



# Høgskulen på Vestlandet

## Bacheloroppgave

ØKB3113

### Predefinert informasjon

<b>Startdato:</b>	21-04-2020 09:00	<b>Termin:</b>	2020 VÅR
<b>Slutt dato:</b>	08-05-2020 14:00	<b>Vurderingsform:</b>	Norsk 6-trinns skala (A-F)
<b>Eksamensform:</b>	Bacheloroppgave - med muntlig presentasjon		
<b>SIS-kode:</b>	203 ØKB3113 1 PRO-1 2020 VÅR HAUGESUND		
<b>Intern sensor:</b>	(Anonymisert)		

### Deltaker

**Kandidatnr.:** 326

### Informasjon fra deltaker

**Tittel \*:** Estimering av implisitte priser for miljøvariabler. En empirisk analyse ved bruk av data fra Haugalandet.

**Engelsk tittel \*:** Estimating implicit prices of environmental variables. An empirical analysis using data from Haugalandet.

**Sett hake dersom ja**  
**besvarelsen kan brukes**  
**som eksempel i**  
**undervisning?:** Ja

**Egenerklæring \*:** Ja  
**Inneholder besvarelsen** Nei  
**konfidensielt**  
**materiale?:**

**Jeg bekrefter at jeg har ja**  
**registrert**  
**oppgavetittelen på**  
**norsk og engelsk i**  
**StudentWeb og vet at**  
**denne vil stå på**  
**vitnemålet mitt \*:**

### Gruppe

**Gruppenavn:** (Anonymisert)

**Gruppenummer:** 12

**Andre medlemmer i**  
**gruppen:** 310

Jeg godkjenner autalen om publisering av bacheloroppgaven min \*

Ja

Er bacheloroppgaven skrevet som del av et større forskningsprosjekt ved HVL? \*

Nei

Er bacheloroppgaven skrevet ved bedrift/virksomhet i næringsliv eller offentlig sektor? \*

Nei



Høgskulen  
på Vestlandet

## BACHELOROPPGAVE

Estimering av implisitte priser for miljøvariabler.  
En empirisk analyse ved bruk av data fra  
Haugalandet.

Estimating implicit prices of environmental  
variables. An empirical analysis using data from  
Haugalandet.

**Tom Lund-Andersen og Cristina Birgithe  
Tvihaug**

Økonomi og administrasjon

Fakultet for økonomi og samfunnsvitenskap / Institutt for  
økonomi og administrasjon / Finansiering og  
økonomistyring

Veileder: Liv Osland

Innleveringsdato: 22.05.2020

Vi bekrefter at arbeidet er selvstendig utarbeidet, og at referanser/kildehenvisninger til alle kilder som er brukt i arbeidet er oppgitt, jf. Forskrift om studium og eksamen ved Høgskulen på Vestlandet, § 12-1.

## Sammendrag

Denne oppgaven fokuserer på å finne de implisitte prisene for miljøvariabler på Haugalandet. Datasettet som er benyttet stammer originalt fra et arbeidskrav i emnet «By- og regionaløkonomi», men er blitt videre revidert av to bacheloroppgaver etter dette. Anglevik og Øvretveit (2018) utviklet en basemodell som vi brukte som utgangspunkt. Det ble benyttet analyseverktøy for å gjennomføre regresjonsanalyse, VIF-test og Breusch-Pagan-test gjennom oppgaven.

Vi analyserte basemodellen med seks miljøvariabler; støynivå, avstand til sjø, sjøfeste, solforhold, strandlinje og turmuligheter. På grunnlag av resultatene lagde vi en rangering som skulle danne utgangspunktet for vår endelige modell. Vi inkluderte deretter strandlinje, solforhold og sjøfeste i basemodellen, og resultatet av det ble vår endelige modell. Støynivå, avstand til sjø og turmuligheter ble utelatt fra den endelige modellen på grunn av resultatene vi fikk i analysene. En annen viktig del av oppgaven går på feilspesifisering. Det var korrelasjon mellom flere uavhengige variabler, målefeil, skjevhet og stor variasjon i antall observasjoner. Når dette er tilfelle, kan det føre til forventningsskjevne resultat.

Det er vanskelig å konkludere med hva de implisitte prisene for miljøvariablene blir i modellen, på grunn av at vår modell mest sannsynlig er feilspesifisert. Det vi derimot kan konkludere med, er at ved å inkludere *strandlinje*, *solforhold* og *sjøfeste* retter det opp en del av skjevhetene til *naust* og *utsikt*. De implisitte prisene for *naust* og *utsikt* representerer i større grad seg selv, og fanger ikke opp *strandlinje*, *solforhold* og *sjøfeste*. Selv om regresjonskoeffisientene til *naust* og *utsikt* er redusert, har de fremdeles høy økonomisk verdi. De er også statistisk signifikante og vi kan derfor med 95 % sikkerhet si at når disse variablene er til stede, vil de øke verdien på boligen betydelig. Vi kan ikke si noe om den økonomiske verdien for *strandlinje*, *solforhold* eller *sjøfeste* på grunnlag av at disse, i vår modell, ikke er statistisk signifikante. Det vi tar med oss er videre, er at fortegnene til variablenes regresjonskoeffisienter stemmer overens med våre antagelser og hva tidligere forskning sier.

## Abstract

This thesis focuses on finding the implicit prices of environmental variables on Haugalandet. The dataset that has been used was originally a course requirement in the subject «By- og regionaløkonomi» but has been further revised by two bachelor theses afterwards. Anglevik and Øvretveit (2018) developed a base model, that we used as the starting point for our thesis. We used analytic tools to conduct a regression analysis, VIF-test and Breusch-Pagan-test throughout the thesis.

We analysed the base model with six environmental variables: noise level, the distance to the sea, oceanfront, sunshine, shoreline and hiking opportunities. Based on these results we made a ranking which was going to form the basis of our final model. We then included shoreline, sunshine and oceanfront to the base model, and the result was our final model. Noise level, the distance to the sea and hiking opportunities was omitted from the final model based on the results we got in the analyses. Another important part of the thesis is incorrect specification. There was correlation between multiple independent variables, measurement error, bias and there was great variation in the number of observations for each variable. When this is the case, it can lead to biased results.

It is hard to conclude what the implicit prices of the environmental variables will be based on our model, seeing that our model most likely is incorrectly specified. What we, on the other hand, can conclude is that by including *shoreline*, *sunshine* and *oceanfront* we correct some of the bias for *boathouse* and *view*. The implicit prices for *boathouse* and *view* represent themselves to a greater extent and are not affected by the influence of variables such as *shoreline*, *sunshine* or *oceanfront*. Even though the regression coefficients for *boathouse* and *view* are reduced, they still have high economic value. They are also statistically significant, and we can therefore say with 95 % certainty that when these variables are present, they will increase the value of the house considerably. We are unable to comment on the economic value of *shoreline*, *sunshine* or *oceanfront* on the basis that these in our model, are not statistically significant. What we bring from this thesis is that the sign of the variables' regression coefficients are consistent with our assumptions and what previous research says.

## Forord

Denne oppgaven er et avsluttende arbeid ved bachelorstudiet i økonomi og administrasjon på Høgskulen på Vestlandet, våren 2020.

Arbeidet med denne oppgaven ble langt fra slik vi hadde sett for oss. Det har vært en utfordrende vår, både med tanke på oppgaven og situasjonen ellers i samfunnet. Det å ikke kunne møtes for å arbeide sammen, samt det å ikke kunne snakke direkte med veileder og andre ansatte ved høgskolen, har gjort arbeidet med oppgaven vanskeligere. Vi hadde en jevn progresjon i starten, før det ble flere uker med relativt lite arbeid. Det har resultert i at de siste ukene har vært meget hektiske for å komme i mål med oppgaven. For å få kunne fullføre oppgaven har vi sett oss nødt til å be om utsettelse i to uker. Med tanke på situasjonen rundt, har det ført til at det er en del ting vi ikke har fått tid til å gjennomføre. Vi hadde blant annet planer om å bruke QGIS til å innhente mer data om variablene og gjennomføre robuste standardavvik.

Vi ønsker å rette en stor takk til vår fabelaktige veileder, Liv Osland. Hun har gjennom hele perioden kommet med råd til hvordan vi skal håndtere de utfordringene vi har stått overfor og gitt oss gode tilbakemeldinger på det vi har gjort. Vi ønsker også å gi en stor takk til Arnstein Gjestland for all den hjelp vi har fått. Uten hans kunnskap og hjelp hadde arbeidet med og rundt dataprogrammene tatt vesentlig lenger tid, og resultatet hadde sett annerledes ut.

# Innholdsfortegnelse

Sammendrag.....	i
Abstract.....	ii
Forord .....	iii
Figuroversikt .....	v
Tabelloversikt.....	v
Analyseoversikt.....	vi
1 Innledning .....	1
2 Teori.....	2
2.1 Boligmarkedet som studeres.....	4
2.2 Den hedoniske metode .....	7
2.3 Budrente-modellen .....	9
2.4 Miljøvariabler .....	11
2.4.1 Utsikt.....	13
2.4.2 Støynivå.....	14
2.4.3 Avstand til sjø.....	14
2.4.4 Sjøfeste .....	14
2.4.5 Strandlinje .....	14
2.4.6 Solforhold.....	15
2.4.7 Turmuligheter .....	15
3 Metode .....	16
3.1 Regresjonsanalyse .....	16
3.1.1 Enkel lineær regresjonsanalyse .....	16
3.1.2 Minste kvadraters metode.....	17
3.1.3 Forventningsrett og skjevhet.....	18
3.1.4 Konsistens .....	19
3.1.5 Prediksjon.....	19
3.1.6 Best Linear Unbiased Estimator (BLUE) .....	20
3.1.7 Korrelasjon .....	21
3.1.8 Multikollinearitet .....	21
3.1.9 T-verdi.....	22
3.1.10 Intern validitet.....	22
3.2 Dataprogrammer.....	24
3.2.1 Excel .....	24
3.2.2 RStudio.....	24
3.3 Datagrunnlag for analysen .....	24
3.3.1 Datasett.....	24
3.3.2 Basemodellen.....	27
3.3.3 Variabler.....	29
3.3.4 Deskriptiv statistikk av miljøvariablene .....	30
4 Resultat .....	35
4.1 Basemodell.....	35
4.2 Basemodell med støynivå.....	36
4.3 Basemodell med avssjø .....	39

4.4 Basemodell med sjofeste .....	42
4.5 Basemodell med solforhold .....	44
4.6 Basemodell med strandlinje .....	46
4.7 Basemodell med turbuligheter .....	48
5 Feilspesifisering .....	51
5.1 Korrelasjon .....	51
5.2 Målefeil .....	52
6 Rangering av miljøvariablene .....	54
6.1 Regresjonsanalyse av basemodellen inkludert strandlinje og solforhold .....	55
6.2 Regresjonsanalyse av basemodellen inkludert strandlinje, solforhold og sjofeste .....	56
7 Drøfting av endelig modell .....	59
8 Konklusjon .....	62
9 Videre forskning .....	64
10 Litteraturliste .....	65

## Figuroversikt

Figur 1: prisutvikling i Norge de siste 17 årene (Eiendom Norge, 2020) .....	3
Figur 2: Illustrasjon av kommunene i Haugalandet (Haugalandet_no) .....	4
Figur 3: Utviklingen i antall boliger i hver kommune (Statistisk Sentralbyrå, u.d.) .....	5
Figur 4: Gjennomsnittlig pris per år for boligene i datasettet .....	6
Figur 5: Den hedoniske prisfunksjonen som likevektsfunksjon, og illustrerer likevekt (*) for to konsumenter (Osland, 2016) .....	8
Figur 6: Budrentekurver for en husholdning (McCann, 2013, s. 122) .....	10
Figur 7: Effektene av miljøvariasjoner på budrentekurver (McCann, 2013, s. 127) .....	11
Figur 8: 95% av arealet under normalfordeling ligger innenfor 1,96 standardavvik fra gjennomsnittet (Wikipedia, 2020) .....	22
Figur 9: Prosentvis fordeling av antall observasjoner i hver kommune .....	25
Figur 10: Gjennomsnittlig kvadratmeterpris for enebolig per kommune (Statistisk Sentralbyrå, u.d.) .....	26

## Tabelloversikt

Tabell 1: Oppsummering av skjevhet for $\beta_1$ når $x_2$ er utelatt (Wooldridge, 2016, s. 79) .....	19
Tabell 2: Tommelfingerregel for korrelasjon (Johannessen, Christoffersen, & Tufte, 2011, s. 322) .....	21
Tabell 3: Gjennomsnittlig kvadratmeterpris i datasettet .....	25



Tabell 4: Basemodellens variabler (Anglevik & Øvretveit, 2018).....	28
Tabell 5: Region- og kommunesenter (Anglevik & Øvretveit, 2018).....	28
Tabell 6: Tilgjengelighetsvariabler (Anglevik & Øvretveit, 2018).....	29
Tabell 7: Variablene som undersøkes .....	29
Tabell 8: Deskriptiv analyse av støynivå.....	30
Tabell 9: Deskriptiv analyse av avssjo.....	31
Tabell 10: Deskriptiv analyse av sjofeste.....	31
Tabell 11: Deskriptiv analyse av solforhold.....	32
Tabell 12: Deskriptiv analyse av strandlinje.....	32
Tabell 13: Deskriptiv analyse av turmuligheter.....	33
Tabell 14: Korrelasjonsanalyse av uavhengige variabler .....	33

## Analyseoversikt

Analyse 1. Regresjonsanalyse av basemodellen til Anglevik og Øvretveit (2018).....	35
Analyse 2. VIF-verdier for basemodellen. ....	36
Analyse 3. Breusch-Pagan-test av basemodellen. ....	36
Analyse 4. Regresjonsanalyse av basemodellen inkludert støynivå.....	38
Analyse 5. VIF-verdier for basemodellen inkludert støynivå.....	39
Analyse 6. Breusch-Pagan-test av basemodellen inkludert støynivå. ....	39
Analyse 7. Regresjonsanalyse av basemodellen inkludert avssjo.....	40
Analyse 8. VIF-verdier for basemodellen inkludert avssjo.....	42
Analyse 9. Breusch-Pagan-test av basemodellen inkludert avssjo.....	42
Analyse 10. Regresjonsanalyse av basemodellen inkludert sjofeste.....	43
Analyse 11. VIF-verdier for basemodellen inkludert sjofeste.....	44
Analyse 12. Breusch-Pagan-test av basemodellen inkludert sjofeste.....	44
Analyse 13. Regresjonsanalyse av basemodellen inkludert solforhold.....	45
Analyse 14. VIF-verdier for basemodellen inkludert solforhold.....	45
Analyse 15. Breusch-Pagan-test av basemodellen inkludert solforhold.....	46
Analyse 16. Regresjonsanalyse av basemodellen inkludert strandlinje.....	47
Analyse 17. VIF-verdier for basemodellen inkludert strandlinje.....	48
Analyse 18. Breusch-Pagan-test av basemodellen inkludert strandlinje. ....	48
Analyse 19. Regresjonsanalyse av basemodellen inkludert turmuligheter.....	49
Analyse 20. VIF-verdier for basemodellen inkludert turmuligheter. ....	50
Analyse 21. Breusch-Pagan-test av basemodellen inkludert turmuligheter.....	50
Analyse 22. VIF-verdier for basemodellen inkludert strandlinje og solforhold.....	56
Analyse 23. Regresjonsanalyse av basemodellen inkludert strandlinje, solforhold og sjofeste.....	57
Analyse 24. VIF-verdier for basemodellen inkludert strandlinje, solforhold og sjofeste. ....	58

## 1 Innledning

Kjøp og salg av boliger er et tema vi forholder oss til hver eneste dag. Det dukker alltid opp en boligannonse et sted, eller at noen vi kjenner har kjøpt en bolig. Vi er i aldersgruppen hvor mange enten har tenkt å kjøpe seg eller allerede har kjøpt sin første bolig, noe som gjør dette temaet høyst aktuelt for oss.

Haugalandet består av kommunene Haugesund, Karmøy, Tysvær, Bokn, Vindafjord, Sauda, Sveio og Etne. De to sistnevnte kommunene tilhører Vestland fylke, men Bhuller (2009) trekker frem andre definisjoner på Haugalandet hvor disse kommunene er inkludert. Anglevik og Øvretveit (2018) argumenterer for å inkludere Sauda av den grunn at det vil være en god utkantkommune for regionsenteret, Haugesund.

Oppgavens problemstilling er å finne ut hva de implisitte prisene for miljøvariabler er for en bolig. Miljøvariablene som er inkludert er; utsikt, støynivå, avstand til sjø, sjøfeste, solforhold, strandlinje og turmuligheter. Vi skal gjennom oppgaven finne ut om de har en positiv eller negativ innvirkning på boligprisen. Vi vil også se på hva som skjer med variablene dersom de inneholder skjevheter, målefeil eller om de korrelerer med allerede eksisterende variabler i basemodellen. Målet er å til slutt få på plass en endelig modell som tar for seg de miljøvariablene som er av betydning.

## 2 Teori

Ifølge Kotler & Keller (2016, s. 37) er et marked ofte beskrevet som en samling av kjøpere og selgere som utfører transaksjoner i forbindelse med et bestemt produkt eller en bestemt produktgruppe. Enklere sagt er et marked en samling av tilbud og etterspørsel, som for eksempel boligmarkedet.

En bolig er en boenhet som er bygd eller ombygd som helårs privatbolig for en eller flere personer og som har egen atkomst uten å gå gjennom en annen bolig (Statistisk Sentralbyrå, 2020). En bolig blir regnet som både et konsumgode og formuesobjekt, som betyr at man både kjøper boligjenestene som boligen produserer, samtidig som man investerer i formuesobjektet (NOU 2002:2, 2002, s. 17). Bolig skiller seg ut fra andre goder fordi det er et heterogent gode, har lang levetid og det forekommer høye kostnader. Selv om boliger kan ligne på hverandre, er det variasjon når det kommer til blant annet alder, størrelse og beliggenhet (O'Sullivan, 2007).

Ulike typer bolig er enebolig, tomannsbolig, rekkehus, leilighet og hybel. Ettersom alle boliger er forskjellige, er boligmarkedet et komplisert og uoversiktlig marked. Det består av mange delmarkeder, hvor aktørene er ulike og beslutningene som blir tatt ofte har en svært lang tidshorisont (NOU 2002:2, 2002, s. 17). Boligmarkedet kan deles inn på mange ulike vis. Eksempler på delmarkeder som det kan deles inn etter er boligtype, alder og kvalitet, samtidig som man ofte skiller mellom eksisterende bolig og nybygg. Man skiller her blant annet mellom kjøpe- og leiemarkedet (Smith, Rosen, & Fallis, 1988, s. 30). Man kan også definere boligmarkedet geografisk, for eksempel dersom det er mange som pendler innad i en region, eller inn til en region, regnes dette som ett marked (Bhuller, 2009). Noen definerer Haugalandet på denne måten, når de velger å inkludere Etne og Sveio i regionen.

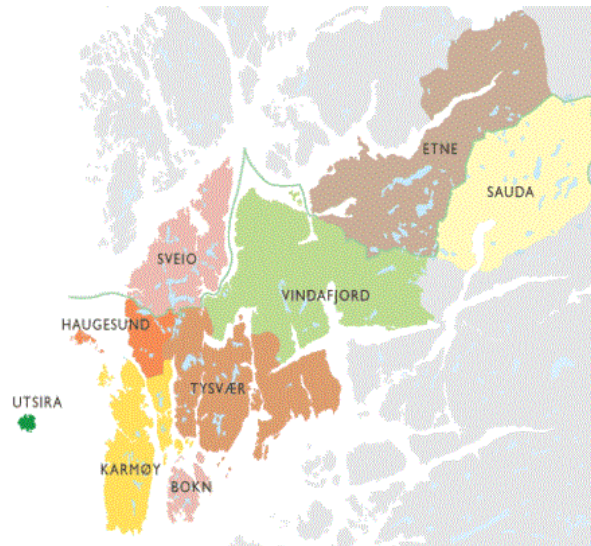


Figur 1: Prisutviklingen i Norge de siste 17 årene (Eiendom Norge, 2020).

Det er over 2,5 millioner registrerte boliger i Norge (Statistisk Sentralbyrå, 2020). Flertallet av husholdningene i Norge bor i en bolig de eier, og verdien av boligen utgjør ofte brorparten av husholdningenes samlede formue. Hovedsakelig bestemmes boligprisene av tilbud og etterspørsel. Faktorer som påvirker etterspørselssiden er blant annet arbeidsledighet, inntektsutvikling, demografi, renter og kredittvekst. For eksempel har arbeidsledighet betydning for inntekten i økonomien og størrelsen på inntekten påvirker igjen etterspørselen (Jansen, 2011). Vi kan i figur 1 se at det har vært en jevn og stigende trend for boligprisene i Norge de siste 17 årene, men med noen mindre fall.

Oljeprisfallet i 2014 er en viktig faktor for boligprisene på Haugalandet, men hadde liten innvirkning på det nasjonale boligmarkedet. Selv om oljeprisen falt, fortsatte boligprisene å stige, og i 2015 så man en 7 % økning på landsbasis. Oljeprisfallet påvirket i liten grad på nasjonalt nivå, men påvirket Rogaland negativt. Rundt 40 % av alle arbeidsplasser i Rogaland hadde tilknytning til petroleumsnæringen, og var dermed det fylket med størst andel knyttet til næringen. Arbeidsledigheten økte mer i Rogaland, enn i resten av landet, som også førte til lavere befolkningsvekst. Disse to faktorene førte til at Rogaland var en av de eneste fylkene i landet som hadde nedgang i boligprisene i 2015 (Senneset, 2015).

## 2.1 Boligmarkedet som studeres

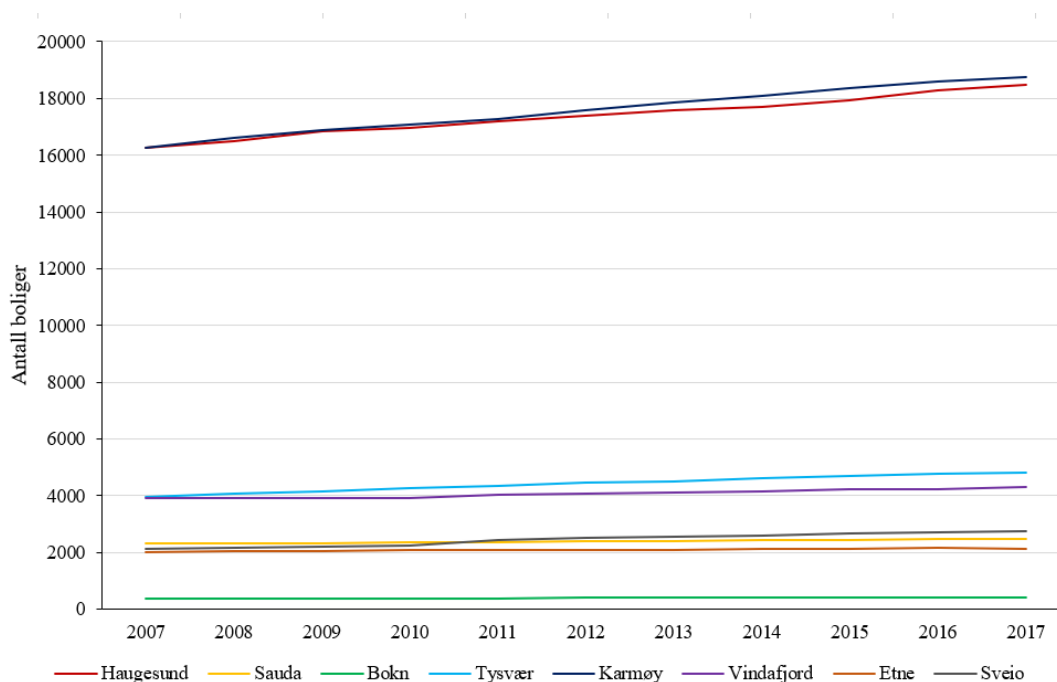


Figur 2: Illustrasjon av kommunene på Haugalandet (*Haugalandet\_no*).

Haugalandet er en region i Sørvest-Norge og består av Rogalandskommunene Haugesund, Karmøy, Tysvær, Bogn, Utsira og Vindafjord. Den ytre delen av Haugalandet, med unntak av Utsira, har en del større befolkning enn resten av regionen, særlig i og rundt Haugesund. Dette, samt at en del næringsliv også er lokalisert i og rundt kommunen, har medført at Haugesund er blitt det naturlige sentrum for regionen (Thorsnæs, 2017). Haugalandet er en multisentrisk økonomi, der det er mange “undersentre” som fungerer som lokale samlingspunkter for næringsliv og kommersiell aktivitet (McCann, 2013, s. 133). For eksempel er Kopervik kommunesenteret i Karmøy kommune, men det er likevel næringsområder på blant annet Norheim, Bygnes og Vea.

Vi gjør som Anglevik og Øvretveit (2018) og velger å inkludere de daværende Sunnhordlandskommunene Sveio og Etne, samt Sauda som ligger i Ryfylke, som en del av regionen. Disse kommunene tas med av den grunn at det flere steder argumenteres for at Sveio og Etne tilhører Haugalandet. Sauda vil fungere som en god utkantkommune for det sentrale arbeidsområdet i Haugesundsområdet og det vil øke spredningen i avstandsfølsomheten. Det vil si hvor villig man er til å pendle langt. Det tar nesten dobbelt så lang tid å pendle fra Haugesund til Sauda som det tar å pendle til Etne. De resterende kommunene på Haugalandet har kortere reisetid fra Haugesund, og derfor gir Sauda bedre spredning.

I Bhuller (2009) finner Anglevik og Øvretveit (2018) støtte for den nevnte seleksjonen av kommuner. Sveio, Etne og Sauda ligger i samme arbeidsmarked som Haugalandet og Haugesund fungerer som et regionsenter også for disse kommunene. Når det er mulighet for at arbeidsmarkedsregioner kan krysse fylkesgrensene, brukes en regel om at en kommune må ha 10 % eller mer utpendling til en naboregion for at den skal sammenslås med den regionen. På tidspunktet for inndeling av arbeidsmarkedet presentert i Bhuller (2009) var det mer enn 40 % som pendlet til Haugesundregionen fra Sveio, og 18 % fra Etne, og disse kan derfor slås sammen med Haugesundregionen. Vi velger å ekskludere Utsira fra vår analyse, selv om kommunen er en del av Haugalandet. Det er på grunn lag av at det er en svært liten kommune med 198 innbyggere (Statistisk Sentralbyrå, u.d.). Det vil gjøre at kommunen vil gi lite utslag i vår analyse, da datagrunnlaget er betydelig mindre enn i de resterende kommunene. Anglevik og Øvretveit (2018) og Hovland, Kvalvågnes og Kvalevåg (2019) tok også valget om å ekskludere Utsira fra analysene.

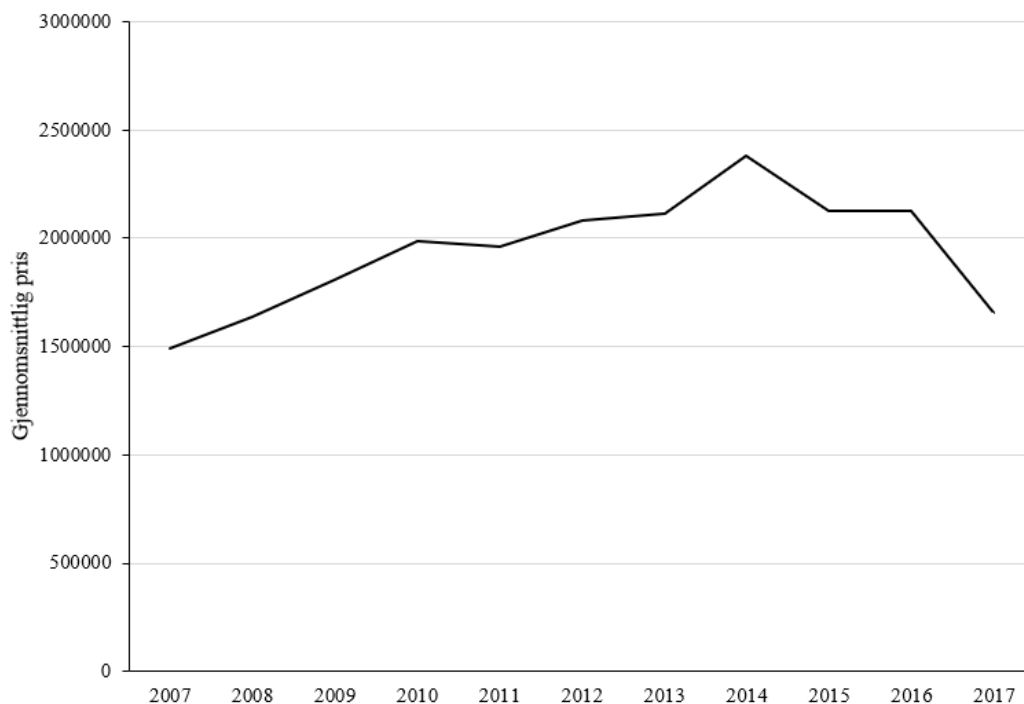


Figur 3: Utviklingen i antall boliger i hver kommune (Statistisk Sentralbyrå, u.d.).

Som illustrert i figur 3 har utviklingen i antall boliger i regionen vært svakt stigende i tiåret som studeres. Maksimal økning på ti år er omtrent 2 000 boliger, som var økningen i både Haugesund og Karmøy, mens de resterende kommunene hadde betydelig lavere økning. Man kan sammenligne antall boliger med befolkningsutviklingen i de ulike kommunene. Haugesund og Karmøy hadde også størst befolkningsvekst på Haugalandet fra årene 2010 til 2017, der

Haugesund hadde en økning på over 3 000 og Karmøy en økning på over 2 500 innbyggere. Tysvær økte tredje mest i både befolkning og antall boliger, hvor økningen var på omtrent 1 000 både for bolig og innbyggere (Statistisk Sentralbyrå, u.d.).

Som tidligere nevnt hadde oljeprisfallet større betydning for Rogaland, enn det hadde nasjonalt. I Haugesund var det 5,5 % av arbeidsstyrken som var helt arbeidsledige i november 2015, sammenlignet med 4,3 % i Stavanger (Sundberg, 2015). Haugalandet er et mindre urbant område enn Stavangerregionen. Stavangerregionen er et storbyområde som har bedre tilgang til alternative arbeidsplasser, sammenlignet med Haugalandet. Det kan være en av grunnene til at Haugalandet hadde mer arbeidsledighet enn Stavanger i 2015. Regionen er mer sårbar mot fall i viktige næringer, på grunn av at den har færre alternativer enn i Sør-fylket. På Haugalandet er det mange arbeidsplasser som er tilknyttet den maritime sektoren, og viktige aktører i regionen er blant annet Equinor, Gassco, Aibel, Sjøfartsdirektoratet og Kystverket. De ulike næringene i regionen er knyttet tett sammen, og store aktører påvirker byen i større grad. For eksempel kan Equinor ha lite jobb, som videre kan føre til at Aibel får færre prosjekter, som også fører til at det kan gå dårlig med andre næringer i regionen. Derfor hadde oljeprisfallet en større påvirkning i vår region enn resten av landet.



Figur 4: Gjennomsnittlig boligpris for boligene i datasettet.

Figur 4 illustrerer gjennomsnittlig salgspris på boliger per år på Haugalandet i tiårsperioden vi studerer. Vi har ikke tatt hensyn til kvalitetsforskjeller, men kun fokusert på salgspris og omsetningsår. Det er en jevn prisøkning i regionen i flere år, men etter året 2014 begynner prisene å falle. Det antar vi at kan ha en sammenheng med oljeprisfallet i 2014. Fra 2007 til 2014 økte gjennomsnittsprisen på boliger med nesten en million. Dersom vi studerer figur 4 kan vi se at prisene i 2017 nesten var tilbake på samme prisnivå som i 2007.

Det bygges mange nye boliger rundt omkring på Haugalandet, både i form av leilighetskomplekser, tomannsboliger, rekkehus og eneboliger. Vi ser at utbyggere som blant annet Berge Sag AS, A. Utvik AS, Odd Hansen AS og Askeland Eiendomsutvikling AS bygger opp nye nabolag i hele regionen. Berge Sag bygger for eksempel ferdige boliger i både Haugesund, Karmøy, Tysvær og Sauda (Berge Sag AS, u.d.).

## 2.2 Den hedoniske metode

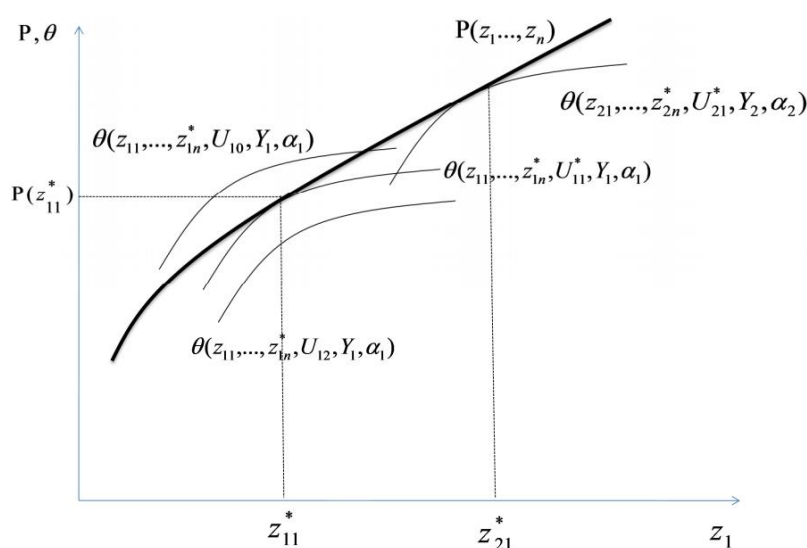
Vi skal i denne oppgaven bruke hedonisk metode for å beregne markedsverdien for ulike miljøvariabler. Hedonisk metode er mye anvendt i studier av boligmarkedet. Metoden tar utgangspunkt i at det heterogene godet, som i dette tilfellet er bolig, er karakterisert ved ulike egenskaper eller attributter, og at det er de ulike attributtene som gir glede eller nytte. Attributter for bolig kan være størrelse, nærhet til friluftsområde og utsikt (Osland, 2001). Hedonisk metode antar at markedet for bolig er preget av konkurranse, hvor begge sider av markedet har full informasjon og ønsker høyest mulig nytte eller fortjeneste (Osland, 2016).

I forklaringen på hedonisk metode følger vi Oslands (2016) beskrivelse ganske tett. Det er samspillet mellom tilbuds- og etterspørselssidene som gir den hedoniske prisfunksjonen på bolig. Totalprisen på boligen blir en funksjon av mengden attributter og deres implisitte priser. Som tidligere nevnt observeres ikke de implisitte prisene til godene direkte, men de beregnes indirekte via totalprisen på boligen. Her inngår både goder som er knyttet til selve boligen, for eksempel areal og alder, og attributter knyttet til lokaliseringen. Dette kan være både avstandsvariabler og sosiale faktorer. Totalprisen blir en funksjon av egenskapene ved boligen og omgivelsene og deres implisitte priser. Den hedoniske prisfunksjonen formuleres generelt slik:

$$P = f(B, M, T)$$



I denne funksjonen er  $P$  markedsprisen, som vi finner ved en funksjon av  $B$ ,  $M$  og  $T$ . Hvor  $B$  er direkte egenskaper ved boligen (størrelse, alder, boligtype),  $M$  er mengden/kvaliteten på ulike miljøgoder og  $T$  som for eksempel kan være tilgjengelighetsfaktorer (avstand til sentrum, skole, osv.). I modellen er utgangspunktet at nevnte egenskaper er viktig og har verdi for husholdninger. En forutsetning i hedonisk metode er at markedet skal være fritt og uregulert, noe som i stor grad er tilfredsstillende for boligmarkedet i Norge (Osland, 2016). Modellen brukes også på utleiemarkedet, men vi skal kun fokusere på markedet hvor det kjøpes og selges boliger.



Figur 5: Den hedoniske prisfunksjonen som likevektsfunksjon, og illustrerer likevekt (\*) for to konsumenter (Osland, 2016).

Osland (2016) forklarer den hedoniske prisfunksjonen som en likevektsfunksjon som dekkes av budfunksjoner til hver konsument. Figur 5 presenterer en ikke-lineær hedonisk prisfunksjon  $P(Z)$  for ulike nivåer av  $z_1$ , som i dette eksempelet er luftkvalitet. Den vertikale aksene i figuren viser totalprisen på bolig og den horisontale aksene viser mengden av  $z_1$ . Økt mengde av miljøgodet  $z_1$  vil gi bedret luftkvalitet. En antagelse i figuren er at konsumenten er optimalt tilpasset for alle andre egenskaper. I figuren ser vi at prisen på bolig ( $P$ ) stiger når vi øker mengden av  $z_1$ .

Konsumentenes nytte maksimeres dersom  $U_i = (z_1, \dots, z_n, X, \alpha_i)$  maksimeres gitt at  $Y_i = X + P(Z)$ . I denne sammenheng vil  $P$  fremdeles være prisen på bolig,  $Z$  være de ulike egenskapene ved en bolig,  $Y$  er inntekt og  $X$  er et privat konsumgode, hvor prisen settes lik 1.  $\alpha_i$  beskriver

preferansene og behovene til konsumentene. Figur 5 illustrerer også budfunksjonene ( $\theta$ ) til to konsumenter. For konsument 1 illustreres tre kurver som gir ulike nyttenivå. Desto lavere budkurven ligger i diagrammet, desto høyere nyttenivå oppnås. Nykten er maksimert når man kommer på den lavest oppnåelige budkurven og den tangerer den hedoniske prisfunksjonen, altså vil konsumenten betale en lavere pris for et gitt nivå av luftkvalitet. Konsument 2 har høyere preferanser for miljøgodet og/eller høyere inntekt, og betaler høyere pris og i tillegg får bolig i et område med bedre luftkvalitet. Når det finnes et stort antall konsumenter, vil den hedoniske prisfunksjonen dekkes av ulike konsumenters laveste oppnåelige budkurver (Osland, 2016).

Anglevik og Øvretveit (2018) uttrykker en lineær hedonisk boligprismodell slik:

$$P = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \dots + \beta_N * X_N + \varepsilon$$

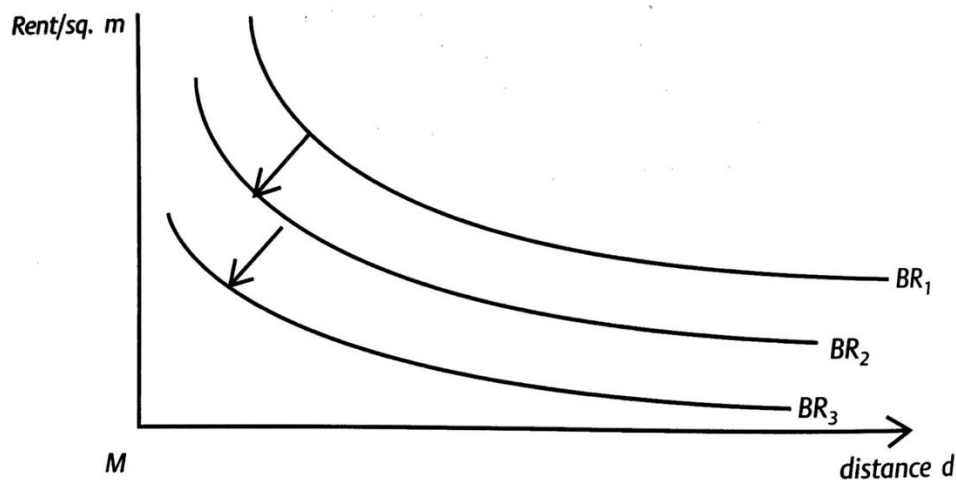
I denne formelen er  $P$  er den avhengige variabelen, som i dette tilfellet er prisen på en gitt enebolig.  $\beta_0$  er konstantleddet.  $X_N$  representerer alle de uavhengige variablene, som for eksempel boligalder, primærrrom og omsetningsår.  $\beta_N$  representerer regresjonskoeffisientene, som forklarer oss hvor mye den avhengige variabelen, boligpris, endres dersom den uavhengige variabelen øker med en måleenhet. Regresjonskoeffisientene blir våre implisitte priser, på for eksempel variabelen utsikt.  $\varepsilon$  er det stokastiske restleddet og angir avviket mellom den observerte og predikerte prisen på en gitt enebolig. Disse avvikene kalles for residualer eller restledd (Anglevik & Øvretveit, 2018). Vår endelige boligprismodell blir presentert etter at vi har gjennomført regresjonsanalysene for å avgjøre hva som skal inkluderes og ikke.

### 2.3 Budrente-modellen

Budrente-modellen er viktig å forstå når vi skal studere prisendringer ved endring i avstand og hvordan en negativ miljøvariabel som forurensning kan ha en negativ innvirkning på boligpriser. Datasettet vi benytter har definert både avstand til regionsenter og avstand til kommunesenter, og dersom boligprisen avtar med økt avstand kan dette knyttes direkte til teorien om budrente-modellen. Vi antar også at miljøonder påvirker boligprisene i negativ forstand, noe som illustreres i figur 7 fra budrente-teori.

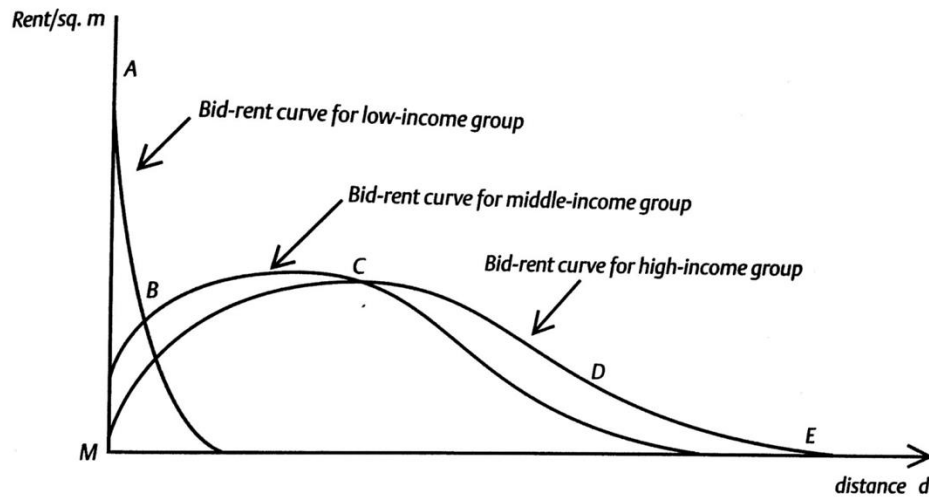
Budrente-modellen forklarer hvordan avstanden til et markedssenter,  $M$ , påvirker boligprisene, og her forutsetter modellen at det er en monosentrisk geografi. Ved hjelp av budrenteteori finner

man ut hva en bedrift eller husholdning er villig til å betale per enhet areal, for å være lokalisert innen en viss distanse fra M, og fortsatt oppnå et nivå av profitt (McCann, 2013, s. 114). Modellen er en forenklet versjon av virkeligheten, og har derfor en del forutsetninger. En av forutsetningene er det kun finnes ett markedssenter, der all sysselsetting finner sted. Modellen antar også at alt land er homogent og tilgjengeligheten til land er fast. En annen antagelse er at alt land er eid av fraværende utleiery og tildeles den som er villig til å betale høyest leie (McCann, 2013, s. 121).



Figur 6: Budrentekurver for en husholdning (McCann, 2013, s. 122).

McCann argumenterer for at det er en fallende budrentekurve for husholdninger med økt avstand til markedssenteret.  $\frac{-t}{S}$  angir helningen til husholdningens budrentekurve. Her er  $t$  transportkostnader og  $S$  er landareal, som øker ved økende avstand til M. En økende  $S$  er årsaken til at helningen på budrentekurven blir slakere med økt avstand. Dette fører til at budrentekurven for en husholdning blir konveks (McCann, 2013, s. 121). I figur 6 illustreres budrentekurver til en husholdning og nytten er høyest ved kurve  $BR_3$ , fordi prisen på land er lavest i den kurven (McCann, 2013, s. 122).



Figur 7: Effektene av miljøvariasjoner på budrentekurver (McCann, 2013, s. 127).

I figur 7 ser vi variasjonene i avstand til markedssenter og inntektsgrupper, når vi har miljøonder med i beregningen. Vi kan anta at det i markedssenteret er størst forurensing, både fra trafikk og lokale fabrikker. Lavinntektsgruppen bor nær markedssenteret, av den grunn at de ikke har råd til kostnader som forekommer ved å pendle over lange avstander. Middels- og høyinntektsgruppen kan være villige og har mulighet til å betale høyere leie for en rekke lokasjoner, og kan dermed skaffe land lenger borte fra sentrum. Årsaken til dette er at landets miljøkvalitet vil øke med avstand fra sentrum, det vil lide mindre av skadelige effekter av forurensing. Middels- og høyinntektsgruppens budrentekurver vil derfor ha positiv helning over en stor avstand, fordi disse inntektsgruppene vil være villige til å betale høyere leie for å unngå at miljøet har skader på grunn av forurensing (McCann, 2013, s. 127).

Når man når en viss avstand vil imidlertid de lokale effektene av forurensningen i sentrum være ubetydelige, og leieprisen med hensyn til avstand vil være som ved figur 6. Man må huske at forholdet mellom miljøkvalitet og den urbane leiegradienten er mer kompleks enn figur 7 (McCann, 2013, s. 127).

## 2.4 Miljøvariabler

Verdien av en boligeiendom avhenger både av dens fysiske egenskaper og egenskaper knyttet til lokalisering. Viktige fysiske egenskaper som antall rom, alder på eiendommen, bygningens tilstand og eiendommens størrelse er vanligvis lett målbare (Bourassa, Hoesli, & Sun, 2004).

Miljøvariabler kan deles inn i to hovedgrupper, miljøgoder og miljøonder, og er variabler som ikke direkte omsettes i markedet. Dersom miljøgoder eller miljøonder påvirker boligprisen på

enten positiv eller negativ måte, kapitaliseres godene i boligprisen. Produktet bolig har lang levetid, og markedsprisen angir nåverdi av de fremtidige tjenestene den kan gi (Freeman, Herriges, & Kling, 2014, s. 323). Forventninger om den fremtidige kvaliteten og den tilgjengelige mengden av disse miljøgodene eller miljøøndene påvirker altså kapitaliseringsverdien (Osland, 2016). Noen miljøgoder, som frisk luft og rent vann er en del av vårt forbruk og påvirker velferden direkte. Andre miljøgoder kan derimot anses som innsatsfaktorer i produksjonsprosesser, og bidrar dermed indirekte til varer og tjenester vi er vant til å skaffe oss i markedet (NOU 2012:16, 2012, ss. 47-48).

Miljøgoder er generelt sett kollektive goder, men de deler vanligvis andre egenskaper som kompliserer evalueringsproblemet. Dette kommer av at de kan ha flere likhetstrekk og er vanskelige å måle. Kollektive goder er goder som kan konsumeres av alle, og en persons konsum av godet påvirker ikke andres konsum av samme gode (Idsø, 2014). Miljøgoder utgjør et uklart sett med naturskapte trekk som deler noen eller alle av følgende fem kjennetegn.

(1) Forbrukeren kan få annen verdi enn bruksverdi fra eksistensen av et gode: verdien av et område i landskapet til en forbruker er ikke nødvendigvis begrenset til den gleden vedkommende får ved å besøke lokasjonen. (2) Endringer i tilbudet eller tilgjengeligheten til godene er sannsynligvis mer eller mindre irreversible: dersom et miljøgode er ødelagt, kan godet ikke erstattes av en perfekt erstatning. (3) Tilbudet og etterspørselen etter godene er i stor grad usikre. (4) Tilførselen av goder er begrenset og kan være unike, uten eventuelle erstatninger. (5) Et gode vil sannsynligvis være et naturlig trekk, og ikke menneskeskapt (Green, Tunstall, N'Jai, & Rogers , 1990).

Miljøgoder, som blant annet utsikt, nærhet til natur og parker, sol og stillhet, kan påvirke prisen på en bolig. Disse har positiv innvirkning på nytten til konsumentene. Også nabolaget rundt en bolig kan være en påvirkningsfaktor. Miljøønder som støy, forurensing og fare for flom og ras kan også påvirke boligprisene, men dette vil, i motsetning til miljøgoder, ha en negativ innvirkning på nytten (Osland, 2016). Både antallet miljøgoder og tilgjengeligheten til miljøgoder varierer fra sted til sted, likt som for miljøønder. Som tidligere nevnt er et miljøgode tilgjengelig for alle, og noens nytte av godet hindrer ikke nytten andre har av det. De fleste miljøgoder kan dermed ikke deles opp og gjøres tilgjengelig kun for enkeltindivider. På grunn av dette omsettes ikke miljøgoder direkte i markedet og har heller ikke markedspriser (Navrud, 2016).

Investeringsprosjekter kan påvirke mengden eller kvalitet på miljøgoder og miljøonder (Osland, 2016). Store prosjekter innen for eksempel samferdsel, energiproduksjon og energiforsyning vil ofte ha negative effekter på fellesgoder som blant annet vann- og luftkvalitet, stillhet, økosystemer og kulturminner. Motsatt vil investeringer for å verne naturområder ha positiv virkning på miljøgoder (Navrud, 2016).

Når geografisk nærhet til miljøvariablene er viktige for boligprisene, kan vi finne markedsverdien, altså estimere en markedspris på de lokale miljøvariablene (Osland, 2016). Man kan blant annet bruke hedonisk metode for å estimere verdien, også kalt implisitt pris, på miljøvariablene, noe vi kommer mer inn på senere. Miljøvariablene vi skal se nærmere på er utsikt, støynivå, strandlinje, solforhold, avstand til sjø, sjøfeste og turmuligheter.

#### 2.4.1 Utsikt

Utsikt blir sett på som et miljøgode som kan ha innvirkning på boligprisene, og boligkjøpere ønsker utsikt hovedsakelig av estetiske grunner (Bourassa, Hoesli, & Sun, 2004). Det å definere en variabel som utsikt er vanskelig. En av grunnene til at det er vanskelig å definere er fordi hvert enkelt menneske har ulike preferanser på hva de liker og ikke. Selv om noen liker å ha utsikt til havet, kan andre like å ha utsikt over et byområde. Utsikt har tidligere blitt definert som utsikt til sjø, innsjø, fjell eller skog, og en slik definisjon er kanskje den folk flest tenker på (Sirmans, Machperson, & Zietz, 2005).

Selv om god utsikt ofte er assosiert med enkel tilgang til naturen, bør de to effektene tydelig skilles. Tilgang til naturen tillater forskjellige fysiske aktiviteter og kan måles med avstandsvariabler. For eksempel avstand til sjø eller skogsområder. Utsikt er imidlertid vanskeligere å måle og definere. Nye studier bruker en rekke dummy variabler for å måle forskjellige typer og kvaliteter av utsikt. I de fleste empiriske studier er resultatet at utsikt har positive effekter på verdien av en bolig, selv om andre studier observert ubetydelige effekter. Det er flere studier som har definert utsikt dårlig, som kan komme av blant annet målefeil eller av en liten prøvestørrelse, og dermed kan få feil resultater (Bourassa, Hoesli, & Sun, 2004).

Bourassa et al. (2004) definerer to typer utsikt, over vann og land. Det defineres også et utsiktsomfang på tre nivåer: bred, middels og smal, som vil si hvor vid utsikt en eiendom har. For eiendommer med havutsikt, kan man også beregne avstanden til kysten. Lokaliseringen til

en eiendom vil ha påvirkning på boligprisen, som blant annet det generelle utseendet av nabolaget som boligen tilhører. Derfor argumenteres det for at utsikt ikke bør begrenses til naturlige trekk som hav, skog og fjell, men bør også inkludere fordelene forbundet med et attraktivt landskap i nærheten av en eiendom og med kvaliteten på omkringliggende bygninger.

#### 2.4.2 Støynivå

Støy er en uønsket lyd. Det deles ofte inn i to typer støy. Den første er irriterende støy, som kan være et ventilasjonsanlegg eller viften i en PC. Den andre er skadelig støy som er målt over 80 dB (Arbeidstilsynet, u.d.). Man skiller også mellom støyplager i og utenfor boligen. Støy i boligen kan være tv, radio og de som er nevnt for irriterende støy, mens støy utenfor boligen er blant annet trafikk- og industristøy (Amundsen & Aasvang, 2006, s. 1). Navrud og Strand (2011) finner at støy har en signifikant negativ virkning på boligpriser. En bolig vil få en gjennomsnittlig prisreduksjon på nesten 7 % dersom støynivået øker fra 55 til 70 desibel. Reduksjonen varierer derimot blant annet for typen bolig.

#### 2.4.3 Avstand til sjø

Det er flere grunner til at man ønsker en bolig nær kysten, blant annet at det gir nytte i form av rekreasjonsverdi gjennom bruksområder som bading, svømming, soling og surfing (Conroy & Milosch, 2011). Conroy og Milosch (2011) estimerte effekten av boligprisene av å være lokalisert nær kysten og fant at nærhet til kysten har en stor og positiv effekt på verdien på en bolig. Deres hedoniske regresjonsmodell illustrerte at en økning på en amerikansk mil fra kysten vil føre til en reduksjon i verdi på 8 680 amerikanske dollar. Reduksjonen forblir den samme over en lengre avstand, men den blir lengre enn seks amerikanske mil, vil avstanden bli ubetydelig for prisen.

#### 2.4.4 Sjøfeste

Som Conroy og Milosch (2011) forklarer, er det å ha en bolig med nærhet til kysten positivt for verdien. Dersom en bolig har sjøfeste, vil avstanden til sjø være tilnærmet lik null. Vi kan dermed hevde at å ha sjøfeste vil øke boligens verdi.

#### 2.4.5 Strandlinje

Strandlinje gir, i liket med sjøfeste og nærhet til kysten, muligheter for bading, svømming, soling og surfing, som igjen gir nytte (Conroy & Milosch, 2011). Rinehart og Pompe (1994)

finner at en bredere strand øker en boligs verdi. Selv om de analyserte nærhet til strand, vil verdien for en bolig med egen strand også øke. Det skjer av den grunn at avstand til strand vil være lik null.

#### 2.4.6 Solforhold

Solforhold er målt i antall timer en bolig har sol i løpet av en dag. I *Valuing Sunshine* (2017) har de gjennom et utvalg på 5 000 solgte boliger funnet at gode solforhold kan øke prisen på en bolig. For hver ekstra time med sollys på en eiendom i løpet av en dag, øker prisen med 2,4 % av totalverdien. De konkluderte med at økningen i boligverdi kan være både høyere og lavere avhengig av klima, topografi, bystørrelse og inntekt (Fleming, Grimes, Lebreton, Maré, & Nunns, 2017).

#### 2.4.7 Turmuligheter

Turmuligheter kan være fjell- eller skogstur, eller tur på fortauet langs veien. Mulighetene varierer for ulike områder, og i bysentrum kan det for eksempel være lenger til skog og fjell, enn for boliger som ligger ute på landet. Tyrväinen (1997) finner at avstand til skogsområder som er i nærheten av en bolig er signifikant, og har en positiv påvirkning på boligpriser. Dersom mengden skogsområder økte i boligområdet, samt nærhet til elv og rekreasjonsområde, ville også prisene på leiligheter øke.



## 3 Metode

### 3.1 Regresjonsanalyse

#### 3.1.1 Enkel lineær regresjonsanalyse

Regresjonsanalyse er en metode for å undersøke hvordan gjennomsnittsverdien på en avhengig variabel varierer med en eller flere uavhengige variabler. I forbindelse med denne oppgavens problemstilling kan den enkle lineære regresjonsanalysen benyttes til å beskrive sammenhengen mellom en avhengig variabel (pris på bolig, P) og en uavhengig variabel (X) som en lineær sammenheng. En lineær sammenheng kan skrives som følgende:

$$P = b_0 + b_1X_1$$

hvor  $b_0$  angir verdien på den avhengige variabelen når den uavhengige variabelen har en verdi lik 0 og  $b_1$  beskriver hvor mye den avhengige variabelen endres dersom den uavhengige øker med en måleenhet.  $X_1$  er den uavhengige variabelen (Johannessen, Christoffersen, & Tufte, 2011, ss. 335-337).

Ifølge Johannessen et al. (2011, ss. 254-257) er det fire forutsetninger for en god regresjonsanalyse:

1. Forutsetninger om modellspesifikasjon.

Alle uavhengige variabler må være relevante, og alle relevante variabler må være inkludert. Dersom en relevant variabel er utelatt og den korrelerer med de inkluderte, vil vi få skjevhet i de parameterne som estimeres.

2. Forutsetninger om residualer.

Residualer ( $\epsilon$ ) er avvikene mellom regresjonslinjen og de observerte verdiene. Det er fire forutsetninger om residualer; restleddet skal være normalfordelt, ha konstant varians, det skal ikke være korrelert med noen av forklaringsvariablene i regresjonsmodellen og restleddet fra en observasjon skal ikke være korrelert med restleddet fra en annen observasjon.

3. Fravær av multikollinearitet.

Det vil si at det ikke kan være en perfekt lineær sammenheng mellom to eller flere av de uavhengige variablene.

#### 4. Fravær av målefeil.

Variablene må være målt uten målefeil. Det er mindre alvorlig med målefeil i avhengig variabel enn i uavhengige variabler. Målefeil i avhengig variabel vil påvirke hypotesetestingen og føre til lavere forklart varians ( $R^2$ ). Dersom det er målefeil i en uavhengig variabel, for eksempel utsikt, vil det øke sannsynligheten for gale estimater og gale hypotesetester.

### 3.1.2 Minste kvadraters metode

Fugleberg & Kristianslund (1995) lister opp fem antagelser om en lineær regresjonsmodell.

1. *Eksistens*: Forventningen til  $P$  avhenger av verdien på  $X$ ,  $E(P|X)$ .
2. *Linearitet*: Forventningsverdien til  $Y$ ,  $E(P|X)$ , er en lineær funksjon av  $X$ .
3. *Uavhengighet*:  $\varepsilon$ -verdiene er statistisk uavhengige av hverandre.
4. *Homoskedastisitet*: Variansen til  $\varepsilon$  er den samme for alle  $X$ .
5. *Normalfordeling*: For enhver fast verdi av  $X$  har  $\varepsilon$  en normal fordeling.

Antagelsene 1 til 4 må være oppfylt for å få best mulig estimater og for å kunne beregne regresjonskoeffisienten  $\beta$  benyttes minste kvadraters metode. Antagelse 5 om normalitet får implikasjoner for fordelingen til testobservatoren ved hypotesetesting og konfidensintervall for parameterne i regresjonsmodellen. Her er det viktig at ved moderate avvik fra antagelsen om normalfordeling vil konklusjonene som baseres på denne antagelsen fremdeles ha tilnærmet gyldighet. Når antagelsene 1-5 er oppfylt, er den avhengige variabelen,  $P$ , for hver uavhengig variabel,  $X$ , normalfordelt og den har samme varians for alle  $X$  (Fugleberg & Kristianslund, 1995, ss. 34-35).

Minste kvadraters metode sikrer at summen av de vertikale avstandene mellom punktene i analysen og selve regresjonslinjen er minst mulig. Siden noen av punktene ligger over linjen, mens andre punkter ligger under linjen sier man at de nuller hverandre ut. For å unngå at de nuller hverandre ut kvadreres de først for deretter å summeres (Thrane, 2017, s. 31). Vi betegner  $\hat{\beta}$  som estimert parameter. For å kunne regne ut  $\hat{\beta}$  brukes følgende formel:

$$\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X})(P_i - \bar{P})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Etter å ha regnet ut kan en gå videre til å regne ut  $\hat{\alpha}$ . Formelen blir da

$$\hat{\alpha} = \bar{P} - \hat{\beta}\bar{X}$$

Formlene som brukes her gir oss estimatorene for  $\beta$  og  $\alpha$ , som kalles for minste kvadraters metode-estimatorene (Fugleberg & Kristianslund, 1995, s. 40).

### 3.1.3 Forventningsrett og skjevhet

Fugleberg og Kristianslund (1995) beskriver en estimator,  $\hat{\theta}$  for  $\theta$ , som forventningsrett dersom  $E(\hat{\theta}) = \theta$ . Det vil si at en estimator er forventningsrett, dersom gjennomsnittet av estimatene går mot  $\theta$ . Hvis en estimator er forventningsrett, har sannsynligheten en forventet verdi lik parameteren den antas å estimere. Dersom en relevant variabel er utelatt fra modellen, og den korrelerer med en variabel som er inkludert, vil parameteren til den inkluderte variabelen ikke bli forventningsrett.

En måte å sammenligne estimater på som ikke nødvendigvis er forventningsrette, er å beregne den gjennomsnittlige kvadratiske feilen eller MSE (mean squared error). Hvis  $\hat{\theta}$  er et estimat av  $\theta$ , er MSE av  $\hat{\theta}$  definert som (Wooldridge, 2016, s. 680);

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

MSE måler hvor langt, i gjennomsnitt, en estimator avviker fra  $\theta$ . Det kan vises gjennom (Wooldridge, 2016, s. 680):

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + Skjevhet^2$$

hvor  $Skjevhet^2 = (E(\hat{\theta}) - \theta)^2$  (Wooldridge, 2016, s. 677).

MSE avhenger av variansen og skjevheten (Wooldridge, 2016, s. 680), skjevheten er altså et mål for hvordan forventningen til estimatoren avviker fra den sanne parameterverdien (Fugleberg & Kristianslund, 1995, s. 31). Skjevhet kan oppstå ved å utelate variabler som korrelerer med en annen fra estimeringen, og koeffisienten vil da bli betydelig høyere eller lavere enn forventet. Man vil rette opp skjevheten ved å inkludere denne variabelen. En annen grunn til skjevhet er målefeil, som også vil påvirke koeffisientene (Stock & Watson, 2007).

Tabell 1: Oppsummering av skjevhet for  $\beta_1$  når  $x_2$  er utelatt (Wooldridge, 2016, s. 79)

	$Korr(x_1, x_2) > 0$	$Korr(x_1, x_2) < 0$
$\beta_2 > 0$	Positiv skjevhet	Negativ skjevhet
$\beta_2 < 0$	Negativ skjevhet	Positiv skjevhet

Tabell 1 beskriver om skjevheten til  $\beta_1$  når en annen uavhengig variabel  $\beta_2$  er utelatt. For eksempel, skjevheten til  $\beta_1$  vil være positiv dersom  $\beta_2 > 0$ , at den har en positiv påvirkning for modellen, og at  $x_1$  og  $x_2$  har en positiv korrelasjon (Wooldridge, 2016, s. 79).

### 3.1.4 Konsistens

Konsistens er også en viktig egenskap ved estimatorer. Wooldridge (2016, s. 681) forklarer at når størrelsen på prøven øker mot uendelig vil man kunne si hvor langt estimatoren er fra parameteren den antas å estimere. Konsistens handler om at estimatorene med høy sannsynlighet skal gi et estimat som er nær den sanne verdien av parameteren når vi øker  $n$  (Fugleberg & Kristianslund, 1995, s. 31). La  $\hat{\theta}$  være en estimator for  $\theta$  som baserer seg på stikkprøver  $Y_1, Y_2, \dots, Y_n$  av størrelsen  $n$ . Når,  $\hat{\theta}$ , er en konsistens estimator av  $\theta$  for alle  $\varepsilon > 0$

Det gir følgende formler:

$$P(|\hat{\theta} - \theta| > \varepsilon) \rightarrow 0 \text{ når } n \rightarrow \infty \quad (\text{Wooldridge, 2016, s. 681})$$

og

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1 \quad (\text{Fugleberg \& Kristianslund, 1995, s. 31})$$

Dersom  $\hat{\theta}$  er konsistent med  $\theta$  kalles det for en sannsynlighetsgrense for  $\hat{\theta}$ , også skrevet som  $plim(\hat{\theta})$  (Wooldridge, 2016, s. 681). Et viktig element i konsistens er at når størrelsen ( $n$ ) øker blir fordelingen av  $\hat{\theta}_n$  lik den sanne verdien til  $\theta$  (Fugleberg & Kristianslund, 1995, s. 32).

### 3.1.5 Prediksjon

Prediksjon er en del av regresjonslinjen, og en bruker regresjonslinjen til å predikere den forventede verdien på den avhengige variabelen når den uavhengige variabelen har en bestemt verdi (Johannessen, Christoffersen, & Tufte, 2011, s. 342). Hvis det lineære forholdet er perfekt,

vil det si at alle verdier for X gir en eksakt verdi for P, anses sammenhengen å være deterministisk. All variasjon i P forklares ved variasjon i X. Siden disse tilfellene er sjeldne benytter man et feilledd eller en residual,  $\varepsilon$ , slik at man tar høyde for at feilledd er en mulighet (Walope, Myers, Myers, & Ye, 2012, ss. 389-391). For å regne ut hva residualene er, trekkes predikert verdi for avhengig variabel fra observert verdi.  $\varepsilon$  vil være residualen for respondentene,  $P_i$  vil være den observerte verdien for den avhengige variabelen for respondentene, mens  $\hat{P}_i$  vil være den predikerte verdien på den avhengige variabelen etter å ha brukt regresjonslinjen (Johannessen, Christoffersen, & Tufte, 2011, ss. 342-344). Formelen ser dermed slik ut:

$$\varepsilon = P_i - \hat{P}_i \text{ (Johannessen, Christoffersen, \& Tufte, 2011, ss. 342-344).}$$

Videre erstatter vi konstantene i den lineære sammenhengen  $b_0$  og  $b_1$  med ukjente parameter  $\alpha$  og  $\beta$ . I tillegg må vi legge til residualen, slik at feilledd inkluderes i formelen. Formelen blir dermed (Fugleberg & Kristianslund, 1995, ss. 33-34):

$$P = \alpha + \beta X + \varepsilon$$

Det antas at sannsynligheten for at en gitt X er over linjen  $P = \alpha + \beta X$  er like stor som at en gitt X er under linjen. På grunnlag av det kan vi deretter anta at  $E(\varepsilon) = 0$ , som igjen gir at  $E(P|X) = \alpha + \beta X$ . Hvordan observasjoner av X-P varierer på linjen spesifiseres ved antagelser på residualen. Det ligger en antagelse om at P avhenger av X som en lineær hovedtendens, altså  $E(P|X) = \alpha + \beta X$ . Denne delen kalles også for den "sanne" underliggende linjen, som man finner ved å estimere  $\alpha$  og  $\beta$  (Fugleberg & Kristianslund, 1995, ss. 33-34).

### 3.1.6 Best Linear Unbiased Estimator (BLUE)

Rundt mange problemstillinger er det av interesse å estimere lineære kombinasjoner av  $\beta_0, \dots, \beta_{n-1}$ , som er  $t'\beta$ , hvor t er hvilken som helst  $n \times 1$  vektor av kjente kombinasjoner, men den kan ikke være null. Best Linear Unbiased Estimator (BLUE) forklares slik; den beste lineære forventningsrette estimatoren for  $t'\beta$  er

1. en lineær funksjon av den observerte vektoren Y, det vil si en funksjon av formen  $a'Y + a_0$  hvor a er en  $n \times 1$ -vektor av konstanter og  $a_0$  er en skala og
2. den objektive estimatoren for  $t'\beta$  med den minste variansen (Moser, 1996, s. 86).

### 3.1.7 Korrelasjon

Korrelasjon er et mål for samvariasjon mellom variablene. Pearsons  $r$  (Pearsons produktmomentkorrelasjon) er et mye brukt korrelasjonsmål og angir hvor sterk lineær sammenheng det er mellom to variabler. Det finnes to typer korrelasjon, positiv og negativ. Dersom det er positiv korrelasjon vil høye verdier på en variabel bety at det også er høye verdier på den andre variabelen, og omvendt med lave verdier. Det er negativ korrelasjon dersom høye verdier på en variabel gir lave verdier på den andre variabelen. Pearsons  $r$  angir både typen og styrken av samvariasjon. Pearsons  $r$  er en standardisert koeffisient som varierer mellom  $-1$  og  $+1$ . Ved korrelasjon på  $1$  og  $-1$  vil det være en veldig høy korrelasjon, mens det ved en korrelasjon på  $0$  vil det ikke eksistere en lineær sammenheng (Johannessen, Christoffersen, & Tufte, 2011, ss. 320-322). Styrken til Pearsons  $r$  kan beskrives slik:

Tabell 2: Tommelfingerregel for korrelasjon (Johannessen, Christoffersen, & Tufte, 2011, s. 322)

R	Korrelasjon
0,00 - 0,19	Veldig svak
0,20 - 0,39	Svak
0,40 - 0,69	Moderat
0,70 - 0,89	Høy
0,90 - 1,00	Meget høy

### 3.1.8 Multikollinearitet

Johannessen et al. (2011) argumenterer for at multikollinearitet kan oppstå hvis det er en perfekt eller tilnærmet perfekt lineær sammenheng mellom to eller flere av de uavhengige variablene. Det kan medføre at de uavhengige variablene blir ustabile og at de mister presisjon (Allison, 1999) (Kennedy, 2003). Hvis det er en høy korrelasjon mellom de uavhengige variablene blir det vanskelig for regresjonsanalysen å fastslå hvilken uavhengig variabel variasjon i modellen kommer fra (Thrane, 2017, s. 90).

Variance inflation factor (VIF) brukes til å finne multikollinearitet i en regresjonsanalyse. VIF er en numerisk verdi som strekker seg fra én og oppover. Den viser hvilke variabler som korrelerer og hvor mye de korrelerer (Statistics How To, 2015).

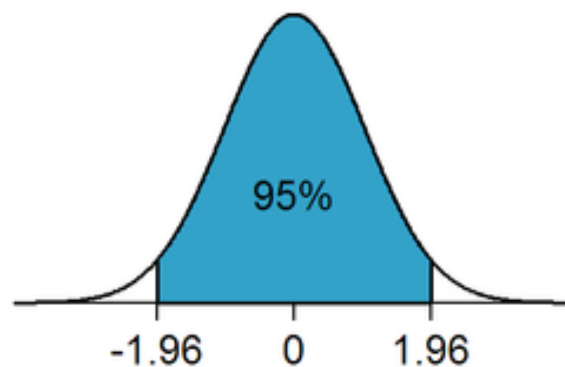
En god huskeregel for å tyde VIF-verdier er:

- 1 = ikke korrelert
- Mellom 1 og 5 = moderat korrelasjon
- Høyere enn 5 = sterk korrelasjon (Statistics How To, 2015)

Det vil alltid være diskusjon rundt hvor stor VIF-verdien må være for at den skal skape et problem. Vanligvis sies det at verdien må overstige 10 for å skape et problem, mens andre, mer konservative forskere, fremmer en verdi på 2,5 (Statistics How To, 2015).

### 3.1.9 T-verdi

Når man gjennomfører statistisk signifikanstesting, finner man ut om variablene man bruker er innenfor nullhypotesens gyldighetsområde. Ligger variabelen utenfor gyldighetsområde (figur 8) forkaster man nullhypotesen og beholder alternativ hypotese. Det man finner er t-verdien og p-verdien til variabelen. Det er mest vanlig å bruke en t-verdi på 1,96 som vil si at man bruker et signifikansnivå eller p-verdi på 5 %. Har variabelen en t-verdi på 1.96 kan man med 95 % sikkerhet si at det er korrekt å forkaste nullhypotesen (Thrane, 2017, s. 78).



Figur 8: 95% av arealet under normalfordeling ligger innenfor 1,96 standardavvik fra gjennomsnittet (Wikipedia, 2020).

### 3.1.10 Intern validitet

Validitet handler om i hvilken grad man kan trekke gyldige konklusjoner ut fra resultatene i en studie. Intern validitet er muligheten en studie gir til at resultatene kan forklares gjennom den bestemte hypotesen, og for at den interne validiteten skal være høy, må man ha kontroll over eventuelle bias (Grønmo, 2018).

Det er fem trusler mot den interne validiteten av regresjonsstudier, og vi skal fokusere på truslene 1-3.

1. Utelatt variabel skjevhet.
2. Feil funksjonsform.
3. Feil-i-variabelskjevhet.
4. Prøvevalg-skjevhet.
5. Samtidig kausalitet skjevhet (Stock & Watson, 2007, ss. 313-314).

Feilspesifisering i modellarbeidet kan føre til at vi får feil resultater. Den første trusselen til intern validitet av regresjonsstudier er utelatt variabel skjevhet, som oppstår hvis en utelatt variabel både er en avgjørende faktor for vår avhengige variabel (Y) og korrelerer med minst én inkludert variabel. Dette kan for eksempel skje dersom vi utelater solforhold i våre analyser og dette korrelerer med utsikt. Det finnes flere løsninger på utelatt variabel skjevhet, og en er så enkel som å inkludere den utelatte variabelen, dersom den kan måles (Stock & Watson, 2007, ss. 316-317).

Den andre trusselen er feil funksjonsform og denne delen er for omfattende for oss å se på. Vi kan likevel forklare kort hvordan det oppstår. Feil funksjonsform oppstår åpenbart hvis funksjonsformen ikke stemmer (Stock & Watson, 2007, ss. 318-319). Dersom et viktig ledd som korrelerer med avhengig variabel og flere uavhengige variabler og blir utelatt vil modellen bli feil (Berge, 2004, s. 3).

Den tredje og siste trusselen vi skal se på er feil-i-variabelskjevhet. Økonomiske data har i virkeligheten ofte målefeil. Dette kan for eksempel være feil i dataregistrering i administrative data, huske-feil i undersøkelser, tvetydige spørsmål og intensjonelt falske respons problemer med undersøkelser. Det er flere løsninger på feil-i-variabler skjevhet. Den første er å få bedre data, noe som kan være veldig vanskelig. Man kan også utvikle en spesifikk modell for å prosessere målefeil. Dette er bare mulig hvis det er mye kunnskap om bakgrunnen til målefeilen (Stock & Watson, 2007, ss. 319-322).



## 3.2 Dataprogrammer

### 3.2.1 Excel

I gjennomførelsen av våre analyser ble det blant annet brukt Microsoft Excel. Excel er en del av Microsoft Office pakken. Det er et regneark hvor man enkelt kan utføre mange ulike typer tester, kalkulasjoner og lignende. Planen var opprinnelig å bruke Excel til alle analysene våre, men selv om Excel er et nyttig verktøy, har det også sine begrensninger. En av begrensningene er at man ikke kan ha mer enn 16 variabler i regresjonsanalysen. Vi var nødt til å velge et annet dataprogram for å gjennomføre testene og analysene ønsket å utføre. Derimot er Excel brukt til all behandling av datasett og klargjøring av disse, samt til tabeller, figurer og deskriptiv analyse. Gjennom tre år på Høgskulen på Vestlandet har vi fått god kjennskap til dataprogrammet Excel.

### 3.2.2 RStudio

For å kunne gjennomføre analysene ble RStudio benyttet. Vi gjennomførte regresjonsanalyser, VIF-test og Breusch-Pagan-test. RStudio er et integrert utviklingsmiljø for R. Det er et dataprogram som støtter direkte koding, samt verktøy for plotting, feilsøking og håndtering av arbeidsområde (RStudio, u.d.). Det ble en bratt læringskurve, da vi aldri har anvendt dette dataprogrammet tidligere.

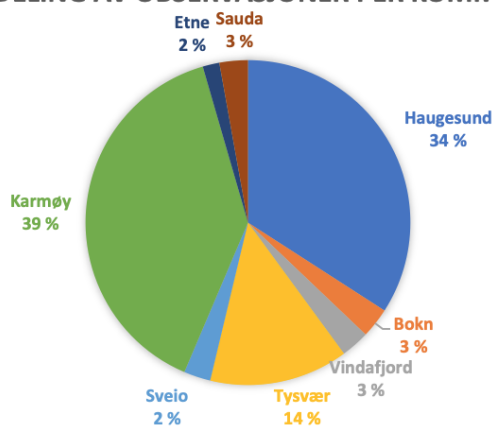
## 3.3 Datagrunnlag for analysen

### 3.3.1 Datasett

Datasettet som er benyttet i oppgaven ble som poengtert innledningsvis utarbeidet som et arbeidskrav i emnet «By- og regionaløkonomi». Dette har blitt videre arbeidet med først av Anglevik og Øvretveit (2018), og deretter Hovland, Kvalvågnes og Kvalevåg (2019). I til Anglevik og Øvretveit (2018) var det 618 observasjoner med 79 variabler, mens det i Hovland, Kvalvågnes og Kvalevåg (2019) var 601 observasjoner med 62 variabler. Datasettet fra Hovland et al. (2019) inneholdt færre variabler, da dette datasettet var blitt prosessert, hvor variabler med store mangler knyttet til observasjonene var fjernet. For at vi skulle kunne bruke begge datasettene til å gjennomføre analysen, var vi nødt til å samkjøre datasettene. Det gjorde at vi fikk tilgang til alle variablene, og samlet deretter de variablene vi ville se nærmere på i et eget regneark. Det har vært store kvalitetsforskjeller på variablene i datasettet. Det har for noen variabler vært mange observasjoner, og dataene som er hentet inn er korrekte data. For eksempel er variabelen *primaerrom* en variabel det er enkelt å samle data inn på. For andre

variabler har det derimot vært mangelfulle datasett, og dataene som er hentet inn er gjort gjennom en subjektiv vurdering. Det er tenkelig at resultatene vil bære preg av at datasettet har mangler.

#### FORDELING AV OBSERVASJONER PER KOMMUNE



Figur 9: Prosentvis fordeling av antall observasjoner i hver kommune.

Figur 9 er en illustrasjon av hvordan observasjonene er fordelt i de ulike kommunene. Over halvparten av observasjonene i vårt datasett er fra kun Haugesund og Karmøy. Tre av åtte kommuner utgjør totalt 87 % av alle observasjonene. Det derfor er stor differanse mellom antall observasjoner i kommunene. Vi har for eksempel høyest antall observasjoner fra Karmøy kommune, med 235 observasjoner, men kun 10 observasjoner fra Etne kommune, som er kommunen med færrest observasjoner.

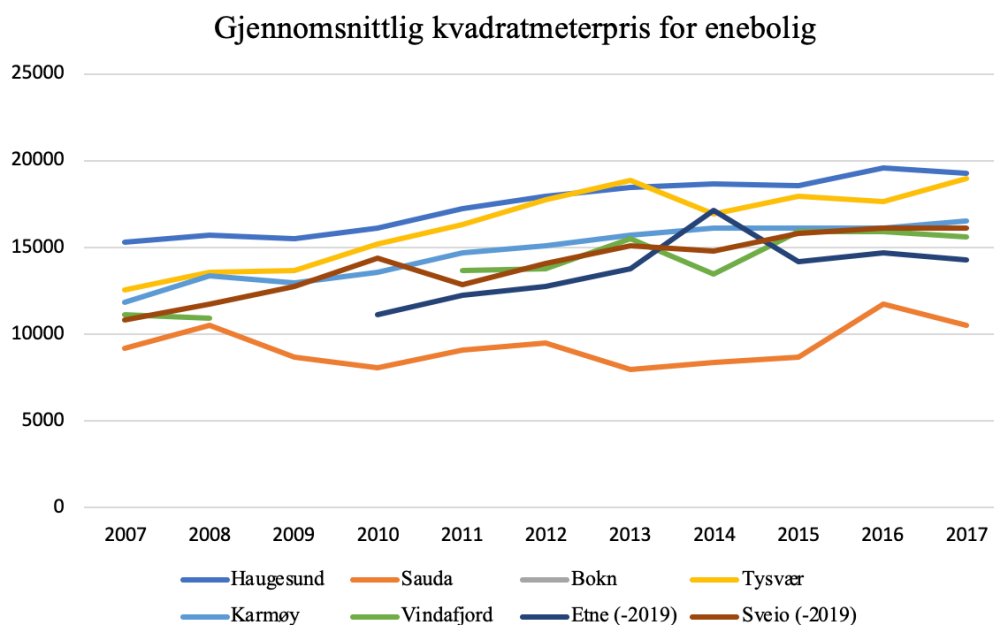
Tabell 3: Gjennomsnittlig kvadratmeterpris i datasettet.

Kommune	Gjennomsnittlig Kvm-pris
<i>Haugesund</i>	20 269,53
<i>Karmøy</i>	16 419,43
<i>Tysvær</i>	18 637,17
<i>Bokn</i>	14 657,37
<i>Vindafjord</i>	18 857,54
<i>Sveio</i>	18 135,99
<i>Etne</i>	16 681,92
<i>Sauda</i>	9 884,10

I tabell 3 er det beregnet gjennomsnittlig kvadratmeterpris for alle boligene i vårt datasett, inndelt etter kommune. For å finne gjennomsnittlig kvadratmeterpris må vi dividere salgsprisen på primærrommet til en boenhet. Deretter summeres kvadratmeterprisene for hver boenhet i gjeldende kommune, og til slutt dividere det på antall observasjoner. Vi har funnet total gjennomsnittlig kvadratmeterpris for hver kommune og ikke differensiert for boligtype, da dette ikke var vårt fokus. Tolv observasjoner er utelatt fra beregningen, på grunn av at vi mangler informasjon om variabelen primærrom.

Med utgangspunkt i budrente-modellen forventer vi høyest gjennomsnittlig kvadratmeterpris i Haugesund, og dette blir bekreftet med våre data. Vi observerer at det ikke er store forskjeller i pris mellom de fleste kommunene, men at prisen mer eller mindre faller desto lenger reisetid man har fra bolig til regionsenteret. Sauda har lengst reisetid til Haugesund, og har mye lavere kvadratmeterpris enn de resterende kommunene. Sauda har kun 9 884 kr/kvm, noe som er under halvparten av kvadratmeterprisen i Haugesund kommune.

Som tidligere illustrert, hadde vi ikke mange observasjoner fra de fem nederste kommunene i tabellen ovenfor, og derfor kan det være feil i resultatet. Det er også andre faktorer som kan føre til variasjoner i prisen, for eksempel boligalder og boligtype. Det er blant annet inkludert både andels- og selveierboliger i datasettet, som vil gi større forskjeller i kvadratmeterprisen.



Figur 10: Gjennomsnittlig kvadratmeterpris for enebolig per kommune (Statistisk Sentralbyrå, u.d.).

Figur 10 illustrerer utviklingen i gjennomsnittlig kvadratmeterpris for eneboliger i de ulike kommunene. Vi valgte å se på eneboliger i dataene fra SSB, fordi dette ga det beste bildet av vår region, selv om det mangler data for alle år i Bokn kommune, og for to år i Etne og Vindafjord. Vi kunne også valgt å studere blokkleiligheter eller småhus, men disse hadde kun informasjon om tre og to kommuner. Det viktig å poengtere at vi ikke differensierte for type bolig da vi beregnet kvadratmeterpris, men over halvparten av observasjonene i vårt utvalg er eneboliger.

Dersom vi sammenligner kvadratmeterprisene fra vårt utvalg med figur 10 kan vi observere likhetstrekk til populasjonen. I figur 10 kan vi observere at for de fleste årene har Haugesund høyest kvadratmeterpris, og Sauda lavest, noe som er likt for våre data. Gjennomsnittlig i våre observasjoner har både Sveio, Etne og Vindafjord høyere kvadratmeterpris enn Karmøy, noe som ikke er tilfellet når man observerer hele populasjonen. Hvis vi studerer verdiene på kvadratmeterprisen, og ikke rekkefølgen på de ulike kommunene, kan vi observere at gjennomsnittlig er kvadratmeterprisene i de ulike kommunene lavere fra SSB. Dette kommer nok av at vi har inkludert alle boligtyper, og størrelsen på utvalget.

### 3.3.2 Basemodellen

Vi har tatt utgangspunkt i den endelige modellen til Anglevik og Øvretveit (2018), som også er kalt "basemodellen". Denne modellen ble utarbeidet gjennom flere modeller og variabler, men de endte opp med en modell som består av variablene; *avsreg*, *prosarb60*, *avskom*, *primaerrom*, *garasje*, *omsetningsaar*, *naust*, *boligalder* og *utsikt*. Videre har også Hovland, Kvalvågnes og Kvalevåg (2019) tatt utgangspunkt i den endelige modellen. De innhentet data fra flere variabler for å undersøke om disse påvirket boligprisene og om forklaringskraften til modellen økte. I deres oppgave undersøkte de variablene; *boligtype*, *bruksareal*, *renovering*, *anneks*, *soverom*, *solforhold*, *sjøfeste*, *eiendomsskattesats*, *bunnfradrag*, *vann*, *avløp*, *feiing*, og *avfall*.

Tabell 4: Basemodellens variabler (Anglevik &amp; Øvretveit, 2018).

Variabel	Forklaring
<i>Avsreg</i>	Avstand til regionsenter målt i minutter med bil
<i>Prosarb60</i>	Prosentandelen av arbeidsplasser i regionen som kan nås innen 60 minutter med bil
<i>Avskom</i>	Avstand til kommunesenter målt i minutter med bil
<i>Primaerrom</i>	Oppholdsrom i boligen; stue, kjøkken, soverom og bad
<i>Garasje</i>	Kodet med 0 og 1
<i>Omsetningsaar</i>	Hvilket år boligen ble solgt
<i>Naust</i>	Kodet med 0 og 1
<i>Boligalder</i>	Alderen på boligen det året den ble solgt
<i>Utsikt</i>	Rangert mellom 0,0 og 1,0

For variablene som er brukt i basemodellen og hvor mange observasjoner i prosent disse har, ser vi at de har en relativt høy prosentandel av observerte verdier. Dette vil si at de fleste observasjonene i vårt datasett har informasjon om alle variablene i basemodellen. For de andre variablene vi har valgt å se på er det i gjennomsnitt en lavere prosentandel av observerte verdier. *Garasje* og *Naust* er kodet med 0 og 1. Koden 1 brukes dersom boligen har garasje/naust, og 0 brukes dersom den ikke har.

Tabell 5: Region- og kommunesenter (Anglevik &amp; Øvretveit, 2018).

Kommune	Regionsenter	Kommunesenter
<i>Haugesund</i>	Haugesund	Haugesund sentrum
<i>Karmøy</i>	Haugesund	Kopervik
<i>Tysvær</i>	Haugesund	Aksdal
<i>Bokn</i>	Haugesund	Føresvik
<i>Vindafjord</i>	Haugesund	Sveio sentrum
<i>Sveio</i>	Haugesund	Ølen
<i>Etne</i>	Haugesund	Etne sentrum
<i>Sauda</i>	Haugesund	Sauda sentrum

I basemodellen måles avstand til både regionsenter og kommunesenter for de ulike observasjonene. Som tidligere sagt har alle kommunene på Haugalandet samme regionsenter. Haugesund er tett befolket og en stor del næringsliv er lokalisert i og rundt kommunen, som har

ført til at Haugesund blir det naturlige sentrum i regionen. Alle kommunene har derimot egne kommunesenter, illustrert i tabell 5, som er administrasjonssenteret i hver kommune.

### 3.3.3 Variabler

Tabell 6: Tilgjengelighetsvariabler (Anglevik & Øvretveit, 2018).

Variabel	Forklaring
Arb	Antall arbeidsplasser registrert i postnummeret
Syss	Antall sysselsatte med registrert bosted i postnummeret
Arb/Syss	Antall arbeidsplasser sysselsatt i postnummeret
Arbx	Antall arbeidsplasser innenfor x antall minutter med bil ( $x = 45$ og $x = 60$ )
Tidarbx	Vektet gjennomsnittlig reisetid til arbeidsplasser som nås innenfor x antall minutter med bil ( $x = 45$ og $x = 60$ )
Prosarbx	Prosentandelen av arbeidsplasser i regionen som kan nås innenfor x antall minutter med bil ( $x = 45$ og $x = 60$ )

Variablene *arbx*, *tidarbx* og *prosarbx* inneholder opprinnelig verdier for  $x = 45$ ,  $x = 60$  og  $x = 90$ . I Anglevik og Øvretveit (2018) argumenteres det for at NOU 2011:3 (2011) definerer et arbeidsmarked som “*det samlede tilbud av, og etterspørsel etter, arbeidskraft 60 minutter med bil*” og at “*den gjennomsnittlige arbeidsreisen i 2009 var i overkant av 21 minutter*”. På grunnlag av dette velger de å ta bort  $x = 90$ . Selv om det finnes personer på Haugalandet som har over 60 minutt til jobben, kommer vi frem til samme vurdering som Anglevik og Øvretveit (2018) om å ta med variabelen *prosarb60*. Fra tabell 6 bruker vi kun variabelen *prosarb60* i våre analysen. Vi tar utgangspunkt i basemodellen og skal analysere de variablene som er nevnt i tabell 7.

Tabell 7: Variablene som undersøkes.

Variabel	Forklaring
<i>Stoynivaa</i>	Kodet med 0 og 1
<i>Avssjo</i>	Avstand til sjø i meter
<i>Sjofeste</i>	Kodet med 0 og 1
<i>Solforhold</i>	Rangert mellom 0,0 og 1,0
<i>Strandlinje</i>	Antall meter i strandlinje
<i>Turmuligheter</i>	Rangert mellom 0 og 10

Variablene vi undersøker i denne oppgaven er *utsikt*, *støynivaa*, *avssjo*, *sjofeste*, *solforhold*, *strandlinje* og *turmuligheter*. *Støynivaa* og *sjofeste* fremstilles som dummy variabler og er, i likhet med *garasje* og *naust*, kodet med 0 og 1. Dersom *støynivaa* er kodet med 1 har boligen høyt støynivå, og kodes den med 0 har den lavt støynivå. *Støynivaa* var opprinnelig rangert fra 0 til 10 i datasettet. De som var over 5, og hadde lite støy, ble kodet 0, mens de med mye støy (under 5) ble kodet med 1. *Solforhold* er subjektiv vurdert på en skala fra 0 til 10. Variablene *avssjo* og *strandlinje* er begge målt i antall meter. *Turmuligheter* er rangert mellom 0 og 10, hvor 10 betyr at man rundt boligen har veldig gode turmuligheter. Vi så på muligheten for å også fremstille turmuligheter som dummy variabel, kodet med 0 og 1. Siden alle observasjonene som inneholdt data på *turmuligheter* hadde en verdi på 6,0 og høyere, så vi det som lite hensiktsmessig å kode det som dummy variabel.

### 3.3.4 Deskriptiv statistikk av miljøvariablene

Tabell 8: Deskriptiv analyse av *støynivaa*.

<i>støynivaa</i>	
Gjennomsnitt	0,0078125
Standardfeil	0,00551343
Median	0
Modus	0
Standardavvik	0,08821487
Utvalgsvarians	0,00778186
Kurstosis	125,468411
Skjevhet	11,2466982
Område	1
Minimum	0
Maksimum	1
Sum	2
Antall	256

*Støynivaa* har data på 256 av de 601 observasjonene vi har i datasettet. Det utgjør 42,6 %. «Sum» blir 2, noe som betyr at kun to av 256 observasjoner har et høyt støynivå, resten har et lavt støynivå. Utvalgsvariansen er på 0,00778. og standardavviket er på 0,0882.

Tabell 9: Deskriptiv analyse av avssjo.

<i>avssjo</i>	
Gjennomsnitt	354,730594
Standardfeil	41,3958371
Median	10
Modus	0
Standardavvik	612,602446
Utvalgsvarians	375281,757
Kurstosis	7,58423833
Skjevhet	2,51892476
Område	3600
Minimum	0
Maksimum	3600
Sum	77686
Antall	219

*Avssjo* har observasjoner på 219 av 601 observasjoner, som utgjør 36,44 %. Observasjonene ligger på mellom 0 og 3600 meter til sjøen. Gjennomsnittet er 354,73. Medianen for *avssjo* er kun 10. Årsaken til det store avviket mellom gjennomsnitt og median er at verdien 0 er den som er observert flest ganger for variabelen, i tillegg til at det er et stort avvik mellom minimum og maksimum. Utvalgsvariansen er 375281,76 og standardavviket er 612,60.

Tabell 10: Deskriptiv analyse av sjofeste.

<i>sjofeste</i>	
Gjennomsnitt	0,02828619
Standardfeil	0,00676832
Median	0
Modus	0
Standardavvik	0,16592737
Utvalgsvarians	0,02753189
Kurstosis	30,6463859
Skjevhet	5,70477097
Område	1
Minimum	0
Maksimum	1
Sum	17
Antall	601



Det er registrert data for *sjofeste* på alle observasjonene i datasettet. Gjennomsnittet til *sjofeste* er 0,0283. Summen av alle observasjonene er 17, som vil si at kun 17 av 601 observasjoner har sjøfeste. Utvalgsvariansen er 0,0275 og standardavviket er 0,166.

Tabell 11: Deskriptiv analyse av *solforhold*.

<i>solforhold</i>	
Gjennomsnitt	0,63310105
Standardfeil	0,00683182
Median	0,6
Modus	0,7
Standardavvik	0,1636787
Utvalgsvariens	0,02679072
Kurstosis	-0,6045709
Skjevhet	-0,160242
Område	0,7
Minimum	0,2
Maksimum	0,9
Sum	363,4
Antall	574

Det er registrert 574 observasjoner av *solforhold*. Observasjonene ligger på mellom 0,2 og 0,9, og gjennomsnittet er på 0,633. Utvalgsvariansen er på 0,0268 og standardavviket er på 0,164.

Tabell 12: Deskriptiv analyse av *strandlinje*.

<i>strandlinje</i>	
Gjennomsnitt	0,20798669
Standardfeil	0,10398336
Median	0
Modus	0
Standardavvik	2,54918342
Utvalgsvariens	6,49833611
Kurstosis	149,609854
Skjevhet	12,2586701
Område	35
Minimum	0
Maksimum	35
Sum	125
Antall	601

*Strandlinje* har også informasjon på alle observasjonene i datasettet, og varierer mellom 0 og 25 meter med strandlinje. Gjennomsnittet er 0,208 og modus er 0. Altså har flertallet av observasjonene ingen strandlinje. Utvalgsvariansen er på 6,498 og standardavviket er på 2,549.

Tabell 13: Deskriptiv analyse av *turmuligheter*.

<i>turmuligheter</i>	
Gjennomsnitt	8,92554745
Standardfeil	0,05017831
Median	9
Modus	9,4
Standardavvik	0,83059881
Utvalgsvariens	0,68989439
Kurstosis	1,34896179
Skjevhet	-1,1626037
Område	4
Minimum	6
Maksimum	10
Sum	2445,6
Antall	274

*Turmuligheter* har data på 274 av 601 observasjoner, som er rett under halvparten med 45,59 %. Variabelen er rangert mellom 0 og 10, men vår data har kun verdier mellom 6 og 10. Gjennomsnittet er dermed høyt, og er på 8,92. Verdien som er mest observert er dermed også høy, og modus er 9,4. Utvalgsvariansen tilnærmet lik 0,69 og standardavviket er 0,831.

Tabell 14: Korrelasjonsanalyse av uavhengige variabler.

	<i>naust</i>	<i>utsikt</i>	<i>stoynivaa</i>	<i>turmuligheter</i>	<i>strandlinje</i>	<i>avssjo</i>	<i>solforhold</i>	<i>sjofeste</i>
<i>naust</i>	1							
<i>utsikt</i>	0,13609834	1						
<i>stoynivaa</i>	-0,00981761	0,04968714	1					
<i>turmuligheter</i>	-0,00813903	-0,10894724	<b>-0,26510555</b>	1				
<i>strandlinje</i>	<b>0,41887861</b>	0,12833298	-0,00785064	-0,00818706	1			
<i>avssjo</i>	-0,01179378	-0,03419786	#DIV/0!	0,17122891	-0,03602166	1		
<i>solforhold</i>	0,09536459	<b>0,49506169</b>	0,05554212	0,00143872	0,07506728	-0,07135783	1	
<i>sjofeste</i>	<b>0,48756982</b>	0,13972884	-0,0148781	-0,01555782	<b>0,36039678</b>	-0,09472888	0,11545435	1

I tabell 14 har vi gjennomført en korrelasjonsanalyse, hvor korrelasjonen mellom de ulike variablene i datasettet presenteres. Vi har uthevet de resultatene som har en korrelasjon fra svak og oppover, som vi kan se i tabell 2, som er tommelfingerregelen til Johannessen, Christoffersen

og Tufte (2011). Variablene som besitter den største korrelasjonen er *utsikt* og *solforhold*, og er tilnærmet lik 50 %. Det betyr at det er en moderat samvariasjon mellom de variablene.

For *naust* er det to variabler som korrelerer; *strandlinje* og *sjofeste*. Korrelasjonen naust har med disse variablene er moderat, og er på henholdsvis 41,88 % og 48,75 %. Vi observerer også at variablene *strandlinje* og *sjofeste* har en svak korrelasjon med 36 %. At de tre variablene korrelerer er noe vi muligens kunne si på forhånd, siden det er vanskelig å ha et naust uten å ha sjøfeste og hvis en har naust har en som oftest strandlinje.

De siste variablene som utviser en svak korrelasjon, er mellom *stoynivaa* og *turmuligheter*. Korrelasjonen ligger på -26,51 % som vil si at hvis en bolig er i et område med støy, vil boligen ha mindre tilgang til turmuligheter enn boliger i områder med mindre støy.

## 4 Resultat

### 4.1 Basemodell

Analyse 1 blir utgangspunktet for de resterende analysene, og er en regresjonsanalyse av basemodellen til Anglevik og Øvretveit (2018). Resultatene vi har fått i analysen er ikke helt identiske med det Anglevik og Øvretveit (2018) og Hovland, Kvalvågnes og Kvalevåg (2019) fikk i sine analyser.

*Analyse 1. Regresjonsanalyse av basemodellen til Anglevik og Øvretveit (2018).*

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	593688.5	308378.6	1.925	0.054750	.
avsreg	-9459.1	2739.1	-3.453	0.000599	***
prosarb60	486.1	3977.0	0.122	0.902770	
avskom	-5528.8	4268.7	-1.295	0.195820	
primaerrom	7702.7	507.6	15.174	< 2e-16	***
garasje	293098.2	55491.4	5.282	1.88e-07	***
omsetningsaar2008	191405.6	140342.2	1.364	0.173205	
omsetningsaar2009	212326.1	142275.7	1.492	0.136212	
omsetningsaar2010	408925.8	128544.1	3.181	0.001554	**
omsetningsaar2011	506148.8	130010.8	3.893	0.000112	***
omsetningsaar2012	639969.3	122493.9	5.224	2.53e-07	***
omsetningsaar2013	591732.8	120636.9	4.905	1.25e-06	***
omsetningsaar2014	833470.2	105334.4	7.913	1.53e-14	***
omsetningsaar2015	759356.2	148745.0	5.105	4.65e-07	***
omsetningsaar2016	617362.5	190929.7	3.233	0.001301	**
omsetningsaar2017	483727.7	216228.7	2.237	0.025702	*
naust	1363592.1	264523.7	5.155	3.61e-07	***
boligalder	-11013.4	838.2	-13.140	< 2e-16	***
utsikt	606144.2	134702.5	4.500	8.40e-06	***
---					

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 566400 on 520 degrees of freedom  
(71 observations deleted due to missingness)  
Multiple R-squared: 0.6205, Adjusted R-squared: 0.6073  
F-statistic: 47.23 on 18 and 520 DF, p-value: < 2.2e-16

Anglevik og Øvretveit (2018) opererte med 548 boligobservasjoner, mens Hovland et al. (2019) opererte med 539 observasjoner. I analyse 1 er det fjernet 71 boligobservasjoner på grunn av mangelfull data. Det gjør at vi ender opp med et antall på 530 observasjoner. Det skal sies at alle variablene har samme fortegn, men regresjonskoeffisienten, t-verdien og p-verdien har ulike verdier i alle analysene. Dette kan skyldes at vi har bearbeidet datasettet ulikt og dermed fått ulike verdier. Dersom vi sammenligner med sistnevnte er vår *omsetningsaar2017* statistisk

signifikant, men deres ikke er det. Motsatt er deres *intercept* statistisk signifikant, mens vår ikke er det.

*Analyse 2. VIF-verdier for basemodellen.*

avsreg	5.290175	1
prosarb60	5.172492	1
avskom	1.090357	1
primaerrom	1.291277	1
garasje	1.286472	1
omsetningsaar	1.327316	10
naust	1.080317	1
boligalder	1.120163	1
utsikt	1.126565	1

I både Anglevik og Øvretveit (2018) og Hovland Kvalvågnes og Kvalevåg (2019) diskuterer de problemene knyttet de romlige variablene *avsreg*, *prosarb60* og *avskom*. De beskriver det som at det er en høy negativ korrelasjon mellom *avsreg* og *prosarb60*, og at de ved hjelp av VIF-verdier, presentert i analyse 2, bekrefter muligheten for multikollinearitet mellom disse.

Hovland et al. (2019) sjekker også for heteroskedastisitet ved hjelp av Breusch-Pagan-test. Grunnen til at de sjekker for det er på grunn av heteroskedastisitet bryter med en av forutsetningene som er listet opp i 3.1.1.

*Analyse 3. Breusch-Pagan-test av basemodellen.*

```
data: base
BP = 88.428, df = 18, p-value = 2.762e-11
```

Hovland et al. (2019) tok Breusch-Pagan-test av basemodellen, slik vi gjorde i analyse 3. Breusch-Pagan-testen tok utgangspunkt i at det ikke fantes heteroskedastisitet, og P-verdien fra testen sammenlignes med signifikansnivået på 5 %. Testen deres ga en p-verdi lavere enn 0,05, i likhet med vår analyse 3, som da tilsier at det finnes heteroskedastisitet i basemodellen (Hovland, Kvalvågnes, & Kvalevåg, 2019, s. 64).

## 4.2 Basemodell med stoynivaa

I denne analysen inkluderes variabelen *stoynivaa*. Navrud og Strand (2011) finner at støy har en signifikant negativ virkning på boligpriser. Vi antar dermed at støy vil påvirke negativt i

våre analyser. Resultatene i analyse 4 bekrefter at variabelen får en negativ påvirkning for boligprisen og den får en regresjonskoeffisient på -282 858. Dette viser at støynivå har en høy økonomisk verdi, men at den tilfredsstillende ikke signifikansnivået for t-verdi og p-verdi.

Analyse 4 gir en forklaringskraft,  $R^2$ , på 0,6049 og en R-justert på 0,5685. Sammenlignet med analyse 1, som gir en  $R^2$  på 0,6205 og en R-justert på 0,6073, har forklaringskraften til modellen blitt redusert ved å inkludere *stoynivaa*. Det er også viktig å bemerke at det er et høyt antall observasjoner som er slettet på grunn av manglende data. Det fører til at det kun gjenstår 226 observasjoner i analysen. Når antall observasjoner reduseres kan det få utslag for de andre variablene sammenlignet med resultatet fra analyse 1.

Vi observerer først at regresjonskoeffisienten til *avsreg* har blitt redusert fra -9 459 til -9 670. Det i seg selv har ikke særlig betydning for en boligpris, men det som er mer interessant er hvordan t-verdien og p-verdien har økt. Den gikk fra å ha en t-verdi på -3,453 og en p-verdi på 0,00059, som vil si at den var godt innenfor signifikansnivået, til nå å ha en t-verdi på -2,024 og en p-verdi på 0,0443. Variabelen er fremdeles statistisk signifikant, men ikke i nærheten av hva den var i analyse 1. Videre ser vi at *prosarb60* har skiftet fortegn og har nå en koeffisient på -2 994. Det går mot vår antagelse om at variabelen skal ha et positivt fortegn grunnet den høye negative korrelasjonen med *avsreg*. Derimot viser t-verdien og p-verdien at den fortsatt er et stykke fra å være statistisk signifikant. Variabelen *avskom* endres koeffisienten fra -5 528 til -6 255. Endringen i koeffisienten har heller ingen særlig betydning for boligprisen, totalt sett, og endringen i t-verdien og p-verdien fører heller ikke til at den blir statistisk signifikant. For variabelen *primaerrom* øker regresjonskoeffisienten til 8 381, men t-verdien reduseres. Den har fortsatt en høy t-verdi på 9,312 og dermed innenfor signifikansnivået. Dersom vi ser på regresjonskoeffisienten og t-verdien til *garasje* har begge falt i verdi. Regresjonskoeffisienten er redusert, men har fortsatt høy økonomisk verdi. Variabelen er enda statistisk signifikant, da t-verdien er over 1,96 og p-verdien er under 0,05.

## Analyse 4. Regresjonsanalyse av basemodellen inkludert stoynivaa.

```

-----
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    687145.4   519186.0    1.324 0.187134
avsreg         -9670.4    4778.8   -2.024 0.044303 *
prosarb60      -2994.1    6596.9   -0.454 0.650400
avskom        -6255.4    9816.5   -0.637 0.524684
primaerrom     8381.3     900.1    9.312 < 2e-16 ***
garasje       259201.1    95079.8    2.726 0.006960 **
omsetningsaar2008 220680.2   269121.2    0.820 0.413162
omsetningsaar2009 196480.7   261074.2    0.753 0.452558
omsetningsaar2010 336228.3   206900.0    1.625 0.105675
omsetningsaar2011 570774.6   236850.4    2.410 0.016838 *
omsetningsaar2012 753453.9   215973.8    3.489 0.000594 ***
omsetningsaar2013 645505.0   202647.9    3.185 0.001670 **
omsetningsaar2014 776584.7   160294.4    4.845 2.49e-06 ***
omsetningsaar2015 745145.0   255490.2    2.917 0.003932 **
omsetningsaar2016 645471.2   314875.9    2.050 0.041639 *
omsetningsaar2017 342941.8   429700.4    0.798 0.425734
naust         2062918.7   478514.1    4.311 2.52e-05 ***
boligalder    -12809.5     1580.0   -8.107 4.55e-14 ***
utsikt        782573.8   212974.4    3.674 0.000304 ***
stoynivaa     -282857.7   462389.4   -0.612 0.541391
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 631400 on 206 degrees of freedom
(375 observations deleted due to missingness)
Multiple R-squared:  0.6049,    Adjusted R-squared:  0.5685
F-statistic: 16.6 on 19 and 206 DF,  p-value: < 2.2e-16

```

For de ulike variablene for *omsetningsaar* øker regresjonskoeffisienten til 2008, 2011, 2012, 2013, 2015 og 2016, mens regresjonskoeffisienten for de resterende årene går ned. For 2008 og 2009 beholder de en t-verdi og en p-verdi som ikke er statistisk signifikant. For 2010 derimot ser en at t-verdien reduseres fra 3,181 til 1,625 og en p-verdi fra 0,0015 til 0,1056. Det vil si at 2010 gikk fra å være statistisk signifikant, til å ikke lenger være det. Variabelen 2011 har en lignende tendens, men beholder en t-verdi og p-verdi innenfor signifikansnivået på 5 %. Den samme tendens har vi også for 2012, 2013, 2014, 2015 og 2016. Både t-verdien og p-verdien endres, men for alle disse variablene opprettholder de signifikansnivået på 5 %. Derimot for 2017 reduseres t-verdien og p-verdien øker så mye at variabelen ikke lenger er statistisk signifikant.

For variabelen *naust* øker regresjonskoeffisienten, og har dermed den største endringen i analyse 4. Det vi ser er at t-verdien reduseres og p-verdien stiger, men den er fortsatt innenfor signifikansnivået på 5 %. For *boligalder* er det en mindre endring i koeffisienten, men en stor

endring i t-verdien hvor den reduseres fra -13,14 til -8,107. Selv med en stor endring i t-verdien er både t-verdien og p-verdien statistisk signifikant. For den siste variabelen i basemodellen, *utsikt*, har også den en stor økning i koeffisienten. Det vil også for denne variabelen ha seg at t-verdien og p-verdien reduseres, men som mange av de andre variablene er den fremdeles innenfor signifikansnivået.

*Analyse 5. VIF-verdier for basemodellen inkludert stoynivaa.*

	GVIF	Df
avsreg	4.064041	1
prosarb60	4.103004	1
avskom	1.555174	1
primaerrom	1.372356	1
garasje	1.252137	1
omsetningsaar	2.273946	10
naust	1.138486	1
boligalder	1.142049	1
utsikt	1.100659	1
stoynivaa	1.063051	1

Siden vi antok at *stoynivaa* skulle ha en innvirkning på boligprisen, så vi på muligheten for om variabelens manglende statistiske signifikans skyldes multikollinearitet. I analyse 5 observeres ingen VIF-verdier som skulle tilsi at det foreligger multikollinearitet når variabelen *stoynivaa* er inkludert. Multikollinearitet mellom *avsreg* og *prosarb60* har blitt redusert sammenlignet med analyse 2.

*Analyse 6. Breusch-Pagan-test av basemodellen inkludert stoynivaa.*

```
data: nystoy
BP = 36.189, df = 19, p-value = 0.01
```

Ved å ta en Breusch-Pagan-test ser vi at p-verdien er lavere enn signifikansnivået på 5 % som gjør at vi må forkaste nullhypotesen. Det tyder på at foreligger heteroskedastisitet i denne modellen.

### 4.3 Basemodell med avssjo

I analyse 7 inkluderes en ny miljøvariabel, *avssjo*. Denne inkluderes på grunnlag av at vi fra tidligere forskning vet at det gir nytteverdi å ha en bolig nær kysten, og at økt avstand fra sjøen



vil føre til en reduksjon i boligverdi (Conroy & Milosch, 2011). Med grunnlag i dette antar vi at variabelen skal ha et negativt fortegn, om går ut på at til lenger boligen er fra sjøen, desto mer reduseres prisen. Vi observerer at variabelen har et negativt fortegn, men regresjonskoeffisienten er på -63. Det tilsier at den har en minimal økonomisk verdi. Den har en t-verdi på -0,821 og en p-verdi på 0,4128, noe som gjør at signifikansnivået på 5 % ikke er tilfredsstillt.

Sammenligner vi forklaringskraften fra analyse 1 og analyse 7 kan vi se at  $R^2$  øker til 0,672 og 0,6381. Det betyr at variabelen bidrar til å øke forklaringskraften til modellen med over 5 %. Det som er å merke seg ved denne variabelen er mangelen på observasjoner. Det har i denne analysen blitt fjernet 417 observasjoner på grunn av manglende data, som vil si at vi kun har 184 observasjoner igjen. På grunnlag av dette kan man anta at flere variabler vil få større endringer i både regresjonskoeffisient, t-verdi og p-verdi.

*Analyse 7. Regresjonsanalyse av basemodellen inkludert avssjo.*

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	308542.69	625854.07	0.493	0.622638
avsreg	-809.50	5709.19	-0.142	0.887411
prosarb60	3555.69	7427.85	0.479	0.632755
avskom	-11973.73	9472.31	-1.264	0.207893
primaerrom	7276.52	804.81	9.041	2.97e-16 ***
garasje	213784.33	93109.64	2.296	0.022866 *
omsetningsaar2008	192456.72	217599.57	0.884	0.377672
omsetningsaar2009	341116.83	229716.98	1.485	0.139368
omsetningsaar2010	462753.91	199889.37	2.315	0.021778 *
omsetningsaar2011	427299.13	235147.95	1.817	0.070915 .
omsetningsaar2012	672422.23	212729.17	3.161	0.001855 **
omsetningsaar2013	566621.60	190475.41	2.975	0.003349 **
omsetningsaar2014	816314.07	166683.97	4.897	2.21e-06 ***
omsetningsaar2015	1099888.38	259813.91	4.233	3.72e-05 ***
omsetningsaar2016	1541956.65	424668.42	3.631	0.000371 ***
naust	2351155.94	356646.58	6.592	4.98e-10 ***
boligalder	-12086.26	1592.37	-7.590	1.86e-12 ***
utsikt	636840.31	224823.29	2.833	0.005162 **
avssjo	-63.27	77.07	-0.821	0.412812

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 548500 on 174 degrees of freedom  
 (417 observations deleted due to missingness)  
 Multiple R-squared: 0.672, Adjusted R-squared: 0.6381  
 F-statistic: 19.8 on 18 and 174 DF, p-value: < 2.2e-16

Sammenligner vi analyse 7 med analyse 1, bekrefter det vår antagelse om at det ville bli merkbare endringer i modellen. Alle variablene har det fortegnet vi antok at de ville ha. Variabelen *avsreg* får i analyse 7 en kraftig endring i koeffisienten og får en verdi på -810. Den går også fra å være statistisk signifikant til nå å ha en t-verdi på -0,142 og en p-verdi på 0,8874. Regresjonskoeffisienten *prosarb60* har også en større endring, der den øker fra 486 til 3 556. Selv om den har en større endring i koeffisienten, er variabelen fortsatt ikke statistisk signifikant. For *avskom* har vi en lignende situasjon, der regresjonskoeffisienten mer enn doubler seg til -11 974. For t-verdien øker den i absoluttverdi fra 1,264 til 1,295 og p-verdien øker fra 0,1958 til 0,2078. For *primaerrom* er det en mindre endring i både regresjonskoeffisienten og p-verdien, mens for t-verdien reduseres den fra 15,174 til 9,041. Selv med en stor nedgang i t-verdi er variabelen innenfor signifikansnivået på 5 %. Sammenligner vi resultatene for *garasje* ser vi at det gir utslag for t-verdien og p-verdien. Der den i analyse 1 hadde en t-verdi på 5,282 og en p-verdi på 1,88e-07, får den i analyse 7 verdier på 2,296 og 0,0228. På tross av dette, er den fremdeles statistisk signifikant.

For *omsetningsaar* har en del av variablene opprettholdt t-verdi og p-verdi innenfor signifikansnivået, men for 2011 har t-verdien blitt redusert til 1,817 og er dermed ikke lenger statistisk signifikant. Vi kan også se at 2015 og 2016 har en kraftig økning i regresjonskoeffisienten, mens 2017 forsvinner fra analysen.

Variablene *naust* og *boligalder* har blitt mer statistisk signifikante, men det er kun endringen i regresjonskoeffisienten til *naust* som har større betydning for den økonomiske verdien. For den siste variabelen i den opprinnelige basemodellen ser vi at *utsikt* har en tilnærmet lik koeffisient, mens signifikansnivået øker. Selv med en solid økning er variabelen fortsatt statistisk signifikant.

Analyse 8. VIF-verdier for basemodellen inkludert *avssjo*.

	GVIF	Df
avsreg	7.728797	1
prosarb60	7.144403	1
avskom	1.268150	1
primaerrom	1.351003	1
garasje	1.363166	1
omsetningsaar	1.922122	9
naust	1.248647	1
boligalder	1.272108	1
utsikt	1.269649	1
avssjo	1.570325	1

Vi beregner VIF-verdiene i analyse 8 for å undersøke om det er en mulighet for multikollinearitet i modellen etter å ha inkludert *avssjo*. Resultatet viser en økning i multikollinearitet mellom *avsreg* og *prosarb60*.

Analyse 9. Breusch-Pagan-test av basemodellen inkludert *avssjo*.

```
data: avssjo
BP = 48.524, df = 18, p-value = 0.0001258
```

Ved å ta en Breusch-Pagan-test ser vi at p-verdien er lavere enn signifikansnivået på 5 % som gjør at vi må forkaste nullhypotesen. Det tyder på at det foreligger heteroskedastisitet i denne modellen.

#### 4.4 Basemodell med sjofeste

Vi antar at om en bolig har *sjofeste* vil det påvirke boligprisen i positiv forstand, og inkluderer derfor variabelen *sjofeste* i analysen. Med grunnlag i Conroy og Milosch (2011) som fant at økning i avstand til kysten fører til en reduksjon i boligprisene, kan vi anta at dersom en bolig har *sjofeste*, vil prisen øke. Vi ser i analyse 10 at variabelen har et positivt fortegn på 174 943 som tilsier at hvis en bolig har *sjofeste* øker boligverdien med 174 493 kroner. Selv om koeffisienten er det vi antar viser t-verdien og p-verdien at variabelen ikke er statistisk signifikant. I analysen gir de oss verdier på henholdsvis 0,981 og 0,3272.

Ser vi videre på hva variabelen tilfører modellen i form av forklaringskraft er det minimalt. Hvor den i analyse 1 hadde en  $R^2$  lik 0,6205 og en R-justert lik 0,6073 får den i analyse 10 en  $R^2$  lik 0,6212 og en R-justert like 0,6073. Ved å ta med variabelen *sjofeste* øker

forklaringskraften med 0,05 %, mens for R-justert er den på stedet hvil. Det som er positivt, er at det ikke forsvinner flere observasjoner i denne regresjonsanalysen sammenlignet med regresjonsanalysen av bare basemodellen. Også her forsvinner det 71 observasjoner på grunn av mangel på data, og en har dermed et utvalg på 530. Siden det ikke forsvinner observasjoner antar vi at resultatene i analyse 10 skal være relativt like analyse 1.

*Analyse 10. Regresjonsanalyse av basemodellen inkludert sjofeste.*

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	580803.6	308669.8	1.882	0.060445	.
avsreg	-9234.8	2748.8	-3.360	0.000838	***
prosarb60	760.6	3987.0	0.191	0.848789	
avskom	-5774.3	4276.2	-1.350	0.177494	
primaerrom	7732.4	508.6	15.204	< 2e-16	***
garasje	292491.3	55496.9	5.270	2.00e-07	***
omsetningsaar2008	175189.7	141318.4	1.240	0.215654	
omsetningsaar2009	204576.9	142500.3	1.436	0.151711	
omsetningsaar2010	400085.1	128864.7	3.105	0.002009	**
omsetningsaar2011	506932.2	130018.1	3.899	0.000109	***
omsetningsaar2012	632855.2	122713.2	5.157	3.57e-07	***
omsetningsaar2013	585714.2	120797.5	4.849	1.64e-06	***
omsetningsaar2014	833229.3	105338.6	7.910	1.56e-14	***
omsetningsaar2015	759671.5	148750.8	5.107	4.60e-07	***
omsetningsaar2016	604928.3	191357.5	3.161	0.001663	**
omsetningsaar2017	482609.6	216239.7	2.232	0.026052	*
naust	1193390.8	316399.0	3.772	0.000181	***
boligalder	-11014.6	838.2	-13.141	< 2e-16	***
utsikt	590248.6	135679.4	4.350	1.64e-05	***
sjofeste	174943.0	178416.9	0.981	0.327282	
---					

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 566500 on 519 degrees of freedom  
(71 observations deleted due to missingness)  
Multiple R-squared: 0.6212, Adjusted R-squared: 0.6073  
F-statistic: 44.79 on 19 and 519 DF, p-value: < 2.2e-16

Sammenligner vi de to analysene bekreftes vår antagelse. Resultatet vi får ved å inkludere *sjofeste* er tilnærmet lik resultatet vi fikk for basemodellen. Det er ingen endringer, av stor betydning, når det kommer til regresjonskoeffisienten, t-verdi eller p-verdi for variablene i modellen. Derfor er det ingen som endrer fortegn eller at deres statistiske signifikans endres.

Analyse 11. VIF-verdier for basemodellen inkludert sjofeste.

	GVI	F	Df
avsreg	5.327072	1	1
prosarb60	5.198115	1	1
avskom	1.094105	1	1
primaerrom	1.295860	1	1
garasje	1.286633	1	1
omsetningsaar	1.379096	10	10
naust	1.545469	1	1
boligalder	1.120165	1	1
utsikt	1.142881	1	1
sjofeste	1.540182	1	1

Siden resultatene i analyse 10 gir tilnærmet like resultat antar vi at VIF-verdiene også skal være tilnærmet lik de verdiene som var i analyse 2. Vi beregner verdiene og resultatet bekrefter våre antagelser.

Analyse 12. Breusch-Pagan-test av basemodellen inkludert sjofeste.

```
data: sjofeste
BP = 87.879, df = 19, p-value = 7.864e-11
```

Ved å ta en Breusch-Pagan-test ser vi at p-verdien er lavere enn signifikansnivået på 5 % som gjør at vi må forkaste nullhypotesen. Det tyder på at foreligger heteroskedastisitet i denne modellen.

#### 4.5 Basemodell med solforhold

I den neste analysen legger vi til variabelen *solforhold*, som vi også antar skal ha en positiv påvirkning på boligprisen. Denne variabelen inkluderes fordi vi vet fra tidligere forskning at en økning i antall soltimer per dag gir en økning i boligpris (Fleming, Grimes, Lebreton, Maré, & Nunns, 2017). Som analyse 13 viser, har variabelen en høy økonomisk verdi, men tilfredsstillende ikke signifikansnivå på 5 %. Forklaringskraften til modellen etter å ha inkludert *solforhold* er 0,6226 og en R-justert på 0,6074. Det gir en økning fra analyse 1 og viser at variabelen har en betydning for modellen.

## Analyse 13. Regresjonsanalyse av basemodellen inkludert solforhold.

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	515202.2	317838.1	1.621	0.105634	
avsreg	-9130.7	2757.9	-3.311	0.000995	***
prosarb60	600.3	3978.4	0.151	0.880122	
avskom	-5627.3	4269.6	-1.318	0.188090	
primaerrom	7585.3	520.5	14.572	< 2e-16	***
garasje	291740.1	55505.4	5.256	2.15e-07	***
omsetningsaar2008	194645.4	140372.9	1.387	0.166150	
omsetningsaar2009	211186.1	142274.8	1.484	0.138322	
omsetningsaar2010	409425.1	128540.3	3.185	0.001534	**
omsetningsaar2011	505213.5	130009.2	3.886	0.000115	***
omsetningsaar2012	646003.6	122632.4	5.268	2.03e-07	***
omsetningsaar2013	593026.4	120639.1	4.916	1.19e-06	***
omsetningsaar2014	832187.0	105338.0	7.900	1.68e-14	***
omsetningsaar2015	772213.5	149273.5	5.173	3.29e-07	***
omsetningsaar2016	638268.4	192021.4	3.324	0.000950	***
omsetningsaar2017	512420.8	218045.9	2.350	0.019144	*
naust	1361132.3	264524.8	5.146	3.79e-07	***
boligalder	-10996.6	838.3	-13.118	< 2e-16	***
utsikt	534127.9	152107.5	3.512	0.000484	***
solforhold	190014.7	186444.7	1.019	0.308608	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 566400 on 519 degrees of freedom  
(71 observations deleted due to missingness)  
Multiple R-squared: 0.6212, Adjusted R-squared: 0.6074  
F-statistic: 44.8 on 19 and 519 DF, p-value: < 2.2e-16

Det er små endringer totalt sett i denne analysen. I analyse 13 er det ingen store endringer i regresjonskoeffisienten, t-verdi eller p-verdi for *avsreg*, *prosarb60*, *avskom*, *primaerrom* og *garasje*, sammenlignet med hvordan verdiene var i analyse 1. Tilfellet er det samme for alle omsetningsår, der den eneste endringen som kan være av betydning er at variabelen *omsetningsaar2016* har blitt mer statistisk signifikant. *Naust* og *boligalder* har små endringer i regresjonskoeffisient, t-verdi og p-verdi. *Utsikt* har fått lavere økonomisk verdi, i tillegg til at t-verdien og p-verdien har falt. Variabelen er imidlertid fortsatt statistisk signifikant.

## Analyse 14. VIF-verdier for basemodellen inkludert solforhold.

	GVIF	Df
avsreg	5.363413	1
prosarb60	5.176600	1
avskom	1.090915	1
primaerrom	1.357762	1
garasje	1.287214	1
omsetningsaar	1.392096	10
naust	1.080407	1
boligalder	1.120598	1
utsikt	1.436610	1
solforhold	1.552869	1

Vi beregner VIF-verdier for basemodellen inkludert *solforhold*, men resultatene gir ingen nevneverdige nye funn.

*Analyse 15. Breusch-Pagan-test av basemodellen inkludert solforhold.*

```
data: sol  
BP = 85.769, df = 19, p-value = 1.847e-10
```

Ved å ta en Breusch-Pagan-test ser vi at p-verdien er lavere enn signifikansnivået på 5 % som gjør at vi må forkaste nullhypotesen. Det tyder på at foreligger heteroskedastisitet i denne modellen.

#### 4.6 Basemodell med strandlinje

I analyse 16 legger vi til miljøvariabelen *strandlinje* i basemodellen. Tidligere forskning viser at en bredere strand gir en økning i boligens verdi (Rinehart & Pompe, 1994), samtidig som Conroy og Milosch (2011) fant at det gir økt nytteverdi å ha en bolig nær kysten. Vi antar derfor at variabelen skal ha et positivt fortegn på regresjonskoeffisienten, som forteller oss at for hver ekstra meter med strandlinje en bolig har, øker prisen.

Analyse 16 bekrefter vår antagelse om positiv regresjonskoeffisient. Regresjonskoeffisienten til *strandlinje* er 17 665. Til tross for positiv regresjonskoeffisient, har den likevel lav økonomisk verdi. T-verdien er 1,705 og p-verdien er 0,0888, ergo tilfredsstillende ikke *strandlinje* signifikansnivå, og man kan ikke med 95 % sikkerhet si at strandlinje påvirker boligprisene. Modellen har en forklaringskraft,  $R^2$ , på 0,6226 og R-justert på 0,6088. Dersom man sammenligner dette med analyse 1, kan man se at forklaringskraften og R-justert har økt. Dette betyr at ved å inkludere variabelen *strandlinje* forbedres modellen.

## Analyse 16. Regresjonsanalyse av basemodellen inkludert strandlinje.

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	596209.7	307818.2	1.937	0.053301	.
avsreg	-9353.9	2734.8	-3.420	0.000675	***
prosarb60	693.2	3971.6	0.175	0.861512	
avskom	-6215.1	4279.8	-1.452	0.147053	
primaerrom	7699.2	506.7	15.194	< 2e-16	***
garasje	293152.4	55390.0	5.293	1.78e-07	***
omsetningsaar2008	192663.2	140087.5	1.375	0.169630	
omsetningsaar2009	210635.2	142018.9	1.483	0.138642	
omsetningsaar2010	407363.6	128312.3	3.175	0.001589	**
omsetningsaar2011	507647.3	129776.1	3.912	0.000104	***
omsetningsaar2012	623687.1	122642.3	5.085	5.13e-07	***
omsetningsaar2013	591906.4	120416.4	4.915	1.19e-06	***
omsetningsaar2014	831822.0	105146.2	7.911	1.55e-14	***
omsetningsaar2015	760776.2	148475.3	5.124	4.23e-07	***
omsetningsaar2016	620688.7	190590.6	3.257	0.001201	**
omsetningsaar2017	485503.8	215835.8	2.249	0.024905	*
naust	1143319.8	293951.3	3.889	0.000114	***
boligalder	-10990.1	836.7	-13.134	< 2e-16	***
utsikt	587199.7	134914.4	4.352	1.62e-05	***
strandlinje	17664.9	10360.6	1.705	0.088790	.

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 565400 on 519 degrees of freedom  
(71 observations deleted due to missingness)  
Multiple R-squared: 0.6226, Adjusted R-squared: 0.6088  
F-statistic: 45.06 on 19 and 519 DF, p-value: < 2.2e-16

Det har oppstått små endringer fra analyse 1 ved å legge til variabelen *strandlinje*. Dette kan komme av at ingen observasjoner er slettet ved å inkludere variabelen. Regresjonskoeffisientene til *avsreg*, *prosarb60*, *avskom*, *primaerrom* og *garasje* har små endringer ved å tilføye *strandlinje* til modellen. *Strandlinje* har dermed minimal betydning for økonomisk verdi for disse variablene i basemodellen. T-verdi og p-verdi på disse variablene har heller ingen store endringer. Både regresjonskoeffisientene, t-verdi og p-verdi for alle årene som inngår i *omsetningsaar* har små endringer og dermed ingen endringer av betydning. De omsetningsårene som var statistisk signifikante i analyse 1, er fremdeles statistisk signifikante.

Variabelen *naust* har derimot en større endring i regresjonskoeffisienten. Den er redusert fra 1 363 583 til 1 143 319. T-verdi og p-verdi er redusert, men *naust* tilfredsstillers fremdeles signifikansnivået på 5 % og variabelen er like statistisk signifikant som i analyse 1. Vi ser dermed at regresjonskoeffisienten til *naust* blir betraktelig mindre når *strandlinje* inkluderes i modellen. *Boligalder* og *utsikt* har i likhet med de tidligere nevnte variablene, lite endringer



som følge av å legge til *strandlinje*. Det er ingen endringer av betydning på hverken økonomisk verdi eller signifikansnivå.

*Analyse 17. VIF-verdier for basemodellen inkludert strandlinje.*

	GVIF	Df
avsreg	5.292873	1
prosarb60	5.177335	1
avskom	1.100087	1
primaerrom	1.291298	1
garasje	1.286473	1
omsetningsaar	1.346524	10
naust	1.338945	1
boligalder	1.120463	1
utsikt	1.134258	1
strandlinje	1.308210	1

I analyse 17 beregner vi VIF-verdiene til basemodellen inkludert *strandlinje*. Vi finner ingen nevneverdig endring som kan skape problemer med multikollinearitet.

*Analyse 18. Breusch-Pagan-test av basemodellen inkludert strandlinje.*

```
data: strand
BP = 93.172, df = 19, p-value = 9.048e-12
```

Ved å ta en Breusch-Pagan-test ser vi at p-verdien er lavere enn signifikansnivået på 5 % som gjør at vi må forkaste nullhypotesen. Det tyder på at foreligger heteroskedastisitet i denne modellen.

#### 4.7 Basemodell med turmuligheter

I analyse 19 inkluderes variabelen *turmultipheter* i analyse 1. Denne variabelen inkluderes fordi vi vet fra Tyrväinen (1997) at nærhet til skogsområder har en positiv innvirkning på boligpriser. Vi antar derfor at *turmultipheter* skal ha et positivt fortegn på regresjonskoeffisienten, og at desto bedre turmuligheter det er i nærheten av en bolig, desto høyere vil prisen være. Dersom vi ser på analyse 19 kan vi observere at regresjonskoeffisienten er positiv, slik som vi antok i forveien. Regresjonskoeffisienten har en verdi på 87 424, som vil si at for hvert nivå av bedre turmuligheter vil prisen på en bolig, ut fra våre resultater, øke med denne verdien. T-verdien er 1,673 og p-verdien er 0,103, noe som ikke tilfredsstillt signifikansnivået. Vi kan derfor ikke

konkludere om *turmuligheter* påvirker boligprisene. Modellen i analyse 19 har en  $R^2$  på 0,6501 og en R-justert på 0,6203. Dersom vi sammenligner  $R^2$  og R-justert med analyse 1, har disse økt. Man kan da anta ut fra forklaringskraften og R-justert at ved å inkludere variabelen *turmuligheter*, vil modellen ha en høyere forklaringsevne. Ved å inkludere *turmuligheter* vi 367 observasjoner, som er over halvparten av alle observasjonene vi har. Ettersom vi mister mange observasjoner kan det få utslag på de resterende variablene, og resultatene i analysen kan bære preg av det.

*Analyse 19. Regresjonsanalyse av basemodellen inkludert turmuligheter.*

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-452545.73	685416.97	-0.660	0.509776
avsreg	-6782.22	4055.85	-1.672	0.095886 .
prosarb60	70.59	5810.28	0.012	0.990317
avskom	1812.19	8153.07	0.222	0.824307
primaerrom	7795.14	732.35	10.644	< 2e-16 ***
garasje	253255.72	80308.74	3.154	0.001835 **
omsetningsaar2008	413409.28	230955.72	1.790	0.074812 .
omsetningsaar2009	337817.10	233213.03	1.449	0.148872
omsetningsaar2010	485359.83	180285.97	2.692	0.007638 **
omsetningsaar2011	681868.01	205755.58	3.314	0.001073 **
omsetningsaar2012	882789.65	178782.75	4.938	1.55e-06 ***
omsetningsaar2013	735203.83	177015.25	4.153	4.67e-05 ***
omsetningsaar2014	941455.83	147142.54	6.398	9.13e-10 ***
omsetningsaar2015	792500.83	232820.05	3.404	0.000787 ***
omsetningsaar2016	429776.69	301314.55	1.426	0.155169
omsetningsaar2017	381467.79	376288.73	1.014	0.311794
naust	2273499.28	350355.69	6.489	5.51e-10 ***
boligalder	-11533.40	1405.77	-8.204	1.84e-14 ***
utsikt	645053.11	187584.23	3.439	0.000697 ***
turmuligheter	87423.76	53412.48	1.637	0.103090

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 560200 on 223 degrees of freedom  
 (367 observations deleted due to missingness)  
 Multiple R-squared: 0.6501, Adjusted R-squared: 0.6203  
 F-statistic: 21.81 on 19 and 223 DF, p-value: < 2.2e-16

Det første vi oppdager er at skjæringspunktet har falt fra 593 689 til -452 546, som vil si at boligen renses for kvalitetsforskjeller vil ha en negativ verdi. Regresjonskoeffisientene som har økt i analyse 19 er blant annet *avsreg*, *avskom* og *primaerrom*. *Avskom* tilfredsstillers ikke lenger signifikansnivå, på grunn av at t-verdi er redusert og p-verdi økt. *Prosarb60* har også fått lavere t-verdi og høyere p-verdi, i tillegg til at den har fått enda mindre økonomisk verdi. Tilfellet er likt for variabelen *garasje*, med unntak av at *garasje* fortsatt tilfredsstillers signifikansnivået.

Regresjonskoeffisientene til flertallet av *omsetningsaar* har økt en del, bortsett fra 2016 og 2017 hvor verdien er redusert. 2016 og 2017 er ikke lenger statistisk signifikante, og 2011 er blitt mindre statistisk signifikant. For alle tre årene er det en reduksjon i t-verdi, samt en økning i p-verdi. Regresjonskoeffisientene til både *naust* og *utsikt* har også økt ved å inkludere *turmuligheter*. Det har vært en økning for både t-verdien og p-verdien for *naust*, men for *utsikt* er tilfellet motsatt. Til tross for ulike endringer i t-verdi og p-verdi tilfredsstillers begge variablene fortsatt signifikansnivået. For *boligalder* er koeffisienten mer negativ. For denne variabelen har både t-verdi og p-verdi hatt en nedgang, men den er enda statistisk signifikant.

Analyse 20. VIF-verdier for basemodellen inkludert *turmuligheter*.

	GVIF	Df
avsreg	6.404877	1
prosarb60	6.085522	1
avskom	1.502369	1
primaerrom	1.290370	1
garasje	1.228299	1
omsetningsaar	2.238304	10
naust	1.159052	1
boligalder	1.330563	1
utsikt	1.114656	1
turmuligheter	1.157972	1

Vi beregner VIF-verdier for basemodellen inkludert *turmuligheter*. Vi registrerer at *omsetningsaar* får en liten økning i verdi, men ikke nok til at det vil skape problemer.

Analyse 21. Breusch-Pagan-test av basemodellen inkludert *turmuligheter*.

```
data: tur
BP = 53.01, df = 19, p-value = 4.664e-05
```

Ved å ta en Breusch-Pagan-test ser vi at p-verdien er lavere enn signifikansnivået på 5 % som gjør at vi må forkaste nullhypotesen. Det tyder på at foreligger heteroskedastisitet i denne modellen.

## 5 Feilspesifisering

### 5.1 Korrelasjon

For å kunne beregne de implisitte prisene for miljøvariablene, er det viktig å legge til de variablene som korrelerer med de allerede eksisterende variablene. Selv om noen variabler i seg selv kan ha lav økonomisk verdi, utgjør de en stor del dersom de blir fanget opp i en annen variabel.

I analyse 13, hvor vi legger til variabelen *solforhold*, kan vi se at t-verdien og p-verdien til *utsikt* mister statistisk signifikans. Ser vi disse resultatene opp mot tabell 14 kan vi se at det er en moderat korrelasjon mellom disse variablene. Vi kan dermed tenke oss at den implisitte prisen til *utsikt* i basemodellen, fanger opp variabelen *solforhold* og at den implisitte prisen til *utsikt* blir unormalt høy i forhold til hva den egentlig skulle vært. Ved å inkludere *solforhold* vil man rette opp en del av skjevheten og feilspesifiseringen som tidligere fantes, og *utsikt* vil være nærmere den implisitte prisen på kun *utsikt* alene.

For å si noe om skjevheten til *utsikt* ser vi på tabell 1. Tabellen viser at, siden *solforhold* har en regresjonskoeffisient større enn 0 og korrelasjonen mellom *utsikt* og *solforhold* er positiv, vil det forekomme en positiv skjevhet for den uavhengige variabelen *utsikt*. Ved å inkludere *solforhold* vil vi rette opp en del av skjevheten som er for variabelen. Dersom vi ikke inkluderer *solforhold*, vil ikke parameteren til *utsikt* være forventningsrett.

For den uavhengige variabelen *naust* oppstår det endringer når vi legger til variablene; *avssjo*, *sjofeste* og *strandlinje*. I analyse 7 får vi en nevneverdig endring i koeffisienten, hvor den implisitte prisen øker med bortimot én million, samt at den blir mer signifikant. På en annen side får vi en motsatt endring når vi legger til *sjofeste* og *strandlinje*. Her avtar både koeffisienten og signifikansnivået til variabelen *naust*. Vi kan dermed se for oss at disse variablene korrelerer med *naust*, og det får vi delvis bekreftet i tabell 14. Tabell 14 at *sjofeste* og *strandlinje* har en moderat positiv korrelasjon med *naust*, mens *avssjo* har en meget svak negativ korrelasjon. Det betyr at den implisitte prisen på *naust*, fanger opp mer enn bare verdien på *naust* og har dermed uventet høy økonomisk verdi. Sammenligner vi med *garasje*, gir den en mye lavere økonomisk verdi.

For at vi skal kunne se om det er en positiv eller negativ skjevhet for *naust* ser vi på resultatene til de nevnte variablene ovenfor. Resultatene til *sjofeste* og *strandlinje* viser at koeffisientene har en positiv påvirkning for modellen, samt at begge variablene har en positiv korrelasjon med *naust*. Dette vil da tilsi at *naust* har en positiv skjevhet til disse variablene. For *avssjo* kan vi se en lav negativ koeffisient og at korrelasjonen mellom *naust* og *avssjo* er minimalt negativ. Ifølge tabell 1 gir dette at *naust* også har en positiv skjevhet til *avssjo*. Parameteren til *naust* vil ikke være forventningsrett dersom *sjofeste* og *strandlinje* er utelatt fra modellen.

Det vi også kan se i tabell 14 er at *sjofeste* og *strandlinje* har en svak positiv korrelasjon. Å inkludere begge i en endelig modell vil gjøre at den implisitte prisen til begge variablene blir bedre og at variabelen kun omhandler seg selv. Det samme gjelder for variablene *turmuligheter* og *stoynivaa*. Her viser tabell 14 derimot en svak negativ korrelasjon, som sier at har du hvis du har over middels høyt støynivå har du mindre tilgang til turmuligheter. Inkluderer vi kun en av disse i den endelige modellen vil variablene enten få en for høy eller for lav implisitt pris.

For *sjofeste* og *strandlinje* vil det forekomme en positiv skjevhet uansett hvilken av dem som blir sett på som  $\beta_1$  og  $\beta_2$ . Begge variablene har en positiv virkning på modellen, samt at korrelasjonen mellom de er positiv. Derimot blir det annerledes for *turmuligheter* og *stoynivaa*. Det vil her avhenge av hvilken som er  $\beta_1$  og hvilken som er  $\beta_2$ . Et eksempel kan være at *turmuligheter* er  $\beta_1$  og *stoynivaa* er  $\beta_2$ . Korrelasjonen mellom variablene er negativ, og *stoynivaa* har en negativ påvirkning på modellen. Det resulterer i positiv skjevhet for *turmuligheter*. Snur vi derimot på eksempelet og lar *stoynivaa* være  $\beta_1$  og *turmuligheter* være  $\beta_2$  vil vi enda ha at korrelasjonen er negativ, men siden *turmuligheter* har en positiv påvirkning for modellen viser tabell 1 at det tilsier en negativ skjevhet for *stoynivaa*.

## 5.2 Målefeil

En del av variablene i datasettet er hentet inn gjennom subjektive vurderinger og derfor er muligheten for målefeil til stede. Både *utsikt*, *solforhold*, *stoynivaa* og *turmuligheter* er i dette datasettet rangert ut fra en subjektiv vurdering. Målefeil i den uavhengige variabelen vil gi større utslag for modellens forklaringskraft.

Tidligere forskning gjennomført av Navrud og Strand (2011) finner at støy har en signifikant negativ virkning på boligpriser. Vi har inkludert *stoynivaa* i analyse 4, hvor resultatet blir at

støy har høy økonomisk verdi, men er ikke statistisk signifikant. Dette stammer mest sannsynlig fra målefeil, fordi støynivået var fra Anglevik og Øvretveit (2018) subjektivt rangert fra 0 til 10. Vi endret *stoynivaa* til en dummyvariabel, med 0 og 1, hvor skillet gikk på verdien 5. I tabell 8 ser vi at kun to observasjoner er registrert med mye støy. Det blir derfor stor skjevhet i analysen, fordi kun to av 256 observasjoner er registrert med et høyt støynivå.

Det er også flere variabler som fører til at vi mister en del observasjoner i analysene. Vi har totalt 601 observasjoner, og variablene *avssjo*, *stoynivaa* og *turmuligheter* har under halvparten av dette. *Avssjo* har kun 219 observasjoner, *stoynivaa* har 256 observasjoner og *turmuligheter* har 279 observasjoner. Dette kan føre til skjevhet i analysene ved at vi mister observasjoner når disse variablene inkluderes. Konsekvenser av dette kan være at variabler som skulle vært statistisk signifikante, ikke blir det. Vi kan også få motsatte tilfeller. Den økonomiske verdien kan også bli feil. For eksempel har vi ingen boligobservasjoner med *avssjo* som er omsatt i 2017, som fører til at *omsetningsaar2017* forsvinner i analyser hvor *avssjo* inkluderes som uavhengig variabel.

## 6 Rangering av miljøvariablene

På bakgrunn av resultatene fra analysene med de seks variablene vi har studert, har vi kommet frem til en rangering av disse. Rangeringen er gjort på grunnlag av følgende kriterier; hvilke variabler som er mest signifikant, har best forklaringskraft og har mest økonomisk verdi. Ved å legge til variablene vil vi til slutt få på plass en endelig modell som vil gi oss de implisitte prisene på miljøvariablene vi har inkludert.

Variabelen *stoynivaa* gir som nevnt en høy økonomisk verdi, mens t-verdi og p-verdi ikke tilfredsstiller kravene for signifikansnivå. Det at en variabel ikke er statistisk signifikant kan være avgjørende for om en variabel inkluderes eller ikke, men det kan argumenteres for å inkludere den likevel. Blant annet Navrud og Strand (2011) finner at støy og andre eksterne effekter har en signifikant negativ påvirkning på boligpriser.

Dersom en variabel legges til i modellen og forklaringskraften reduseres, er vanlig praksis at disse ikke inkluderes videre i modellen. Dette er tilfelle for variabelen *stoynivaa*. Det ble prøvd å endre koding for variabelen, men til ingen nytte. Vi velger dermed, på bakgrunn av våre resultater og vårt datasett, å ikke inkludere *stoynivaa* i vår endelige modell.

Ved å inkludere *avssjo* i basemodellen vil det forsvinne 417 observasjoner fra analysen. Det kan føre til at beregningen av de implisitte prisene blir unøyaktig. Variabelen *avssjo* har også en lav økonomisk verdi, samt at den ikke tilfredsstiller signifikansnivået på 5 %. Vi ønsker å opprettholde et høyt antall observasjoner i den endelige modellen, og velger dermed å ikke inkludere *avssjo* i vår endelige modell.

Resultatene ved å inkludere variabelen *turmuligheter* gir et negativt skjæringspunkt i analysen. Dersom *turmuligheter* inkluderes i en endelig modell, vil det føre til at bruken av modellen vil vi ende opp med et feilaktig resultat. Vi velger dermed, på bakgrunn av våre resultat og vårt datasett, å ikke inkludere *turmuligheter* i vår endelige modell.

De resterende variablene er rangert i følgende rekkefølge: *strandlinje*, *solforhold* og *sjofeste*. Vi bygger videre på analyse 16 med å legge til én og én variabel. Til slutt sitter vi igjen med vår endelig modell.

## 6.1 Regresjonsanalyse av basemodellen inkludert strandlinje og solforhold

I analyse 21 er både *strandlinje* og *solforhold* inkludert i analysen av basemodellen. Vi kan se at både *strandlinje* og *solforhold* høy positiv økonomisk verdi, men at ingen av variablene tilfredsstillt signifikansnivået. Forklaringskraften til analysen er på 0,6233 og en R-justert på 0,6088. Vi konstaterer dermed med at ved å inkludere variablene, *strandlinje* og *solforhold* øker forklaringskraften til modellen.

Analyse 21. Regresjonsanalyse av basemodellen inkludert strandlinje og solforhold.

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	519168.7	317276.1	1.636	0.102378	
avsreg	-9032.2	2753.6	-3.280	0.001107	**
prosarb60	803.9	3973.1	0.202	0.839732	
avskom	-6307.2	4280.8	-1.473	0.141259	
primaerrom	7584.1	519.6	14.596	< 2e-16	***
garasje	291819.2	55405.7	5.267	2.04e-07	***
omsetningsaar2008	195834.4	140122.7	1.398	0.162834	
omsetningsaar2009	209527.6	142022.7	1.475	0.140737	
omsetningsaar2010	407863.9	128312.8	3.179	0.001568	**
omsetningsaar2011	506719.6	129778.9	3.904	0.000107	***
omsetningsaar2012	629716.5	122789.4	5.128	4.14e-07	***
omsetningsaar2013	593174.8	120422.6	4.926	1.13e-06	***
omsetningsaar2014	830573.6	105153.2	7.899	1.70e-14	***
omsetningsaar2015	773384.7	149007.1	5.190	3.02e-07	***
omsetningsaar2016	641183.2	191684.4	3.345	0.000882	***
omsetningsaar2017	513650.7	217655.6	2.360	0.018649	*
naust	1142359.0	293951.8	3.886	0.000115	***
boligalder	-10973.7	836.9	-13.112	< 2e-16	***
utsikt	516649.7	152184.8	3.395	0.000739	***
strandlinje	17548.4	10361.2	1.694	0.090930	.
solforhold	186475.7	186121.7	1.002	0.316859	
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 565400 on 518 degrees of freedom (71 observations deleted due to missingness)					
Multiple R-squared: 0.6233, Adjusted R-squared: 0.6088					
F-statistic: 42.86 on 20 and 518 DF, p-value: < 2.2e-16					

Vi bemerker oss at det ikke har vært store endringer i modellen ved å legge til en variabel til. For variablene *avsreg*, *prosarb60* og *avskom* er det små økninger i regresjonskoeffisienten og t-verdi. P-verdi til *avsreg* har økt, men den er redusert for de andre. For *primaerrom* og *garasje* er det en liten reduksjon i koeffisienten og t-verdien, men de opprettholder sitt signifikansnivå på 5 %. Det er ingen endringer av betydning for *omsetningsaar*, bortsett fra for 2016 og 2017. Vi registrerer at 2016 er blitt mer signifikant, mens 2017 er blitt mindre signifikant. Selv med en dårligere p-verdi er variabelen fortsatt signifikant. Variablene *naust*, *boligalder* og *utsikt* får



endringer i regresjonskoeffisienten, t-verdien og p-verdien, men alle variablene er fremdeles statistisk signifikante.

*Analyse 22. VIF-verdier for basemodellen inkludert strandlinje og solforhold.*

	GVI	F	Df
avsreg	5.365805	1	1
prosarb60	5.181344	1	1
avskom	1.100594	1	1
primaerrom	1.357765	1	1
garasje	1.287215	1	1
omsetningsaar	1.412303	10	10
naust	1.338959	1	1
boligalder	1.120889	1	1
utsikt	1.443246	1	1
strandlinje	1.308375	1	1
solforhold	1.553065	1	1

Vi beregner VIF-verdier for å se på om det oppstår multikollinearitet ved å legge til *strandlinje* og *solforhold*. VIF-verdiene viser at det fortsatt kun er hos *avsreg* og *prosarb60* at det foreligger en høy nok verdi til at det kan skape problemer.

## 6.2 Regresjonsanalyse av basemodellen inkludert strandlinje, solforhold og sjofeste

I analyse 23 er variabelen *sjofeste* inkludert analysen med *strandlinje* og *solforhold*. Vi kan se at *sjofeste*, sammen med *strandlinje* og *solforhold*, har en høy økonomisk verdi for modellen. Det som også er felles for de variablene som er lagt til, er at de ikke innfrir signifikansnivået på 5 %. Resultatene fra  $R^2$  og R-justert forteller at forklaringskraften til analysen har økt.

## Analyse 23. Regresjonsanalyse av basemodellen inkludert strandlinje, solforhold og sjofeste.

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	513464.0	317553.0	1.617	0.106501	
avsreg	-8895.8	2762.3	-3.220	0.001361	**
prosarb60	978.9	3983.6	0.246	0.805982	
avskom	-6434.7	4287.2	-1.501	0.133983	
primaerrom	7610.2	521.3	14.599	< 2e-16	***
garasje	291451.2	55437.3	5.257	2.14e-07	***
omsetningsaar2008	184272.0	141228.3	1.305	0.192548	
omsetningsaar2009	204253.6	142309.4	1.435	0.151812	
omsetningsaar2010	401746.5	128696.1	3.122	0.001899	**
omsetningsaar2011	507221.9	129848.7	3.906	0.000106	***
omsetningsaar2012	625418.0	123016.8	5.084	5.17e-07	***
omsetningsaar2013	588896.4	120650.4	4.881	1.41e-06	***
omsetningsaar2014	830557.2	105208.1	7.894	1.76e-14	***
omsetningsaar2015	772956.4	149086.2	5.185	3.11e-07	***
omsetningsaar2016	631368.6	192329.5	3.283	0.001098	**
omsetningsaar2017	511501.2	217792.2	2.349	0.019221	*
naust	1036160.6	333181.8	3.110	0.001975	**
boligalder	-10976.6	837.3	-13.109	< 2e-16	***
utsikt	509799.4	152598.7	3.341	0.000896	***
strandlinje	16522.9	10476.3	1.577	0.115367	
solforhold	178105.0	186627.2	0.954	0.340360	
sjofeste	122412.0	180467.1	0.678	0.497881	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 565700 on 517 degrees of freedom  
(71 observations deleted due to missingness)  
Multiple R-squared: 0.6236, Adjusted R-squared: 0.6084  
F-statistic: 40.79 on 21 and 517 DF, p-value: < 2.2e-16

Ved å inkludere *sjofeste* har det også totalt sett vært små endringer fra analyse 16, og mange av de samme variablene er endret i samme retning som den analysen. For variablene *avsreg*, *prosarb60* og *primaerrom* har det vært en liten økning i regresjonskoeffisient og t-verdi. P-verdi har økt for *avsreg*, men fått minimal reduksjon for de to andre variablene. Disse endringene er små og av liten betydning i det store bildet, spesielt når ingen er blitt mer eller mindre statistisk signifikant. Variablene *avskom* og *garasje* har fått en liten reduksjon i regresjonskoeffisient og t-verdi. P-verdi til *avskom* er redusert, men har fått en økning for *garasje*. Også for disse variablene har endringene liten betydning for variabelen og har ikke påvirket hvor statistisk signifikant de er. Det er i likhet med de fleste andre variabler, ingen endringer av betydning for *omsetningsaar*, bortsett fra *2016*. Der har variabelen mistet litt statistisk signifikant, og er tilbake på omtrent samme nivå som før vi inkluderte *solforhold* i analyse 21. Regresjonskoeffisienten til *naust* har hatt en reduksjon, men er fremdeles den variabelen med høyest økonomisk verdi. T-verdi har hatt en reduksjon og p-verdi en økning,

men dette påvirker ikke variabelens statistiske signifikans. *Boligalder* og *utsikt* får små endringer ved å inkludere *sjofeste*, og deres statistiske signifikans er dermed ikke blitt påvirket.

*Analyse 24. VIF-verdier for basemodellen inkludert strandlinje, solforhold og sjofeste.*

	GVI	F	Df
avsreg	5.394381		1
prosarb60	5.203169		1
avskom	1.102716		1
primaerrom	1.365217		1
garasje	1.287339		1
omsetningsaar	1.469498	10	
naust	1.718402		1
boligalder	1.120919		1
utsikt	1.449595		1
strandlinje	1.336203		1
solforhold	1.559885		1
sjofeste	1.580043		1

Vi sjekker om mangelen på statistisk signifikans blant miljøvariablene skyldes multikollinearitet. I analyse 24 observeres det ingen VIF-verdier som skulle tilsi at det foreligger multikollinearitet i analysen, bortsett fra de vi kommenterte i analyse 2.

## 7 Drøfting av endelig modell

Etter å ha analysert de miljøvariablene vi ønsket å studere, har disse nå blitt rangert slik at vi til slutt har fått vår endelige modell. Den tar utgangspunkt i Anglevik og Øvretveit (2018) sin basemodell, og vi inkluderer variablene; *strandlinje*, *solforhold* og *sjofeste*.

Vår endelige modell blir dermed seende slik ut:

$$P = \beta_0 + \beta_1 * avsreg + \beta_2 * prosarb60 + \beta_3 * avskom + \beta_4 * primaerrom + \beta_5 * garasje + \beta_6 * omsetningsaar + \beta_7 * naust + \beta_8 * boligalder + \beta_9 * utsikt + \beta_{10} * strandlinje + \beta_{11} * solforhold + \beta_{12} * sjofeste$$

Ingen av de tre miljøvariablene som er inkludert er statistisk signifikante. Vi kan dermed ikke konkludere med at *strandlinje*, *solforhold* og *sjofeste* har betydning for boligprisen. Årsaken til at disse variablene ikke er signifikante, kan skyldes flere ting. Det er etter vår vurdering tenkelig at en av årsakene for *strandlinje* og *sjofeste* kan være mangelen på observasjoner, samt målefeil. Det er kun fire observasjoner som har *strandlinje* og 17 observasjoner som har *sjofeste*. Det kan skape skjevheter for datasettet, og det blir vanskelig å tilfredsstille signifikansnivået på 5 %. Angående *solforhold* sin manglende signifikans, diskuterer både Anglevik og Øvretveit (2018) og Hovland, Kvalvåagnes og Kvalevåg (2019) hva dette kan skyldes. Sistnevnte trekker frem at *solforhold* er blitt kodet ved å bruke en subjektiv vurdering på en skala fra 0,0 – 1,0, men at det argumentet faller litt gjennom med tanke på at *utsikt* er kodet på samme måte. Anglevik et al. (2018) trekker frem at det er lite sannsynlig at de subjektive vurderingene av hver enkelt observasjon knyttet til *solforhold* og *utsikt* ville avvike mellom f.eks. 0,3 og 0,9, og at dette dermed var et redusert problem. Hovland et al. (2019) argumenterer derimot for at dette blir et problem da de subjektive vurderingene samler seg rundt 0,5 og 0,7, og åpner dermed for stor usikkerhet knyttet til de subjektive kodingene.

Det vi derimot kan være mer sikre på, er at ved å inkludere disse variablene retter vi opp en del av skjevheten som var i variablene *utsikt* og *naust* i basemodellen. Med tanke på at disse variablene korrelerer med de variablene vi har lagt til, bør vi kunne hevde at *utsikt* og *naust* fanger opp de nye variablene. Dermed blir koeffisienten og den implisitte prisen for *utsikt* og *naust* for høy i basemodellen.

Vi kan se i tabell 14 at variablene *utsikt* og *solforhold* har en moderat korrelasjon. Sammenligner vi resultatene i vår endelige modell, analyse 23, med resultatene i analyse 1 kan vi se at den implisitte prisen på *utsikt* er redusert med om lag 100 000. Ved å inkludere *solforhold*, kan vi hevde den implisitte prisen til *utsikt* representerer mer seg selv.

For *naust* kan vi i tabell 14 se at korrelasjonen til *strandlinje* og *sjofeste* er moderat. Ved å inkludere dere disse i vår endelige modell kan vi se av den implisitte prisen til *naust* er redusert. Vi kan også her hevde at *naust* fanger opp færre variabler i vår endelig modell, og at det er grunnen til at den implisitte prisen er redusert.

Variabelen *avsreg* er statistisk signifikant, men det er ikke *prosarb60* og *avskom*. Regresjonskoeffisienten til *avsreg* er -8 896, som vil si at dersom avstanden til regionsenter økes med ett minutt vil verdien på en bolig reduseres med 8 896 kroner. Regresjonskoeffisienten til *prosarb60* forteller oss hvor mye en bolig vil øke eller redusere i verdi dersom prosentandelen av arbeidsplasser som nås innen 60 minutter med bil endres med 1 %. Regresjonskoeffisienten til variabelen er på 979 og den har lav økonomisk verdi. Koeffisienten betyr at verdien på en bolig øker med 979 kroner dersom andelen arbeidsplasser som nås innen 60 minutter med bil, øker med 1 %. Den siste avstandsvariabelen er *avskom*, som har en regresjonskoeffisient på -6 435. Dette forteller oss at dersom avstanden til kommunesenteret øker med 1 minutt vil verdien på en bolig reduseres med 6 435 kroner.

Både *primaerrom* og *garasje* har en statistisk signifikant innvirkning på boligprisene. Regresjonskoeffisienten til primærrom er 7 610. Det vil si, at dersom alt annet holdes konstant, vil prisen på en bolig øke med 7 610 kroner dersom primærrommet øker med én kvadratmeter. *Garasje* har høy økonomisk verdi og har en regresjonskoeffisient på 291 541. Dersom en bolig har garasje, vil prisen øke med 291 451 kroner. *Boligalder* er statistisk signifikant og variabelen har en regresjonskoeffisient på -10 977. Dette betyr at dersom alderen til en bolig øker med ett år, vil prisen reduseres med 10 977 kroner.

Regresjonskoeffisienten til *naust* er 1 036 161, som forteller oss at dersom en bolig har naust vil den øke i pris med 1 036 161 kroner. *Strandlinje* har en regresjonskoeffisient på 16 523. En bolig vil øke med 16 523 kroner i pris for hver meter med strandlinje den har. Regresjonskoeffisienten til *sjofeste* er 122 412, som betyr at dersom en bolig har sjøfeste, vil prisen øke med 122 412 kroner.

For *utsikt* er regresjonskoeffisienten på 509 799 og for *solforhold* er den 178 105. Siden både *utsikt* og *solforhold* er rangert fra skala 0,0 til 1,0. Det betyr at hvis en bolig er rangert som 1,0 på utsikt og solforhold vil boligens verdi øke med 509 799 kroner og 178 105 kroner.

Vi har ikke satt inn regresjonskoeffisienten til omsetningsår i den hedoniske boligprismodellen. Dette kommer av at regresjonskoeffisientene for hvert år er ulike, og en observasjon er solgt kun en gang. Variabelen *omsetningsaar2007* er utelatt fra modellene fordi den er brukt som referanseår. Når vi skal sette inn en implisitt pris for omsetningsår i den endelige modellen, velger vi regresjonskoeffisienten for året boligen ble solgt. Dersom en bolig ble solgt i 2011 vil boligprisen øke med 507 222 kroner.

Ved å ta i bruk de implisitte prisene fra analyse 23, vil vår endelig modell se slik ut:

$$\begin{aligned} P = & 513\,464 - 8\,896 * \textit{avsreg} + 979 * \textit{prosarb60} - 6\,435 * \textit{avskom} + 7\,610 \\ & * \textit{primaerrom} + 291\,451 * \textit{garasje} + X * \textit{omsetningsaar} + 1\,036\,161 \\ & * \textit{naust} - 10\,977 * \textit{boligalder} + 509\,799 * \textit{utsikt} + 16\,523 * \textit{strandlinje} \\ & + 178\,105 * \textit{solforhold} + 122\,412 * \textit{sjofeste} \end{aligned}$$

## 8 Konklusjon

Formålet med denne oppgaven var å finne de implisitte prisene for miljøvariabler. Vi hadde på forhånd antagelser om hvordan miljøvariablene ville påvirke boligprisene, og fant flere artikler som underbygget disse. Vi hadde blant annet en oppfatning om at mye støy ville føre til en reduksjon i boligpris.

Gjennom oppgaven har vi analysert de valgte variablene med basemodellen til Anglevik og Øvretveit (2018). Vi analyserte først hver enkelt variabel, slik at vi fikk oversikt over hvordan hver enkelt påvirket basemodellen. Vi kunne deretter lage en rangering som ville bli utgangspunktet for vår endelige modell. Rangeringen ble gjort med utgangspunkt i hvilke variabler som var mest signifikant, hadde høyest økonomisk verdi og som økte forklaringskraften i modellen mest. Rangeringen ble slik; strandlinje, solforhold og til slutt sjøfeste. Tre variabler ble utelatt fra den endelige modellen, og disse var støy, turmuligheter og avstand til sjø. Dette er på grunn av at de enten reduserte forklaringskraft, førte til tap av over halvparten av observasjonene eller at den ga negativt konstantledd.

Analysene viste at ingen av variablene var statistisk signifikante, men flere av dem var likevel viktige å inkludere i analysen. Vi valgte å inkludere *strandlinje*, *solforhold* og *sjøfeste* på grunnlag av at disse økte forklaringskraften og hadde høy økonomisk verdi. Selv om de ikke var statistisk signifikante, korrelerte de med variabler fra basemodellen som var statistisk signifikante. *Naust* korrelerte med både *strandlinje* og *sjøfeste*, og vår oppfatning var at deres verdi ble fanget opp i den implisitte prisen til *naust* ved å ekskludere dem fra analysen. Ved å inkludere både *strandlinje* og *sjøfeste* førte det til at den økonomiske verdien til *naust* ble redusert. Situasjonen ble den samme for *utsikt* ved å inkludere *solforhold*. Ved å inkludere de tre variablene fikk vi vår endelige modell og variablenes implisitte priser.

Både korrelasjon og målefeil, kan føre til at resultater i modellen blir feilspesifisert. Fra tidligere forskning er flere av variablene vi undersøker statistisk signifikante, men med vårt datasett ble ikke variablene signifikante. For eksempel, variabelen *støynivaa*, er subjektivt vurdert og blir ikke statistisk signifikant i våre analyser, men vi vet fra Navrud og Strand (2011) at støy har en signifikant negativ innvirkning på boligpriser.

Det er vanskelig å konkludere med hva de implisitte prisene for miljøvariablene blir i modellen, på grunn av at vår modell mest sannsynlig er feilspesifisert. Det vi derimot kan konkludere

med, er at ved å inkludere *strandlinje*, *solforhold* og *sjofeste* retter det opp en del av skjevhetene til *naust* og *utsikt*. De implisitte prisene for *naust* og *utsikt* representerer i større grad seg selv, og fanger ikke opp *strandlinje*, *solforhold* og *sjofeste*. Selv om regresjonskoeffisientene til *naust* og *utsikt* er redusert, har de fremdeles høy økonomisk verdi. De er også statistisk signifikante og vi kan derfor med 95 % sikkerhet si at når disse variablene er til stede, vil de øke verdien på boligen betydelig. Vi kan ikke si noe om den økonomiske verdien for *strandlinje*, *solforhold* eller *sjofeste* på grunnlag av at disse, i vår modell, ikke er statistisk signifikante. Det vi tar med oss er videre, er at fortegnene til variablenes regresjonskoeffisient stemmer overens med våre antagelser og hva tidligere forskning sier.



## 9 Videre forskning

Da vi startet arbeidet med oppgaven hadde vi planer om å bruke QGIS for å innhente data om utsikt. QGIS ville gjort det mulig å innhente data på alle observasjonene og målet var da å innhente en mer objektiv vurdering av hvilken utsikt en bolig hadde. En mulighet hadde da vært å teste og sammenligne resultatene mellom de objektive og de subjektive vurderingene. Ettersom situasjonen i samfunnet ble som den ble, strandet hele planen om å bruke QGIS og målet forflyttet seg til å bruke det datasettet vi hadde.

Det hadde vært interessant å sett om en kunne innhente objektive data fra flere variabler, som *turmuligheter* og *solforhold*. En kunne også her ha sammenlignet resultatene fra de objektive vurderingene med de subjektive vurderingene, og sett på om det endret signifikansnivået til variablene.

Skulle vi gjort oppgaven på ny, hadde vi nok prøvd å samle inn data på de variablene som manglet data, samt at vi hadde prøvd å få tak i flere observasjoner. Det kunne ha gjort at resultatene hadde sett annerledes ut, og resultatene kunne ha gitt et enda mer korrekt bilde av hva som påvirker boligprisene. Vi vet for eksempel av støynivå og turmuligheter har påvirkning på boligprisen, men med vårt datasett svekket det mer modellen, mer enn det forbedret den.

## 10 Litteraturliste

- Allison, P. D. (1999). *Multiple Regression. A Primer*. Thousand Oaks, CA: Pine Forge Press.
- Amundsen, A., & Aasvang, G. (2006). *Trafikkstøy i boliger. Virkninger av fasadeisolerings tiltak etter grenseverdiforskriften (TØI rapport 836/2006)*. Oslo: Transportøkonomisk Institutt.
- Anglevik, R., & Øvretveit, T. (2018). *Romlige boligmarkedsvirkninger av T-forbindelsen (Bachelorgrad)*. Høgskulen på Vestlandet, Haugesund.
- Arbeidstilsynet. (u.d.). *Støy*. Hentet fra <https://www.arbeidstilsynet.no/tema/stoy/>
- Berge Sag AS. (u.d.). *Ferdig bolig*. Hentet fra <https://www.bergesag.no/ferdig-bolig/oversikt>
- Berge, E. (2004). *Anvendt statistisk dataanalyse i samfunnsvitenskap*. Hentet fra <http://www.erlingberge.no/SOS3003L05H04.pdf>
- Bhuller, M. S. (2009). *Inndeling av arbeidsmarkeder i Norge*. Hentet fra Statistisk Sentralbyrå: [https://www.ssb.no/a/publikasjoner/pdf/notat\\_200924/notat\\_200924.pdf](https://www.ssb.no/a/publikasjoner/pdf/notat_200924/notat_200924.pdf)
- Bourassa, S., Hoesli, M., & Sun, J. (2004). What's in a view? I *Environment & Planning A* (Vol. 36(8), ss. 1427-1450).
- Conroy, S., & Milosch, J. (2011). An Estimation of the Coastal Premium for Residential Housing Prices in San Diego County. I *The Journal of Real Estate Finance and Economics* (Vol. 42(2), ss. 211-228).
- Eiendom Norge. (2020). *Prisutvikling*. Hentet fra <https://eiendommnorge.no/boligprisstatistikk/statistikkbank/>
- Fleming, D., Grimes, A., Lebreton, L., Maré, D., & Nunns, P. (2017). *Valuing Sunshine*. I *Motu Working Paper 17-13*. New Zealand: Motu Economic and Public Policy Research.
- Freeman, A., Herriges, J., & Kling, C. (2014). *The Measurement of Environmental and Resource Values (3. utg.)*. New York: PFF PRes.
- Fugleberg, O., & Kristianslund, I. (1995). *Innføring i regresjonsanalyse og multivariate metoder*. Oslo: Bedriftsøkonomens Forlag A/S.
- Green, C., Tunstall, S., N'Jai, A., & Rogers, A. (1990, Juni). Economic evaluation of environmental goods. *Project Appraisal*, 5(2), ss. 70-82.
- Grønmo, S. (2018, Februar 20). *Validitet*. Hentet fra Store Norske Leksikon: <https://snl.no/validitet>
- Haugalandet\_no. (u.d.). *Haugalandet*. Hentet fra [https://twitter.com/haugalandet\\_no](https://twitter.com/haugalandet_no)

- Hovland, M. N., Kvalvågnes, M., & Kvalevåg, J. B. (2019). *Eiendomsskatt, kommunale avgifter og radons påvirkning på boligpriser på Haugalandet (Bachelorgrad)*. Høgskulen på Vestlandet, Haugesund.
- Idsø, J. (2014, Mai 27). *Kollektive goder*. Hentet fra Store Norske Leksikon: [https://snl.no/kollektive\\_goder](https://snl.no/kollektive_goder)
- Jansen, E. S. (2011, Desember 5). *Hva driver utviklingen i boligprisene?* Hentet fra Statistisk Sentralbyrå: <https://www.ssb.no/priser-og-prisindekser/artikler-og-publikasjoner/hva-driver-utviklingen-i-boligprisene>
- Johannessen, A., Christoffersen, L., & Tufte, P. A. (2011). *Forskningsmetode for økonomisk-administrative fag (3. utg.)*. Oslo: Abstrakt forlag AS.
- Kennedy, P. (2003). *A Guide to Econometrics*. Cambridge: MIT Press.
- Kotler, P., & Keller, K. L. (2016). *Markedsføringsledelse (4. utg.)*. Oslo: Gyldendal Akademisk.
- McCann, P. (2013). *Modern Urban and Regional Economics (2. utg.)*. Oxford: Oxford University Press.
- Moser, B. K. (1996). Linear Models. A Mean Model Approach. I *Probability and Mathematical Statistics*. Academic Press.
- Navrud, S. (2016). Miljøversetting. Verdsettingsmetoder og verdioverføring. I K. Hagen, & G. Volden (red.), *Investeringsprosjekter og miljøkonsekvenser* (ss. 26-54). Trondheim: Ex ante akademisk forlag.
- Navrud, S., & Strand, J. (2011). Using hedonic pricing for estimating compensation payment for noise and other externalities from new roads. I J. Bennett (red.), *The International Handbook on Non-Market Environmental Valuation* (ss. 14-36). Cheltenham: Edward Elgar Publishing.
- NOU 2002:2. (2002, Februar 15). *Boligmarkedene og boligpolitikken*. Hentet fra <https://www.regjeringen.no/contentassets/80899d9e55ef499c86359694e816207f/no/pdfa/nou200220020002000dddpdfa.pdf>
- NOU 2011:3. (2011, Mars 9). *Kompetansearbeidsplasser - drivkraft for vekst i hele landet*. Hentet fra <https://www.regjeringen.no/contentassets/a90bf4e1e5f740ea9f810ff4e1519fb4/no/pdfs/nou201120110003000dddpdfs.pdf>
- NOU 2012:16. (2012, Oktober 3). *Samfunnsøkonomiske analyser*. Hentet fra <https://www.regjeringen.no/contentassets/5fce956d51364811b8547eebdbcde52c/no/pdfs/nou201220120016000dddpdfs.pdf>

- O'Sullivan, A. (2007). *Urban Economics*. Boston: McGraw-Hill/Irwin.
- Osland, L. (2001). Den hedoniske metoden og estimering av attributtpriser. *Norsk økonomisk tidsskrift*, 115(1), ss. 1-22.
- Osland, L. (2016). Verdsetting av lokale miljøgoder ved bruk av hedoniske priser. I K. Hagen, & G. Volden (red.), *Investeringsprosjekter og miljøkonsekvenser* (ss. 55-70). Trondheim: Ex ante akademisk forlag.
- Rinehart, J., & Pompe, J. (1994). Adjusting the Market Value of Coastal Property for Beach Quality. I *The Appraisal Journal* (Vol. 62(4), ss. 604-409). Chicago: Appraisal Institute.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets. Product Differentiation in Pure Competition. I *The Journal of Political Economy* (Vol. 82(1), ss. 24-55). Chicago: The University of Chicago Press.
- RStudio. (u.d.). *RStudio*. Hentet fra <https://rstudio.com/products/rstudio/>
- Senneset, K. (2015, Desember 11). *Boligprisene og oljeprisfall*. Hentet fra Prognosesenteret: <https://prognosesenteret.no/boligprisene-og-oljeprisfall/>
- Sirmans, G., Machperson, D., & Zietz, E. (2005). The Composition of Hedonic Pricing Models. I *Journal of Real Estate Literature* (Vol. 13(1), ss. 3-43). American Real Estate Society.
- Smith, L., Rosen, K., & Fallis, G. (1988). Recent Developments in Economic Models of Housing Markets. I *Journal of Economic Literature* (Vol. 26(1), ss. 29-64). American Economic Association.
- Statistics How To. (2015, September 21). *Variance Inflation Factor*. Hentet fra <https://www.statisticshowto.com/variance-inflation-factor/>
- Statistisk Sentralbyrå. (u.d.). 06035: *Selveierboliger. Gjennomsnittlig kvadratmeterpris og antall omsetninger (K) 2002 - 2019*. Hentet fra <https://www.ssb.no/statbank/table/06035/chartViewLine/>
- Statistisk Sentralbyrå. (u.d.). 06265: *Boliger, etter bygningstype (K) 2006 - 2020*. Hentet fra <https://www.ssb.no/statbank/table/06265/chartViewLine/>
- Statistisk Sentralbyrå. (u.d.). 07459: *Alders- og kjønnsfordeling i kommuner, fylker og hele landets befolkning (K) 1986 - 2020*. Hentet fra <https://www.ssb.no/statbank/table/07459/tableViewLayout1/>,
- Statistisk Sentralbyrå. (2020, April 16). *Boliger*. Hentet fra <https://www.ssb.no/bygg-bolig-og-eiendom/statistikker/boligstat?fane=om>
- Stock, J., & Watson, M. (2007). *Introduction to Econometrics* (2. utg.). Addison Wesley Publishing Company.

- Sundberg, J. D. (2015, November 27). *I 10 norske kommuner er ledigheten 5 prosent eller mer*. Hentet fra E24: <https://e24.no/norsk-oekonomi/i/GGreJ4/i-10-norske-kommuner-er-ledigheten-fem-prosent-eller-mer>).
- Thorsnæs, G. (2017, August 20). *Haugalandet*. Hentet fra Store Norske Leksikon: <https://snl.no/Haugalandet>
- Thrane, C. (2017). *Regresjonsanalyse. En praktisk tilnærming*. Oslo: Cappelen Damm AS.
- Tyrväinen, L. (1997). The amenity value of the urban forest: an application of the hedonic pricing method. I *Landscape and Urban Planning* (Vol. 37(3-4), ss. 211-222). Elsevier BV.
- Walope, R., Myers, R., Myers, S., & Ye, K. (2012). *Probability & Statistics for Engineers and Scientists (9. utg.)*. Boston: Pearson Education Limited.
- Wikipedia. (2020, Mars 15). *1,96*. Hentet fra <https://en.wikipedia.org/wiki/1.96>
- Wooldridge, J. M. (2016). *Introductory Econometrics. A Modern Approach (6. utg.)*. Boston: Cengage Learning.