



Forprosjekt om boligformuesmodellen

TALL

SOM FORTELLER

NOTATER / DOCUMENTS

2022/1

Erlend Eide Bø, Per Medby, Odd Erik Nygård og Mona Takle

I serien Notater publiseres dokumentasjon, metodebeskrivelser, modellbeskrivelser og standarder.

© Statistisk sentralbyrå

Publisert: 23. mars 2022

Rettet 11.04.2022, side 12

ISBN 978-82-587-1495-5 (elektronisk)

ISSN 2535-7271 (elektronisk)

Standardtegn i tabeller	Symbol
Ikke mulig å oppgi tall Tall finnes ikke på dette tidspunktet fordi kategorien ikke var i bruk da tallene ble samlet inn.	.
Tallgrunnlag mangler Tall er ikke kommet inn i våre databaser eller er for usikre til å publiseres.	..
Vises ikke av konfidensialitetshensyn Tall publiseres ikke for å unngå å identifisere personer eller virksomheter.	:
Desimaltegn	,

Forord

Den eksisterende boligverdsettingsmodellen utarbeidet av SSB har vist seg å være utsatt for endringer i kommunegrenser. På oppdrag fra Finansdepartementet har SSB gjennomført et begrenset forprosjekt for å undersøke hvordan verdsettelse av boliger i større grad kan frigjøres fra kommunegrenser.

Notatet er skrevet av Per Medby og Mona Takle fra Seksjon for eiendoms-, areal- og primærnæringsstatistikk, og Erlend Eide Bø og Odd Erik Nygård fra Forskningsavdelingen. Erik Engelién og Sturla Løkken har gitt nyttige bidrag underveis.

Statistisk sentralbyrå, 16. mars 2022

Per Morten Holt

Sammendrag

Dette notatet beskriver mulige endringer i verdsettelsesmodellen for å frigjøre seg fra kommunestrukturen, og hvor mye treffsikkerheten kan forbedres ved disse endringene. Vi fokuserer på tre fylker med forskjellige karakteristikk: Oslo, Hedmark og Møre og Romsdal.

Statistisk sentralbyrå (SSB) har tidligere utviklet en modell for å beregne markedsverdier for boliger som ble tatt bruk i 2009, og siden har blitt brukt til å beregne formuesgrunnlaget for boliger. Flere kommuner bruker også dette formuesgrunnlaget ved utskrivning av eiendomsskatt.

Boligverdi beregnes separat for forskjellige regioner og bygningstyper (beregningsområde). De enkelte beregningsområdene er inndelt i soner basert på prisnivå. Hvert beregningsområde deles opp i prissoner, der hovedregelen er at soner består av hele kommuner. Kommunereformen i 2020 førte til at modellen treffer dårligere enn før, fordi kommuner i forskjellige soner ble slått sammen.

Vi utarbeider en modell basert på grunnkretser, og en maskinlæringsalgoritme for å gruppere grunnkretsene til grupper med tilstrekkelig antall observasjoner. De 14 000 grunnkretsene i Norge deles opp i ca. 1 200 grupperinger, separat for bygningstype. For hver kombinasjon av geografisk område og bygningstype kjøres så en separat hedonisk regresjon for å predikere boligverdi.

Den grunnleggende avveiningen ved aggregering av grunnkretser er at man ønsker å skape små områder som er så homogene som mulig, samtidig som man er avhengig av et visst antall omsetninger i den hedoniske regresjonen. Det resulterer i mange, geografisk små grupper i sentrale strøk, og geografisk større grupper i mindre sentrale strøk. Slik utnyttes den store mengden informasjon i byer til å bedre predikere lokale prisnivåer.

Resultater fra grunnkretsmodellen vises, og sammenlignes med resultater fra den eksisterende modellen for både 2019 og 2020. Grunnkretsmodellen gir tydelige forbedringer i prediksjonskraften for beregnet boligverdi. Dette gjelder for alle tre fylker, og for fire forskjellige mål på prediksjonskraft. I tillegg predikerer grunnkretsmodellen boligverdien for dyre boliger i Oslo mye bedre. Dette har vært en kjent svakhet ved den eksisterende modellen. Resultatene for den nye modellen er svært like for 2019 og 2020, noe som tyder på at sammenslåing av kommuner har liten effekt på modellen, og at den er relativt stabil over tid.

Vi diskuterer også kort andre metoder som kan brukes for å bedre utnytte informasjonen som er tilgjengelig i dagens data.

Notatet avsluttes med en diskusjon av resultatene, samt gjenstående utfordringer ved bruken av grunnkretsmodellen for beregning av boligverdi for hele boligmassen.

Innhold

Forord	3
Sammendrag	4
1. Bakgrunn og avgrensninger	6
1.1. Effekter av regionreformen	7
1.2. Andre svakheter ved modellen.....	9
2. Modell tilpasset endringer i kommunestruktur	11
2.1. Maskinlæring på grunnkretser	11
2.2. Sentralitetsindeks på grunnkrets nivå	12
2.3. Alternative metoder	13
3. Resultater	15
3.1. 2019	15
3.2. 2020	16
3.3. Dyre boliger	18
4. Oppsummering og gjenstående utfordringer	20
4.1. Gjenstående utfordringer.....	20
Referanser	22
Vedlegg A: Ekstra figurer	23
Vedlegg B: Modell med sentralitetsindeks for grunnkretser	29

1. Bakgrunn og avgrensninger

Statistisk sentralbyrå (SSB) har utviklet en metode for å beregne markedsverdier for boliger som ble tatt i bruk i 2009. Siden modellen ble etablert har SSB på oppdrag fra Skatteetaten brukt modellen til å beregne formuesgrunnlaget for boliger. Fra og med 2014 har flere kommuner tatt i bruk dette formuesgrunnlaget ved utskrivning av eiendomsskatt.

Boligverdi beregnes separat for hver kombinasjon av boligtype og region (beregningsområde). De enkelte beregningsområdene er inndelt i soner basert på prisnivå. Det kjøres en regresjon per beregningsområde med soner som forklaringsvariabler sammen med tettstedsinndeling, byggeår og areal. I modellen er hovedregelen at hele kommunen tilhører én sone, og ingen soner går på tvers av kommunegrenser. Unntaket er de fire største byene, der bydeler kan tilhøre ulike prissone. Modellen er nærmere beskrevet i Medby og Takle (2021).

Kommuneinndelingen ble vesentlig endret per 1.1.2020. Det førte til at modellen traff dårligere enn før i noen regioner, fordi kommuner i forskjellige soner ble slått sammen. På denne bakgrunn fikk SSB i oppdrag av Finansdepartementet å vurdere et alternativ til dagens verdsettelsesmodell som i mindre grad er basert på administrative grenser slik at den blir mer robust overfor endringer i kommunestruktur.

Hovedformålene med dette notatet er å kartlegge muligheter for å frigjøre seg fra bindingen til kommunestruktur som finnes i den eksisterende modellen, og å undersøke hvor mye treffsikkerheten kan forbedres ved slike endringer. Resultatene vil også kunne fungere som et sammenligningsgrunnlag i forbindelse med en eventuell større revisjon av dagens verdsettelsesmodell, ved f.eks. bruk av maskinlæring. Det vises i denne sammenhengen til prosjektet med utvikling av en ny verdsettelsesmodell for fritidsbolig i regi av PWC, som i et høringsnotat er skissert implementert fra og med inntektsåret 2024 (PWC, 2020; Finansdepartementet, 2021).

Vi undersøker følgende regioner:

- Oslo
- Hedmark (Innlandet)
- Møre og Romsdal

Valgene av hhv. Hedmark og Møre og Romsdal skyldes at Hedmark er upåvirket av kommunereformen, mens Møre og Romsdal er sterkt påvirket av kommunereformen. Valget av Oslo skyldes mest at vi er interesserte i å undersøke hva som kan oppnås ved større grad av geografisk differensiering av modellen.

Vi utforsker her muligheten for å basere seg på grunnkretser som geografisk enhet, som et alternativ til dagens kommunebaserte modell. Grunnkretser er en mer stabil enhet enn kommuner. Formålet med å dele kommuner inn i grunnkretser er å lage små, stabile geografiske enheter som kan gi et fleksibelt grunnlag for å arbeide med og presentere regionalstatistikk. Norge er delt inn i ca. 14 000 grunnkretser. For å gruppere grunnkretser til enheter med tilstrekkelig antall observasjoner bruker vi en maskinlæringsalgoritme. Vi har også prøvd ut en enklere variant der prissoner erstattes av sentralitetsnivået til grunnkretser, men denne metoden viser seg å ikke gi tilfredsstillende resultater.

Vi diskuterer også kort andre metoder som kan brukes for å bedre utnytte informasjonen som er tilgjengelig i dagens data, men som krever noe mer ressurser til analyse og utvikling.

Den eksisterende modellen benytter få forklaringsvariabler. I dette prosjektet viser vi hvordan informasjonen fra disse forklaringsvariablene kan utnyttes mer effektivt. Datagrunnlaget fra Finn.no gir også mulighet for å inkludere flere forklaringsvariabler enn de som faktisk brukes, og det er mulig å hente ut ytterligere informasjon fra andre kilder. Et viktig hensyn i så fall er at de samme opplysningene må også foreligge for hele boligmassen når verdiene skal beregnes.¹ Videre var det ved innføring av modellen et uttrykt ønske om at den skulle være enkel og transparent. I hvilken grad denne restriksjonen fortsatt gjelder er usikkert.

1.1. Effekter av regionreformen

Inndelingen i prissoner som brukes i dagens modell er lite robust i forhold til endringer i kommunestruktur. I modellen er landet delt inn i:

- 21 regioner for eneboliger (2019-fylker og storbyer)
- 11 regioner for småhus (storbyer, Akershus og fylkesgrupper)
- 11 regioner for leiligheter (storbyer, Akershus og fylkesgrupper)

Beregningene gjennomføres separat for hver kombinasjon av boligtype og region (beregningsområde). Innenfor de enkelte regionene er kommunene inndelt i soner basert på prisnivået. Soneinndelingen er ulik for hvert beregningsområde. Sone 4 for eneboliger i en region er ikke det samme som sone 4 for leiligheter i samme region osv. Det kjøres en regresjon per boligtype per region med soner som forklaringsvariabler sammen med tettstedsinndeling, byggeår og areal. I modellen er hovedregelen at hele kommunen tilhører én sone, og ingen soner går på tvers av kommunegrensene. Unntaket er de fire største byene, der bydeler kan tilhøre ulike prissone.

Fra og med 1. januar 2020 ble det gjort vesentlige endringer i kommunestrukturen. 112 tidligere kommuner ble berørt av kommunereformen fra dette tidspunktet. Mange av sammenslåingene har bestått i at en sentral kommune med høyere boligprisnivå ble slått sammen med en eller flere omegnskommuner med et lavere prisnivå. Som ved tidligere sammenslåinger har dette stort sett blitt håndtert ved at sonen for den nye kommunen har blitt satt lik sonen for den mest folkerike «gamle» kommunen. Dette fordi de små kommunene i liten grad påvirker gjennomsnittlig pris i den store kommunen. Dette innebærer at boliger i noen randkommuner vil få til dels stor økning i de beregnede boligverdiene selv om boligverdiene reelt sett ikke har økt.

For noen mer jevnstore kommuner ble den nye kommunen plassert midt mellom sonene til den/de gamle kommunen(e). Kommunesammenslåinger kan også påvirke kommuner som ikke har blitt gjenstand for sammenslåing ved at det kan bli færre kommuner i de gjenstående sonene. I noen tilfeller har soner blitt slått sammen som følge av dette.

Når en folkerik kommune slås sammen med omegnskommuner kan dette medføre at mer av prisvariasjonen er innad i soner, og mindre mellom soner. Dette vil isolert sett føre til redusert forklaringskraft i modellen.²

Tabell 1 viser hvordan modellens forklaringskraft for eneboliger har endret seg fra 2019 til 2020. Tabellen viser at forklaringskraften har falt kraftig i regionene der det har vært mange sammenslåinger (for eneboliger er regionene de gamle fylkene). Dette gjelder f.eks. i regioner som Sogn og Fjordane, Vest-Agder, Trøndelag og Møre og Romsdal. Alle disse har hatt mange sammenslåinger mellom folkerike kommuner og omegnskommuner. Sterkest nedgang i

¹ Om informasjon mangler for enkelte observasjoner kan verdien imputeres, men dette bør ikke gjelde en stor andel av observasjonene. Å imputere verdier for alle ikke omsatte boliger basert på omsatte boliger er trolig en dårlig ide.

² Forklaringsgraden eller forklaringskraften er her målt ved R^2 . R^2 ligger mellom 0 og 1. Den forteller oss hvor mye variasjonen i avhengig variabel den estimerte modellen fanger opp. $R^2 = 1$ betyr at all variasjon i avhengig variabel fanges opp av forklaringsvariablene. $R^2 = 0$ betyr at ingen variasjon i avhengig variabel fanges opp av forklaringsvariablene.

forklaringskraften hadde Sogn og Fjordane, hvor den falt med nesten 20 prosent. I Sogn og Fjordane har de to «gamle» kommunene med de høyeste boligprisene, Sogndal og Førde, blitt slått sammen med flere kommuner med lavere prisnivå.

Tabell 1 Forklaringskraft 2019 og 2020 og prosentvis endring i forklaringskraft. Eneboliger

Fylke	2019	2020	Prosentvis endring i forklaringskraft
Østfold	0,570	0,570	-0,02
Akershus	0,754	0,709	-5,94
Oslo	0,713	0,660	-7,40
Hedmark	0,632	0,635	0,46
Oppland	0,587	0,600	2,13
Buskerud	0,621	0,615	-1,00
Vestfold	0,444	0,427	-3,74
Telemark	0,495	0,495	0,10
Aust-Agder	0,403	0,399	-0,97
Vest-Agder	0,516	0,459	-11,09
Rogaland	0,649	0,628	-3,27
Stavanger	0,368	0,338	-8,29
Hordaland uten Bergen	0,542	0,512	-5,50
Bergen	0,533	0,519	-2,70
Sogn og Fjordane	0,521	0,417	-19,98
Møre og Romsdal	0,617	0,563	-8,78
Trøndelag uten Trondheim	0,621	0,564	-9,24
Trondheim	0,518	0,489	-5,73
Nordland	0,626	0,609	-2,72
Troms	0,765	0,758	-0,88
Finnmark	0,592	0,568	-4,17

Modellen beregnes i 21 regioner for eneboliger (2019-fylker og storbyer). Tabell fra Medby og Takle (2021).

Samtidig ser vi at forklaringskraften er tilnærmet uendret i regioner som Hedmark og Oppland som er upåvirket/lite påvirket av kommunereformen. Det er imidlertid ikke fall i forklaringskraften i alle regionene hvor det har vært kommunesammenslåinger. Østfold og Telemark er eksempler på dette. Det kan skyldes at sammenslåingene der skjedde mellom kommuner hvor prisnivået i utgangspunktet ikke varierte veldig sterkt.

Tabell 2 viser hvordan modellens forklaringskraft for småhus har endret seg fra 2019 til 2020. Tabellen viser at forklaringskraften ikke faller like kraftig som den gjør i flere av regionene for eneboliger. Det er heller ikke en like iøynefallende sammenheng mellom fall i forklaringskraft og kommuneendringer som for eneboligenes del. Regionen med det kraftigste fallet, Trondheim, er riktignok påvirket av kommunesammenslåing. I Trondheim falt forklaringskraften med om lag 7 prosent. En forklaring på at kommunesammenslåinger har hatt mindre effekt for denne boligtypen, er at sonene i utgangspunktet var større enn sonene for eneboliger.

Tabell 2 Forklaringskraft 2019 og 2020 og prosentvis endring i forklaringskraft. Småhus

Fylke	2019	2020	Prosentvis endring i forklaringskraft
Oslo	0,792	0,781	-1,44
Akershus	0,765	0,728	-4,86
Østfold, Buskerud, Vestfold og Telemark	0,680	0,657	-3,38
Hedmark og Oppland	0,739	0,733	-0,72
Agder og Rogaland	0,712	0,710	-0,27
Stavanger	0,549	0,537	-2,31
Hordaland, Sogn og Fjordane og Møre og Romsdal	0,638	0,608	-4,72
Bergen	0,562	0,533	-5,18
Trøndelag uten Trondheim	0,724	0,723	-0,07
Trondheim	0,572	0,532	-7,03
Nord-Norge	0,752	0,750	-0,17

Modellen beregnes i 11 regioner for småhus (storbyer, Akershus og 2019 fylker/fylkesgrupper). Tabell fra Medby og Takle (2021).

Tabell 3 viser hvordan modellens forklaringskraft for leiligheter har endret seg fra 2019 til 2020. Tabellen viser at forklaringskraften stort sett faller enda mindre enn den gjør for småhus. Heller ikke for leiligheter er det noen iøynefallende sammenheng mellom fall i forklaringskraft og kommuneendringer. Regionen med det kraftigste fallet, Agder og Rogaland, er riktignok påvirket av kommunesammenslåing, her faller forklaringskraften med 4,65 prosent. Som for småhus er større soner en mulig forklaring på at kommunesammenslåinger har hatt mindre effekt. En annen mulig forklaring er at leiligheter, som ofte er konsentrert i byer og tettsteder, har mindre prisvariasjon mellom sentrum og periferi enn eneboliger.

Tabell 3 Forklaringskraft 2019 og 2020 og prosentvis endring i forklaringskraft. Leiligheter

Fylke	2019	2020	Prosentvis endring i forklaringskraft
Oslo	0,802	0,799	-0,34
Akershus	0,701	0,685	-2,27
Østfold, Buskerud, Vestfold og Telemark	0,607	0,597	-1,63
Hedmark og Oppland	0,698	0,693	-0,67
Agder og Rogaland	0,578	0,551	-4,65
Stavanger	0,611	0,591	-3,19
Hordaland, Sogn og Fjordane og Møre og Romsdal	0,602	0,578	-3,87
Bergen	0,744	0,732	-1,65
Trøndelag uten Trondheim	0,635	0,613	-3,56
Trondheim	0,713	0,688	-3,45
Nord-Norge	0,748	0,744	-0,63

Modellen beregnes i 11 regioner for leiligheter (storbyer, Akershus og 2019-fylker/fylkesgrupper). Tabell fra Medby og Takle (2021).

1.2. Andre svakheter ved modellen

Håndteringen av kommunereformen er ikke den eneste svakheten med dagens boligformuesmodell selv om dette er hovedfokus i forprosjektet. Modellen underestimerer også boligverdiene for dyre boliger i sentrale strøk. Modellen benytter i dag få forklaringsvariabler. Innen hvert beregningsområde tas det hensyn til boligens størrelse målt som P-ROM, og dummyvariabler for prissone (bydel/grupper av bydeler og kommune/grupper av kommuner), boligens alder inndelt i fire intervaller, en syvdelt gruppering av tettsteder etter innbyggertall, samt omsetningsår.

$$\ln(kvmpri) = konst + \beta_1 \ln(p_rom) + \beta_2 prissone_p + \beta_3 ald_g + \beta_4 tskode_s + \beta_5 \text{år}_t + \varepsilon$$

Inkludering av flere forklaringsvariabler har blitt lansert som en mulig kilde til forbedring. Datagrunnlaget fra Finn.no inneholder flere variabler som ikke utnyttes (som for eksempel peis/ildsted og balkong/terrasse), men disse har vi bare for omsatte boliger. Det finnes mye mindre data om boliger som ikke omsettes enn om de som omsettes. **For å bruke modellen for å fastsette boligverdier for alle boliger kreves det at slike data må være tilgjengelige for alle boliger.** Modellen ble da også i sin tid bevisst utformet med få forklaringsvariabler for å minimere innhenting av data fra skatteytterne.

Når det gjelder mange av forklaringsvariablene fra finn.no er det vanskelig å se for seg muligheter for å avlede tilsvarende variabler for den totale boligmassen. Bruk av geodata er imidlertid lettere fordi innhenting kan baseres på koordinater som er tilgjengelige for (tilnærmet) alle boliger.

Problemet med underestimering av boligverdier i Oslo ble nylig analysert i en masteroppgave (Vinje, 2020). Her drøftes det hva som kan oppnås ved en annen modell for boligpriser. I sammendraget konkluderes det med at «Resultatene viser at det er mulig å forbedre SSB sin boligprismodell betraktelig uten å kreve at brukerne av modellen må innhente mer informasjon om boligene enn i dag».

Maskinlæringsmodeller skisseres i oppgaven som mulig løsning. I oppgaven vises det at bruk av slike modeller kan gi økt forklaringskraft, men noe underestimering av dyre boliger var det fortsatt. Oppgaven fokuserte bare på Oslo og det er uvisst om resultatene kan generaliseres til resten av landet. Oppgaven berørte naturlig nok heller ikke virkninger av kommunereformen som er

hovedfokus i forprosjektet. I avsnitt 3.3 viser vi at samme type forbedring av verdsetting for dyre boliger kan oppnås ved hjelp av tradisjonelle hedoniske regresjonsmodeller i kombinasjon med en maskinlæringsmetode for å effektivt utnytte lokal geografisk informasjon.

Et annet problem når det gjelder underestimering av verdien på dyre boliger er at undersøkelser tyder på at de aller dyreste boligene er sterkt underrepresenterte i datagrunnlaget fra finn.no, noe som ikke vil kunne avhjelpest ved en bedre modell (Medby og Takle, 2020). Underrepresentasjonen er for øvrig enda større for de aller billigste boligene.

Dette forprosjektet ble startet opp i juni 2021. Hurdalsplattformen inneholder følgende punkt: *Metoden for verdsetting av dyre eiendommer gjennom SSBs sjablongmodell må oppdateres for å reflektere eiendommenes reelle verdi. Øvrige eventuelle endringer i eiendomsskatten må utredes i samarbeid med kommunene.* Fullstendig revisjon av modellen er ikke bestilt på nåværende tidspunkt, og problemet med lave estimerte boligverdier i sentrale strøk omtales ikke spesifikt i oppdragsbeskrivelsen. Ut over det som oppnås ved en mer fingradert inndeling vil vi ikke behandle underestimering av boligverdier i sentrale strøk i dette forprosjektet.

2. Modell tilpasset endringer i kommunestruktur

Som vist i forrige avsnitt, er dagens modell lite robust med hensyn til kommuneendringer. Prosjektet har som mål å undersøke hvordan verdsettelsen av boliger i større grad kan frigjøre seg fra kommunegrensene.

En metode er å undersøke hvordan prissonene kan fastsettes på en mer fleksibel måte, ved å basere seg på koordinatene til de omsatte boligene. I teorien kan boligverdier beregnes ved hjelp av boligadressen til hver enkelt bolig, hvor hver adresse har et sett koordinater. Dette er en metode som omtales i Ahlfeldt et al. (2021). Denne typen metode utnytter tilgjengelig informasjon på en svært fleksibel måte, men utvikling krever mer ressurser enn tilgjengelig i dette prosjektet.

I dette notatet utforsker vi derfor muligheten for å knytte sammen små områder (grunnkretser) ved hjelp av maskinlæringsalgoritmer for å optimalisere avveiningen mellom å bruke geografisk homogene eiendommer, og å få nok observasjoner. Dette omtales nærmere i avsnitt 2.1.

En annen, enklere metode er å basere seg på en kombinasjon av sentralitet og grunnkretser som erstatning for dagens kommunebaserte soner. Dvs. å bruke grunnkretser og en fingradert sentralitetsindeks på grunnkretsnivå som erstatning for soneinndelingen i dagens modell. Vi utforsker også denne muligheten, omtalt i avsnitt 2.2. I avsnitt 2.3 diskuteres flere mulige metoder.

2.1. Maskinlæring på grunnkretser

Ved å frigjøre seg fra kommunegrensene får man muligheten til å skape geografiske enheter som er mer fleksible. I tillegg vil endringer av kommunegrensene slå mindre direkte ut i boligverdsettning. Her har vi valgt å bruke grunnkretser som geografiske grunnenheter i stedet for kommuner. Formålet med grunnkretser beskrives som:³

Formålet med å dele kommuner inn i grunnkretser er å ha små stabile geografiske enheter som kan gi et fleksibelt grunnlag for å arbeide med og presentere regionalstatistikk. Grunnkretser består av et geografisk sammenhengende område som er mest mulig ensartet når det gjelder natur og næringsgrunnlag, kommunikasjonsforhold og bygningsmessig struktur.

Fordelen ved å bruke grunnkretser fremfor for eksempel et rutenett er at grunnkretser er laget for å dekke homogene områder. De følger i stor grad naturlige geografiske grenser som elver, motorveier, endringer i boligstruktur o.l. Ulempen ved grunnkretser er at de ikke er konstante (se grundigere omtale i avsnitt 4).

Grunnkretser er små enheter (det er ca. 14 000 grunnkretser i Norge). En enkelt grunnkrets mangler derfor ofte antallet boligtransaksjoner som er nødvendig for å estimere en regresjonsmodell, selv en med få forklaringsvariabler. I vår situasjon vil det særlig være mange landlige grunnkretser som har få observasjoner av leiligheter og rekkehus. Derfor er man avhengig av å slå sammen flere grunnkretser til større enheter.⁴

Vi bruker en enkel maskinlæringsalgoritme kalt beslutningstre (Decision tree) for å sette sammen grunnkretser som ligner på hverandre. At grunnkretser ligner baseres på geografisk plassering (gjennomsnitt for omsetninger innad i grunnkretsen), kommune og fylke. Den avhengige variabelen er kvadratmeterpris. Selv om modellen bare testes ut for tre regioner brukes observasjoner for hele

³ <https://www.ssb.no/a/metadata/conceptvariable/vardok/135/nb>.

⁴ Selv kommuner kan ha for få transaksjoner for å estimere en modell; i den eksisterende verdsettelsesmodellen slås mange kommuner sammen i grupper.

landet. Det vil si at vi tillater grunnkretser i de aktuelle regionene å tilhøre grupperinger med grunnkretser i andre regioner. Dette vil gi bedre resultater enn ved å begrense gruppene kun til de tre regionene. Grupperingen gjøres separat for forskjellige boligtyper (enebolig, småhus og leiligheter).

For hver enkelt gruppe gjøres så en hedonisk regresjon, med hovedsakelig de samme variablene som i den eksisterende modellen. Det vil si at man har en separat regresjon for hvert enkelt geografisk område og bygningstype. Sonevariablene og tettstedsinndelingen er overflødige i dette rammeverket, men vi legger til dummy-variabler for hver enkelt grunnkrets. Totalt sett blir det i overkant av 1 200 regresjoner for hele landet med formen:

$$\ln(kvmpri) = konst + \beta_1 \ln(p_{rom}) + \beta_2 ald_g + \beta_3 \text{år}_t + \beta_4 grkrets_t + \varepsilon$$

Å ha mange separate regresjoner gir større fleksibilitet fordi alle parameterverdiene i modellen kan være forskjellige for hver kombinasjon av bygningstype og gruppering. I den eksisterende modellen begrenses sammenhengen mellom f.eks. kvadratmeterpris og areal til å være lik for alle boliger av en boligtype i en hel region.

Den grunnleggende avveiningen ved aggregering av grunnkretser er at man ønsker å skape områder som er så homogene som mulig (noe som tilsier små enheter) samtidig som man er avhengig av et visst antall omsetninger i den hedoniske regresjonen (noe som tilsier større enheter). Vi krever at det er minimum 500 omsetninger i en gruppe av grunnkretser, for å få et tilfredsstillende antall observasjoner for den hedoniske regresjonen. Det resulterer i mange, geografisk små grupper i sentrale strøk, og geografisk større grupper i mindre sentrale strøk. Slik utnytter man den store mengden informasjon (antall transaksjoner) i byer til å bedre predikere lokale prisnivåer.

2.2. Sentralitetsindeks på grunnkrets nivå

SSB utarbeider en sentralitetsklassifisering av kommunene. Fram til og med 2017 var klassifiseringen basert på kommunens beliggenhet i forhold til tettsteder og disse tettstedenes størrelse.⁵ Høyest sentralitet hadde kommunene med «landsdelscenter»-funksjoner og kommunene rundt disse. Sentraliteten hadde da fire hovednivåer, mellom 0 og 3. Fra og med 2018 ble det innført en ny standard for sentralitet, basert på nærhet til arbeidsplasser og servicefunksjoner, uten bruk av tettsteder i klassifiseringen (Høydahl, 2017). Denne er beregnet som en indeks, der sentralitetsverdien ligger mellom 0 (kun teoretisk mulig) og 1 000. Disse verdiene er gruppert sammen til seks klasser. Den mest sentrale kommunen vil alltid ha verdi 1 000 (Oslo) og de minst sentrale kommunene har verdier ned mot 300.

Selv om sentralitetsindeksen opprinnelig knyttes til kommunenivå, er grunnkrets den minste byggesteinen underveis i produksjonsprosessen. Dette gjør at man i prinsippet også vil ha en indeks per grunnkrets. Pt er indeksen på grunnkrets nivå ikke publisert og altså ikke «offisiell statistikk». Men muligens kan indeksen på grunnkrets nivå gjøres tilgjengelig for alle. Det er ennå ikke besluttet hvor ofte sentralitetsindeksen skal oppdateres. Indeksen ble introdusert i 2017 og deretter oppdatert i 2020, i etterkant av regionreformen med mange nye kommuner. I dette forprosjektet er sentralitetsindeksen per 2020 brukt. Det er i dag noen grunnkretser som har sentralitet '0', for eksempel der det ikke finnes veiforbindelse og kretser uten bosetting. Figur B1 i vedlegg B viser den fylkesvise fordelingen av grunnkretser etter sentralitet. Oslo har en sentralitet som ligger rundt 1000, med noen få unntak. Troms og Finnmark har derimot en gjennomsnittlig sentralitet på rundt

⁵ <https://www.ssb.no/klasse/klassifikasjoner/128/versjon/468/koder>.

500 med relativt stor spredning. I både Hedmark og Møre og Romsdal har mediangrunnkretsen en sentralitetsindeks på mellom 600 og 700.

En enkel metode for å løsrive seg fra kommunegrenser er å erstatte sonene knyttet til kommunegrenser i dagens modell med sentralitetsindeksen på grunnkrets nivå. Rent konkret beholder vi regionene fra dagens modell (21 for eneboliger, 11 for småhus og leiligheter), og alle de andre variablene i den hedoniske regresjonene, men erstatter sone med sentraliteten til grunnkretsen en eiendom ligger i. En ulempe her er at vi antar en lineær funksjonsform for sentralitet innad i hver region. Denne metoden viser seg å ikke gi resultater som forbedrer dagens metode i særlig grad, men den vil være robust med hensyn til endringer i kommunestrukturen. Resultatene omtales i vedlegg B.

2.3. Alternative metoder

En måte å helt løsrive seg fra de administrative inndelingene på kan være å dele landet i kvadrater av lik størrelse. SSB publiserer f.eks. årlig noen typer statistikk for ruteinndelingene 250m*250m, 1km*1km og 5km*5km. Man kunne tenkt seg at rutene utgjorde den minste byggesteinen i modellen. Man må fortsatt finne en metode for å slå sammen ruter, da noen ruter vil ha for få observasjoner for å estimere priser. Denne metoden unngår problemer som kan oppstå ved at kommuner eller grunnkretser endres. En ulempe med denne metoden er imidlertid at rutene kan dele opp områder som naturlig skulle hørt sammen. Dersom et nabolag tilfeldigvis blir splittet at rutegrensene, kan det oppstå et tilfeldig skille mellom boligverdiene på hver side av grensen.

Det går an å tenke seg metoder som løsriver seg enda mer fra administrative enheter. Hill og Scholz (2018) kombinerer en «spline surface» med en parametrisk hedonisk modell for Sydney.⁶ Den parametriske delen tar inn over seg fysiske attributter ved boligen slik som i vår modell (størrelse, alder osv.), mens den ikke-parametriske delen tar inn over seg betydningen av beliggenhet. De kan dermed konstruere en topografisk flate som beskriver hvordan prisene varierer med de geografiske koordinatene (lengdegrad og breddegrad), gitt de fysiske attributtene. Dermed sammenlignes denne modellen med en enklere modell hvor spline-funksjonen erstattes med dummyvariabler for postnummer og region. Det er interessant å merke seg at ved å bruke postnummer får de om lag samme treffsikkerhet som ved bruk av en spline-funksjon. Årsaken er at postnummer er så finmasket at det er lite å vinne på andre typer metoder for å ta hensyn til boligens beliggenhet. Dette fordrer derimot at det er nok observasjoner innenfor hvert postnummer, noe som ikke er et problem i tilfellet med en storby som Sydney.

Ahlfeldt et al. (2021) bruker en lignende tilnærming med en kombinasjon av parametrisk, hedonisk regresjon og ikke-parametrisk romlig regresjon (spatial regression). Her kan man teoretisk beregne verdien separat for hver enkelt eiendom. Algoritmen de har utviklet bruker i stor grad salg i nærområdet der det er mange tilgjengelige observasjoner, og i større grad en parametrisk regresjon der det er få nærliggende observasjoner. Dermed utnyttes tilgjengelig informasjon på best mulig måte: Der det er mange lignende observasjoner i data settes verdi basert på verdien av boliger helt i nærheten, noe som fanger opp lokal variasjon i priser, men modellen har også en struktur som prissetter boliger som ikke har noe direkte sammenligningsgrunnlag i data.

Metodene vi har omtalt til nå baserer seg alle på å benytte informasjon om salg av nærliggende boliger på en bedre måte enn eksisterende modell. Alle de nevnte modellene vil kunne brukes med data som er tilgjengelig i dag (gitt at data for x- og y-koordinater er utfyllende).

⁶ Man kan tenke seg en «spline surface» som høydekurver som ligger over et kart, og beskriver prisnivået for en «normal» bolig på et gitt punkt.

En annen retning er å bruke et større antall forklaringsvariabler. En moderne tilnærming her er å inkludere et stort antall mulige variabler og interaksjonsledd, og la en maskinlæringsmodell bestemme formen på modellen (se Mullainathan og Spiess, 2017 for en grundigere gjennomgang). Dette vil være en tilnærming som ligner den nye modellen for verdsetting av fritidsboliger (PWC, 2020). Ulempen med en slik modell er at datatilgang kan være utfordrende, da den krever kjennskap til forklaringsvariabler for alle boliger, ikke bare omsatte boliger. Det vil trolig bety en betydelig kostnad for datainnsamling og behandling. Det er videre uklart hvor stor gevinst som kan oppnås ved å inkludere en mengde lokasjonsvariable utover en optimalisering av bruken av geografisk plassering dersom man har godt med omsetningsdata. PWC (2020) finner tydelig bedre forklaringskraft ved en maskinlæringsmodell enn ved en lineær modell for fritidsboliger, men utforsket dessverre ikke en lineær modell med mer enn en region.

3. Resultater

Her viser vi resultater for den ny-utviklede modellen beskrevet i avsnitt 2.2 sammenlignet med den eksisterende modellen. Vi bruker det samme datagrunnlag som i den eksisterende modellen, ca 800 000 boliger omsatt gjennom Finn.no over en periode på 10 år. Først viser vi resultater for år 2019 (dvs. at vi bruker data for perioden 2010 – 2019). Å vise modellene for 2019 er best egnet for sammenligning, da det gir en bedre illustrasjon av forbedringspotensialet ved den nye metodikken. Da sammenligner man den nye modellen med en "fungerende" versjon av den eksisterende modellen. Vi vil også vise resultater fra den enklere oppdateringen av modellen hvor soner erstattes med en sentralitetsindeks. Senere viser vi også resultater for 2020, og vi ser nærmere på hvordan modellene fungerer for dyre boliger i Oslo.

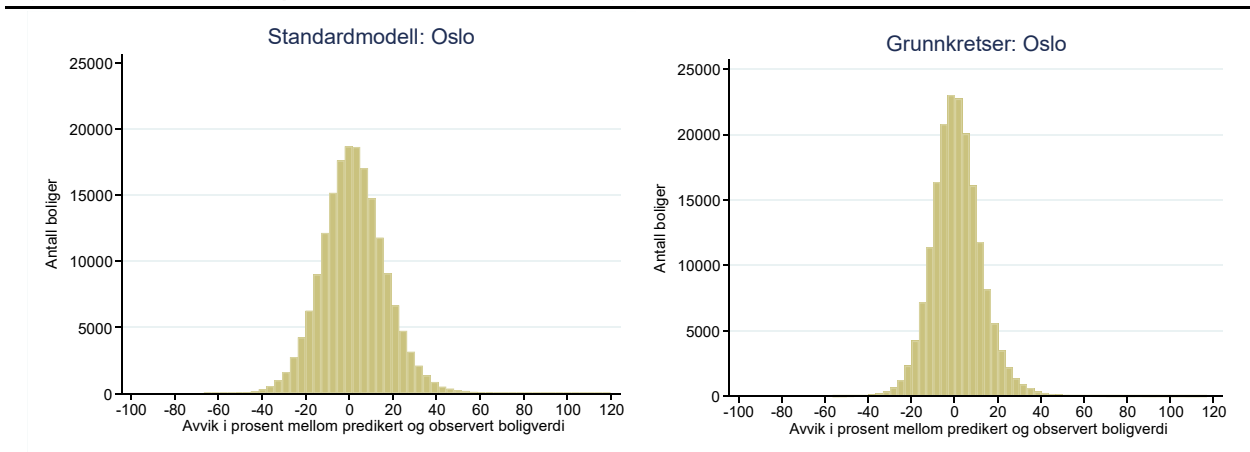
3.1. 2019

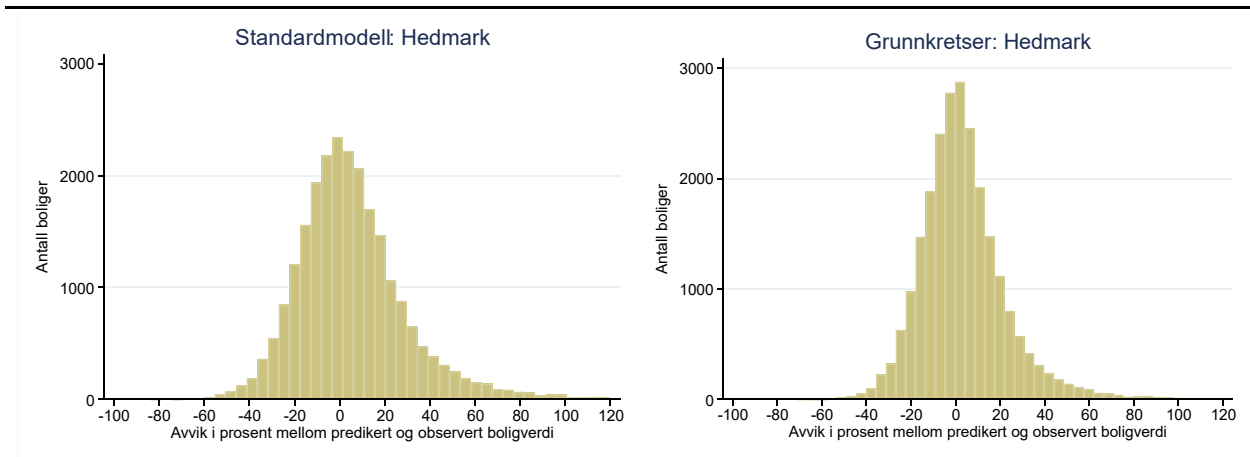
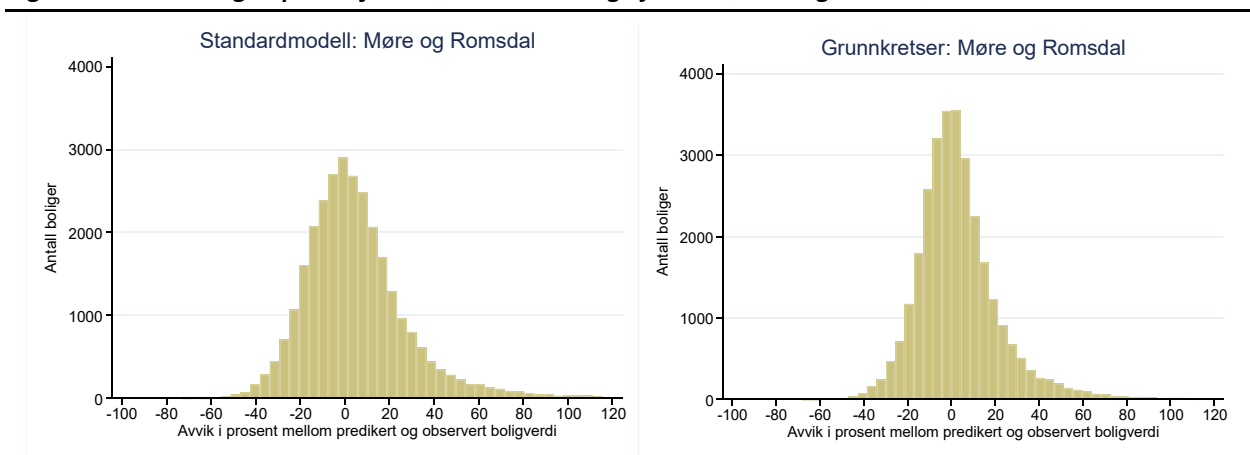
Ved grupperingen av grunnkretser som beskrevet i avsnitt 2.1 dannes totalt 1 263 grupperinger i Norge, hvorav 341 i de aktuelle regionene (henholdsvis 243, 48 og 50 i Oslo, Hedmark og Møre og Romsdal). Grupperingene av grunnkretser for de tre fylkene vises i vedlegg A, figur A1.

Deretter kjøres 341 separate hedoniske regresjoner for å predikere priser for boliger i hver gruppe i de tre regionene. Predikerte priser fra modellen sammenlignes så med transaksjonsprisene fra Finn-datasettet.

Figur 1 – 3 viser fordelingen av prediksjonsfeil for den eksisterende modellen, og for modellen basert på gruppering av grunnkretser for henholdsvis Oslo, Hedmark og Møre og Romsdal. Det er tydelig fra figurene at standardavviket for prediksjonsfeil er lavere ved grunnkretsmodellen, for alle tre fylker.

Figur 1 Fordeling av prediksjonsfeil, eksisterende og ny modell, Oslo, 2019



Figur 2 Fordeling av prediksjonsfeil, eksisterende og ny modell, Hedmark, 2019**Figur 3** Fordeling av prediksjonsfeil, eksisterende og ny modell, Møre og Romsdal, 2019

I Tabell 4 viser vi oppsummerte tall for treffsikkerheten til den eksisterende og den nye modellen. For alle tre regioner, og alle fire mål viser modellen basert på grunnkretser mindre feil. Standardavviket for feilen blir redusert med cirka 20 prosent i alle regionene. R^2 stiger med 5,3 prosent for Oslo, og 4,3 prosent for Hedmark og Møre og Romsdal.

Tabell 4 Gjennomsnitt, median og standardavvik for prosentvis feil, og R^2 , 2019

Modell	Oslo		Hedmark		Møre og Romsdal	
	Eks.	Ny	Eks.	Ny	Eks.	Ny
Gjennomsnitt	2,43	1,53	5,75	3,69	5,46	3,57
Median	1,72	0,71	2,53	1,44	1,94	0,95
Standardavvik	17,3	13,8	25,7	20,3	25,2	20,3
R^2	0,805	0,848	0,751	0,783	0,746	0,778

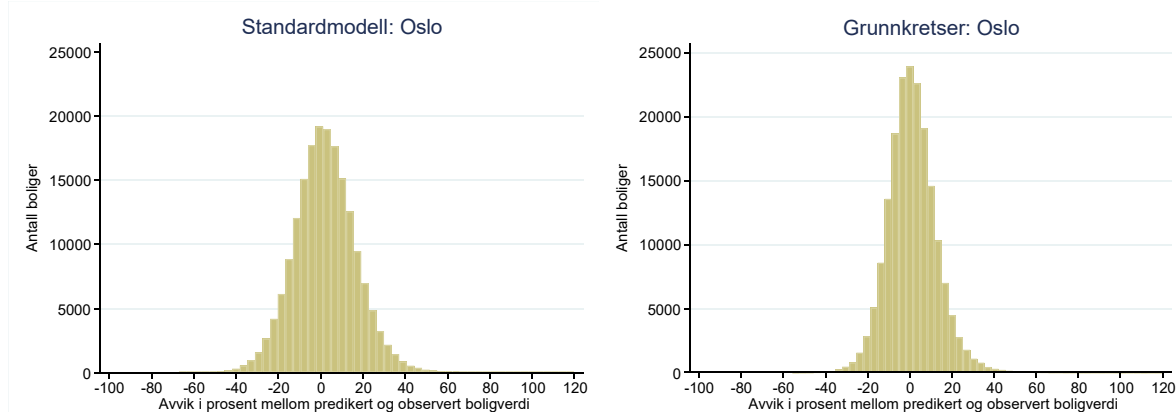
Figurer som viser fordelingen av absolutte avvik i prediksjonene finnes i Vedlegg A, figur A5 – A7.

3.2. 2020

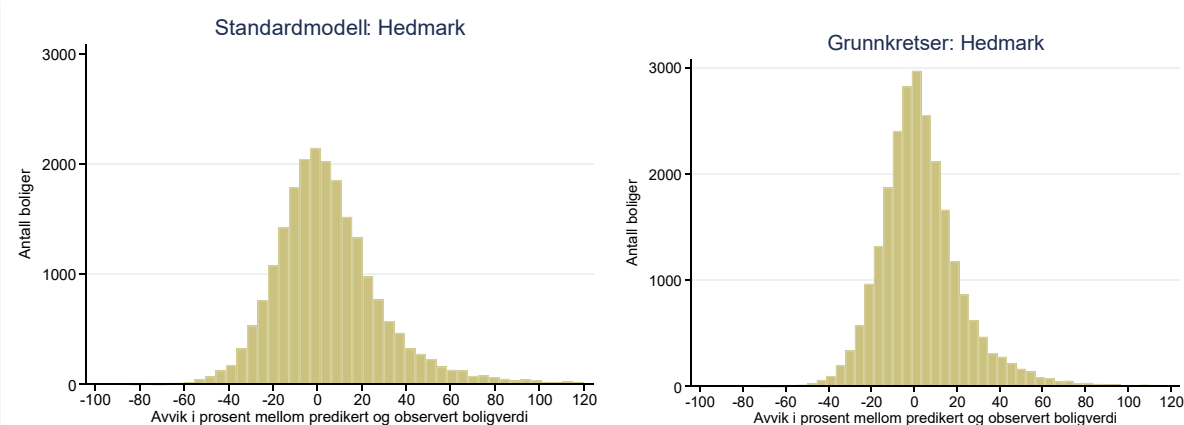
Her viser vi tilsvarende resultater som i avsnitt 3.1 for år 2020 (dvs. data fra 2011 – 2020). Figur 4 – 6 viser fordelingen av standardfeil for den eksisterende og den nye modellen. Hvis man sammenligner resultatene for den eksisterende modellen er det tydelig at modellen treffer dårligere i 2020. Det gjelder særlig for Møre og Romsdal, hvor flere kommuner har blitt slått sammen, men også for Oslo.

For den nye modellen dannes totalt 1 208 grupperinger i Norge, hvorav 342 i de aktuelle regionene (henholdsvis 242, 55 og 45 i Oslo, Hedmark og Møre og Romsdal).

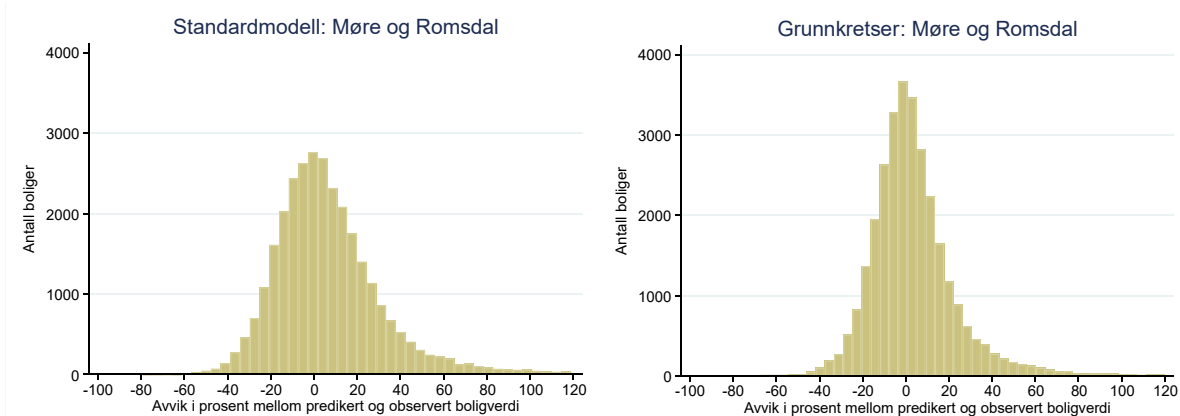
Figur 4 Fordeling av prediksjonsfeil, eksisterende og ny modell, Oslo, 2020



Figur 5 Fordeling av prediksjonsfeil, eksisterende og ny modell, Hedmark, 2020



Figur 6 Fordeling av prediksjonsfeil, eksisterende og ny modell, Møre og Romsdal, 2020



I Tabell 5 viser vi oppsummerte tall for treffsikkerheten til den eksisterende og den nye modellen. Også i 2020 gir modellen basert på grunnkretser mindre feil for alle tre regioner og alle fire mål. Standardavviket for feilen blir redusert med cirka 20 prosent i Oslo og Hedmark, og med nesten 25 prosent i Møre og Romsdal. R^2 forbedres med henholdsvis 5,9, 3,8 og 7,5 prosent.

Figurer som viser fordelingen av absolutte avvik i prediksjonene finnes i Vedlegg A, figur A8 – A10.

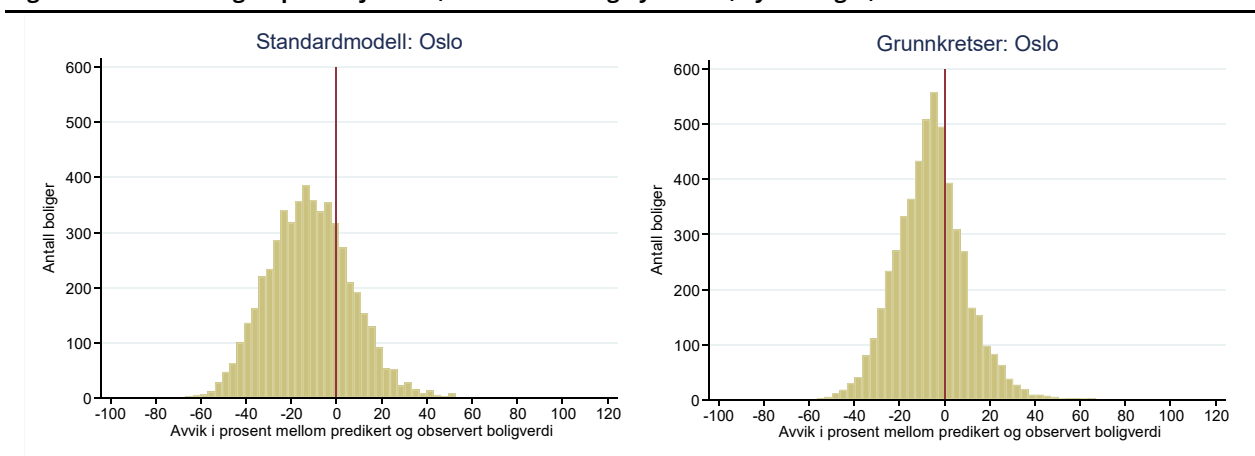
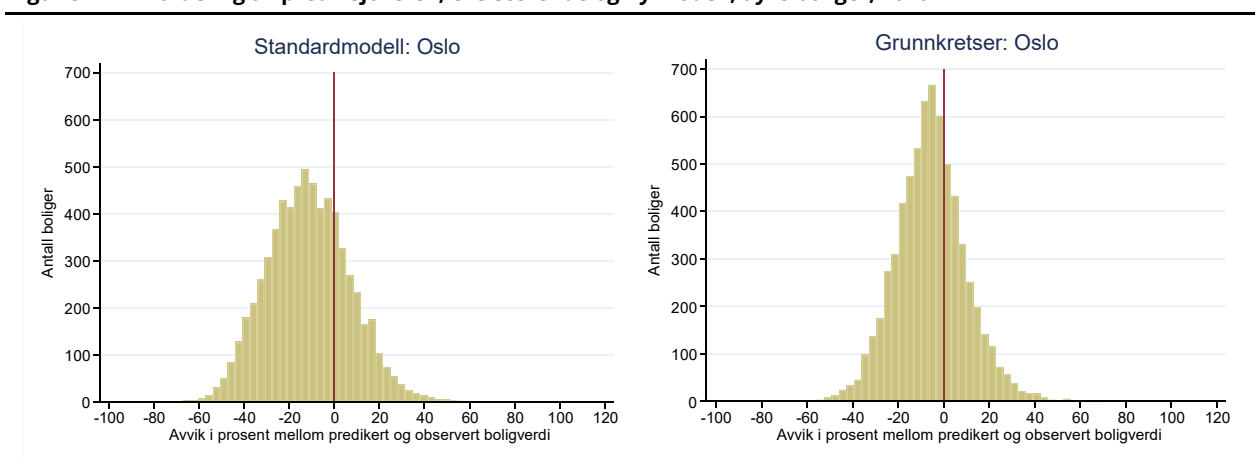
Tabell 5 Gjennomsnitt, median og standardavvik for prosentvis feil, og R², 2020

Modell	Oslo		Hedmark		Møre og Romsdal	
	Eks.	Ny	Eks.	Ny	Eks.	Ny
Gjennomsnitt	2,40	1,49	5,55	3,74	7,32	3,68
Median	1,75	0,69	2,25	1,34	3,17	0,95
Standardavvik	17,2	13,6	26,0	20,6	27,3	20,7
R ²	0,802	0,849	0,757	0,786	0,721	0,775

3.3. Dyre boliger

Det har vært en del oppmerksomhet i den senere tid om at den eksisterende modellen for boligverdsetting underestimerer verdien på dyre boliger. Derfor undersøker vi her om modellen med grunnkretsgrupper forbedrer prediksjon av priser for dyre boliger i Oslo. Dyre boliger definerer vi som boliger som selges for over 10 millioner (inkludert fellesgjeld).

Figur 7 og 8, og tabell 6 viser at den nye modellen gir en kraftig forbedring i prediksjonen for dyre boliger. Gjennomsnittlig og median prosentvis feil blir omtrent halvert, mens standardavviket reduseres med rundt 15 prosent. R² øker med nesten 40 prosent, fra et veldig lavt nivå.

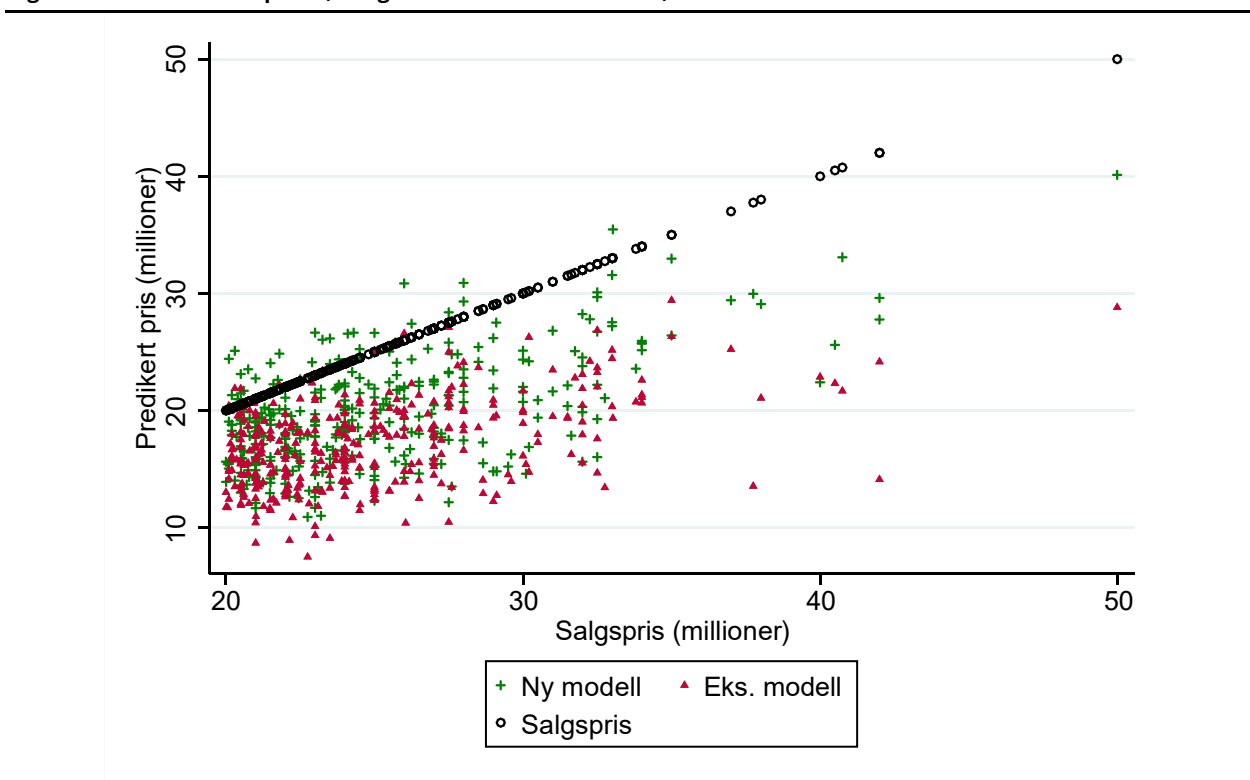
Figur 7 Fordeling av prediksjonsfeil, eksisterende og ny modell, dyre boliger, 2019**Figur 8 Fordeling av prediksjonsfeil, eksisterende og ny modell, dyre boliger, 2020**

Tabell 6 Gjennomsnitt, median og standardavvik for prosentvis feil, og R², dyre boliger i Oslo

Modell	2019		2020	
	Eks.	Ny	Eks.	Ny
Gjennomsnitt	-11,75	-6,03	-11,41	-5,57
Median	-12,07	-6,07	-11,87	-5,85
Standardavvik	18,1	15,4	18,2	15,4
R ²	0,491	0,673	0,485	0,676

Som en illustrasjon på økt treffsikkerhet sammenligner vi i figur 9 transaksjonsprisene for boliger solgt over 20 millioner i 2020-datasettet (369 observasjoner) med predikert pris fra eksisterende og ny modell. De svarte sirkelene er faktisk pris, og danner en 45°-linje. Punkter under denne linjen har for lav predikert pris, punkter over for høy. Det er tydelig hvordan de grønne kryssene med prediksjoner fra den nye modellen ligger nærmere de reelle markedsverdiene enn prediksjonene fra den eksisterende modellen.

Figurer som viser fordelingen av absolutte avvik i prediksjonene finnes i Vedlegg A, figur A11 – A12.

Figur 9 Predikerte priser, boliger over 20 millioner i Oslo, 2020

Noen av de dyreste boligene som selges blir ikke annonsert via Finn. Det er vanskelig å vurdere hvordan den nye modellen vil treffe for disse, men siden den tillater mer variasjon i prisene på forskjellige nabolag vil den trolig også her gi noe bedre verdsetting.

4. Oppsummering og gjenstående utfordringer

Et stort antall kommuner ble slått sammen i 2020. Det har medført dårligere prediksjoner fra den eksisterende boligverdsettingsmodellen. Hovedmålet ved dette notatet er å undersøke hvordan man kan basere seg på mindre, og mer robuste geografiske enheter enn kommuner.

Dagens modell benytter få forklaringsvariabler. Datagrunnlaget gir mulighet for å inkludere flere forklaringsvariabler enn de som faktisk brukes, men som nevnt må de samme opplysningene også foreligge for hele boligmassen når verdiene skal beregnes. Videre var det ved innføring av modellen, et uttrykt ønske om at den skulle være enkel og transparent. I hvilken grad denne restriksjonen fortsatt gjelder vil i noen grad styre fremtidig valg av modell.

Dette notatet gir en beskrivelse av en maskinlæringsmetode som bruker grunnkretser som geografisk grunnenhet ved vurdering av boligverdier. Nærliggende grunnkretser aggregeres til grupper med lignende prisnivå. Størrelsen på gruppene bestemmes av at det kreves et minimum antall transaksjoner i hver gruppe. Grupperingen gjøres separat for forskjellige bygningstyper. Det resulterer i mange, geografisk små grupper i sentrale strøk, og geografisk større grupper i mindre sentrale strøk. En hedonisk regresjon, lignende den eksisterende modellen, kjøres så separat for hver enkelt grunnkretsgruppe. Slik utnytter man den store mengden informasjon (antall transaksjoner) i byer til å bedre predikere lokale prisnivåer.

Resultatene, beskrevet i avsnitt 3, er lovende, og viser tydelig bedre treffsikkerhet enn den eksisterende modellen, for flere forskjellige mål og for alle tre fylker inkludert i prosjektet. Treffsikkerheten er kraftig forbedret også når man sammenligner med den eksisterende boligverdsettingsmodellen i 2019, før kommunesammenslåingsproblematikken. Resultatene er svært like for 2019 og 2020, noe som tyder på at sammenslåing av kommuner har liten effekt på modellen, og at den er relativt stabil over tid. Modellen gir også en kraftig forbedring av prediksjonene for dyre boliger i Oslo, selv om de fortsatt verdsettes noe lavt.

En alternativ modell, som erstatter de kommunebaserte prissonene i modellen med sentralitetsindeksen på grunnkretsnivå, er svært enkel å implementere. Selv om den frigjør seg fra kommunestrukturen er resultatene ikke så lovende, da den kun gir en svak forbedring over den eksisterende modellen for Møre og Romsdal, et fylke med svært mange kommunesammenslåinger, i 2020. Sentralitetsindeksmodellen gir gjennomgående dårligere resultater enn den eksisterende modellen i 2019. Modellen er heller ikke egnet for store byer, da de har liten variasjon i sentralitet, den vil dermed kreve at det etableres en alternativ modell for byregioner.

4.1. Gjenstående utfordringer

Den nye modellen med grunnkretsgrupper presentert i dette notatet gir markert bedre resultater enn den eksisterende verdsettingsmodellen. Å bruke modellen for produksjon av boligverdier vil imidlertid kreve at man tar stilling til noen problemstillinger som vi ikke har behandlet i dette notatet.

Vi har valgt å bruke minimum 500 observasjoner per gruppering av grunnkretser. Det tallet er noe tilfeldig. Analyser av 2019-data med minimum 1000 og 250 observasjoner per gruppe viser kun mindre endringer i resultater (marginalt dårligere for alle treffsikkerhetsmål for 1000 observasjoner per gruppe, marginalt bedre for de fleste mål for 250 observasjoner). Avveiningen mellom økt fleksibilitet (ved flere grupperinger) og større risiko for påvirkning fra ekstreme enkeltobservasjoner ved bruk av mindre grupper kan undersøkes grundigere.

Selv om grunnkretser er ment å være stabile over tid hender det at grunnkretser splittes eller endres. De fleste endringene i grunnkretsstrukturen er oppdeling i mindre kretser, noe som er mest

vanlig i utbyggingsstrøk. I Ullensaker kommune ble for eksempel 7 grunnkretser delt i to eller 3 nye grunnkretser mellom 2017 og 2018. Endringer i kommunegrenser vil føre til at grunnkretsene får nytt nummer (de 4 første sifrene i grunnkretsnummeret er kommunenummer), men selve grunnkretsen består som regel.

I dette notatet har vi brukt grunnkretsdata fra 2018.⁷ Noen boligtransaksjoner er ikke koblet til en av 2018-grunnkretsene, enten fordi boligen lå i en annen grunnkrets på tidspunktet for boligtransaksjon eller pga. manglende utfylling. Disse boligtransaksjonene er plassert i en 2018-grunnkrets ved hjelp av geokoordinater. En eventuell modell basert på grunnkretser må inneholde en rutine for å håndtere endringer i grunnkretser.

Datasettene som brukes til beregningene dekker boligtransaksjoner over ti år. En del grunnkretser har ikke noen transaksjoner over disse periodene. Det kan skyldes at det ikke finnes boliger i grunnkretsen (grunnkretser uten transaksjoner i Oslo er f.eks. Bjerkebanen, Sjursøya og diverse grunnkretser i sentrum og Marka), men det kan også være grunnkretser der det finnes et fåtall boliger som omsettes sjelden. Særlig for boligtypene småhus og leiligheter finnes det et større antall landlige grunnkretser uten observerte transaksjoner (se figur A2 i vedlegg A). Alle grunnkretser bør tildeles en sone. Mulige måter å plassere grunnkretser uten transaksjoner er til den sonen som har lengst grense til grunnkretsen, eller til den tilstøtende sonen med lavest predikert pris.

Det må også vurderes hvorvidt grunnkretsgruppene skal oppdateres fra år til år, eller med for eksempel 5- eller 10-årsintervaller, og hvilke konsekvenser det får med hensyn til stabilitet i ligningsverdier over tid. Vi har gjort en enkel beregning på korrelasjonen mellom predikerte verdier for eiendommer i datasettet som predikeres i både 2019- og 2020-modellen (transaksjoner 2011 – 2019).

Tabell 7 Korrelasjon mellom predikerte verdier 2019 og 2020

Modell	Oslo		Hedmark		Møre og Romsdal	
	Eks.	Ny	Eks.	Ny	Eks.	Ny
Korrelasjon	0,9999	0,9981	0,9989	0,9947	0,9819	0,9939

Resultatene vises i tabell 7. Korrelasjonene mellom predikerte priser er noe lavere for den nye modellen enn for den eksisterende i fylkene uten endring i kommunegrenser, men tydelig høyere for Møre og Romsdal som hadde mange kommunesammenslåinger. Selv om stabiliteten bør undersøkes grundigere er disse resultatene lovende.

⁷ Versjonid 1.5_20180115

Referanser

Ahlfeldt, Gabriel M., Stephan Heblich og Tobias Seidel (2021): «Micro-geographic property price and rent indices», upublisert notat, datert 07.07.2021:
[https://personal.lse.ac.uk/ahlfeldg/WP/GA_SH_TS - HPI.pdf](https://personal.lse.ac.uk/ahlfeldg/WP/GA_SH_TS_-_HPI.pdf)

Finansdepartementet (2021): «Høringsnotat – nytt verdsettelsessystem for fritidsboliger i formuesskatten», Saksnr. 21/3144, datert 07.10.2021.

Hill, Robert J. og Michael Scholz (2018): «Can geospatial data improve house price indexes? A hedonic imputation approach with splines», *Review of Income and Wealth* 64 (4): 737-756.

Høydahl, Even (2017): «Ny sentralitetsindeks for kommunene», Statistisk sentralbyrå, Notat 2017/40.

Medby, Per og Mona Takle (2020): «Modell for beregning av boligformue: Oppdatert med tall for 2019», Statistisk sentralbyrå, Notat 2020/9.

Medby, Per og Mona Takle (2021): «Modell for beregning av boligformue: Oppdatert med tall for 2020», Statistisk sentralbyrå, Notat 2021/10.

Mullainathan, Sendhil og Jann Spiess (2017): «Machine learning: an applied econometric approach», *Journal of Economic Perspectives*, 31 (2): 87-106.

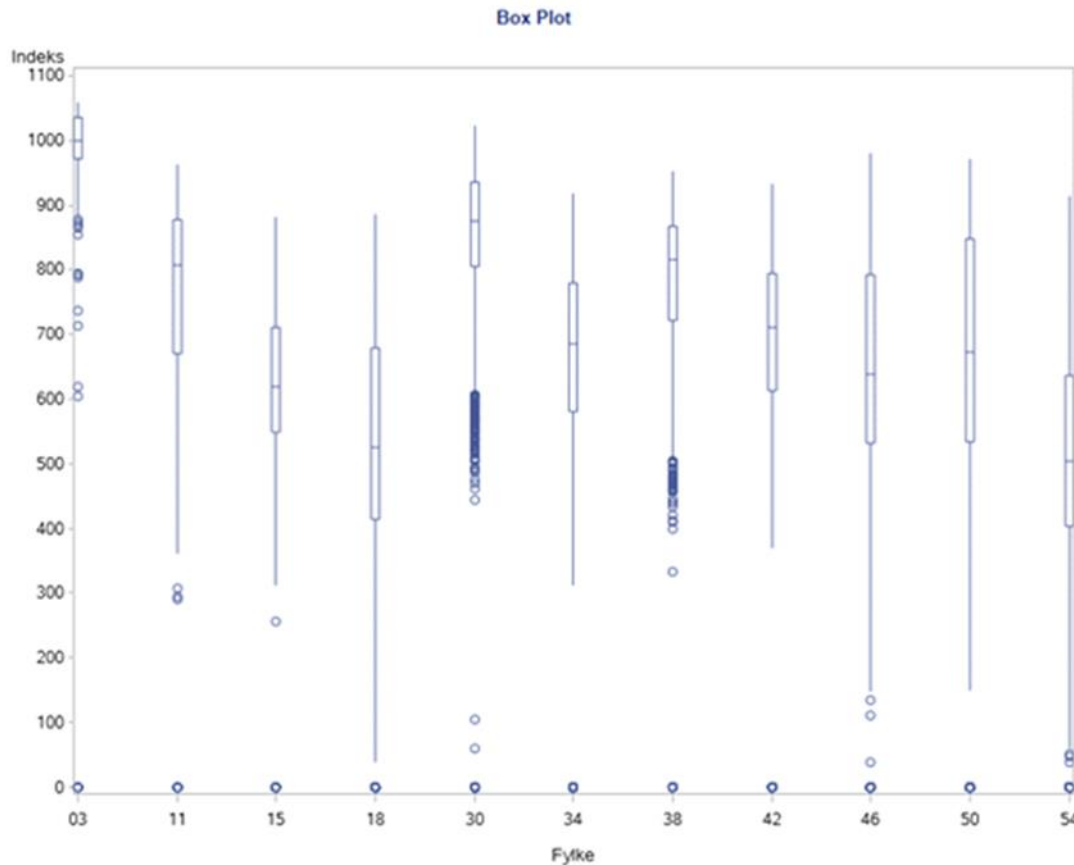
PWC (2020): «Utvikling og implementering av nytt system for verdsettelse av fritidsboliger»,
<https://www.regjeringen.no/contentassets/e9f862f850474dca89826e25e4b01314/pwc-rapport-utvikling-og-implementering-av-nytt-system-for-verdsettelse-av-fritidsboliger.pdf>.

Vinje, Thomas (2020): «Hva oppnår en med en ny modell for boligpriser», Masteroppgave Economics, UiO, juni 2020.

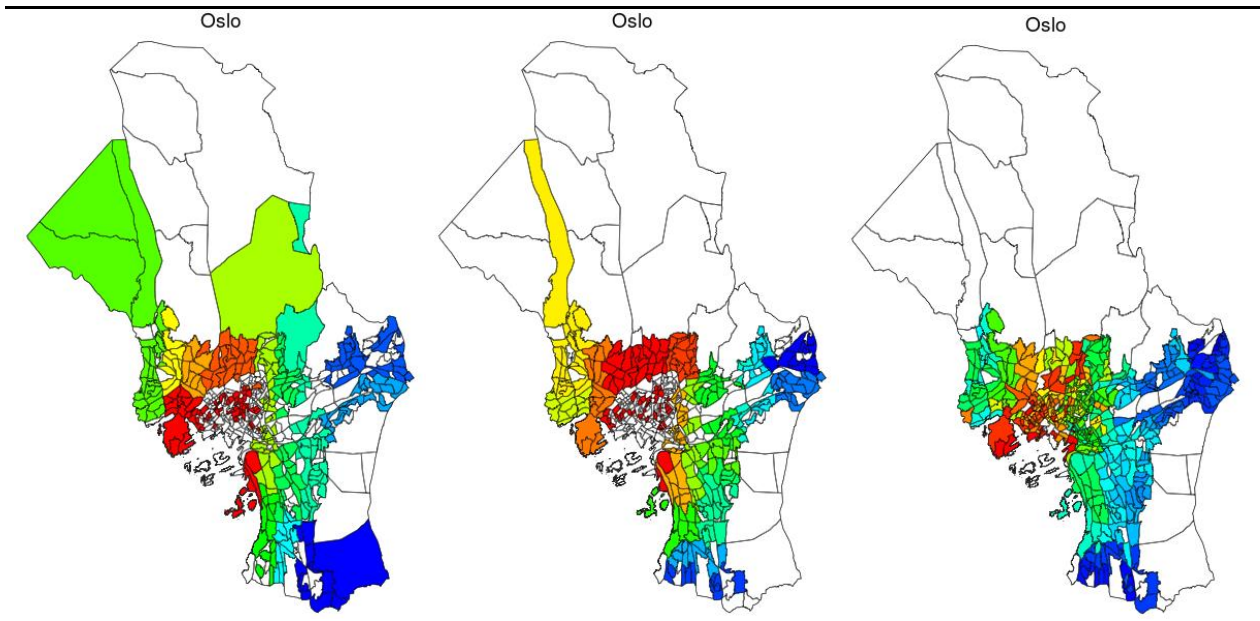
Vedlegg A: Ekstra figurer

Figur A1 viser hvordan sentralitetsindeksen for grunnkretser fordeler seg i forskjellige fylker i 2020. Den horisontale streken i hver boks er medianen.

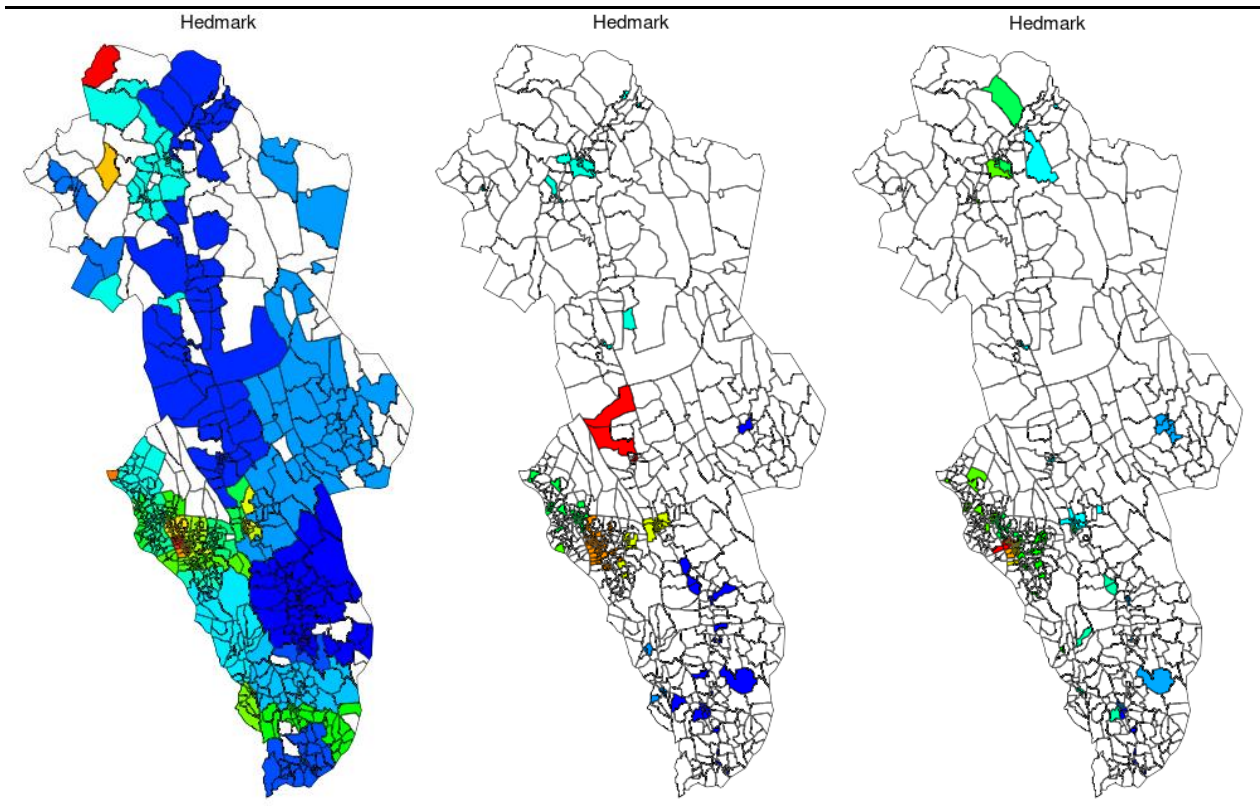
Figur A1 Fordeling av sentralitetsindeks for grunnkretser, fylker



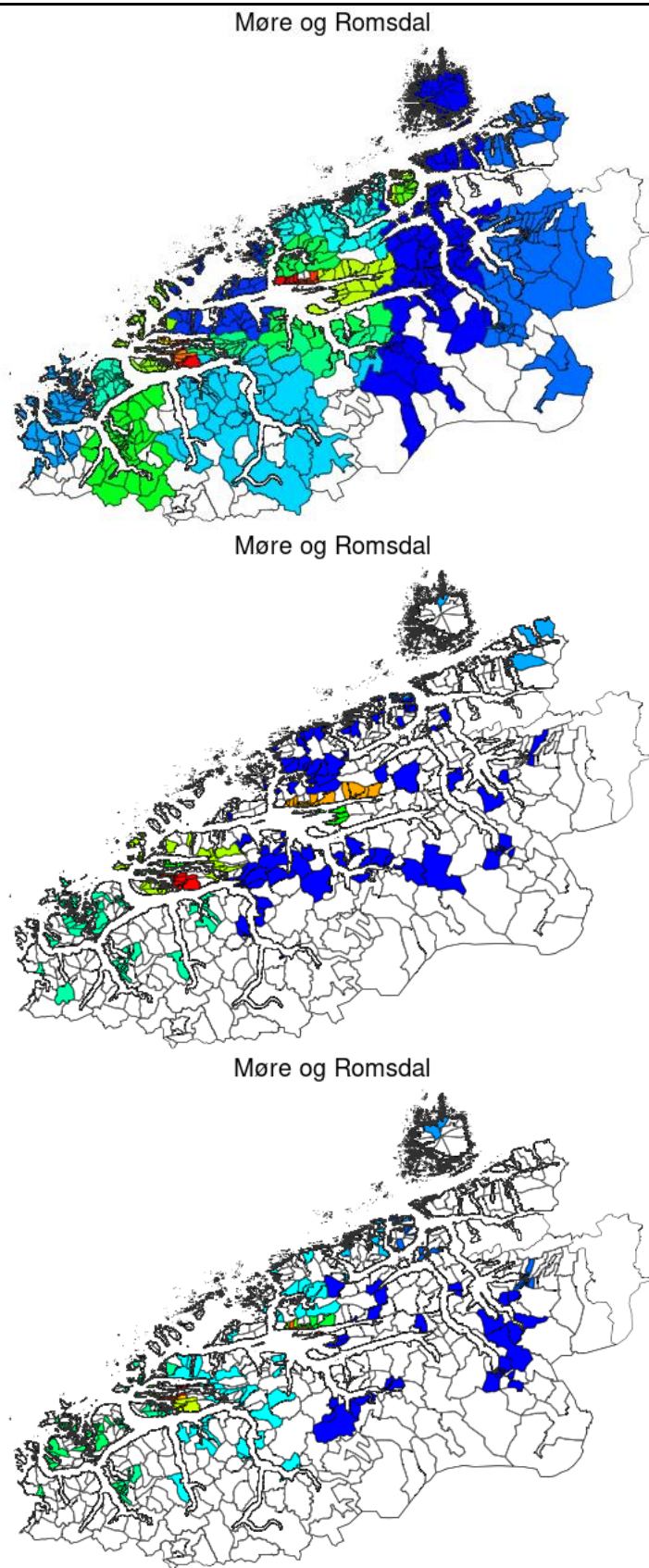
Figur A2 Kart over grunnkretser med predikert kvadratmeterpris, boligtype, Oslo



Figur A3 Kart over grunnkretser med predikert kvadratmeterpris, boligtype, Hedmark



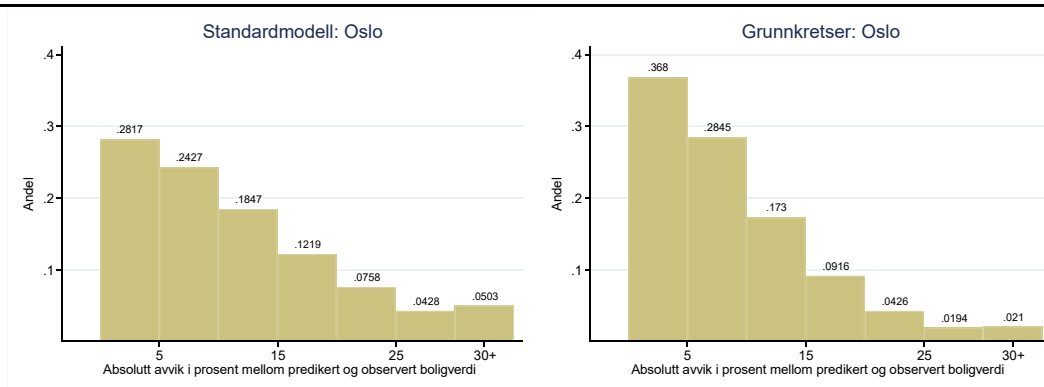
Figur A4 Kart over grunnkretser med predikert kvadratmeterpris, boligtype, Møre og Romsdal



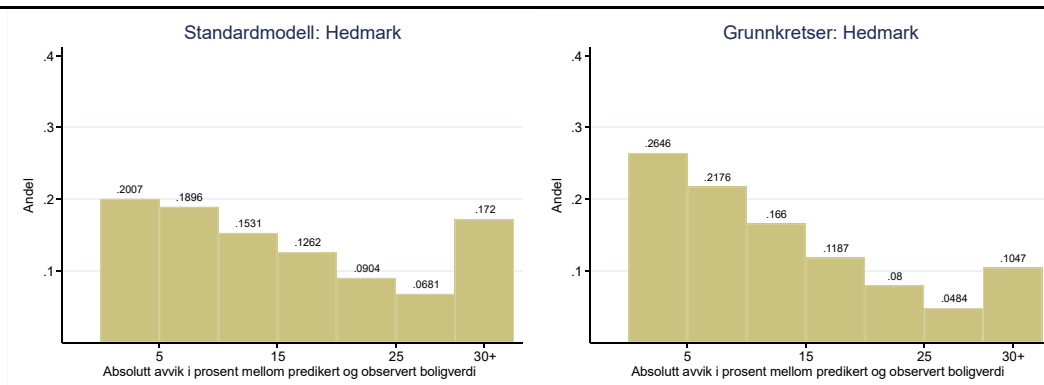
Figur A2 – A4 viser predikerte kvadratmeterpriser for forskjellige grunnkretsgrupperinger. Varme farger er høye priser, kalde farger lave. Kartene viser resultater for respektiv eneboliger, småhus og leiligheter i hvert av de tre fylkene.

2019

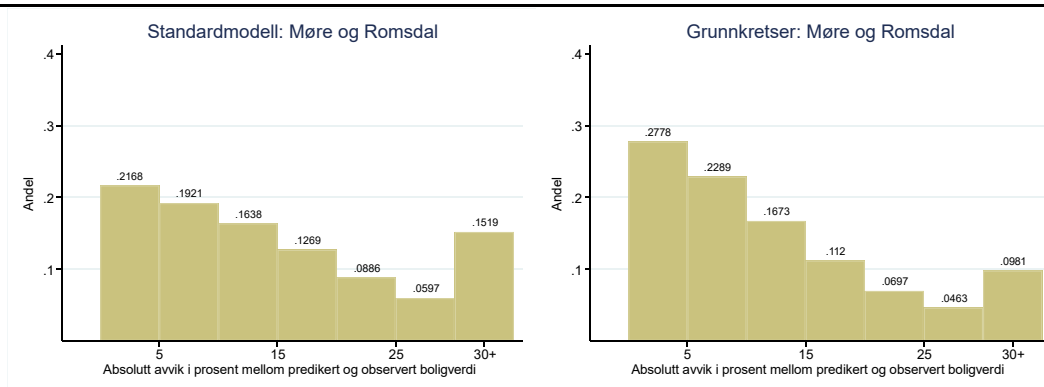
Figur A5 Fordeling av absolutte prediksjonsfeil, eksisterende og alternativ modell, Oslo, 2019



Figur A6 Fordeling av absolutte prediksjonsfeil, eksisterende og alternativ modell, Hedmark, 2019



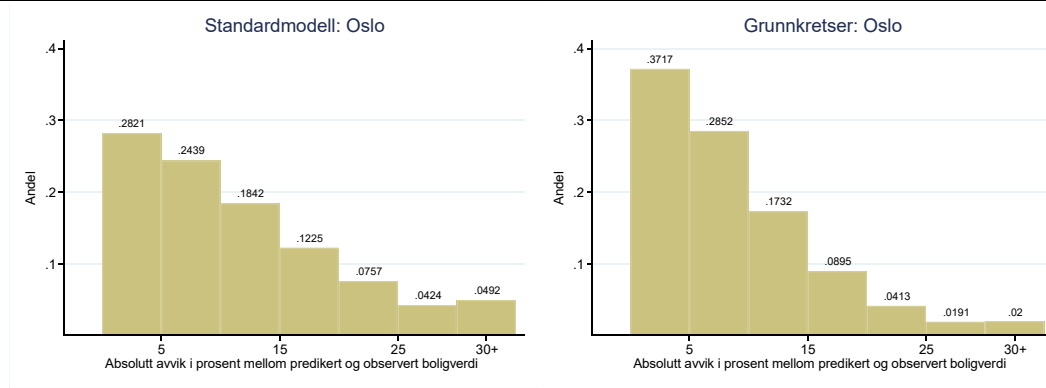
Figur A7 Fordeling av absolutte prediksjonsfeil, eksisterende og alternativ modell, Møre og Romsdal, 2019



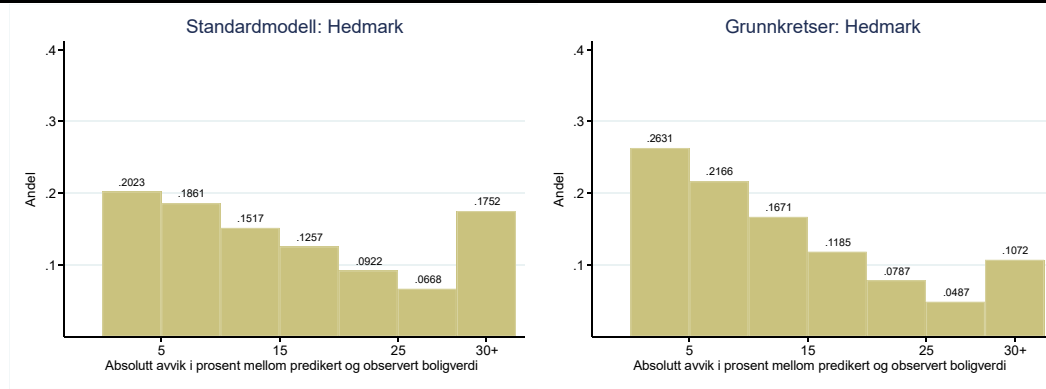
Figur A5 – A7 viser fordelingen av absolutte prosentvis prediksjonsfeil, for den eksisterende og nye modellen, og de tre fylkene i 2019. Den siste stolpen i hver figur, merket 30+, inneholder alle observasjoner med absolutt prediksjonsfeil større enn 30 prosent.

2020

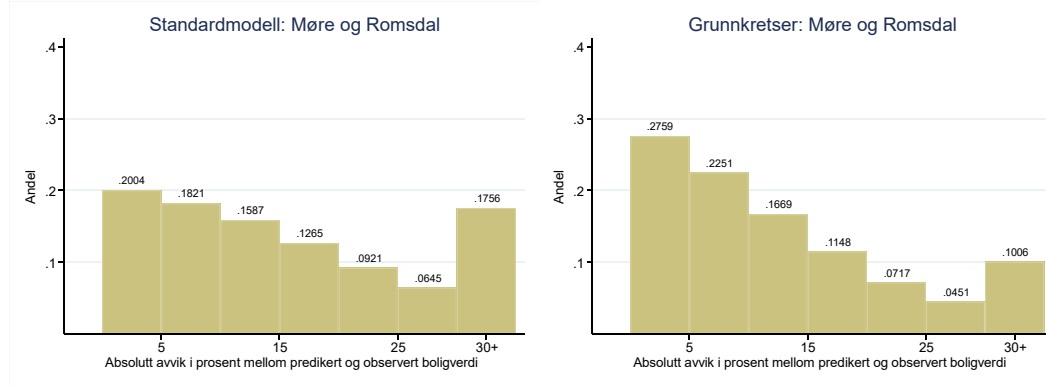
Figur A8 Fordeling av absolutte prediksjonsfeil, eksisterende og alternativ modell, Oslo, 2019



Figur A9 Fordeling av absolutte prediksjonsfeil, eksisterende og alternativ modell, Hedmark, 2019



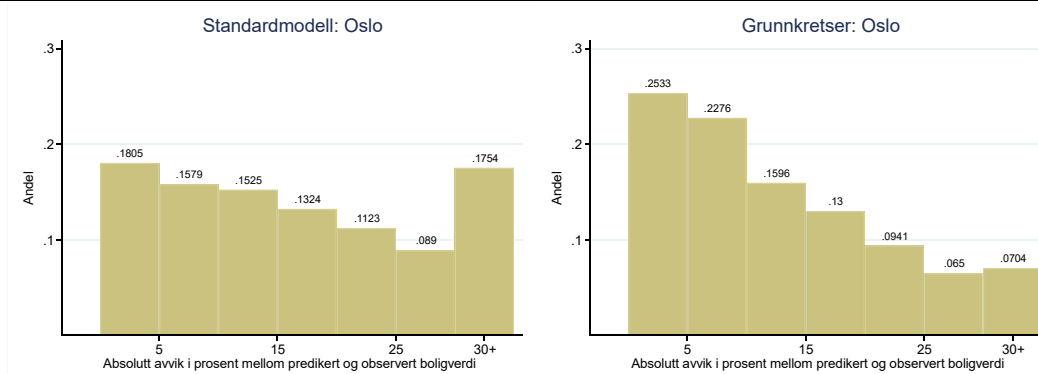
Figur A10 Fordeling av absolutte prediksjonsfeil, eksisterende og alternativ modell, Møre og Romsdal, 2019



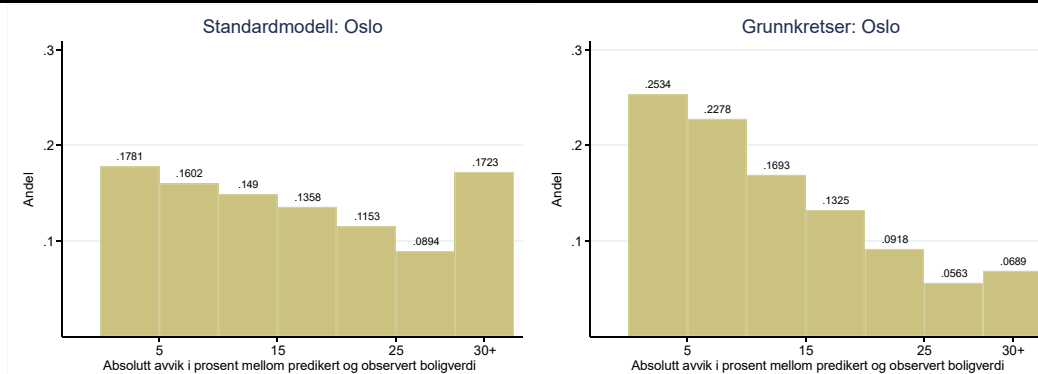
Figur A8 – A10 viser fordelingen av absolutte prosentvis prediksjonsfeil, for den eksisterende og nye modellen, og de tre fylkene i 2020. Den siste stolpen i hver figur, merket 30+, inneholder alle observasjoner med absolutt prediksjonsfeil større enn 30 prosent.

Dyre boliger

Figur A11 Fordeling av absolutt prediksjonsfeil, eksisterende og ny modell, dyre boliger, 2019



Figur A12 Fordeling av absolutt prediksjonsfeil, eksisterende og ny modell, dyre boliger, 2020



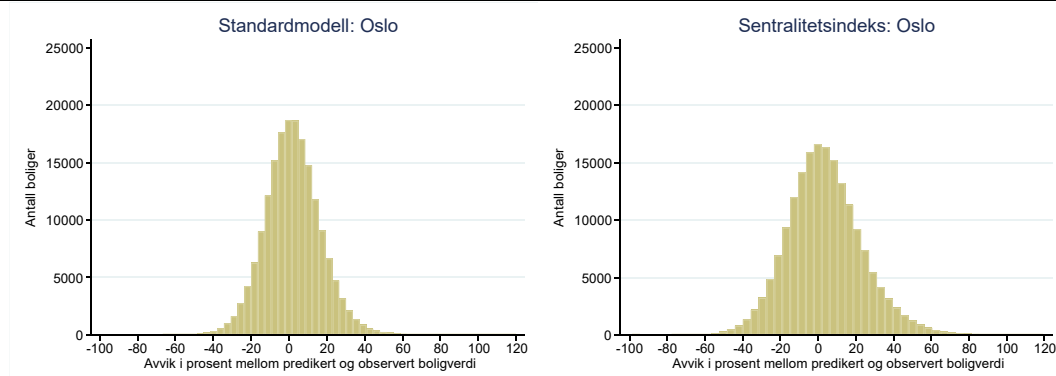
Figur A11 – A12 viser fordelingen av absolutte prosentvis prediksjonsfeil, for dyre boliger i Oslo i den eksisterende og nye modellen, i 2019 og 2020. Den siste stolpen i hver figur, merket 30+, inneholder alle observasjoner med absolutt prediksjonsfeil større enn 30 prosent.

Vedlegg B: Modell med sentralitetsindeks for grunnkretser

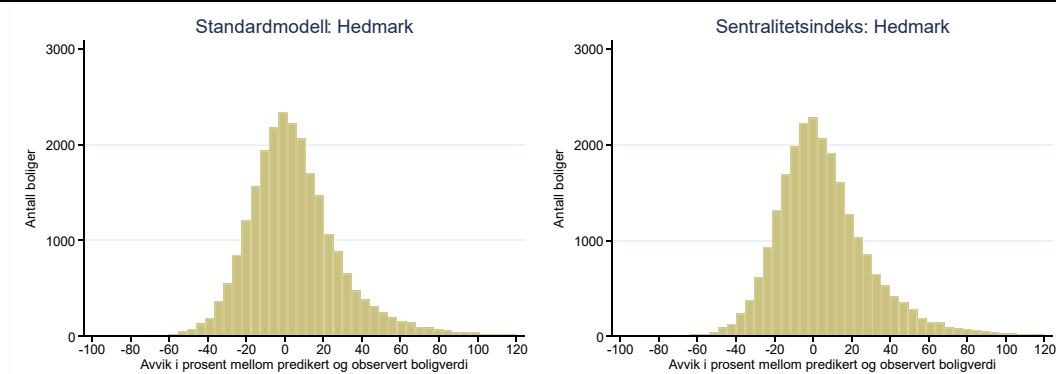
Her vises resultatene fra regresjoner der prissonene i regionene i den eksisterende modellen byttes ut med sentralitetsindeksen på grunnkrets nivå. Metoden er nærmere beskrevet i avsnitt 2.2. Resultatene er presentert på samme måte som for modellen med grunnkretsgrupper i avsnitt 3.

2019

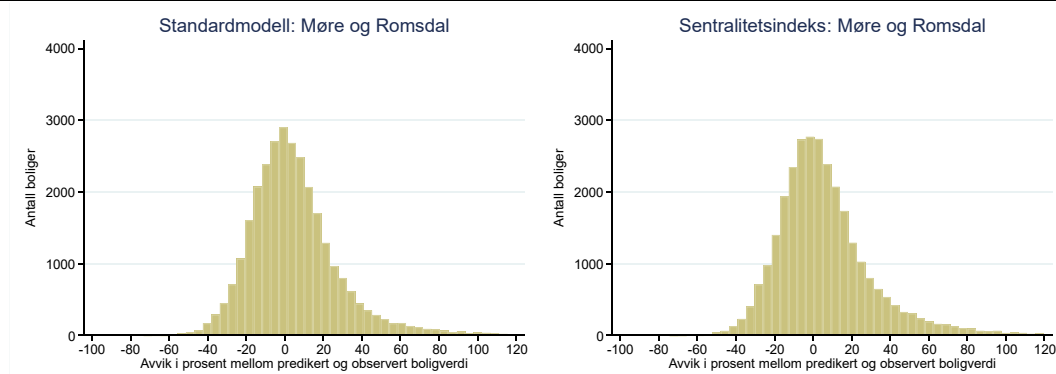
Figur B1 Fordeling av prediksjonsfeil, eksisterende og alternativ modell, Oslo, 2019



Figur B2 Fordeling av prediksjonsfeil, eksisterende og alternativ modell, Hedmark, 2019



Figur B3 Fordeling av prediksjonsfeil, eksisterende og alternativ modell, Møre og Romsdal, 2019



Figur B1 – B3 viser fordelingen av prediksjonsfeil for den eksisterende modellen og den alternative modellen med sentralitetsindeks for 2019. Tabell B1 viser en oppsummering av forskjellige mål på treffsikkerhet. Den alternative modellen treffer mye dårligere for Oslo, noe som er naturlig ettersom det er svært lite variasjon i sentralitet mellom grunnkretser i Oslo. For Hedmark treffer de to

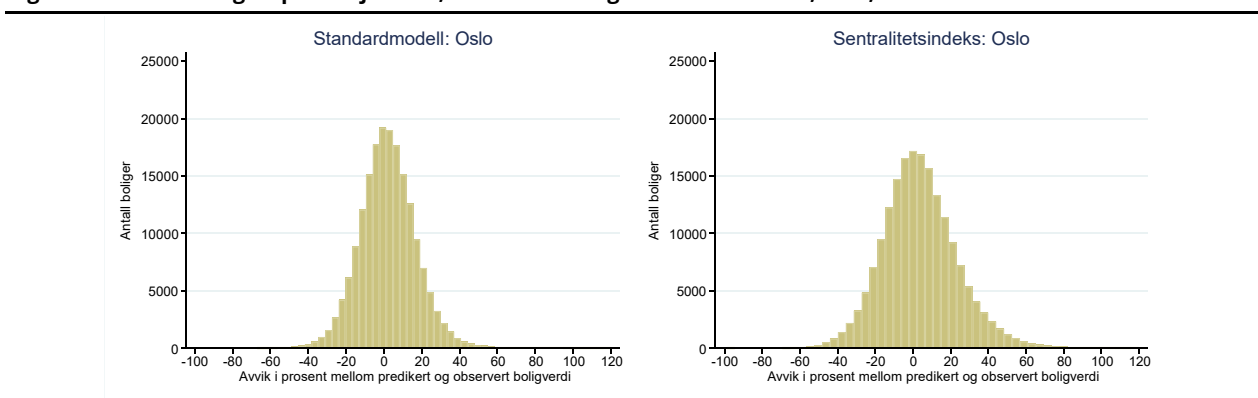
modellene omtrent like godt, mens den alternative modellen treffer noe dårligere enn den eksisterende modellen for Møre og Romsdal.

Tabell B1 Gjennomsnitt, median og standardavvik for prosentvis feil, og R², 2019

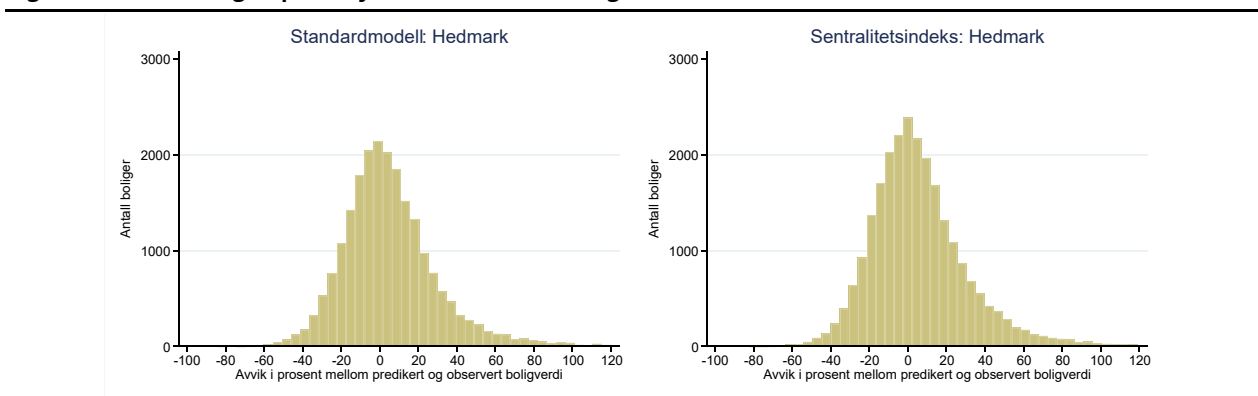
Modell	Oslo		Hedmark		Møre og Romsdal	
	Eks.	Alt. ny	Eks.	Alt. ny	Eks.	Alt. ny
Gjennomsnitt	2,43	4,06	5,75	6,05	5,46	6,73
Median	1,72	2,63	2,53	2,45	1,94	2,40
Standardavvik	17,3	21,8	25,7	25,8	25,2	26,9
R ²	0,805	0,673	0,751	0,748	0,746	0,724

2020

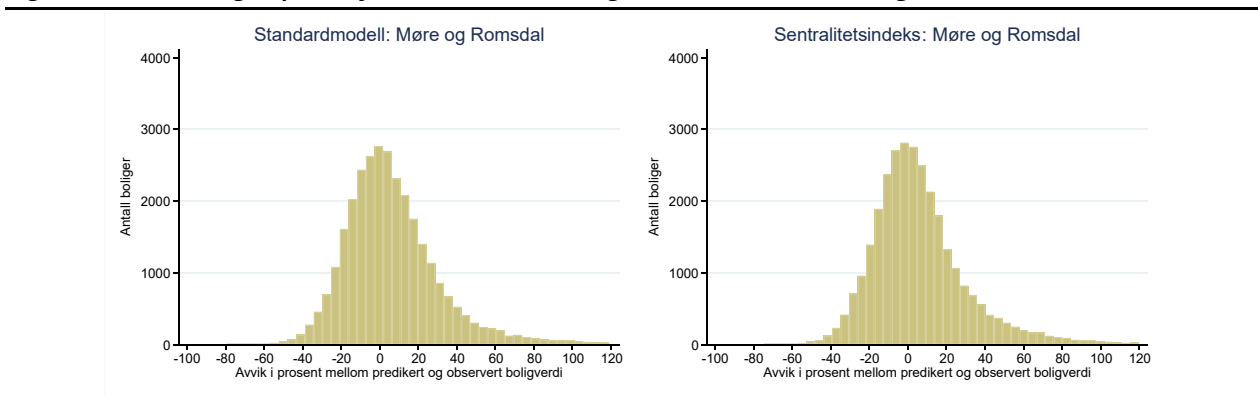
Figur B4 Fordeling av prediksjonsfeil, eksisterende og alternativ modell, Oslo, 2020



Figur B5 Fordeling av prediksjonsfeil, eksisterende og alternativ modell, Hedmark, 2020



Figur B6 Fordeling av prediksjonsfeil, eksisterende og alternativ modell, Møre og Romsdal, 2020



Figur B4 – B6 viser fordelingen av prediksjonsfeil for den eksisterende modellen og den alternative modellen med sentralitetsindeks for 2020. Tabell B2 viser en oppsummering av forskjellige mål på treffsikkerhet. Den alternative modellen treffer igjen dårligere for Oslo, og omtrent like godt for Hedmark. Møre og Romsdal har i 2020 hatt mange kommunesammenslåinger, og den eksisterende modellen treffer betraktelig dårligere enn i 2019. Allikevel treffer den alternative modellen kun moderat bedre enn den eksisterende modellen for Møre og Romsdal.

Tabell B2 Gjennomsnitt, median og standardavvik for prosentvis feil, og R², 2020

Modell	Oslo		Hedmark		Møre og Romsdal	
	Eks.	Alt. ny	Eks.	Alt. ny	Eks.	Alt. ny
Gjennomsnitt	2,40	4,00	5,55	6,16	7,32	6,98
Median	1,75	2,56	2,25	2,60	3,17	2,58
Standardavvik	17,2	21,6	26,0	26,0	27,3	27,4
R ²	0,802	0,670	0,757	0,756	0,721	0,722