ligger innenfor den røde markeringen på bildet. Antall bosatte per rute 250m $\times$250m. Fargen på ruten og antall bosatte per farge er:
Lys gul: 1-39
Mørk gul:40-104
Orange:105-183
Rød:184-303
Burgunder:304-525.
Der som det ikke er noen farge, er det heller ingen bosatte.
Den mørkeste fargen ser vi spesielt i byens sentrum, og litt i de forksjellige bydelene. Da særlig Lund og på Hånes. Ellers utmerker Grim, Tinnheia, Hellemyr, Slettheia og Vågsbygd seg med en mørk oransje farge som sier at det er ganske stor tetthet også her.



Figur 2.1 Bosettingsmønsteret i Kristiansand. SSB,(2010a).

Boliger etter bygningstype

Antall boliger var passert 2,3 millioner i Norge, 1. januar 2009. Boligblokk utgjør hele 22,4 % av alle boliger i Norge.
I tabellen nedenfor kan vi se hvordan situasjonen er for Kristiansand. I 1990 var andelen blokkleiligheter (blokk, leiegård eller annet boligbygg med 3 etasjer eller mer) på 24,53 %. I 2001 var denne gått ned til å være 20,10 %.

Tabell 2.1 Andel boligblokker i Kristiansand

	Antall boliger i alt:	Blokk, leiegård eller annet boligbygg med 3 etasjer eller mer:	Andel:
1990	27 338	6 706	24,53 %
2001	31 866	6 404	20,10 %

SSB (2010a),Tabell 15.

Det ble i sammen undersøkelse undersøkt hvor mange boliger det var etter byggeår, i Kristiansand, og i 2001 var det flest bygninger fra perioden 1961-1970:

Tabell 2.2 Antall boliger etter byggeår i Kristiansand

år	**I alt**	**-1990**	**1901-1920**	**1921-1940**	**1941-1945**	**1946-1960**	**1961-1970**	**1971-1980**	**1981-1990**	**1991-2001**
2001	31 866	1 264	746	2 166	121	5 901	6 571	5 953	4 873	4 271

SSB, (2010a), Tabell 19.

Boforhold i Kristiansand Kommune:

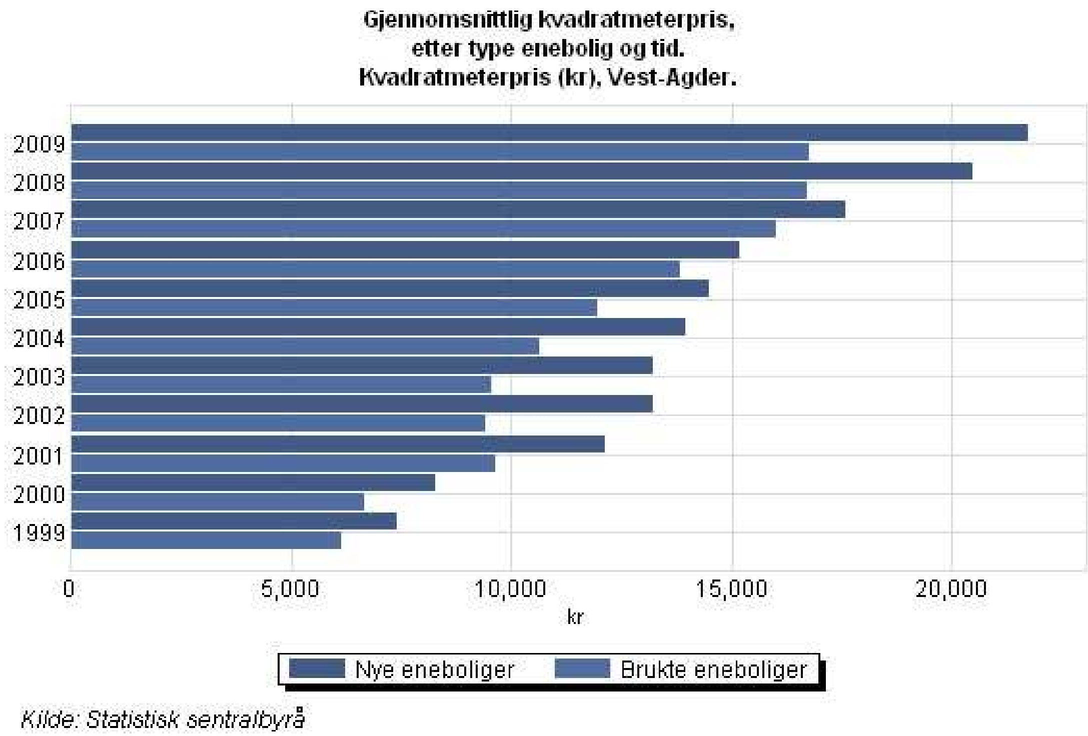
Nedenfor i tabell 2.3 kan vi se litt om boforhold i Kristiansand, i forhold til Vest-Agder fylke og resten av landet. Tabellen er fra 2001, og den viser at andel bosatte i boligblokk/ bygård er 13 %, nesten dobbelt så mye som resten av fylket, og litt over gjennomsnittet for hele landet. Den viser også at andelen som er bosatte i boliger bygd etter 1961 er på hele 71,1 %.

Tabell 2.3 Boforhold

	Kommunen	Fylket	Landet
Andel bosatte i blokk/bygård. 2001. Prosent	13,0	6,6	12,8
Andel bosatte i bolig bygd etter 1961. 2001. Prosent	71,1	70,7	66,9

.
SSB, (2010a).

I senere kapitler skal vi komme mer inn på at alder på en bolig ofte påvirker salgsprisen negativt. I figur 2.2 er priser på nye og brukte eneboliger sammenlignet. Man kan anta at prisene på leiligheter har hatt en lignende sammenheng. Figuren forteller oss at nye eneboliger får høyere pris per kvadratmeter enn det brukte eneboliger får. Eksempelvis kunne man i 2009 få omkring 22 000 kroner per kvadratmeter for en ny enebolig, mot omkring 17 000kr per kvadratmeter for en enebolig som var brukt:



Figur 2.2 Kvadratmeterpriser for nye og brukte eneboliger
SSB, (2010d), tabell 03364.

Litt om prisindekser

Prisindekser fanger opp bevegelser i priser på varer og tjenester over bestemte perioder. Og vi er jo nettopp ute etter å se på prisutviklingen på leiligheter over tid. I kapittel syv lager vi prisindekser ut i fra vårt datamateriale.
Prisene på varene eller tjenestene som det skal lages prisindekser for blir satt i forhold til et såkalt basisår. Den mest kjente indeksen her i landet er konsumprisindeksen som regnes ut av Statistisk Sentralbyrå.
Caplex (2010).
Noen eksempler på indekser:

Utvikling i ***kroneverdien*** med utgangspunkt i Konsumprisindeksen:
1000 kr. i januar 1990 tilsvarte 1551,89 kr. i januar 2010.
Dette tilsvarer en prisstigning fra januar 1990 til januar 2010 på 55,2 %.
(Grunnlaget for dette er de månedlige konsumprisindeksene fra ssb.)
SSB, (2010c).

I følge ***byggekostnadsindeks*** kalkulatoren:
1000 kroner i januar 1990 tilsvarte 1904, 80 kroner i januar 2010.
Dette tilsvarer en økning i byggekostnad fra januar 1990 til januar 2001 på hele 90,5 %
SSB, (2010b).

Til slutt tar vi også med boligprisindeksen som er regnet ut for salg av selveierleiligheter. I tabell 2.4 er år 2000 basisåret. Vi vil sammenligne disse resultatene vi finner her, med de indeksene vi finner fra vårt datamateriale for Kristiansand senere.
Økningen i prisen i hele landet under ett, fra 1991 til 2008 er: (166-43,9)/43,9= 2,781, hele 278, 1 %! Sammenlignet med prisstigningen i kroneverdien og byggekostnadsindeksen, er det altså ganske mye.

Tabell 2.4 Boligprisindeksen

Årstall:	Hele landet	Stavanger, Bergen og Trondheim
1991	43,9	45,8
1992	40,7	43,1
1993	41,2	45,5
1994	48,7	52,1
1995	51,1	55,1
1996	56,6	60,6
1997	64,3	66,9
1998	73,4	75,6
1999	83,3	84,1
2000	100	100
2001	107,6	110,8
2002	114,6	122,6
2003	115,9	132,8
2004	130,6	158
2005	140	171,9
2006	163,2	202,4
2007	177,3	219,6
2008	166	203,4
2009		

SSB, (2010d), tabell 05600.

.

3. Teori: Prisdannelse i boligmarkedet

I dette kapittelet tar jeg for meg den eiendomsøkonomiske teoridelen. Dette går på hvordan prisene blir dannet i markedet. Først litt om Alonso og hans modellby, og så teorien om hedonistiske priser.

3.1 Alonsos enkle modellby

I 1964 tok William Alonso for seg en enkel modellby for å beskrive hvordan bolig- leie og tomtepriser er i en by.

Modellen tar for seg en by som er monosentrisk. Byen har et bysenter i midten av byen, hvor alle jobbene er ligger. Bygningsstrukturen i denne byen er gitt, med lik tomtestørrelse for alle. For å komme til jobb pendler alle langs en rett linje inn til sentrum. Dette betaler de en transportkostnad *k* per kilometer per år. Avstanden mellom hjem og sentrum for den enkelte måles i variabelen d. Antar at bygrensa kalles for b, slik at om man bor på bygrensa er d=b. Husholdningene er identiske, på den måte at alle har lik lønn. Inntekten Y blir fordelt på pendlingskostnader, husleie og annet konsum. Også husene er identiske. Husleien er avhengig av avstand fra sentrum, R(d). Det du betaler ekstra i pendlingskostnad for å bo lengre fra sentrum, får du igjen ved tilsvarende lavere husleie. Dette skriver vi som –kd.

Annet konsum, x, vil da være likt over alt: $x = x^0$.

Nedenfor vises husleiegradienten, som viser at leie i byens sentrum er dyrere enn leie i utkanten av byen. Husleiegradienten er altså en linje som viser forholdet mellom husleia og avstand fra sentrum:

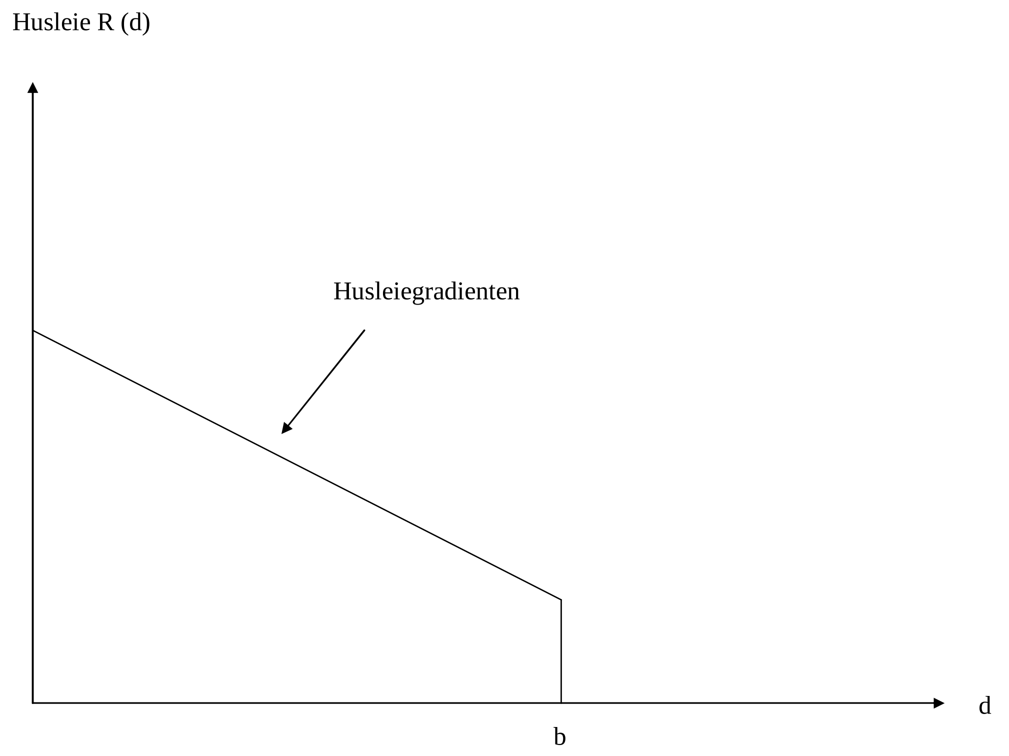
$$R(d) = y - kd - x^0 = y - kd - y + kb + (r^a q + c) \Rightarrow$$

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

Hvor r^a er avkastning per mål eller jordleie, og q er tomtearealet.

Ved å derivere husleiegradienten med hensyn på avstand fra sentrum d, finner vi at helningen på linjen er: $\frac{\partial R(d)}{\partial d} = -k$.

I figur 3.1 kan vi se hvordan en slik husleiegradient kan se ut:



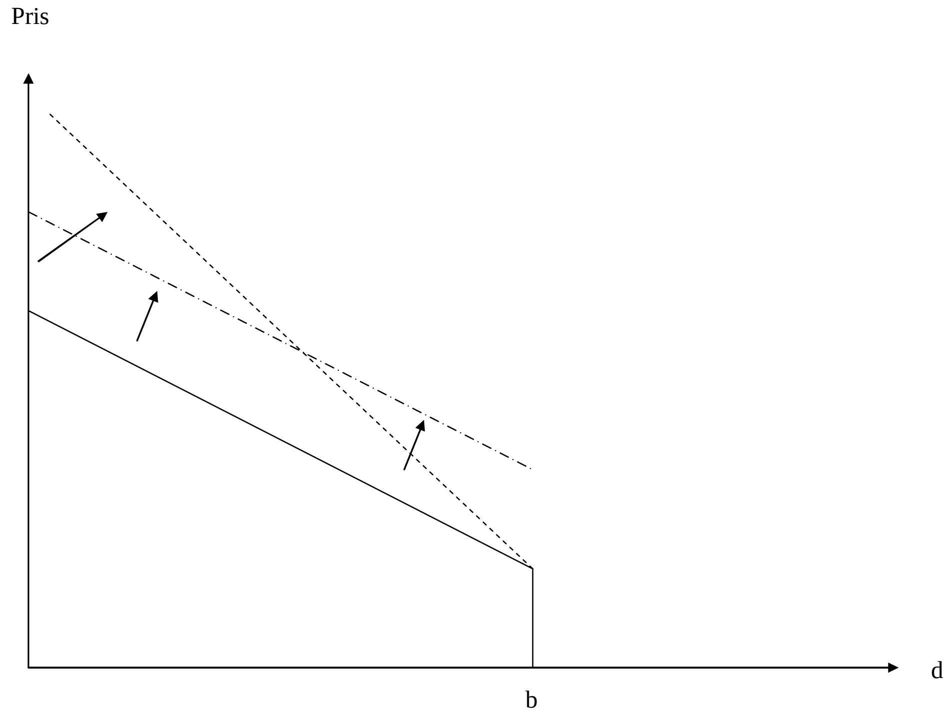
Figur 3.1 Fallende husleie ut fra sentrum

Nåverdien av husleien over uendelig tid gir prisen ved salg av boligen. Dermed kan vi anta at helningen på prislinja er lik helningen på husleiegradienten.

I figur 3.2 er det tegnet inn to ulike forslag til hva som kan skje over tid dersom prisene stiger.

Det ene forteller at prisene vil stige jevnt langs hele radiusen til byen, det vil si at det skjer et skift som går parallelt utover. Da vil både boliger som selges sentralt i byen og boligene som selges litt mer usentralt i byen, ha lik prosentvis prisvekst.

Den andre forteller at det vil være lavere prosentvis prisvekst, jo lengre mot byens grense du nærmer deg.

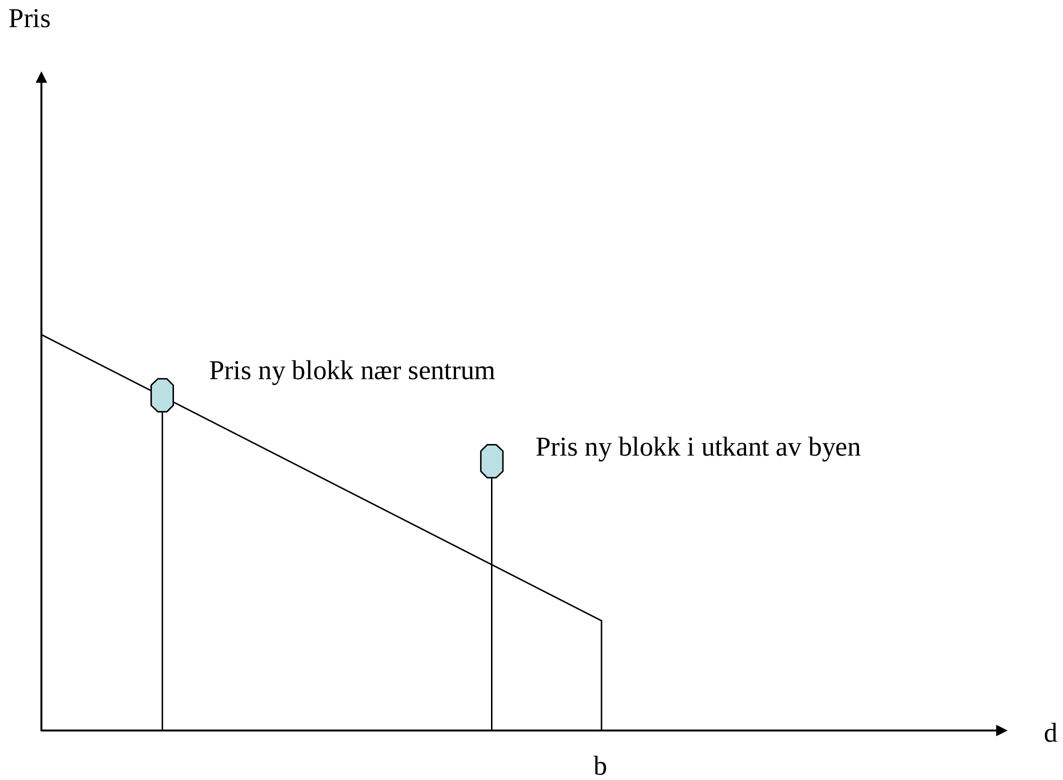


Figur 3.2 To typer prisstigningstilfeller

Vi vet ikke om disse forslagene er reelle, men dette vil vi forsøke å finne ut.

Hvordan kan det skje at det er større prosentvis prisvekst i sentrum av byen? Vi kan tenke oss følgende situasjon (se figur 3.3 for illustrasjon):

To boligblokker blir bygget på samme tid. En blokk ganske nær (eller i) sentrum, mens en annen bygges i utkanten av byen (nærme bygrensen). Kostnaden for å bygge blokkene på de to stedene skiller bare litt.
I sentrum vil prisnivået for en blokkleilighet ved salg, ligge på likt nivå med prisnivået den ble bygget for. I utkanten vil det ikke være tilfelle. Her vil blokkleiligheten selges til en høyere pris, som ligger over prisnivået på andre blokkleiligheter i utkanten.
En grunn er at det er mindre tilbud av slike blokkleiligheter her, og prisene vil dermed presses opp av større etterspørsel. Denne etterspørselen kommer blant annet fra at noen husholdninger som er bosatt der, kan ønske å flytte fra sin enebolig til en blokkleilighet, i det området de bor i. Dette kan for eksempel være eldre mennesker som ikke lengre orker å holde en hel enebolig i stand lengre, men som liker seg godt i nabolaget og ikke vil flytte vekk. Prisen kan da bli unaturlig høy, og det kan være grunn til å regne med en lavere prisstigning over tid her.



Figur 3.3 Pris på ny boligblokk i forhold til prisnivå

3.2 Den hedonistiske metoden

Hedonisme kommer av ordet hedone, som betyr lyst eller glede. Osland (2001). Den hedonistiske hypotesen sier at verdien på godet kommer av nytte- bærende attributter eller karakteristika. Rosen (1987). Den hedonistiske metoden blir ofte brukt til å beskrive for eksempel boliger. Verdien på en bolig kommer av flere ting. Er den relativt ny, nyoppusset, har kort avstand til skole og butikk, har stort boareal og så videre, vil verdien ofte bli høyere. Disse tingene som påvirker verdien til boligen, kaller vi attributter.

Når vi snakker om hedonisme er det som regel heterogene goder vi snakker om. Et heterogent gode er et gode har sine spesielle kjennetegn som gjør det helt unikt. Vi kan derfor si at en bolig er et heterogent gode fordi to boliger aldri kan være helt like. I noen tilfeller kan de være like i innhold og utseende, men beliggenheten til en bolig kan aldri være helt lik en annen bolig.
I studier av boligmarkedet blir ofte den hedonistiske metoden brukt. Ved å bruke metoden kan man blant annet finne prisutvikling over tid. Osland (2001).

Prisindeksene på de forksjellige attributtene blir observert indirekte ved å se på totalprisen. En marginal partiell økning i mengden av et attributt vil da gi en økning i den samlete prisen. Osland (2001).

Forutsetninger for modellen

Jeg bygger mye på teorien til Osland(2001) og Theisen (2009) resten av denne teoridelen.

Vi tar for oss beskrivelsen av et frikonkurranse marked.
Det antas at det finnes et stort antall boliger på markedet, og at du kan velge mellom ulike attributtvektorer kontinuerlig. Sammensetningen av attributter i en bestemt bolig, kaller vi en attributtvektor. Så dette kan forstås som at det finnes et uendelig antall med ulike boliger, fordi ingen er helt like. Nye boliger bygges hver dag, og det vil alltid være en annen sammensetning av attributter tilgjengelig.
For hver enkelt av de mange små aktørene er det ikke mulig å påvirke markedsforholdene og prisene alene. Søke-, transaksjons og flyttekostnader vil heller ikke påvirke i tilpasningen. Det er full informasjon blant både konsumenter og produsenter, når det gjelder priser og hva slags attributter hver bolig inneholder.

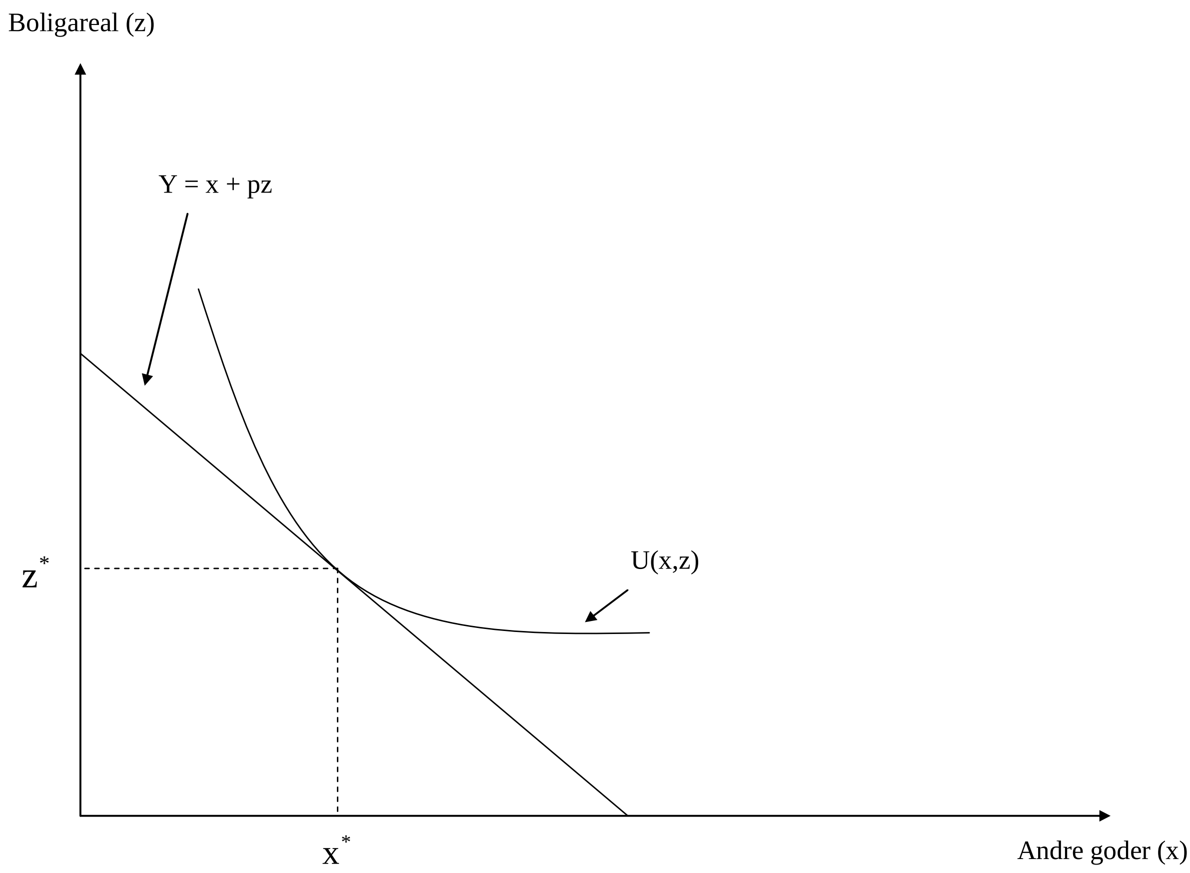
Konsumenten vil maksimere sin nytte, og produsenten (tilbyderen) vil maksimere sin profitt. Hvor en gitt pris tilfredsstiller begge samtidig, vil vi finne en likevekt, og et salg.

3.2 Likevekt på etterspørselssiden

Vi vil nå ta for oss hvordan en konsument kan få maksimert sin nytte. Å maksimere sin nytte i dette tilfellet, vil gå ut på at konsumenten skal få den beste mulige sammensetningen av attributter ved boligen han ønsker, for lavest mulig pris.

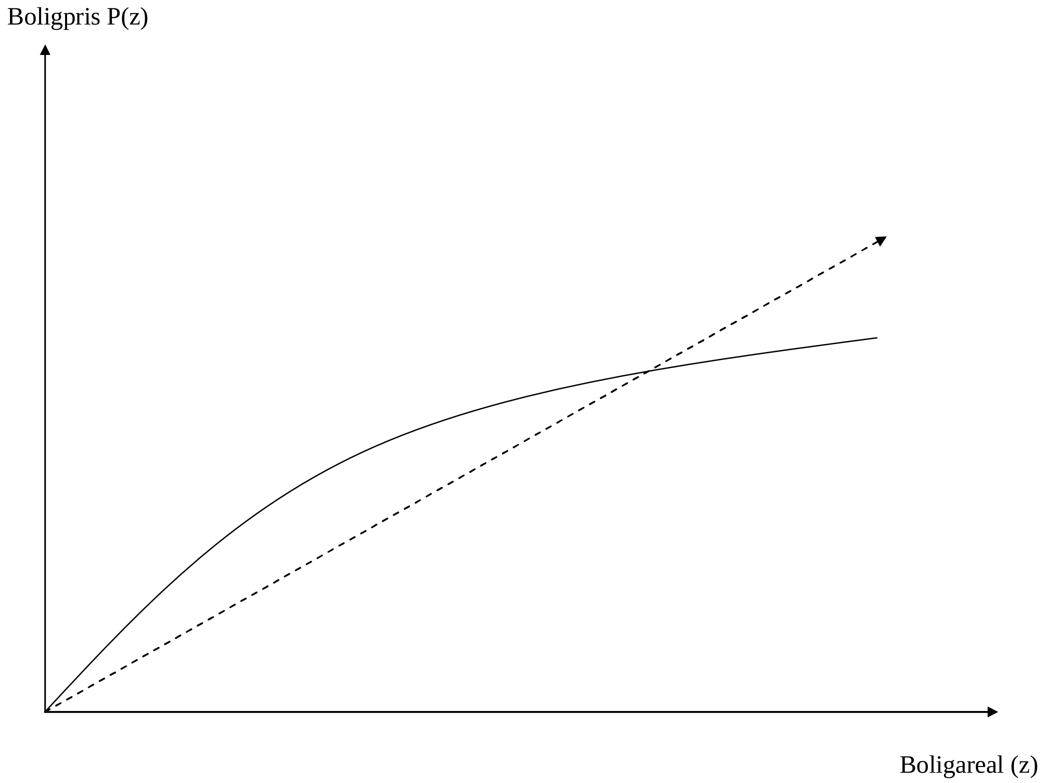
Konsumentens tilpasning

Til å begynne med vil vi vise på en enkel måte hvordan en konsument kan maksimere sin nytte. Se figur 3.4 under.
Vi antar en lineær budsjettbetingelse og kun to goder, noe som betyr at hele konsumentens inntekt fordeles på godene. Det ene godet kaller vi *boligareal*, z, og det andre godet representerer alle andre goder i ett gode, x. Budsjettlinja forteller hvor store mengder konsumenten kan kjøpe av hvert gode, ved et gitt budsjett Y. Budsjettbetingelsen blir da: Y = x + z.
Dersom konsumenten velger å bruke hele budsjettet sitt på gode x, vil det største mengden han kan kjøpe, finnes der hvor budsjettlinja krysser x-aksen.
En nyttefunksjon viser de ulike sammensetningene av godene, som gir lik nytte for konsumenten. Nyttefunksjonen kaller vi for U, og den er en funksjon av de to godene x og z. Det vil si at U viser sammensetningene av x og z som gir samme nytte.
Der nyttefunksjonen tangerer budsjettlinja, vil konsumenten finne sin optimale plassering, og maksimere sin nytte. De optimale mengdene av godene er da z^* og x^*.



Figur 3.4 To goder

En prisfunksjon for boligattributter er ofte ikke lineær. Noe som antagelig er mer realistisk kan vi se i figur 3.5. Vi ser her på prisfunksjonen til en bolig, når boligens pris kun er avhengig av ett attributt, nemlig boligarealet. Formen til prisfunksjonen er konkav, og det vil si at prisen stiger jo større boligen blir, men den stiger avtakende. Dette kan forklares både med at produksjonskostnadene per ekstra produserte kvadratmeter blir lavere, og at betalingsvilligheten til konsumenten per ekstra kvadratmeter, vil synke etter hvert som boligen blir ”stor nok”.



Figur 3.5 Prisfunksjon

Som vi tidligere har sagt, vil en konsument velge å maksimere sin nytte i sitt valg. I en ny nyttefunksjon har konsumenten nytte av alle boligattributtene, ikke bare boligareal. Denne nyttefunksjonen kan vi skrive slik: $U(x, z_1, ..., z_n)$. Det forutsettes at denne funksjonen er

konkav.

z angir mengden av attributt *i,* hvor *i = 1, …, n.*

x er alle andre goder som konsumeres, og Z angir mengden av de ulike attributtene en bolig har, og kalles derfor attributtvektoren:

(1) $Z = z_1, \ldots, z_n$

Prisen på boligen antar vi som en funksjon av attributtvektoren: P(Z).

Konsumentens inntekt er Y, og budsjettbetingelsen blir da:

(2) $Y = x + P\ (Z)$

Ved å snu på budsjettbetingelsen slik at x = Y – P(Z), kan vi sette dette uttrykket for x inn i nyttefunksjonen U(x, z_1, ..., z_n). Attributtene maksimeres så i nyttefunksjonen:

$$\underbrace{Max}_{z_1,...,z_n}\{U(Y-P(Z), z_1,..., z_n)\}$$

Da får vi dette resultatet:

(3) $\frac{\partial U}{\partial x}\left(-\frac{\partial P}{\partial z_i}\right)+\frac{\partial U}{\partial z_i}=0 \Rightarrow \frac{U_{z_i}}{U_x}=\frac{\partial P}{\partial z_i}=P_i$ hvor *i*= 1, …, n

Skriver vi om dette uttrykket får vi:

$\frac{U_{z_i}}{P_i}=\frac{U_x}{1}$, som er Gossens 2. lov. Dette betyr at i optimum skal nytten av siste krone være lik i alle anvendelser.

Konsumentens budfunksjon

Vi introduserer nå budfunksjonen. Det er en funksjon som forteller prisen en konsument er villig til å betale for en bolig, θ_j. (Hvor j benevner den enkelte konsument= 1, …, n).

Budfunksjonen forteller hva den maksimale betalingsvilligheten er, ved gitt konstant nyttenivå og inntekt, og for forskjellige typer hus eller attributtvektorsammensetning.

For et bestemt nyttenivå U^*, kan vi skrive:

(4) $U^*(y^* - \theta_j, z_1, \ldots, z_n, \alpha_j) = U^*$

Hvor α_j, er den enkelte konsuments preferanseparameter. En konsument kan for eksempel foretrekke noen ekstra kvadratmeter og bo på, mens en annen foretrekker å bo nærmest mulig sentrum.

Løser (4) med hensyn på θ_j, og finner da budfunksjonen:

(5) $\theta_j = \theta_j\ (Z, Y_j, U_j, \alpha_j)$

Vi skal så vise hvordan budfunksjonen utledes.
Ved å bruke de optimale verdiene for z og x, slik at Z^*, og sette disse inn i nyttefunksjonen får vi:

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

Ved å anta en gitt inntekt og et konstant nyttenivå vil prisen man betaler være lik den maksimale betalingsvilligheten. (Det vil si at $\theta_j = P(Z)$). Da kan vi sette dette inn i nyttefunksjonen:

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

Siden funksjonen skal være tilpasset alle inntekts og nyttenivå, kan vi skrive den mer generelt:

(8) $\theta = \theta\ (Z, Y_j, U_j,\ \alpha_j)$

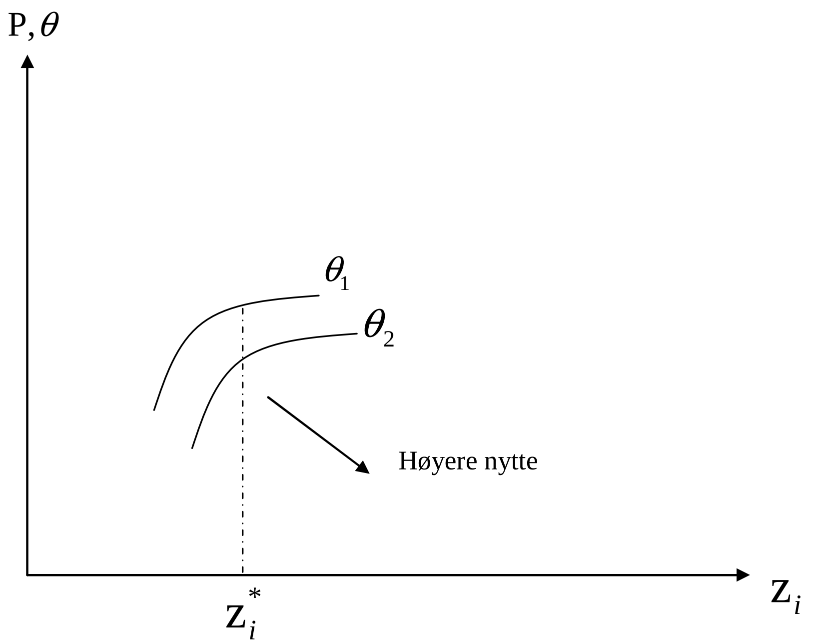
Ved å derivere den implisitte av (7), med hensyn på attributt *i* får vi:

(9) $\dfrac{\partial \theta_j}{\partial z_i} = \dfrac{\partial U_j / \partial z_i}{\partial U_j / \partial X} > 0$

Dette resultatet viser at den marginale betalingsvilligheten er positiv. Det vet vi fordi begge grensenyttene er positive.

Siden forutsetningen sier at budfunksjonen er strengt konkav, kan vi derfor vise at $\frac{\partial^2 \theta_j}{\partial z_i^2}$ < 0.

Hvert nyttenivå til en konsument gir en budfunksjon. Dette kan vi se grafisk i figuren nedenfor.



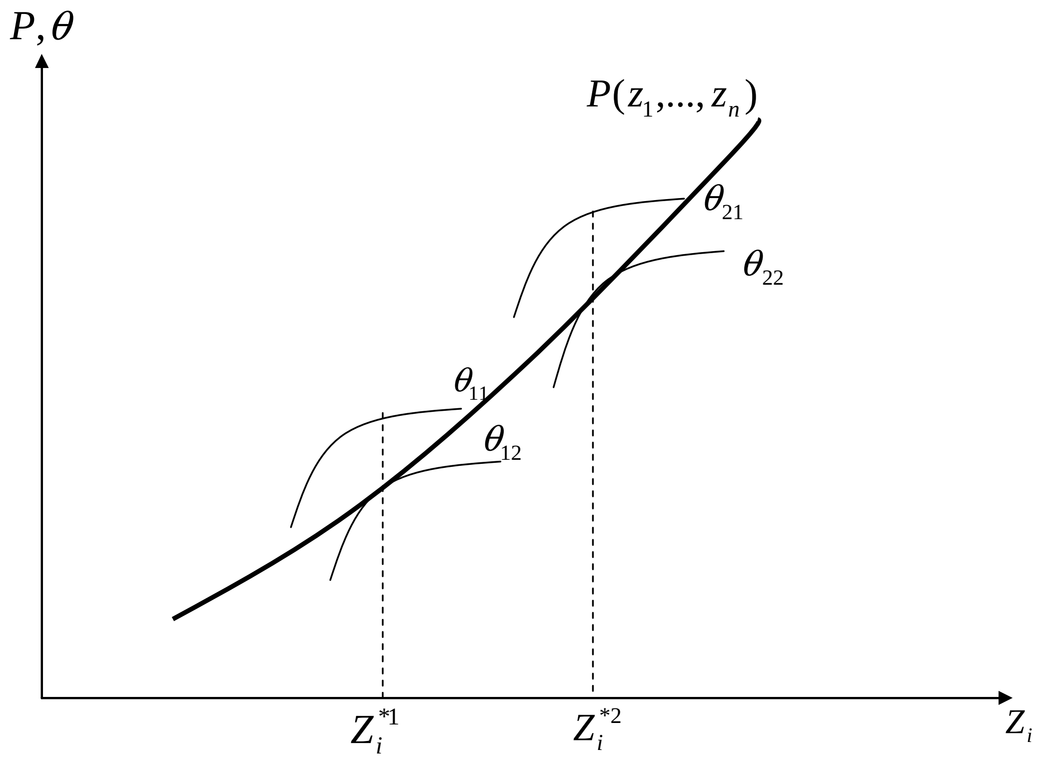
Figur 3.6 Budfunksjoner

Ser av figuren at dersom man holder mengden av attributt *i* fast, vil nytten øke når prisen reduseres.
Likevekt (salg) oppstår når den implisitte prisen er lik det implisitte budet, altså når:

(10) $\theta_j(Z^*, Y_j, U_j^*, \alpha_j)$ = P (Z)

Den laveste prisen konsumenten kan få kjøpt en bolig med attributtvektor Z er da P(Z), og denne prisen er den samme som konsumentens maksimale betalingsvillighet, $\theta_j(Z^*, Y_j, U_j^*, \alpha_j)$.
I figur 3.7 tar vi for oss to husholdninger (konsumenter) med ulik innekt og/eller preferanser. Husholdning 1 har budfunksjonen θ_1, og husholdning 2 har budfunksjonen θ_2. Her er det tegnet inn to budfunksjoner til hver husholdning. Der hvor budfunksjonen tangerer den eksogent gitte prisen P(Z), finner vi den optimale mengden av attributt *i* for hver husholdning.

For husholdning 1 er denne z_i^{*1}, og for husholdning 2 er det z_i^{*2}. Vi kan se at husholdning 2 etterspør mer av z_i enn husholdning 1.



Figur 3.7 Budfunksjoner med den eksogent gitte hedonistiske prisfunksjonen

3.3 Likevekt på tilbudssiden

Produsentens tilpasning

På tilbudssiden forutsetter vi at det er mange små og profittmaksimerende bedrifter. Antallet av produserte enheter av en gitt boligtype, og sammensettinger av attributter kan på kort sikt endres. Nyetableringer og nedleggelser av bedrifter tillates på lang sikt. Til forenkling tar vi kun for oss på kort sikt.

Prisen en bedrift kan få for en bolig med en bestemt attributtvektor, Z, oppfattes av hver enkelt bedrift som eksogent gitt i markedet, uavhengig av antall boliger bedriften tilbyr. Kostnadsfunksjonen C(M, z_1, …, z_n, β) antas å være konveks og stigende av antall boliger M. I produksjonen av attributter er grensekostnadene positive men avtakende. Faktorpriser eller produksjonsteknologi for den enkelte bedrift er underliggende faktorer i kostnadsminimerings- problemet og representeres ved β som er en vektor av disse skiftparametrene. Alle de kontinuerlige variasjonene i attributtene kommer av at det er så mange bedrifter på markedet. Man antar at hver bedrift har komparative fortrinn som de i produksjonen bruker til å spesialisere seg.

Profitten (π) til hver enkelt bedrift er:

(11) $\pi = MP(Z) - C(M, z_1, ..., z_n, \beta)$

Profitten maksimeres ved å velge optimale antall boliger M^*, og mengde attributt z_i. For hver enkelt bedrift blir det derfor:

(12) $\underbrace{Max}_{M, Z_1, .., Z_n} \{MP(Z) - C(M, Z, \beta)\}$

Førsteordensbetingelsene for maksimal fortjeneste blir da:

(13) $\frac{\partial \pi}{\partial z_i} \Leftrightarrow P_i = \frac{C_i}{M}$ (hvor i= 1, …, n)

(14) $\frac{\partial \pi}{\partial M} \Leftrightarrow P(Z) = C_M$

Hver bedrift bør etter ligning(13) velge en slik sammensetning av boligattributt slik at den implisitte prisen for et gitt boligattributt blir lik grensekostnaden per bolig ved en partiell økning i dette attributtet.

Hver bedrift bør etter ligning (14) øke produksjonen helt til prisen for en ekstra bolig produsert er lik grenseinntekten ved å selge den.
Optimalt antall boliger, M^*, og optimal attributtvektor, Z^*, blir bestemt av disse betingelsene.

Produsentens offerfunksjon

Offerfunksjonen definerer den laveste prisen en produsent er villig til å tilby en bolig for, med gitt attributtvektor, optimalt antall boliger og konstant profittnivå.

For et gitt profittnivå har vi:

(15) $\pi = M\Phi - C(M, Z, \beta)$

Hvor β er som tidligere nevnt en skiftparameter som forteller noe om den enkelte bedrift. Det kan være bedre tilgang til innsatsfaktorer enn andre, eller et naturlig fortrinn som for eksempel at de eier områder som er lettere å bygge ut.

Implisitt av denne, finner vi offerfunksjonen:

(16) $\Phi(Z, \pi, \beta)$

Vi skal nå beskrive offerfunksjonen i optimum. Dette gjøres ved å ta utgangspunktet i de optimale verdiene for attributter, Z^*, antall boliger, M^*, og profittnivå, π^*. Av dette får vi profittfunksjonen:

(17) $\pi^* = M^* P(Z^*) - C(M^*, Z^*, \beta)$

Setter offerfunksjonen inn i profittfunksjonen og får:

(18) $\pi^* = M^* \Phi(\pi^*, Z^*, \beta) - C(M^*, Z^*, \beta)$

Deriverer vi (18) med hensyn på antall boliger, M, og attributtvektor, Z, får vi førsteordensbetingelsene:

(19) $\frac{\partial \pi^*}{\partial M} \Leftrightarrow \Phi(\pi^*, Z^*, \beta) = C_M$

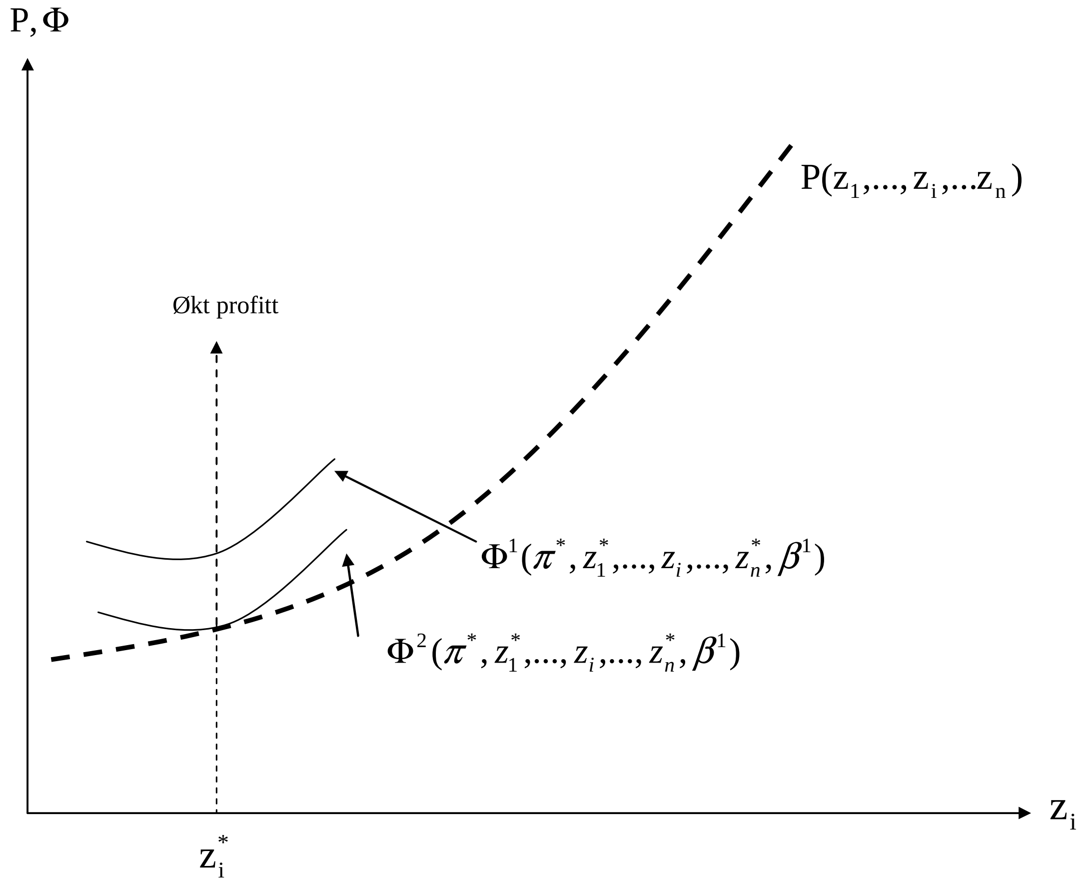
(20) $\frac{\partial \pi^*}{\partial Z_i} \Leftrightarrow \Phi_i(\pi^*, Z^*, \beta) = \frac{C_i}{M}$ (hvor *i* = 1, …, n)

Ser at disse resultatene at i optimum:
(19) - skal offerprisen være lik grensekostnaden ved å produsere en bolig ekstra.
(20) - bør attributtvektoren velges slik at den implisitte prisen for attributt *i* blir lik grensekostnaden per bolig ved en partiell økning av det aktuelle attributtet.
Et annet kjennetegn ved produsentens optimum, er også at offerprisen må være lik den faktiske prisen på boligen. Det vil si:
$\Phi(\pi^*, Z^*, \beta) = P(Z^*)$

Til slutt løser vi (19) med hensyn på M, og setter svaret inn i (18), går M bort. Gjennom profittfunksjonen finner vi da at den implisitt definerer offerfunksjonen:

(21) $\Phi(\pi^*, Z^*, \beta)$

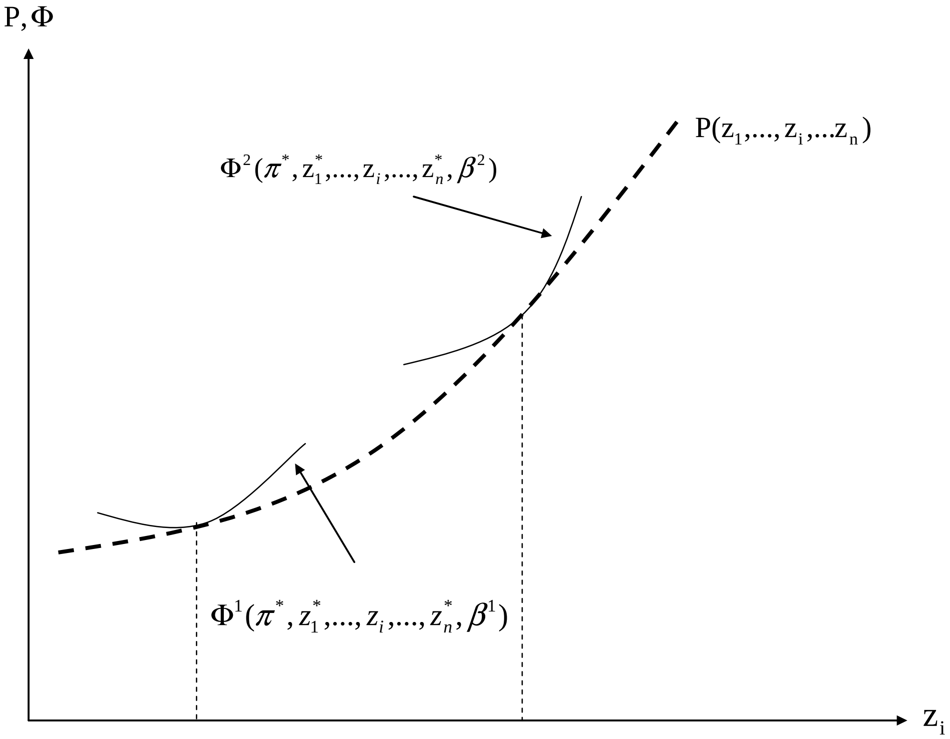
I figur 3.8 kan vi vise en bedrift og to av dens offerfunksjoner. Den offerfunksjonen som tangerer den eksogent gitte prisfunksjonen, vil gi den optimale mengde attributt *i*. Ved å holde mengden attributt *i* fast, vil profitten øke jo høyere pris bedriften kan få for boligen.



Figur 3.8 Offerfunksjon og den eksogent gitte prisfunksjonen

En forklaring på hvorfor offerkurvene er konvekse; jo mindre attributt *i* boligen inneholder, og til jo høyere pris den kan bli solgt for gjør at produsentens profitt øker og han havner dermed på et høyere profittnivå. Prisen må altså stige mer per ekstra attributt *i*, for at produsenten skal være på samme profittnivå.

I figur 3.9 har vi tatt med to produsenter, og den viser at det er ulik verdi på skiftparameteren, β, som gjør at de får ulike offerfunksjoner. Produsent 2 vil for eksempel tilby større boliger.

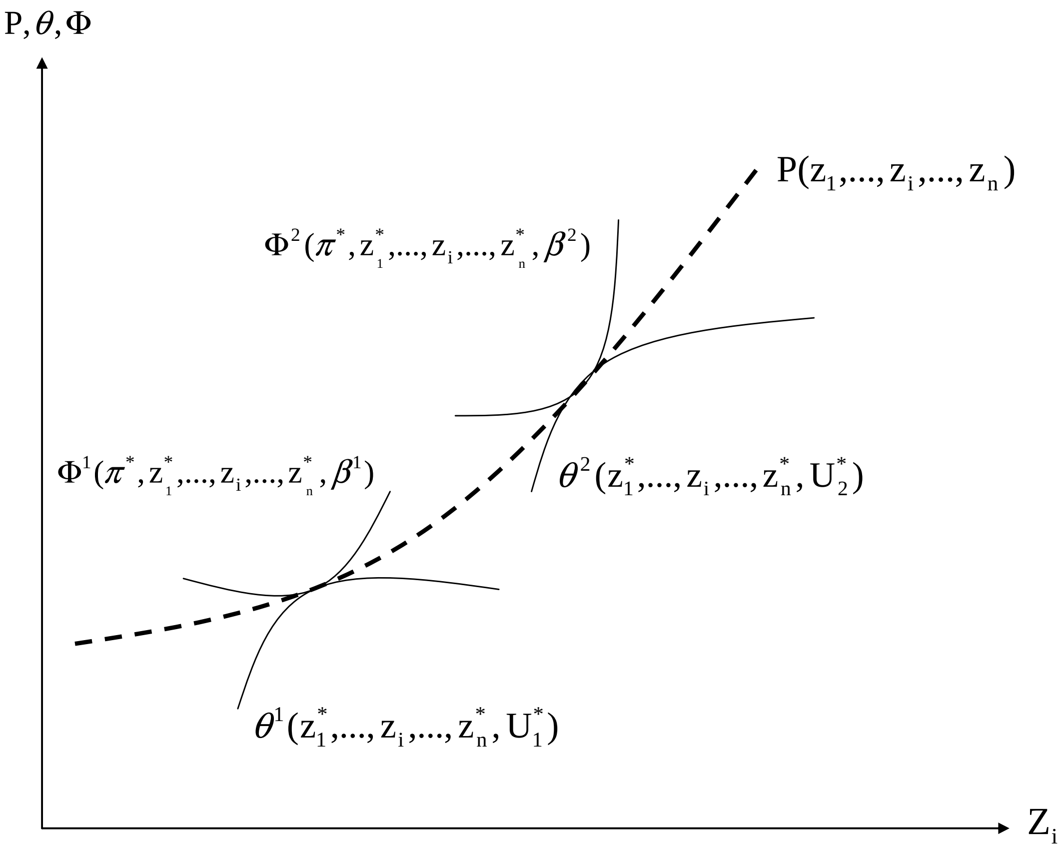


Figur 3.9 Offerfunksjoner og eksogent gitt prisfunksjon

Likevekt oppstår for produsenten når offerkurven tangerer den eksogent gitte prislinja. Dette får man ved førsteordensbetingelsene, (13), og (20). Likevekten krever at førsteordensbetingelsene er oppnådd, i tillegg til at $\Phi(Z^*, \pi^*, \beta) = P(Z^*)$. Det vil si at offerprisen skal være lik den eksogent gitte prisfunksjonen.

3.4 Markedslikevekt

I markedslikevekt tar vi konsumentens budfunksjon og produsentens offerfunksjon inn i samme figur, det vil si at de møtes i markedet. Der budfunksjonene og offerfunksjonene tangerer hverandre, er konsumenten og bedriften enige om en pris. Det oppstår da et salg i dette punktet Alle tangeringspunktene skaper til sammen den hedonistiske prisfunksjonen.



Figur 3.10 Hedonistisk prisfunksjon, offerfunksjoner og budfunksjoner

Betingelsene som gjelder for markedslikevekt:

Fra konsumentsiden må budet være lik som den eksogene prisen: $\theta_i = P_i$

Fra produsentsiden må offeret være likt som den eksogene prisen: $P_i = \frac{1}{M} C_i = \Phi_i$

I likevekt må det da være slik at begge disse oppfylles på likt, det vil si:

$$\theta_i = P_i = \frac{1}{M} C_i = \Phi_i$$

Vi har funnet at både konsument og produsent tangerer den eksogent gitte prislinja i optimum. Dette forteller at for både budfunksjonene og offerfunksjonene, utgjør denne prislinja en omhyllingskurve av disse funksjonene. Den hedonistiske prisfunksjonen oppstår som en

konstruksjon av at bud- og offerfunksjonene tangerer hverandre i likevekt.
Noen annen tilpasning enn på tangeringspunktet mellom budkurven og offerkurven vil ikke kunne finnes, dette fordi det alltid vil finnes andre konsumenter som på grunn av inntekt eller preferanser, derfor har høyere betalingsvillighet for akkurat denne boligen.

4. Økonometrisk modell og hypotesetester

I kapittel tre kom vi frem til hvordan en salgspris blir dannet i markedet I dette kapittelet skal vi velge hvilken funksjonsform prisen må ha, for å fange opp mest mulig av variasjonen fra vårt datamateriale. Det vil si den funksjonsformen som beskriver vårt datamateriale best mulig. I dette kapittelet skal vi derfor gjennomgå teori om den økonometriske modellen.
I tillegg tar vi med teori om hypotesetester, og utledning av de hypotesene vi ønsker å teste for.

4.1 Økonometrisk modell

Fra kapittel 3 fant vi at prisen på en bolig var avhengig av attributtene. Denne prisformen kan vi skrive slik:
P = P(z_1, …, z_i, …, z_n), hvor i = 1,…, n, og n er antall attributter.
Siden attributtene påvirker prisen, og ikke motsatt, kan vi kalle attributtene for uavhengige variabler. Den avhengige variabelen er da prisen.

I tillegg til de uavhengige variablene, bør det være et restledd med i prisfunksjonen som fanger opp de variasjonene i dataene som modellen vår ikke fanger opp. Dette leddet kaller vi for et stokastisk restledd. En funksjon med restledd vil se slik ut:
P = P(z_1, …, z_i, …, z_n, ε)
Restleddet må være normalfordelt, og med en forventningsverdi lik null.
For å finne den funksjonsformen som beskriver dataene best mulig, må vi teste de forskjellige typene funksjonsformer som er mest sannsynlig, og velge den som har oppfylt restleddsforutsetningene best mulig.
I tillegg er det en fordel med homoskedastisitet. Det vil si at for alle casene skal det være konstant varians i restleddet.

To mulige prisformer som med størst sannsynlig kan forklare våre data best, er:

1. Enkel lineær modell:

(4.1) P = $\beta_0 + \beta_1 z_1 + \beta_2 z_2 + ... + \beta_n z_n + \varepsilon$

2. Dobbelt logaritmisk modell:

(4.2) P = $\beta_0 z_1^{\beta_1} z_2^{\beta_2} z_3^{\beta_3} e^{\beta_4 z_4 + \beta_5 z_5 + \varepsilon}$

Den lineære modellen

β_0 er konstantleddet i prisen, og den påvirkes ikke av attributtene. β_i, hvor *i* går fra 1 til n, (ettersom hvor mange attributter det er) er attributtenes koeffisienter. Koeffisienten forteller hvor stor økning det blir i prisen, dersom attributtet øker med en enhet.

Fordelen med å bruke denne lineære modellen er at den er enkel å estimere, og at koeffisientene er enkle å tolke.

Ulempene med funksjonsformen er at attributtprisene er konstante, uavhengig av hvilket nivå de har. Dette er ofte ikke rimelig, siden prisene ofte reduseres med mengde.

En annen ting er at elastisiteten blir mindre jo høyere prisen er, El $_{z_1} P = \beta_1 \frac{z_1}{P}$, noe som kan være urimelig. Tilslutt kan det også være et problem dersom restledddet ikke er normalfordelt.

Den dobbeltlogaritmiske modellen

Ved å ta den naturlige logaritmen til denne formen får vi:

(4.3) Ln P = $\ln\beta_0 + \beta_1 \ln z_1 + \beta_2 \ln z_2 + \beta_3 \ln z_3 + \beta_4 z_4 + \beta_5 z_5 + \varepsilon$

De to attributtene som det her ikke er tatt logaritmen til, er dummyvariabler.

En dummyvariabel er en variabel som sier at et bestemt kjennetegn som kan påvirke prisen, er tilstede eller ikke. Eksempler på dummyvariabler som vi kommer til å bruke er salgsår, salgsmåned og postnummer.

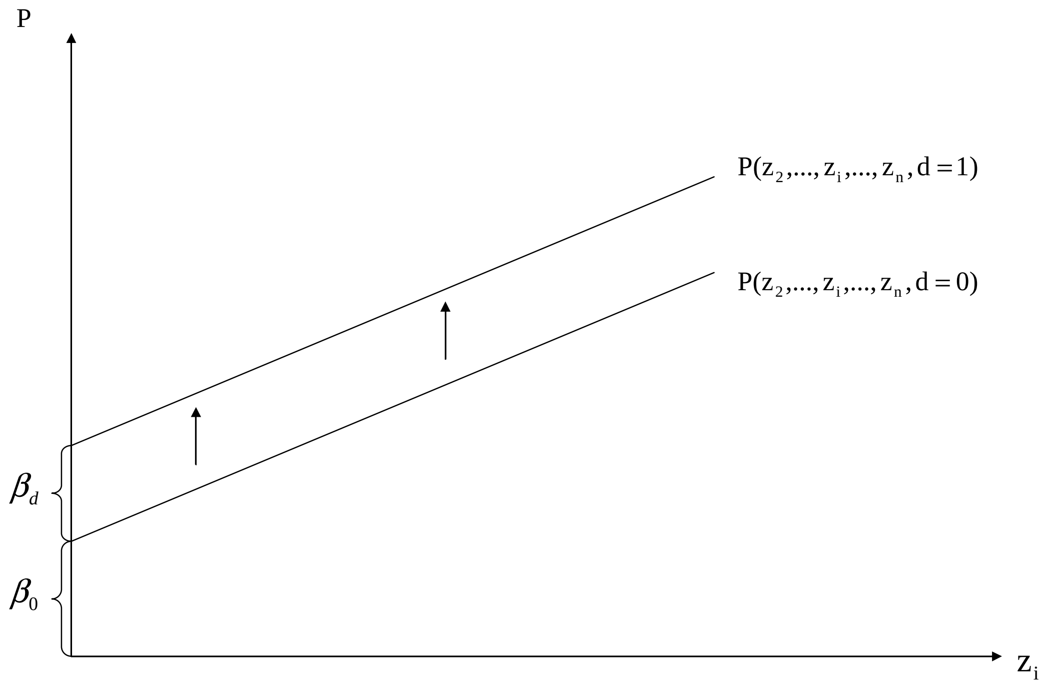
Koeffisientene i den dobbeltlogaritmiske formen kan tolkes som elastisiteten til attributtene. Den forteller hva den prosentvise endringa i prisen er, ved en prosents økning i attributtet.

Fordelen med den dobbeltlogaritmiske modellen når vi tar den naturlige logaritmen, er at den blir lineær, og derfor blir den lett å estimere. Det vil også være et samspill mellom de

uavhengige variablene/attributtene. En annen fordel er at koeffisientene til salgsårsdummyene, kan vi bruke til å lage prisindekser av. Dette skal vi se på i kapittel 7.
Vi skal se litt på hva som skjer med prisen dersom mengden av ett attributt øker. Som vi har nevnt ovenfor, kan vi også kalle attributter for uavhengige variabler. Disse uavhengige variablene kan variere på to måter:

- kan anta uendelige verdier, vi kaller dem kontinuerlige
- kan bare anta to verdier (0 og 1), vi kaller dem dummyvariabler

Eksempel på en uavhengig variabel som varierer kontinuerlig i en bolig, er arealet, eller antall kvadratmeter en bolig inneholder.
En dummyvariabel setter vi i utgangspunktet lik (0), men hvis det spesielle kjennetegnet inntreffer settes den lik (1). Dersom kjennetegnet inntreffer, vil man se dette som et skift i prisfunksjonen, hvor avstanden på skiftet avhenger av størrelsen på koeffisienten:

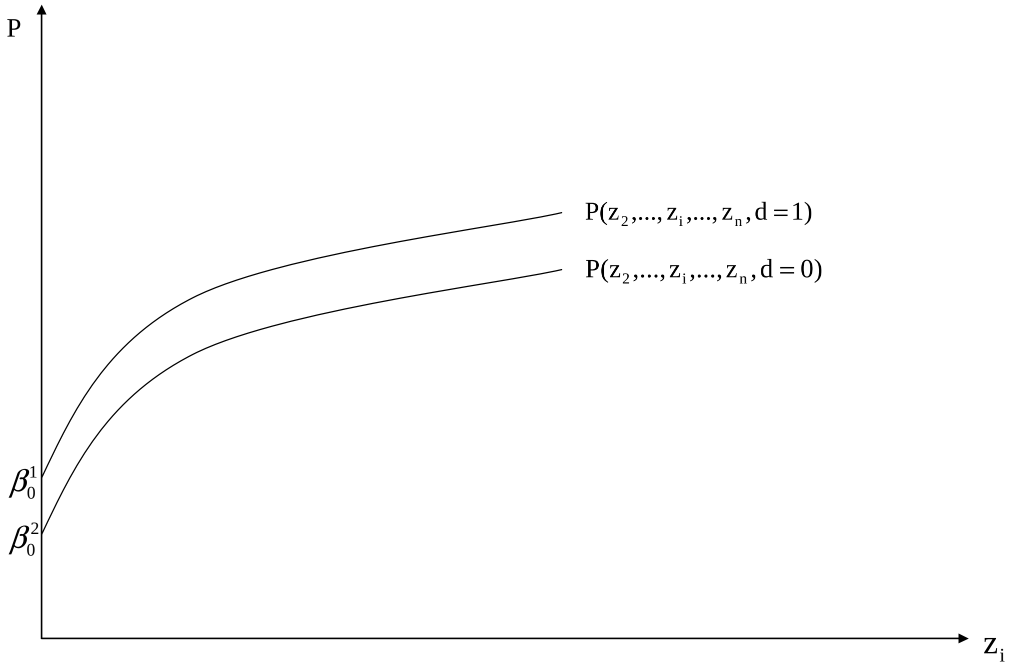


Figur 4.1 Virkningen av en dummyvariabel

For å unngå multikollinearitet holdes alltid en i hvert sett av dummyvariabler utenfor modellen, det vil si, dersom man bruker en månedsdummy, vil man bare ta med 11 måneder og ikke 12.
I denne figuren vises det at det skjer et positivt skift når en dummyvariabel er til stede, men det kan like godt være et negativt skift. Det kommer an på hvordan variabelen ved å være tilstede, påvirker prisen. Dersom dummyvariabelen er garasje, vil tilstedeværelsen av denne

antagelig gi et positivt skift, og dersom dummyen ”dårlig nabolag” er tilstede, vil det gi et negativt skift.

Det kan også være interessant å se hvordan dummyvariabelen påvirker en dobbellogaritmisk form. Den dobbeltlogaritmiske formen er en ikke-lineær funksjon, og vi vil se at avstanden når et skift oppstår, er større jo lengre ute i diagrammet man beveger seg. Men det vil være samme forholdstall på linjene (med og uten dummyen) og til z_i aksen:



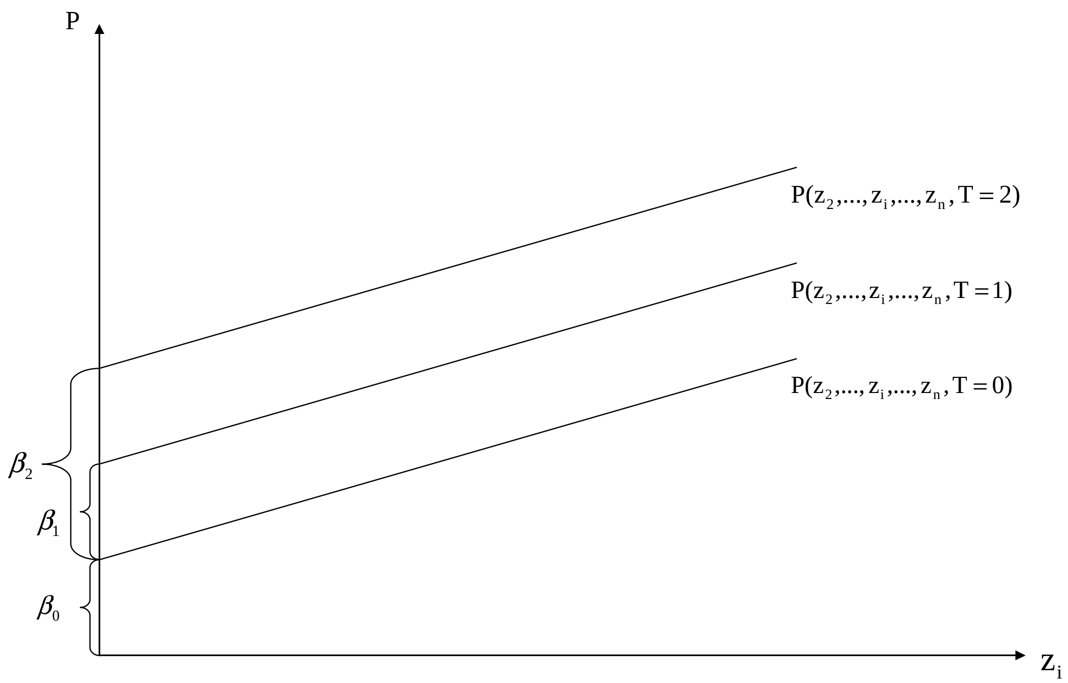
Figur 4.2 Virkningen av en dummy for en ikke- lineær prisfunksjon

Dersom man bruker logaritmeformen til den dobbeltlogaritmiske formen, vil det bli en rett linje, som den lineære vist ovenfor.

Når man måler prisutvikling over tid, vil det på grunn av inntektsendring og generell inflasjon over perioden, kunne føre til endring av verdisetting av de ulike boligattributtene. Tidsdummyer er dummyer som gir informasjon om tidspunktet salget har skjedd på, og om disse tidspunktene har påvirket prisen. Eksempler på dette kan være dummyer for salgsår og salgsmåneder. Koeffisientene til disse dummyene gir oss hvor stor påvirkning tidspunktet har for salgsprisen.

I kapittel to nevnte vi blant annet prisindekser, og at det stort sett er slik at priser stiger over tid.

I figur 4.3 er det nettopp prisstigning over tid, som er tilfellet. Her har vi tatt en periode på tre år(år T= 0,1 og 2). Salgsårsdummyenes koeffisienter betegnes med β_i, hvor *i* står for året T som gjelder. Koeffisientene forteller hvor stort skift det blir i konstantleddet til prisen, ved at salgsårsdummyen inntreffer. Av figuren kan vi se at hvert år gir et større skift i prisen enn det foregående året. Dette kan vi se på koeffisientene, fordi: (β_2) >(β_1) >(β_0).



Figur 4.3 Virkningen av en tidsdummy over tid

4.2 Hypotesetester

For å teste hypotesene fra kapittel 4, kan vi ta i bruk regresjonsanalyse. Regresjonsanalyse går blant annet ut på å teste hvilken påvirkning de forskjellige uavhengige variablene har på den avhengige variabelen. Vi kaller prisen for den avhengige variabelen, og attributtene, eller dummyene for uavhengige variabler.
Vi bygger mye på teorien til (Hegre, 2008) og (Jungeilges, 2008) i denne delen om hypotesetester.

En hypotese består av en nullhypotese (H0) og en alternativhypotese (H1). Hypotesen som er satt opp gir to forslag til hva som kan være sant. Vi skal finne ut hvilken av hypotesene som har størst sannsynlighet for å være den sanne.
Nullhypotesen testes ved hjelp av forskjellige verdier man finner av regresjonsanalysen. Ut ifra disse verdiene kan man bestemme om nullhypotesen skal forkastes og alternativhypotesen skal anses for mer sannsynlig, eller om nullhypotesen skal beholdes.

For å unngå at en tilfeldig observasjon fra vårt datamateriale gjør at man velger H1 når H0 er riktig, setter vi en kritisk grense slik at sannsynligheten for å velge feil er under et valgt nivå. Dette nivået kaller vi signifikansnivået α .
Vanligvis settes signifikansnivået til $\alpha = 0.05$, og det gjøres også her. Signifikansnivået måles opp mot p-verdien til den enkelte uavhengige variabelen som vi finner i regresjonsanalysen. P-verdien må være lavere enn 0,05, for at det skal være signifikans.

Det finnes flere forskjellige tester man kan utføre, og det skilles blant annet mellom ensidige og tosidige tester.
En ensidig test kan kjennetegnes ved at alternativhypotesen ikke er det motsatte av nullhypotesen. Eksempel på en ensidig test er:

H0: $\beta_i = 0$

H1: $\beta_i > 0$

En t-verdi er en annen måte å teste signifikansen. Den gir samme svar på signifikansen, som p-verdien. Den kritiske t-verdien for en ensidig test er 1.645, og dersom t-verdien er større enn dette er det statistisk signifikans. Betaen(eller koeffisienten) til den uavhengige variabelen som testes må ha samme fortegn som i hypotesen. Fra eksempelet på den ensidige testen, betyr dette at betaen må være større enn null. Vi kan også se på konfidensintervallet til den uavhengige variabelen. Konfidensintervallet viser hvilke verdier betaen ligger mellom, med 95 % sannsynlighet. Dersom intervallet kun dekker negative verdier, vet vi med 95% sannsynlighet at betaen ikke kan være positiv.

En tosidig test kjennetegnes ved at alternativhypotesen er det motsatte av nullhypotesen. Eksempel på en tosidig test er:

H0: $\beta_i = 0$

H1: $\beta_i \neq 0$

Den kritiske t-verdien til en tosidig test er på 1.96. Det vil si at hvis den uavhengige variabelens beta er ulik null, og t-verdien er høyere enn 1.96, forkastes nullhypotesen, og man beholder alternativhypotesen.

Det finnes også noe som kalles for en f-test, og den kan teste om alle koeffisientene samtidig ikke har noen innvirkning på den avhengige variabelen (prisen).

H0: $\beta_1 = \beta_2 = ... = \beta_i = 0$

H1: minst en $\beta_i \neq 0$

Ved en f-test kan vi også se på p-verdien. Dersom den er under 0,05 kan vi forkaste H0.
En annen måte å teste signifikansen til en f-test, vil være å sammenligne f-verdien som vi kan finne i regresjonsanalysen (F), mot en f-verdi($F\alpha, p, n-p-1$) vi kan finne i en F-tabell for signifikans. (α er signifikansnivået du ønsker å teste for, n er antall observasjoner, og p er antall uavhengige variabler)

Det finnes tabeller som gir f-verdier både på 10 %, 5 % og 1 % ‘s signifikansnivå. For å kunne bruke tabellen må man også vite hva frihetsgradene er:

Frihetsgrad v_1 = p

Frihetsgrad v_2 = n-p-1

F-verdien (F) fra regresjonsanalysen beregnes slik: F=$\frac{MSR}{MSE} = \frac{SSR}{p-1} / \frac{SSE}{n-p-1}$

Hvor:
SSR= sum squared residuals: Variasjon i prisen forklart av modellen
SSE= sum squared errors: Variasjon i prisen forklart av restleddet
MSR= mean square residual: gjennomsnittsfeil i variasjonen forklart av modellen
MSE= mean square errors: gjennomsnittsfeil i variasjonen forklart av restleddet

Dersom F > $F\alpha, p, n-p-1$ skal vi forkaste nullhypotesen.

Dersom man forkaster nullhypotesen, kan vi konkludere at minst en uavhengig variabel påvirker den avhengige variabelen. Dersom man ikke forkaster nullhypotesen, betyr det at dataene ikke kan fremskaffe bevis på at det er et lineært forhold mellom den avhengige og minst en uavhengig variabel.

.

4.3 Utledning av hypoteser

I denne delen av kapitlet skal vi utlede de hypotesene vi ønsker å få svar på, i tilknytning til vår problemstilling.

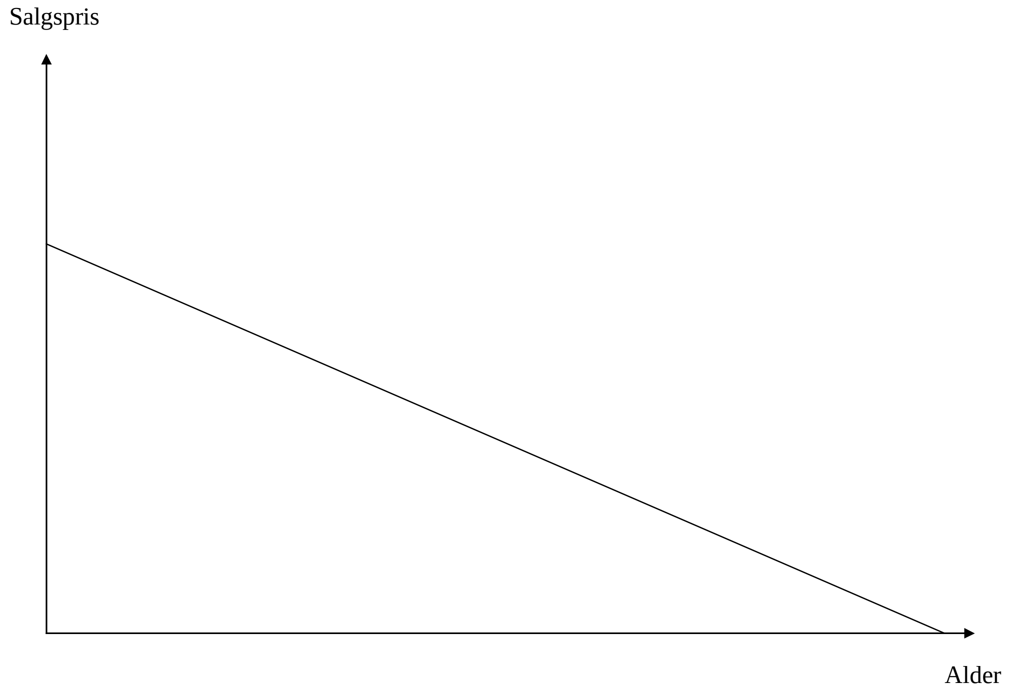
På grunn av at vi ser på prisutvikling over et lengre tidsrom, er alder på leiligheten på salgstidspunktet, en faktor vi vil se nærmere på. Alderen er beregnet som salgsår minus byggeår. Vi ønsker å finne ut om alderen påvirker salgsprisen negativt, eller om den ikke har noen betydning for salgsprisen. Dette gir oss følgende hypotese:

<u>Hypotese 1:</u>

H0: Alder har ingen betydning på salgsprisen

H1: Alder påvirker salgsprisen negativt

I figur 4.4 er alternativhypotesen skissert. Prisen vil da synke med alderen(antar en lineær prisfunksjon).



Figur 4.4 Sammenheng mellom salgspris og alder

Når det utføres en regresjonsanalyse, vil alle variablene som er med i analysen ha en koeffisient som forteller hvor mye akkurat den variabelen påvirker salgsprisen med. En positiv prisutvikling forutsetter at denne koeffisienten blir større med tiden. Vi ønsker dermed å sjekke om koeffisientene til salgsår blir større for hvert år. Koeffisienten til salgsåret kaller vi her, for enkelthetsskyld, α .

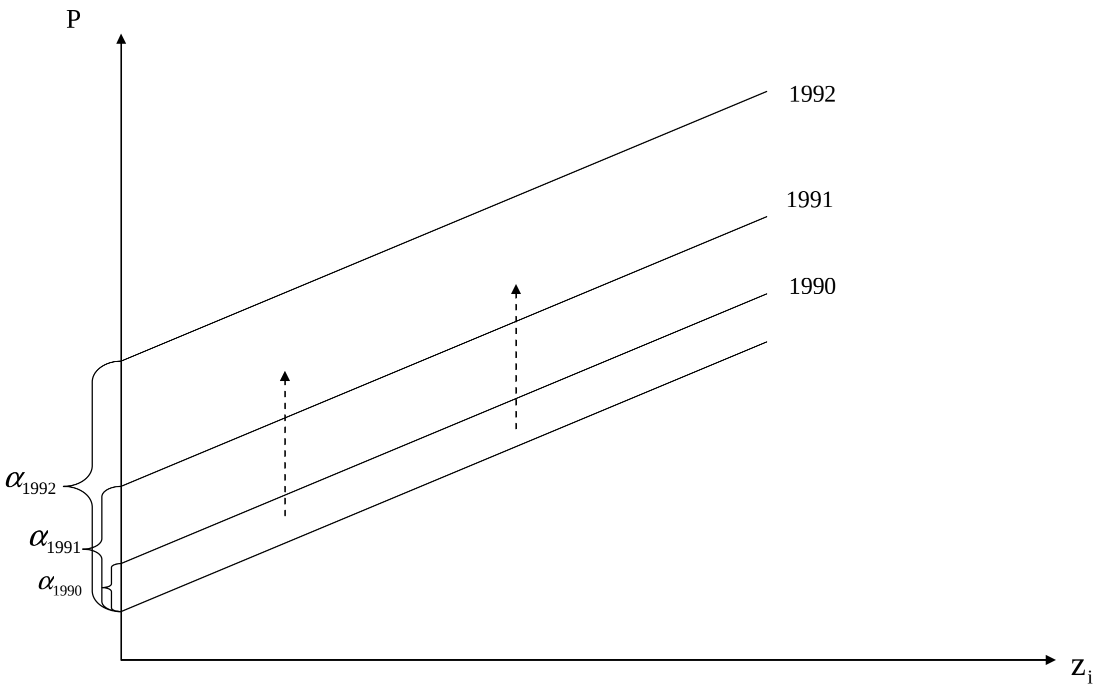
α_t er koeffisienten på tidspunkt(salgsår) *t*, og α_{t+1} betegner da salgsåret etter.

<u>Hypotese 2:</u>

H0: $\alpha_t = \alpha_{t+1}$, koeffisientene er like hvert år

H1: $\alpha_t < \alpha_{t+1}$, koeffisientene blir større for hvert år

I figur 4.5 kan vi se alternativhypotesen. Der kan vi se at α 'en til hvert salgsår blir større, og det fører dermed til at salgsprislinjen får større og større skift oppover.(Antar her en lineær prisfunksjon).



Figur 4.5 Økte salgsårskoeffisienter med tiden

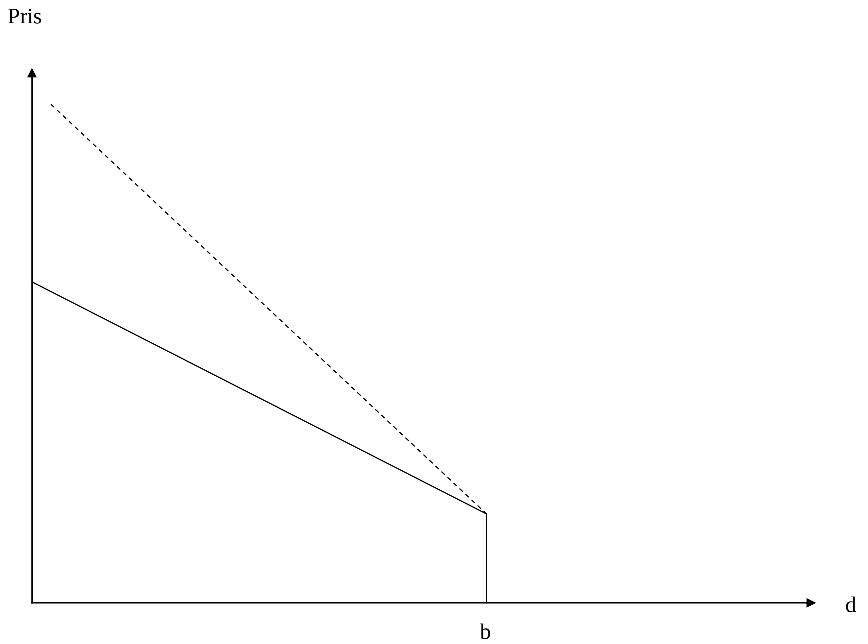
Ofte er det slik at man ser at salg av boliger gjort i en by, har høyere priser enn boliger solgt på landet. Men selv om prisene er ulike kan jo prisutviklingen ha vært lik, det vil si at prisene stiger prosentvis likt. Dette ønsker vi og undersøke nærmere, men i stedet for å se på sammenhengen mellom en by og på landet, ser vi på sammenhengen mellom sentralt og usentralt i en by. Det er da interessant å vite om også prisutviklingen over tid vil være forskjellig eller lik. Dette munner ut den siste hypotesen vår, som går direkte på problemstillingen:

<u>Hypotese 3:</u>

H0: Prisutviklingen i byens periferi har holdt tritt med prisutviklingen i byens senter

H1: Prisutviklingen i byens periferi har ikke holdt tritt med prisutviklingen i byens senter.

Denne hypotesen kan vi ta med bakgrunn i teorien i begynnelsen av kapitlet, om Alonso. Vanligvis tenker en seg at den prosentvise prisutviklinga er høyere sentralt enn usentralt, siden prisene ofte er høyere sentralt. I figuren nedenfor er akkurat denne antagelsen skissert. Her vil prisutviklingen over tid gi et høyere prisnivå sentralt, enn usentralt.
(b = bygrense, d = avstand fra sentrum).



Figur 4.6 Høyere prosentvis prisutvikling i Sentrum

5. Datamaterialet

5.1 Innsamling

For å finne ut om prisutvikling på leiligheter over tid, trengte jeg opplysninger om salg som har vært av leiligheter over en lengre periode.

En enkel og grei måte på å skaffe slike opplysninger på, er fra et selskap som heter Eiendomsverdi AS. Dette selskapet overvåker og registrerer blant annet alle salg av eiendommer som er registrert i det offentlige eiendomsregisteret i Norge. På internettsiden deres ligger en database som vi blant annet kan finne informasjon om enkelteiendommer, områder/utvalg av eiendommer og eiendommer med kriteriet nybygg. Søker vi på en enkelt eiendom, får vi opp de siste salgsopplysninger det er på denne. I tillegg til dette legges det ut nyheter om eiendomsmarkedet generelt.

Databasen er slik at man kan lage egne utvalg, og på den måten kunne få opplysninger fra for eksempel et spesielt område man er interessert i. Kriterier for utvalget er:

- Perioden (enkeltdato, eller datointervall)

- Status: solgt, registrert eller i markedet.

- Område:
- Fylke
- Kommune/Bydel
- Poststed (postnummer og sted)
- Boligtype:
- Type bolig (enebolig, leilighet, rekkehus, tomannsbolig, fritidsbolig)
- Fra- til- verdier på areal, byggeår og pris
- Eierform bolig
- Eierform tomt

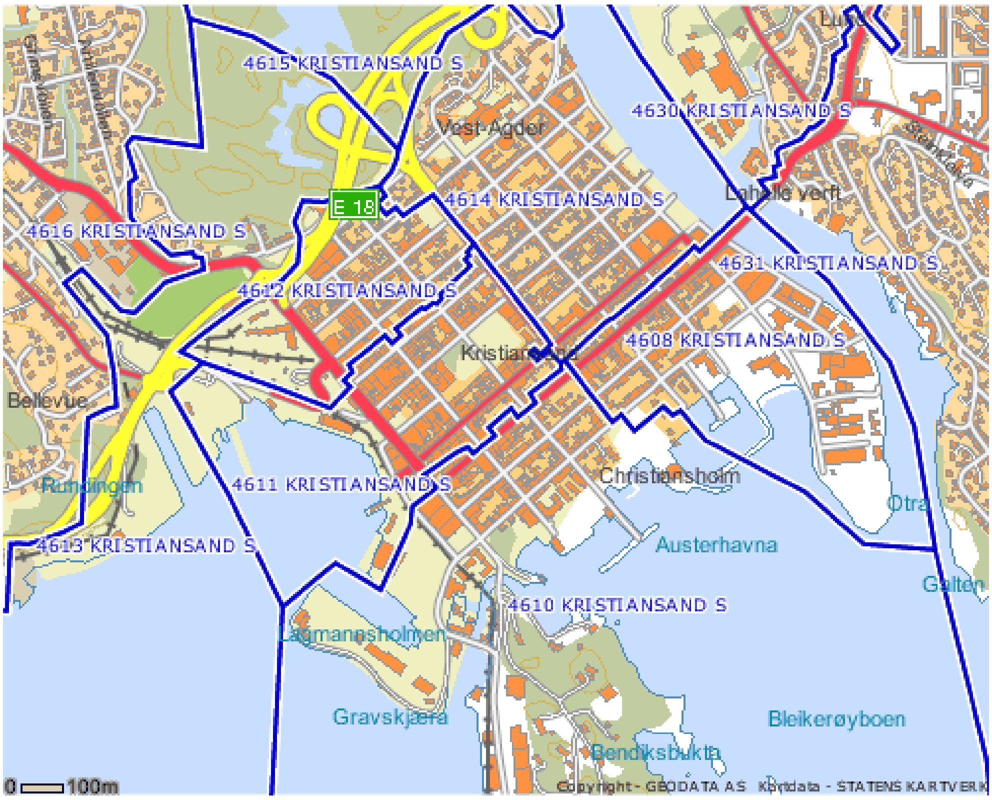
Kriteriene jeg valgte for mitt utvalg var:
Periode: 1. Januar 1985- 12. Februar 2010, status: solgt, Vest-Agder fylke, Kristiansand kommune, boligtype leilighet, med selveier form.

Perioden som utvalget ble valgt fra, kommer av at Eiendomsverdi kun har registrert salg fra 1.januar 1985. Alle leiligheter som var solgt i perioden var nødvendig informasjon for å se på prisutviklingen, derav ble ikke flere av kriteriene for utvalget satt til noe spesielt.
Eierformen selveier ble valgt etter råd fra veileder. Dette fordi det ofte kan bli vanskelig å beregne reell pris når det er fellesgjeld i tillegg til en salgspris, som ofte finnes i andels- og borettslag.
Grunnen til at boligtype leilighet ble valgt, kommer av at det e leiligheter vi ønsker å se på.

Under ”område” kan det plukkes ut hvilke poststed som ønskes opplysninger om salg fra, som nevnt ovenfor. På denne måten kunne det lages to utvalg, et for Kristiansand Sentralt, og et for Kristiansand Usentralt, ved å velge to grupper med postnummer.
De to utvalgene fungerte slik at det ene bare fanget opp salg som hadde skjedd i Kristiansand sentralt, og det andre søket fanget opp salg som hadde funnet sted i Kristiansand usentralt.
På Postens hjemmesider fant vi kart over Kristiansand med postnummergrenser, dette brukte vi for å velge hvilke postnummer som skulle være med i de to utvalgene.
For søket til Kristiansand sentralt, valgte vi å bruke alle postnumrene som fantes i sentrum av byen, totalt 5 stk, for å få flest mulig case derfra. Og for Kristiansand usentralt, valgte vi 6 andre postnumre fra forskjellige bydeler, som ikke grenset direkte til postnummergrensene i sentrum. På denne måten innrettet vi dataene slik at de to gruppene kan skilles fra hverandre.

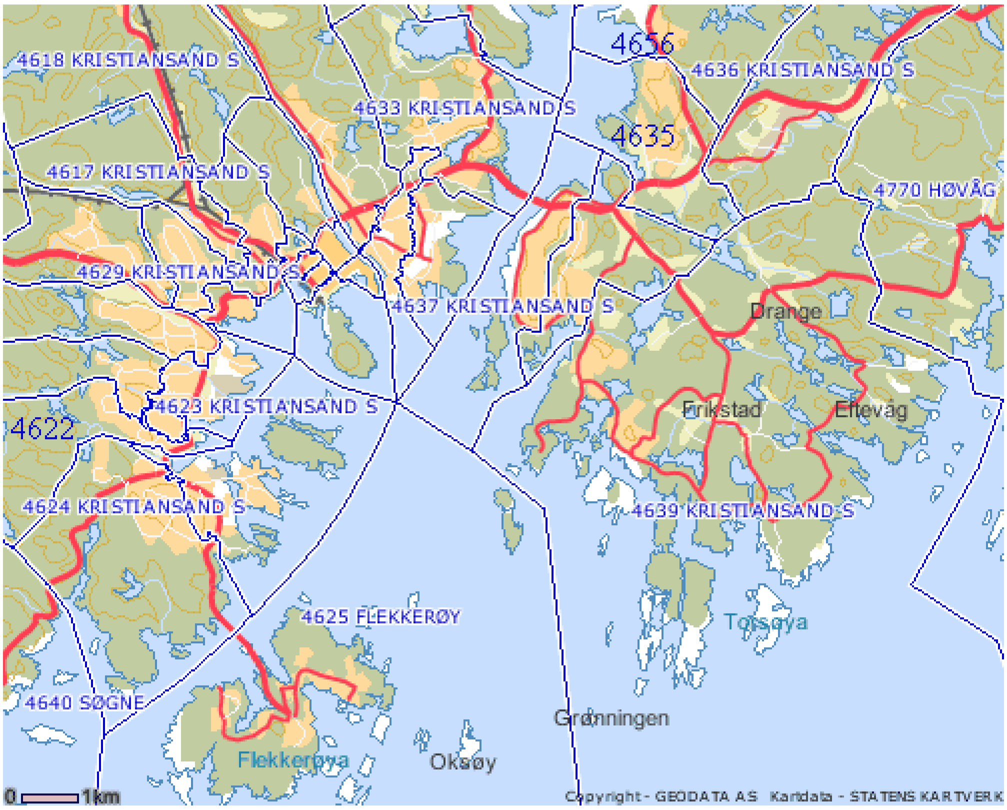
Postnumrene valgt til å representere sentral beliggenhet (kvadraturen) er:

- 4610 Kvadraturen (Dette dekker også Odderøya som ikke er en del av Kvadraturen, men siden det dekker en såpass stor del av Kvadraturen (ca. ¼ del), så tar vi det med)
- 4611 Kvadraturen
- 4612 Kvadraturen
- 4614 Kvadraturen
- Hadde også med postnummer 4608, men det viste seg at det ikke var registrert noen salg på dette postnummeret, derfor er det blitt utelatt av resten av oppgaven.



Figur 5.1 Postnummergrensekart, Sentralt
Posten,(2010).

Postnumrene valgt til å representere usentral beliggenhet er:
4622 Vågsbygd
4623 Voie/Brøvik/Møvik
4624 Voie
4635 Øvre Timenes, Hånes
4639 Korsvik, Dvergsnes
4656 Hamresanden



Figur 5.2 Postnummergrensekart, Usentralt (De postnumrene som ikke stod på kartet, er skrevet på med mørkeblå skrift der de hører hjemme: 4622, 4635 og 4656) Posten,(2010).

Eiendomsverdi gir flere opplysninger om hver solgte leilighet. I et enkelt søk med de kriteriene som er lagt til grunn kommer følgende opplysninger (6stk):

- Dato for salg
- Adresse
- BOA/P-rom
- Byggeår
- Pris
- Megler

I tillegg er det mulig å gå inn på hvert enkelt salg eller case, og finne flere opplysninger. Legger ikke vekt på dem her, i og med at det er begrenset med tid på oppgaven.

Det som er viktig, er å ta med de opplysningene om salget, som man tror gir størst innvirkning på prisen, og som kan gi best mulig svar på det vi prøver å finne ut. Det vil si blant annet hvilke attributter ved boligen man tror har størst innvirkning på prisen. I vårt tilfelle er

tidsverdier, som salgstidspunktet, og alder på leiligheten viktige. Tidsverdier er viktige siden vi ser på prisutvikling over tid.
I tillegg er det greit å vite hvor stor leiligheten er, siden de fleste undersøkelser sier at areal er viktig for prisen. Jo større leiligheten er, jo høyere pris. Har vi størrelsen kan vi også finne en pris per kvadratmeter, som er en grei måte å sammenligne prisene på leilighetene.

For ikke å gjøre analysen for komplisert, har vi valgt ut de opplysningene som kom frem ved et enkelt søk, bortsett fra hvilken megler som ble brukt.

5.2 Bearbeiding

Å bearbeide datamaterialet kan blant annet bety at vi må endre litt på noen av dataene, for at de skal være til nytte for vårt formål. Det kan også være tilpassing, og rensing av dataene. Ovenfor nevnte vi de seks opplysningene som fulgte av hver salgscase, og nedenfor beskrives hva som gjøres med dem:

- Salgsdato: Skiller ut salgsmåneden og salgsåret fra hvert salgscase. Disse lager vi til dummyvariabler, med dummyer for hvert månedsnummer og for hvert år. På den måten kan vi sjekke om de begge har påvirkning på prisen hver for seg.

- Adresse: Skiller ut postnummeret fra hvert salgscase. Dette lager vi til en dummyvariabel. En dummy for hvert postnummer.

- BOA/P-rom: (Betydning: Antall kvadratmeter til boligformål) Dette er en kontinuerlig variabel, som vi vil kalle for kvadratmeter fra nå.

- Byggeår: Dette er en kontinuerlig variabel, som vi beholder slik den er.

- Pris: Dette er en kontinuerlig variabel, og vi vil kalle den for salgspris heretter.

- I tillegg: Vi lager en kontinuerlig variabel som vi kaller for alder.

Hvilken megler som ble brukt, har de fleste undersøkelser vist seg og ikke hatt noe særlig påvirkning på prisen, og derfor tas ikke denne variabelen med. Denne opplysningen manglet dessuten i en stor del av materialet.

Som tidligere nevnt i kapittel 4, er prisen en avhengig variabel, og denne blir påvirket av de uavhengige variablene. Vår avhengige variabel er da salgsprisen. De uavhengige variablene er:

<u>Salgsmåned</u>: Er det januar, eller er det juli? Denne er trukket ut som en egen variabel for å se om det kan være forskjell på prisen på grunn av hvilken måned den selges i.

<u>Salgsår</u>: Variabelen er viktig først og fremst for å se hvordan prisen har utviklet seg over tid. Har prisnivået steget, sunket eller stått stille over perioden?
Men den er også viktig for å beregne alderen på leiligheten på salgstidspunktet, siden alder kan si noe om hvilken stand boligen er i. Jo eldre bygning, jo større sjanse for at det må regnes med reparasjoner og oppussing og lignende.

<u>Postnummer:</u> Variabelen forteller oss om leiligheten ligger sentralt eller usentralt. Dette er viktig fordi vi ønsker å se på forskjellen i prisutviklingen på de to stedene.

<u>Kvadratmeter:</u> Variabelen forteller hvor stort bruksareal det er i leiligheten. Størrelsen på en bolig har ofte en stor utslagskraft på prisen. Ved å vite hvor stor en leilighet er, kan man finne pris per kvadratmeter, og dermed gjøre prisen mer sammenlignbar.

<u>Byggeår:</u> Variabelen forteller hvilket år leiligheten ble bygget.

<u>Alder</u>: = Salgsår – Byggeår.
Alderen på salgstidspunktet gir mer korrekt informasjon om leiligheten enn å se kun på byggeåret. Dette fordi perioden vi ser på går over tid. Variabelen kan blant annet si noe om standarden på leiligheten. Som regel gir økt alder på leiligheten, en lavere pris.

Noen av salgscasene manglet en eller flere av disse opplysningene. Blant annet var det ikke oppgitt antall kvadratmeter, byggeår og salgspris på noen. Disse casene dropper vi, siden det ville blitt mye ekstra jobb å fylle disse hullene.
I tillegg var det noen case som oppgav at leiligheten var solgt året før den var ferdigstilt. Dette gjør at alderen på leiligheten ville bli negativ. Så disse dropper vi også for enkelhelts skyld.
Når det gjelder manglende/gale opplysninger, kan dette føre til at man feiltolker situasjonen, og dermed danner seg feilaktige konklusjoner. Det kan også gi fullstendig feil svar. Derfor må man gjøre noe med disse dataene, enten å søke for å få på plass de opplysningene som mangler, eller å forkaste dem som ikke brukbare case.
I og med at det forelå såpass mange case, så valgte jeg å ta ut de casene som manglet opplysninger. Antall case som brukes videre til analysering er:

Tabell 5.1 Antall case Sentralt

SENTRALT: antall case i utgangspunktet:	**1981**
- antall case der byggeår > salgsår	2
- antall case uten oppgitt antall kvadratmeter	629
- antall case uten oppgitt byggeår	33
- antall case uten sikker salgspris	26
= Antall case fra sentralt videre til analyse	**1291**

Tabell 5.2 Antall case Usentralt

USENTRALT: antall case i utgangspunktet:	**708**
- antall case der byggeår > salgsår	1
- antall case uten oppgitt antall kvadratmeter	172
- antall case uten oppgitt byggeår	1
- antall case uten sikker salgspris	13
= Antall case fra usentralt videre til analyse	**521**

Totalt sentrale og usentrale case blir da= 1812

Tidsperioden som vi velger å se på:
Det er kun registrert et salg på de kriteriene nevnt ovenfor, i tidsrommet 1.januar 1985 til 31. desember 1989. Å ha kun et salg fra et år vil ikke gi noen stor mening i estimeringen vi skal gjøre senere, så dette salget droppes for enkelhets skyld. Tilfeldigvis manglet akkurat dette salget opplysning om byggeår, og derfor er det allerede trukket ut og regnet med i tabellene ovenfor.

5.3 Presentasjon

I denne delen beskrives de dataene jeg har funnet. Måten jeg skal gjøre dette på, er gjennom deskriptiv statistikk. Det vil si at dataene organiseres og fremstilles på en måte som skal være lette å forstå og tolke.
Statistikk forbindes ofte med ulike diagrammer, tabeller og grafiske figurer. Det er gjennom å bruke slike metoder man omformer dataene man har, til noe som gir mening.
En frekvenstabell kan benyttes dersom man ønsker å se hvor ofte en type verdi eller variabel dukker opp. Den teller opp og ordner dataene. Et stolpediagram kan vise det samme. Aarsland (2009)

Noen statistiske uttrykk:

- Gjennomsnittsverdi: summen av alle verdiene på observasjonene deles på antall observasjoner.

- Varians: gir spredningen i verdiene til observasjonene våre. Varians = $\frac{1}{n}((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- Standardavvik: måler avvik fra gjennomsnittverdien. Denne finnes ved å ta kvadratroten av variansen.

- Minimum og maksimums verdier: det vil si den minste og den største verdien i tallmaterialet.
  Ubøe og Jørgensen(Side 24-25)

I tabell 5.3 er det en oversikt over alle variablene og de statistiske verdiene deres(deskriptiv statistikk).
For dummyvariablene vil minimumsverdien alltid være 0, og maksimumsverdien alltid være 1. En 0 forklarer at kriteriet ikke er oppfylt, mens 1 at det er.
For alder er det en minimumsverdi på -1. Dette fordi noen av leilighetene antagelig er solgt året før de er ferdigstilte.
Månedsdummyene er gjengitt med navnet på måneden akkurat i denne tabellen, (ellers i oppgaven står det MND1, MND2,…, MND12), og salgsårsdummyene er gjengitt som AR1990, AR1992 og så videre. Postnummerdummyene kjennetegnes med PNR.
For mine 1812 case finner jeg følgende statistiske verdier på mine variabler:

Tabell 5.3 Deskriptiv statistikk

Variabel	Gjennomsnitt	Standard Avvik	Minimum	Maximum
Salgsmåned	6.747792	3.331471	1	12
Salgsår	2003.087	4.883931	1990	2010
Postnummer	4617.36	8.459612	4610	4656
Kvadratmeter	65.64128	29.52006	18	285
Byggeår	1965.114	42.44461	1740	2009
Salgspris	1339518	854676.8	150000	7200000
Alder	37.97241	42.11641	0	267
Januar	.0684327	.2525566	0	1
Februar	.066777	.2497044	0	1
Mars	.0811258	.2731036	0	1
April	.080574	.2722548	0	1
Mai	.0717439	.2581347	0	1
Juni	.0960265	.2947089	0	1
Juli	.0756071	.2644414	0	1
August	.1181015	.3228174	0	1
September	.0866446	.2813912	0	1
Oktober	.0960265	.2947089	0	1
November	.0761589	.2653255	0	1
Desember	.0827815	.2756277	0	1
AR1990	.00883	.0935782	0	1
AR1991	.0187638	.1357272	0	1
AR1992	.0115894	.107058	0	1

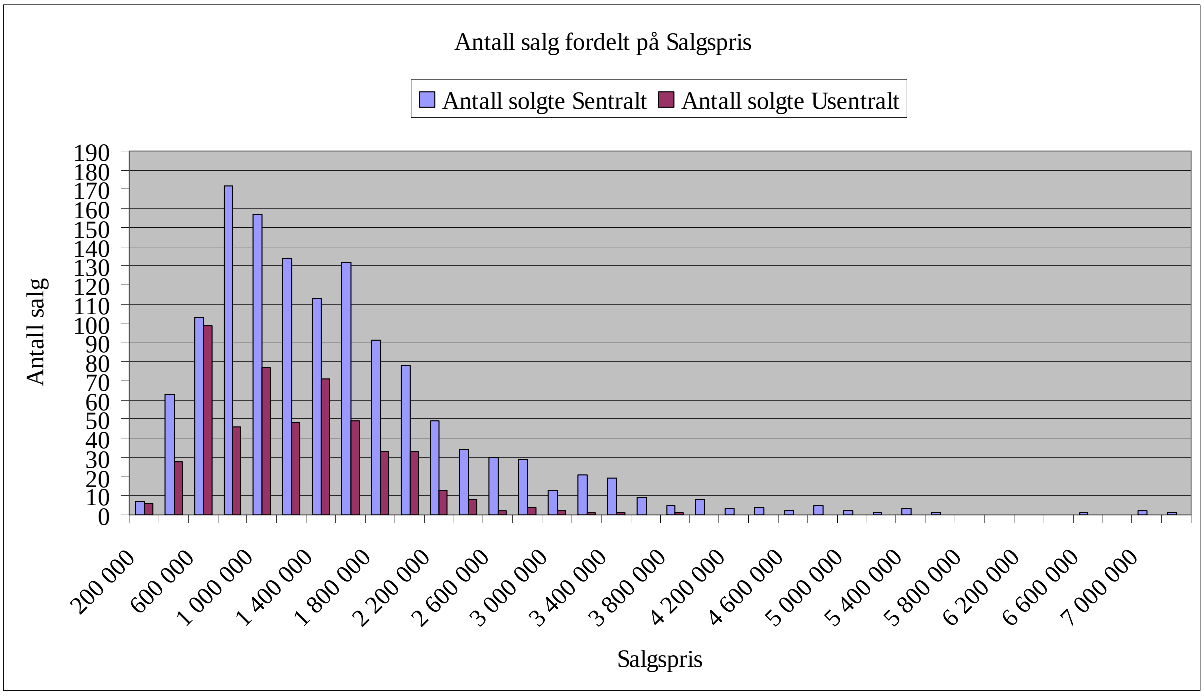
AR1993	.0149007	.1211889	0	1
AR1994	.0209713	.1433278	0	1
AR1995	.0226269	.148752	0	1
AR1996	.0375276	.1901032	0	1
AR1997	.0353201	.1846385	0	1
AR1998	.0264901	.1606318	0	1
AR1999	.0364238	.187394	0	1
AR2000	.0331126	.1789799	0	1
AR2001	.040287	.1966858	0	1
AR2002	.044702	.2067058	0	1
AR2003	.0529801	.2240557	0	1
AR2004	.0932671	.2908867	0	1
AR2005	.1081678	.3106779	0	1
AR2006	.1125828	.3161694	0	1
AR2007	.1065121	.3085771	0	1
AR2008	.0800221	.2714023	0	1
AR2009	.0877483	.2830067	0	1
AR2010	.0071744	.0844207	0	1
PNR4610	.192053	.3940232	0	1
PNR4611	.0309051	.1731083	0	1
PNR4612	.0695364	.2544343	0	1
PNR4614	.4199779	.4936912	0	1
PNR4622	.0104857	.1018893	0	1
PNR4623	.0695364	.2544343	0	1
PNR4624	.0800221	.2714023	0	1
PNR4635	.080574	.2722548	0	1
PNR4639	.0452539	.2079178	0	1
PNR4656	.0005519	.023492	0	1

Salgspris

Figur 5.3 viser antall boliger solgt på salgspris. Har valgt å vise salg sentralt og salg usentralt i hver sine farger, slik at vi kan se forskjellen.

Dette er vist ved intervall på 200 000 kroner. For eksempel er det flest antall solgte leiligheter sentralt, (omkring 170stk) og de er solgt for mellom 801 000 – 1 000 000 kroner.

Flest antall solgte leiligheter solgt usentralt er i intervallet 401 000kroner-600 000kroner (nesten 100 stk).

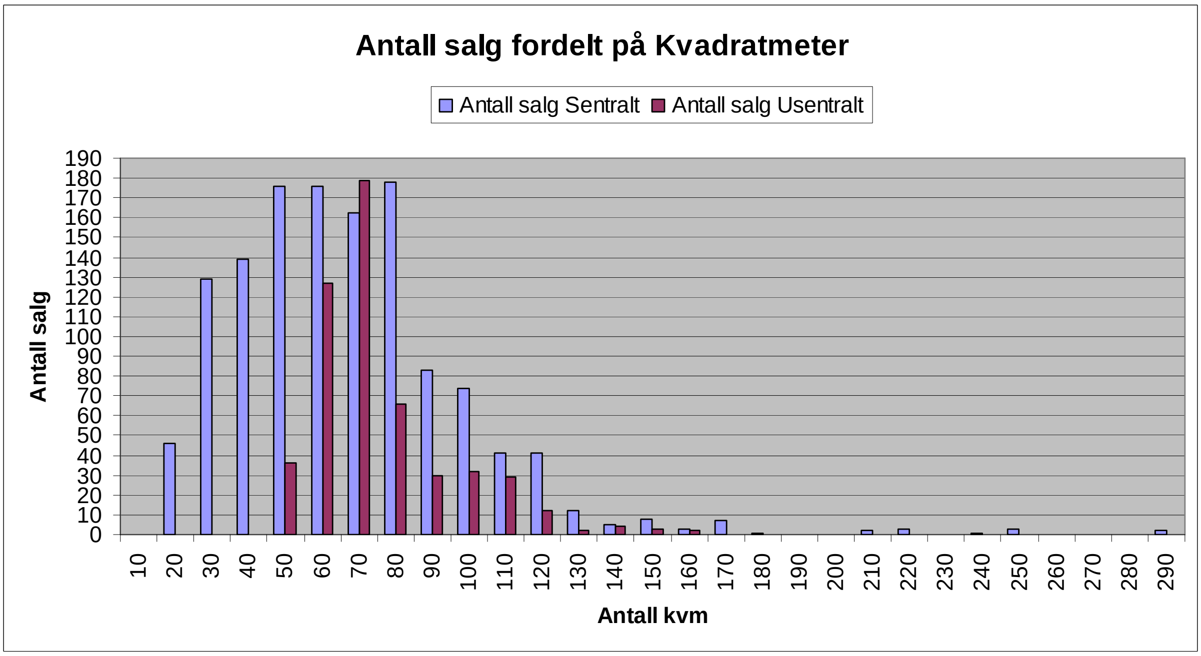


Figur 5.3 Antall salg fordelt på salgsprisen

Kvadratmeter

I figuren 5.4 kan vi se trenden for hvor mange kvadratmeter som vanligvis blir solgt. Igjen kan vi se sentralt og usentralt hver for seg. Legger merke til at fordelingen til salg sentralt er lenger til venstre i bildet. Dette gjenspeiler plassrealiteten, om at å kjøpe sentralt gir liten plass, til en dyrere pris. Flest observasjoner er det likevel på leiligheter fra 71 – 80 kvadratmeter for de som er solgt sentralt, noe som er høyere antall enn de solgt usentralt: 61-70 kvadratmeter.

(Men husker at det er flere case fra salg som er skjedd sentralt, enn det er for salg usentralt.)

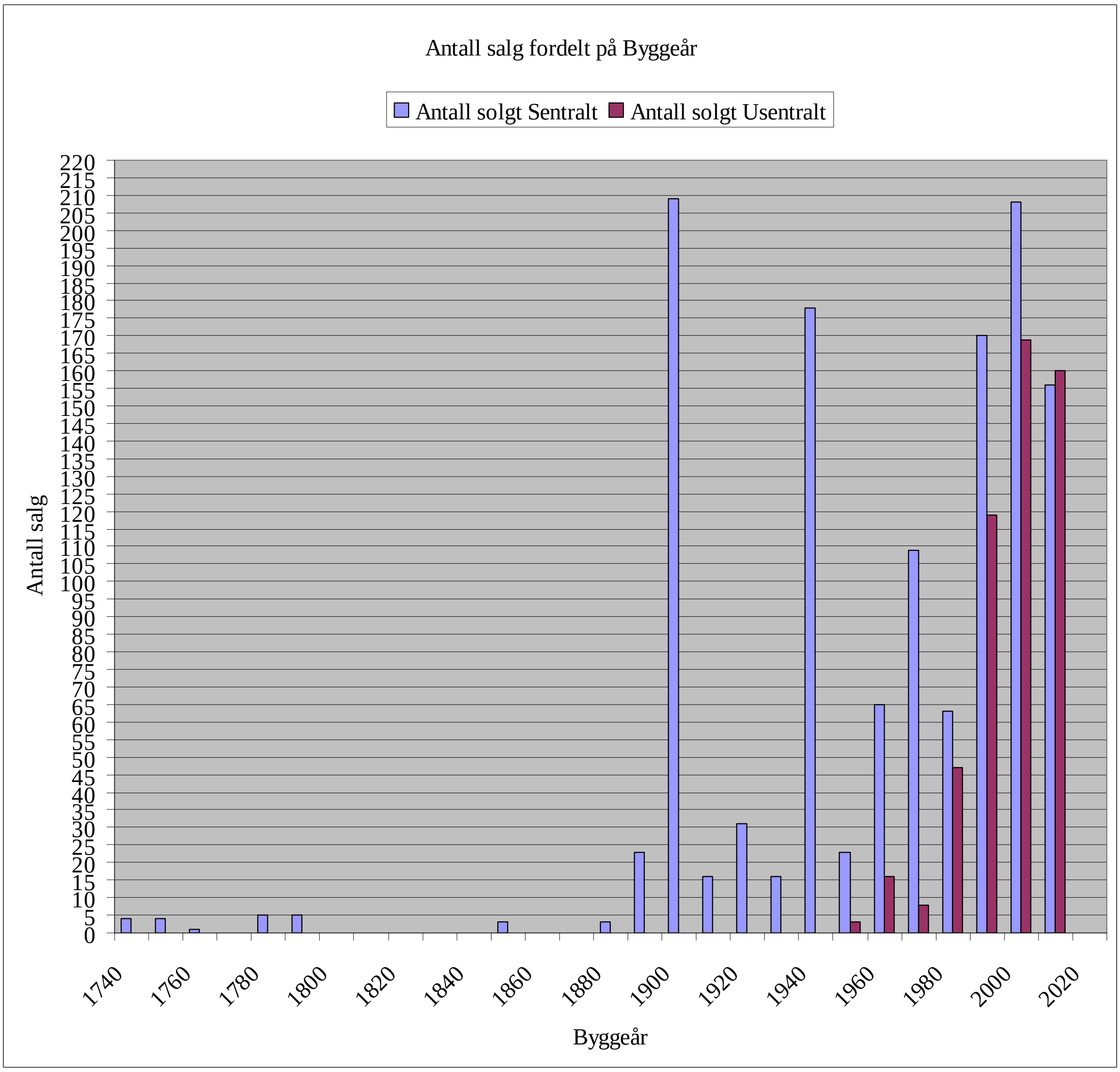


Figur 5.4 Antall salgscase fordelt på størrelsen av leiligheten

Byggeår

I figur 5.5 kan vi se et diagram som forteller hvor mange leiligheter det er solgt, fordelt på byggeåret de er bygget i. Fordelingen viser at leiligheter sentralt er bygd tidligere, enn de som er bygd usentralt.

I Kristiansand sentrum er de fleste husene bygd sammen i rekker i hvert kvartal, og naboer så tett inntil kan føre til at riving og nybygging kan bli ekstra kostbart. Da er det kanskje lettere å fikse opp det som allerede er, i stedet for å bygge et nytt. I tillegg er også det meste av plassen i sentrum utbygd, slik at ledige tomter er sjeldne. Disse tingene forekommer ikke like ofte usentralt.

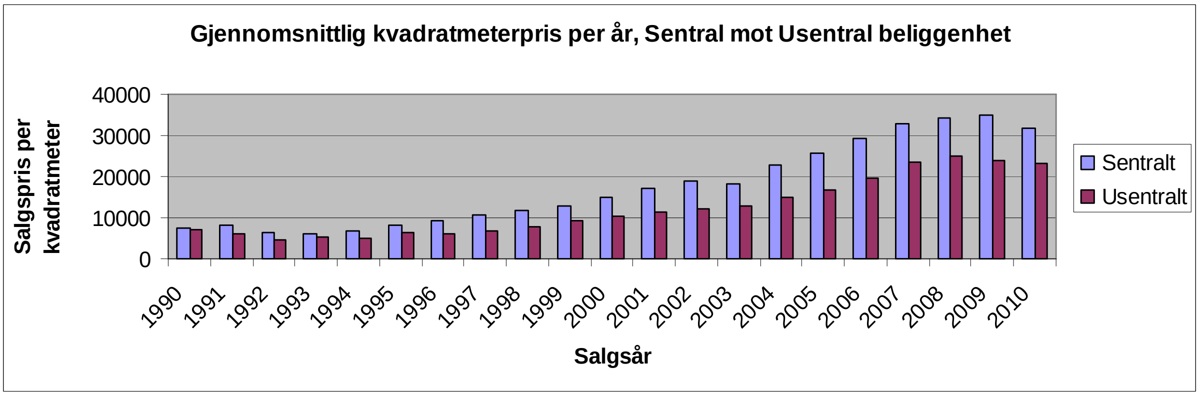


Figur 5.5 Antall salgscase fordelt på de forskjellige byggeårene

Gjennomsnittlig kvadratmeterpris

I figur 5.6 kan vi se at gjennomsnittlig kvadratmeterpris har mer eller mindre økt jevnt og trutt gjennom hele perioden, med unntak av et par nedturer. Vi ser nedgang i 1992-1993 og en liten nedgang i år 2010. Nedgangen kan skyldes enkelte observasjoner som gir store utslag, blant annet at det i 2010 var et ekstra kort salgstidsrom(frem til 12.februar). Eventuelt kan det være endringer i økonomien på tidspunktet.

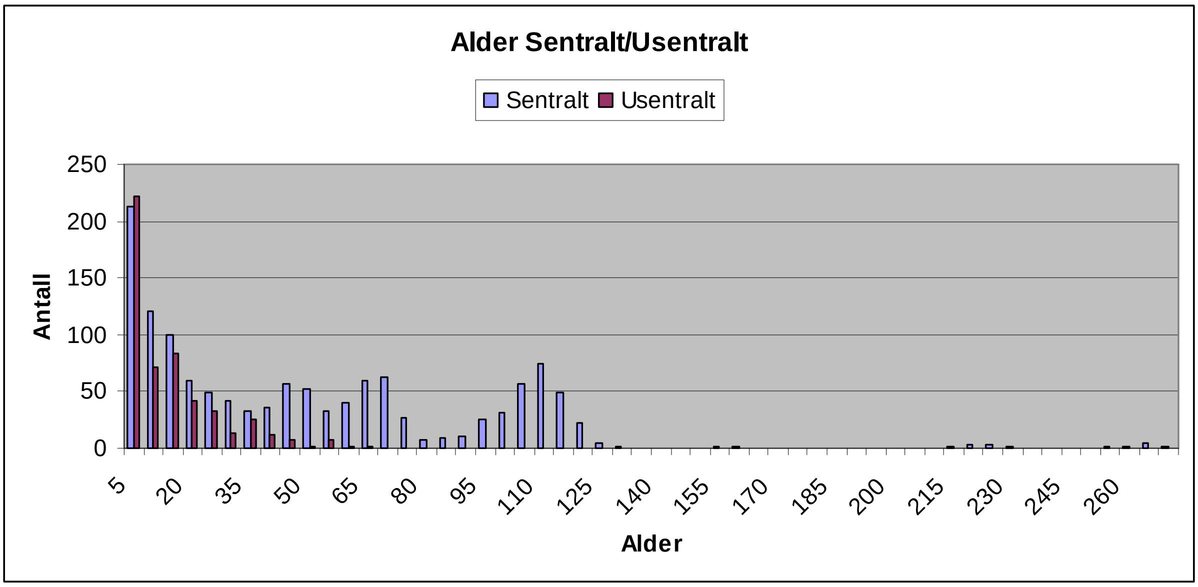
Sammenligner vi prisnivået sentralt mot usentralt, ser vi at dette er høyere sentralt i byen. Men selve prisutviklingen over tid, ser ut til å være ganske lik.



Figur 5.6 Gjennomsnittlig kvadratmeterpris per år

Leilighetens alder på salgstidspunktet

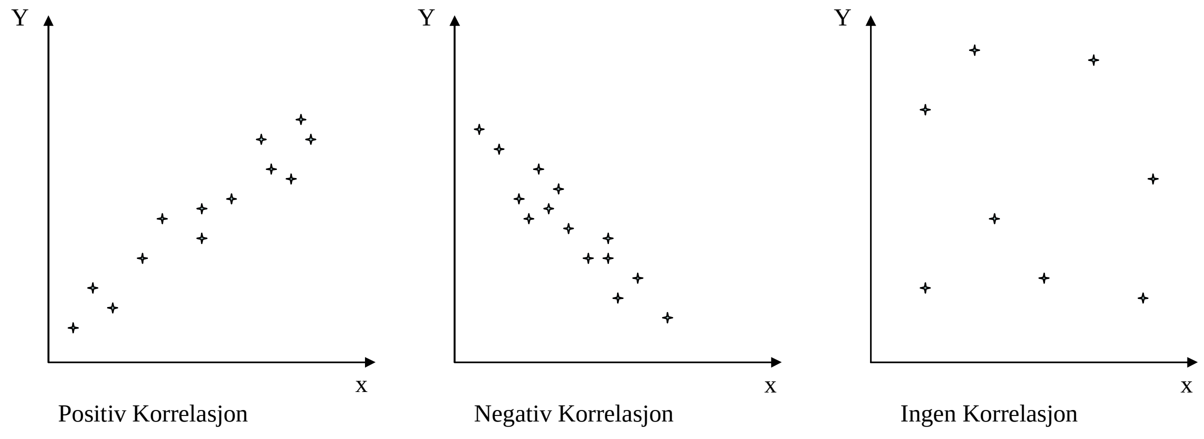
I figur 5.7 kan vi se fordelingen av solgte leiligheter fordelt på alder.
Fordelingen på alder til leiligheter solgt sentralt, er nokså spredt utover, med flere topper. De aller fleste solgt når de er 5-10 år gamle, men en del er også solgt når de er 50 og 110 år gamle.
For usentral beliggenhet er aldersfordelingen mest konsentrert rundt 5-10 år gamle leiligheter. Så selv om sentralt/usentralt har forskjellige fordelinger, er det likevel 5-10 år gamle leiligheter som selges hyppigst på begge stedene.



Figur 5.7 Antall salg fordelt på leilighetens alder på salgstidspunktet

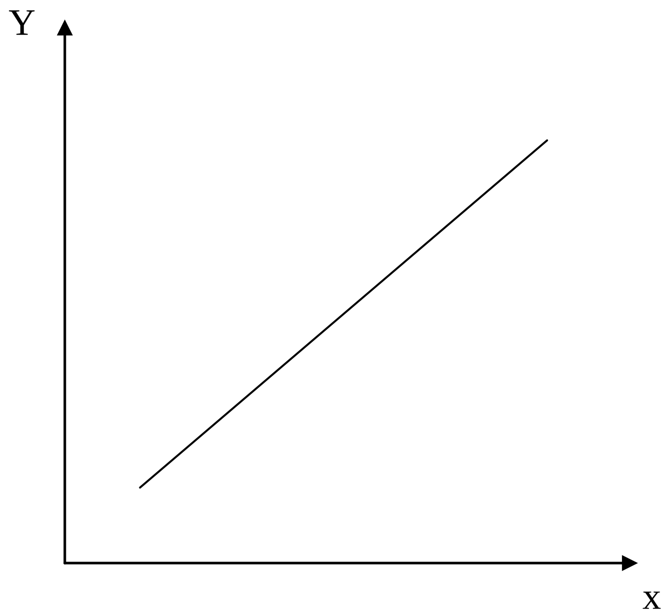
5.4 Multikollinearitet

Dersom verdiene til to variabler varierer sammen, er det korrelasjon mellom dem. Det kan være både positiv og negativ korrelasjon mellom to variabler.
Det er positiv korrelasjon når en økning i verdien på den ene variabelen, alltid svarer til en økning i verdien til den andre variabelen. Det samme gjelder dersom verdien på den ene variabelen synker, da gjør den andre variabelen det også.
Styrken av korrelasjonen mellom to variabler måles på en skala fra -1 til +1. Positiv korrelasjon har en styrke mellom 0 og 1. Dersom en korrelasjonsverdi på 1 oppstår, er det perfekt positiv korrelasjon, og da øker eller reduserer begge variablenes verdier akkurat likt.
Det er negativ korrelasjon når en økning i verdien på den ene variabelen, alltid svarer til en reduksjon i verdien til den andre variabelen. Negativ korrelasjon er det man ofte ønsker å oppnå ved å ha flere typer aksjer i en portefølje. På den måten så taper man ikke mye penger, fordi en aksjes tap motlignes en annen aksjes gevinst. Styrken på den negative korrelasjonen ligger mellom -1 og 0, hvor -1 er perfekt negativ korrelasjon. Perfekt negativ korrelasjon er når variablenes verdier varierer nøyaktig likt hver sin vei.
Ingen korrelasjon betyr at variablene ikke varierer samtidig, og dette betegnes med verdien 0.
Nedenfor vises de forskjellige typene med korrelasjon, med hensyn på de to variablene x og y:



Figur 5.8 Korrelasjon

Multikollinearitet er en annen betegnelse for perfekt korrelasjon. Det blir problemer dersom det er mye korrelasjon mellom variablene. Det ene problemet er å skille variablenes effekter fra hverandre, og det andre er at signifikansverdiene blir upålitelige. I figur 5.9 kan vi se multikollinearitet mellom variabel x og y:



Figur 5.9 Multikollinearitet

Multikollinearitet er noe som oppstår sjeldent, men dersom vi måler korrelasjonen til en variabel med seg selv, vil vi naturlig nok få det.

I korrelasjonsmatrisen i tabell 5.4, har vi tatt med noen av dummyvariablene for år, i tillegg til variablene kvadratmeter, byggeår, alder, og salgspris. Her kan vi se at alle variablene korrelerer helt perfekt med seg selv. Grunnen til at det ikke står noen verdi på oversiden for diagonalen, er fordi disse verdiene er like de som står på nedsiden, bare speilvendt. Det vil si, korrelasjon mellom år 1991 og 1990 (-0,0130), er lik korrelasjonen mellom år 1990 og 1991.

De to variablene det er høyest korrelasjon mellom, er alder og byggeår. Her er det nesten perfekt negativ korrelasjon (-0,9934), og det er jo ikke så rart. Fordi jo nærmere byggeåret er år 2010, jo lavere blir alderen på leiligheten (verdiene til variablene beveger seg i motsatt retning).
Ser vi på korrelasjonen mellom alder og kvadratmeter(0,0534), er denne positiv. Det er ikke en sterk korrelasjon, men det tyder på at jo eldre leiligheten er, jo større blir den.
Salgsprisvariabelen varierer negativt med årsdummyene og positivt med resten av variablene.
Det at salgsprisen korrelerer positivt med alder, er jo litt merkelig, siden man ofte antar at

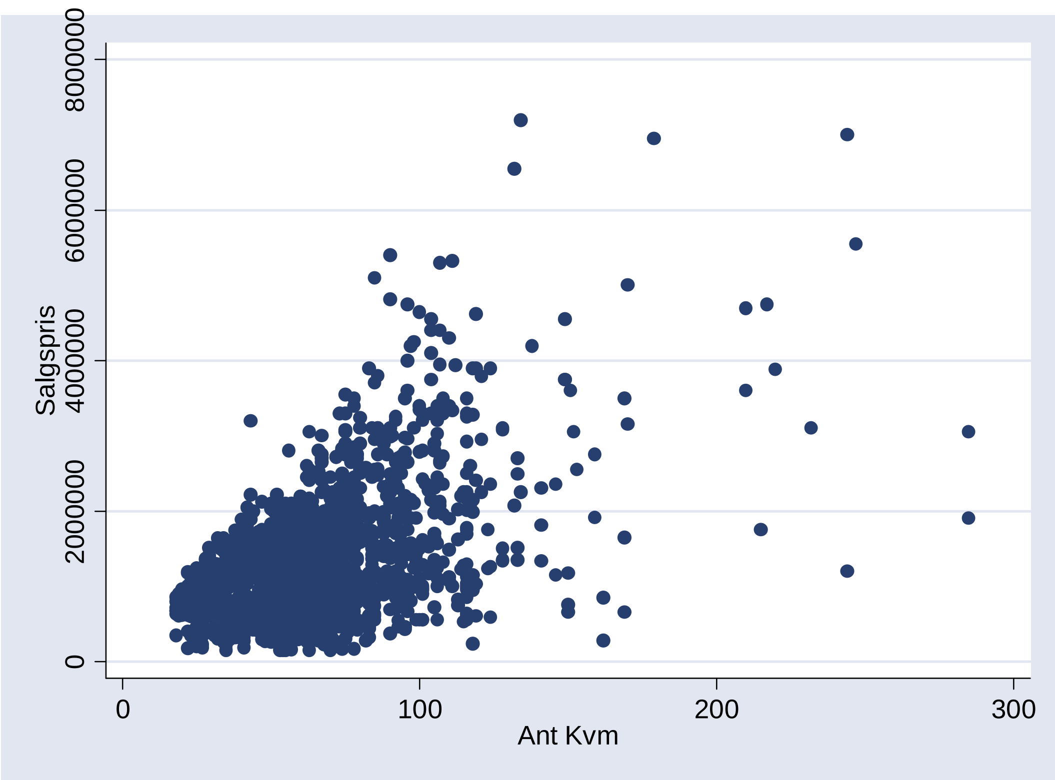
alderens verdi vil gå i motsatt retning av salgsprisens verdi. Det vil si, jo eldre leiligheten er, jo lavere pris. Men det var jo ikke så sterk korrelasjon (0,0063).

Tabell 5.4 Korrelasjonsmatrise

	AR1990	AR1991	Kvm	Byggeår	Alder	Salgspris
AR1990	1.0000					
AR1991	-0.0131	1.0000				
Kvm	-0.0052	0.0116	1.0000			
Byggeår	-0.0364	0.0071	-0.0627	1.0000		
Alder	0.0073	-0.0469	0.0534	-0.9934	1.0000	
Salgspris	-0.0954	-0.1408	0.5106	0.0609	0.0063	1.0000

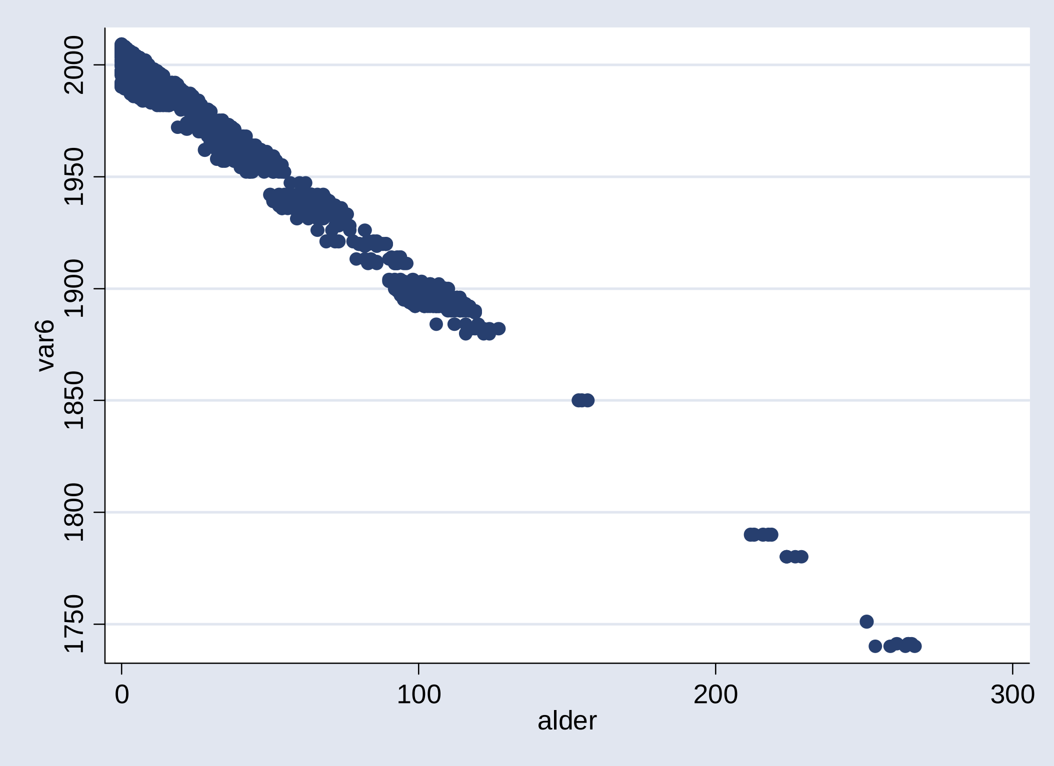
Nedenfor vises det grafisk i to figurer hvordan korrelasjonen mellom variabelen kvadratmeter og salgspris, og mellom alder og byggeår ser ut.

I den første figuren, 5.8, kan vi se et visst mønster, og tendens til positiv korrelasjon, men likevel ikke den store systematikken mellom variablene.



Figur 5.10 Korrelasjon mellom antall kvadratmeter og salgspris

Mellom variablene alder og byggeår i figur 5.9 er det sterk negativ korrelasjon, som vi kan se.



Figur 5.11 Korrelasjon mellom alder på salgstidspunktet og byggeår

6. Analyse og Estimeringsresultater

I dette kapitlet skal teorien fra kapittel 4 brukes og testes. Det ble nevnt to former for modeller, hvor den ene var en lineær form, og den andre en dobbeltlogaritmisk form. Vi skal teste for de ulike forutsetningene som kjennetegner en god modell, og ut i fra det velge en av formene. Denne brukes så til å teste hypotesene fra slutten av kapittel 4.
Regresjonsprogrammet som brukes er STATA.

6.1 Innledende analyser

Lineær regresjon med en variabel

Til å begynne med presenteres den enkle regresjonen med kun en uavhengig variabel. Den avhengige variabelen er salgspris, og den uavhengige variabelen velges til kvadratmeter. Kjøres det en regresjon på dette fås følgende data ut, se tabell 6.1 nedenfor.
Den øverste delen av tabellen er en analyse av variansen til modellen. Den forteller hvor mye av variansen som kan forklares av modellen i seg selv, og hvor mye som forklares gjennom restleddet. For at modellen skal være god, må modellen ha mest mulig av forklaringskraften. Dette kan vi enkelt finne ved å se på R^2. Dette tallet på 0,2607, sier at 26,07 % av variasjonen kan forklares av modellen. Forklaringskraften, som R^2 også kalles for, er alltid mellom 0 og 1. Resten av variasjonen i salgsprisen må da forklares av feilleddet *e*(1-0,2607=0,7393).
Verdien til R^2 blir høyere jo flere uavhengige variabler du putter inn i modellen. Men dette tas det hensyn til, i tallet under, som er en justert R^2. Denne verdien synker dersom modellen blir tillagt uavhengige variabler som har liten effekt på den avhengige variabelen.
Som tidligere nevnt må en t-verdi være minst 1,96 og en p-verdi under 0,05 for at det skal være signifikans med 95 % sannsynlighet. Disse forteller det samme, så fra nå ser vi bare på p-verdien.
P-verdien i tabell 6.1 for kvadratmeter er lik null. Dette betyr at antall kvadratmeter virkelig påvirker salgsprisen.
I nederste delen av tabellen er det gitt en koeffisient til kvadratmeter. Koeffisienten er det samme som β_1 i modellen, altså hvor mye variabelen kvadratmeter påvirker prisen med, og z_1 er antall kvadratmeter.

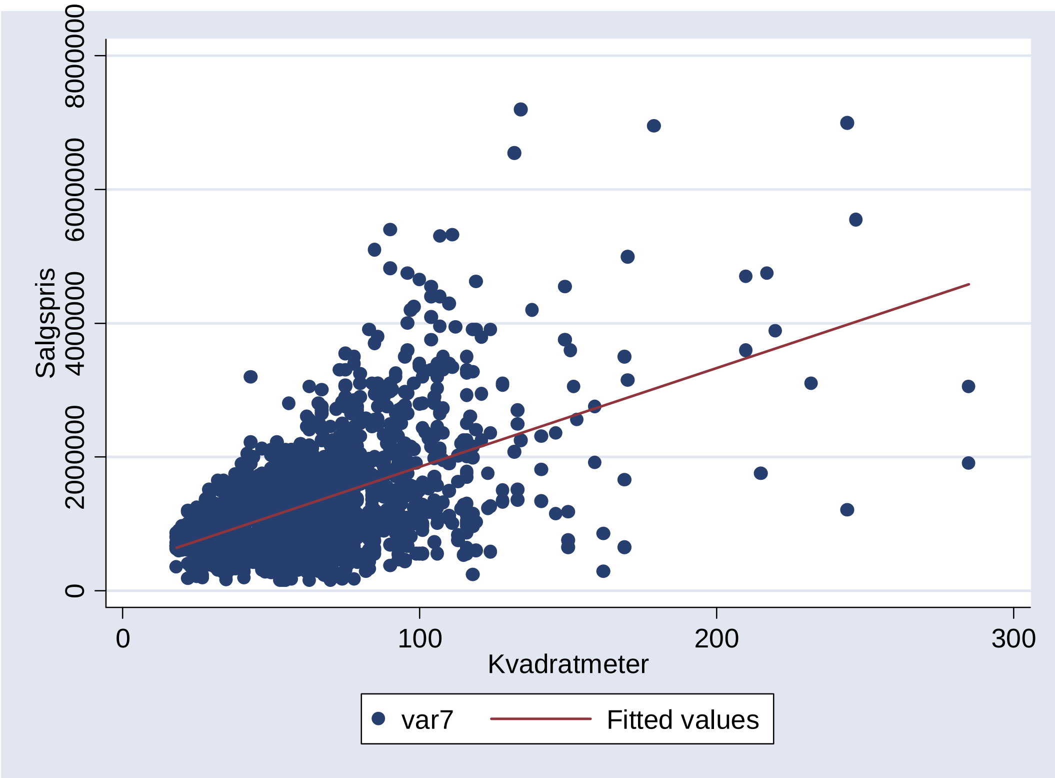
Denne er på 14 782,98. Koeffisienten sier hvor mye i kroner, antall kvadratmeter påvirker salgsprisen med ved en ekstra kvadratmeter.
Standardavviket på 585,13 forteller hvor mye +/- variasjonen fra gjennomsnittet eller koeffisienten blir, med 95 % sannsynlighet.
Konstanten sier hvor mye leiligheten vil bli solgt for selv uten den uavhengige variabelen. (som i dette tilfellet er tilnærmet lik 0 kroner).

Tabell 6.1 Lineær regresjon med en uavhengig variabel

Kilde	SS	df	MS		
Modell	3.4489e+14	1	3.4489e+14	Antall obs =	1812
Residual	9.7800e+14	1810	5.4033e+11	F(1, 1810) =	638.29
				Sanns. > F =	0.0000
				R^2 =	0.2607
				Justert R^2 =	0.2603
Totalt	1.3229e+15	1811	7.3047e+11	$\sqrt{MSE}$ =	7.4e+05

Salgspris	Koef.	Std.Avv.	t	P>|t|	[95% Konf.	Intervall]
Kvadratmeter	14782.98	585.1313	25.26	0.000	13635.38	15930.59
Konstant	369143.8	42112.1	8.77	0.000	286550.3	451737.2

Under i figur 6.1 har vi med den samme figuren som viser korrelasjonen mellom antall kvadratmeter og salgspris, som i figur 5.8. Den røde linja er regresjonslinja, og viser hvordan plottet ville vært, dersom kvadratmeter var den eneste variabelen som påvirket salgsprisen. Den ville da vært helt lineær, og stigningstallet på linja ville vært lik koeffisienten til kvadratmeter, i vårt tilfelle 14 782,98.



Figur 6.1 Regresjonslinje for salgspris og kvadratmeter

Lineær regresjon med en dummyvariabel

Vi skal nå ta med enda en ny type variabel i regresjonen vår. Variabelen er en dummy for salgsår. Akkurat denne som vi tar med her, lages spesielt for å vise på en enkel måte hvordan en slik variabel virker. Vi lager en dummyvariabel for salgsår, som går på om salget har foregått etter år 2000 (det vil si fra år2000-år 2010), eller før år 2000 (år1990-år1999).

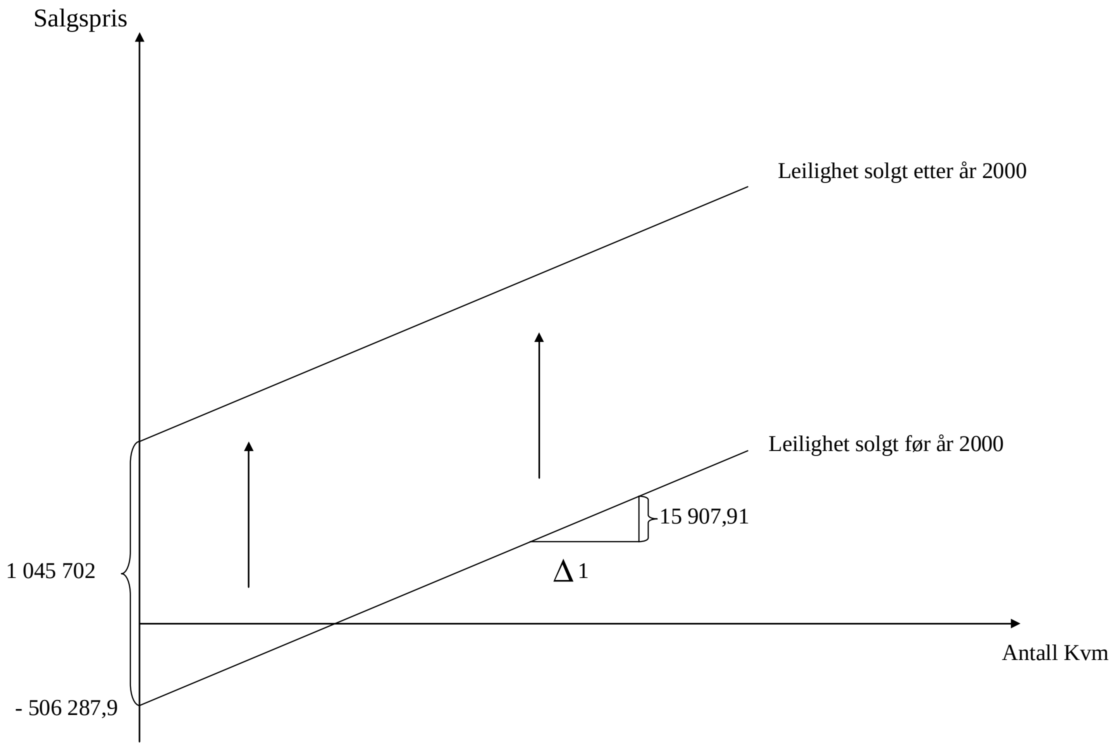
En dummyvariabel er som tidligere nevnt ikke kontinuerlig, og kan kun ha verdien 0 eller 1. Dummyvariabelen fungerer ikke på samme måte som en kontinuerlig variabel, den gir i stedet en bestemt verdi som plusses på eller trekkes fra salgsprisens konstantledd, alt etter som den er positiv eller negativ. Hver tidsdummy har sin egen verdi den bidrar med. Danton (2005).
Ved å ta med den nye dummyen i modellen, kan vi se av tabell 6.2 at R^2 er blitt styrket i forhold til bare å ha med kvadratmeter, men den er svekket i forhold til å ha variabelen salgspris som en kontinuerlig variabel som under forrige avsnitt.
P-verdien er likevel null også her, noe som tyder på at dummyen er signifikant, og at den påvirker salgsprisen.

Tabell 6.2 Regresjon med dummyvariabel

Source	SS	df	MS		
Model	6.9746e+14	2	3.4873e+14	Antall obs	= 1812
Residual	6.2543e+14	1809	3.4573e+11	F(2, 1809)	= 1008.67
				Sanns > F	= 0.0000
				R^2	= 0.5272
				Justert R^2	= 0.5267
Totalt	1.3229e+15	1811	7.3047e+11	$\sqrt{MSE}$	= 5.9e+05

Salgspris	Koef.	Std. Avv.	t	P>|t|	[95% Konf.	Interval]
Kvadratmeter	15907.91	469.3739	33.89	0.000	14987.34	16828.49
salgett~2000	1045702	32745.62	31.93	0.000	981479.1	1109926
Konstant	-506287.9	43430.85	-11.66	0.000	-591467.7	-421108

I figur 6.2 ser vi hvordan salgsprisen reagerer dersom dummyen inntreffer, det vi si, når det er etter år 2000. (I denne figuren antar vi også at det er kun de to variablene, kvadratmeter og tidsdummyen, som påvirker prisen.) Koeffisienten til dummyen legger til konstanten en verdi på 1 045 702, og gir dermed salgsprislinjen et skift oppover. Stigningstallet til linja er fortsatt koeffisienten til kvadratmeter, og her er denne på 15 907,91.



Figur 6.2 Lineær modell, virkningen av en dummy

6.2 Lineær form

Den generelle formen husker vi fra ligning (4.1):

P = $\beta_0 + \beta_1 z_1 + \beta_2 z_2 + ... + \beta_n z_n + \varepsilon$

Nå tas alle de uavhengige variablene med, i tillegg til dummyvariabler for måneder, år og postnummer. Den første i rekken av dummyvariablene droppes for å unngå multikollinearitet. I tabell 6.3 nedenfor ser vi at R^2 sier at modellen med de variablene vi har med nå forklarer 77,02 % av all variasjonen i prisen. Dette betyr at 22,98 % forklares av restleddet, noe som kanskje er litt mye siden alle variablene nå er med.

Ved å se på p-verdiene til alle variablene, finner vi først av kvadratmeter fremdeles er en signifikant variabel med verdi lik null. Derimot er ingen av månedsvariablene signifikante, unntatt august (0,035) og oktober (0,04). Salgsårene er ikke signifikante til og med år 1998, men etter dette er de det. Alle postnumrene er signifikante med p-verdi tilnærmet lik null for alle, og alder er også signifikant.

Tabell 6.3 Hele modellen i lineær regresjon

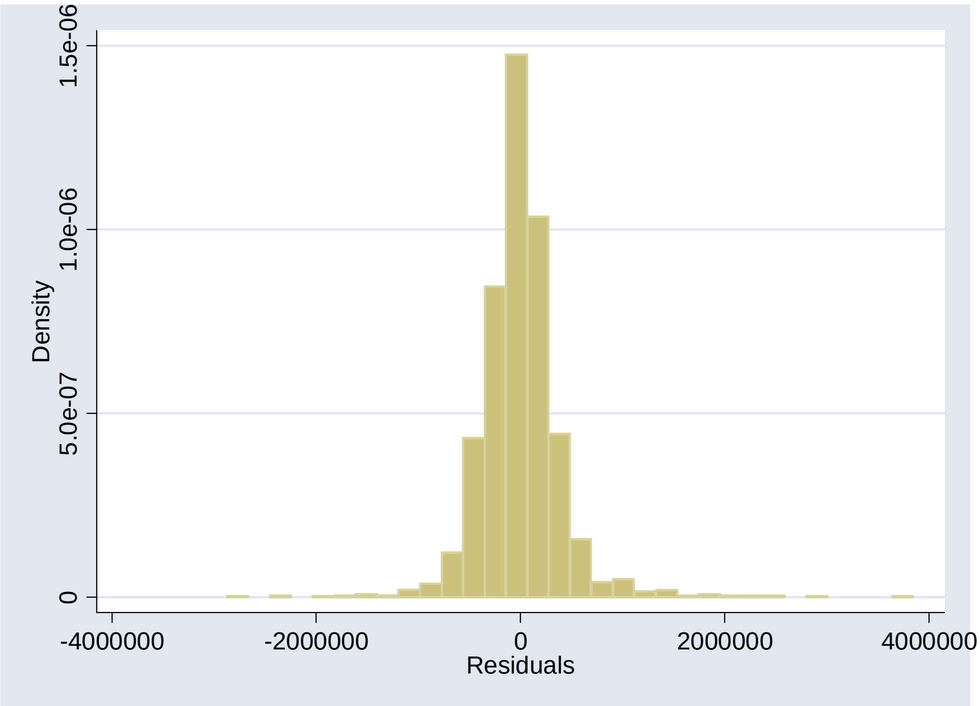
Kilde	SS	df	MS
Model	1.0189e+15	42	2.4261e+13
Restledd	3.0394e+14	1769	1.7181e+11
Totalt	1.3229e+15	1811	7.3047e+11

Antall obs	= 1812
F(42, 1769)	= 141.20
Sanns > F	= 0.0000
R^2	= 0.7702
Justert R^2	= 0.7648
$\sqrt{MSE}$	= 4.1e+05

Salgspris	Koef.	Std. Avv.	t	P>|t|	[95% Konf.	Interval]
Kvadratmeter	16463.16	349.8789	47.05	0.000	15776.94	17149.38
MND2	27571.18	53576.37	0.51	0.607	-77508.47	132650.8
MND3	26657.07	51719.21	0.52	0.606	-74780.12	128094.3
MND4	65879.68	51859.22	1.27	0.204	-35832.11	167591.5
MND5	75572.91	53216.32	1.42	0.156	-28800.57	179946.4
MND6	50389.19	49796.48	1.01	0.312	-47276.94	148055.3
MND7	9998.795	52680.87	0.19	0.849	-93324.5	113322.1
MND8	101782.6	48154.84	2.11	0.035	7336.271	196229
MND9	74284.15	51249.66	1.45	0.147	-26232.12	174800.4
MND10	145485.7	50207.49	2.90	0.004	47013.49	243958
MND11	90322.79	52652.9	1.72	0.086	-12945.66	193591.2
MND12	-9458.606	51929.96	-0.18	0.855	-111309.1	92391.92
AR1991	111788.4	127058.9	0.88	0.379	-137412.9	360989.8
AR1992	-9124.139	138624	-0.07	0.948	-281008.1	262759.9
AR1993	-140363.9	131139	-1.07	0.285	-397567.6	116839.7
AR1994	-69787.77	124207.1	-0.56	0.574	-313396	173820.4

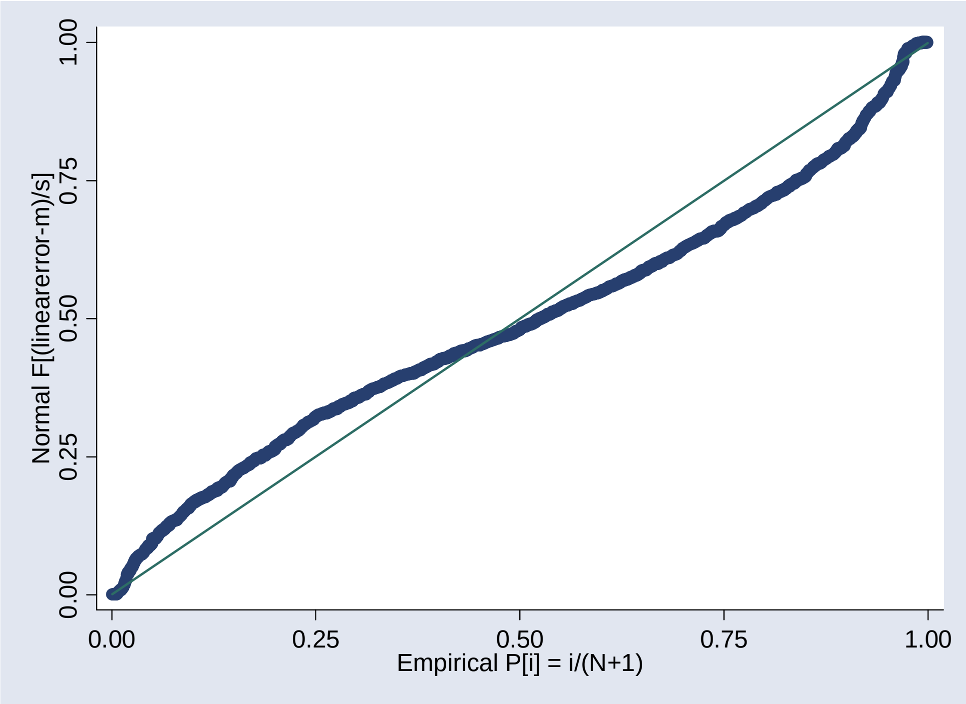
AR1995	43938.47	122814.3	0.36	0.721	-196938	284814.9
AR1996	63875.19	116049.1	0.55	0.582	-163732.6	291483
AR1997	150927.4	116332.6	1.30	0.195	-77236.43	379091.1
AR1998	221103.3	120155.2	1.84	0.066	-14557.89	456764.5
AR1999	344644.4	116200.9	2.97	0.003	116738.9	572550
AR2000	509179.9	117175	4.35	0.000	279363.9	738995.9
AR2001	593294.7	115013.3	5.16	0.000	367718.4	818871
AR2002	686136	113928.1	6.02	0.000	462688.1	909583.9
AR2003	685621.2	112719.3	6.08	0.000	464544.2	906698.2
AR2004	922360.9	109271.4	8.44	0.000	708046.2	1136676
AR2005	1048988	108380	9.68	0.000	836421.7	1261554
AR2006	1283866	108206.8	11.86	0.000	1071640	1496093
AR2007	1550899	108281.1	14.32	0.000	1338526	1763271
AR2008	1545362	109622.7	14.10	0.000	1330358	1760365
AR2009	1694257	109139.5	15.52	0.000	1480201	1908313
AR2010	1543822	158143.6	9.76	0.000	1233654	1853990
PNR4611	-205089.4	60363.89	-3.40	0.001	-323481.4	-86697.33
PNR4612	-278439.2	44593.23	-6.24	0.000	-365900.2	-190978.3
PNR4614	-382015.4	27787.06	-13.75	0.000	-436514.3	-327516.5
PNR4622	-780795.2	98592.97	-7.92	0.000	-974166.2	-587424.3
PNR4623	-749876.8	45036.23	-16.65	0.000	-838206.6	-661546.9
PNR4624	-845317.2	42731.34	-19.78	0.000	-929126.5	-761508
PNR4635	-702011	42902.4	-16.36	0.000	-786155.8	-617866.3
PNR4639	-737450.4	53787.28	-13.71	0.000	-842943.7	-631957.1
PNR4656	-876570.7	417387.2	-2.10	0.036	-1695195	-57946.63
alder	-2562.764	261.6252	-9.80	0.000	-3075.892	-2049.637
Konstant	-224108.1	117124	-1.91	0.056	-453824	5607.858

I figur 6.3 ser vi fordelingen av restleddet, *e*, og ser at det har et gjennomsnitt rundt 0.



Figur 6.3 Restleddsfordeling med lineær modell

I figur 6.4 er et normalskråplott, som også forteller hvordan restleddet er i forhold til den normalfordelingslinja. Hadde restleddet vært helt riktig fordelt, ville den blå linjen ligget akkurat oppå normalfordelingslinja, det vil si at den ville vært symmetrisk. Ser at den avviker noe, og at den er noe spissere fordelt enn normalfordelingen. At den er spissere forklares av at formen på den ser ut som en speilvendt s, og det betyr at fordelingen er mer konsentrert enn normalfordelingen. Dersom den hadde vært formet som en s riktig vei, hadde det vært motsatt(mindre konsentrert enn normalfordelingen).



Figur 6.4 Normalskråplott lineær form

Forsøk på å forbedre den lineære modellen

På grunn av at månedsdummyene ikke hadde mer enn 2 måneder som var signifikante, vil vi nå kjøre regresjonen ved å lage månedsdummyene om til kvartalsdummyer. Da får vi totalt 4 dummyer.

Av tabell 6.4 kan vi se at både R^2, og den justerte R^2 er gått ned. Dette tyder på at å gjøre om månedsdummyene til kvartalsdummyer, ikke har forbedret modellen. Vi kan se at kvadratmeter og alder fortsatt er signifikant, i tillegg til alle salgsårene fra og med 1999. Kun fjerde kvartal av kvartalsdummyene er signifikante. Dette kan jo ha noe med at nettopp oktober måned hadde den beste p-verdien under den lineære modellen (0,004). Denne måneden ligger i fjerde kvartal. August var også signifikant, men verdien dens var ikke god nok til å gi signifikans for hele tredje kvartal. Her er postnummer 4656 ikke signifikant, men ellers er resten signifikante.

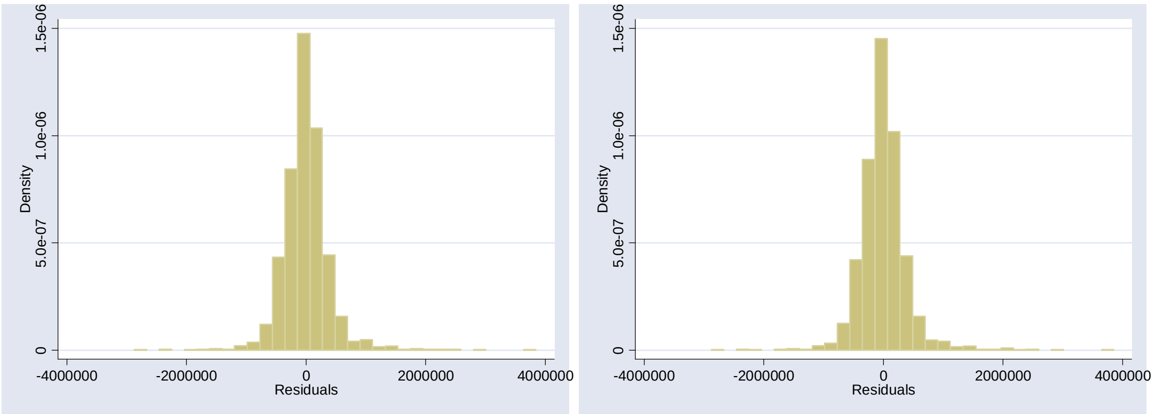
Tabell 6.4 Forsøk på å forbedre den lineære modellen

Kilde	SS	df	MS
Model	1.0163e+15	34	2.9891e+13
Restledd	3.0658e+14	1777	1.7253e+11
Totalt	1.3229e+15	1811	7.3047e+11

Antall obs	=	1812
F(34, 1777)	=	173.26
Sanns > F	=	0.0000
R^2	=	0.7682
Justert R^2	=	0.7638
$\sqrt{MSE}$	=	4.2e+05

Salgspris	Koef.	Std. Avv.	t	P>|t|	[95% Konf.	Interval]
Kvadratmeter	16516.94	349.8089	47.22	0.000	15830.86	17203.02
andrekvartal	43953.35	29219.19	1.50	0.133	-13354.24	101260.9
tredjekvar~l	48920.51	28645.64	1.71	0.088	-7262.172	105103.2
fjerdekvar~l	60023.61	29188.49	2.06	0.040	2776.228	117271
AR1991	121871.8	126849.3	0.96	0.337	-126917.6	370661.3
AR1992	-7930.692	138510.2	-0.06	0.954	-279590.7	263729.3
AR1993	-133523.6	131143.2	-1.02	0.309	-390734.6	123687.5
AR1994	-70499.93	124254.7	-0.57	0.571	-314200.7	173200.9
AR1995	58647.82	122742.7	0.48	0.633	-182087.3	299383
AR1996	66058.94	116144.4	0.57	0.570	-161735.1	293853
AR1997	151920.9	116323.1	1.31	0.192	-76223.6	380065.4
AR1998	222712.3	120203.5	1.85	0.064	-13042.77	458467.4
AR1999	349260.8	116183.6	3.01	0.003	121389.9	577131.7
AR2000	521629.2	117141	4.45	0.000	291880.6	751377.9
AR2001	588872.9	114926.8	5.12	0.000	363467.1	814278.8
AR2002	695701.3	113855.9	6.11	0.000	472395.8	919006.8
AR2003	702731.1	112494.8	6.25	0.000	482095.1	923367
AR2004	933278.4	109072.9	8.56	0.000	719353.8	1147203
AR2005	1036679	108230.5	9.58	0.000	824406.7	1248952
AR2006	1293358	108089.9	11.97	0.000	1081361	1505355
AR2007	1555820	108258	14.37	0.000	1343493	1768146
AR2008	1549491	109649.4	14.13	0.000	1334435	1764546
AR2009	1697427	109126.8	15.55	0.000	1483397	1911457
AR2010	1536799	156616.2	9.81	0.000	1229628	1843971
PNR4611	-203362.3	60379.35	-3.37	0.001	-321784.3	-84940.28
PNR4612	-279023.2	44591.68	-6.26	0.000	-366480.8	-191565.5
PNR4614	-385901.9	27796.9	-13.88	0.000	-440420	-331383.8
PNR4622	-788681.2	98739.96	-7.99	0.000	-982339.9	-595022.5
PNR4623	-749099.5	44913.65	-16.68	0.000	-837188.6	-661010.3
PNR4624	-839126	42738.25	-19.63	0.000	-922948.6	-755303.5
PNR4635	-701247.2	42813.45	-16.38	0.000	-785217.2	-617277.2
PNR4639	-733438.4	53486.81	-13.71	0.000	-838342	-628534.7
PNR4656	-815754.5	417501.4	-1.95	0.051	-1634600	3090.931
alder	-2564.73	260.906	-9.83	0.000	-3076.445	-2053.015
Konstant	-212127	111144.4	-1.91	0.056	-430114.4	5860.378

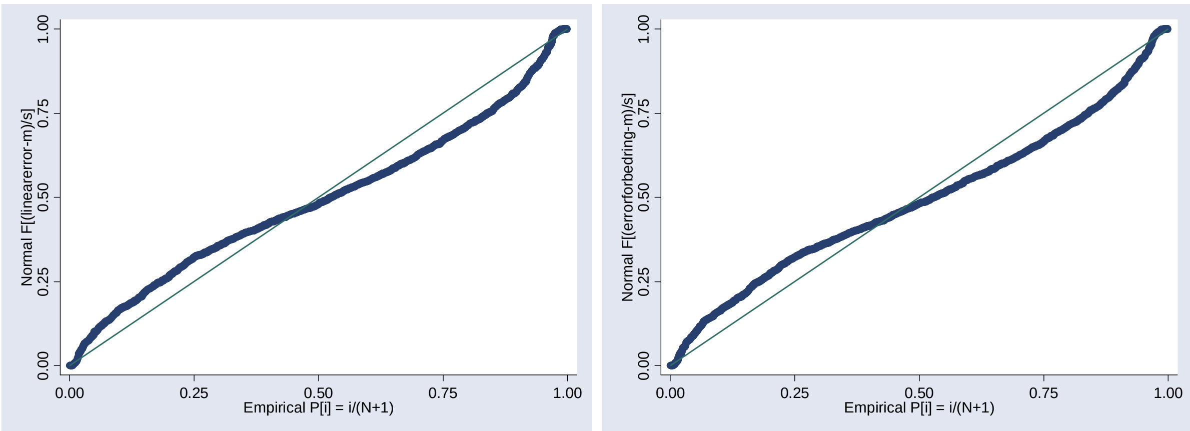
Av figur 6.5 kan vi se fordelingen av restleddet. Dette er så godt som identisk som før vi innførte kvartalsdummyer.



1. Før *2. Etter*

Figur 6.5 Restleddsfordelingen, før og etter kvartalsdummyer ble innført

Vi har også tatt med normalskråplottet i figur 6.6. Som i figur 6.5 er dette tilnærmet likt både før og etter innføring av kvartalsdummy.



.

1. Før *2. Etter*

Figur 6.6 Normalskråplott, før og etter kvartalsdummyer

6.3 Dobbeltlogaritmisk form

.

Den dobbeltlogaritmiske formen husker vi fra ligning (4.2):

P = $\beta_0 z_1^{\beta_1} z_2^{\beta_2} z_3^{\beta_3} e^{\beta_4 z_4 + \beta_5 z_5 + \varepsilon}$

Tar vi den naturlige logaritmen til denne får vi ligning (4.3):

Ln P = $\ln \beta_0 + \beta_1 \ln z_1 + \beta_2 \ln z_2 + \beta_3 \ln z_3 + \beta_4 z_4 + \beta_5 z_5 + \varepsilon$

For å kunne gjøre en regresjon med en slik form, må vi lage noen nye variabler til de som er kontinuerlige. Lnsalgspris, lnkvm og lnalder. Dummyvariablene trenger vi ikke å gjøre noe med.

For lnalder gjør vi et lite triks med STATA, slik at de leilighetene som er 0 år, også blir med i analysen. Ln0 gir ikke noe svar, og dermed droppes de salgscasene av STATA. Dette ønsker vi ikke, så vi går inn i datamaterialet, og forandrer alle salgscasene som har alder 0, til å ha en alder på 0,1. Dette tallet går det an å regne logaritmen til, så da unngår vi å miste disse casene det gjelder for.

Koeffisientene til de kontinuerlige variablene blir da prosenttall, som forteller hvor stor endring i prosent salgsprisen endrer seg ved en % økning i den enkelte variabelen.

Dummyvariablenes prosentverdi forteller hvor stor endring i prosent de vil gi salgsprisen dersom de inntreffer.

Av tabell 6.5 kan vi se at denne modellen har en forklaringskraft på 88,75 %, noe som er høyere enn for den lineære funksjonen. Men man kan ikke trekke noe konklusjon på at den ene formen er bedre enn den andre ut ifra forklaringskraften, siden det er to ulike funksjonsformer.

Siden kvartalsdummyene ikke gjorde at den lineære regresjonen noe bedre, har vi her gått tilbake til vanlige månedsdummyer.

Både lnkvm og lnalder ser ut til å være signifikante, med p-verdi lik null.

Når det gjelder månedsdummyene er august, september, oktober og desember måned signifikante, noe som er mer enn for den lineære, som bare hadde to signifikante måneder.

Alle årsdummyene utenom 1991, 1994 og 1995 er signifikante. Dette er igjen mer enn for den lineære modellen. Tilslutt er også alle postnumrene signifikante.

Tabell 6.5 Regresjon med Dobbeltlogaritmisk form

Kilde	SS	df	MS		
Model	622.56892	42	14.8230695	Antall obs =	1812
Restledd	78.928625	1769	.044617651	F(42, 1769) =	332.22
				Sanns. > F =	0.0000
				R^2 =	0.8875
				Justert R^2 =	0.8848
Totalt	701.497545	1811	.387353697	$\sqrt{MSE}$ =	.21123

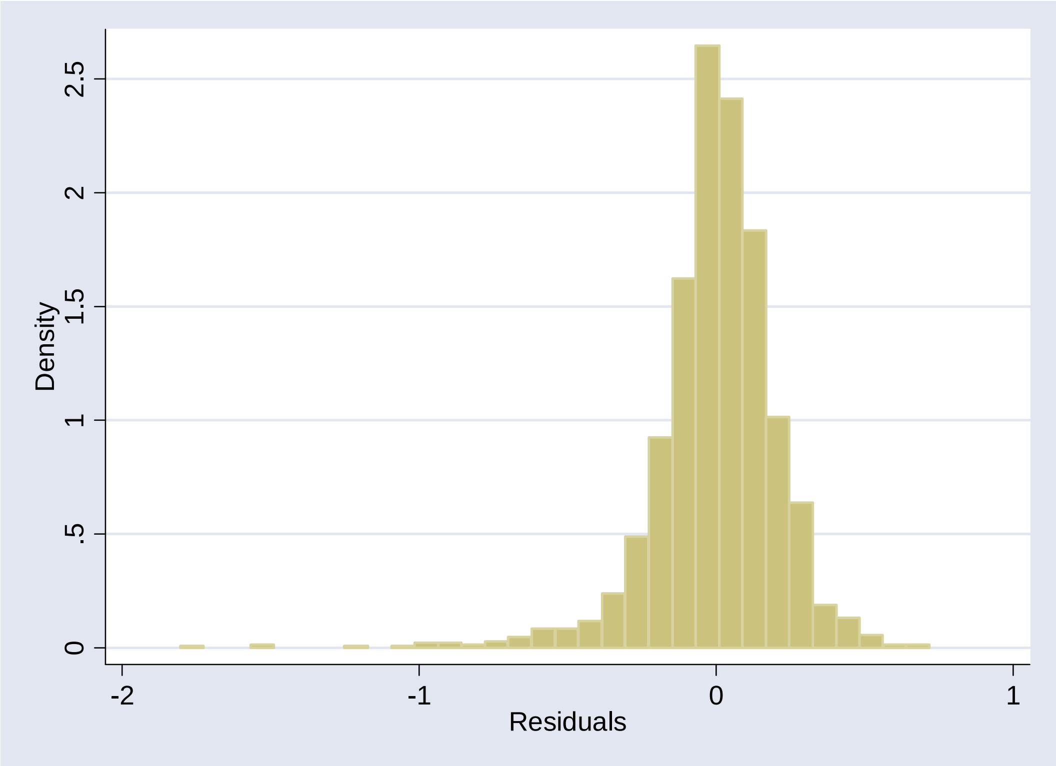
lnpris	Koef.	Std. Avv.	t	P>|t|	[95% Konf.	Interval]
lnkvm	.7265141	.0119973	60.56	0.000	.7029837	.7500445
lnalder	-.0399632	.0030311	-13.18	0.000	-.0459081	-.0340183
MND2	.0094138	.02729	0.34	0.730	-.0441103	.0629379
MND3	.0027477	.0263575	0.10	0.917	-.0489474	.0544429
MND4	.0211216	.0264218	0.80	0.424	-.0306996	.0729428
MND5	.0367272	.0271023	1.36	0.176	-.0164286	.089883
MND6	.0275349	.0253793	1.08	0.278	-.0222418	.0773115
MND7	.0259678	.0268391	0.97	0.333	-.0266719	.0786075
MND8	.0575696	.0245492	2.35	0.019	.0094211	.1057181
MND9	.05474	.0261008	2.10	0.036	.0035483	.1059316
MND10	.074143	.025625	2.89	0.004	.0238846	.1244013
MND11	.048387	.0268366	1.80	0.072	-.0042477	.1010217
MND12	.0637377	.0265312	2.40	0.016	.0117019	.1157735
AR1991	.0458583	.0649399	0.71	0.480	-.0815087	.1732254
AR1992	-.1804272	.0706312	-2.55	0.011	-.3189567	-.0418977
AR1993	-.1922571	.0668439	-2.88	0.004	-.3233585	-.0611557
AR1994	.013756	.0633391	0.22	0.828	-.1104713	.1379833
AR1995	.1095964	.0625952	1.75	0.080	-.0131719	.2323648
AR1996	.1652838	.0591582	2.79	0.005	.0492564	.2813112
AR1997	.3427918	.0593018	5.78	0.000	.2264828	.4591009
AR1998	.4724874	.0613045	7.71	0.000	.3522505	.5927243
AR1999	.5574476	.0592584	9.41	0.000	.4412237	.6736715
AR2000	.739007	.0597642	12.37	0.000	.6217912	.8562228
AR2001	.7862421	.0586576	13.40	0.000	.6711966	.9012877
AR2002	.9290361	.0580938	15.99	0.000	.8150964	1.042976
AR2003	.9512117	.057514	16.54	0.000	.8384092	1.064014
AR2004	1.073737	.0556752	19.29	0.000	.9645412	1.182933
AR2005	1.18441	.0552631	21.43	0.000	1.076023	1.292798
AR2006	1.346657	.0551572	24.41	0.000	1.238477	1.454837
AR2007	1.516323	.0552025	27.47	0.000	1.408054	1.624591
AR2008	1.524362	.0558944	27.27	0.000	1.414736	1.633988
AR2009	1.545814	.0556398	27.78	0.000	1.436687	1.654941
AR2010	1.555774	.0806543	19.29	0.000	1.397586	1.713961
PNR4611	-.0721849	.0307713	-2.35	0.019	-.1325369	-.011833
PNR4612	-.1040658	.0226288	-4.60	0.000	-.1484477	-.0596838
PNR4614	-.1770066	.0142149	-12.45	0.000	-.2048863	-.1491268
PNR4622	-.6187447	.050203	-12.32	0.000	-.7172081	-.5202812
PNR4623	-.499035	.0227704	-21.92	0.000	-.5436947	-.4543752

PNR4624	-.5558793	.0213617	-26.02	0.000	-.5977762	-.5139824
PNR4635	-.4880544	.0216607	-22.53	0.000	-.5305377	-.4455711
PNR4639	-.4517572	.0279689	-16.15	0.000	-.5066128	-.3969016
PNR4656	-.6405464	.2131621	-3.00	0.003	-1.058622	-.2224703
Konstant	10.24321	.0782587	130.89	0.000	10.08973	10.3967

.

Figuren under, 6.7, gir restleddsfordelingen ved dobbeltlogaritmisk form. Fordelingen er ikke like symmetrisk som ved den lineære modellen, og den har en hale på høyre side. Det vil si at det er litt flere tilfeller med lave verdier(Aarnes,2000).

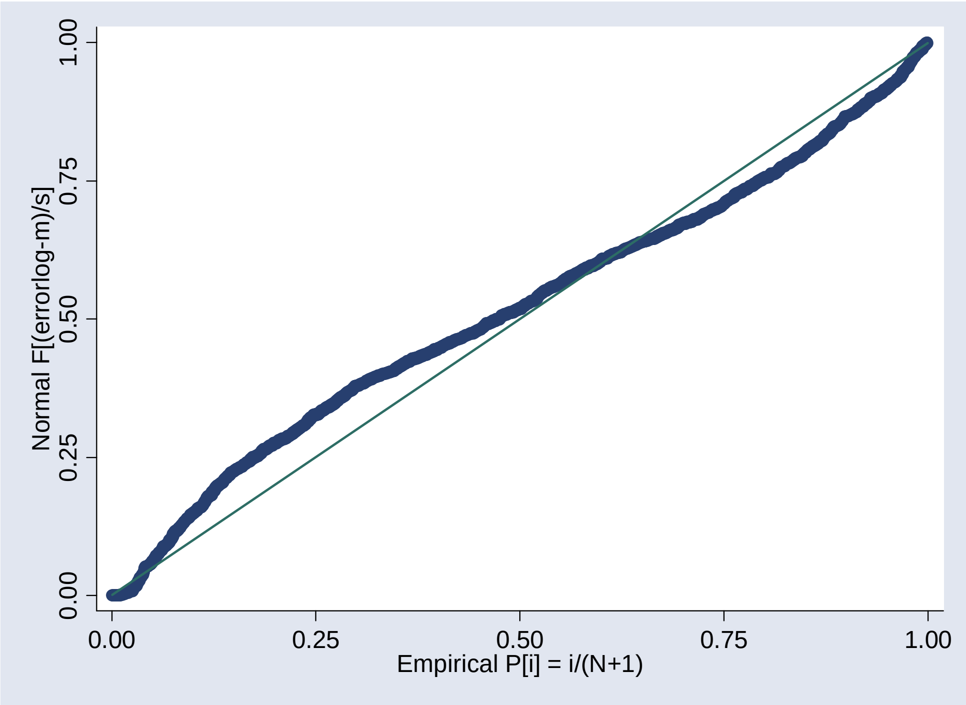
Forventningsverdien ser likevel ut til å være tilnærmet lik 0.



.

Figur 6.7 Restleddsfordeling, dobbeltlogaritmisk

Skråplottet i figur 6.8 viser at det er et nokså likt resultat som med den lineære formen. Også her er det en spissere fordeling enn normalfordelingen. Men vi kan se at den omvendte s'en krysser linjen på et litt senere tidspunkt, enn for den lineære modellen.



Figur 6.8 Normalskråplott restledd, med dobbellogaritmisk form

6.4 Valg av form

Etter å ha sett på regresjonsresultatene til begge formene, får vi ikke et veldig klart og tydelig bilde på hvilken modellform som er best til å bruke videre. Begge formene må anses som gode nok, og med grunnlag i restleddsfordelingen kan det se ut som at den lineære formen er marginalt bedre enn den dobbeltlogaritmiske. Men velger likevel å bruke den dobbeltlogaritmiske formen, siden det var såpass liten forskjell, og siden det er et enklere/ mer praktisk utgangspunkt for videre beregning.

7. Prisutvikling

I dette kapittelet skal vi ta for oss prisutvikling over tid. Dette er jo en sentral del av oppgaven vår. Vi er interessert i å vite forskjellen mellom sentralt i en by, og mer perifere/ usentrale deler av byen.
Koeffisientene til salgsårene forteller om denne utviklingen. For å få egne koeffisienter for alle salgsårene, både for sentralt og usentralt, kjøres regresjonen i to separate omganger. Når vi kjører regresjonen for sentralt, droppes alle casene med postnummer som er fra usentrale områder, og likeså droppes casene med postnumrene for de sentrale områdene når vi kjører regresjon for usentralt.

For sentralt gjelder postnummer 4610, 4611, 4612 og 4614. Også her må første dummy i rekken (her er det 4610) droppes. I tabell 7.1 kan vi se resultatet for sentralt. Forklaringskraften til modellen, for case sentralt er 88,09 %. Dersom vi ser på alle variablene, er de fleste signifikante. Men for dummyene er månedene mai og oktober de eneste signifikante. For salgsårene er de signifikante fra og med år 1996, årene før dette er de ikke. Her vil vi også ta med eksempel på hvordan vi kan regne ut prisen på en leilighet, ved hjelp av modellen. Vi velger å finne prisen på en leilighet som er 100 kvadratmeter stor, er 5 år gammel, solgt på postnummer 4611, i oktober måned, år 2008:

$$\text{Lnsalgspris} = 10{,}21751 + (0{,}7323177 \times \ln 100) - (0{,}0460051 \times \ln 5) - (0{,}0739313 \times 1) + (0{,}0756914 \times 1) + (1{,}508624 \times 1) = 15{,}02629939$$

$$\text{Salgspris} = e^{15{,}02629939}$$

$$\text{Salgspris} = 3\,356\,131{,}02 \text{ Kroner}$$

Tabell 7.1 Regresjon for Sentralt

Kilde	SS	df	MS
Model	448.619091	36	12.4616414
Restledd	60.6821781	1254	.048390892
Totalt	509.30127	1290	.394807186

Antall obs	1291
F(36, 1254)	257.52
Sanns. > F	0.0000
R^2	0.8809
Justert R^2	0.8774
$\sqrt{MSE}$	.21998

lnpris	Koef.	Std. Avv.	t	P>|t|	[95% Konf.	Interval]
lnkvm	.7323177	.0132595	55.23	0.000	.7063043	.758331
lnalder	-.0460051	.0037534	-12.26	0.000	-.0533686	-.0386415
PNR4611	-.0739313	.0322285	-2.29	0.022	-.137159	-.0107035
PNR4612	-.1102556	.0238239	-4.63	0.000	-.1569946	-.0635165
PNR4614	-.1743023	.014906	-11.69	0.000	-.2035457	-.1450589
MND2	.0110557	.0341769	0.32	0.746	-.0559946	.078106

MND3	.0079257	.0331769	0.24	0.811	-.0571627	.073014
MND4	.0262326	.033265	0.79	0.430	-.0390287	.0914939
MND5	.0693865	.0326889	2.12	0.034	.0052556	.1335175
MND6	.0493736	.0317431	1.56	0.120	-.0129017	.111649
MND7	.0322669	.0326408	0.99	0.323	-.0317698	.0963035
MND8	.0544405	.0307139	1.77	0.077	-.0058159	.1146968
MND9	.0494247	.0325673	1.52	0.129	-.0144677	.1133171
MND10	.0756914	.0320542	2.36	0.018	.0128056	.1385772
MND11	.0718337	.0332	2.16	0.031	.0067001	.1369674
MND12	.0478084	.0321597	1.49	0.137	-.0152844	.1109013
AR1991	.0236012	.0821991	0.29	0.774	-.1376618	.1848641
AR1992	-.0814167	.0826727	-0.98	0.325	-.2436088	.0807753
AR1993	-.1927226	.0733116	-2.63	0.009	-.3365494	-.0488958
AR1994	.0426657	.0722729	0.59	0.555	-.0991234	.1844547
AR1995	.066715	.0707395	0.94	0.346	-.0720657	.2054957
AR1996	.193588	.0654963	2.96	0.003	.0650937	.3220824
AR1997	.377946	.0654138	5.78	0.000	.2496134	.5062785
AR1998	.5124534	.069301	7.39	0.000	.3764948	.648412
AR1999	.5648471	.0663893	8.51	0.000	.4346008	.6950935
AR2000	.7651597	.0665181	11.50	0.000	.6346608	.8956587
AR2001	.8047606	.0650146	12.38	0.000	.6772113	.9323099
AR2002	.955517	.0636539	15.01	0.000	.8306371	1.080397
AR2003	.9647988	.0632378	15.26	0.000	.8407352	1.088862
AR2004	1.101675	.0613258	17.96	0.000	.9813624	1.221987
AR2005	1.222471	.0601383	20.33	0.000	1.104488	1.340453
AR2006	1.35294	.059935	22.57	0.000	1.235356	1.470524
AR2007	1.525654	.0601173	25.38	0.000	1.407712	1.643595
AR2008	1.508624	.0610235	24.72	0.000	1.388905	1.628344
AR2009	1.544826	.0605269	25.52	0.000	1.426081	1.663571
AR2010	1.548422	.0949432	16.31	0.000	1.362157	1.734688
Konstant	10.21751	.0863515	118.32	0.000	10.0481	10.38692

For usentrale/perifere deler av byen tar jeg med casene med postnumrene som hører til dem. Nemlig 4622, 4623, 4624, 4635 og 4639. I tabell 7.2 kan vi se resultatet for usentralt. Forklaringskraften er litt høyere her enn for sentralt, 91,27 %. Det må jo bety at modellen vår passer litt bedre for de salgscasene som er usentralt, enn de som er sentralt.

Det er det signifikans for alle postnumrene bortsett fra postnummer 4624 og 4656. For månedsdummyene er det kun signifikans for september, oktober og desember. For årsdummyene er det signifikans fra år 1999 og til 2010.

Vi tar de samme kriteriene som for sentralt bortsett fra at leiligheten blir solgt på postnummer 4639, og regner ut hva prisen blir for en usentral leilighet i følge modellen:

$$\text{Lnsalgspris} = 10,26897 + (0,6136259 \times \ln 100) - (0,033566 \times \ln 5) + (0,2163767 \times 1) + (0,0785848 \times 1) + (1,405819 \times 1) = 14,74157981$$

$$\text{Salgspris} = e^{14,74157981}$$

$$\text{Salgspris} = 2\,524\,566{,}21 \text{ Kroner}$$

Ser av de to priseksemplene at prisen på en tilsvarende leilighet som er lokalisert usentralt i Kristiansand, kan selges for 831 564, 81 kroner dyrere dersom den er lokalisert sentralt i byen.

Tabell 7.2 Regresjon for Usentralt

Kilde	SS	df	MS
Model	159.493627	38	4.1972007
Restledd	15.2586088	482	.031656865
Totalt	174.752235	520	.336061991

Antall obs	=	521
F(38, 482)	=	132.58
Sanns. > F	=	0.0000
R^2	=	0.9127
Justert R^2	=	0.9058
$\sqrt{MSE}$	=	.17792

lnpris	Koef.	Std. Avv.	t	P>|t|	[95% Konf.	Interval]
lnkvm	.6136259	.0353634	17.35	0.000	.5441405	.6831113
lnalder	-.033566	.0054114	-6.20	0.000	-.0441989	-.0229331
PNR4623	.1281566	.0453705	2.82	0.005	.0390081	.2173051
PNR4624	.0584609	.0448395	1.30	0.193	-.0296442	.146566
PNR4635	.1291511	.0448674	2.88	0.004	.0409912	.2173109
PNR4639	.2163767	.0483283	4.48	0.000	.1214166	.3113368
PNR4656	-.0319596	.1874716	-0.17	0.865	-.4003221	.3364028
MND2	.0239764	.0423132	0.57	0.571	-.0591647	.1071175
MND3	.0048652	.0406573	0.12	0.905	-.0750222	.0847526
MND4	.0332329	.0403541	0.82	0.411	-.0460588	.1125245
MND5	-.0774318	.0465726	-1.66	0.097	-.1689422	.0140787
MND6	-.0047971	.0398231	-0.12	0.904	-.0830454	.0734513
MND7	.0314688	.0457691	0.69	0.492	-.0584629	.1214005
MND8	.060819	.0384268	1.58	0.114	-.0146856	.1363237
MND9	.0838789	.0407433	2.06	0.040	.0038224	.1639354
MND10	.0785848	.0400362	1.96	0.050	-.0000823	.1572519
MND11	.0202245	.0437056	0.46	0.644	-.0656526	.1061016
MND12	.1355119	.0456083	2.97	0.003	.0458962	.2251275
AR1991	-.0819481	.1838269	-0.45	0.656	-.4431491	.2792529
AR1992	-.5465194	.1934599	-2.82	0.005	-.9266483	-.1663906
AR1993	-.2811797	.2008247	-1.40	0.162	-.6757797	.1134202
AR1994	-.1785415	.1881713	-0.95	0.343	-.548279	.191196
AR1995	.0278199	.187593	0.15	0.882	-.3407812	.3964209
AR1996	-.0452099	.1860935	-0.24	0.808	-.4108645	.3204447
AR1997	.1144151	.1864552	0.61	0.540	-.2519503	.4807804
AR1998	.2486091	.1871069	1.33	0.185	-.1190368	.616255
AR1999	.3761175	.1855914	2.03	0.043	.0114494	.7407857
AR2000	.526861	.1870588	2.82	0.005	.1593096	.8944123
AR2001	.5776159	.1854777	3.11	0.002	.2131711	.9420607
AR2002	.7021292	.1860957	3.77	0.000	.3364701	1.067788
AR2003	.7651462	.1850244	4.14	0.000	.4015921	1.1287
AR2004	.8781712	.1823064	4.82	0.000	.5199578	1.236385
AR2005	.9339685	.1822152	5.13	0.000	.5759343	1.292003
AR2006	1.182906	.1829356	6.47	0.000	.8234563	1.542356
AR2007	1.354334	.1831046	7.40	0.000	.994552	1.714116
AR2008	1.405819	.1836473	7.65	0.000	1.04497	1.766667
AR2009	1.407281	.1839061	7.65	0.000	1.045924	1.768638
AR2010	1.449203	.2043881	7.09	0.000	1.047602	1.850805
Konstant	10.26897	.241544	42.51	0.000	9.794363	10.74358

I tabell 7.3 har vi trukket ut koeffisientene for salgsårene fra sentralt og usentralt, og gjort dem om til prisindekser.

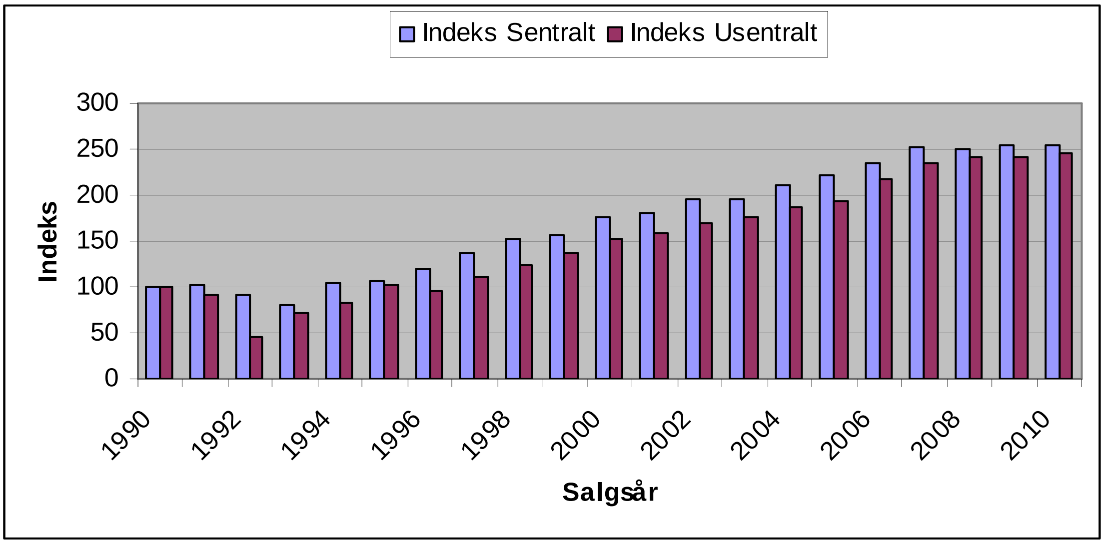
Ved å ikke ta med salgsår 1990 i regresjonen, unngår vi multikollinearitet, og 1990 blir basisåret med verdi lik 100. Alle salgsårskoeffisientene viser da prosentvis endring i forhold til dette året.

For å ta et eksempel kan vi se at koeffisienten til salgsår 1997 er 0,378 for sentralt. Det betyr at i år 1997 var det en økning i salgsprisen på 37,8 % fra år 1990. På denne måten kan vi se på hvert salgsår hvilken utvikling det har vært i forhold til basisåret 1990.

Tabell 7.3 Omgjøring fra salgsårskoeffisienter til prisindekser, med basisår 1990

Salgsår:	Sentral Koeffisient	Indeks Sentralt:	Usentral Koeffisient	Indeks Usentralt:
1990	-	100,0	-	100,0
1991	.0236012	102,4	-.0819481	91,8
1992	-.0814167	91,9	-.5465194	45,3
1993	-.1927226	80,7	-.2811797	71,9
1994	.0426657	104,3	-.1785415	82,1
1995	.066715	106,7	.0278199	102,8
1996	.193588	119,4	-.0452099	95,5
1997	.377946	137,8	.1144151	111,4
1998	.5124534	151,2	.2486091	124,9
1999	.5648471	156,5	.3761175	137,6
2000	.7651597	176,5	.526861	152,7
2001	.8047606	180,5	.5776159	157,8
2002	.955517	195,6	.7021292	170,2
2003	.9647988	196,5	.7651462	176,5
2004	1.101675	210,2	.8781712	187,8
2005	1.222471	222,5	.9339685	193,4
2006	1.35294	235,3	1.182906	218,3
2007	1.525654	252,6	1.354334	234,4
2008	1.508624	250,9	1.405819	240,6
2009	1.544826	254,5	1.407281	240,7
2010	1.548422	254,8	1.449203	244,9

I figur 7.1 er indeksene for sentralt og usentralt målt opp mot hverandre, for å se hvor lik utviklingen i prisen har vært dem imellom. Vi kan se at prisnivået sentralt stort sett ligger et lite hakk over usentralt.



Figur 7.1 Sammenligner Indekser sentralt mot usentralt

.
I tabell 7.4 nedenfor er indeksene for sentralt og usentralt i Kristiansand, satt sammen med indeksene fra kapittel to, tabell 2.4. Indeksene fra kapittel to hadde basisår i år 2000. For å gjøre indeksene sammenlignbare, må de ha samme basisår. Dette gjør vi om nå, ved å velge 1990 til basisår for disse også. I og med at det ikke var noen observasjon på indeksen fra år 1990, tar vi en fortsetning om at prisnivået i 1990 er lik 1991. Da vi i forutsetningen antar at prisnivået var lik i 1991 som i 1990, blir da indeksen lik 100 begge disse årene.
Indeksen for hele landet med basisår i år2000, var på 43,9 i 1991, i 1992 var den på 40,7. Ny indeks for 1992 omregnet til å ha 1990 som basisår blir da:

$$\frac{40,7}{43,9} \times 100 = 92,7$$

Og indeks for 1993 omregnet med 1990 som basisår blir:

$$\frac{41,2}{43,9} \times 100 = 93,8$$

Vi kan se at frem til og med år 2001, har det stort sett vært samme rangering av hvilke steder som har høyest prisnivå. Høyest prisnivå har det vært hele landet sett under ett, så kommer de 3 byene på andre plass, sentralt i Kristiansand på tredje plass, og til slutt usentralt i Kristiansand, på fjerde plass.
Fra år 2002 og til og med år 2008 har det vært de 3 byene som har ligget på prisnivå- toppen. Deretter følger hele landet, så sentralt, og til slutt usentralt.
Opplysninger om prisindekser for 2009 og 2010 var ikke tilgjengelige fra tabell 2.4, så derfor kan vi ikke sammenligne de siste årene. Indeksen fra 2010 er basert fra observasjoner til omtrent midten i februar, vi antar derfor at prisstigningen for 2010 er den samme for resten av året.

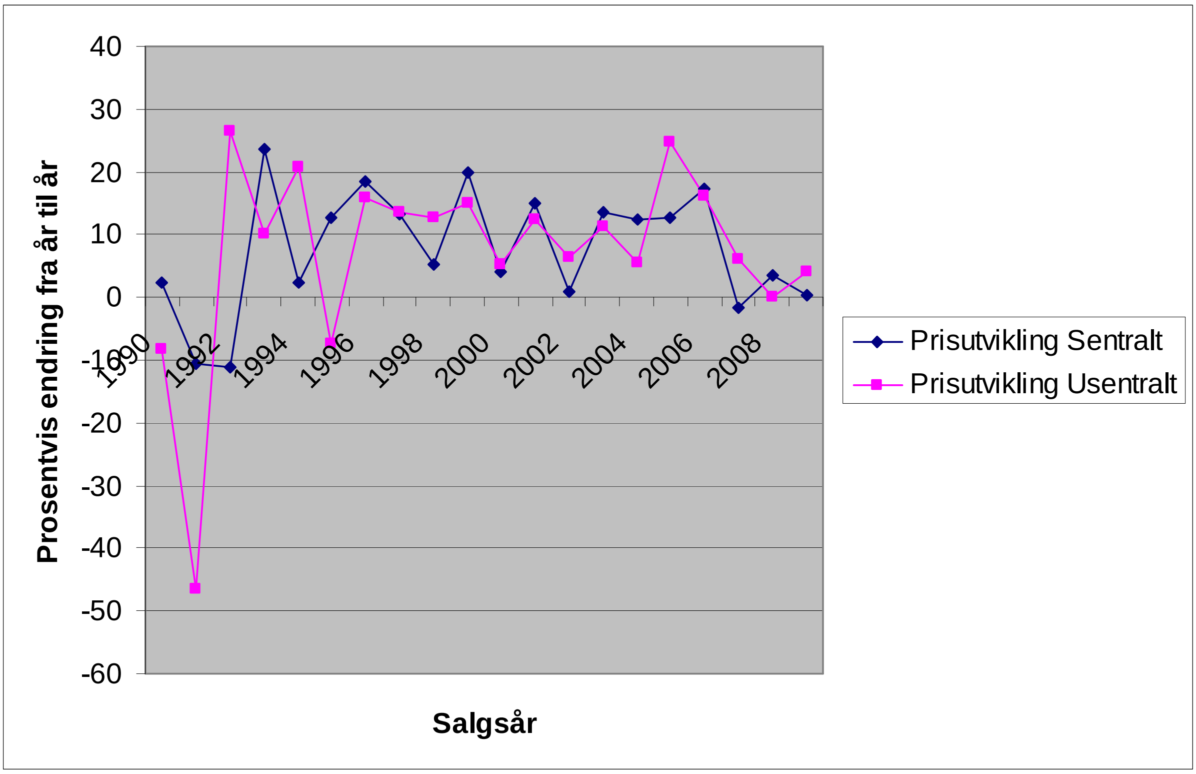
Tabell 7.4 Indekser fra Kristiansand sammenlignet med hele landet, og noen store byer

Salgsår:	Hele Landet:	Stavanger, Bergen og Trondheim	Sentralt:	Usentralt:
1990	100	100	100	100
1991	100	100	102,4	91,8
1992	92,7	94,1	91,9	45,3
1993	93,8	99,3	80,7	71,9
1994	110,9	113,8	104,3	82,1
1995	116,4	120,3	106,7	102,8
1996	128,9	132,3	119,4	95,5
1997	146,5	146,1	137,8	111,4
1998	167,2	165,1	151,2	124,9
1999	189,7	183,6	156,5	137,6
2000	227,8	218,3	176,5	152,7
2001	245,1	241,9	180,5	157,8
2002	261	267,7	195,6	170,2
2003	264	290	196,5	176,5
2004	297,5	345	210,2	187,8
2005	318,9	375,3	222,5	193,4
2006	371,7	441,9	235,3	218,3
2007	403,9	479,5	252,6	234,4
2008	378,1	444,1	250,9	240,6
2009	-	-	254,5	240,7
2010	-	-	254,8	244,9

Når det gjelder vår hovedhypotese er det ikke bare interessant å se hvilket sted som har det høyeste prisnivået, men utviklingen fra år til år.

For eksempel kan vi se fra år 1994 til 1995 at prisstigningen har vært på 2,4 % (106,7-104,3) sentralt, men på hele 20,7 % (102,8-82,1) for usentralt.

Disse prosentvise endringene fra år til år kan vi se i figuren nedenfor, 7.2. Som vi kan se, ligger ikke linjen til den prosentvise endringen for sentralt over linja for usentralt hele tiden. De går litt på kryss og tvers, det ene året er det mer økning sentralt, og det andre er det mer økning usentralt.



Figur 7.2 Prosentvis endring fra år til år

8. Hypotesetesting

I dette kapitlet skal hypotesene som vi utledet i kapittel 4 testes. Dette skal vi gjøre ved hjelp av den dobbeltlogaritmiske regresjonsformen. Gjennom utdataene til regresjonene vil vi da se om det er empirisk støtte til nullhypotesen, eller om vi må forkaste denne.

Hypotese 1: Alders påvirkning på salgsprisen

H0: Alder har ingen betydning på salgsprisen
H1: Alder påvirker salgsprisen negativt

Dette er en ensidig test fordi vi i alternativhypotesen ikke er interessert i om alderens påvirkning kun er forskjellig, men at den faktisk er negativ. Ved bruk av ensidige tester er den kritiske t-verdien på 1,645.
Nedenfor har jeg klippet ut den delen av regresjonen som gjelder for alderen på leiligheten, fra tabell 6.5. Først ser vi på koeffisienten og ser at den er negativ(-0,0399632). Dette tyder på at alder påvirker salgsprisen negativt. Ved å se på p-verdien sjekker vi om dette kan stemme. P-verdien på 0, gir gode signaler på at alderen faktisk påvirker salgsprisen negativt. Dette kan vi også se på konfidensintervallet, som kun dekker negative verdier.

Tabell 8.1 Utdrag fra tabell 6.5, regresjonsanalysen med alder

lnpris	Koef.	Std. Avv.	t	P>|t|	[95% Konf.	Interval]
lnalder	-.0399632	.0030311	-13.18	0.000	-.0459081	-.0340183

Konklusjon: Nullhypotesen forkastes, og vi kan akseptere alternativhypotesen, som sier at alder påvirker salgspris negativt.

Hypotese 2: Utvikling av årsdummy- koeffisienter over tid

H0: $\alpha_t = \alpha_{t+1}$, koeffisientene er like hvert år

H1: $\alpha_t < \alpha_{t+1}$, koeffisientene blir større for hvert år

Også dette er en ensidig test, fordi vi ønsker å teste om koeffisientene blir større for hvert år. I tabell 8.2 er et utdrag fra regresjonsresultatet fra tabell 6.5, som går på salgsårs- dummyene. Her må vi ta for oss hvert eneste år og se på p-verdiene for å sjekke om de er statistisk signifikante.

Tabell 8.2 Utdrag fra tabell 6.5, regresjonen med Salgsårsdummyer

lnpris	Koef.	Std. Avv.	t	P>|t|	[95% Konf.	Interval]
AR1991	.0458583	.0649399	0.71	0.480	-.0815087	.1732254
AR1992	-.1804272	.0706312	-2.55	0.011	-.3189567	-.0418977
AR1993	-.1922571	.0668439	-2.88	0.004	-.3233585	-.0611557
AR1994	.013756	.0633391	0.22	0.828	-.1104713	.1379833
AR1995	.1095964	.0625952	1.75	0.080	-.0131719	.2323648
AR1996	.1652838	.0591582	2.79	0.005	.0492564	.2813112
AR1997	.3427918	.0593018	5.78	0.000	.2264828	.4591009
AR1998	.4724874	.0613045	7.71	0.000	.3522505	.5927243
AR1999	.5574476	.0592584	9.41	0.000	.4412237	.6736715
AR2000	.739007	.0597642	12.37	0.000	.6217912	.8562228
AR2001	.7862421	.0586576	13.40	0.000	.6711966	.9012877
AR2002	.9290361	.0580938	15.99	0.000	.8150964	1.042976
AR2003	.9512117	.057514	16.54	0.000	.8384092	1.064014
AR2004	1.073737	.0556752	19.29	0.000	.9645412	1.182933
AR2005	1.18441	.0552631	21.43	0.000	1.076023	1.292798
AR2006	1.346657	.0551572	24.41	0.000	1.238477	1.454837
AR2007	1.516323	.0552025	27.47	0.000	1.408054	1.624591
AR2008	1.524362	.0558944	27.27	0.000	1.414736	1.633988
AR2009	1.545814	.0556398	27.78	0.000	1.436687	1.654941
AR2010	1.555774	.0806543	19.29	0.000	1.397586	1.713961

For å unngå multikollinearitet, er som vanlig den første av dummyene ikke med i regresjonen, derfor får vi ikke noe koeffisient til dette året. I 1991 kan vi derfor ikke vite om koeffisienten har vokst fra året før, men p-verdien er ikke lav nok til at dette salgsåret har signifikans. I

1994 og 1995 har begge voksende koeffisienter, men p-verdien er for høy til at de er signifikante. Etter dette er resten av salgsårsdummyene stigende, og signifikante.

Konklusjon: Skulle nullhypotesen være riktig, måtte alle koeffisientene til salgsårene være like. Dersom alternativhypotesen skulle være riktig, måtte alle koeffisientene til salgsårene være større enn for året før. Nullhypotesen kan i hvert fall forkastes, siden ingen av koeffisientene er like. Men, siden det ikke er signifikans for alle årene på at disse har stigende koeffisienter, kan vi heller ikke anta at alternativhypotesen er helt riktig.

Hypotese 3: Prisutvikling over tid

H0: Prisutviklingen i byens periferi har holdt tritt med prisutviklingen i byens senter

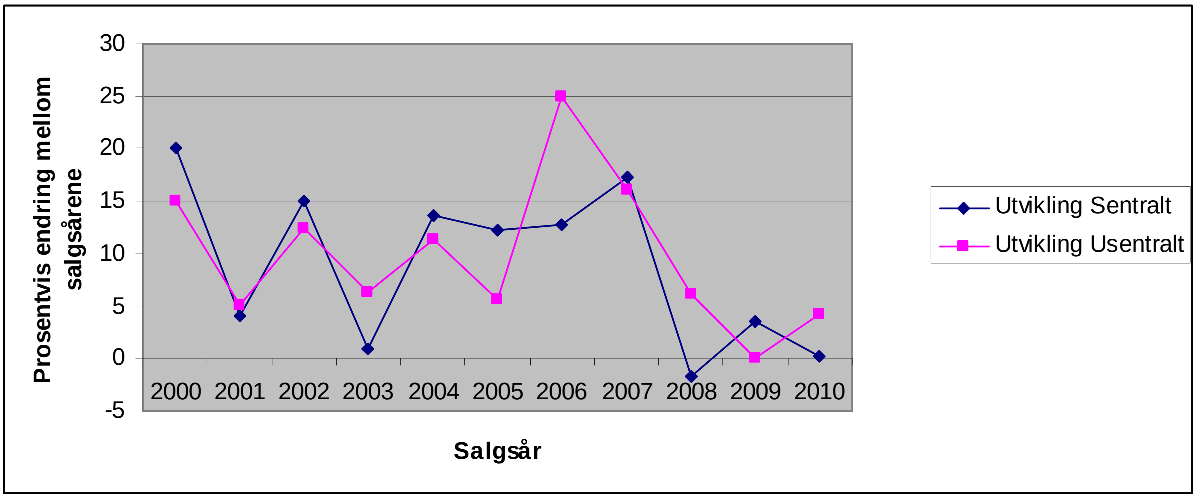
H1: Prisutviklingen i byens periferi har ikke holdt tritt med prisutviklingen i byens senter.

Til denne hypotesen må vi bruke den prosentvise endringen i prisindeksene, som det er mellom hvert salgsår.

Først tester vi om alle salgsårene er signifikante. De årene som ikke er signifikante, er det jo ikke noe særlig poeng at vi sammenligner. Vi finner fra tabell 7.1(sentralt) og 7.2(usentralt) følgende:

- Sentralt er år 1991-1995 ikke signifikante
- Usentralt er år 1991 - 1998 ikke signifikante.

Fra 1999-2010, er alle år signifikante for begge steder. Utviklingen i denne perioden er tatt med i figur 8.1 nedenfor. Dette er et utdrag av figur 7.2, men i figur 8.1 er det bare perioden med signifikante salgsår som er med.
Hadde det vært lik prosentvis utvikling fra år til år på begge steder, hadde linjene ligget oppå hverandre. Det kan vi se at de ikke gjør. Men vi kan se at utviklingen har vært tilnærmet lik frem til år 2002. Sentralt er det stort sett lavere prosentvis utvikling, når endringen er liten for begge, og høyere prosentvis økning når endringen er høy for begge. I 2006 skjer det noe radikalt usentralt (høy prosent endring), mens det sentralt er ganske liten endring.



Figur 8.1 Utdrag av figur 7.2

Konklusjon: Vi kan forkaste nullhypotesen, det er ikke lik utvikling av prisen begge steder, og vi kan anta at alternativhypotesen stemmer.

9. Konklusjon

Resultater

Formålet med denne undersøkelsen var å se om det var forskjell i prisutviklingen sentralt og usentralt i en by. Vi kom frem til at det var en forskjell, om enn minimal. Der den prosentvise veksten var høy, var den høy for begge, og der den var lav, var den lav for begge. Forskjellen var at for sentralt var det høyere prosentvis vekst når begge var høye, og lavere prosentvis vekst når begge var lave. Fra cirka år 2004 til 2007 var det ikke like stor tendens til at de varierte samme vei. Da stod den prosentvise veksten omtrent stille på et nivå for sentralt, mens det var kraftig vekst usentralt.

Så svaret på problemstillingen vår er at det er forskjellig prisutvikling, mellom sentralt og usentralt i Kristiansand.

Kritiske vurderinger

For det første burde vi kanskje tatt med flere av variablene fra eiendomsverdi, som kunne hatt påvirkning på både prisen og de andre variablene. For det andre kunne sampel størrelsen vært mye større, slik at vi fikk mer data og jobbe med. Da hadde analysene våre kanskje fanget opp mer enn det de gjorde. Vi kunne blant annet latt være å droppe alle casene som manglet opplysninger, i og med at noen av disse opplysningene antagelig hadde vært mulig å spore opp. Vi kunne også brukt flere byer, ikke bare Kristiansand.

I tillegg kan unøyaktighet ved behandling av datamaterialet, være en kilde til forstyrrelse. For eksempel ved overføringen fra Eiendomspris til STATA. Det kan også være unøyaktighet i den kilden som vi fikk tak i datamaterialet vårt fra.

Referanser:

Aarnes, H. (2000). *Litt Statistikk*. Lastet ned 03.06.10, fra: http://www.bio.uio.no/plfys/haa/littav/stat.htm#normalfordeling

Aarsland, L.G. (2009)*Leilighetskompleks i små lokalsamfunn- virkninger i eiendomsmarkedet" En analyse basert på data fra Egersund*. Universitetet i Agder.

Caplex. (2010). Artikkel. Lastet ned 12.06.10, fra: http://www.caplex.no/Web/ArticleView.aspx?id=9328455

Danton, E. & Siljan, C . (2005). *Faktorer som påvirker boligprisene i Kristiansand.* Universitetet i Agder.

Eiendomsverdi (2010) Lastet ned 12.02.10, fra: http://www.eiendomsverdi.no/

Hegre, H. (2008) *Hypotesetesting.* Lastet ned 12.04.10 fra: http://folk.uio.no/hahegre/index_files/Forelesninger/Hypotesetesting.pdf Universitetet i Oslo.

Jungeilges, J. (2008) Forelesningsnotater i Econometrics, ME-408. Universitetet i Agder.

Jørgensen, K. & Ubøe, J. (2004) Beskrivende statistikk. *Statistikk for Økonomifag.* 2.utgave, 1. opplag 2004. Utgiver: Gyldendal Norsk Forlag AS. s. 24-25

Kristiansand Kommune (2010). Historisk Utvikling. Lastet ned 20.04.10, fra: http://www.kristiansand.kommune.no/Om-Kristiansand/Historisk-utvikling/

Osland, L. (2001) Den hedonistiske metoden og estimering av attributtpriser. *Norsk Økonomisk tidsskrift*, 115, s.1-22.

Posten (2010) Lastet ned 28.04.10, fra: http://www.posten.no/Kundeservice/Kart?visible=bedrift;pib;post;dpg&checked=pib;post;dpg

Robertsen, K. & Theisen, T. (2009) Forelesningsnotater i Eiendomsøkonomi, BE-409. Universitetet i agder.

Rosen, S. (1987) Hedonic prices and Implicit Markets: Product Differentiation in Pure Competition. *The economic of Housing 1*. s. 34-55

Statistisk Sentralbyrå (2010a.). Folke og boligtellingen 2001. Lastet ned 13.03.10, fra: http://www.ssb.no/fob/kommunehefte/1001/fob_1001_tabeller.pdf

Statistisk Sentralbyrå (2010b). Byggekostnadsindeks. Lastet ned 13.03.10, fra: http://www.ssb.no/vis/bkibol/calc.cgi

Statistisk Sentralbyrå(2010c). Konsumprisindeksen. Lastet ned 13.03.10, fra:
http://www.ssb.no/vis/kpi/kpiregn.html

Statistisk Sentralbyrå (2010d). Statistikkbanken. Lastet ned 13.03.10, fra:
http://statbank.ssb.no/statistikkbanken/Default_FR.asp?PXSid=0&nvl=true&PLanguage=0&tilside=selectvarval/define.asp&Tabellid=05600

Store Norske Leksikon (2010) Lastet ned 12.06.10, fra:
http://www.snl.no/korrelasjon/statistikk

Vedlegg

Programmeringskoder til STATA

Nr 1.

Denne koden (nr1) lå som basis hver gang vi utførte en regresjon i STATA. Nr4 og nr5 måtte vi kjøres separat, siden det er dropformler i disse (Men også disse hadde basis i nr1).
Siden vi ikke lagret navnet på variablene våre, kan vi se på disse under:

Var1=salgscase, var2=Salgsmåned, var3=Salgsår, var4=postnummer, var5=kvm,var6=byggeår, var7=salgspris

```
*lager månedsdummyer

generate MND1 = 0
replace MND1 = 1 if var2 == 1
generate MND2 = 0
replace MND2 = 1 if var2 == 2
generate MND3 = 0
replace MND3 = 1 if var2 == 3
generate MND4 = 0
replace MND4 = 1 if var2 == 4
generate MND5 = 0
replace MND5 = 1 if var2 == 5
generate MND6 = 0
replace MND6 = 1 if var2 == 6
generate MND7 = 0
replace MND7 = 1 if var2 == 7
generate MND8 = 0
replace MND8 = 1 if var2 == 8
generate MND9 = 0
replace MND9 = 1 if var2 == 9
generate MND10 = 0
replace MND10 = 1 if var2 == 10
generate MND11 = 0
replace MND11 = 1 if var2 == 11
generate MND12 = 0
replace MND12 = 1 if var2 == 12

*lager årsdummyer

generate AR1990 = 0
replace AR1990 = 1 if var3 == 1990
generate AR1991 = 0
replace AR1991 = 1 if var3 == 1991
generate AR1992 = 0
replace AR1992 = 1 if var3 == 1992
generate AR1993 = 0
replace AR1993 = 1 if var3 == 1993
generate AR1994 = 0
replace AR1994 = 1 if var3 == 1994
```

```
generate AR1995 = 0
replace AR1995 = 1 if var3 == 1995
generate AR1996 = 0
replace AR1996 = 1 if var3 == 1996
generate AR1997 = 0
replace AR1997 = 1 if var3 == 1997
generate AR1998 = 0
replace AR1998 = 1 if var3 == 1998
generate AR1999 = 0
replace AR1999 = 1 if var3 == 1999
generate AR2000 = 0
replace AR2000 = 1 if var3 == 2000
generate AR2001 = 0
replace AR2001 = 1 if var3 == 2001
generate AR2002 = 0
replace AR2002 = 1 if var3 == 2002
generate AR2003 = 0
replace AR2003 = 1 if var3 == 2003
generate AR2004 = 0
replace AR2004 = 1 if var3 == 2004
generate AR2005 = 0
replace AR2005 = 1 if var3 == 2005
generate AR2006 = 0
replace AR2006 = 1 if var3 == 2006
generate AR2007 = 0
replace AR2007 = 1 if var3 == 2007
generate AR2008 = 0
replace AR2008 = 1 if var3 == 2008
generate AR2009 = 0
replace AR2009 = 1 if var3 == 2009
generate AR2010 = 0
replace AR2010 = 1 if var3 == 2010

*lager postnummerdummyer

generate PNR4610 = 0
replace PNR4610 = 1 if var4 == 4610
generate PNR4611 = 0
replace PNR4611 = 1 if var4 == 4611
generate PNR4612 = 0
replace PNR4612 = 1 if var4 == 4612
generate PNR4614 = 0
replace PNR4614 = 1 if var4 == 4614
generate PNR4622 = 0
replace PNR4622 = 1 if var4 == 4622
generate PNR4623 = 0
replace PNR4623 = 1 if var4 == 4623
generate PNR4624 = 0
replace PNR4624 = 1 if var4 == 4624
generate PNR4635 = 0
replace PNR4635 = 1 if var4 == 4635
generate PNR4639 = 0
replace PNR4639 = 1 if var4 == 4639
generate PNR4656 = 0
replace PNR4656 = 1 if var4 == 4656

*lager en ny kontinuerlig uavhengig variabel, Alder lik Salgsår minus
Byggeår

generate alder = var3 - var6
```

```
*dropper de casene som mangler opplysninger/har feil opplysninger

drop if var6>var3

drop if var5<0

drop if var6<0

drop if var8<0

*lager en pris for den dobbeltlogaritmiske formen

generate lnpris = log(var7)

generate lnkvm = log(var5)

*for at ikke jeg skal miste alle casene som har alder lik 0 når stata tar
logaritmen på alder, så setter jeg disse lik 0,1
replace alder=0.1 if alder==0

generate lnalder = log(alder)


* kommando for å kunne utføre regresjoner med alle variablene mine
set matsize 100
--
*lager deskriptiv statistikk tabell

summarize var2 var3 var4 var5 var6 var7 MND1 MND2 MND3 MND4 MND5 MND6 MND7
alder MND8 MND9 MND10 MND11 MND12  AR1990 AR1991 AR1992 AR
1993 AR1994 AR1995 AR1996 AR1997 AR1998 AR1999 AR2000 AR2001 AR2002 AR2003
AR2004 AR2005 AR2006 AR2007 AR2008 AR2009 AR2010 PNR4610 P
NR4611 PNR4612 PNR4614 PNR4622 PNR4623 PNR4624 PNR4635 PNR4639 PNR4656

*korrelasjonsmatrise mellom noen av variablene
correlate AR1990 AR1991 var5 var6 alder var7

*Korrelasjonsfigur mellom salgsprisen og antall kvadratmeter

twoway (scatter var7 var5), ytitle(Salgspris) xtitle(Antall Kvadratmeter)

*Korrelasjon mellom byggeår og alder
twoway (scatter var6 alder), ytitle(Byggeår) xtitle(Alder)

*regresjon med en variabel, kvadratmeter
regress var7 var5

*regresjonslinje i korrelasjonsplott med salgspris og kvadratmeter
twoway (scatter var7 var5) (lfit var7 var5), ytitle(Salgspris)
xtitle(Antall Kvadratmeter)

*lineær regresjon, hele modellen
regress var7 var5 MND2 MND3 MND4 MND5 MND6 MND7 MND8 MND9 MND10 MND11 MND12
AR1991 AR1992 AR1993 AR1994 AR1995 AR1996 AR1997 AR1998
AR1999 AR2000 AR2001 AR2002 AR2003 AR2004 AR2005 AR2006 AR2007 AR2008
AR2009 AR2010 PNR4611 PNR4612 PNR4614 PNR4622 PNR4623 PNR4624 P
NR4635 PNR4639 PNR4656 alder


*lager figur på normalfordeling av restleddet til den lineære modellen
```

```
predict lerror, resid
histogram lerror

*lager normalskråplottfigur for restleddet

pnorm lerror

*regresjon med den dobbeltlogaritmiske formen

regress lnpris lnkvm lnalder var5 MND2 MND3 MND4 MND5 MND6 MND7 MND8 MND9
MND10 MND11 MND12 AR1991 AR1992 AR1993 AR1994 AR1995 AR1996 AR1997AR1998
AR1999 AR2000 AR2001 AR2002 AR2003 AR2004 AR2005 AR2006 AR2007 AR2008
AR2009 AR2010 PNR4611 PNR4612 PNR4614 PNR4622 PNR4623 P
NR4624 PNR4635 PNR4639 PNR4656

*lager figur på normalfordeling av restleddet til den dobbeltlogaritmiske
modellen

predict derror, resid
histogram derror

*Så et skråplott for den dobbeltlogaritmiske modellen
pnorm derror
```

Nr. 2

```
* lager dummy for salg som har skjedd før og etter år 2000.

generate salgetter2000 = 0
replace salgetter2000 = 1 if
var3==2000|var3==2001|var3==2002|var3==2003|var3==2004|var3==2005|var3==200
6|var3==2007|var3==2008|var3==2009|var3==2010
generate salgfør2000 = 0
replace salgfør2000 = 1 if
var3==1990|var3==1991|var3==1992|var3==1993|var3==1994|var3==1995|var3==199
6|var3==1997|var3==1998|var3==1999

*regresjon med dummy for salg etter år 2000

regress var7 var5 salgetter2000
```

Nr. 3

```
* Forsøk på forbedring av modellen, ved å lage kvartalsdummyer

generate førstekvartal = 0
replace førstekvartal = 1 if var2==1|var2==2|var2==3

generate andrekvartal = 0
replace andrekvartal = 1 if var2==4|var2==5|var2==6

generate tredjekvartal = 0
replace tredjekvartal = 1 if var2==7|var2==8|var2==9

generate fjerdekvartal = 0
replace fjerdekvartal = 1 if var2==10|var2==11|var2==12
```

```
*regresjon med kvartalsdummyer

regress var7 var5 andrekvartal tredjekvartal fjerdekvartal MND2 MND3 MND4
MND5 MND6 MND7 MND8 MND9 MND10 MND11 MND12 AR1991 AR1992 AR1993 AR1994
AR1995 AR1996 AR1997 AR1998 AR1999  AR2000 AR2001 AR2002 AR2003 AR2004
AR2005 AR2006 AR2007 AR2008 AR2009 AR2010 PNR4611
PNR4612 PNR4614 PNR4622 PNR4623 PNR4624 PNR4635 PNR4639 PNR4656 alder


*normalfordeling av restledd ved forsøk på forbedring av modellen

predict kvartalerror, resid

histogram kvartalerror

*skråplott av forbedringsforsøket

pnorm kvartalerror
```

Nr. 4

```
* Case med postnummer fra usentrale områder droppes, slik at vi kan gjøre
regresjon bare på case sentralt

drop if PNR4622==1
drop if PNR4623==1
drop if PNR4624==1
drop if PNR4635==1
drop if PNR4639==1
drop if PNR4656==1

*utfører regresjonen for salgscase sentralt
regress  lnpris lnkvm lnalder MND2 MND3 MND4 MND5 MND6 MND7 MND8 MND9 MND10
MND11 MND12 AR1991 AR1992 AR1993 AR1994 AR1995 AR1996 AR1997 AR1998 AR1999
AR2000 AR2001 AR2002 AR2003 AR2004 AR2005 AR2006 AR2007 AR2008 AR2009
AR2010 PNR4611 PNR4612 PNR4614
```

Nr. 5

```
* Case med postnummer fra sentrale områder droppes, slik at vi kan gjøre
regresjon bare på
case usentralt

drop if PNR4610==1
drop if PNR4611==1
drop if PNR4612==1
drop if PNR4614==1

*utfører en regresjon for salgscase usentralt

regress  lnpris lnkvm lnalder MND2 MND3 MND4 MND5 MND6 MND7 MND8 MND9 MND10
MND11 MND12 AR1991 AR1992 AR1993 AR1994 AR1995 AR1996 AR1997 AR1998 AR1999
AR2000 AR2001 AR2002 AR2003 AR2004 AR2005 AR2006 AR2007 AR2008 AR2009
AR2010 PNR4623 PNR4624 PNR4635 PNR4639 PNR4656
```