



Big Data and Little Norway

En empirisk studie av mulige sammenhenger mellom Google-søk og boligprisen

Haakon Resaland og Ingvild Kristoffersen Rooth

Veileder: Ola Honningdal Grytten

Masterutredning, Institutt for Samfunnsøkonomi

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i økonomi- og administrasjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer innestår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.

Sammendrag

Denne masterutredningen vil utforske sammenhengen mellom Google-søk og det norske boligmarkedet. Vi har benyttet Google-søkedata til å konstruere åtte variabler basert på søkeord relatert til boligmarkedet. Variablene reflekterer individenes samlede forventning til egen handling i boligmarkedet i fremtiden. Sentrale makroøkonomiske variabler er valgt ut for å indikere fundamentale verdier i den norske økonomien.

Datagrunnlaget er basert på Google-data fra Google Trends, sentrale makroøkonomiske variabler benyttet til boligprismodeller fra Norges Bank og boligprisindeksens verdi over tid fra Eiendom Norge. Den analyserte perioden strekker seg fra 2007 til 2017.

Vi har estimert seks modeller for å studere sammenhengen mellom Google-søk og boligprisen. Modellene er estimert ved å benytte minste kvadraters metode, for perioden 2007 til 2017 og for 2007 til 2016, sistnevnte for å undersøke perioden før boliglånsforskriftene ble endret. Teori om aktivabobler danner det teoretiske grunnlaget for masterutredningen.

Resultatene viser at Google-søk ikke er godt egnet til å forklare boligprisutviklingen i Norge alene, men at de gjør en god jobb dersom de kombineres med makroøkonomiske verdier. Den estimerte modellen med Google-søk er bedre til å forutse boligprisfall enn modellen med makroøkonomiske faktorer, og kan i tillegg benyttes til å identifisere euforiske bobletendenser i boligmarkedet.

Abstract

This master thesis explores the potential of digital searches in predicting growth in the Norwegian real estate market. Based on similar studies from other countries this thesis has constructed six models in order to test Googles ability to explain, forecast and predict bubbles in the Norwegian market.

Our findings somewhat support our hypothesis that Google queries can perform better than macroeconomic variables in explaining the movements of the real estate market. Google variables was not as able to accurately explain the price index, but seemed able to predict future falls in the index better than the baseline model. We also find that a hybrid model using both macroeconomic and Google variables outperform either “clean” model. Finally, we found some evidence that studying the difference between the two “clean” models could indicate bubble tendencies in the market.

Forord

Denne masterutredningen er skrevet som en del av masterstudiet i økonomi og administrasjon ved Norges Handelshøyskole (NHH). Utledningen inngår som et selvstendig arbeid i vår hovedprofil Economics, og utgjør 30 studiepoeng. Arbeidet ble gjennomført våren 2018.

I løpet av vårt studieløp ved NHH har vi fulgt kurs som har omhandlet boligmarkedet. Disse kursene vekket vår interesse for temaet. *Big Data* er svært aktuelt i disse dager, og vi kunne ikke la muligheten til å undersøke den spennende sammenhengen mellom boligmarkedet og *Big Data* gå fra oss. Interessen rundt boligprisen er svært stor i Norge, og potensielle nye estimeringsmetoder for boligprisen er derfor alltid spennende å undersøke. Dette gjorde oss ekstra motiverte for å besvare oppgavens problemstilling.

Vi vil rette en stor takk til vår engasjerte veileder Ola Honningdal Grytten, som har gitt konstruktive og svært raske tilbakemeldinger gjennom hele semesteret. Vi setter stor pris på interessen han har vist for oppgaven, og hans tilgjengelighet har vært over all forventning. Vi ønsker i tillegg å takke Erling Risa for faglige innspill, samt Bjørn Erik Naug i Norges Bank for tilsendt datasett og Renate Berentsen i DNB Eiendom for raske svar på spørsmål relatert til emnet.

Videre ønsker vi å takke venner og familie for deres støtte i masterutredningsprosessen. En spesiell takk rettes til Rebekka og Audun. Til slutt vil vi takke NHH og NHHS for noen fantastiske år i Bergen.

Bergen, juni 2018

Haakon Resaland

Ingvild Kristoffersen Rooth

Innholdsfortegnelse

SAMMENDRAG	2
ABSTRACT	3
FORORD	4
INNHALDSFORTEGNELSE	6
FIGURLISTE	10
TABELLISTE	11
1. INTRODUKSJON	12
1.1 MÅL.....	12
1.2 BAKGRUNN	12
1.3 RELEVANS	13
1.4 MASTERUTREDNINGENS STRUKTUR	13
2. TEORETISK FUNDAMENT	15
2.1 DEFINISJONER	15
2.1.1 <i>Aktivaboble</i>	15
2.1.2 <i>Euforiske og ikke-euforiske bobler</i>	16
2.1.3 <i>Krakk</i>	16
2.2 SYV-TRINNS DYNAMISKE KRISEMODELL	17
3. LITTERATURGJENNOMGANG	20
3.1 GOOGLE-SØK BENYTTET TIL PREDIKSJON	20
3.1.1 <i>Prediksjoner om nåtiden</i>	20
3.1.2 <i>Prediksjoner om fremtiden</i>	21
3.1.3 <i>Kritikk av bruk av Google-søk til prediksjon</i>	21
3.2 GOOGLE-SØK BENYTTET TIL PREDIKSJON AV BOLIGMARKEDET	22
3.2.1 <i>Prediksjon av internasjonale boligmarkeder</i>	22
3.2.2 <i>Forbedring av tradisjonelle makroøkonomiske modeller</i>	23

3.3	FORSKINGSPØRSMÅL.....	24
4.	DATASETT OG UTVELGELSE AV VARIABLER.....	26
4.1	DATAKILDER OG BESKRIVELSE AV DATASETTET.....	26
4.1.1	<i>Boligprisindeksen fra Eiendom Norge.....</i>	<i>26</i>
4.1.2	<i>Google-variable fra Google Trends.....</i>	<i>27</i>
4.1.3	<i>Makroøkonomiske variable fra Norges Bank.....</i>	<i>28</i>
4.2	UTVELGELSE AV VARIABLER.....	28
4.2.1	<i>Kriterier for utvelgelse av Google-søkeord.....</i>	<i>29</i>
5.	DATADESKRIPSJON.....	32
5.1	BOLIGPRISINDEKSEN.....	32
5.2	UTVALGTE GOOGLE-VARIABLER.....	33
5.2.1	<i>Etterspørselsdrevde variable.....</i>	<i>33</i>
5.2.2	<i>Tilbudsdrevde variable.....</i>	<i>35</i>
5.3	UTVALGTE MAKROØKONOMISKE VARIABLER.....	37
6.	EMPIRISK METODE.....	39
6.1	VALG AV METODE.....	39
6.2	REGRESJONSMODELLER.....	41
6.2.1	<i>Regresjonsmodell med Google-variable.....</i>	<i>41</i>
6.2.2	<i>Regresjonsmodell med makroøkonomiske variable.....</i>	<i>41</i>
6.2.3	<i>Regresjonsmodell med både Google- og makrovariable.....</i>	<i>42</i>
6.3	KONVERTERING FRA KVARTAL TIL MÅNED.....	43
6.4	UTFORDRINGER MED TIDSSERIEDATA.....	43
6.4.1	<i>Ikke-stasjonære data.....</i>	<i>43</i>
6.4.2	<i>Autokorrelasjon.....</i>	<i>47</i>
6.4.3	<i>Heteroskedastisitet.....</i>	<i>48</i>

7.	EMPIRISK ANALYSE	49
7.1	FORVENTNINGSMODELLER	49
7.1.1	<i>Regresjonsresultater med forventningsmodellene.....</i>	<i>50</i>
7.1.2	<i>Forventningsmodeller benyttet til å forutsi boligprisen</i>	<i>55</i>
7.2	FUNDAMENTALMODELLER	57
7.2.1	<i>Regresjonsresultater med fundamentalmodellene.....</i>	<i>58</i>
7.2.2	<i>Fundamentalmodeller benyttet til å forutsi boligprisen</i>	<i>62</i>
7.3	HYBRIDMODELLER	64
7.3.1	<i>Regresjonsresultater med hybridmodellene</i>	<i>65</i>
7.3.2	<i>Hybridmodeller benyttet til å forutsi boligprisen</i>	<i>68</i>
7.4	OPPSUMMERING	70
8.	ANALYSE AV BOBLETENDENSER	72
8.1	ANALYSE AV TOPPUNKTER BASERT PÅ EMPIRISKE RESULTATER	72
8.2	ANALYSE AV TOPPUNKTET I 2017 BASERT PÅ TEORI	74
9.	DISKUSJON	76
9.1	DRØFTING AV FORSKNINGSSPØRSMÅLENE	76
9.1.1	<i>Forskningsspørsmål 1</i>	<i>76</i>
9.1.2	<i>Forskningsspørsmål 2</i>	<i>78</i>
9.1.3	<i>Forskningsspørsmål 3</i>	<i>80</i>
9.2	BEGRENSNINGER	80
9.3	FORSLAG TIL VIDERE ARBEID	81
10.	KONKLUSJON	82
	REFERANSER	85
	APPENDIKS A – DATASETT OG VARIABLER	90
	A1: HVORDAN GOOGLE TRENDS FUNGERER	90

A2: HVORDAN GOOGLE ADWORDS FUNGERER	91
A3: OVERSIKT OVER SAMTLIGE GOOGLE-SØKEORD VI HAR VURDERT.....	92
A4: BESKRIVELSE AV VARIABLER BENYTTET I MASTERUTREDNINGEN.....	97
A5: MAILKORRESPONDANSE MED DNB EIENDOM.....	98
A6: JUSTERING AV VARIABLER.....	101
APPENDIKS B – UAVHENGIGE VARIABLER	102
B1: GRAFISK FREMSTILLING AV GOOGLE-VARIABLENE	102
B2: GRAFISK FREMSTILLING AV MAKROVARIABLENE	105
APPENDIKS C – REGRESJONSUTSKRIFTER	108
C1: REGRESJONSRESULTATER MED FORVENTNINGSMODELLEN	108
C2: REGRESJONSRESULTATER MED FUNDAMENTALMODELLEN.....	109
C3: REGRESJONSRESULTATER MED HYBRIDMODELLEN.....	110

Figurliste

Figur 1: Boligprisindeksen, kilde: Eiendom Norge	32
Figur 2: Trend- og sesongjustert prisindeks. Årstallet på x-aksen er satt til januar det respektive året. Kilde: Eiendom Norge	46
Figur 3: Residualverdien til modell 3, 4 og 5 mot et etterslep av residualet på en periode. ..	47
Figur 4: Prisindeksen og Forventningsmodellen for 2007 til 2017. Årstallene på x-aksen indikerer januar det respektive året.	55
Figur 5: Prisindeksen og Forventningsmodellen for 2007 til 2016. Årstallene på x-aksen indikerer januar det respektive året.	57
Figur 6: Boligprisindeksen og Lønn i perioden 2007 til 2017. Årstallet på x-aksen indikerer januar det respektive året.....	59
Figur 7: Boligprisindeksen og estimert fundamentalmodell for perioden 2006 til 2017. Årstallet på x-aksen er satt til januar det respektive året.....	63
Figur 8: Boligprisindeksen og estimert fundamentalmodell for perioden 2006 til 2016. Årstallet på x-aksen er satt til januar det respektive året.....	64
Figur 9: Estimert hybridmodell og boligprisindeksen for 2006 til 2017. Årstallet på x-aksen er satt til januar det respektive året.....	69
Figur 10: Estimert hybridmodell og boligprisindeksen for 2006 til 2016. Årstallet på x-aksen er satt til januar det respektive året.....	70
Figur 11: Estimert fundamentalmodell og forventningsmodell, samt boligprisindeksen, for perioden 2007 til 2017. Årstallene på x-aksen indikerer januar det respektive året.	73

Tabelliste

Tabell 1: Kriterier for utvalgelse av Google-søkeord.	29
Tabell 2: Konstruerte Google-variabler	33
Tabell 3: Regresjonsresultater for forventningsmodellen, for data for hele perioden (Forventning) og frem til og med desember 2016 (Forventning2016).	54
Tabell 4: Estimert Fundamentalmodell og Fundamentalmodell2016.	62
Tabell 5: Estimert hybridmodell for 2007 til 2016 (Hybrid) og for 2007 til 2016 (Hybrid2016).	68

1. Introduksjon

1.1 Mål

Denne oppgaven har som hensikt å besvare følgende problemstilling:

«Kan Google-søk benyttes til å forklare prisutviklingen i det norske boligmarkedet?»

Med «forklare» over mener vi om Google-søk i form av uavhengige variabler kan forklare boligprisindeksen i regresjonsmodeller. Individens forventning til egen fremtidig handling i boligmarkedet antas samlet reflektert i deres Google-søkeatferd. Problemstillingen tar dermed indirekte for seg om forventninger til fremtidig handling i boligmarkedet kan forklare boligprisutviklingen i Norge.

For å besvare denne problemstillingen har vi benyttet Google-søkedata til å konstruere åtte variabler basert på søkeord relatert til boligmarkedet. Variablene reflekterer individenes samlede forventning til egen handling i boligmarkedet i fremtiden. Vi benytter i tillegg makroøkonomiske faktorer og data om den norske boligprisindeksen. De makroøkonomiske variablene, heretter også referert til som makrovariabler, representerer variabler benyttet i Norges Banks modeller for å predikere boligprisen, og er inkludert som sammenligningsgrunnlag for Google-søk.

Vi vil i masterutredningen ha et hovedfokus på å forklare hvor godt Google-søk kan forklare boligprisen i det norske markedet. Vi vil i tillegg presentere analyser for om Google-søk kan benyttes til å forutsi fremtidig utvikling i boligprisen og identifisere bobler i boligmarkedet

1.2 Bakgrunn

Boligkjøp er for mange nordmenn den største investeringen han eller hun foretar seg, og mye kapital bindes opp i boligen. Særegent ved Norge, i forhold til andre land, er det store andelen som eier fremfor å leie. Mens 77 prosent av nordmenn bor i en bolig husholdningen selv eier (SSB, 2017), gjelder tilsvarende kun for 62 prosent av svensker og 58 prosent av dansker (Sæby, 2014). Boligprisen er derfor viktig for en stor andel nordmenn, enten om man allerede eier eller planlegger å kjøpe eller selge. Om man planlegger å ta opp et lån i millionklassen

for å gjøre sitt livs største investering er det viktig å ha informasjon om hvilken retning man tror prisene vil bevege seg. Vi har besluttet å benytte offentlig tilgjengelig data for å estimere modeller som kan benyttes til å forutsi boligprisutviklingen, fordi modellen da kan benyttes av alle.

Vi antar at søkemotorer benyttes i boligkjøps- eller salgsprosessen for et stort antall individer, og det er derfor rimelig å anta at digitale søk vil gjenspeile individenes samlede forventninger til fremtidig handling i boligmarkedet. Google er den desidert mest brukte søkemotoren i Norge, med en markedsandel på over 90 prosent (Statscounter, 2018). Dette er grunnen til at vi har valgt å kun benytte Google-data til å representere digitale søk i vår masterutredning. Digitale søk vil derfor være synonymt med Google-søk i denne utredningen.

Offentlig tilgjengelig Google-data har de siste årene vist seg å være svært nyttig til å forklare boligprisutviklingen i flere land. Funnene viser at ved å inkludere variabler som representerer søkedata for spesifikke søkeord eller -uttrykk kan tradisjonelle boligmarkedsmodeller forbedres. Modellene med Google-søk har vist seg å være gode til å beskrive boligprisen i nåtid, som kan være nyttig da sentrale boligstørrelser ofte publiseres en gang i måneden eller sjeldnere. Det har også vist seg at bruk av digitale søk kan være nyttig til å forutsi fremtidig vekst eller fall, og til å identifisere bobler før de sprekker.

1.3 Relevans

Tidligere studier på sammenhengen mellom Google-søk og boligmarkedet er foretatt i USA, Nederland og England. Det er i tillegg utført masterutredninger og mindre omfattende studier i flere land, blant annet våre naboland Finland og Sverige. Denne masteroppgaven bidrar til forskningen med oss bekjent de første resultatene for det norske markedet. Big Data er svært aktuelt i dagens digitaliserte samfunn, og ny forskning på det norske boligmarkedet er alltid relevant da det utgjør en stor andel av nordmenns formue og økonomi.

1.4 Masterutredningens struktur

Denne utredningen vil følge følgende struktur. I kapittel to vil masterutredningens teoretiske fundament presenteres med et fokus på teori om aktivabobler. Deretter følger en

litteraturgjennomgang som tar for seg tidligere forskning relevant for vår problemstilling. Forsknings spørsmål vi ønsker å studere videre i utredningen presenteres mot slutten av kapittel tre. Kapittel fire tar for seg datakilder og hvilket grunnlag utvelgelsen av variabler er basert på. I kapittel fem blir de utvalgte variablene beskrevet. Kapittel seks forklarer den empiriske metoden benyttet i masterutredningen, samt regresjonsmodellene som utvikles for å besvare problemstillingen. Kapittel syv presenterer den empiriske analysen, basert på regresjonsresultater og grafiske fremstillinger. Deretter følger en analyse om bobletendenser i det norske markedet basert på det teoretiske fundamentet og den empiriske analysen. Kapittel ni diskuterer forsknings spørsmålene, samt begrensninger og forslag til videre arbeid. Til slutt presenteres konklusjonen som vil besvare problemstillingen og presentere de viktigste funnene.

2. Teoretisk fundament

I dette kapitlet presenterer vi sentrale definisjoner relatert til aktivabobler og en modell som forklarer stegene i oppbygningen av bobler. Dette vil anvendes i masterutredningens analyse i kapittel åtte. Vi vil først definere en aktivaboble, euforiske og ikke-euforiske bobler og krakk, før vi tar for oss Grytten og Hunnes' dynamiske syvtrinns krisemodell.

2.1 Definisjoner

Definisjonene vi har inkludert er essensielle for å forstå hvordan bobler oppstår og utfolder seg. Der det er relevant vil definisjonen utdypes med et spesielt fokus på boligmarkedet.

2.1.1 Aktivaboble

En aktivaboble skapes av vekst i markedet grunnet tro på fremtidig gevinst i markedet over aktivas faktiske verdi. Nobelprisvinner Joseph Stiglitz beskriver bobler slik: «Dersom årsaken til at prisen er høy i dag, *bare* skyldes at investorene tror at salgsprisen er høy i morgen – når «fundamentale» faktorer ikke rettferdiggjør en slik pris – da eksisterer det en boble» (Grytten & Hunnes, 2016). Grytten gir en lignende definisjon av finansielle bobler, som «handel av objekter i stort volum, til pris med signifikant avvik fra fundamentale verdier» (Grytten & Hunnes, 2016).

De to bobledefinisjonene gir innsikt i sentrale kjennetegn ved aktivabobler. Et fellestrekk er en overdreven tro på at veksten i markedet vil fortsette fremover, og at veksten i markedet drives av psykologiske faktorer fremfor fundamentale verdier.

Aktivabobler i boligmarkedet er beskrevet mer spesifikt av Nobelprisvinner Robert Shiller og Karl Case i artikkelen «Is There a Bubble in the Housing Market?» (2003). I artikkelen presenterer Shiller og Case flere tendenser som kan være tegn på en boligboble. En tendens beskrevet er førstegangskjøpere som frykter at hvis de ikke kommer seg inn på boligmarkedet nå, vil de aldri få det til. Tankegangen baserer kjøpet av bolig på troen om at prisene vil stige i fremtiden fremfor betjeningsmulighet av lånet eller dagens verdi av boligen.

Artikkelen beskriver at undervurderingen av hvor mye man må spare for å kjøpe bolig kan være en annen tendens på en aktivaboble. Boligens verdi kommer til å vokse i fremtiden, og

bolig kan dermed anses som en form for sparing. Denne tankegangen baserer seg på at markedet kun kan bevege seg i én retning.

En tredje tendens beskrevet i artikkelen er at kjøpere vil være villig til å kjøpe boliger som de klassifiserer som «for dyre», i troen på at den fremtidige prisstigningen vil kompensere for dette.

De tre tendensene beskrevet av Case og Shiller har til felles at kjøper har sterk tro på at boligprisen vil fortsette å stige i fremtiden. Dette tillater dem å investere enten mer, investere tidligere eller ta på seg mer risiko enn de ville gjort dersom de ikke hadde en slik sterk tro på fremtidig prisvekst.

2.1.2 Euforiske og ikke-euforiske bobler

Bobler kan kategoriseres som euforiske eller ikke-euforiske (Grytten & Hunnes, 2016). De beskriver en euforisk boble som en boble som «kjennetegnes ved at den raske prisveksten ikke kan forklares utfra fundamentale forhold». Robert Shiller definerer spekulative bobler lignende, som «en situasjon hvor midlertidig høye priser er opprettholdt hovedsakelig av investorers entusiasme, heller enn de konsistente estimatene av den reelle verdien» (Shiller, 2015).

En ikke-euforisk boble er på den annen side en sterk vekst i pris grunnet endringer i de fundamentale forholdene (Grytten & Hunnes, 2016). De trekker frem høye oljepriser eller befolkningsvekst som eksempler på fundamentale verdier som kan drive ikke-euforiske bobler. Videre beskriver de hvordan krakket i en ikke-euforisk boble skiller seg fra krakket for en euforisk boble: Når en ikke-euforisk boble sprekker representerer det en korreksjon i et overopphetet marked, og er ikke styrt av mani.

2.1.3 Krakk

Grytten og Hunnes (2016) beskriver et krakk som «når en boble sprekker, det vil si at prisene på aktiva kollapser. Situasjonen preges av langt større tilbuds- enn etterspørselsvolum». Videre forklarer de at «et krakk kan være langt mer begrenset enn en finanskriser».

En annen definisjon på et finansielt krakk, gitt av Hyman Minsky og Charles Kindleberger, sier at et finansielt krakk er et «betydelig og raskt fall i finansielle størrelser utover normal

korreksjon; har ringvirkninger til økonomien forøvrig; fallet er større enn en realøkonomisk betraktning skulle tilsi; må forklares ut i fra psykolog» (Grytten, 2008). Definisjonen indikerer at et krakk er et raskt fall i priser som følge av endret psykologi i markedet.

2.2 Syv-trinns dynamiske krisemodell

I dette underkapitlet vil vi presentere Grytten og Hunnes' (2016) dynamiske syvtrinns krisemodell, fra deres bok «Krakk og Kriser» dersom ikke annet er spesifisert. Modellen bygger hovedsakelig på Kindleberger sin modell, men inkluderer i tillegg elementer fra empirisk kriseforskning og Minskys og Eichengreens modeller. Vi vil ikke gå inn på Kindleberger, Minsky og Eichengreens modeller i detalj da disse modellene i seg selv ikke er benyttet i denne masterutredning.

Grytten og Hunnes' syv-trinns dynamiske krisemodell forklarer hvordan en økonomisk krise utvikler seg i syv steg, fra forstyrrelser i markedet til krise og spredning til resten av økonomien. Det forklares at skillene mellom stegene ikke nødvendigvis er tydelig, og at bobler kan forekomme hvor flere av stegene ikke er tilfellet.

1. Forstyrrelse

Som Kindleberger antar Grytten og Hunnes i sin modell at forstyrrelsen forårsakes av et eksogent sjokk. Endringer i økonomisk politikk, innovasjon, krig og innvandring trekkes frem som potensielle eksogene faktorer som kan føre til en forstyrrelse i markedet. Forstyrrelsen vil føre til endringer enten på tilbud- eller etterspørselssiden av økonomien. Det utelukkes ikke i modellen at forstyrrelsen kan medføre endringer både på tilbuds- og etterspørselssiden i samme periode. Grytten og Hunnes trekker frem penge- og kredittpolitikk som beveger seg i en ekspansiv retning som en mulig forstyrrelse som kan få fart på økonomien. Dette vil gjøre det lettere å få tilgang på kreditt som kan benyttes til investering.

2. Overoppheting

Forstyrrelsen kan føre til en overoppheting i økonomien grunnet troen på at forstyrrelsen kan føre til en langsiktig trendendring i økonomien. Euforien fører til en stadig større spekulasjon i aktiva og finansielle instrumenter, og det økende tempoet i økonomien leder til høyere etterspørsel etter kreditt for å hente ut gevinsten ved prisveksten.

3. Bobleøkonomi

Bobleøkonomien kjennetegnes ved at investorer fortsetter å investere til tross for at de har forstått at økonomien er overopphetet. Årsaken til at de ikke stanser investeringene er troen på at en «*greater fool*» vil være villig til å kjøpe de overprisede aktivaene i fremtiden – man antar man vil få solgt den videre til en høyere pris i fremtiden.

I en bobleøkonomi kan de fundamentale verdiene stagnere. Veksten i markedet vil da fortsette utelukkende som følger av spekulasjon og enkel tilgang på kreditt, og ikke på grunn av endringer i de fundamentale verdiene. Investorer kan fortsatt tjene på at prisene vokser, men samfunnet og økonomien som helhet vil lide i fremtiden ettersom veksten bygger på spekulasjon.

Grytten og Hunnes forklarer videre at det kan være vanskelig å oppdage når økonomien går inn i denne fasen. Reinhart og Rogoff (2008) forklarer i sin artikkel «This time is different» at folk ofte ønsker å begrunne bobletendenser med fundamentale verdier.

4. Nervøsitet

Bobleøkonomien leder til en periode med nervøsitet, ettersom stadig flere oppfatter at økonomien befinner seg på et unaturlig høyt nivå. De oppfatter aktiva som overpriset, og banker blir restriktive på kreditttilgangen fordi de anser risikoen for å låne ut penger som høyere enn før. I den nervøse fasen vil prisene svinge i stor grad ettersom all ny informasjon vil tolkes som antydning på kontinuerlig vekst eller at et vendepunkt nærmer seg. Institusjonelle investorer og stater kan gripe inn for å berolige markedene.

5. Vendepunkt

Vendepunktet vil komme når pessimismen tar overhånd. Investorer begynner å selge i frykt for at aktiva vil være mindre verdt i fremtiden. Stadig flere selger fordi pessimismen sprer seg. Antallet selgere i forhold til kjøpere øker drastisk, som leder til panisk stemning, og prisene kollapser i markedene. Grytten og Hunnes forklarer at det hovedsakelig er tre grunner til at prisene kollapser: kreditten tørker ut, investorer vil vente med å investere til de har troen på prisvekst, og investorer vil kjøpe når aktiva er billigst, altså på bunn. Dette bidrar til å forsterke prisfallet.

6. Krise

Krisen inntreffer som en ekkoeffekt av vendepunktet. Stadiet preges av at stadig flere ønsker å selge siden alle andre selger, og stadig færre vil kjøpe siden ingen andre vil kjøpe. Markedet

stuper og investorer blir stadig mer pessimistiske. Fallet fører til konkurser, som får bankene til å redusere kredittilbudet ytterligere. Bedrifter vil da få problemer med å lønne sine ansatte, og oppsigelser blir nødvendig. Dette er en selvforsterkende, ond spiral mot en resesjon eller en mulig depresjon.

7. Spredning

Den syvende og siste fasen vil ikke være tilfellet for alle kriser. Noen krakk forekommer kun på børsen. Hvis de er kortvarende trenger det ikke nødvendigvis føre til spredning til resten av økonomien. En krise i finanssektoren eller boligmarkedet kan derimot spre seg til store deler av økonomien og få alvorlige konsekvenser nasjonalt, og mulig internasjonalt. Krisen kan spre seg til andre markeder og lede til flere vendepunkt og krakk. Spredningen kan forekomme mellom ulike industrier eller på landsbasis. Et godt eksempel er Subprime-krisen i USA, som førte til en global finanskris.

3. Litteraturgjennomgang

I dette kapitlet vil vi først presentere Google-søk benyttet til prediksjon i forskjellige markeder, før vi spisser oss inn mot boligmarkedet. Avslutningsvis presenteres tre forskningsspørsmål vi ønsker å studere i analysearbeidet senere i masterutledningen.

3.1 Google-søk benyttet til prediksjon

Prediksjon kan dreie seg om å predikere hvordan tilstanden er i nåtid, eller fremtidig utvikling. Dette underkapitlet vil først fokusere på Google-søkedata benyttet til å predikere nåtiden, før prediksjon av fremtiden presenteres. Til slutt vil kritikk av Google-søk benyttet til prediksjon presenteres.

3.1.1 Prediksjoner om nåtiden

I 2009 publiserte Googles forskningsavdeling en artikkel om hvordan Google-data kan benyttes til å forklare utviklingen i forskjellige markeder i nåtid, såkalt *nowcasting* (Varian & Choi, 2009). I artikkelen viser forskerne til fire markeder de har studert: detaljhandel, bil, bolig og reise. For samtlige markeder viser modeller som inkluderer variabler basert på Google-søk en økt forklaringsgrad sammenlignet med modeller som ikke inkluderer Google-søk. Forskerne fremhever fordelen ved at datamaterialet er lett tilgjengelig og publiseres etter 48 timer, fremfor å vente flere uker eller måneder på at myndigheter skal publisere data. Varian og Choi trekker også frem at datamaterialet myndighetene publiserer ofte justeres i etterkant, som medfører ytterligere etterslep i publiseringen av data.

Forskere ved Google og The Center for Disease Control and Prevention (CDC) publiserte samme år som Varian og Choi en artikkel om Google som *nowcaster* influensautbrudd basert på Google Flu Trends-modellen (Ginsberg et al., 2009). Artikkelen hevder at «fordi den relative hyppigheten av visse søk er sterkt korrelert med prosentandelen av legebesøk fra pasienter med influensalignende symptomer, kan vi nøyaktig estimere nåværende ukentlig influensaaktivitet i hver region i USA, med et rapporteringsetterslep på én dag».

I etterkant av publiseringen av de mye omtalte artiklene av Varian og Choi og Ginsberg et al. har det blitt utført en rekke forsøk på å benytte Google-data til *nowcasting*. Google-søkemateriale har blitt benyttet til å substituere eller supplere modeller som måler trender i bilindustrien i Chile (Carrière-Swallow & Labbé, 2011), arbeidsledighet i Polen, Slovakia, Tsjekkia og Ungarn (Pavlicek & Kristoufek, 2015) og privatkonsum i USA (Kholodilin, Podstawski & Siliverstovs, 2010). Samtlige artikler trekker frem fordelene ved det korte publiseringsetterslepet av Google-data sammenlignet med makroøkonomiske modeller. Dataene kan derfor benyttes til å gi et bilde av nåtiden uten en forsinkelse på flere uker eller måneder.

3.1.2 Prediksjoner om fremtiden

Det siste tiåret har Google-søk sitt potensiale for prediksjon av fremtiden vekket nysgjerrighet i academia (Jun, Yoo & Choi, 2017). Jun, Yoo og Choi argumenter for at Google-søk potensielt kan predikere fremtidige trender, basert på hvilke temaer som søkes på. Det har blitt gjort forsøk på å predikere fremtidig utvikling i arbeidsledighet i USA (D'Amuri & Marcucci, 2017), globalt bensinkonsum (Yu, Zhao, Tang & Yang, 2018) og unormal aksjeavkastning på Wall Street (Joseph, Wintoki & Zhang, 2011) ved bruk av Google-data. Alle disse forsøkene har gitt statistisk signifikante resultater.

Det har hersket uenighet om hvor god den prediktive evnen til Google-søk for aksjeavkastning er. Ifølge Preis, Reith og Stanley (2010) gir det ikke signifikant høyere avkastning å benytte Google-data til å predikere enn andre metoder. Preis, Mot og Stanley fant i 2013 derimot at Google-søk kan predikere fremtidig volum på aksjehandel. I 2016 fant Bijl, Kringhaug, Molnár & Sandvik at et stort antall søk på selskapsnavn førte til negativ avkastning, i kontrast til tidligere artikler som fant at Google-søkevolum var positivt korrelert med avkastning på kort sikt. Det er med andre ord ikke entydig om Google-søk er godt egnet til å predikere fremtidig utvikling i aksjemarkedet eller ei.

3.1.3 Kritikk av bruk av Google-søk til prediksjon

I 2012 og 2013 slo prognosene for influensautbrudd basert på Google Flu Trends-modellen (Lazer, Kennedy, King & Vespignani, 2014) feil. I artikkelen fremhever forskerne hvordan en modell basert på Big Data gikk fra å være et ypperlig prediksjonsverktøy noen år tidligere til å slå så feil i 2012 og 2013. En forklaring på at modellen slår feil kan være at Google's algoritme for å undersøke hvilke søkeord som er assosiert er mangelfull (Lazer et al., 2014).

Artikkelen påpeker hvordan algoritmen feilaktig inkluderer flere søkeord i modellen som ikke er relatert til influensa, for eksempel «high school basketball». Det er vanskelig å tro at dette søkeordet har noe med influensa å gjøre.

I artikkelen påpeker forskerne i tillegg problematikken ved å inkludere søkeord som «cough» og «fever» i Google Flu Trends-modellen, da dette er symptomer for andre sykdommer i tillegg til influensa (Lazer et al., 2014). Forskerne trekker frem at modellen vil predikere at mange har disse symptomene, uten være i stand til å skille forkjølelse- eller matforgiftningstilfeller fra influensatilfeller. Dette vil ifølge Lazer et al. medføre at modellen overestimerer det faktiske omfanget av influensa.

I samme artikkel trekker forskerne frem hvordan all medieoppmerksomheten rundt «bird flu» og «swine flu» kan påvirke søkeaktivitet. Nyhetsreportasjer som nevner influensa kan føre til at flere benytter Google til å skaffe seg informasjon om influensa (Lazer et al., 2014). En økning i Google-søk etter influensarelaterte søkeord trenger derfor ikke indikere at de faktisk har influensa.

3.2 Google-søk benyttet til prediksjon av boligmarkedet

Internasjonalt er det utført flere studier på prediksjonsevnen til Google-søk på boligmarkedet. Vi vil i dette underkapitlet starte med å presentere forskningsartikler om det nederlandske, amerikanske og britiske boligmarkedet. Grunnen til at disse markedene er valgt ut er ganske enkelt at de er de eneste markedene der forskningsartiklene er på engelsk.

Flere av artiklene undersøker i tillegg om Google-søkedata kan benyttes til å forbedre tradisjonelle boligmarkedsmodeller basert på sentrale makroøkonomiske variabler. Dette vil vi ta for oss i andre del av dette underkapittel.

3.2.1 Prediksjon av internasjonale boligmarkeder

Det amerikanske boligmarkedet

En av de mest diskuterte artiklene om sammenhengen mellom boligmarkedet og Google-søk er en forskningsartikkel av Wu og Brynjolfsson (2015) for NBERs utgivelse «Economic Analysis of the Digital Economy». Artikkelen tar for seg perioden 2004 til 2011, og forsøker å avdekke om Google-søk kan predikere solgt volum og prisen i det amerikanske boligmarkedet. De finner ved hjelp av en enkel lineær regresjonsmodell at søk på spesifikke

ord på Google er korrelert med både antall solgte hus og boligprisindeksen. Wu og Brynjolfsson viser til at ved å benytte Google-søk utkonkurreres flere eksperter på området, blant annet organisasjonen for eiendomsmeglere i USA.

Ole Martin Eidjord og Are Oust (2017) ved NTNU undersøkte om Google-søk kan benyttes til å predikere boligpriser og identifisere bobler i det amerikanske boligmarkedet. Studien benytter paneldata og inkluderer kvartalsvise observasjoner for samtlige amerikanske stater, i tidsperioden 2004 til 2016. Med utgangspunkt i boligboblen i 2006-2007 testet de 204 ulike Google-søkeord som indikatorer for prisutviklingen og eksistensen av aktivabobler i boligmarkedet. Artikkelen konkluderte med at søkeordet «Real Estate Agent» (eiendomsmegler) kan inkluderes i tradisjonelle boligprismodeller for å forbedre modellene. Artikkelen viser også at søk på «Housing Bubble» (boligboble) er en ledende indikator for bobler i boligmarkedet.

Det nederlandske boligmarkedet

van Veldhuizen, Vogt og Voogts (2016) har undersøkt hvor godt søk på det nederlandske ordet «hypotheek» (boliglån) på Google kan forklare månedlige boligtransaksjoner i Nederland. Hypotesen forskerne jobber ut ifra hevder at fremtidige boligkjøpere benytter Google til å skaffe seg informasjon om potensielle eiendommer, lokasjoner og finansieringsmuligheter før de gjennomfører et kjøp. Forskerne fant at antall søk på «hypotheek» er signifikant korrelert med antall boligtransaksjoner samme måned, samt én eller flere måneder tilbake i tid. Forskerne hevder at det tyder på at søkeatferd på Google kan gi nyttig informasjon om fremtidig adferd i boligmarkedet.

Det britiske boligmarkedet

Bank of England undersøkte i 2011 om søkevolum på nett kan benyttes som indikator på den økonomiske aktiviteten i Storbritannia (McLaren & Shanbhogue, 2011), og tar for seg arbeidsmarkedet og boligmarkedet spesifikt. For boligmarkedet benytter de søkeordet «Estate Agent» (eiendomsmegler) som indikator. Forskningsresultatet viste at søkeordet var signifikant positivt korrelert med boligprisene i Storbritannia mellom 2004 og 2011.

3.2.2 Forbedring av tradisjonelle makroøkonomiske modeller

For å undersøke hvor nyttig Google-søkedata er til å predikere pris og solgt volum i boligmarkedet kan det være hensiktsmessig å undersøke om Google-data kan bidra til å

forbedre eksisterende boligmarkedsmodeller. Wu & Brynjolfsson (2015) fant at bruk av Google-søkedata signifikant forbedrer prognosene for boligmarkedets utvikling sammenlignet med prediksjonene fra National Association of Realtors. Også i Nederland fant Nederlandsche Bank at forklaringsgraden til prediksjonsmodellen for antall boligtransaksjoner økte med fire prosentpoeng ved å inkludere Google-søk som forklaringsvariabel (van Veldhuizen, Vogt & Voogts, 2016). Det samme gjelder det britiske boligmarkedet, hvor *baseline*-modellen ble forbedret signifikant ved å inkludere en variabel for Google-søk (McLaren & Shanbhogue, 2011).

Dietzel, Braun og Schaefer ved Universitetet i Regensburg utformet i 2014 tre modeller basert på det amerikanske boligmarkedet: en *baseline*-modell basert på sentrale makroøkonomiske variabler, en modell som utelukkende inkluderer Google-søkevariabler, og en tredje modell som en kombinasjon av de to. Resultatene viste at modellen med Google-søk er bedre til å predikere enn modellen med makroøkonomiske variabler. Den tredje modellen gir best resultater, som viser at prediksjonsevnen til boligmarkedsmodeller kan forbedres signifikant ved å inkludere gratis og lett tilgjengelig Google-data (Dietzel, Braun og Schaefer, 2014).

3.3 Forskingsspørsmål

Basert på litteraturgjennomgangen og det teoretiske fundamentet om boligbobler har vi kommet frem til tre forskningsspørsmål vi ønsker å studere i vår empiriske analyse. Dette vil danne grunnlag for å besvare vår problemstilling.

Flere av de presenterte artiklene i litteraturgjennomgangen peker på at Google-variabler kan forbedre tradisjonelle boligprismodeller. Vi ønsker å undersøke om dette kan være tilfellet for det norske boligmarkedet, og har utformet følgende forskningsspørsmål:

Spørsmål 1: Kan boligprismodeller kun basert på makroøkonomiske variabler forbedres ved å inkludere Google-variabler?

Spørsmål 1a: Kan Google-variabler forklare boligprisutviklingen bedre enn makrovariabler alene?

Spørsmål 1b: Kan Google-variabler i kombinasjon med makrovariabler forklare boligprisutviklingen bedre enn makrovariabler alene?

Hovedfokuset i vår utredning er å undersøke om Google-søk kan forklare boligprisens utvikling. Vi ønsker i tillegg å undersøke om empiriske resultater fra vårt arbeid kan benyttes til å indikere fremtidig boligprisoppgang og identifisere bobletendenser i boligmarkedet.

For å undersøke om Google-søk kan benyttes til å forutsi fremtidige prisfall i boligmarkedet ønsker vi å undersøke følgende forskningsspørsmål:

Spørsmål 2: Kan Google-søk forutsi fall i boligprisen bedre enn makroøkonomiske variabler?

For å undersøke om Google-søk kan benyttes til å identifisere bobletendenser i boligmarkedet ønsker vi å undersøke følgende forskningsspørsmål:

Spørsmål 3: Kan vi identifisere bobletendenser i det norske boligmarkedet ved å sammenligne Google- og makroøkonomiske variabler?

Vi vil undersøke de tre forskningsspørsmålene i den empiriske analysen og analysen av bobletendenser, for deretter å diskutere resultatene i kapittel syv. Hovedfokuset i vår masterutredning vil være på forskningsspørsmål 1. Forskningsspørsmål 2 og 3 vil studeres basert på resultater fra spørsmål 1.

4. Datasett og utvelgelse av variabler

Dette kapitlet vil først beskrive hvilke datakilder som benyttes til å konstruere datasettet som benyttes i den empiriske analysen. Deretter beskrives metoden for å velge ut variabler. Datakildene og metoden for å velge ut variabler vil beskrives med hensyn til validitet og reliabilitet, dette for å sikre et robust og hensiktsmessig datasett til vårt formål. Reliabilitet sier noe om hvor stabile og konsistente resultatene er (Wooldridge, 2016). Validitet sier noe om hvor godt variabelen måler det den har som hensikt å måle (Wooldridge, 2016).

Boligprisindeksen produseres av Eiendom Norge, Google-søkedata fra Google Trends og makroøkonomiske variabler fra Norges Bank. All data er offentlig tilgjengelig. Datasettet er sortert kronologisk etter måneder, og er enten hentet ut på månedsbasis eller konvertert fra kvartal til måned. Første observasjon i det sammensatte datasettet er januar 2007, ettersom dette er første året data fra Google er stabile nok til å kunne anvendes i analyser. Siste observasjon er desember 2017, den siste observasjonen før masterutredningen ble påbegynt. Dette utgjør til sammen 132 observasjoner. Grunnen til at vi har valgt måned som tidsseriekategorisering er for å få et relativt stort antall observasjoner.

4.1 Datakilder og beskrivelse av datasettet

I dette underkapitlet vil de tre datakildene presenteres. Hver av datakildene vil veies opp mot reliabilitet og validitet.

4.1.1 Boligprisindeksen fra Eiendom Norge

Eiendom Norge mottar utarbeidet data fra Eiendom Verdi, og har ansvaret for å utarbeide og publisere månedlig prisstatistikk i samarbeid med Eiendomsverdi AS og Finn.no (Eiendom Norge, 2018). Den hyppige publiseringen og det stabile datagrunnlaget gjør prisindeksen til en reliabel variabel.

Datasettet er basert på «salg formidlet av meglere og annonsert gjennom Finn.no» (Eiendom Norge, 2018). Omsetninger som kan anses som av annen karakter enn bolig eller fritidsbolig, slik som tomter og garasjer, tas ut av datagrunnlaget ved automatiserte rutiner hos Eiendomsverdi AS (Eiendom Norge, 2018). Det innebærer at indeksen ikke inkluderer samtlige eiendomstransaksjoner foretatt i Norge, som er viktig ettersom vi kun ønsker å ta for

oss bruktboliger i denne utredningen. Det styrker argumentet om at prisindeksen har sterk validitet.

4.1.2 Google-variabler fra Google Trends

Google er en utradisjonell datakilde, og vi vil gi en grundig forklaring av hvordan data er samlet inn og behandlet før publiseringen av statistikk på Google Trends. Dette er gjort for å understreke hvorfor søkeord har potensiale som forklaringsvariabler for boligprisindeksen.

Google Trends

Google Trends angir statistikk over søk og skaleres mellom null og 100 i forhold til den perioden hvor ordet har blitt søkt på mest. Appendix A1 viser et stillbilde av informasjon Google Trends kan gi. Tidsperioden kan enten sorteres etter dager, uker, måneder eller år. Dersom tidsaksen for eksempel settes til måneder vil statistikken vise populariteten til søkeordet for den respektive måneden relativt til måneden med flest søk. Dersom søkeordet var svært lite populært relativt i forhold til den mest populære måneden eller dersom søkevolumet faller under en minstegrense (Google Help1) settes verdien for måneden til null.

Google Trends eliminerer dupliserte søk fra statistikken før den publiseres (Google Help2). Det innebærer at dersom en person søker på et ord gjentatte ganger over en kort tidsperiode vil det ikke registreres som mer enn ett søk, som styrker validiteten til Google Trends som datakilde. Statistikken vil representere forbrukernes interesser og hensikter, og vil ikke kunne manipuleres av enkeltindivider.

En utfordring med Google-data er at verdiene kan variere basert på hva som settes som sluttdato. Verdiene fluktuerer svakt for hver måned om man for eksempel setter mai 2018 eller desember 2017 som sluttdato for perioden man ønsker å studere. Google har ingen forklaring på sine sider av dette avviket. Den inkonsistente målingen av Google-data svekker reliabiliteten til datakilden.

Google Adwords

For å undersøke reliabiliteten og validiteten til data fra Google Trends ytterligere har vi tatt i bruk Google Adwords. Appendix A2 viser et stillbilde av informasjonen Google Adwords tilbyr).

En potensiell ulempe med Google Trends' bruk av skalering til å angi søkeordets popularitet er at man ikke har mulighet til å hente ut de absolutte søkertallene for perioden. For å kvalitetssikre søkeordene vi ønsker å benytte som variabler er det derfor nødvendig å undersøke om de absolutte søkertallene er tilfredsstillende store. På Google Adwords angis gjennomsnittlig antall søk i intervallene én til 10, 10 til 100, 100 til 1000, og så videre. Det gjør det mulig å underbygge reliabiliteten til søkeord ved å utelukke søkeord med få absolutte søk.

En annen fordel ved Google Adwords er muligheten for å undersøke hvor sterk konkurranse det er om å kjøpe annonser på hvert av søkeordene. Konkurransesgraden kategoriseres i tre grupper: høy, middels og lav. Vi antar at bedrifters etterspørsel etter å annonsere på søkeordene i høy-kategorien er basert på hva selskapene tror de vil tjene penger på. Vi anerkjenner at bedrifter har bedre kunnskap om forbrukernes antatte preferanser enn det vi selv har, og at søkeordene med høy konkurranse er søkeordene som anses for å representere forbrukernes interesser. Søkeord med høy konkurranse antas derfor å ha sterk validitet.

4.1.3 Makroøkonomiske variabler fra Norges Bank

Norges Bank utarbeider hvert kvartal data for variabler de anser som viktige i å forklare utviklingen i boligmarkedet. Vi mottok data for disse variablene fra Norges Bank ved Bjørn Erik Naug på mail. Samtlige variabler er sentrale makroøkonomiske variabler, og er mål på den norske økonomien som helhet. Vi anser data tilsendt fra Norges Bank som en reliabel da de selv benytter dette datasettet i egne beregninger for boligprisutviklingen i Norge.

4.2 Utvelgelse av variabler

Her vil vi presentere på hvilket grunnlag vi har valgt ut den avhengige variabelen, prisindeksen, og de uavhengige variablene fra Google Trends. Vi anser det ikke som nødvendig å gå nærmere inn på utvalget av variablene fra Norges Bank da de allerede er vurdert som relevante for boligmarkedet av Jacobsen og Naug i deres boligprismodell (Jacobsen & Naug, 2004). Variablene er vurdert opp mot validitet og reliabilitet.

4.2.1 Kriterier for utvelgelse av Google-søkeord

For å studere om Google-søk kan forklare boligprisen har vi konstruert åtte Google-variabler basert på Google-søkeord. Det er disse Google-variablene vi vil benytte til å utforme regresjonsmodeller i kapittel seks, og variablene vil beskrives i kapittel fem. Utvelgelsen av søkeord for å konstruere Google-variablene er foretatt ved at fire egendefinerte kriterier må tilfredsstilles, dette for å sikre relevans i forhold til boligmarkedet, reliabilitet og validitet. Kriteriene er presentert i tabell 1. Appendiks A3 viser hvilke søkeord som er ekskludert basert på hvilke kriterier.

Utgangspunktet for utvelgelsen av Google-søkeord var 84 søkeord vi kom frem til på grunnlag av tidligere litteratur, økonomisk teori og egen intuisjon. En oversikt over disse er å finne i appendiks A3. Hensikten var å dekke et bredt spekter av søkeord som potensielt kan være relevante for boligmarkedet. Både tilbuds- og etterspørselsrelaterte ord ble inkludert.

Kriterier	Forklaring
(1) Økonomisk relevans	For å sikre kausalitet og validitet i analysene våre er alle ord som anses som lite relevant for boligmarkedet eliminert.
(2) Tilstrekkelig datakvalitet på Google Trends	Søkeord som har et ustabil søkeantall er eliminert. Dette er søk som har stor grad av nullverdier og store svingninger over korte perioder på Google Trends.
(3) Tilstrekkelig absolutte søkertall	>1000 gjennomsnittlig antall søk for å sikre at analysen baserer seg på et tilstrekkelig antall søk på Google. Vi anser mer enn 1000 gjennomsnittlige søk som et tilstrekkelig stort datagrunnlag. Vurdert ved å benytte Google Adwords.
(4) Konkurranse om søkeordet	Kategorisert som «høy» eller «middels» på Google Adwords, for å sikre at søkeordet reflekterer forbrukernes preferanser.

Tabell 1: Kriterier for utvelgelse av Google-søkeord.

Kriterie (1) *Økonomisk relevans* eliminerer ord vi anser som mindre relevante for å forklare boligprisen. Dette gjelder ord vi har undersøkt basert på tidligere litteratur, slik som «renovasjon» og «krakk», og søkeord som blir for generelle til å kunne forklare boligprisen, for eksempel «gjeld» – årsaken til at man søker på «gjeld» kan være mange, ikke utelukkende fordi man skal kjøpe bolig. Ordene som gjenstår etter denne elimineringsrunden er ord vi har

grunnlag til å tro har tilknytning til boligmarkedet. Kriteriet er inkludert for å sikre søkeordenes validitet.

Kriterie (2) *Tilstrekkelig datakvalitet på Google Trends* utelukker søkeord som er av for dårlig kvalitet til å gi meningsfylt og brukbar statistikk på Google Trends. Dette inkluderer blant annet «kjøpe bolig i Bergen», som har lavt søkerantall, og «Notar Eiendomsmegler», som gikk konkurs i perioden. Her testes søkeordenes reliabilitet. Kun ord som har konsistente og stabile resultater blir inkludert.

Kriterie (3) *Tilstrekkelig absolutte søkertall* sikrer at Google-søkeord som er inkludert har et tilstrekkelig stort antall søk. Google Trends vil gi verdien null til søk som er foretatt av få personer, men siden de ikke publiserer informasjon om hva «ikke tilstrekkelig med søk» innebærer, kontrollerer vi ved hjelp av Google Adwords at søkeordene har stort nok søkevolum. De enkeltstående søkeordene ekskluderes dersom de har mindre enn 1000 gjennomsnittlige søk.

Der vi har konstruert Google-variabler sammensatt av flere søkeord setter vi kravet på 100 gjennomsnittlige søk per ord. Vi antar at forventet antall søk for hvert ord i kategorien 100 til 1000 ligger på 500, og ved å sette sammen to eller flere ord til en Google-variabel vil summen av antall absolutt søk overstige 1000. Sammensettingen av flere søkeord til en Google-variabel blir utdypet i kapittel fem.

Kriterie 3 undersøker reliabiliteten til Google-søkene. Her kreves det at søkevolumet er tilstrekkelig stort, for å sikre at de ikke gjøres svært ustabile av mindre hendelser.

Kriterie (4) *Konkurranse om søkeordet* utelukker ord som ikke er tett knyttet til en beslutning i boligmarkedet. Vi ønsker kun å inkludere ord som reflekterer søkerens intensjon om å kjøpe eller selge bolig, og eliminerer derfor ord som bedrifter ikke anser som interessante å annonsere på. Ord som «boligannonse» er kategorisert som “lav”, som betyr at få bedrifter er interesserte i å annonsere på dette ordet. Det indikerer at ordet sjelden leder til kjøp blant Google sine brukere. Vi har kun inkludert søkeord kategorisert som “høy” eller “middels”. Kriteriet er inkludert for å sikre søkeordenes validitet.

For kriterie 4 er det gjort unntak for Google-variabelen som representerer navn på eiendomsmeglerselskaper. Disse søkeordene kategoriseres som “lav”. Vi antar at dette ikke er på grunn av mangel på tilknytning til en beslutning i boligmarkedet, men heller at selskap ser det som ulønnsomt å annonsere på konkurrenters navn.

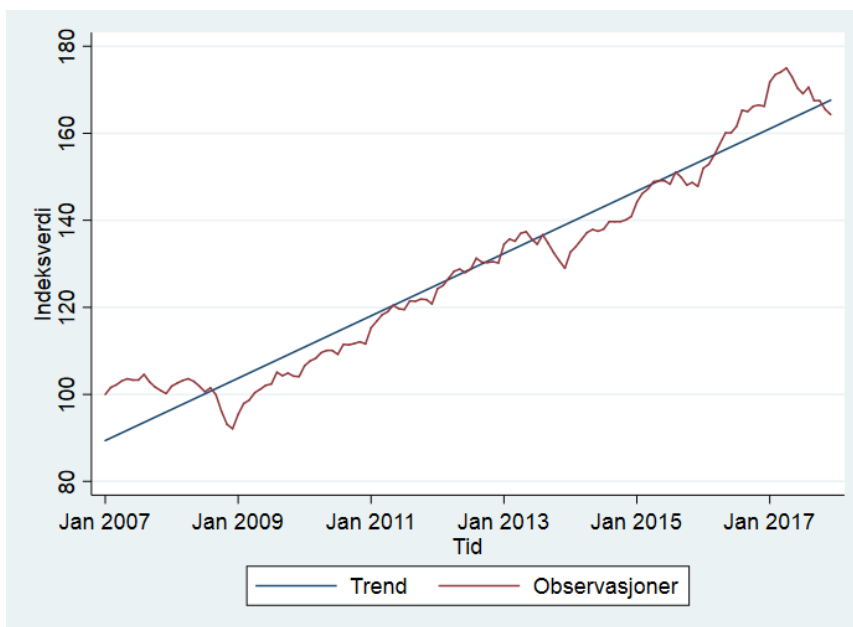
5. Datadeskripsjon

I dette kapitlet vil den avhengige og de uavhengige variablene beskrives med fokus på deres tilknytning til boligmarkedet. De uavhengige variablene vil også knyttes til tilbuds- eller etterspørselssiden av markedet. Reliabilitet og validitet ble diskutert for datasettene i forrige kapittel, og vil i dette kapitlet kun diskuteres i tilfeller der eksterne faktorer kan bidra til å svekke reliabiliteten og validiteten.

Først vil boligprisindeksen beskrives, før de utvalgte Google-variablene presenteres. Til slutt vil de makroøkonomiske variablene diskuteres. Samtlige variabler benyttet i masterutredningen er gitt en kort beskrivelse i appendiks A4.

5.1 Boligprisindeksen

I denne utredningen vil Eiendom Norges Boligprisindeks benyttes som mål på boligprisutviklingen i det norske markedet, og som avhengig variabel. Når det er snakk om boligmarkedets utvikling i Norge i dag er det som regel boligprisindeksen det er snakk om. Veksten i boligprisen har fulgt en oppadgående trend siden 2007, se figur 1.



Figur 1: Boligprisindeksen, kilde: Eiendom Norge

Boligprisindeksen er hentet fra Eiendom Norges nettsider. Januar 2007 er benyttet som referanseperiode (januar 2007 = 100). For hver påfølgende måned oppgis boligprisen som en

prosentvis endring i indeksenivået fra måneden før (Eiendom Norge, 2018). Boligprisindeksen beregnes og publiseres av Eiendom Norge for den foregående måneden på den tredje virkedagen hver måned.

5.2 Utvalgte Google-variabler

De 17 søkeordene presentert i kolonne tre i tabell 2 tilfredsstillende de fire kriteriene beskrevet i tabell 1. Av disse har vi konstruert åtte uavhengige variabler. Variablene vil bli omtalt som Google-variabler. Seks av Google-variablene er konstruert av flere søkeord, hvor vi antar at intensjonen bak søkeordene er den samme. Det er for eksempel liten grunn til å tro en som søker etter Krogsveen eller Eiendomsmegler 1 har ulik intensjon med søket sitt i de fleste tilfeller – begge vil frem til en spesifikk eiendomsmegler.

Variabelnavn	Inkluderer følgende søkeord
<i>Boliglån</i>	boliglån
<i>Boliglånsrente</i>	boliglånsrente + boligrente
<i>Eiendomsmegler</i>	eiendomsmegler
<i>Selskapsnavn</i>	dnb eiendom + krogsveen + eiendomsmegler 1
<i>Verdivurdering</i>	verdivurdering + gratis verdivurdering + verdivurdering av bolig
<i>Boligkjøp</i>	kjøpe bolig + kjøpe hus + kjøpe leilighet
<i>Boligsalg</i>	salg av bolig + selge bolig
<i>Finn</i>	finn eiendom + finn eiendom no

Tabell 2: Konstruerte Google-variabler

5.2.1 Etterspørselsdrevende variabler

Variabler ansees som etterspørselsdrevet når intensjonen Google-søket anses for å være kjøpsfokuseret. Hvis søket fokuserer på kjøp virker det rimelig at etterspørselen etter boliger øker, som vil føre til en økning i boligprisen i henhold til økonomisk teori.

Boliglån

De fleste som skal kjøpe bolig er avhengig av å ta opp lån, og det er derfor naturlig at Google-søk etter boliglån forekommer før man skal kjøpe bolig. Søkeren vil ønske å finne tilbud og

informasjon om hva de forskjellige bankene har å tilby. Det antas i tillegg at søkene etter boliglån foregår en god stund før selve kjøpet finner sted, ettersom prosessen fra innhenting av informasjon om boliglånsmarkedet, boliglånsdiskusjoner med banker og letingen etter en passende bolig vil ta noe tid. Søkene etter boliglån kan derfor forekomme lang tid før det vil kunne gi utslag i boligmarkedet.

En utfordring med *Boliglån* er graden av boliglån som går til refinansiering. Ifølge Finanstilsynets rapport om boliglån fra 2018 gikk kun omtrentlig 30 prosent av alle boliglån til boligkjøp. Til sammenlikning ble over 50 % knyttet til refinansiering. Dette svekker validiteten til variabelen, ettersom sammenhengen mellom Google-søk og beslutningen i boligmarkedet ikke er like sterk.

Til tross for dette svekkes ikke validiteten så mye som det kan virke, ettersom over 80 prosent av lånesaker som innebar refinansieringen skjer med lån fra samme bank. Det er derfor langt mindre sannsynlig at refinansieringsandelen av søkene på «boliglån» er stor, da man ikke vil ha behov for å oppsøke informasjon om bankenes ulike vilkår på samme måte som første gang man skal søke etter lån. Så lite som ti prosent av boliglån gikk til refinansiering av boliglån fra andre banker (Finanstilsynet, 2018). Disse kan påvirke søkene, men variabelen virker fortsatt å ha tilfredsstillende validitet.

Boliglånsrente

På lik linje med boliglån er de fleste som skal kjøpe bolig opptatt av renten på lånet de må ta opp. Søk på disse ordene antas foretatt av forbrukere som ønsker å ta opp et boliglån for å kjøpe bolig, og som for eksempel ønsker å sammenligne ulike rentesatser hos bankene. Også her antas det at man undersøker rentenivået i god tid før man foretar et boligkjøp, og at det derfor blir naturlig å teste effekten av etterslep. Grunnen til at denne variabelen ikke er kombinert med *Boliglån* er at en grafisk fremstilling av variablene viste at de ikke korrelerte i stor nok grad, se appendiks B1. Vi har derfor valgt å beholde dem som separate forklaringsvariabler.

Det antas også for *Boliglånsrente* at den kan bli påvirket av graden av refinansiering, men at det heller ikke i dette tilfellet representerer store validitetsproblemer.

Boligkjøp

Boligkjøp er satt sammen av søkeordene «kjøpe bolig», «kjøpe hus» og «kjøpe leilighet».

Intensjonen bak de tre søkeordene er den samme, å kjøpe bolig, men type bolig varierer. De noe ulike søkeordene inkludert i variabelen ble valgt ut for å inkludere en så stor andel av markedet som mulig. Det er usikkert når i kjøpsprosessen de fleste søkene forekommer, men vi antar at det skjer relativt tett opp mot kjøpet ettersom det er relatert til selve transaksjonen.

Finn

Finn-variabelen er satt sammen av «finn eiendom» og «finn eiendom no». I Norge annonseres de fleste boliger på Finn.no. Ifølge Berentsen fra DnB Eiendom i Bergen ble 99 prosent av alle boliger fra hennes avdeling annonsert på Finn (se mailkorrespondanse i appendiks A5). Selv om selgere kan ha motivasjon til å søke for å sjekke prisnivået i sitt nabolag antar vi at de fleste som søker på disse søkeordene er interessert i å kjøpe bolig.

Et potensielt problem ved å inkludere *Finn* er Finn.no sin monopolsituasjon og navnegjenkjenning. Som vi ser av den grafiske fremstillingen av verdier for Finn-søkeordene (appendiks B1) har søk etter Finn på Google variert i stor grad siden januar 2007. Etter en jevn økning frem mot 2012 har søkertallene vært svært ujevne. Man ville forventet en jevnere økning slik som for de fleste andre variabler beskrevet i denne seksjonen (se appendiks B1 for en komplett oversikt over Google-variablenes utvikling over tid), da vi antar at bruk av Google generelt har økt over tid. Nedgangen fra 2010 og den ujevne søkehistorikken etter dette skyldes antakeligvis ikke at Finn har mistet markedsandeler eller posisjon i det norske markedet, men heller at Finn Eiendom har blitt et synonym for boligannonser. Dette vil gjøre det mer naturlig for personer å gå direkte inn på Finn.no enn å søke via Google. Det kan medføre at søk etter Finn vil falle til tross for økende interesse for Finn Eiendom. Dette svekker reliabiliteten til *Finn* som forklaringsvariabel. *Finn* er likevel inkludert i analysen fordi Finn.no har en såpass sterk tilknytting til boligmarkedet. I fremtiden burde denne vurderes dersom reliabiliteten blir svekket ytterligere.

5.2.2 Tilbudsdrevende variabler

Eiendomsmegler

Både selger og kjøper av bolig kan benytte seg av megler, men vi antar det som vanligst å benytte megler når man skal selge. Vi anser derfor google-søk på «eiendomsmegler» som en tilbudsdreven variabel, ettersom den som søker på dette er på leting etter en eiendomsmegler til å selge boligen sin. Når søk på eiendomsmegler øker er det logisk å forvente at det vil komme flere boliger på markedet, antakeligvis med et etterslep fra noen uker til et år (se

mailkorrespondanse med DnB Eiendom i appendiks A5). Eiendomsmegler er forøverig den eneste Google-variabelen som har en nedadgående trend, se figur i appendiks B1.

Selskapsnavn

Selskapsnavn er en variabel konstruert av flere søkeord, hvor intensjonen bak søkene antas tilnærmet lik. Den består av «DNB eiendom», «Krogsveen» og «Eiendomsmegler 1». Dette er navnene på de største nasjonale eiendomsmeglerselskapene i Norge (Parr, 2014). Andre selskap er ikke inkludert enten grunnet svakt søkegrunnlag (lokale eiendomsmeglere) eller vanskeligheter med å skille søkeordet fra søk med andre intensjoner (for eksempel «Partners»). Søkeordene er slått sammen til en variabel fordi vi antar at personer søker på de forskjellige selskapene i samme steg i boligsalgprosessen.

En utfordring ved å kombinere de forskjellige selskapene til en variabel er at et individ kan søke på de forskjellige navnene uten at de blir eliminert av Google som dupliserte søk. Hvis et individ søker på navnene for å sammenligne priser og tjenester vil dette komme til syne i statistikken til Google Trends som tre separate søk, heller enn ett. Dette svekker validiteten til variabelen noe. Vi har likevel valgt å inkludere variabelen da vi ikke tror dette vil svekke variabelens validitet i stor grad.

Verdivurdering

Variabelen *Verdivurdering* er sammensatt av søkeordene «verdivurdering», «gratis verdivurdering» og «verdivurdering av bolig». Vi antar at intensjonen bak søkene er noenlunde lik, og vi har derfor valgt å slå sammen de ulike søkeordene til en variabel. Det antas at de som søker etter verdivurdering på Google er i en prosess hvor de planlegger å selge boligen sin noen måneder senere.

Boligsalg

Variabelen *Boligsalg* representerer søk på «selge bolig» og «salg av bolig». Som for *Boligkjøp* er det vanskelig å definere konkret når søkene forekommer i forhold til salget, men dette burde ikke ekskludere variablene fra oppgaven.

En utfordring ved tolkningen av *Boligsalg* er muligheten for at personer som skal kjøpe søker på salg av bolig for å finne personer som selger bolig, og motsatt. Dette representerer en generell utfordring med Google-søk: Man kan ikke være helt sikker på intensjonen til søkeren.

5.3 Utvalgte makroøkonomiske variabler

De følgende variablene er valgt basert på Jacobsen og Naugs (2004) boligprismodell. En grafisk fremstilling av utviklingen i variablene over tid er å finne i appendiks B2.

Boligmasse

Boligmasse er et absolutt mål på det samlede antallet bruktboliger i Norge den gitte måneden. Det anses for å være en relativt stabil størrelse på kort sikt. Vi ser på utviklingen i boligmarkedet på kort sikt, men har likevel valgt å inkludere boligmasse som en variabel i tilfelle plutselig store økninger eller reduksjoner i boligmassen skulle forekomme i perioden vi studerer. Boligmasse anses som en tilbudsidefaktor, og det er derfor ifølge økonomisk teori nærliggende å anta at en økning i boligmasse fører til reduserte boligpriser.

Lønn

Variabelen *Lønn* representerer nordmenns inntekt, og er inkludert fordi det er sterkt knyttet til deres evne til å betjene lån og deres incentiver til å kjøpe bolig. For de fleste nordmenn er lønn den største inntektskilden, og er derfor en avgjørende faktor når man bestemmer seg for å kjøpe bolig og hvor mye man kan by i budrunder. Det antas at en økning i lønn vil føre til en prisøkning i boligmarkedet. *Lønn* er dermed en etterspørselsfaktor i boligmarkedet. Variabelen måles som en husholdnings lønnsinntekt per kvartal, hvor en husholdning regnes som alle personer bosatt i boligen.

KPI-JAE

KPI-JAE er et mål på norske prisendringer, men inkluderer ikke endringer i energi- og avgiftskostnader. Vi har valgt å inkludere KPI-JAE i stedet for KPI fordi vi tror energipriser og avgiftsendringer har lite å si for etterspørselen etter boliger. Strømpriser er dessuten svært volatile og avhenger i stor grad av sesong, og ved å benytte KPI-JAE vil sesongvariasjonene dempes. Dette kan bidra til å styrke reliabiliteten til variabelen. Det er naturlig å anta at en økning i det generelle prisnivået er assosiert med redusert etterspørsel etter boliger fordi man generelt får dårligere råd. Dette vil igjen lede til fallende boligpriser.

Arbeidsledighet

Ledigheten i arbeidsmarkedet er inkludert som variabelen *Arbeidsledighet*. Det antas at høy ledighet indikerer at mange i samfunnet ikke har råd til å kjøpe bolig, som vil lede til et fall i boligprisene som følge av redusert etterspørsel. Arbeidsledighet kan påvirke boligmarkedet på

to måter. Den første og mest åpenbare effekten er at personer som blir arbeidsledige ikke vil ha råd til å kjøpe ny bolig.

Den andre mer indirekte effekten ved en økning i arbeidsledighet er at befolkningen generelt føler seg mer utrygge på sin egen økonomiske fremtid. Hvis personer rundt deg får sparken kan dette føre til en usikkerhetsfaktor som fører til et fall i etterspørsel av boliger. Selv om effekten ikke er like direkte kan den føre til ringvirkninger for flere enn de som rammes direkte.

Rente

Variabelen *Rente* angir bankenes gjennomsnittlige utlånsrente på lån til husholdninger. Ved en økning i boliglånsrenten vil kostnadene ved å kjøpe bolig øke. Dette fører til et fall i etterspørselen etter bolig og derfor et fall i boligprisene. Renten har holdt seg ganske stabil på rundt fire prosent etter finanskrisen.

Skattesats

Skatt måler marginal skatt på kapitalinntekter og -utgifter. Det er relevant å inkludere variabelen fordi skatten gjelder for inntekt ved utleie av bolig og gevinst ved salg av bolig. Skattenivået kan dermed påvirke interessen etter boligkjøp som investeringsobjekt. Det antas at økt skatt gir redusert etterspørsel, og dermed redusert boligpris.

Forventningsindikatoren til TNS Gallup

Den siste uavhengige variabelen i datasettet er *TNS Gallup*, en forventningsindikator som måler husholdningenes forventninger til egen og landets økonomi. Indikatoren baserer seg på en spørreundersøkelse som gjennomføres hvert kvartal, og regnes ut som differansen mellom prosentandelen optimistiske og pessimistiske svar. En vekst i optimismen i samfunnet vil tyde på at personer er tryggere på fremtiden og mer villig til å investere i bolig. En økning i optimisme vil derfor føre til en økning i etterspørsel etter bolig, og dermed økte boligpriser.

6. Empirisk metode

I dette kapitlet vil vi forklare den empiriske metoden vi har benyttet for å undersøke forskningsspørsmålene. Det første underkapitlet vil presentere regresjonsmetoden vi benytter for å estimere tre ulike boligprismodeller. I det andre underkapitlet vil vi presentere de tre regresjonsmodellene, før vi beskriver metoden for å konvertere kvartalsbaserte data fra Norges Bank til månedsbaserte. Til slutt vil vi teste og eventuelt korrigere for kjente utfordringer med tidsseriedata. Samtlige variabler med beskrivelse av hvordan de er justert er tilgjengelig i appendiks A6.

6.1 Valg av metode

Vi vil benytte minste kvadraters metode, heretter omtalt som OLS, for å estimere de lineære regresjonsmodellene i kapittel syv. Valget av OLS-metoden som estimeringsmetode er gjort på bakgrunn av litteratur på området, som har funnet at enkle lineære regresjonsmodeller presterer like godt eller bedre enn mer sofistikerte ikke-lineære modeller (Wu & Brynjolfsson, 2015). I tillegg benytter Jacobsen & Naug (2004) minste kvadraters metode i sin empiriske modell for boligprisene i Norge.

Problemstillingen vi ønsker å studere er om Google-søk kan forklare utviklingen i det norske boligmarkedet, og vi velger å estimere tre enkle modeller for å besvare spørsmålet. Dersom modellene viser seg å kunne forklare prisutviklingen godt er de sannsynligvis gode på å forutsi fremtidige boligprisfall i tillegg, samt identifisere bobletendenser. Dette vil adresseres i de to påfølgende analysekapitlene.

Den generelle OLS-regresjonsmodellen vi vil benytte kan skrives som følger:

$$Prisindeks_t = \beta_0 + \beta_1 \mathbf{X}_t + \beta_2 \mathbf{Z}_t + u_t, \quad t = 1, 2, \dots, n \quad (1)$$

Regresjonsmodellen er en multippel regresjonsmodell, med flere uavhengige variabler. \mathbf{X}_t er en vektor som representerer Google-søkeordvariablene i måned t , og vektoren \mathbf{Z}_t består av de makroøkonomiske variablene vi vil benytte. u_t er feilleddet som fanger opp uobserverte faktorer som påvirker *Prisindeks*, samt eventuelt feil bruk av funksjon og målefeil i de

observerte variablene i måned t . Koeffisientene β_1 og β_2 representerer endringen i *Prisindeks* som følge av én enhets endring i henholdsvis X_t og Z_t .

Vi har studert hvilke modeller tidligere litteratur på området har benyttet, og ser at flere inkluderer etterslep av den avhengige variabelen i modellene sine. Vårt hovedfokus i masterutredningen er å undersøke om Google-søk kan forklare boligprisutviklingen, og mener det av den grunn er mer hensiktsmessig å ikke inkludere etterslep av den avhengige variabelen. Vi anser det ikke som interessant å undersøke hvilken effekt tidligere verdi av boligprisen har på nåværende.

Vi ønsker derimot å undersøke om *Prisindeks* kan reagere med et etterslep på endringer i Google-variablene, da vi anser det som naturlig at planleggingen og sondering av boligmarkedet resulterer i et boligkjøp eller boligsalg senere enn inneværende måned. Hvor stort etterslepet er vil variere. Vi foretok noen antakelser om lengden på etterslepet for Google-variablene i forrige kapittel. Det ville vært naturlig å benytte en Finite Distributed Lag Model, hvor man tillater at en eller flere uavhengige variabler påvirker *Prisindeks* med et etterslep på en eller flere perioder (Wooldridge, 2016).

Siden etterslep vil bli hensyntatt ved å benytte et glidende gjennomsnitt for de foregående elleve månedene som sesongjusteringsmetode for Google-variablene anser vi det som overflødig å inkludere etterslep som separate variabler i regresjonsmodellen vår. Denne metoden presenteres i neste underkapittel. Modellen vi benytter er derfor en Static Model fremfor en Finite Distributed Lag Model, hvor kun de uavhengige variablene i inneværende periode er inkludert (Wooldridge, 2016).

Vi ønsker å presentere endringene i *Prisindeks* og de uavhengige variablene på prosentform eller andre intuitivt forståelige måleenheter. Boligprisindeksen er konstruert som prosentvise endringer eller som endringer i indeksnivået, og endringene er dermed allerede forståelige. For de uavhengige variablene velger vi å benytte logaritmisk skala for variabler hvor det er naturlig. Dette gjelder samtlige Google-variabler og to makroøkonomiske variabler. Den generelle regresjonsmodellen vil basert på dette endres til følgende form:

$$Prisindeks_t = \beta_0 + \beta_1 \ln(X_t) + \beta_2 \ln(Z_t^a) + \beta_3 (Z_t^b) + u_t, \quad t = 1, 2, \dots, n \quad (2)$$

Her representerer \mathbf{X}_t en vektor bestående av samtlige Google-variabler. En endring på ett prosent i \mathbf{X}_t fører til en $\frac{\beta_1}{100}$ endring i *Prisindeks*. \mathbf{Z}_t^a representerer en vektor av makrovariablene vi tar logaritmen av, og \mathbf{Z}_t^b er en vektor av makrovariablene som ikke er tatt logaritmen av. Dette innebærer at en ett prosent endring i \mathbf{Z}_t^a leder til en $\frac{\beta_2}{100}$ endring i *Prisindeks*, og en enhets endring i \mathbf{Z}_t^b leder til en β_3 endring i *Prisindeks*.

6.2 Regresjonsmodeller

For å studere forskningsspørsmålene er det utarbeidet tre regresjonsmodeller vi vil basere analysen i neste kapittel på, etter inspirasjon fra Dietzel, Braun og Schaefers ved Universitetet i Regensburg (2014). Den første baserer seg kun på Google-variabler, deretter følger en modell utelukkende bestående av makroøkonomiske uavhengige variabler. Den siste er en kombinasjon av begge typer variabler.

6.2.1 Regresjonsmodell med Google-variabler

Regresjonsmodellen som kun består av *Prisindeks* og de utvalgte Google-variablene presentert i forrige kapittel er inkludert primært for å undersøke forskningsspørsmål 1b om Google-søk kan forklare boligprisutviklingen bedre enn makroøkonomiske variabler alene.

Følgende regresjonsmodell har kun Google-variabler som uavhengige variabler:

$$\begin{aligned} \text{Prisindeks}_t = & \beta_0 + \beta_1 \ln(\text{Boliglån}_t) \\ & + \beta_2 \ln(\text{Boliglånsrente}_t) + \beta_3 \ln(\text{Eiendomsmeidler}_t) \\ & + \beta_4 \ln(\text{Selskapsnavn}_t) + \beta_5 \ln(\text{Verdivurdering}_t) + \beta_6 \ln(\text{Boligkjøp}_t) \\ & + \beta_7 \ln(\text{Boligsalg}_t) + \beta_8 \ln(\text{Finn}_t) + u_t, \quad t = 1, 2, \dots, n \end{aligned} \quad (3)$$

$\frac{\beta_1}{100}, \frac{\beta_2}{100}, \frac{\beta_3}{100}, \frac{\beta_4}{100}, \frac{\beta_5}{100}, \frac{\beta_6}{100}, \frac{\beta_7}{100}$ og $\frac{\beta_8}{100}$ representerer endringen i *Prisindeks* som følge av en ett prosent økning i Google-variablene.

6.2.2 Regresjonsmodell med makroøkonomiske variabler

Følgende modell er inkludert som sammenlikningsgrunnlag for modellene som inkluderer Google-variabler, modell 3 og 5. Ved å undersøke hvor godt en modell som utelukkende inkluderer makroøkonomiske variabler presterer, kan vi sammenligne med prestasjonsevnen

til modeller som inkluderer Google-søk. På denne måten får vi et bedre grunnlag for å kunne bekrefte eller avkrefte at Google-søk kan benyttes til å forklare boligprisen.

Regresjonsmodellen med de makroøkonomiske variablene tar følgende form:

$$\begin{aligned} \text{Prisindeks}_t = & \beta_0 + \beta_1(\text{KPI_JAE}_t) + \beta_2(\text{Arbeidsledighet}_t) + \beta_3(\text{Rente}_t) \\ & + \beta_4(\text{Skatt}_t) + \beta_5(\text{TNSGallup}_t) + \beta_6 \ln(\text{Boligmasse}_t) \\ & + \beta_7 \ln(\text{Lønn}_t) + u_t, \quad t = 1, 2, \dots, n \end{aligned} \quad (4)$$

Tolkningen av koeffisientenes effekt på *Prisindeks* vil være lik som for Google-variablene for samtlige makrovariabler. Makrovariablene som ikke er på den logaritmiske skalaen er skalert før modellene estimeres, slik at en økning på ett prosentpoeng i den respektive variabelen gir en $\frac{\beta_i}{100}$ endring i *Prisindeks*.

6.2.3 Regresjonsmodell med både Google- og makrovariabler

For å undersøke forskningsspørsmål 1a om at Google-søk i kombinasjon med makroøkonomiske variabler kan forklare boligprisutviklingen bedre enn makroøkonomiske variabler alene har vi inkludert en regresjonsmodell som kombinerer modell 3 og 4.

Følgende regresjonsmodell kombinerer makroøkonomiske variabler og Google-variabler:

$$\begin{aligned} \text{Prisindeks}_t = & \beta_0 + \beta_1 \ln(\text{Boliglån}_t) \\ & + \beta_2 \ln(\text{Boliglånsrente}_t) + \beta_3 \ln(\text{Eiendomsmeidler}_t) \\ & + \beta_4 \ln(\text{Selskapsnavn}_t) + \beta_5 \ln(\text{Verdivurdering}_t) + \beta_6 \ln(\text{Boligkjøp}_t) \\ & + \beta_7 \ln(\text{Boligsalg}_t) + \beta_8 \ln(\text{Finn}_t) + \beta_9(\text{KPI_JAE}_t) \\ & + \beta_{10}(\text{Arbeidsledighet}_t) + \beta_{11}(\text{Rente}_t) + \beta_{12}(\text{Skatt}_t) \\ & + \beta_{13}(\text{TNSGallup}_t) + \beta_{14} \ln(\text{Boligmasse}_t) + \beta_{15} \ln(\text{Lønn}_t) + u_t, \\ & t = 1, 2, \dots, n \end{aligned} \quad (5)$$

De inkluderte variablene og deres effekt på boligprisindeksen er de samme som forklart for modell 3 og 4.

Modell 5 består av relativt mange variabler. Dette kan redusere frihetsgradene, men kan samtidig bidra til å redusere faren for *omitted variable bias* – at relevante variabler ikke inkluderes i modellen og dermed påvirker *Prisindeks* gjennom feilledet. Det potensielle problemet ved å inkludere for mange variabler er adressert og diskutert senere i utredningen.

6.3 Konvertering fra kvartal til måned

Datasettet fra Norges Bank består av kvartalsbaserte dataserier. Da boligprisindeksen og Google-variablene er månedsbaserte observasjoner må vi konvertere tallene fra kvartalsvise til månedsbaserte. Oust og Eidjord (2017) valgte å løse tilsvarende utfordring med datasettet i sin forskning ved å benytte metoden *Cubic-Spline Interpolation*. Vi har valgt å benytte samme metode da vi jobber med liknende data. Den svekker validiteten til makrovariablene, ettersom konvertering av data fra kvartal til måned alltid vil introdusere unøyaktighet. Cubic-spline interpolation-metoden kan kun gi estimat på data mellom kjente datapunkter (Columbia, 2010).

Alternativet til å konvertere til månedsbaserte tall ville vært å konvertere de månedsbaserte variablene benyttet i oppgaven til kvartalsvise. For å sikre at datasettet vårt består av et tilstrekkelig antall observasjoner har vi valgt å benytte måneder og ikke kvartal som tidsserievariabel.

Vi har kvartalsvise data for hele 2017, men Cubic-Spline Interpolation medfører at vi mister den siste månedsobservasjonen, slik at den siste månedlige observasjonen blir november 2017 for de makroøkonomiske variablene. Unntaket er *Skatt*, da denne endres ved gitte statlige vedtak. Det er derfor ikke hensiktsmessig å konvertere til månedlige tall på samme måte.

6.4 Utfordringer med tidsseriedata

Tidsseriedata utøver ofte autokorrelasjon, ikke-stasjonaritet og heteroskedastisitet som følge av deres ikke-randomiserte dataformat. For at OLS skal være en god estimator må vi undersøke om vi har problemer med disse egenskapene, og eventuelt korrigere for det. Vi tester henholdsvis modell 3, 4 og 5.

6.4.1 Ikke-stasjonære data

For å unngå spuriøse sammenhenger i modeller med tidsserier er det nødvendig å operere med data der trend er eliminert (Wooldridge, 2016). Dersom to variabler beveger seg i samme retning kan det hende at tid er den største forklaringsfaktoren, og ikke at den uavhengige variabelen faktisk forklarer variasjonen i den avhengige variabelen. Som vi poengterte i kapittel fem følger de fleste Google-variabler og boligprisindeksen samme trendutvikling.

Flere av de makroøkonomiske variablene vi har inkludert følger også en positiv trend, slik som *Boligmasse* og *Lønn*. Uten å gjøre variablene stasjonære kan man feilaktig tolke veksten i boligprisen med at nordmenn googler mer i dag enn for ti år siden. Dette virker lite sannsynlig – det virker mer sannsynlig at de har en felles trendutvikling som følger tiden. For å unngå å tolke slike spuriøse sammenhenger som kausale må vi eliminere effekten av trend på dataseriene vi opererer med før vi estimerer modellene i neste kapittel.

Av figur 1 ser vi tydelige tegn til at boligprisutviklingen følger en positiv trend. For å undersøke formelt om vi har et problem med ikke-stasjonære data har vi benyttet en Augmentet Dickey-Fuller-test, og finner at vi ikke kan forkaste nullhypotesen om at prosessen har unit root, som forventet. Dette innebærer at vi ikke kan konkludere med at vi har en stasjonær prosess, og vi velger å fjerne trenden i prisindeksen fra dataserien for å unngå spuriøse resultater. Eliminering av trend vil bli forklart senere i kapitlet.

Sesongjustering

Av samme grunn som for data med en sterk tidstrend er det nødvendig å sesongjustere data hvis det er noen grunn til å tro at den utøver et gjentakende sesongmønster hvert år.

Vi har i masterutredningen benyttet Simple Moving Average av tolv måneder på samtlige Google-variabler. Av grafene for Google-søkenes utvikling i appendiks B1 ser vi tydelige tegn til at søkevolumet følger et sesongmønster basert på at januar alltid er lavere enn månedene før og etter. Det er dessuten logisk at flere benytter Google visse måneder enn andre, for eksempel at færre benytter det i sommerferiemånedene, og at dette vil skape et gjentakende sesongmønster.

For å sesongjusterte Google-data har vi valgt et ikke-vektet glidende gjennomsnitt på inneværende måned og de foregående elleve månedene¹, kalt Simple Moving Average. Grunnen til at vi har valgt å benytte glidende gjennomsnitt er som forklart tidligere at datapunktene kan variere noe ettersom når man laster ned dataseriene fra Google Trends. Dette kan delvis korrigeres for ved å benytte et glidende gjennomsnitt. Google-data har dessuten

¹ Følgende formel er benyttet for å sesongjustere samtlige Google-variabler, basert på *The Science of Financial Market Trading* (2003) av Don K. Mak:

$$\text{Sesongjustert verdi} = \frac{1}{12} \sum_{i=0}^{11} (X_{12-i})$$

noen manglende verdier frem til 2008. Ved å benytte glidende gjennomsnitt kan vi kompensere for dette.

Vi har valgt å beregne gjennomsnittet av inneværende og de elleve foregående månedene heller enn å benytte et sentrert glidende gjennomsnitt av hensyn til hensikten med masterutredningen. Vi ser for oss at en av de største fordelene ved å benytte Google-data fremfor tradisjonelle makroøkonomiske variabler er at dataene er tilgjengelig i tilnærmet nåtid. Hvis oppgaven hadde benyttet et sentrert glidende gjennomsnitt ville det stilt krav til at vi hadde fremtidig data for å kunne beskrive dagens boligpris.

Dersom vi hadde benyttet en kortere periode enn tolv måneder ville vi ikke ha tatt hensyn til sesongjustering i like stor grad. Ved å inkludere tolv måneder inkluderes samtlige sesonger i gjennomsnittsberegningen. Grunnen til at vi ikke har valgt en lengre periode enn tolv måneder er hovedsakelig fordi det kan gjøre sesongjusteringen mindre korrekt dersom flere måneder inkluderes, da samme måned i to ulike år inkluderes.

En fordel ved sesongjusteringsmetoden vi har valgt er at den demper effekten av verdier som er svært høye eller svært lave, såkalte *outliers*. Det styrker reliabiliteten til Google-variablene. Som nevnt tidligere er Google-data fra tidligere perioder preget av sterk variasjon, med enkelte observasjoner lik null etterfulgt av svært høye verdier. Med glidende gjennomsnitt vil dataene bli mer stabil uten at all variasjon i variablene elimineres. Vi ønsker å beholde noe variasjon i de uavhengige variablene for å kunne forklare variasjonen i *Prisindeks*.

En ulempe ved metoden vi benytter sammenlignet med metoden Eiendom Norge benytter er at vi ikke gis anledning til å korrigere for spesielle hendelser i løpet av året, slik som påskeferien og sommerferien, da vektingen av de ulike månedene er lik. Dette er en svakhet man bør være klar over, men som vi ikke anser som et stort problem.

Vi anser det ikke som nødvendig å sesongjusterte de makroøkonomiske variablene da de ikke virker å ha et sterkt sesongmønster. Prisindeksen er allerede sesongjustert av Eiendom Norge.

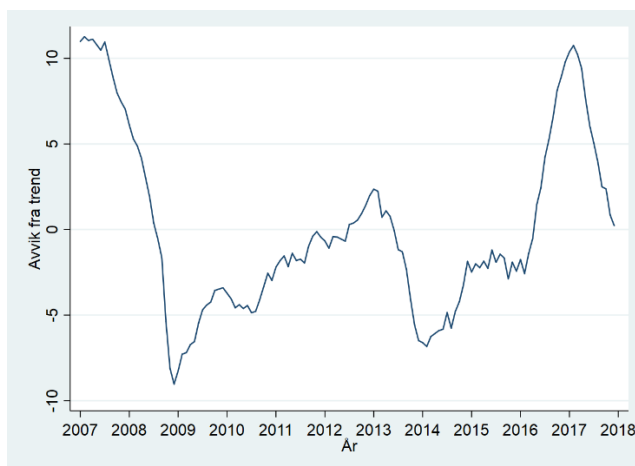
Ideelt sett burde vi ha sesongjustert samtlige variabler etter samme metode som den Eiendom Norge benytter. Vi ikke har tilgang på detaljert informasjon om metoden de benytter, og anser derfor valgte løsning som den mest hensiktsmessige.

Eliminering av trendstøy

I perioden vi studerer har man sett en tilnærmet kontinuerlig vekst i boligprisene. Flere av de uavhengige variablene utøver også en trend i større eller mindre grad, som vi tydelig ser i appendiks B.

For å unngå at en eventuell spuriøs sammenheng mellom økning i bruken av internett og økning i boligprisen vurderes som kausal har vi valgt å fjerne effekten av tidstrender fra prisindeksen, samtlige Google-variabler og makrovariablene boligmasse, lønnsinntekt og KPI-JAE da vi antar at disse variablene følger en positiv trend over tid, basert på den grafiske fremstillingen av variablene i appendiks B. På denne måten vil forklaringsgraden kunne tolkes som i hvor stor grad de uavhengige variablene faktisk kan forklare variasjonen i boligprisen.

Figur 2 viser prisindeksens avvik fra trend. Målet var å gjøre dataserien stasjonær, og vi skulle gjerne sett at den i større grad fluktuerte rundt null. Vi anser den likevel som stasjonær nok til å utføre regresjoner, som bekreftes på et fem prosent signifikansnivå ved å teste den trendjusterte serien med Augmented Dickey Fuller.



Figur 2: Trend- og sesongjustert prisindeks. Årstallet på x-aksen er satt til januar det respektive året. Kilde: Eiendom Norge

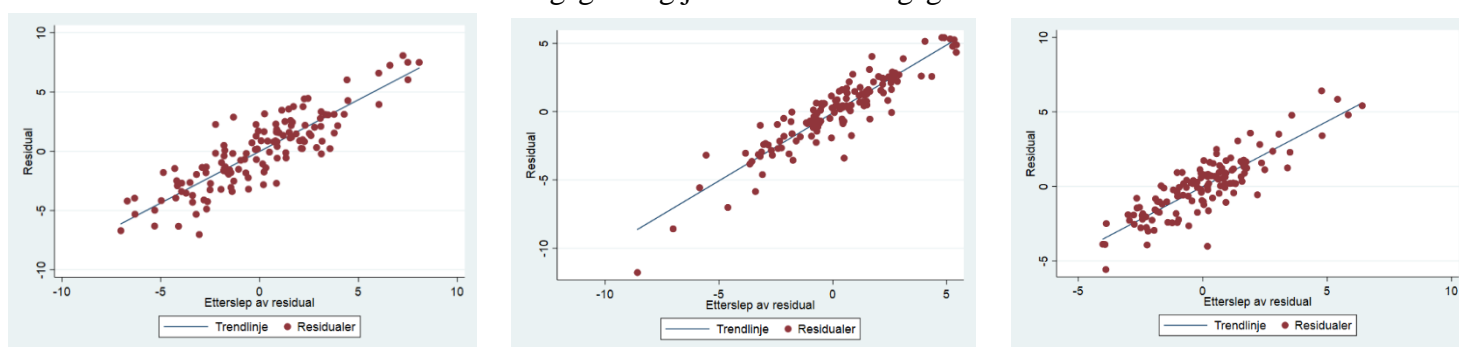
Grafene som demonstrerer estimert boligprisindeks og virkelig boligprisindeks videre i oppgaven tolkes som prisnivå utover trenden, da trendstøy og sesongmønstre er fjernet. Null

på y-aksen indikerer at prisen ligger på samme nivå som trenden, positive y-verdier er indekspoeng over trend og negative y-verdier er indekspoeng under trend.

6.4.2 Autokorrelasjon

Autokorrelasjon er et problem dersom feilleddene til en tidsseriemodell i ulike tidsperioder er korrelerte (Wooldridge, 2016). Ved å studere figur 3 ser vi tydelig at autokorrelasjon er et problem for henholdsvis modell 3, 4 og 5 etter at trend og gjentakende sesongmønstre er fjernet fra samtlige variabler der vi anser det som nødvendig. Samtlige grafer utøver tydelige tegn på positiv autokorrelasjon, som bekreftes ved å teste for det formelt. Det er ikke overraskende at modell 4 og 5, basert på makroøkonomiske variabler, utøver autokorrelasjon, med tanke på at vi har benyttet Cubic-Spline Interpolation for å konvertere kvartalsdata til månedsdata – autokorrelasjon er et kjent problem ved denne metoden, da datapunktene som interpoleres naturlig nok vil være systematisk relatert til hverandre (Columbia Economics, 2010).

Autokorrelasjon må korrigeres for før OLS kan benyttes på tidsseriedataene, og vi vil benytte Newey West-standardavvik videre i utredningen. Dette gjør inferens robust mot både heteroskedastisitet og autokorrelasjon. Det vil ikke påvirke koeffisientene sammenlignet med en OLS-regresjon, men vil kunne påvirke standardavvikene, t-verdier og p-verdier. R^2 og justert R^2 blir ikke påvirket siden koeffisientene ikke påvirkes, og videre i utredningen har vi derfor benyttet R^2 og justert R^2 fra OLS-regresjoner basert på de samme variablene når vi diskuterer modellenes forklaringsgrad og justerte forklaringsgrad.



Figur 3: Residualverdien til modell 3, 4 og 5 mot et etterslep av residualen på en periode.

6.4.3 Heteroskedastisitet

Homoskedastisitet innebærer at variansen til feilleddet er konstant og ikke avhenger av de uavhengige variablene. Ved å benytte Breusch-Pagan/Cook-Weisberg-test for heteroskedastisitet i STATA² for sesong- og trendjustert data finner vi at nullhypotesen om konstant varians ikke kan forkastes for modell 3, men kan forkastes på et fem prosent signifikansnivå for modell 4 og 5. Dette innebærer at vi har et problem med heteroskedastisitet for modell 4 og 5, som må korrigeres før vi kan benytte OLS. Ved å benytte Newey West-standardavvik, som beskrevet i forrige avsnitt, vil standardavvikene gjøres robuste mot heteroskedastisitet, og vil dermed løse problemet med heteroskedastisitet.

² Vi har benyttet testen *estat hettest* i Stata for de tre regresjonsmodellene for å teste om vi har et problem med heteroskedastisitet.

7. Empirisk analyse

I analysearbeidet oppdaget vi at de empiriske resultatene varierte i stor grad avhengig av om observasjoner fra 2017 var inkludert eller ikke. Vi har derfor inkludert regresjonsresultater og analyse basert på hele perioden, fra 2007 til 2017, og fra 2007 til 2016. De estimerte modellene for perioden 2007 til 2016 vil få 2016 som et tillegg i modellnavnet for å skille dem fra modellene estimert på hele perioden.

Ved å estimere modell 3, 4 og 5 (se appendiks C) inkluderes svært mange uavhengige variabler. Dette reduserer antall frihetsgrader. Variablene inkludert i de tre modellene er variabler vi anså som potensielt relevante i å forklare boligprisutviklingen, og vi ønsker nå å undersøke hvilke som har en signifikant effekt på boligprisindeksen. Vi har derfor kun valgt å beholde statistisk signifikante variabler i de estimerte modellene som presenteres i dette kapitlet.

Underkapitlene blir presentert i følgende rekkefølge: To estimerte modeller med Google-variabler (forventningsmodeller), to modeller basert på makrovariabler (fundamentalmodeller) og to modeller basert på både makro- og Google-variabler (hybridmodeller). For de tre underkapitlet vil modellene først estimeres, før en analyse av modellene fremstilt grafisk vil benyttes til å indikere om Google-søk kan benyttes til å forutse boligprisnedganger.

7.1 Forventningsmodeller

Vi har valgt å kalle de estimerte modellen utelukkende bestående av Google-variabler for forventningsmodeller. Innledningsvis i masterutredningen beskrev vi hvordan vi anser Google-søk på boligrelaterte ord å reflektere forventninger til fremtidig handling i boligmarkedet, som er grunnlaget for navnevalget. Aggregert vil dette gjør Google-variablene til mål på investorers forventninger til handlinger i markedet. Det kan derfor argumenteres for at en økning eller fall i søk på disse ordene vil reflektere samfunnets forventninger til hva som kommer til å skje i boligmarkedet. Derfor har vi kalt modellen basert på Google-variablene en forventningsmodell.

For å finne den beste modellen basert på Google-variablene alene starter vi med å estimere modell 3 fra kapittel seks, som inkluderer samtlige Google-variabler, se regresjonsutskrift i appendiks C1. Vi kommer frem til følgende finjusterte regresjonsmodell, hvor samtlige uavhengige variabler har en signifikant effekt på boligprisindeksen, for det komplette datasettet (for perioden 2007 til 2017):

$$\begin{aligned} \text{Prisindeks}_t &= \beta_0 + \beta_1 \ln(\text{Boliglån}_t) + \beta_2 \ln(\text{Eiendomsmegler}_t) + u_t, \\ t &= 1, 2, \dots, n \end{aligned} \quad (6)$$

Vi kaller modell 6 for Forventningsmodellen (forkortet til Forventning).

Ved å benytte samme metode på data fra 2007 til 2016 blir flere variabler signifikante, og vi får følgende regresjonsmodell:

$$\begin{aligned} \text{Prisindeks}_t &= \beta_0 + \beta_1 \ln(\text{Boliglån}_t) + \beta_2 \ln(\text{Eiendomsmegler}_t) \\ &+ \beta_3 \ln(\text{Selskapsnavn}_t) + \beta_4 \ln(\text{Boligsalg}_t) \\ &+ \beta_5 \ln(\text{Boliglånsrente}_t) + u_t, \quad t = 1, 2, \dots, n \end{aligned} \quad (7)$$

Vi kaller modell 7 for Forventningsmodellen2016 (forkortet Forventning2016).

7.1.1 Regresjonsresultater med forventningsmodellene

Her vil variablene inkludert i forventningsmodellene presenteres med hensyn til fortegn og størrelse på de estimerte koeffisientene. Vi har i tillegg inkludert en kort analyse av om fortegnet samsvarer med hva som forventes basert på økonomisk teori.

Fullstendig datasett: 2007 til 2017

Ved å estimere modell 6 finner vi følgende finjusterte modell basert på Google-variabler (se tabell 3 for regresjonsutskrift):

$$\widehat{\text{Prisindeks}}_t = -33,73 \text{Boliglån}_t + 49,03 \text{Eiendomsmegler}_t \quad (8)$$

Det er ikke overraskende at *Boliglån* og *Eiendomsmegler* begge er signifikante. Litteraturen presentert i kapittel 3 trekker frem nettopp disse ordene som signifikante i andre land. Søk på boliglånsrelaterte ord og «eiendomsmegler» fremheves som de beste variablene til å forutsi boligprisen i England og Nederland. Det som er bemerkelsesverdig er at fortegnene til variablene er motsatt av hva som kan forventes etter økonomisk teori, ettersom *Boliglån*

forventes å være knyttet til etterspørsel og *Eiendomsmegler* forventes å være en tilbudsvariabel.

Boliglån

Koeffisienten for *Boliglån* i tabell 3 viser i motsetning til antagelsen basert på økonomisk teori i dette tilfelle at en økning i søk etter boliglån fører til at boligprisen faller. Prisene faller med 0,34 indekspoeng når boliglånsvariabelen øker med ett prosent.

En forklaring kan være at boligmarkedet er et spesielt marked i den forstand at en stor andel av deltakerne både er kjøpere og selgere i samme periode. Det innebærer at aktivitet fra personer som søker etter typiske etterspørselsvariabler, slik som *Boliglån*, også tyder på at de vil bidra til aktivitet på tilbudssiden av markedet i nærmeste fremtid.

Eiendomsmegler

Variabelen *Eiendomsmegler* kan tolkes både som en tilbudsside- og etterspørselssidefaktor, avhengig av om de som søker etter dette på Google ønsker en megler for å bistå salg eller kjøp av bolig. Koeffisienten i tabell 3 viser at *Eiendomsmegler* har et positivt fortegn, som indikerer at *Eiendomsmegler* er relatert til etterspørselssiden av markedet. Dette kan bety at vår antakelse om at de fleste som benytter seg av eiendomsmegler skal selge er feil, og at det heller er motsatt.

Utviklingen i søk på *Eiendomsmegler* følger et kontraintuitivt mønster, ved at populariteten synker over perioden vi ser på. Dette kan være en mulig forklaring på hvorfor den estimerte koeffisienten for *Eiendomsmegler* er positiv. En annen forklaring kan være, som med boliglånsøk, at samme individ er kjøper og selger av bolig i samme marked, og at Google-søk etter typiske tilbuds- og etterspørselsfaktorer dermed kan overlapse.

Variabelen er statistisk signifikant på et 0,1 prosents nivå, og en ett prosents økning i søk på eiendomsmegler øker boligprisen med litt under et halvt indekspoeng.

Avkortet datasett: 2007 til 2016

Ved å estimere modell 7 finner vi følgende finjusterte modell basert på Google-variabler for det avkortede datasettet (se tabell 3 for regresjonsutskrift), estimert på perioden 2007 til 2016:

$$\widehat{Prisindeks}_t = -35,22Boligl\ddot{a}n_t + 53,43Eiendomsme\ddot{g}ler_t - 32,94Selskapsnavn_t + 9,477Boligsalg_t + 8,650Boligl\ddot{a}nsrente_t \quad (9)$$

Fem av de opprinnelig utvalgte \ddot{a}tte Google-variablene er signifikante p\ddot{a} et fem prosents niv\ddot{a}. I denne modellen er begge variabler som omhandler boligl\ddot{a}n, *Boligl\ddot{a}n* og *Boligl\ddot{a}nsrente*, signifikante. Dette er i og for seg logisk ettersom litteraturen viser at boligl\ddot{a}nsrelaterte s\ddot{o}k er gode til \ddot{a} forutsi vekst i boligmarkedet. Mer interessant er det at disse tilsynelatende like variablene har motsatt fortegn. Det samme er tilfellet for *Eiendomsme\ddot{g}ler* og *Selskapsnavn*, som begge er signifikante. Den femte variabelen inkludert i modellen er *Boligsalg*. Det er interessant at denne er inkludert da den ikke ser ut til \ddot{a} ha v\ddot{e}rt viktig i andre land, basert p\ddot{a} litteraturgjennomgangen. Denne variabelen virker \ddot{a} p\ddot{a}virke prisindeksen motsatt av det som forventes fra \ddot{o}konomisk teori.

Vi vil n\ddot{a} diskutere variablene i modellen. *Boligl\ddot{a}n* og *Eiendomsme\ddot{g}ler* er allerede diskutert for det fullstendige avsnittet og vil ikke diskuteres igjen, da de har samme fortegn i modell 8 og 9 og de estimerte koeffisientene er ikke nevneverdig forskjellige.

Selskapsnavn

At s\ddot{o}k p\ddot{a} spesifikke eiendomsme\ddot{g}lerselskaper f\ddot{o}rer til et fall i boligpriser virker intuitivt logisk, ettersom det vanligvis er selger av boligen som benytter seg av eiendomsme\ddot{g}ler. Tolkningen av variabelen er mer logisk enn ved \ddot{a} tolke den lignende variabelen, *Eiendomsme\ddot{g}ler*. En \ddot{o}kning p\ddot{a} ett prosent for *Selskapsnavn* f\ddot{o}rer til et fall p\ddot{a} 0,33 indekspoeng for boligprisindeksen. Det er verdt \ddot{a} bemerke det relativt h\dd{o}ye standardavviket, p\ddot{a} hele 14,35, nesten halvparten av koeffisientens st\dd{o}rrelse.

Boligsalg

For s\ddot{o}k p\ddot{a} salg av bolig er fortegnet motsatt av hva man skulle anta i henhold til \ddot{o}konomisk teori. Forklaringen kan ligge i at et stort antall boligtransaksjoner i Norge involverer personer som i n\dd{e}rmeste fortid eller fremtid har v\dd{e}rt p\dd{a} motsatt side i en liknede transaksjon: De fleste som k\dd{j}per bolig i Norge har solgt eller har planer om \dd{a} selge boligen sin like etter de har k\dd{j}pt (Eiendomsverdi, 2018). Dette kan forklare hvorfor et s\dd{o}k p\dd{a} salg av bolig ogs\dd{a} kan tolkes som en ettersp\dd{o}rselsvariabel. Dermed virker det logisk at boligprisen stiger ved \dd{o}kt s\dd{o}keaktivitet etter salg av bolig.

En annen forklaring kan være at søkeren har motsatt hensikt av hva ordlyden skulle tilsi. Vi beskrev dette som en potensiell utfordring med denne variabelen og *Boligkjøp* i kapittel fem. Søk på salg av bolig kan foretas an en person som leter etter en som selger bolig, heller enn at personen selv planlegger å selge.

Av tabell 3 kan vi se at en økning på ett prosent i *Boligsalg* fører til en økning på 0,095 indekspoeng i boligprisindeksen, altså en mindre effekt enn for *Eiendomsmegler*, *Selskapsnavn* og *Boliglån*.

Boliglånsrente

Ifølge tabell 3 vil en økning i søk på «boliglånsrente» på ett prosent føre til en økning på 0,087 indekspoeng i boligprisindeksen, også denne effekten mindre enn for *Eiendomsmegler*, *Selskapsnavn* og *Boliglån*. Fortegnet til *Boliglånsrente* er logisk i henhold til økonomisk teori: En økning i søk fører til en økning i pris, ettersom en økende interesse for boliglånrentens nivå kan tyde på at flere ønsker å ta opp boliglån med intensjonen om å kjøpe hus.

Denne variabelen er mer økonomisk intuitiv enn den liknende variabelen *Boliglån*, som har motsatt fortegn. Det kan tenkes at personer som søker etter renter skiller seg fra de som søker kun etter boliglån, men dette er vanskelig å finne bevis på.

Sammenligning av Forventning og Forventning2016

De to estimerte modellene, *Forventning* og *Forventning2016*, har ikke oppsiktsvekkende store standardavvik, sett bort fra *Selskapsnavn*. Dette indikerer at estimatoren benyttet er effektiv.

Det er interessant å bemerke at den justerte forklaringsgraden stiger med ti prosentpoeng ved å ekskludere 2017 fra datagrunnlaget modellene estimeres på. Dette anses som en betydelig forskjell.

Øvrige bemerkninger

Av Google-variablene inkludert i modell 3 er det tre variabler som verken er signifikante i modell 8 eller 9. Metoden for å korrigere for ustabil data og sesongjustering for Google-variablene beskrevet i forrige kapittel kan ha gjort variabler som kun er signifikante over en kort periode ikke-signifikante som følge av glidingen over tolv måneder. Dersom tidsperioden fra interessen for kjøp eller salg av bolig, i form av et Google-søk, til et gjennomført kjøp eller

salg er kort, for eksempel tre måneder, vil ikke nødvendigvis effekten fanges opp like godt av modellene våre.

Dette er særlig relevant for søk vi antar foretas kort tid før boligtransaksjonen, som *Finn* og *Boligkjøp*. Vi antar dette fordi boligene du søker etter på Finn eller ønsker å finne ved å søke på «kjøp av bolig» eller liknende på Google antakelig ikke vil være der mange måneder frem i tid. For slike variabler er ikke gliding over tolv måneder det mest hensiktsmessige for å justere data, og det burde vurderes om variablene inkludert i modellen heller burde blitt vurdert individuelt når det kommer til valg av glattingslengde.

For de øvrige Google-variablene i tabell 3 virker det rimelig å anta at søk på ordene variablene representerer foretas lenge før boligtransaksjonene gjennomføres. Grunnen til dette er at det ikke finnes noen korte tidsfrister for det variablene representerer: Finansieringsbevis varer i tre måneder og kan fornyes flere ganger uten store problemer, kontakt med eiendomsmegler medfører ikke et umiddelbart krav om salg eller kjøp, og verdivurdering av boligen vil ikke endre seg i stor grad over noen måneder.

Forklaringsvariabel	Forventning Avhengig variabel: Prisindeks (trend- og sesongjustert)	Forventning2016
Boliglån	-33,73*** (3,564)	-35,22*** (4,560)
Eiendomsmegler	49,03*** (8,854)	53,43*** (9,804)
Selskapsnavn		-32,94* (14,35)
Boligsalg		9,477** (3,473)
Boliglånsrente		8,650** (2,834)
Konstantledd	-3,14e-09 (0,567)	-0,343 (0,504)
<i>Observasjoner</i>	132	120
<i>Adj. R²</i>	0,5599	0,6592

Newey West standardavvik i parentes.

* $p < 0,05$ ** $p < 0,01$, *** $p < 0,001$

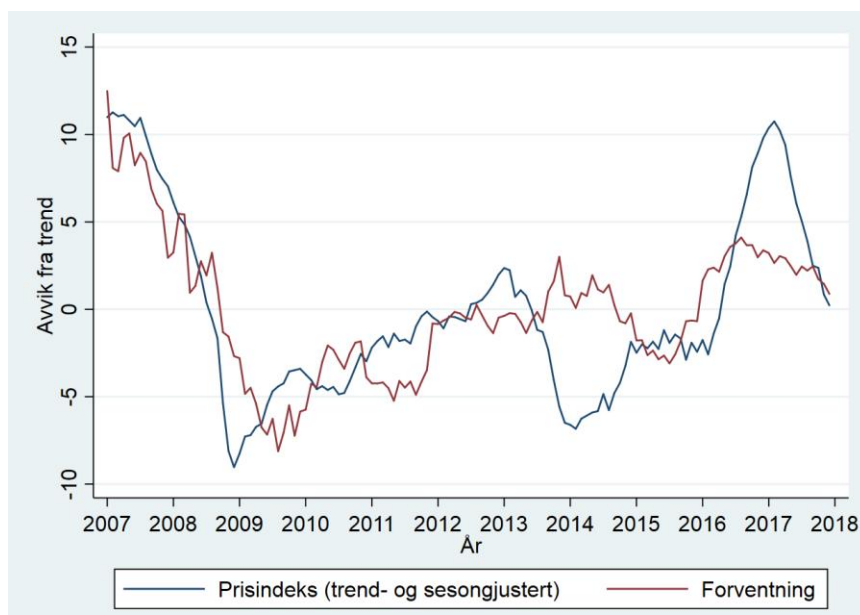
Tabell 3: Regresjonsresultater for forventningsmodellen, for data for hele perioden (Forventning) og frem til og med desember 2016 (Forventning2016).

7.1.2 Forventningsmodeller benyttet til å forutsi boligprisen

Fullstendig datasett: 2007 til 2017

Figur 4 viser boligprisindeksen og den estimerte boligprisindeksen ved å estimere Forventningsmodellen for hele datasettet (Forventning), modell 8. Der grafene ligger tett gir Forventningsmodellen et godt estimat, og det motsatte der avstanden er stor. Modellen sammenfaller godt med prisindeksen i begynnelsen av perioden, til tross for noe etterslep i vekstperioden etter finanskrisen i 2009. Deretter når forventningsmodellen et lokalt bunnpunkt i midten av 2010 uten noe liknende fall i prisindeksen.

Etter dette fallet flater forventningsmodellen noe ut, til tross for at prisene faller over en lengre periode i 2013. Fra 2015 passer modellen godt før den begynner å falle seks måneder før prisfallet i 2017. Totalt passer modellen godt i begynnelsen og slutten av den studerte perioden, mens den avviker mye i perioden fra 2010 til høsten 2014, med kortere perioder hvor den passer godt. Her følger den prisindeksen dårlig, og klarer desto mindre å forutsi boligprisfall.



Figur 4: Prisindeksen og Forventningsmodellen for 2007 til 2017. Årstallene på x-aksen indikerer januar det respektive året.

Det mest interessante å påpeke ved den grafiske fremstillingen er at modellen faller før de to største prisfallene, i 2007 og 2017, men fanger ikke opp prisfallet i 2013. Mens prisindeksen faller fra begynnelsen av 2013 holder Forventningsmodellen seg stabil, og øker til og med noe når prisindeksen stuper i midten av 2013.

Det mest interessante fallet i boligprisen er fallet som starter i april 2017. Etter en kraftig oppgangsperiode fra starten av 2016 snur utviklingen i april 2017, etter at regjeringen innførte strengere boliglånsvilkår ved boliglånsforskriften (Regjeringen, 2016). Vi hevder ikke at den fullstendige forklaringen på prisfallet er boliglånsforskriften, men vi antar at dette kan ha hatt en stor påvirkning. Forskriftene ble innført i januar 2017, men personer med innvilget finansieringsbevis kunne fortsatt ta opp lån på de gamle betingelsene. Et finansieringsbevis har varighet i tre måneder, så det er ikke overraskende at man ser at fallet i boligprisene finner sted i fjerde måned i 2017.

Fallet i Forventningsmodellen begynner allerede i august året før. Det interessante ved dette er at august er fire måneder før endringene i boliglånsforskriftene annonseres. Det betyr at søkerne ikke kan være påvirket av endringen i boliglånskravene. Vi hevder ikke at Forventningsmodellen presterte å forutse regjeringens planer om å endre boliglånsforskriften, men forklaringen kan være at den sterke veksten i markedet frem til 2017 kan ha ført til et mettet boligmarked.

Avkortetdatasett: 2007 til 2016

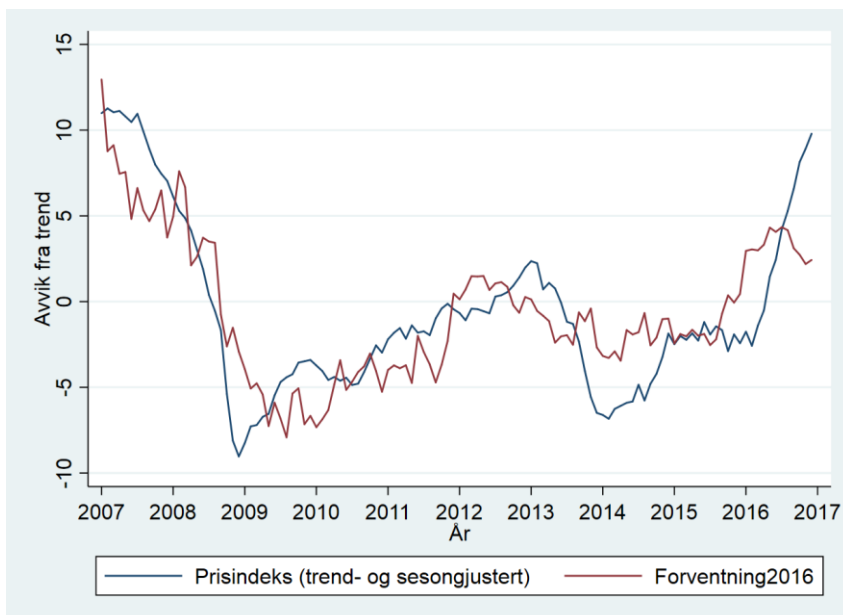
I figur 5 er forventningsmodellen, Forventning2016, estimert på datasettet frem til og med 2016 grafisk fremstilt i samme figur som den virkelige boligprisindeksen. Nå klarer forventningsmodellen å forutse samtlige store prisfall i perioden, også for 2013.

Etter juli 2007 begynner prisindeksen å stupe, antakelig som en ringvirkning av Finanskrisen. Forventning2016-modellen starter å falle allerede fra februar. Deretter faller modellen til midten av 2009, med enkelte vekstperioder innimellom. Den generelle trenden er dog nedadgående hele perioden.

Det andre større fallet i boligprisen fant sted etter toppunktet i januar 2013. Prisveksten faller da frem til februar 2014. Også i dette tilfellet starter nedgangen i Forventning2016 før prisindeksen, selv om fallet for Forventning2016 ikke er like markant som for boligprisen. Nøyaktig når fallet i Forventning2016-modellen begynner er noe diskutabelt. Modellen når et lokalt toppunkt allerede i juni 2012, men det er først i november at en stabil og mer markant nedgang begynner. Perioden fra juni til november kan muligens forklares ved svingninger i datasettet, men fra november er fallet konstant. Dette betyr at fallet i Forventning2016 kommer mellom tre og seks måneder i forkant av fallet i prisindeksen.

Som for forventningsmodellen estimert på hele datasettet ser det ut til at Forventning2016 klarer å forutsi prisfallet i april 2017, men vi vil ikke gå nøyere inn på dette i dette avsnittet da modellen kun strekker seg til desember 2016.

Sammenlignet med figur 4 legger vi merke til noen markante forskjeller. Det mest påfallende er at Forventning2016 følger den virkelige prisindeksen bedre enn Forventning; den har ikke lenger det store avviket i perioden fra 2010 til 2015. En annen markant forskjell er at Forventning2016 faller i forkant av de tre største prisfallene, inkludert fallet i 2013. For hele datasettet viste Forventningsmodellen en generell oppgang i samme periode. Til og med før 2017 begynner Forventning2016 å falle, selv om prisindeksen benyttet til å estimere modellen fortsetter å stige.



Figur 5: Prisindeksen og Forventningsmodellen for 2007 til 2016. Årstallene på x-aksen indikerer januar det respektive året.

7.2 Fundamentalmodeller

Vi kommer frem til de finjusterte modellene basert på de makroøkonomiske variablene på samme måte som for forventningsmodellen. Siden de makroøkonomiske variablene inkludert her er de samme som Jacobsen og Naug (2004) ordrett beskriver som fundamentalverdier i sin modell.

Ved å inkludere signifikante makrovariabler får vi følgende regresjonsmodell for det komplette datasettet, perioden 2007 til 2017:

$$\begin{aligned} \text{Prisindeks}_t &= \beta_0 + \beta_1 \ln(\text{Lønn}_t) + \beta_2 \text{KPI_JAE}_t + \beta_3 \text{Ledighet}_t + \beta_4 \text{Skatt} + u_t, \\ t &= 1, 2, \dots, n \end{aligned} \quad (10)$$

Den samme metoden er utført på data fra 2007 til 2016, som gjør at en ekstra variabel blir signifikant, *Rente*:

$$\begin{aligned} \text{Prisindeks}_t &= \beta_0 + \beta_1 \ln(\text{Lønn}_t) + \beta_2 \text{KPI_JAE}_t + \beta_3 \text{Ledighet}_t + \beta_4 \text{Skatt} \\ &+ \beta_5 \text{Rente}_t + u_t, \quad t = 1, 2, \dots, n \end{aligned} \quad (11)$$

Det bør nevnes at modellen ikke er et forsøk på å gjenskape Jacobsen og Naugs boligprismodell fra 2004, men heller en enkel modell som reflekterer effekten av endringer i sentrale makroøkonomiske variabler på boligprisen.

7.2.1 Regresjonsresultater med fundamentalmodellene

I denne seksjonen vil variablene inkludert i fundamentalmodellen presenteres med hensyn til fortegn og størrelse på de estimerte koeffisientene. Vi har i tillegg inkludert en kort analyse av om fortegnet samsvarer med hva som forventes etter antakelser foretatt i beskrivelsen av variablene i kapittel fem.

Fullstendig datasett: 2007 til 2017

Ved å estimere modell 10 finner vi følgende finjusterte modell basert på makroøkonomiske variabler (se tabell 4 for regresjonsutskrift):

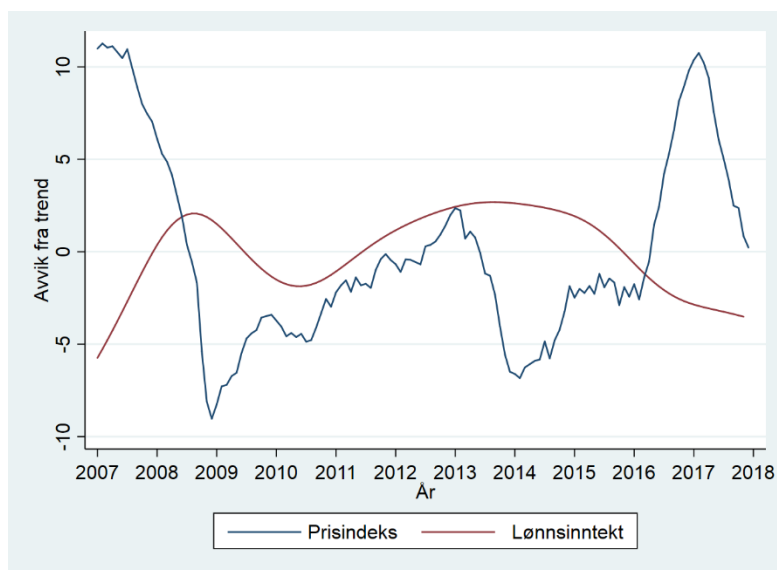
$$\begin{aligned} \widehat{\text{Prisindeks}}_t &= 66,85 - 180,7 \text{Lønn}_t - 276,6 \text{KPI} - \text{JAE}_t - 701,4 \text{Arbeidsledighet}_t \\ &- 178,7 \text{Skatt} \end{aligned} \quad (12)$$

Av de syv variablene som det makroøkonomiske datasettet består av er kun fire signifikante og inkludert i Fundamentalmodellen. *Boligmasse*, *TNSGallup* og *Rente* er ikke signifikante og er derfor utelatt. Koeffisientene i tabell 4 er mer intuitivt forståelige enn for forventningsmodellen, som ikke er veldig overraskende basert på at de fleste makrovariablene per definisjon baserer seg på økonomisk teori.

Lønn

Lønn er den eneste makroøkonomiske variabelen som ikke påvirker boligprisen slik antatt i henhold til økonomisk teori. Av tabell 4 ser vi at en økning i lønn på ett prosent vil føre til et fall i boligprisindeksen på 1,8 indekspoeng, som er relativt stort med tanke på at fortegnet er motsatt av hva vi antok i henhold til økonomisk teori. En skulle tro at en økning i lønninger vil føre til høyere priser i boligmarkedet ettersom økt lønn gir mer å betjene boliglånet med.

En mulig forklaring på det mindre intuitive fortegnet kan ligge i at den studerte perioden inkluderer Finanskrisen og kjølvannet av denne. Figur 6 viser at *Lønn* og *Prisindeks* beveger seg i motsatt retning av hverandre store deler av perioden. Dette kan potensielt forklares av at lønn er «sticky», altså at den endrer seg tregt, og kan forklare hvorfor lønn vokser fram til 2009, rundt tidspunktet der boligmarkedet slutter når bunnen. Deretter begynner lønn å falle, mens boligmarkedet stiger. Med unntak av perioder 2010 til 2013 har lønn og prisindeksen motsatt fortegn. Dette kan antagelig forklares ved at transaksjoner i boligmarkedet kan respondere raskt på eksogene faktorer, mens arbeidsmarkedet i Norge er regulert slik at det er vanskelig å senke lønninger raskt.



Figur 6: Boligprisindeksen og Lønn i perioden 2007 til 2017. Årstallet på x-aksen indikerer januar det respektive året.

KPI-JAE

En økning på en enhet i *KPI-JAE*-verdien fører ifølge modellen til et fall i boligprisindeksen på 2,77 indekspoeng. Dette er logisk i henhold til økonomisk, ettersom en økning i det generelle prisnivået vil gjøre det vanskeligere å betjene et lån fordi man har mindre å rutte med.

Arbeidsledighet

En økning i arbeidsledigheten på ett prosentpoeng fører til at boligprisindeksen faller med rundt syv indeks-poeng ifølge modellen. Dette virker som en svært stor effekt, men det er viktig å huske på at en økning i ledigheten på ett prosentpoeng er svært mye, og forekommer sjelden over korte tidsperioder. Dersom dette skulle skje må det regnes som et sjokk på økonomien, og de er da ikke overraskende at det ville påvirke etterspørselen etter boliger drastisk.

En økning i ledighet fører til reduserte boligpriser ifølge økonomisk teori. Det kan forklares ved at individer som nylig har blitt arbeidsledige ikke har inntektssikkerheten til å ta opp boliglån, og derfor ikke til å kjøpe bolig. Dermed faller etterspørselen, og boligprisene synker.

En annen forklaring kan være at ledigheten kan påvirke boligprisene ved at befolkningen generelt oppfatter en økning av arbeidsledighet som tegn på en usikker fremtid, og utsetter derfor å investere i langtidsprosjekter som bolig. Denne effekten blir antakeligvis delvis fanget opp av forventningsbarometeret til TNS Gallup, men er ikke inkludert i den estimerte modellen over da den ikke er signifikant. Ringvirkningene av arbeidsledigheten til innbyggernes forventninger kan dermed være en forklaring på hvorfor den estimerte koeffisienten er såpass stor.

Skatt

Av tabell 4 ser vi at en økning i skattenivået på ett prosentpoeng fører til et fall i prisindeksen på 1,8 indeks-poeng. Dersom skattenivået øker fører det til redusert kjøpekraft, og dermed synker boligprisene.

Skatten kan påvirke etterspørselen etter boliger på følgende måte: En skatteøkning kan redusere forventet nettoinntekt for utleie av bolig. Dette fører til redusert etterspørsel etter å kjøpe bolig for utleie. Økte skatter vil dessuten alltid medføre redusert disponibel inntekt, som påvirker boligprisene negativt.

Avkortet datasett: 2007 til 2016

Ved å estimere modell 11 finner vi følgende finjusterte modell basert på makrovariabler (se tabell 4 for regresjonsutskrift):

$$\widehat{Prisindeks}_t = 114 - 160,2Lønn_t - 270,1KPI - JAE_t - 1409,8Arbeidsledighet_t - 231,1Skatt_t - 292,1Rente_t \quad (13)$$

Sammenlignet med modell 12 basert på det fullstendige datasettet er en ekstra variabel signifikant, *Rente*. Vi vil kun diskutere denne variabelen i dette avsnittet, da fortegnet er det samme for de øvrige variablene i modell 12 og 13. De estimerte koeffisientene er noenlunde like, med unntak av for *Ledighet*, som har doblet seg. En endring i prisindeksen på 14 indekspoeng når arbeidsledigheten øker med ett prosentpoeng virker drastisk, og noe usannsynlig at kan stemme. Som forklart over er en endring i ledighetsnivået på ett prosentpoeng svært mye, men et påfølgende fall på 14 indekspoeng for prisindeksen er likevel svært høyt.

Rente

Modell 13 forteller at en økning i den gjennomsnittlige utlånsrenten på boliglån med ett prosentpoeng leder til et fall i prisindeksen på nesten tre indekspoeng. Dette samsvarer med våre antakelser om rentens påvirkning på boligprisen. Høyere boliglånsrenter gjør det dyrere å ta opp boliglån, og etterspørselen synker. Dersom boliglånsrenten derimot reduseres betyr dette høyst sannsynlig at renten for sparekontoer også er lav, og at incentivet til å investere i bolig, ofte sekundærbolig, i stedet øker. Dette vil øke etterspørselen, og dermed gi økte priser.

Sammenligning av Fundamental og Fundamental2016

Det første å bemerke seg er at forklaringsgraden har steget drastisk ved å kutte ut 2017 fra estimeringsperioden, fra 71 prosent i Fundamentalmodellen, til 85 prosent i Fundamental2016-modellen. Dette kan tolkes som at Fundamental2016-modellen kan forklare variasjonen i *Prisindeks* bedre enn Fundamental – den estimerte modellen passer bedre til de ekte datapunktene, noe vi også kan se grafisk av figur 8. Her følger den estimerte verdien for prisindeksen den ekte prisindeksen relativt tett. Det er nå ingen perioder der det er store avvik mellom prisindeksen og regresjonsmodellen, slik vi ser det er for Fundamentalmodellen i figur 7.

Begge modeller har relativt lave standardavvik, som indikerer at vi opererer med en effektiv estimator.

Forklaringsvariabel	Makro Avhengig variabel: Prisindeks (trend- og sesongjustert)	Fundamental2016
Lønn	-180,7*** (24,74)	-160.2*** (16.65)
KPI-JAE	-276,6*** (65,72)	-270.1*** (40.92)
Arbeidsledighet	-701,4*** (107,0)	-1409.8*** (168.8)
Skatt	-178,7** (59,05)	-231.1*** (37.34)
Rente		-292.1*** (76.08)
Konstantledd	66.85*** (18.06)	114.0*** (12.89)
<i>Observasjoner</i>	131	120
<i>Adj. R²</i>	0,7125	0,845

Newey West standardavvik i parentes.

* $p < 0,05$ ** $p < 0,01$, *** $p < 0,001$

Tabell 4: Estimert Fundamentalmodell og Fundamentalmodell2016.

7.2.2 Fundamentalmodeller benyttet til å forutsi boligprisen

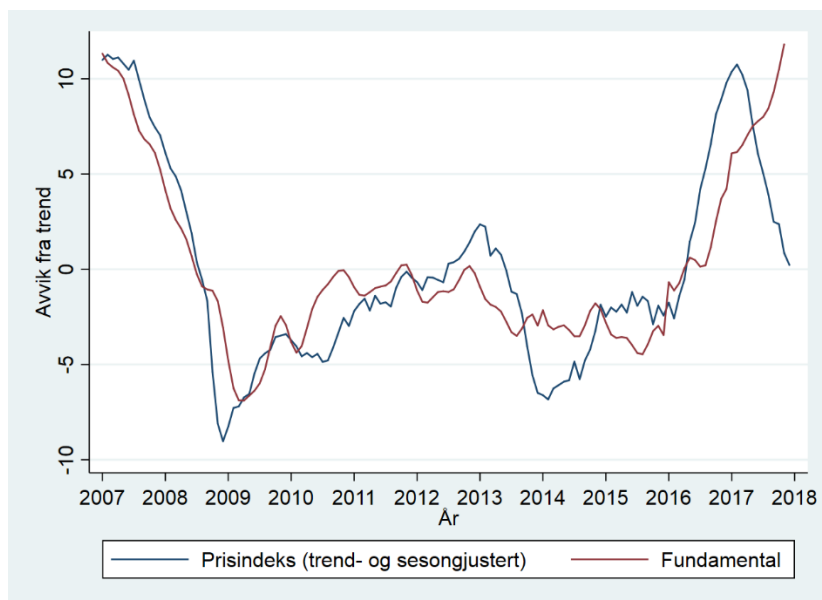
Fullstendig datasett: 2007 til 2017

Fundamentalmodellens estimerte boligpris og den ekte boligprisindeksen er grafisk fremstilt i figur 7, estimert på det fullstendige datasettet. Det ser ut til at den estimerte modellen følger prisindeksen relativt tett.

Prisindeksen og den estimerte prisindeksen faller sammenhengende fra 2007 til 2009, og fundamentalmodellen snur med et par måneders etterslep på boligprisindeksen. Frem til prisfallet i 2013 passer modellen godt, med en liten nedgang like før 2010 i forkant av et kort fall i prisindeksen. Boligprisens andre betydelig fall, i januar 2013, kommer to måneder etter at fallet begynte i fundamentalmodellen.

Modellen holder seg relativt stabil fra slutten av 2013 til begynnelsen av 2016, til tross for at prisindeksen begynner å vokse allerede i 2014. Fra 2016 til slutten av 2017 vokser

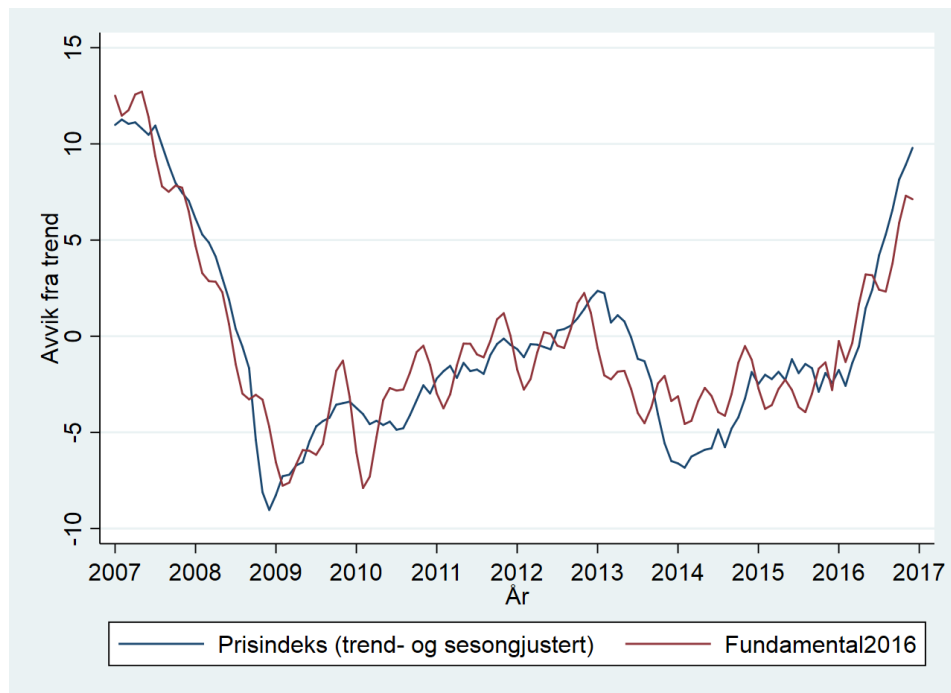
fundamentalmodellen kraftig, til tross for at prisindeksen stuper fra april 2017. Denne perioden tilsvarer også det største avviket mellom fundamentalmodellen og boligprisindeksen, og viser at modellen ikke presterte å forutsi det kraftige prisfallet fra april 2017 – den fortsetter faktisk å stige. Avviket kan antakelig forklare mye av forskjellen i forklaringsgraden når vi benytter det fullstendige datasettet kontra det avkortede.



Figur 7: Boligprisindeksen og estimert fundamentalmodell for perioden 2006 til 2017. Årstallet på x-aksen er satt til januar det respektive året.

Avkortet datasett: 2007 til 2016

Figur 8 viser Fundamental2016 estimerte boligprisindeks i forhold til den virkelige boligprisindeksen. Modellen følger prisindeksen bedre enn vi ser av figur 7, som ikke er overraskende siden 2017, som fundamentalmodellen for det komplette datasettet bommet totalt på, er ekskludert fra datasettet. Vi ser av figur 7 og 8 at de estimerte modellene stort sett er like med tanke på vekst- og nedgangsperioder, men vi ønsker å bemerke at de lokale bunn- og toppunktene er dypere for Fundamental2016 enn for Fundamental. Flere steder, som i 2010, 2012 og 2015, er toppene og bunnpunktene dypere enn prisindeksen.



Figur 8: Boligprisindeksen og estimert fundamentalmodell for perioden 2006 til 2016. Årstallet på x-aksen er satt til januar det respektive året.

Når det gjelder modellens evne til å forutsi boligprisfall bør fallet i 2013 nevnes. Her kan Fundamental2016 forutsi boligprisens fall noen måneder før prisindeksen når toppen, og når bunnen før prisindeksen.

7.3 Hybridmodeller

I dette underkapitlet vil vi presentere de finjusterte regresjonsmodellene basert på både makrovariabler og Google-variabler. Vi tar utgangspunkt i modell 5 fra kapittel seks, som på lik linje med forventningsmodellene og fundamentalmodellene finjusteres basert på å utelukke variabler som ikke er signifikante. Regresjonsutskriften i appendiks C3 viser den estimerte modellen med samtlige Google- og makrovariabler, for avkortet og fullstendig datasett.

Vi har valgt å kalle modellene som estimeres med Google- og makrovariabler for hybridmodeller, dette fordi de anses som en hybrid mellom modellen basert utelukkende på Google-variabler (Forventningsmodellene) og modellen som kun har makrovariabler (Fundamentalmodellen).

Vi kommer frem til følgende regresjonsmodell, hvor samtlige uavhengige variabler har en signifikant effekt på boligprisindeksen, for det komplette datasettet, perioden 2007 til 2017:

$$\begin{aligned}
Prisindeks_t = & \beta_0 + \beta_1 \ln(LØNN_t) + \beta_2 ARBEIDSLEDIGHET_t + \beta_3 SKATT_t \\
& + \beta_4 TNSGALLUP_t + \beta_5 \ln(boliglån_t) + \beta_6 \ln(boliglånsrente_t) \\
& + \beta_7 \ln(BOLIGMASSE_t) + u_t, \quad t = 1, 2, \dots, n
\end{aligned} \tag{14}$$

Variabelnavn med blokkbokstaver markerer makrovariabler, små bokstaver er Google-variabler.

På data fra 2007 til 2016 blir flere variabler signifikante, og vi får følgende regresjonsmodell:

$$\begin{aligned}
Prisindeks_t = & \beta_0 + \beta_1 \ln(LØNN_t) + \beta_2 ARBEIDSLEDIGHET_t + \beta_3 SKATT_t \\
& + \beta_4 TNSGALLUP_t + \beta_5 \ln(boliglån_t) + \beta_6 \ln(boliglånsrente_t) \\
& + \beta_7 \ln(boligkjøp_t) + \beta_8 RENTE_t + u_t, \quad t = 1, 2, \dots, n
\end{aligned} \tag{15}$$

Variabelnavn med blokkbokstaver markerer makrovariabler, små bokstaver er Google-variabler.

7.3.1 Regresjonsresultater med hybridmodellene

Her vil variablene inkludert i hybridmodellene presenteres med hensyn til fortegn og størrelse på de estimerte koeffisientene. Vi har i tillegg inkludert en kort analyse av om fortegnet samsvarer med hva som forventes.

Fullstendig datasett: 2007 til 2017

Ved å estimere modell 14 finner vi følgende finjusterte modell basert på makrovariabler (se regresjonsutskrift i tabell 5):

$$\begin{aligned}
\widehat{Prisindeks}_t = & 62,22 - 120,8LØNN_t - 523,0ARBEIDSLEDIGHET_t - 183,1SKATT_t \\
& + 9,095TNSGALLUP_t - 20,24Boliglån_t + 8,614Boliglånsrente_t \\
& - 580,8BOLIGMASSE_t
\end{aligned} \tag{16}$$

Hele fem av makrovariablene beskrevet i kapittel fem er signifikante i modellen. I tillegg til variablene som også var signifikante i fundamentalmodellen for hele perioden er nå *Boligmasse*- og *TNSGallup*-variablene signifikante. *KPI_JAE* er ikke lenger signifikant. Av Google-variablene er det kun to variabler som nå har en signifikant effekt på boligprisindeksen, begge knyttet til boliglån: *Boliglån* og *Boliglånsrente*. Overraskende nok er ikke lenger *Eiendomsmegler* signifikant, da denne fremheves som viktig i litteraturen.

Variablene som er diskutert tidligere i analysen vil ikke gjentas her, da fortegnet er likt for samtlige og koeffisientene avviker ikke drastisk.

TNS Gallup

Forventningsbarometeret til TNS Gallup, som i de tidligere nevnte modellene ikke har vært signifikant, er nå blitt statistisk signifikant. Optimisme vedrørende egen og landets økonomi som helhet indikerer at man er mer villig til å investere, og dermed øker etterspørselen etter bolig. Et positivt fortegn er dermed logisk. En økning på en enhet i indikatorverdien fører til en økning på 0,09 indekspoeng for boligprisindeksen, en beskjeden påvirkning.

Boligmasse

Boligmasse har også blitt signifikant i hybridmodellen. Modellen sier at en ett prosents økning i boligmassen fører til et fall på 5,8 indekspoeng for boligprisindeksen. Dette virker tilsynelatende som en stor påvirkning. Det må likevel bemerkes at en økning i antall boliger på ett prosent tilsvarer en økning på for eksempel 36.043 boliger i november 2017. En ett prosents-økning er derfor en relativt stor økning, og det er derfor naturlig at det har stor effekt på boligprisen.

Boliglånsrente

For forventningsmodellen estimert på hele perioden var ikke *Boliglånsrente* signifikant. Det er den derimot i hybridmodellen for samme periode. En økning på ett prosent i søk på *Boliglånsrente* leder til en økning på 0,086 indekspoeng i boligprisindeksen.

Avkortet datasett: 2007 til 2016

Ved å estimere modell 15 finner vi følgende finjusterte modell basert på makrovariabler og Google-variabler (se regresjonsutskrift i tabell 5):

$$\begin{aligned} \widehat{Prisindeks}_t = & 68,10 - 105,0LØNN_t - 764,5ARBEIDSLEDIGHET_t - 148,8SKATT_t \\ & + 11,78TNS\ GALLUP_t - 11,00Boliglån_t + 5,633Boliglånrente_t \\ & + 14,1Boligkjøp_t - 189,2RENTE_t \end{aligned} \quad (17)$$

Også her er fem av makrovariablene beskrevet i kapittel fem statistisk signifikante. KPI_JAE er ikke lenger signifikant sammenliknet med Fundamentalmodellen for samme periode, noe TNS Gallup nå er. Av Google-variablene er tre av variablene beskrevet i kapittel fem statistisk signifikante, langt færre enn for Forventningsmodellen estimert på samme tidsperiode. Også

her er det overraskende at *Eiendomsmegler* ikke lenger er signifikant, basert på litteraturgjennomgangen.

Tidligere diskuterte variabler vil ikke gjentas her da verken fortegn eller størrelse har endret seg bemerkelsesverdig. Vi har heller valgt å fokusere på den eneste variabelen som til nå ikke har vært signifikant.

Boligkjøp

Regresjonsresultatene i tabell 5 viser at en økning på ett prosent i søk på ordene representert av *Boligkjøp* fører til en økning på 0,14 indekspoeng for boligprisindeksen. Vi antok at økt søkeaktivitet etter boligkjøpsrelaterte søkeord indikerer økt etterspørsel, og dermed en økning i boligprisene.

Sammenlikning av Hybrid og Hybrid2016

Det er ikke overraskende at flere makrovariabler enn Google-variabler er signifikante i hybridmodellene ettersom både Fundamental og Fundamental2016 har høyere forklaringsgrad enn for tilsvarende forventningsmodeller.

Begge hybridmodellene har høy forklaringsgrad. Modellen på det avkortede datasettet er bedre selv etter at reduksjonen i frihetsgrader er tatt hensyn til ved å beregne den justerte forklaringsgraden.

Forklaringsvariabel	Hybrid Avhengig variabel: Prisindeks (trend- og sesongjustert)	Hybrid2016
Lønn	-120,8*** (14,46)	-105,0*** (15,16)
Arbeidsledighet	-523,0*** (116,5)	-764,5*** (179,3)
Skatt	-183,1*** (53,65)	-148,8** (51,07)
TNS Gallup	9,095* (3,553)	11,78*** (2,080)
<i>Boliglån</i>	-20,24*** (3,692)	-11,00*** (2,600)
<i>Boliglånsrente</i>	8,614*** (1,307)	5,633*** (1,225)
Boligmasse	-580,8*** (109,5)	
<i>Boligkjøp</i>		14,10*** (3,058)
Rente		-189,2** (60,52)
Konstantledd	62,22*** (16,60)	68,10*** (17,84)
<i>Observasjoner</i>	131	120
<i>Adj. R²</i>	0,8288	0,8937

Newey West standardavvik i parentes.

* $p < 0,05$ ** $p < 0,01$, *** $p < 0,001$

Variabler i kursiv er Boligsøksvariabler

Variabler i fet skrift er Makroøkonomiske variabler

Tabell 5: Estimert hybridmodell for 2007 til 2016 (Hybrid) og for 2007 til 2016 (Hybrid2016)

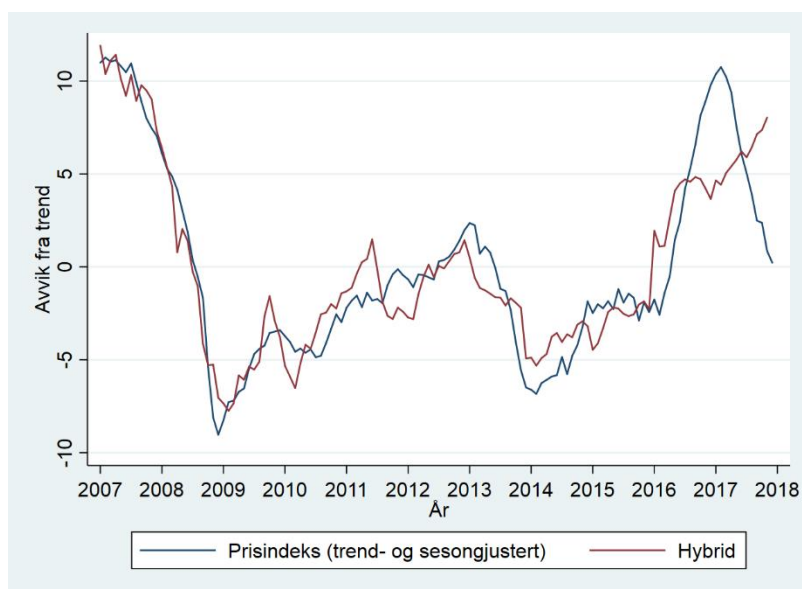
7.3.2 Hybridmodeller benyttet til å forutsi boligprisen

Fullstendig datasett: 2007 til 2017

Figur 9 viser den virkelige boligprisindeksen og den estimerte boligprisen med hybridmodellen, for hele datasettet. Den likner i stor grad på fundamentalmodellen, som ikke

er overraskende da Fundamentalmodellen har høyere forklaringsgrad enn Forventningsmodellen for tilsvarende periode. Hybridmodellen avviker mest fra Fundamentalmodellen fra midten av 2016, heller ikke overraskende da det er i denne perioden Fundamentalmodellen avviker i størst grad fra prisindeksen og ikke presterer å forutsi det kraftige boligprisfallet i april 2017.

Av figur 9 ser vi at modellen har en tendens til å forutsi dype prisfall og høye pristopper som ikke kommer. Dette gjelder spesielt mot slutten av 2009, starten av 2010 og i 2011. Boligprisen følger riktignok de samme svingningene med noen måneders etterslep, men de lokale toppunktene og bunnpunktene er ikke like markante. Likevel ser det ut til at hybridmodellen gjør en god jobb i å forutsi nedganger i boligprisen, sett bort fra prisfallet i 2017.



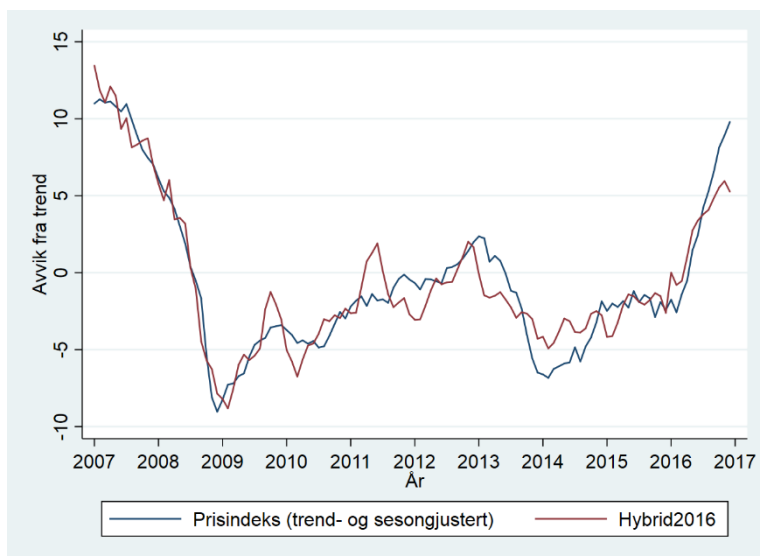
Figur 9: Estimert hybridmodell og boligprisindeksen for 2006 til 2017. Årstallet på x-aksen er satt til januar det respektive året.

Hybridmodellen har i likhet med fundamentalmodellen problemer med å fange opp det kommende fallet i boligprisen i 2017, men viser heller ikke like sterk vekst i samme periode, slik fundamentalmodellen gjorde.

Avkortet datasett: 2007 til 2016

Den estimerte boligprisen for Hybrid2016-modellen i figur 10 er så å si identisk med Hybridmodellen i figur 9. Svingningene er like, fallet i forkant av boligprisens fall er like markante og det er ingen klare tegn til at modellen vil forutsi et fall i 2017.

Det skal nevnes at det ser ut til at den snur i desember 2016, men ved å undersøke hybridmodellen for hele perioden i figur 9 ser vi at også denne har en liten nedgang i desember 2016, før den fortsetter å stige. Vi antar det samme skjer med Hybrid2016-modellen, og at vi derfor ikke kan slå fast at den vil starte å falle markant og dermed forutsi boligprisfallet som kommer i april 2017.



Figur 10: Estimert hybridmodell og boligprisindeksen for 2006 til 2016. Årstallet på x-aksen er satt til januar det respektive året.

7.4 Oppsummering

Av resultatene presentert i kapitlet ser vi at samtlige modeller mer eller mindre godt forklarer boligprisen. Forklaringsgraden stiger for samtlige modeller når 2017 utelates fra den estimerte perioden. Særlig fundamentalmodellen forbedres markant i form av økt forklaringsgrad for det begrensede datasettet.

Hybridmodellene har høyest forklaringsgrad, ikke overraskende siden flere variabler er inkludert enn for fundamental- og forventningsmodellene. Det er mer interessant å bemerke at hybridmodellen hadde såpass mye høyere forklaringsgrad ved å estimere modellene på det fullstendige datasettet, i forhold til Fundamentalmodellen for samme periode. Et interessant funn er at det ser ut til at Forventningsmodellen estimert på data frem til 2016 fortsatt ser ut til forutse fallet før boligprisfallet i 2017.

Flere variabler i forventningsmodellen kan virke kontraintuitive i henhold til økonomisk teori hva gjelder fortegn. Noe av dette kan antagelig forklares ved at en aktør i boligmarkedet ofte både er selger og kjøper i samme tidsperiode.

I neste kapitel vil resultatene fra den empiriske analysen diskuteres i sammenheng med teori om boligbobler, med et spesielt fokus på forventninger.

8. Analyse av bobletendenser

Dette kapitlet vil kombinere Grytten og Hunnes' syvtrinns dynamiske krismodell med tolkninger fra kapittel 7.1.2 og 7.2.2. Hensikten er å studere om Fundamental- og Forventningsmodellene kan sies å identifisere bobletendenser når før et toppunkt i boligmarkedet. I figur 1 ser vi at boligprisen når et toppunkt i 2007, 2013 og 2017. I dette kapitlet ønsker vi å undersøke perioden før disse punktene.

Vi har valgt å ta utgangspunkt modellene basert på det fullstendige datasettet i analysen som følger, ettersom vi ønsker å inkludere prisetallet i 2017 i analysen. Forventning- og Fundamentalmodellene er fremstilt i samme graf i figur 11. Siden vårt datasett ikke inkluderer prisstigningen før fallet i 2007 vil dette toppunktet ikke diskuteres.

8.1 Analyse av toppunkter basert på empiriske resultater

Som nevnt i kapittel to kjennetegnes en euforisk boble ved at «den raske prisveksten ikke kan forklares ut fra fundamentale forhold» (Grytten & Hunnes, 2016). Vi vil basert på dette anse betydelige positive avvik mellom Forventningsmodellen og Fundamentalmodellen som en indikasjon på bobletendenser i boligmarkedet. Begrunnelsen er at vi tolker dette avviket som at potensielle boligkjøpere har større forventninger til markedet enn hva som forklares med fundamentale forhold.

Toppunktet i 2013

I starten av 2012 passerer Forventningsmodellen Fundamentalmodellen og blir liggende over frem til høsten 2012. Avviket er ikke betydelig, men det var første gangen etter 2009 at det skjedde. Vi anser likevel ikke denne forskjellen som grunnlag for å hevde at det var euforiske tendenser i boligmarkedet, da forskjellen er liten, perioden er kort og veksten i forventningsmodellen er beskjedne. Prisstigningen i forkant av 2013 kan dermed ikke sies å være utelukkende forårsaket av folks forventninger om at prisen vil fortsette å stige fremover, som Stiglitz definerer som en aktivaboble (Grytten & Hunnes, 2016).

Det er interessant å studere Fundamentalmodellen i forhold til den faktiske prisindeksen når vi skal studere bobletendenser. I perioden 2010 til 2012 ligger Fundamentalmodellen over prisindeksen, noe som kan bidra til å forklare prisveksten fram til 2013 i form av fundamentale

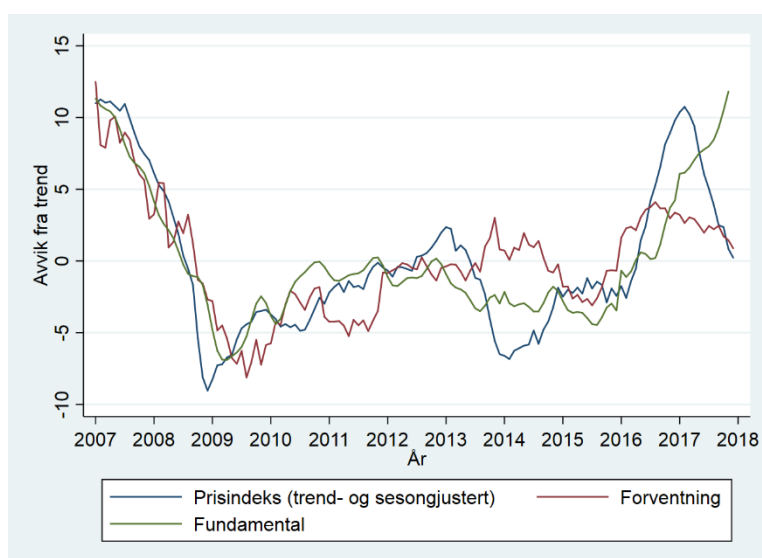
forhold. Det understreker dessuten argumentet om at prisstigningen før 2013 ikke bar preg av euforiske tendenser.

Basert på analysen kan det tyde på at prisoppgangen bar preg av ikke-euforiske tendenser heller enn euforiske tendenser, og at priset etter pristoppen i 2013 heller var en korreksjon av et overopphetet marked enn en euforisk boble som sprakk.

Toppunktet i 2017

I mars 2017 nådde boligprisindeksen et toppunkt. Prisutviklingen i forkant var relativt kraftig, og ender på hele ti indekspoeng over trend. Etter at boligprisene hadde hentet seg inn igjen etter priset i 2013 var veksten omtrent på trend til 2016. I starten av 2016 starter en sterk boligprisvekst som varer helt til toppunktet er nås i mars året etter.

Forventningsmodellen begynner å stige kraftig fra midten av 2015, før den snur i midten av 2016. Deretter følger en nedgangsperiode som varer frem til slutten av 2017. For Fundamentalmodellen ser vi at denne også vokser fra 2015 og følger et noenlunde likt mønster som Forventningsmodellen. Når Forventningsmodellen snur i midten av 2016 fortsetter Fundamentalmodellen å vokse kraftig.



Figur 11: Estimert fundamentalmodell og forventningsmodell, samt boligprisindeksen, for perioden 2007 til 2017. Årstallene på x-aksen indikerer januar det respektive året.

Forventningsmodellen ligger over Fundamentalmodellen i perioden fra 2015 til høsten 2016, som kan indikere at den kraftige prisoppgaven bar preg av euforiske tendenser. Vi ønsker derfor å studere perioden nærmere ved å benytte teori om euforiske bobler.

8.2 Analyse av toppunktet i 2017 basert på teori

For å studere oppbyggingen til toppunktet i 2017 vil det Grytten og Hunnes' syv-trinns dynamiske krisemodell benyttes. Det er viktig å poengtere at Grytten og Hunnes (2016) forklarer at ikke alle trinnene i modellen trenger å være til stede i forkant av en boble, og at trinnene kan glid inn i hverandre.

Det kan være veksten begynte grunnet en kombinasjon av endogene og eksogene forhold. I Oljeprisen nådde sin laveste pris de siste ti årene i januar 2016 og har opplev en stødig vekst deretter. Det kan være at dette endogene sjokket, kombinert med andre fundamental forholdene førte til prisveksten. Som nevnt begynte fundamentalmodellen å vokse allerede i 2015. Det kan være de gode fundamentale forholdene gjorde at prisindeksen tok av. Dette sjokket kan sies å være en forstyrrelse, endog en endogen en, som startet vekstperioden som etter Grytten og Hunnes modell er det første steget i modellen.

Det er vanskelig å finne konkrete tegn til at boligmarkedet opplevde overoppheting etter januar 2016. Som vi tydelig ser av figur 11 vokser boligprisene i takt med Fundamentalmodellen, som kan indikere at prisoppgangen kan forklares med fundamentale faktorer. Det kan derimot diskuteres hvorvidt medieoppslag om hvor mye man kunne tjene på boligmarkedet var en forsterkende effekt for boligprisveksten, og bidro til å skape eufori i markedet. Bare i Dagens Næringsliv ble det i løpet av 2016 skrevet flere artikler enten om personer som hadde tjent store summer på investering i bolig eller hvordan en skulle investere i boligmarkedet (Hartwig, 2016) (Mikalsen, 2016) (Hartwig, 2016). Vi ser også fra Forventningsmodellen at det i denne perioden er sterk vekst, og at modellen ligger over Fundamentalmodellen. Disse faktorene kombinert kan tyde på at det var euforiske tendenser i markedet i 2016.

Av figur 11 er det ingen tegn til at markedet er inne i en bobleøkonomi. Prisveksten i 2016 og tidlig 2017 samsvarer med veksten i Fundamentmodellen. Det kan derfor ikke sies at dette steget i modellen var til stedet. Prisindeksen kan ha vokst noe mer enn de fundamentale modellene skulle tilsi, men det er vanskelig å konkludere med basert på figur 11. Likevel ser vi, som bemerket i forrige underkapittel, at Forventningsmodellen ligger over

Fundamentalmodellen helt til prisene begynner å falle i april 2017. Analysen gjør det vanskelig å konkludere med om boligmarkedet befant seg i en bobleøkonomi eller ikke før pristoppunktet i 2017.

Det observeres ingen nervøsitit boligmarkedet i den studerte perioden. Om dette er et tegn på at markedet bærer preg av en ikke-euforisk boble heller enn en euforisk er vanskelig å si. Som nevnt skriver Grytten og Hunnes at ikke alle stegene må være til stedet for at en euforisk boble skal identifiseres.

Vendepunktet i boligprisene fant sted på våren 2017, etter endringen i boliglånsforskriftene trådte i kraft. Det er tydelig prisene snur, men dette ser ikke ut til å skyldes endringer i de fundamentale verdiene da Fundamentalmodellen fortsetter å vokse. Det interessant her er at Forventningsmodellen hadde snudd allerede før forskriftene ble publisert. Hvis dette er et tegn på at optimismen har snudd til pessimisme kan dette bidra å forklare graden av fallet i boligprisene, hvis ikke selve vendepunktet. Dette kan tyde på at det var euforiske bobletendenser i boligmarkedet før april 2017.

Vi ser ingen tegn til de to siste trinnene i Grytten og Hunnes' syv-trinnsmodell i boligmarkedet. Selv om det ble rapportert om svært få på visninger av boliger etter april 2017, som et tegn på at tilbudet var større enn etterspørselen, kan det ikke hevdes å ha vært noen panisk stemning eller liknende krisefenomener. Media skrev mye om boligprisnedgangen og om få på visninger i perioden. Når ikke engang dette fikk folk til å selge i panikk anser vi dette som et tegn på at vendepunktet ikke var et vendepunkt for en euforisk boble. Det kan eventuelt ha vært en påbegynnende korreksjon for en ikke-euforisk boble.

At boligprisindeksen begynte å stige allerede i løpet av 2018 kan tyde på at prisfallet i 2017 heller var en korreksjon av et overopphetet marked. Dette til tross virker det som oppløpet til fallet bare preg av noen euforiske bobletendenser. Det virker også som forholdet mellom Forventnings- og Fundamentalmodellen kan gi indikasjoner om markedet er mer euforisk enn de fundamentale forholdene skulle tilsi.

9. Diskusjon

I dette kapitlet vil vi starte med å drøfte de tre forskningsspørsmålene presentert i kapittel to. Deretter vil vi legge frem begrensninger ved masterutretningen, før vi kommer med forslag til videre arbeid.

9.1 Drøfting av forskningsspørsmålene

Denne seksjonen vil drøfte de tre forskningsspørsmålene postulert i kapittel to. Vi vil basere drøftingen på de to foregående analysekapitlene.

9.1.1 Forskningsspørsmål 1

Kan boligprismodeller kun basert på makroøkonomiske variabler forbedres ved å inkludere Google-søkr?

Spørsmål 1a

Kan Google-søk forklare boligprisutviklingen bedre enn makrovariabler alene?

Basert på vår empiriske analyse har vi ikke grunnlag til å hevde at Google-variabler kan forklare boligprisutviklingen bedre enn makrovariabler. Dette står i kontrast til tidligere litteratur på området. Dietzel, Braun og Schaefer (2014) fant at modellen kun bestående av Google-variabler gjorde det bedre enn ved kun å benytte sentrale makroøkonomiske variabler.

Både for det fullstendige og det avkortede datasettet presterer de estimerte fundamentalmodellene bedre enn forventningsmodellene i vår empiriske analyse. Til tross for dette kan det argumenteres for at Google-søk kan bidra til å forklare utviklingen i boligprisindeksen. Litteratur på området har fremhevet at fordelene med Google-data er at statistikk gjøres tilgjengelig med et etterslep på maksimum 48 timer. Selv om forventningsmodellen forklarer boligprisen dårligere enn fundamentalmodellen har den likevel fordel av å kunne gi et estimat av boligprisen i nåtid. Forventningsmodellen bør i så måte ikke forkastes som verktøy for å estimere prisene i det norske boligmarkedet. Den kan heller sees som et noe unøyaktig estimat før mer nøyaktige tall blir tilgjengelig.

Valget mellom Fundamentalmodellen og Forventningsmodellen for å estimere boligprisen i nåtid vil avhenge av avveiningen mellom rask tilgjengelighet og nøyaktighet. Dersom førstnevnte er viktigst vil forventningsmodellen kunne passe formålet best, og motsatt dersom nøyaktighet er det viktigste.

Spørsmål 1b

Kan Google-søk i kombinasjon med makrovariabler forklare boligprisutviklingen bedre enn makrovariabler alene?

Det er tydelig fra tabell 4 og 5 at hybridmodellen presterer bedre enn fundamentmodellen både på det fullstendige og det begrensede datasettet. Forklaringsgraden er høyere i begge tilfeller, også etter reduksjonen i frihetsgrader er tatt hensyn til, og standardavviket er ikke betydelig annerledes.

På det begrensede datasettet presterer modellene begge meget god, men Hybridmodellen noe bedre. Mangelen på tydelige eksogene sjokk kan forklare hvorfor modellene er så like. Prisveksten i perioden reflekterer endringen i de fundamentale verdiene. Dette gjør fundamentalmodellen godt egnet. Hybridmodellen gjør det noe bedre, men ikke oppsiktsvekkende.

For det fullstendige datasettet er forskjellene større. 2017 var et utfordrende år å estimere for samtlige modeller. Den justerte forklaringsgraden svekkes i forhold til modellene basert på det avgrensede datasettet for samtlige modeller. Fallet i fundamentalmodellen er på hele 13 prosentpoeng, der hvor hybridmodellen kun faller med fem. Dette er ikke overraskende ettersom fundamentalmodellen går i motsatt retning av boligprisen etter mars 2017. Det er heller ikke overraskende at hybridmodellen, som inkluderer noen Google-variabler knyttet til forventning, ikke faller like mye. For 2016-modellene er forskjellen mellom de to modellene i overkant av seks prosentpoeng i favør av hybridmodellen. Når 2017 inkluderes er forskjellen på ti prosentpoeng. Ved å inkludere flere variabler er det naturlig at forklaringsgraden øker, men vi har sammenlignet de justerte forklaringsgradene, som kontrollerer for tap av frihetsgrader og understøtter at forskjellen mellom hybridmodellen og fundamentalmodellen er markant.

De nevnte forskjellene mellom hybridmodellen og fundamentalmodellen gir innsikt i når Google-søk kan bidra til å forbedre modeller basert på makroøkonomiske variabler. I perioder hvor boligprisen endrer seg mye grunnet faktorer som ikke kan knyttes til endringer i de fundamentale verdiene vil Google-data kunne bidra med verdifull informasjon.

9.1.2 Forskningsspørsmål 2

Kan Google-søk forutsi fall i boligprisen bedre enn makroøkonomiske variabler?

Fullstendig datasett: 2007 til 2017

I kapittel 7.1.2 og 7.2.2 er det beskrevet tilfeller hvor henholdsvis forventings- og fundamentalmodellen forutsier fall i prisindeksen. Forventningsmodellen forutsier ikke fallet i 2013, i motsetning til fundamentalmodellen. På den andre siden forutsier ikke fundamentmodellen prisleilet i 2017 som forventningsmodellen gjør.

Det kraftige prisleilet i 2017 er spesielt interessant å studere. Det ser ut til at forventningsmodellen presterer å forutsi denne trekvart år i forveien. Fundamentalmodellen stiger derimot både i forkant og etterkant av toppunktet i mars 2017, og feiler dermed totalt i å forutsi 2017-prisleilet. Det stilles dermed spørsmål ved hvor god fundamentalmodellen er til å forutsi fall i boligprisen, spesielt dersom denne skyldes eksogene sjokk.

Det er vanskelig å si definitivt at den ene modellen presterer bedre i alle tilfeller enn den andre basert på det fullstendige datasettet. Her kan igjen fordelene med den korte publiseringstiden til Google-variabler kontra makroøkonomiske variabler trekkes frem. Makrovariabler publiseres ofte med et betydelig etterslep, og dersom ikke fundamentalmodellen forutsier et prisfall mange måneder i forkant av det faktiske fallet vil ikke fallet kunne forutsies før det er for sent. For prisleilet i 2013 forutsier fundamentalmodellen boligprisleilet to måneder i forveien, og dersom publiseringsetterslepet av makrovariablene er lenger enn dette ville ikke modellen ha klart å forutsi fallet i 2012.

Avkortet datasett: 2007 til 2016

Som nevnt er fundamentalmodell2016 bedre til å forklare boligprisen enn forventningsmodell2016 i form av høyere justert forklaringsgrad. Det er derfor interessant å oppdage at det motsatte ser ut til å være tilfellet for å forutsi fall. Ved å studere grafen for

forventning2016 i kapittel syv ser vi at den forutsier toppunktene i 2007, 2013 og 2017. Den forutsier i tillegg fallene før fundamental2016 gjør det.

En annen interessant observasjon ved fundamentmodell2016 er at den forutsier et sterkt fall i prisindeksen i 2010, som ikke finner sted i boligprisindeksen. Til sammenligning er fallet dypere og brattere enn fallet før pristoppen i 2013. Det representerer en betydelig svakhet med fundamentmodell2016. Dette «falske» krakket svekker kredibiliteten til modellens evne til å forutsi fall i boligprisen. Det skal nevnes at det er et mindre fall i boligprisindeksen i 2010, men ikke i nærheten av like drastisk som fundamentmodellen forutsier.

Det falske «krakket» i 2010 er ikke unikt for fundamentmodellen2016. Det er det klart mest markante, men liknende tilfeller forekommer ellers for modellen. I 2012 og 2015 viser modellen liknende sterke vekstperioder etterfulgt av tilsvarende sterke fallperioder. Disse fallene varer vanligvis ikke mer enn tre måneder, men i kombinasjon med etterslepet på publiseringen kan dette føre til at det kreves seks måneder for å forutsi et faktisk fall i markedet. Siden modellen ikke forutsier noen av fallene i modellen med mer enn seks måneder kan den sies å ikke være svært nyttig til å forutsi fall i prisindeksen.

Det store avviket mellom modellene i 2016 er også interessant. Det viser tydelig at forventningsmodellen faller fra midten av 2016. Dette står i sterk kontrast til fundamentmodell2016, som stiger kraftig i 2016. Modellenes evne til å forutsi det prisfallet ser ut til å gjelde også når 2017 er ekskludert fra datasettet modellene estimeres på.

Hybridmodellene har ikke blitt drøftet i dybden for forskningsspørsmålet fordi de følger en liknende utvikling som fundamentalmodellen, både for det begrensede og det fullstendige datasettet. Hybridmodellen er generelt mindre påvirket av svingninger enn fundamentalmodellen, men forutsier to sterke fall i boligprisene uten at dette inntreffer. Vi ser ikke like sterk vekst i Hybridmodellen i 2016 som fundamentalmodellen, men den forutsier heller ikke det store prisfallet som inntreffer i april 2017.

Basert på drøftingen av avkortet og fullstendig datasett kan vi konkluderer vi med at Google-søk generelt gjør en bedre jobb i å forutsi fremtidige fall boligprisene, og bedre enn modellen bygget på fundamentale variabler. Særlig forventning2016 presterer bedre enn tilsvarende fundamentalmodell.

9.1.3 Forskningsspørsmål 3

Kan vi identifisere bobletendenser i det norske boligmarkedet ved å sammenligne Google- og makroøkonomiske variabler?

Basert på analysen i forrige kapittel kan det diskuteres hvorvidt sammenligning av Forventnings- og Fundamentalmodellen kan identifisere bobletendenser. Den største utfordringen i dette tilfellet er mangel på en bekreftet boble i løpet av perioden vi studerer.

Det ser ut til at når forventningsmodellen overstiger fundamentalmodellen så kan dette samsvare med mulige tegn på tidlige boblefaser. Prisoppgangen før 2013 bærer preg av ikke-euforiske tendenser og prisoppgangen før april 2017 kan bære preg av euforiske tendenser.

Vi baserer konklusjonen på at forventningsmodellen med Google-variabler indikerer samfunnets forventning til fremtidig atferd i boligmarkedet og fundamentalmodellen med makrovariabler indikerer fundamentale faktorer.

Ved å velge ut Google-søkeord som spesifikt retter seg mot boligmarkedet, som vi har gjort i masterutredningen, vil man få et mål på samfunnets forventning spesifikt for boligprisene og ikke økonomien som helhet.

9.2 Begrensninger

Bruken av Google-data har flere tekniske og teoretiske utfordringer som begrenser denne oppgaven. Datagrunnlaget i perioden fra 2004 til 2006 er meget ustabil. Hvis data fra denne perioden hadde vært mer stabil ville vi hatt muligheten til å undersøke sammenhengen mellom Google-søk og boligprisene før og under finanskrisen, som ville vært interessant med tanke på forskningsspørsmål 2 og 3

En mulig begrensning er utvalget av Google-variabler. Vi har basert utvalget på fire egendefinerte kriterier, samt stilt krav til reliabilitet og validitet, men anerkjenner at Google-søkeord som kan forklare boligprisutviklingen bedre eller like godt kan ha blitt utelatt.

En tredje potensiell begrensning er at vi i stor grad har justert dataseriene vi benytter i oppgaven. Spesielt sesongjusteringen og konverteringen fra kvartal til måned kan introdusere unøyaktigheter i datasettet vårt, og potensielt påvirke vår empiriske analyse og konklusjoner.

Vi har foretatt en antakelse om at Google-søk reflekterer søkerens forventninger til boligprisens utvikling. Andre motivasjonsfaktorer enn forventning kan ligge bak søkene, som nysgjerrighet eller ren tilfeldighet. Det er vanskelig å bekrefte at dette stemmer da vi måtte hatt innsikt i søkernes tankegang før et søk ble foretatt, og kan representere en begrensning for vår oppgave. Det gjelder spesielt for drøftingen av forskningsspørsmål 3.

9.3 Forslag til videre arbeid

Masterutredningen åpner for videre arbeid for å undersøke sammenhenger mellom Google-søk og boligmarkedet, eller til å «ta temperaturen» i andre marked i Norge. Det antas at data fra Google Trends vil bli mer stabil over tid og gi muligheten til å predikere bedre modeller. Et alternativ kan være å undersøke data fra Finn.no, som er spisset mer inn mot boligmarkedet enn Google-søk.

Google-søk sitt potensiale for å forutsi fall i boligprisen har fått støtte i denne oppgaven. En grundigere utredning bør muligens gjøres før det definitivt konkluderes med at Google-søk kan forutsi prisfall. Hvis Google-søk, som er tilgjengelig med et svært kort etterslep, kan bidra til å forutse krakk i boligmarkedet kan økonomien skånes.

Våre analyser av om Google-søk kan benyttes til å identifisere bobletendenser i det norske boligmarkedet bør også utforskes videre med sofistikerte modeller. Her kan det blant annet vurderes om Eidfjord og Oust sin metode kan anvendes på det norske boligmarkedet.

Det kan være interessant å studere problemstillingen vår ved å benytte et paneldatasett for norske regioner dersom datagrunnlaget for regionene forbedrer seg i fremtiden. Regionbasert oppdeling vil potensielt kunne avdekke regionale forskjeller som vår analyse ikke har fanget opp.

10. Konklusjon

For å avslutte masterutredningen vil vi presentere en kort oppsummering av utredningen, presentere våre viktigste funn og til slutt besvare utredningens problemstilling.

Vår masterutredning har studert sammenhengen mellom boligmarksrelaterte søk på Google og boligprisindeksen. Problemstillingen vi har jobbet etter har vært å undersøke om Google-data kan benyttes til å forklare utviklingen i norske boligpriser. Vi har benyttet statistikk over søk på Google og sentrale makroøkonomiske størrelser for å studere problemstillingen.

Det teoretiske fundamentet for utredningen er basert på Grytten og Hunnes sin dynamiske syv-trinns krisemodell. Modellen beskriver hvordan et eksogent sjokk kan føre til en forstyrrelse i markedet. Forstyrrelsen endrer investorers forventninger, som leder til overoppheting i markedet. Ettersom aktiva blir overpriset beveger markedet seg inn i en bobleøkonomi, hvor investorer begynner å innse at prisene er for høye i forhold til fundamentale verdier. Når flere og flere innser dette vil nervøsiteten spre seg, som til slutt leder til et vendepunkt hvor prisene snur. Vendepunktet etterfølges av en krise i markedet, hvor et stort antall selger aktiva. Det kan lede til spredning til andre markeder.

Videre presenteres tidligere litteratur der digitale søk er benyttet til å predikere boligmarkedets utvikling. De siste ti årene er det utført studier på å benytte offentlig tilgjengelig Google-data for å predikere utviklingen i forskjellige markeder, fra det nederlandske boligmarkedet til det chilenske bilmarkedet. Google-søk gir umiddelbar tilgang på data i samtid, en fordel sammenlignet med statistikk fra statlige institusjoner som fremheves i litteraturen. Google-data har blitt benyttet til å predikere trender i samtid og for fremtiden. De fleste artikler finner at Google-søk inkludert som variabler i modeller egner seg godt til prediksjonsformål, men metoden har til tider fått kritikk for at datagrunnlaget inkluderer feil søkeord eller er lett påvirkelig.

Basert på tidligere litteratur, økonomisk teori og våre egne vurderinger har vi valgt ut søkeord vi ønsker å undersøke om kan ha en tilknytning til boligmarkedet. Variablene ble, i tillegg til å undersøke deres økonomiske relevans, vurdert i henhold til reliabilitet og validitet ved ytterligere tre kriterier. Det første kriteriet var at det må være tilstrekkelig antall søk til at Google Trends gir stabil statistikk for analyseformål. Deretter tok vi i bruk Google Adwords for å bekrefte at ordene som velges ut anses som relevante for boligmarkedet av bedrifter. Ord

markert med «lav konkurranse» forkastes, med unntak av søkeord for selskapsnavn. Det siste kriteriet er at det søkeordene har stort nok absolutt søkevolum. Basert på kriteriene konstruerte vi åtte Google-variabler som blir benyttet i regresjonsmodellene videre i utredningen. Syv makroøkonomiske variabler ble i tillegg inkludert som sammenlikningsgrunnlag for Google-variablene.

Tre regresjonsmodeller ble utformet for å studere problemstillingen. Variablene vi benyttet bar tydelig preg av typiske utfordringer med tidsseriedata, som ble tatt hensyn til før estimeringen av regresjonsmodellene i den empiriske analysen. Først eliminerte vi trend- og sesongstøy. Deretter gjorde vi inferens robust mot heteroskedastisitet og autokorrelasjon. Variabler angitt på en absolutt skala ble tatt logaritmen av. Regresjonsmetoden benyttet i masterutredningen er OLS.

I den empiriske analysen konstruerte vi seks modeller for å studere problemstillingen. Tre av modellene ble estimert på et avkortet datasett, fra 2007 til 2016, de øvrige på det komplette datasettet fra 2007 til 2017. Oppdelingen i to datasett ble gjort for å studere modellenes prediksjonsevne før og etter innføringen av boliglånsforskriftene i 2017. For hvert av de to datasettene har vi estimert en modell bestående av Google-variabler (forventningsmodeller), en bestående av makrovariabler (fundamentalmodeller) og en hybrid mellom de to (hybridmodeller). De empiriske resultatene viste at hybridmodellen har best justert forklaringsgrad, men at forventningsmodellen forutser prisfall i markedet bedre enn både fundamental- og hybridmodellen. Dette gjaldt spesielt for datasettet som ekskluderer 2017.

I analysen av bobletendenser i det norske boligmarkedet studerte vi to toppunkt, 2013 og 2017, ved å sammenligne fundamentalmodellen med forventningsmodellen. Konklusjonen var at prisoppgangen før 2013 bar preg av ikke-euforiske tendenser, og oppgangen før 2017 bar preg av euforiske tendenser. Dette ble videre studert ved å benytte Grytten og Hunnes sin syvtrinnsmodell presentert i det teoretiske fundamentet.

I arbeidet med masterutredningen har vi kommet frem til tre funn vi finner spesielt interessante og ønsker å poengtere. For det første fant vi, i motsetningen til tidligere litteratur, at en modell med makrovariabler har høyere justert forklaringsgrad enn en modell basert på Google-søk. Ulempen med makrovariablene er at de publiseres med et betydelig etterslep. Til tross for at modellen med Google-variabler er mindre presis har den likevel et stort fortrinn i at data

publiseres i samtid. Når Google-variabler og makrovariabler kombineres i samme modell forklarer de boligprisutviklingen bedre enn en modell bestående av makrovariabler.

For det andre fant vi at Google-søk kan benyttes til å forutsi større fall i boligprisene. Dette gjelder spesielt for fallet etter innføringen av boligforskriftene i 2017. Fundamentalmodellen varslet om en rekke prisfall som ikke kom, noe forventningsmodellen ikke gjorde.

For det tredje fant vi at forskjellen mellom fundamentalmodellen og forventningsmodellen indikerte de to toppunktene utover trend i den studerte perioden hadde bobletendenser. Oppgangen før 2013 bar preg av ikke-euforiske tendenser, og vendepunktet i 2017 viste euforiske tendenser.

Problemstillingen presentert innledningsvis lød som følger:

«Kan Google-søk benyttes til å forklare prisutviklingen i det norske boligmarkedet?»

Basert på de tre funnene presentert over konkluderer vi med at Google-søk kan benyttes til å forklare prisutviklingen i det norske boligmarkedet. En modell kun bestående av Google-variabler er mindre god, men kombinasjonen av makrovariabler og Google-variabler gjør det svært godt. Modellen med Google-variabler gjør det dessuten bedre enn modellen med makrovariabler i å forutse større fall i boligmarkedet. Vi konkluderer med at Google-søk kan være svært nyttige å benytte i boligprismodeller.

Funnene i denne masteroppgaven bidrar til litteraturen om hvordan Google-søkedata kan benyttes i forbindelse med boligmarkedet. Boliginvesteringer utgjør en stor andel av norske husholdningers økonomi, og det er alltid relevant å finne nye metoder for å forklare utviklingen i boligmarkedet. Boligprisindeksen publiseres i dag med opp til en måneds etterslep. Ved å benytte modeller basert på Google-data kan man danne seg et øyeblikksbilde av boligprisen. Det kan være nyttig for myndigheter, husholdninger og investorer.

Referanser

- Askitas, N. (2016). Trend-Spotting in the Housing Market. *Cityscape*, 18(2), 165-178.
- Bethlehem, J. G. (2018). *Understanding public opinion polls*. CRC Press.
- Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016, 05). Google searches and stock returns. *International Review of Financial Analysis*, 45, 150-156. doi:10.1016/j.irfa.2016.03.015
- Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Icwsn*, 11, 450-453.
- Carrière-Swallow, Y., & Labbé, F. (2011, 11). Nowcasting with Google Trends in an Emerging Market. *Journal of Forecasting*, 32(4), 289-298. doi:10.1002/for.1252
- Case, K. E., & Shiller, R. J. (2003). Is There a Bubble in the Housing Market? *Brookings Papers on Economic Activity*, 2003(2), 299-362. doi:10.1353/eca.2004.0004
- Columbia Economics. (2010) "How Economists Convert Quarterly Data into Monthly: Cubic Spline Interpolation." *Columbia Economics, L.L.C.*, 12 Sept. 2015, columbiaeconomics.com/2010/01/20/how-economists-convert-quarterly-data-into-monthly-cubic-spline-interpolation/.
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, 65(3), 122-135.
- Dietzel, M. A., Braun, N., & Schäfers, W. (2014, 08). Sentiment-based commercial real estate forecasting with Google search volume data. *Journal of Property Investment & Finance*, 32(6), 540-569. doi:10.1108/jpif-01-2014-0004
- Dietzel, M. A., Braun, N., & Schäfers, W. (2014, 08). Sentiment-based commercial real estate forecasting with Google search volume data. *Journal of Property Investment & Finance*, 32(6), 540-569. doi:10.1108/jpif-01-2014-0004
- Dijk, D. V., & Francke, M. (2015). Internet Search Behavior, Liquidity and Prices in the Housing Market. *SSRN Electronic Journal*. doi:10.2139/ssrn.2654170
- D'Amuri, F., & Marcucci, J. (2017, 10). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33(4), 801-816. doi:10.1016/j.ijforecast.2017.03.004

- ECON Nye boliger. (2017). Retrieved from <http://www.samfunnsokonomisk-analyse.no/enb/>
- Eidjord, O. M., & Oust, A. (2017). Can Google Search be Used as a Housing Bubble Indicator?-a US 2006/07 Bubble Case Study. NTNU Business School
- Eiendom Norge (2017). Boligstatistikk Desember 2017.
- Eiendom Verdi (2018, 05). Den norske engstelsesindeksen. *Eiendom Norge*. Hentet 5/31/2018 fra <http://eiendomnorge.no/den-norske-engstelsesindeksen/>
- Fan, C. S., & Wong, P. (1998, 01). Does consumer sentiment forecast household spending? *Economics Letters*, 58(1), 77-84. doi:10.1016/s0165-1765(97)00247-4
- Finanstilsynet (2017). Boliglånsundersøkelse 2017. *Finanstilsynet*.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2008, 11). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014. doi:10.1038/nature07634
- Google Help2. *How Trends data is adjusted - Trend Help*. Hentet fra https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052.
- Google Help1. *Where Trends data comes from- Trend Help*. Hentet fra https://support.google.com/trends/answer/4355213?hl=en&ref_topic=6248052
- Google Trend. Real Estate: Kjøp og Selg. Hentet fra <https://trends.google.com/trends/explore?cat=29&geo=NO&q=kj%C3%B8p.selg>
- Grimm, P. (2010, 12). Social Desirability Bias. *Wiley International Encyclopedia of Marketing*. doi:10.1002/9781444316568.wiem02057
- Groves, R. M. (2005). *Survey errors and survey costs*. Wiley.
- Grytten, O. H., & Hunnes, A. (2016). *Krakk og kriser i historisk perspektiv*. Cappelen Damm akademisk.
- Hartwig, K. (2016.08.06). *Eiendomsmegler: - Selgernes forventninger er skyhøye*. Dagens Næringsliv. Hentet den 20.06.2018 fra: <https://www.dn.no/privat/2016/08/06/1344/Boligmarkedet/eiendomsmegler-selgernes-forventninger-er-skyhøye>
- Hartwig, K. (2016.01.23). *Slik tjener du penger på å investere i eiendom*. Dagens Næringsliv. Hentet den 20.06.2018 fra: <https://www.dn.no/privat/2016/01/23/0935/Investeringer/slik-tjener-du-penger-p-investere-i-eiendom>
- Harzing, A. W. (2009). Rating versus ranking: What is the best way to reduce response and language bias in cross-national research?. *International Business review*.

-
- Google Trends. How Trends data is adjusted - Trends Help hentet den 04.05.2018 fra https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052
- Jacobsen, D. H., & Naug, B. E. (2004). Hva driver boligprisene. *Penger og Kreditt*, 4(04), 229-240.
- Joseph, K., Wintoki, M. B., & Zhang, Z. (2011, 10). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, 27(4), 1116-1127. doi:10.1016/j.ijforecast.2010.11.001
- Jun, S. P., Yoo, H. S., & Choi, S. (2017). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*.
- Kholodilin, K. A., Podstawski, M., & Siliverstovs, B. (2010). Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US. *SSRN Electronic Journal*. doi:10.2139/ssrn.1615453
- Krumpal, I. (2011, 11). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4), 2025-2047. doi:10.1007/s11135-011-9640-9
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014, 03). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176), 1203-1205. doi:10.1126/science.1248506
- Mak, D. K. (2003). *The science of financial market trading*. World Scientific.
- Mclaren, N., & Shanbhogue, R. (2011). Using Internet Search Data as Economic Indicators. *SSRN Electronic Journal*. doi:10.2139/ssrn.1865276
- Mikalsen, B. (2016.06.09).- *Fristende å være med boligprisstigningen videre i Oslo*. Dagens Næringsliv. Hentet den 20.06.2018 fra: <https://www.dn.no/privat/eiendom/2016/06/09/1001/Boligsalg/-fristende--vre-med-boligprisstigningen-videre-i-oslo>
- NEF (2017). Førstegangskjøpende-analyse. NEF
- O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Icwsn*, 11(122-129), 1-2.
- Parr, O. S. (2014). Eiendomsmegler 1 er størst i Norge. Hentet den 28/05/2018 fra <https://www.hegnar.no/Nyheter/Personlig-oekonomi/2014/02/Eiendomsmegler-1-er-stoerst-i-Norge>
- Pavlicek, J., & Kristoufek, L. (2015, 05). Nowcasting Unemployment Rates with Google Searches: Evidence from the Visegrad Group Countries. *Plos One*, 10(5). doi:10.1371/journal.pone.0127084
- Pentland, A., & Heibeck, T. (2010). *Honest signals: how they shape our world*. MIT press.

- Preis, T., Reith, D., & Stanley, H. E. (2010, 11). Complex dynamics of our economic life on different scales: Insights from search engine query data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1933), 5707-5719.
doi:10.1098/rsta.2010.0284
- Preis, T., Moat, H. S., & Stanley, H. E. (2013, 04). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, 3(1). doi:10.1038/srep01684
- Regjeringen (2018/12/14). Fastsetter ny boliglånsforskrift. Pressemelding. Hentet den 5/6/2018 fra <https://www.regjeringen.no/no/aktuelt/fastsetter-ny-boliglansforskrift/id2523967/>
- Reinhart, C., & Rogoff, K. (2008, 03). This Time is Different: A Panoramic View of Eight Centuries of Financial Crises. doi:10.3386/w13882
- Shiller, R. J. (2015, 01). Irrational Exuberance. doi:10.1515/9781400865536
- Statista. (2018). Global market share of search engines 2010-2017 | Statistic. Tilgjengelig fra URL: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>.
- Statistisk Sentralbyrå. (2012). *Statistisk sentralbyrås budsjett for 2012*.
- Statistisk Sentralbyrå. (2017). Stort flertall eier boligen. Hentet den 6/4/2018 fra <https://www.ssb.no/bygg-bolig-og-eiendom/artikler-og-publikasjoner/stort-flertall-eier-boligen>
- Statscounter. (2018). Desktop, Mobile & Tablet Search Engine Market Share Norway. Retrieved from <http://gs.statcounter.com/search-engine-market-share/desktop-mobile-tablet/norway/#monthly-201701-201801>
- Stäubert, A. (2016). Forventningsbarometeret 4. kv. 2016. Kantar TNS. Retrieved from: <http://www.tns-gallup.no/kantar-tns-innsikt/forventningsbarometeret-1-kv-2016/>
- Sæby, I. K. (2014, Februar 16). Nordmenn er selveiere i særklasse. Hentet 6/4/2018 fra <https://www.nrk.no/okonomi/nordmenn-er-selveiere-i-saerklasse-1.11488063>
- Thomson, A. (2011, June 13). Google Can Help Predict House Prices, Unemployment. Retrieved from <https://blogs.wsj.com/economics/2011/06/13/google-can-help-predict-house-prices-unemployment/>
- Varian, H. R., & Choi, H. (2009). Predicting the Present with Google Trends. *SSRN Electronic Journal*.
doi:10.2139/ssrn.1659302
- Veldhuizen, S. V., Vogt, B., & Voogt, B. (2016, 06). Internet searches and transactions on the Dutch housing market. *Applied Economics Letters*, 23(18), 1321-1324.
doi:10.1080/13504851.2016.1153785

Wig, K. (2018). Strid om nye boligsalgstall: – Tolker statistikk som fanden leser bibelen. E24 Hentet den 25.05.2018 fra <https://e24.no/privat/bolig/strid-om-nye-boligsalgstall-tolker-statistikk-som-fanden-leser-bibelen/24246345>

Whitaker, S. D. (2018, 02). Big Data versus a survey. *The Quarterly Review of Economics and Finance*, 67, 285-296. doi:10.1016/j.qref.2017.07.011

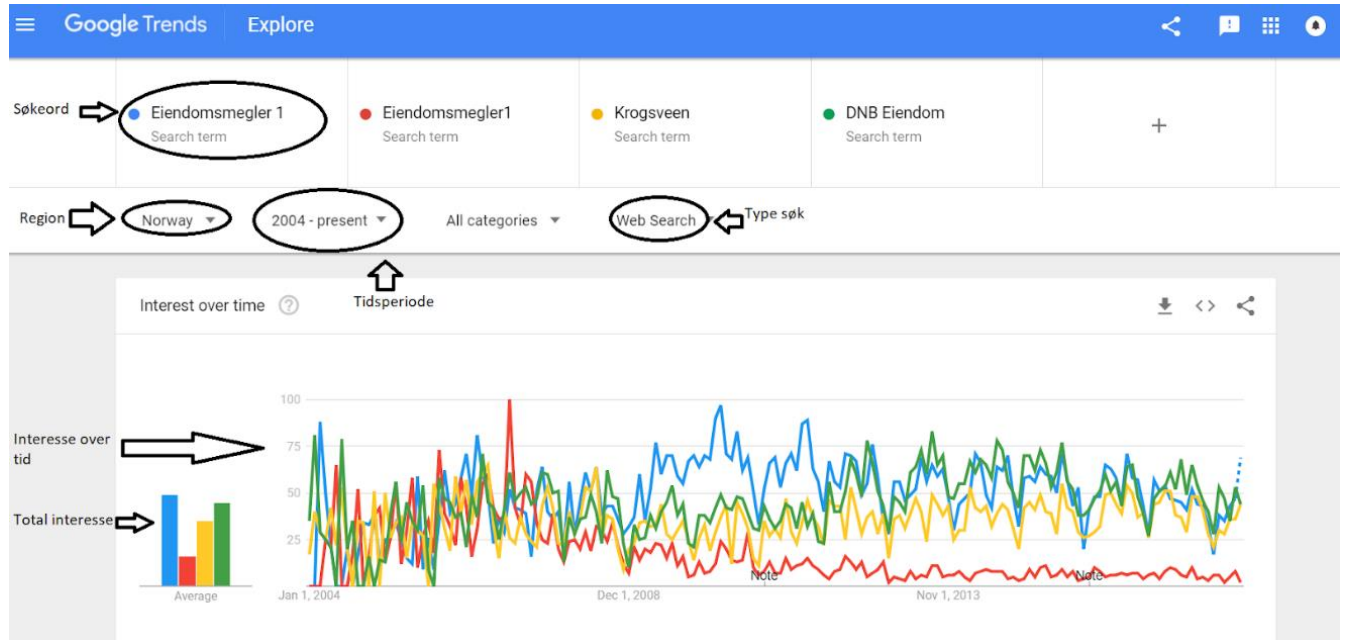
Wooldridge, J. M. (2016). *Introductory econometrics a modern approach*. Thomson/South-Western.

Wu, L., & Brynjolfsson, E. (2015). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. *SSRN Electronic Journal*. doi:10.2139/ssrn.2022293

Yu, L., Zhao, Y., Tang, L., & Yang, Z. (2018, 01). Online big data-driven oil consumption forecasting with Google trends. *International Journal of Forecasting*. doi:10.1016/j.ijforecast.2017.11.005

Appendiks A – Datasett og variabler

A1: Hvordan Google Trends fungerer



Stillbilde av hvordan Google Trends fungerer.

A2: Hvordan Google Adwords fungerer

Keyword ↑	Avg. monthly searches	Competition
beste rente boliglån	100 – 1K	High
boligannonser	100 – 1K	Low
boligkjøp	100 – 1K	Low
boliglån	1K – 10K	High
boliglånsrente	1K – 10K	Medium
boligpriser	1K – 10K	Low
boligrente	100 – 1K	High
egenkapital bolig	100 – 1K	Medium
eiendomsmegler	1K – 10K	High
finn eiendom	10K – 100K	Low
huskjøp	10 – 100	Low
huspriser	100 – 1K	Low
kjøp av bolig	10 – 100	High
kjøpe bolig	100 – 1K	High
kjøpe hus	100 – 1K	High

Eksempel på hvordan Google Adwords fungerer.

A3: Oversikt over samtlige Google-søkeord vi har vurdert

Søkeord	Relevant ifølge økonomisk teori?	God nok kvalitet på Google Trends?	Konkurransen på Google Adwords	Stort nok søkevolum?	Benyttet i oppgaven?	Tilbud (T)/ Etterspørsel (E)
Ny bolig	Nei	-	-	-	-	-
Selge bolig uten megler	Nei	-	-	-	-	-
Notar	Nei	-	-	-	-	-
Partners	Nei	-	-	-	-	-
Boble	Nei	-	-	-	-	-
Krakk	Nei	-	-	-	-	-
Eiendom Norge	Nei	-	-	-	-	-
Eiendom	Nei	-	-	-	-	-
Feriehus	Nei	-	-	-	-	-
FINN	Nei	-	-	-	-	-
FINN.no	Nei	-	-	-	-	-
Gjeld	Nei	-	-	-	-	-
Hage	Nei	-	-	-	-	-
Hytte	Nei	-	-	-	-	-
Innredning	Nei	-	-	-	-	-
Megler	Nei	-	-	-	-	-
Møbler	Nei	-	-	-	-	-
Oppussing	Nei	-	-	-	-	-
Renovasjon	Nei	-	-	-	-	-
Fritidsbolig	Nei	-	-	-	-	-
Kjøpe hvitevarer	Nei	-	-	-	-	-
Kjøpe oppvaskmaskin	Nei	-	-	-	-	-
Kjøpe tørketrommel	Nei	-	-	-	-	-
Kjøpe vaskemaskin	Nei	-	-	-	-	-

Søkeord eliminert på grunn av kriterie (1) Økonomisk relevans.

Søkeord	Relevant ifølge økonomisk teori?	God nok kvalitet på Google Trends?	Konkurransen på Google Adwords	Stort nok søkevolum?	Benyttet i oppgaven?	Tilbud (T)/ Etterspørsel (E)
Avdrag	Ja	Nei	-	-	-	-
Boligpriser 2012	Ja	Nei	-	-	-	-
Boligpriser 2013	Ja	Nei	-	-	-	-
Boligpriser 2014	Ja	Nei	-	-	-	-
Boligpriser 2015	Ja	Nei	-	-	-	-
Boligpriser 2016	Ja	Nei	-	-	-	-
Boligpriser 2017	Ja	Nei	-	-	-	-
Nordvik Partners	Ja	Nei	-	-	-	-
Eiendomsmegler DNB	Ja	Nei	-	-	-	-
Eiendomsmegler Krogsveen	Ja	Nei	-	-	-	-
Kjøp av hus	Ja	Nei	-	-	-	-
Kjøp av leilighet	Ja	Nei	-	-	-	-
Kjøpe bolig i Bergen	Ja	Nei	-	-	-	-
Kjøpe bolig i Oslo	Ja	Nei	-	-	-	-
Kjøpe brukt bolig	Ja	Nei	-	-	-	-
Leilighetskjøp	Ja	Nei	-	-	-	-
Lånebevis	Ja	Nei	-	-	-	-
Boliglånsrente DnB	Ja	Nei	-	-	-	-
Boliglånsrente Danske bank	Ja	Nei	-	-	-	-
Boliglånsrente Skandiabanken	Ja	Nei	-	-	-	-
Boliglånsrente Sbanken	Ja	Nei	-	-	-	-
Boliglånsrente Nordea	Ja	Nei	-	-	-	-

Søkeord eliminert på grunn av kriterie (2) Tilstrekkelig datakvalitet på Google Trends, del (1/2).

Søkeord	Relevant ifølge økonomisk teori?	God nok kvalitet på Google Trends?	Konkurransen på Google Adwords	Stort nok søkevolum?	Benyttet i oppgaven?	Tilbud (T)/ Etterspørsel (E)
Boliglån DnB	Ja	Nei	-	-	-	-
Boliglån Danske bank	Ja	Nei	-	-	-	-
Boliglån Skandiabanken	Ja	Nei	-	-	-	-
Boliglån Sbanken	Ja	Nei	-	-	-	-
Boliglån Nordea	Ja	Nei	-	-	-	-
Boligpriser 2004	Ja	Nei	-	-	-	-
Boligpriser 2005	Ja	Nei	-	-	-	-
Boligpriser 2006	Ja	Nei	-	-	-	-
Boligpriser 2007	Ja	Nei	-	-	-	-
Boligpriser 2008	Ja	Nei	-	-	-	-
Boligpriser 2009	Ja	Nei	-	-	-	-
Boligpriser 2010	Ja	Nei	-	-	-	-
Boligpriser 2011	Ja	Nei	-	-	-	-
Boligstatistikk	Ja	Nei	-	-	-	-
Boligboble	Ja	Nei	-	-	-	-
Boligkrakk	Ja	Nei	-	-	-	-
Finansieringsbevis	Ja	Nei	-	-	-	-

Søkeord eliminert på grunn av kriterie (2) Tilstrekkelig datakvalitet på Google Trends, del (2/2).

Søkeord	Relevant ifølge økonomisk teori?	God nok kvalitet på Google Trends?	Konkurranse på Google Adwords	Stort nok søkevolum?	Benyttet i oppgaven?	Tilbud / Etterspørsel/ Begge
Kjøp av bolig	Ja	Ja	Høy	Nei	Nei	-
Rente boliglån	Ja	Ja	Høy	Nei	Nei	-
Salg av hus	Ja	Ja	Høy	Nei	Nei	-
Selge hus	Ja	Ja	Høy	Nei	Nei	-
Selge leilighet	Ja	Ja	Høy	Nei	Nei	-
Egenkapital bolig	Ja	Ja	Medium	Nei	Nei	-
Boligannonse	Ja	Ja	Lav	Nei	Nei	-
Boligkjøp	Ja	Ja	Lav	Nei	Nei	-
Huskjøp	Ja	Ja	Lav	Nei	Nei	-
Huspriser	Ja	Ja	Lav	Nei	Nei	-
Boligpriser	Ja	Ja	Lav	Nei	Nei	-

Søkeord eliminert på grunn av kriterie (3) Tilstrekkelig absolutte søkertall og kriterie (4) Konkurranse om søkeordet.

Søkeord	Relevant ifølge økonomisk teori?	God nok kvalitet på Google Trends?	Konkurranse på Google Adwords	Stort nok søkevolum?	Benyttet i oppgaven?	Tilbud-/Etterspørsel-proxy
Kjøpe hus	Ja	Ja	Høy	I kombinasjon	Ja	Etterspørsel
Kjøpe leilighet	Ja	Ja	Høy	I kombinasjon	Ja	Etterspørsel
Salg av bolig	Ja	Ja	Høy	I kombinasjon	Ja	Tilbud
Selge bolig	Ja	Ja	Høy	I kombinasjon	Ja	Tilbud
Verdivurdering av bolig	Ja	Ja	Høy	I kombinasjon	Ja	Tilbud
Verdivurdering bolig	Ja	Ja	Høy	I kombinasjon	Ja	Tilbud
Gratis verdivurdering	Ja	Ja	Medium	I kombinasjon	Ja	Tilbud
Kjøpe bolig	Ja	Ja	Høy	Ja	Ja	Etterspørsel
Bolig lån	Ja	Ja	Høy	Ja	Ja	Etterspørsel
Boliglån	Ja	Ja	Høy	Ja	Ja	Etterspørsel
Boliglånsrente	Ja	Ja	Medium	Ja	Ja	Etterspørsel
Boligrente	Ja	Ja	Høy	Ja	Ja	Etterspørsel
Eiendomsmegler	Ja	Ja	Høy	Ja	Ja	Tilbud
Verdivurdering	Ja	Ja	Høy	Ja	Ja	Tilbud
Finn Eiendom	Ja	Ja	Lav	Ja	Ja	Etterspørsel
DnB Eiendom	Ja	Ja	Lav	Ja	Ja	Tilbud
Krogsveen	Ja	Ja	Lav	Ja	Ja	Tilbud
Eiendomsmegler1	Ja	Ja	Lav	Ja	Ja	Tilbud
Eiendomsmegler 1	Ja	Ja	Lav	Ja	Ja	Tilbud

Søkeordene vi har valgt ut til å konstruere Google-variabler i vår masterutredningen. Tilfredsstill alle fire kriterier.

A4: Beskrivelse av variabler benyttet i masterutredningen

Variabelnavn	Forklaring
Boliglån	Googlesøk etter boliglån
Boliglånsrente	Googlesøk etter boliglånsrente
Eiendomsmegler	Googlesøk etter eiendomsmegler
Navn på individuelle eiendomsmeglere	Googlesøk etter dnb eiendom + krogsvveen + eiendomsmegler 1
Verdivurdering	Googlesøk etter verdivurdering + gratis verdivurdering + verdivurdering av bolig
Kjøp av bolig	Googlesøk etter kjøpe bolig + kjøpe hus + kjøpe leilighet
Salg av bolig	Googlesøk etter salg av bolig + selge bolig
Finn Eiendom	Googlesøk etter finn eiendom + finn eiendom no
Boligmasse	Total boligmasse i Norge
Lønnsinntekt	Husholdningenes lønnsinntekter, glattet serie. Årstall er transformert til kvartalstall
KPI-JAE	Konsumprisindeks justert for avgiftsendringer og uten energivarer
Arbeidsledighet	Registrert ledighet
Boliglånsrente	Bankenes utlånsrente på lån til husholdninger. Rente på alle utlån frem til og med 2001.3.
Kapitalskatt	Marginal skatt på kapitalinntekter og -utgifter
TNS Gallup	Forventningsindikatoren til TNS Gallup

Beskrivelse av samtlige variabler benyttet i masterutredningen.

A5: Mailkorrespondanse med DnB Eiendom

Ved Renate Berentsen, Saksbehandler/Megler.

Mail 1:

Hei!

Mitt navn er Haakon Resaland og jeg skriver for tiden masteroppgave ved NHH om boligmarkedet. Vi ser på om det er mulig å predikere endringer i det norske boligmarkedet ved hjelp av Google-søk. Der vi har en blindsoner er den praktiske erfaringen som eiendomsmeglere har. Vi lurte derfor på om dere kunne hjelpe oss ved å svare på fire spørsmål?

Håper å høre fra dere.

Svar 1:

Hei

Send oss gjerne spørsmålene så skal jeg høre med meglerne her på kontoret om de har mulighet.

Mail 2:

Takk for å svare så raskt.

Det viktigste for oss er å få en generell følelse for hva som er vanlig i markedet, ikke eksakte tall. Vi lurte på:

- 1) Andel som tar kontakt via e-post/telefon?
- 2) Andel som ber om verdivurdering av boligen?
- 3) Andel av disse som velger dere som eiendomsmegler?
- 4) Hvor lang tid fra kunden tar kontakt til boligen legges ut på markedet?
- 5) Flest som selger eller kjøper først?
- 6) Generelt hvor høy andel av boligene annonseres på FINN?

Svar 2:

Hei

Har prøvd å svare så godt som mulig, men ta gjerne kontakt igjen om dere lurte på noe mer.

- 1) Andel som tar kontakt via e-post/telefon?

Begge deler er vanlig, men vi får flest digitale henvendelser. De som tar kontakt pr. telefon ringer gjerne direkte til en megler. Ikke så mange som lenger ringer til kontoret.

2) Andel som ber om verdivurdering av boligen?

De fleste som tar kontakt ønsker verdivurdering som et utgangspunkt sammen med et salgstilbud fra megler, men det er også en del kunder som har allerede har bestemt seg for å bruke DNB Eiendom når de tar kontakt.

En del tar også kontakt kun for å få verdivurdering ifbm refinansiering, arveoppgjør, samlivsbrudd.

3) Andel av disse som velger dere som eiendomsmegler?

Ca 50% av våre befaringer ender med salgsoppdrag

4) Hvor lang tid fra kunden tar kontakt til boligen legges ut på markedet?

Alt etter kundens behov. Alt fra 2 uker – 1 år

Vi kan ha boligen salgsklar i løpet av kort tid. Alt avhenger av om taksten er klar og bilder tatt.

5) Flest som selger eller kjøper først?

Flest som kjøper først i Bergen

6) Generelt hvor høy andel av boligene annonseres på FINN?

99%

Mail 3:

Hei Renate,

Mange takk for raske og gode svar.

Et oppfølgingsspørsmål angående hvor lang tid det tar fra kunden tar kontakt til de selger.

Er det mulig å anslå ca. fordeling av hvor lang tid det tar fra kunden tar kunder til boligen er på markedet? Kommer den som regel på markedet i løpet av de første 3 månedene eller er det jevnt fordelt utover det kommende året?

Takk igjen.

Svar 3:

Hei igjen

Synes det er litt vanskelig å anslå fordelingen. Dette er kanskje litt sesongavhengig.

Når først kunden har bestemt seg for å bruke DNB Eiendom til å selge, så går det ganske kort tid fra de tar kontakt til boligen er i markedet. Gjerne 1-2 uker.

Det vi merker er at kundene i større og større grad tar kontakt tidligere i prosessen. Dette skyldes at de gjerne ønsker at megler skal gi de gode råd både i forhold til oppussing/styling av boligen som skal selges, og at de ønsker å bruke megler som en rådgiver ifbm med kjøp og budgivning på ny bolig. Noen tar gjerne kontakt om vinteren, men venter med å markedsføre boligen til våren/sommeren for å få de beste bildene.

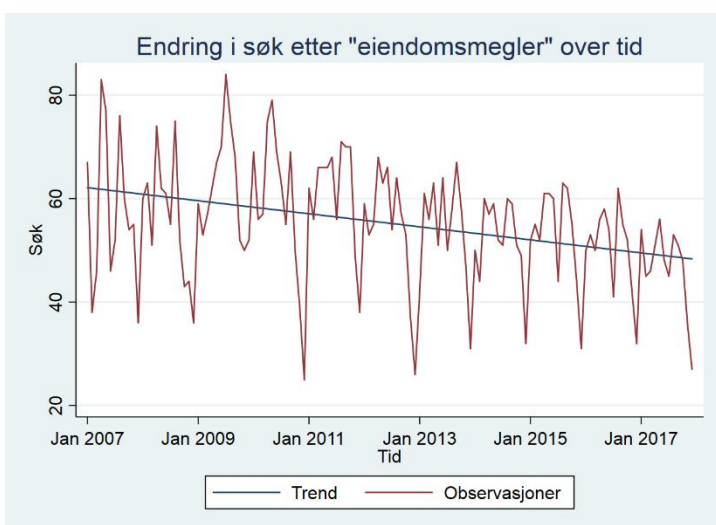
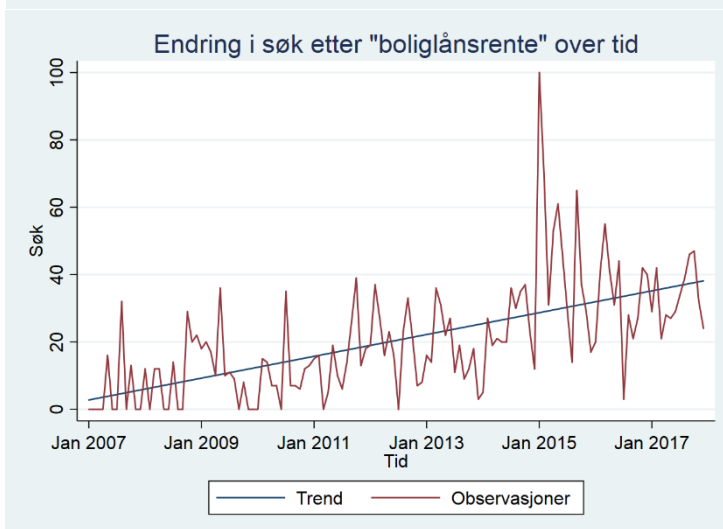
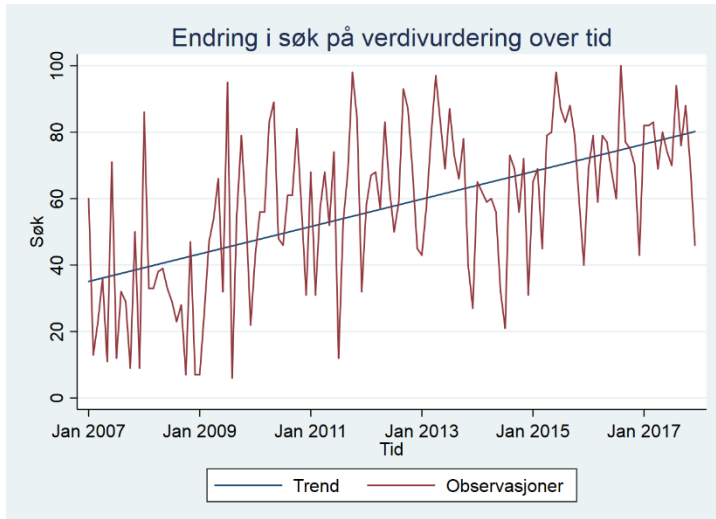
A6: Justering av variabler

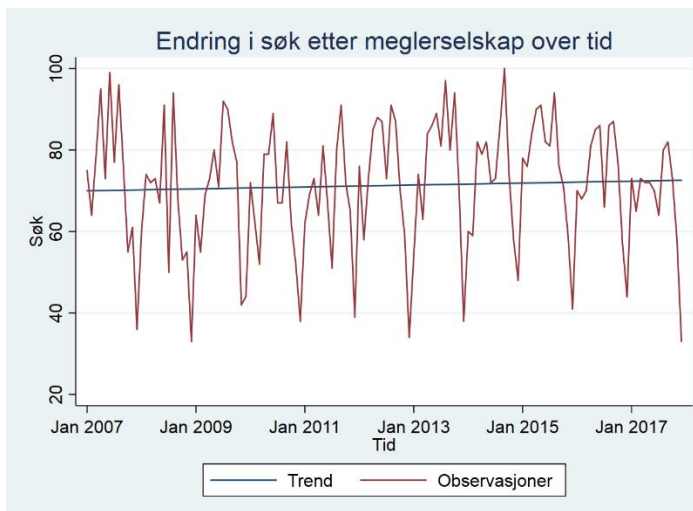
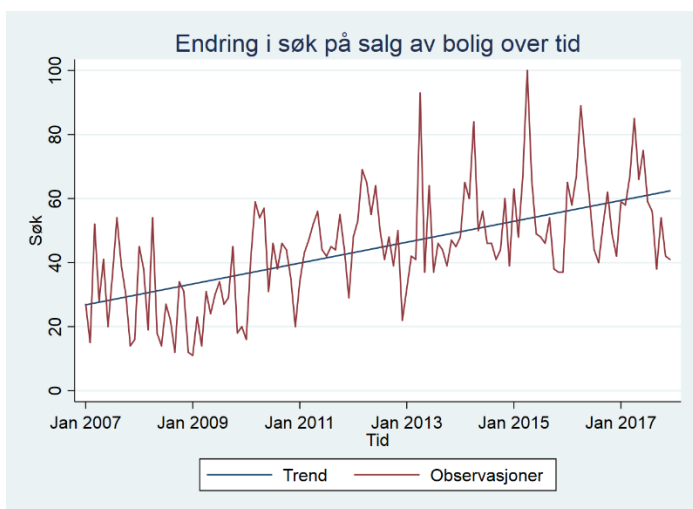
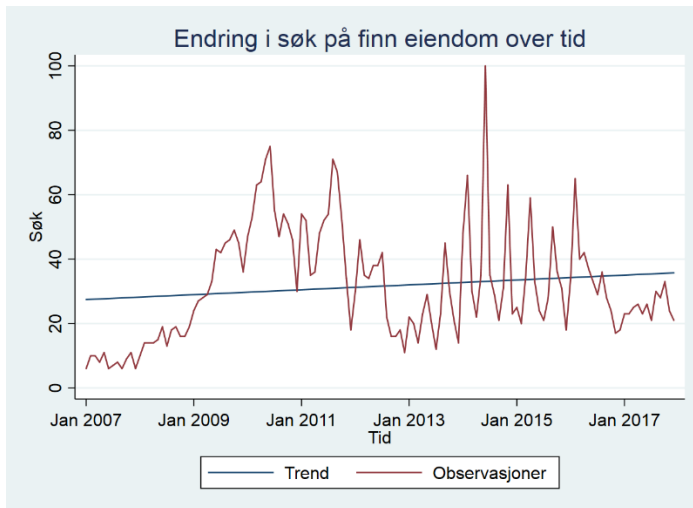
Variabelnavn	Type variabel	Kilde	Justeringer
Prisindeks	Avhengig	Eiendom Norge	Trendjustert. Mottatt som sesongjustert fra Eiendom Norge.
Boliglån	Uavhengig	Google	Trend- og sesongjustert, naturlige logaritmen
Boliglånsrente	Uavhengig	Google	Trend- og sesongjustert, naturlige logaritmen
Eiendomsmegler	Uavhengig	Google	Trend- og sesongjustert, naturlige logaritmen
Selskapsnavn	Uavhengig	Google	Trend- og sesongjustert, naturlige logaritmen
Verdivurdering	Uavhengig	Google	Trend- og sesongjustert, naturlige logaritmen
Boligkjøp	Uavhengig	Google	Trend- og sesongjustert, naturlige logaritmen
Boligsalg	Uavhengig	Google	Trend- og sesongjustert, naturlige logaritmen
Finn	Uavhengig	Google	Trend- og sesongjustert, naturlige logaritmen
Boligmasse	Uavhengig	Makro	Trendjustert, naturlige logaritmen, transformert fra kvartal til måned
Lønn	Uavhengig	Makro	Trendjustert, naturlige logaritmen, transformert fra kvartal til måned
KPI-JAE	Uavhengig	Makro	Trendjustert, transformert fra kvartal til måned
Arbeidsledighet	Uavhengig	Makro	Transformert fra kvartal til måned
Rente	Uavhengig	Makro	Transformert fra kvartal til måned
Skatt	Uavhengig	Makro	Ingen endring
TNSGallup	Uavhengig	Makro	Transformert fra kvartal til måned

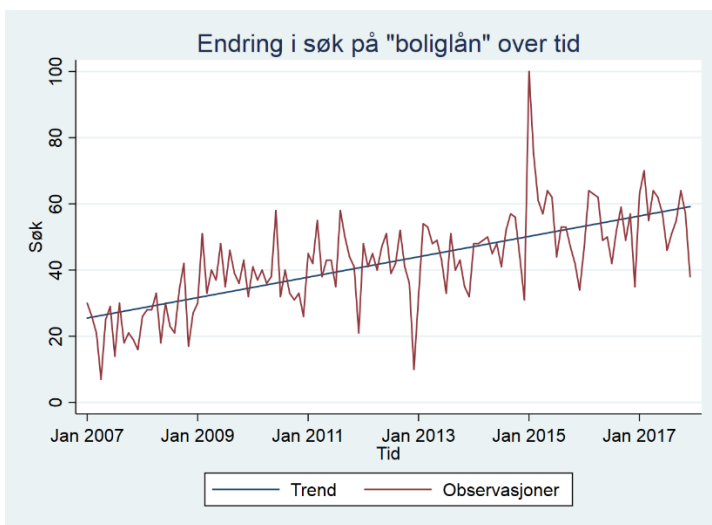
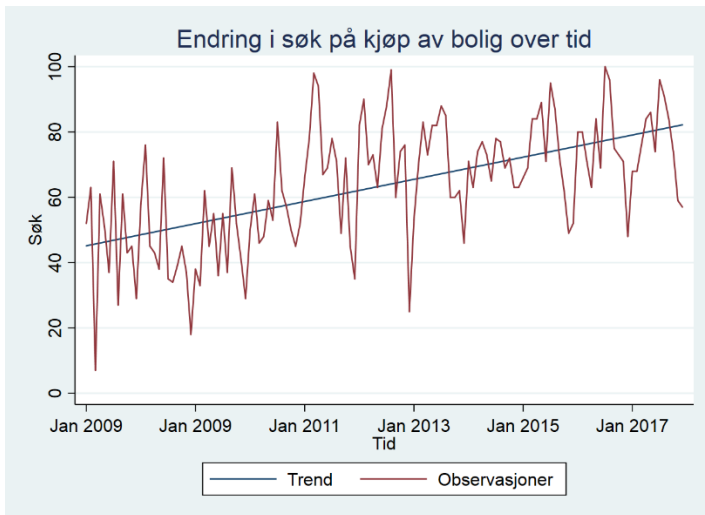
Oversikt over hvordan samtlige variabler er justert før de benyttes til å estimere modeller i analysen.

Appendiks B – Uavhengige variabler

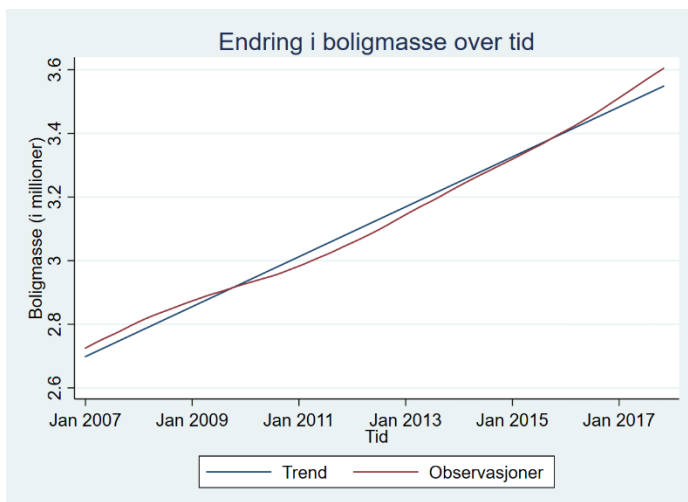
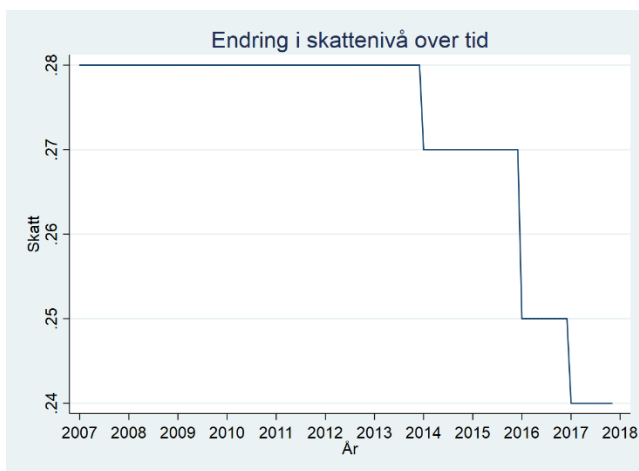
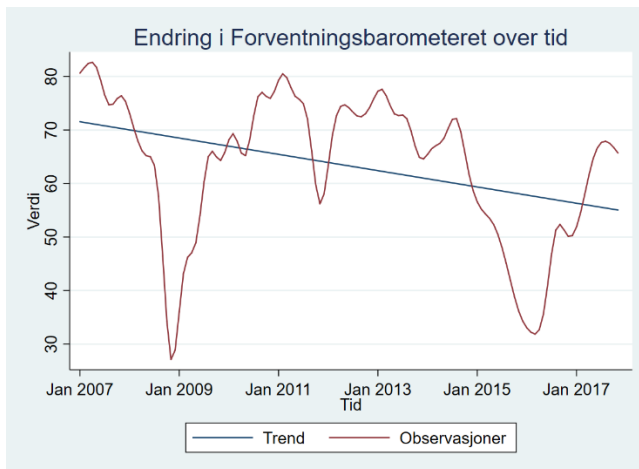
B1: Grafisk fremstilling av Google-variablene

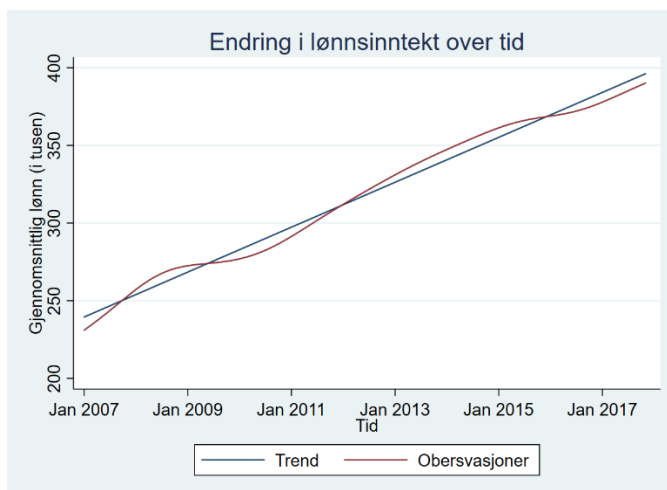
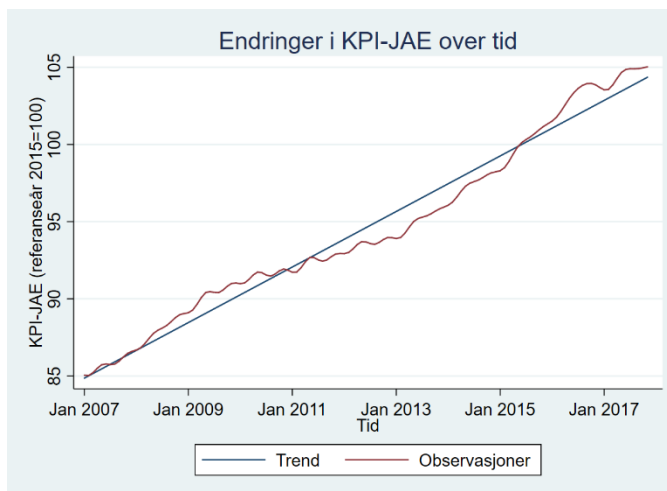
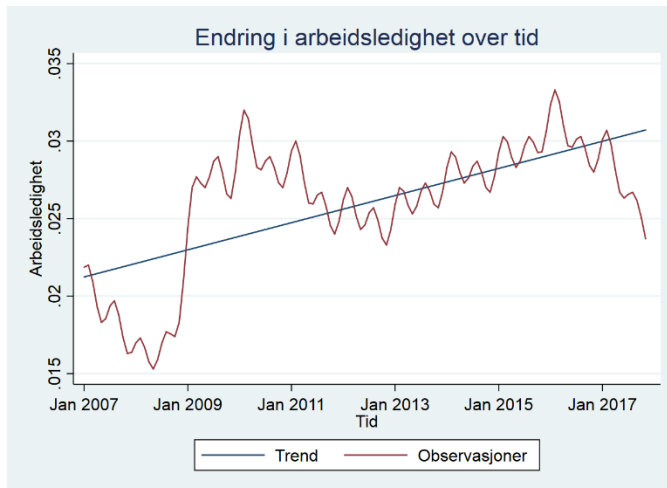


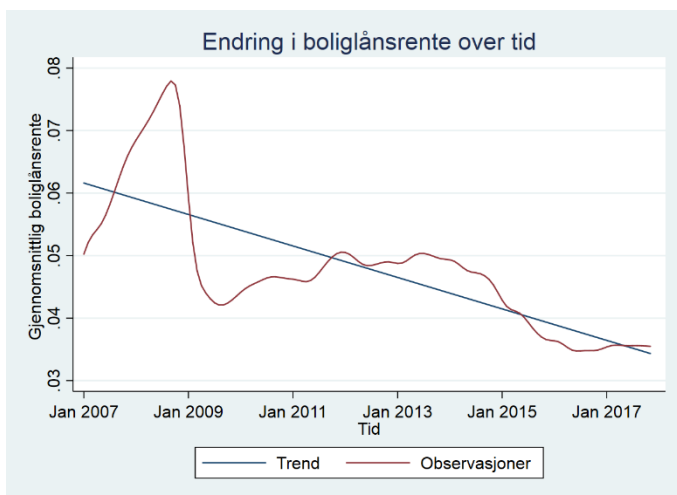




B2: Grafisk fremstilling av makrovariablene







Appendiks C – Regresjonsutskrifter

C1: Regresjonsresultater med forventningsmodellen

	Forventning	Modell 3	Forventning 2016	Forventning 2016 for modell 3
Forklaringsvariabel	Avhengig variabel: Prisindeks (trend- og sesongjustert)			
Boliglån	-33,73*** (3,564)	-39,71*** (10,57)	-35,22*** (4,560)	-41,33*** (11,06)
Eiendomsmegler	49,03*** (8,854)	52,90*** (9,835)	53,43*** (9,804)	47,87*** (9,789)
Selskapsnavn		-34,94* (13,43)	-32,94* (14,35)	-30,89* (14,59)
Boligsalg		8,060 (5,208)	9,477** (3,473)	7,003 (5,451)
Boliglånsrente		9,042* (4,043)	8,650** (2,834)	11,08** (3,645)
Verdivurdering		4,139 (3,095)		4,535 (2,716)
Boligkjøp		-6,966 (6,776)		-2,265 (6,302)
Finn		0,694 (2,680)		2,133 (2,946)
Konstantledd	-3,14e-09 (0,567)	-7,19e-09 (0,507)	-0,343 (0,504)	-0,379 (0,473)
<i>Observasjoner</i>	132	132	120	120
<i>Adj. R²</i>	0,5599	0,6228	0,6592	0,6699

Newey West standardavvik i parentes

* $p < 0,05$ ** $p < 0,01$, *** $p < 0,001$

Modell 3 henviser til modell 3 i kapittel seks, hvor samtlige Google-variabler er inkludert.

Forventning2016 for modell 3 er modell 3 utført på det avkortede datasettet, fra 2007 til 2016.

C2: Regresjonsresultater med fundamentalmodellen

Forklaringsvariabel	Fundamental	Modell 4	Fundamental2016	Fundamental2016 for modell 4
	Avhengig variabel: Prisindeks (trend- og sesongjustert)			
Lønn	-120,8*** (14,46)	-166,4*** (27,09)	-160,2*** (16,65)	-139,2*** (21,12)
KPI-JAE	-523,0*** (116,5)	-268,0** (84,57)	-270,1*** (40,92)	-237,4*** (57,27)
Arbeidsledighet	-183,1*** (53,65)	-950,3*** (227,1)	-1409,8*** (168,8)	-1312,4*** (176,9)
Skatt	9,095* (3,553)	-143,0** (54,50)	-231,1*** (37,34)	-240,6*** (45,05)
Rente		-136,6 (107,7)	-292,1*** (76,08)	-298,5*** (71,73)
Boligmasse		15,28 (118,0)		124,8 (85,47)
TNS Gallup		-0,960 (4,572)		5,953 (3,711)
Konstantledd	66,85*** (18,06)	70,28*** (19,20)	114,0*** (12,89)	113,7*** (12,87)
Observasjoner	131	131	120	120
Adj. R ²	0,7125	0,7145	0,8450	0,8555

Newey West standardavvik i parentes

* $p < 0,05$ ** $p < 0,01$, *** $p < 0,001$

Modell 4 henviser til modell 4 i kapittel seks, hvor samtlige makrovariabler er inkludert.

Fundamental2016 for modell 4 er modell 4 utført på den avkortede datasettet, fra 2007 til 2016.

C3: Regresjonsresultater med hybridmodellen

Forklaringsvariabel	Hybrid	Modell 5	Hybrid2016	Hybrid2016 for modell 5
	Avhengig variabel: Prisindeks (trend- og sesongjustert)			
Lønn	-120,8*** (14,46)	-147,8*** (38,60)	-105,0*** (15,16)	-127,6*** (24,76)
Arbeidsledighet	-523,0*** (116,5)	-382,3 (209,0)	-764,5*** (179,3)	-887,8*** (217,2)
Skatt	-183,1*** (53,65)	-236,3* (95,14)	-148,8** (51,07)	-196,2** (68,89)
TNS Gallup	9,095* (3,553)	10,27** (3,419)	11,78*** (2,080)	10,29*** (2,848)
<i>Boliglån</i>	-20,24*** (3,692)	-11,21 (6,740)	-11,00*** (2,600)	-8,034 (6,294)
<i>Boliglånsrente</i>	8,614*** (1,307)	6,293* (2,865)	5,633*** (1,225)	7,306*** (2,145)
Boligmasse	-580,8*** (109,5)	-874,3* (436,7)		240,7 (350,5)
<i>Boligkjøp</i>		4,861 (6,980)	14,10*** (3,058)	8,378 (5,621)
Rente		111,8 (150,2)	-189,2** (60,52)	-209,7* (105,5)
KPI-JAE		57,37 (140,6)		-181,4 (113,9)
<i>Eiendomsmegler</i>		-3,900 (9,566)		0,978 (8,454)
<i>Selskapsnavn</i>		10,73 (10,80)		-14,49 (9,498)
<i>Verdivurdering</i>		-4,554 (3,081)		2,659 (2,549)
<i>Boligsalg</i>		-0,231 (3,274)		1,074 (3,563)
<i>Finn</i>		-3,455 (2,947)		0,882 (2,389)

Konstantledd	62,22*** (16,60)	67.52*** (19.86)	68,10*** (17,84)	85.61*** (22.82)
<i>Observasjoner</i>	131	131	120	120
<i>Adj. R²</i>	0.8288	0.8317	0.8937	0.8963

Newey West standardavvik i parentes

* $p < 0,05$ ** $p < 0,01$, *** $p < 0,001$

Variabler i kursiv er Boligsøksvariabler

Variabler i fet skrift er Makroøkonomiske variabler

Modell 5 henviser til modell 5 i kapittel seks, hvor samtlige Google-variabler og makrovariabler er inkludert.

Hybrid2016 for modell 5 er modell 5 utført på den avkortede datasettet, fra 2007 til 2016.