Statistisk sentralbyrå
Statistics Norway

Prisindeks for nye eneboliger

Dokumentasjon av datagrunnlag og beregningsmetode

Magnus Espeland og Mona Takle

TALL SOM FORTELLER
NOTATER / DOCUMENTS
2021/34

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

Publisert: 15. oktober 2021

ISBN 978-82-587-1381-1 (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

Statistisk sentralbyrås prisindeks for nye eneboliger er utarbeidet og publisert siden 1. kvartal 1990. Indeksen ble sist dokumentert i 2018 (Haglund 2018).

Den viktigste endringen siden 2018 er at datagrunnlaget ikke lenger innhentes gjennom en skjemaundersøkelse, men hentes fra administrative registre. Fra og med 1. kvartal 2021 er det tinglysningsdata fra grunnboken som sammen med matrikkeldata som utgjør beregningsgrunnlaget for indeksen. Dette notatet gir en grundig gjennomgang av det nye datagrunnlaget og konsekvensene av dette skiftet. Notatet dokumenterer også metoden som brukes for å beregne prisindeks for nye eneboliger, selv om den i hovedsak ikke er endret siden 2018.

Statistisk sentralbyrå, 28. september 2021

Per Morten Holt

Sammendrag

Prisindeks for nye eneboliger måler kvartalsvis prisutvikling på å sette opp en ny enebolig eksklusiv tomt. Sammen med prisindeks for nye flerboliger (småhus og blokkleiligheter) inngår den i publiseringen av prisindeks for nye boliger. Dette notatet gir en oppdatert dokumentasjon av prisindeks for nye eneboligers beregningsmetode og datagrunnlag.

Måling av prisutvikling på nye boliger er utfordrende da sammensetningen av boliger som bygges i to forskjellige perioder sjelden er helt like. Boliger er en type vare som kan være svært ulike med hensyn til alder, beliggenhet, størrelse og kvalitet. For å måle prisutviklingen for boliger er det derfor nødvendig å bruke en metode som justerer for de kvalitetsmessige ulikhetene, og gjør boligene sammenlignbare over tid. I SSB bruker vi karakteristikkprisings-metode sammen med en log-lineær hedonisk funksjon for å måle prisutvikling for boliger. Notatet gir det teoretiske bakteppet for denne metoden.

Ovennevnte metode krever et datamateriale der man har de viktigste prisstyrende karakteristikkene for boligene. For nye eneboliger gir matrikkelen informasjon om to viktige egenskaper ved boligene, nemlig boligens areal og beliggenhet. Fra indeksen ble utviklet på slutten av 1980-tallet har kompletterende opplysninger om eneboligene i tillegg blitt innhentet fra de som eier boligen. Dette har vært en krevende datafangst, særlig for oppgavegiverne. Derfor har det vært ønskelig heller å benytte allerede eksisterende kilder. Fra grunnboken har man tilgang til tinglyste eiendomsoverdragelser med tilhørende prisinformasjon. Denne kilden gir oss opplysninger om nye eneboliger som er omsatt sammen med grunneiendommen. Fra grunnboken kan man ikke hente opplysninger om boliger som er satt opp i etterkant av overdragelse av eiendommen. Til tross for færre eneboliger og færre opplysninger om eneboligene, regner vi den nye datakilden som god nok for å kunne beregne en indeks som viser prisutviklingen på nye eneboliger. Dessuten vil det nye utvalget av eneboliger stemme godt overens med Eurostat sine krav til en indeks for nye eneboliger; den skal ikke omfatte selvbyggerboliger, men reflektere den faktiske markedsprisen.

Det er også blitt foretatt en mindre justering av prissonene som kommunene er delt inn i.

Innhold

1. Innledning

Prisindeks for nye eneboliger måler kvartalsvis prisutvikling på å sette opp en ny enebolig eksklusiv tomtekostnad. Indeksen har blitt publisert av SSB siden 1990. Sammen med prisindeks for nye flerboliger inngår den i publiseringen av prisindeks for nye boliger, som kom på plass som en følge av det europeiske statistikksamarbeidet (Rådsforordning 93/2013 og 2016/792). I tillegg publiserer SSB en prisindeks for brukte boliger som, i tråd med forordningen, vektes sammen med prisindeks for nye boliger til en total boligprisindeks. I forordningen spesifiseres det at den totale boligprisindeksen bør sammenstilles basert på endelige markedspriser betalt av husholdningene (Eurostat 2017 kap. 4.2).

Siden oppstarten på 1990-tallet har karakteristikkprisings-metode sammen med en log-lineær hedonisk funksjon blitt brukt for å måle prisutviklingen på boliger. Dette er internasjonalt anerkjente metoder som brukes av flere statistikkbyråer i verden. Det er også den anbefalte metoden for det europeiske statistikksamarbeidet i regi av Eurostat (Eurostat 2013). I notatet gis en kortfattet teoretisk gjennomgang av metodene samt henvisning til internasjonal litteratur[1].

Til og med 4. kvartal 2020 har det blitt sendt ut et spørreskjema til alle som er hjemmelshavere av en tomt der det er bygget en enebolig. SSB har brukt matrikkelen som utgangspunkt for å definere populasjonen av nyoppførte eneboliger hvert kvartal. Fra matrikkelen har vi hentet opplysninger om bruksareal og beliggenhet for boligene. Gjennom spørreskjemaet har vi innhentet flere opplysninger om de prisdrivende egenskapene til eneboligen. Fra og med 1. kvartal 2021 vil indeksen i stedet baseres på tinglysninger fra grunnboken.

Dette fører til en endring i populasjonen; der vi tidligere hadde med alle nylig ferdigstilte eneboliger, vil vi nå kun få med de som er solgt/tinglyst sammen med tomten. I og med at oppføring av ny enebolig på egen tomt som allerede er eid, ikke tinglyses, blir de følgelig ikke med i statistikken. Dermed utelukkes såkalte selvbyggerboliger, og vi får kun med boliger som kjøpes fra utbygger sammen med tomten. Dette er i tråd med den nevnte Eurostat-forordningen som sier at det er de endelige markedsprisene som skal måles i indeksen.

Grunnen til at det er ønskelig å holde selvbyggerboliger utenfor prisindeksen er at selvbyggernes utgifter ikke reflekterer markedsprisene for enebolig. Figur 1.1 illustrerer dette.

Figur 1.1 Input og output indekser

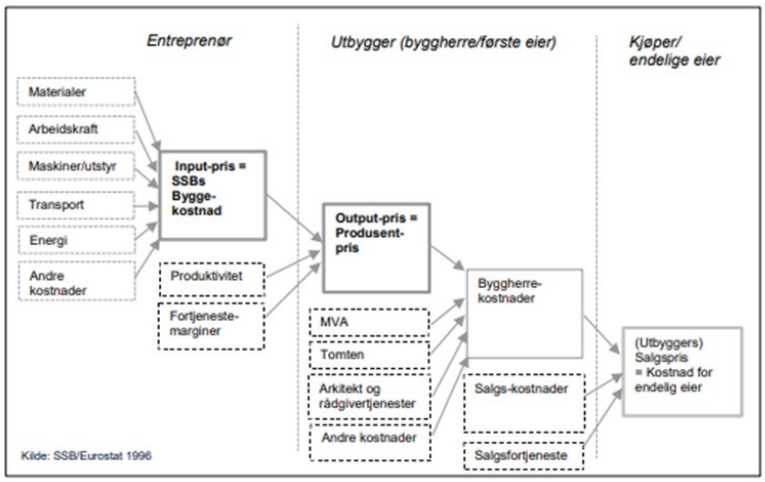


Fig. 1 SSBs begreper vedr. bygging og omsetning av boliger (SSB Rapport 200/28)

[1] Den teoretiske gjennomgangen er utarbeidet i samarbeid med Odd Erik Nygård ved forskningsavdelingen i SSB

For de som leier inn profesjonelle til å gjøre alt arbeidet, vil alle kostnadene i figur 1.1 inngå i den totale salgsprisen til boligen. Selvbyggere derimot, må forholde seg til input-prisene og byggherrekostnader og i mindre grad produktivitet og profittmargin, avhengig av hvor mye arbeid som gjøres selv. Kort oppsummert vil det for all selvbygging ikke være noe salg og derfor ingen fortjenestemargin på salget av boligen. Figur 1.1 illustrerer dette. Ekskludering av selvbyggerboliger gjør at vi får et redusert utvalg sammenlignet med tidligere metode.

På grunn av at det kun brukes registerdata blir utvalget av forklaringsvariabler noe redusert. I spørreskjema kan vi i teorien spørre om alle prisdrivende egenskaper ved eneboligene, mens vi nå er begrenset til å bruke informasjon som allerede eksisterer. I neste avsnitt kommer vi mer inn på hvilken informasjon vi har om eneboligene.

2. Datagrunnlaget

2.1. Matrikkelen

SSBs prisindeks for nye eneboliger skal dekke alle nyoppførte eneboliger i en gitt tidsperiode unntatt selvbyggerboliger. Det finnes ikke en entydig definisjon på hva som er selvbyggerbolig, og tilsvarende finnes det heller ikke én definisjon på hva som er nøkkelferdig enebolig. I mange tilfeller vil det være snakk om ulike grader av hvor mye arbeid eieren selv legger ned i boligen. Som tidligere bruker vi matrikkelen som utgangspunkt, selv om dette både kan omfatte boliger som er oppsatt av en totalentreprenør og boliger som har ulike grader av selvbygging. Vi vil senere filtrere vekk de som ikke har kjøpt tomten og eneboligen som en pakke (i én tinglysning) gjennom kobling mot tinglysningsdata. Dette gir en operasjonell definisjon av populasjonen som innebærer at vi kun har med omsetninger som består av nye eneboliger med tomt.

Eneboligene kategoriseres som i matrikkelen, dvs. eneboliger, enebolig med hybelleilighet, sokkelleilighet o.l. og våningshus. Dette gir oss opplysninger om det er en eller flere boenheter i boligen. Uttrekket omfatter alle eneboliger registrert tatt i bruk i matrikkelen, dvs. eneboliger med bygningsstatus:

- MB – Midlertidig brukstillatelse
- FA – Ferdig attest
- TB – Tatt i bruk

Gitt ovenstående restriksjoner i matrikkelen, er det de siste årene kommet til i underkant av 6 000 nyoppførte eneboliger per år, dvs. 1 400 - 1 500 eneboliger per kvartal. Det er som regel en viss forsinkelse i når boligene registreres i forhold til når de er faktisk tatt i bruk. Gjennomsnittlig forsinkelse er på 247 dager, men som regel går det ikke mer enn 2 dager. Den høye gjennomsnittlige forsinkelsen i matrikkelen skyldes hovedsakelig et fåtall bygg hvor det er veldig lang forsinkelse, mens medianen viser 2 dager forsinkelse som vi da kan anta er normalen.

Fra matrikkelen tar vi med følgende opplysninger om boligen; informasjon om beliggenhet, bruksenheter, bruksareal, antall rom, antall bad, antall wc.

2.2. Tinglysning

Når man kjøper en ny enebolig av en entreprenør, tinglyses kjøpet på lik linje med andre eiendomsomsetninger. SSB mottar tinglysningsdata en gang i måneden, og vi får tilgang til opplysningene måneden etter at transaksjonene blir registrert i grunnboka. Det tar i gjennomsnitt 2 dager før tinglysningen blir registrert.

Når noe tinglyses påfaller det en dokumentavgift, som beregnes som en prosentandel av transaksjonsprisen. Avgiften er 2,5 prosent av eiendommens markedsverdi på tinglysningstidspunktet. Ved første overføring av en nyoppført bygning som ikke er tatt i bruk, blir det bare betalt dokumentavgift av tomteverdien, ikke bygningen. Dette betyr at man gjennom tinglysningsdata kan få informasjon om både prisen på eneboligen og antatt verdi på tomten.

Med dagens regelverk for dokumentavgift for nye boliger kan det tenkes at det foreligger et insentiv til å oppgi en lavere tomteverdi enn det som er riktig for å redusere omkostningene. Vi har fått opplysninger om at Kartverket ikke opplever det som et stort problem i praksis. I de fleste tilfeller med oppføring av nye eneboliger vil det være profesjonelle aktører involvert i egenskap av utbygger og/eller eiendomsmegler. Kartverkets inntrykk er at disse innretter seg etter regelverket og forsøker å føre riktig tomteverdi i skjøtene som sendes inn for tinglysning. Dokumentavgiften vil ofte være

inntatt i kontrakten som er inngått med kjøperen, slik at insentivet for utbygger til å oppgi en lavere verdi vil være begrenset. Det er oppgitt avgiftsgrunnlag som skal legges til grunn for avgiftsberegningen med mindre registerføreren finner at det åpenbart er for lavt. Dersom det oppgitte avgiftsgrunnlaget fremstår som åpenbart for lavt vil skjøtet bli sendt i retur med anmodning om nærmere redegjørelse for avgiftsgrunnlaget.

Tinglysningsdataene inneholder en variabel som kan si om boligen er ny eller ikke. Variabelen kalles «dokumentavgiftsårsak», og to av kategoriene gjelder nye boliger; «Skjøte nyoppført bygning» og «Nyoppført bygning på festet grunn». Ettersom SSB kun har tilgang til denne opplysningen fra og med 2021 brukes en annen metode for å skille ut nye boliger før det. Vi velger da ut boliger der dokumentavgiftsgrunnlaget er mindre enn salgssummen, og vi har valgt å sette grensen til 75 prosent. Dvs. at dersom dokumentavgiftsgrunnlaget er mindre enn 75 prosent av salgssummen regner vi det for en nyoppført bygning. Undersøkelse av tinglyste eneboliger 1. kvartal 2021 viste at denne metoden ville gitt 482 tinglyste eneboliger, mens årsakskoden indikerer 491 tinglyste nye eneboliger. Den gamle metoden samsvarer godt med de nevnte kategoriene for nye boliger.

Det er ønskelig å få med så mange observasjoner som mulig. Derfor tar vi med observasjoner som blir registrert tinglyst også etter kvartalets utløp.

Gitt de nevnte kriteriene for tinglysningsdata;

1. Fritt salg, der anvendelse av grunn er til bolig
2. Boligtypen er «frittliggende enebolig», altså enebolig
3. Redusert avgiftsgrunnlag for dokumentavgift. Fra og med 2021 bruker vi i tillegg en egen variabel fra tinglysningen som sier om bygget er nytt

vil antall observasjoner som hentes fra tinglysningsdataene være på rundt 400 eneboliger i kvartalet.

Fra tinglysningsdataene hentes altså prisen på eneboligen med tomt og prisen på tomten.

2.3. Kobling av de to registrene

Etter at uttaket av nybygde eneboliger er gjort fra matrikkelen, kobler en på tinglysningsdata.

Koblingen av tinglysningsdata og bygningsdata skjer via datafelt som identifiserer grunneiendommen. Først kobles tinglysning mot grunneiendom og deretter fra grunneiendom og bygning.

Tabell 2.1 viser en oversikt over antall observasjoner fra de ulike kildene og koblingen hvert kvartal siden 2019. Fra matrikkelen har det gjennomsnittlig vært rundt 1 400 nye eneboliger som har tilfredsstilt alle kravene hvert kvartal, mens det i gjennomsnitt har vært tinglyst 427 nye eneboliger. Av disse eneboligene er det gjennomsnittlig ca. 300 nye eneboliger hvert kvartal som kobler og som dermed utgjør den endelige populasjonen. Det er ulike årsaker til at enkelte av transaksjonene ikke kobler mellom datakildene. En enebolig kan være registrert som frittliggende bolig i tinglysningsdokumentene, mens den i matrikkelen kan være registret som rekkehus eller vertikal/horisontaldelt tomannsbolig. I slike tilfeller vil observasjonen falle ut ettersom vi tar utgangspunkt i boligtypen som er registrert i matrikkelen. Det er kun de transaksjonene som finnes i begge datasettene som kobler og blir en del av indeksberegningene.

Tabell 2.1 Antall observasjoner i de ulike registrene og ved kobling

Periode	Matrikkel	Tinglysning	Koblet
1kv.2019	1427	356	228
2kv.2019	1504	508	382
3kv.2019	1346	388	300
4kv.2019	1441	484	312
1kv.2020	1467	314	302
2kv.2020	1455	408	336
3kv.2020	1269	496	273
4kv.2020	1503	546	344
1kv.2021	1216	343	244

2.4. Kontroll og editering

Datamaterialet kontrolleres for ekstremverdier. Ekstremverdier kan for eksempel oppstå gjennom feilregistreringer i matrikkel. Et eksempel på en feilregistrering kan være en enebolig som f.eks. blir solgt for 11 millioner kroner og har bruksareal på 3 kvadratmeter. Slike tilfeller vil bli fjernet fra utvalget ettersom vi ikke vil ha korrekt bruksareal.

Editeringen skjer ved hjelp av enkle kontroller. Vi fjerner alle eneboliger der:

- arealet ligger utenfor intervallet 50 – 600 kvadratmeter.
- kvadratmeterprisen ligger utenfor intervallet 7 000 – 90 000 kroner. Maks-grensen her vil naturlig nok justeres med prisnivået. Grensen bør vurderes en gang i året.

I tillegg til disse enkle kontrollene av data for å se om det er noen åpenbare ekstremverdier, har vi også brukt Cook's distance (CD) for å identifisere usannsynlige eller ekstreme verdier. CD blir beregnet for hver observasjon og er et mål på den totale innflytelsen hver observasjon har på de estimerte koeffisientene. Høyere innflytelse er assosiert med ekstreme verdier. En ekstrem verdi kan defineres som en verdi som enten er feil, usannsynlig eller som overstiger en viss grense. Hovedmålet er å eliminere ekstreme verdier i de avhengige variablene så vel som i de viktigste uavhengige variablene, som kan forvride den hedoniske modellen. CD identifiserer transaksjoner som skiller seg ut på grunn av usannsynlige kombinasjoner av ulike variabler. En enebolig på 50 kvadratmeter med 10 rom vil bli identifisert som usannsynlig, da kombinasjonen av de to variablene er svært tvilsom (Brand, 2020). Observasjoner med CD verdi over grenseverdien (4/n) [2] kan anses som usannsynlige, hvor n er antall observasjoner. I testen vår var det ca. 5 prosent, 119 av 2418 observasjoner, som hadde en verdi over grenseverdien. Ved nærmere undersøkelse av disse observasjonene, var det utfordrende å se noe åpenbart feil med transaksjonene. Eurostat (2017) skriver følgende om datakvalitetskontroller og ekstremverdier; *«Utfallet av datakvalitetskontroller trenger ikke nødvendigvis innebære en automatisk avvisning av informasjonen som har blitt signalisert som ekstreme, da de kan skjule viktig informasjon. En ekstrem verdi trenger ikke være en feil observasjon»*. I dette dokumentet er det da lagt vekt på Eurostats utsagn og beholdt alle transaksjonen, da det var utfordrende å se feilen i de transaksjonene som hadde en høy CD verdi over grensen.

2.5. Hva innebærer overgangen fra skjemaundersøkelse til registerdata?

Som en oppsummering kan det være interessant å se på hovedforskjellene i datagrunnlaget før og etter 1. kvartal 2021. Mens populasjonen tidligere var alle eneboliger som var oppført i kvartalet, vil vi nå kun få med de som har kjøpt en tomt med en nyoppført enebolig på. Dette samsvarer godt

[2] Denne grenseverdien brukes blant annet i Sveits og Irlands offisielle eiendomsprisindeks (Brand 2020).

med det Eurostat ønsker, ettersom det er markedsprisen som skal måles. Dersom det er stor grad av egeninnsats som inngår i boligen, kan det være vanskelig å få et riktig bilde av verdien. Det kan innvendes at vi kunne/burde hatt med eneboliger som er bygget av utbygger på en allerede tinglyst tomt, men det er ikke mulig å få til ved bruk av tinglysningsdata ettersom dette ikke tinglyses. Likeledes kan det tenkes tilfeller der eneboliger med tomt selges, men der noe arbeid overlates til eieren. I dette statistikkproduksjonsløpet skal alle eneboligene være godkjent for å ta i bruk fra kommunens side, så man kan anta at det i disse tilfellene er mindre arbeid som gjenstår.

Endringen i populasjonen fører til færre observasjoner. Mens vi tidligere fikk med ca. 500-800 observasjoner i kvartalet, ligger antallet nå på rundt 300. Samtidig er populasjonen nå mer homogen og opplysningene kan antas å være mer riktige (offentlige dokumenter – profesjonelle aktører, ikke privatpersoner).

Et viktig argument for å skifte datagrunnlag er oppgavebyrden. Tiden det tok for å fylle ut et skjema i undersøkelsen kan anslås å være i snitt 20 minutter. Antall innsendte skjema var rundt 800 noe som tilsvarer 250-300 timer i kvartalet – 1000-1200 timer i året. Nesten ett årsverk.

Som nevnt i innledningen reduseres antall forklaringsvariabler noe som følge av overgangen til registerdata fremfor skjemaundersøkelser. Totalt går vi fra 28 til 9 forklaringsvariabler. Det er en ulempe at det ikke lenger kan spørres om alle relevante prisdrivende egenskaper ved boligene. Reduksjonen i antall forklaringsvariabler kan forsvares med at registerdata anses for å være mer pålitelige og potensielle feilkilder i datamaterialet reduseres, samtidig som vi fremdeles har tilgang til de to klart viktigste forklaringsvariablene, bruksareal og beliggenhet. Eksempler på potensielle feilkilder kan være at enkelte velger å ikke svare på undersøkelsen, kun svarer på noen spørsmål eller rapporterer feil opplysninger. Det finnes flere eksempler på feilkilder, men det er nok hovedsakelig disse to som gjaldt skjemaundersøkelsen.

3. Metode for beregning av prisindekser for boliger

3.1. Hedonisk metode

SSB har tradisjon for å bruke hedonisk metode til beregning av prisindekser for boliger. Det teoretiske grunnlaget for SSB sitt arbeid med hedonisk metode bygger hovedsakelig på Rosens modellbeskrivelse og Wigrens undersøkelser av småhusprisene i Sverige (Rosen 1974, Wigren 1986). Metoden forutsetter at det er en sammenheng mellom boligens pris og dens beliggenhet, størrelse og standard. Man ønsker å finne denne sammenhengen, slik at man kan korrigere for forskjell i egenskaper til boligene. Det forhold at boliger oppført i ulike perioder vil være kvalitativt forskjellige bør ikke påvirke en indeks for boligprisen (Wass 1992, Lillegård 1994).

Man ønsker altså å finne en funksjon der prisen er den avhengige variabelen, mens ulike kvaliteter ved boligen er forklaringsvariabler. For å finne hvilke karakteristikker ved boligen som har betydning for markedsprisen, benyttes lineær regresjonsanalyse. Resultatene fra regresjonsanalysen viser hvilke karakteristikker som er statistisk signifikante og tilhørende priskoeffisienter.

Prisindeksen kan defineres som forholdet mellom prisen på to kvalitetsmessig like boliger i det aktuelle kvartalet og et basistidspunkt.

3.2. Prisfunksjon

Man antar at markedsprisen p på en bolig h kan beskrives som en funksjon av boligens K ulike egenskaper z.

$$p_{ht} = f\left(z_{ht}^{1}, \ldots, z_{ht}^{K}, \varepsilon_{h}\right) \tag{3.1}$$

der ε er en stokastisk variabel med forventning lik 0 og konstant variasjon.

En standard spesifikasjon av (3.1) kan være en lineær regresjonsligning estimert for hver tidsperiode t ved minste kvadraters metode (MKM).

$$p_{ht} = \beta_{t}^{0} + \sum_{1}^{K} \beta_{t}^{k} z_{ht}^{k} + \varepsilon_{ht,} \tag{3.2}$$

Her kan β-koeffisientene tolkes som priser for de ulike karakteristikkene ved boligen. Alternativt brukes ofte en log-lineær versjon av ovenstående:

$$\ln p_{ht} = \beta_{t}^{0} + \sum_{1}^{K} \beta_{t}^{k} z_{ht}^{k} + \varepsilon_{ht,} \tag{3.3}$$

Mange av forklaringsvariablene vil være dummy-variabler, men noen vil være kontinuerlige slik som areal. Disse kan også transformeres til logaritmisk form og da tolkes som priselastisitet, dvs. 1 prosent økning i z^k gir β prosent økning i boligprisen.

3.3. Prisindekser basert på hedonisk metode

3.3.1 Indeksmetoder og forskjellige gjennomsnitt

De vanligste metodene for å konstruere prisindekser er Laspeyres og Paasche. De måler den relative endringen i utgifter for en gitt varekurv mellom to perioder. Ved den hedoniske metoden brukt i boligmarkedet vil varekurven bestå av ulike boliger slik at hver vare i indeksen har kvantitet lik 1. På denne måten måles gjennomsnittsprisen for like boliger i ulike perioder.

Problemet i boligmarkedet er at samme bolig aldri eller sjelden kan måles i to etterfølgende perioder og vi må derfor imputere manglende observasjoner. Ved å kjøre en regresjonsmodell basert på data fra periode t kan man imputere, eller predikere, manglende boligprisobservasjoner ved tidspunkt t. Ved bruk av Laspeyres indeks som startpunkt kan boligprisindeksen skrives slik:

$$I_{L0} = \frac{\sum_{h=1}^{H_0} \hat{p}_{ht}}{\sum_{h=1}^{H_0} p_{h0}} \tag{3.4}$$

der H_0 er antall boliger observert i periode *0,* og $\hat{p}$ indikerer at prisen er imputert basert på vår hedoniske regresjonsmodell. Dvs. at vi måler boligprisene i periode *0* mens vi bruker de imputerte prisene for de samme boligene i periode *t*. På samme måte kan vi konstruere en indeks der basisperioden er *t* og imputere prisene i periode *0*, slik at:

$$I_{Lt} = \frac{\sum_{h=1}^{H_t} p_{ht}}{\sum_{h=1}^{H_t} \hat{p}_{h0}} \tag{3.5}$$

Dette er da en Paasche indeks. Disse indeksmetodene kalles aritmetiske da de er additive i sin natur og gir oss endringene i de aritmetiske gjennomsnittene. Disse er da også konsistente med den lineære modellen i (3.2). En annen type er geometriske indekser, som er konsistente med den log-lineære modellen i (3.3). En geometrisk Laspeyres indekstype kan konstrueres slik:

$$I_{G0} = \frac{\prod_{h=1}^{H_0} (\hat{p}_{ht})^{1/H_0}}{\prod_{h=1}^{H_0} (p_{h0})^{1/H_0}} \tag{3.6}$$

og på samme måte kan man tilsvarende konstruere den geometriske versjonen til en Paasche indeks:

$$I_{Gt} = \frac{\prod_{h=1}^{H_t} (p_{\mathrm{ht}})^{1/H_t}}{\prod_{n=1}^{H_t} (\hat{p}_{h0})^{1/H_t}} \tag{3.7}$$

En egenskap ved geometriske gjennomsnitt er at de er symmetriske slik at det er uten betydning hvor prisøkningen skjer. En 10 prosent prisøkning for en dyr bolig vil ha samme effekt på prisindeksen som en tilsvarende økning på en billig bolig. Det aritmetiske gjennomsnittet vekter endringen for den dyrere boligen høyere ved at en 10 prosent økning på den dyre boligen øker indeksen mer enn tilsvarende økning for den billige boligen.

3.3.2 Dobbel imputering og karakteristikkprisings-metode

Som beskrevet ovenfor kan vi bruke imputering når priser mangler for en periode. Enda vi har de faktiske prisene for referanseperioden kan det være en idé å bytte ut disse med predikerte priser,

p_{h0} med $\hat{p}_{h0}$ i (3.4). Dette gjøres i noen grad for å kompensere for en eventuell skjevhet i den hedoniske modellen grunnet utelatte forklaringsvariabler. Siden $\sum_h \hat{p}_{ht} / H_t = \hat{\beta}_t^0 + \sum_k \hat{\beta}_t^k \bar{z}_t^k$ for alle *t*, der $\bar{z}_t^k$ er karakteristikkenes gjennomsnittsverdi i periode *t*, kan Laspeyres indeks i (3.4) omskrives som

$$I_{L0}^D = \frac{\sum_{h=1}^{H_0}[\hat{\beta}_t^0 + \sum_k \hat{\beta}_t^k z_{0h}^k]}{\sum_{h=1}^{H_0}[\hat{\beta}_0^0 + \sum_k \hat{\beta}_0^k z_{0h}^k]} = \frac{\hat{\beta}_t^0 + \sum_k \hat{\beta}_t^k \bar{z}_0^k}{\hat{\beta}_0^0 + \sum_k \hat{\beta}_0^k \bar{z}_0^k} \tag{3.8}$$

Ved å bruke samme fremgangsmåte kan man også bruke dobbel imputering på (3.5), der eneste forskjellen er at vi bruker karakteristikkenes gjennomsnittsverdier for periode t. Samme resonnement kan brukes for de geometriske indeksformlene og vi får da for Laspeyres indeks:

$$I_{L0}^D = \frac{\prod_h (exp(\hat{\beta}_t^0 + \sum_k \hat{\beta}_t^k \bar{z}_0^k))^{1/H_0}}{\prod_h (exp(\hat{\beta}_0^0 + \sum_k \hat{\beta}_0^k \bar{z}_0^k))^{1/H_0}} = exp(\hat{\beta}_t^0 - \hat{\beta}_0^0)\, exp\left[\sum_k (\hat{\beta}_t^k - \hat{\beta}_0^k)\bar{z}_0^k\right] \tag{3.9}$$

eller for Paasche indeks:

$$I_{P0}^D = \frac{\prod_h (exp(\hat{\beta}_t^0 + \sum_k \hat{\beta}_t^k \bar{z}_t^k))^{1/H_t}}{\prod_h (exp(\hat{\beta}_0^0 + \sum_k \hat{\beta}_0^k \bar{z}_t^k))^{1/H_t}} = exp(\hat{\beta}_t^0 - \hat{\beta}_0^0)\, exp\left[\sum_k (\hat{\beta}_t^k - \hat{\beta}_0^k)\bar{z}_t^k\right] \tag{3.10}$$

Dette viser hvordan dobbel imputering for lineære modeller kan reduseres til det som beskrives som karakteristikkprising metode. Hovedessensen ved denne metoden er at man sammenligner prisutviklingen for en «standardisert» bolig over tid.

Hvis vi bruker MKM regresjon som metode i (3.8) vet vi at $\sum_h^{H_0} p_{h0} = \sum_h^{H_0} \hat{p}_{h0}$, fordi residualene vil summeres opp til 0. Det betyr at singel og dobbel imputert indeks sammenfaller for både det aritmetiske og geometriske tilfellet. Hvis vi antar at koeffisientene er konstante over kortere tidsperioder slik at $\beta_0^k = \beta_t^k$ for alle *k*, kan vi vha. MKM estimere en indeks uten å måtte bruke data fra begge perioder. Vi kan da omskrive (3.10) som:

$$I_{IP} = \frac{\prod_{h=1}^{H_t}(p_{ht})^{1/H_t}}{\prod_{h=1}^{H_0}(p_{h0})^{1/H_0}} \quad exp\left[\Sigma_k \hat{\beta}_0^k (\bar{z}_0^k - \bar{z}_t^k)\right] \tag{3.11}$$

Dvs. at vi baserer estimeringen av *β* parameterne på data fra basisperioden. I praksis kan grunnlaget for å estimere *β* parameterne bestå av sammenslåtte data fra flere foregående perioder og vi trenger ikke å kjøre en ny regresjon for hver ny statistikkperiode. Kvalitetsjusteringsfaktoren kan omformuleres til $1 / \exp(\sum_k \hat{\beta}_0^k (\bar{z}_t^k - \bar{z}_0^k))$ slik at (3.11) kan skrives som

$$I_{IP} = \frac{\prod_{h=1}^{H_t}(p_{ht})^{1/H_t}}{\prod_{h=1}^{H_0}(p_{h0})^{1/H_0}} \Big/ exp\left[\Sigma_k \hat{\beta}_0^k (\bar{z}_t^k - \bar{z}_0^k)\right] \tag{3.12}$$

og tolket som en implisitt Paasche indeks, der nevneren på høyre side er en Laspeyres kvantitet indeks (Hill, 2013). For å se dette kan man ved å bruke MKM omskrive ovenstående til:

$$\frac{\prod_{h=1}^{H_t}(p_{th})^{1/H_t}}{\prod_{h=1}^{H_0}(p_{0h})^{1/H_0}} = \frac{\exp\left(\overline{\ln p_t}\right)}{\exp\left(\overline{\ln p_0}\right)} = \frac{\exp\hat{\beta}_t^0 \exp\sum_k \hat{\beta}_t^k \overline{z}_t}{\exp\hat{\beta}_0^0 \exp\sum_k \hat{\beta}_0^k \overline{z}_0} \tag{3.13}$$

noe som indirekte viser at (3.12) og (3.10) er identiske.

3.3.3 Kjeding av indekser

Ved utarbeiding av tidsserier for prisindekser kreves det som regel at man skifter basisperiode regelmessig. Dette er fordi at sammensetningen av varekurvene endres over tid og/eller at vekter endres ved sammenslåing av ulike delindekser. Ved bruk av den hedoniske metoden kan varene i kurven sammenlignes med de ulike koeffisientene i regresjonsligningen. Vi antar tidligere (3.11) at disse koeffisientene er konstante over kortere tidsperioder, men ved utarbeiding av en lengre tidsserie må vi reestimere modellen jevnlig. Samtidig med reestimeringen skiftes også perioden for hvor vi henter snittprisene for forklaringsvariablene ut. Ved enkel imputering vil den senere perioden sammenfalle med perioden vi bruker til å reestimere modellen. Ved dobbel imputering vil disse periodene kunne være forskjellige ved at vi f.eks. bruker en lengre periode for å estimere modellen. Formelen for en kjedet indeks i periode t ser slik ut:

$$I_{kjedet}^t = I_{ny\ base}^t \frac{I_{gammel\ base}^{t-1}}{I_{ny\ base}^{t-1}} \tag{3.14}$$

4. Variabler, inndelinger og begrep

I dette kapittelet går vi gjennom de viktigste variablene som vil brukes i den hedoniske modellen for prisindeks for nye eneboliger.

4.1. Salgssum

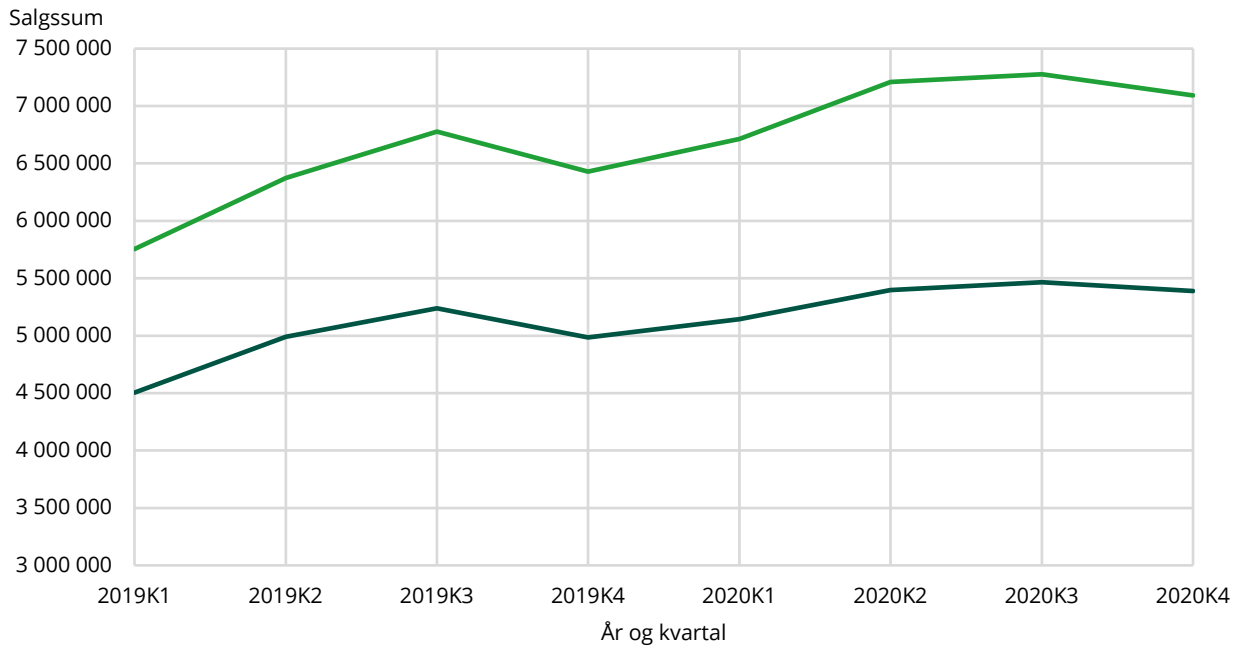
Den avhengige variabelen i regresjonsmodellen er totale kostnader inkludert merverdiavgift eksklusiv tomteverdi. Tomteverdien omfatter både råtomtens verdi og opparbeidelseskostnader. Opparbeidelseskostnader inkluderer blant annet grunnarbeid, sprenging, gravearbeid, planering og kostnader for vei, kloakk og strøm til tomtegrensen og tilknytningsavgift for vann, kloakk og strøm (Skatteetaten, 2021). Tomteverdien settes på en byggeklar tomt og tilsvarer dokumentavgiftsgrunnlaget for nye boliger. Både salgssum (tomt og bygning) og tomteverdi hentes fra tinglysningene. Tabell 4.1 gir en oversikt over salgsverdi uten tomt i populasjonen i tidsperioden 2019 – 2020.

Tabell 4.1 Salgssum ekskludert tomt – Nye eneboliger. 2019K1 – 2020K4

Salgssum ekskludert tomt	
1 000 000 - 1 999 000	26
2 000 000 - 2 999 999	233
3 000 000 - 3 999 999	636
4 000 000 - 4 999 999	507
5 000 000 - 5 999 999	387
6 000 000 - 6 999 999	242
7 000 000 - 7 999 999	107
8 000 000 - 8 999 999	105
9 000 000 - 9 999 999	63
over 10 000 000	112
Totalt	2 418

Figur 4.1 viser en oversikt over den totale gjennomsnittlige salgssummen inkludert og ekskludert tomt i tidsperioden 1. kvartal 2019 til 4. kvartal 2020. Differansen mellom grafene kan sees på som gjennomsnittlig tomtepris.

Figur 4.1 Gjennomsnittlig pris inkludert og ekskludert tomt. 2019K1 – 2020K4



4.2. Areal

Arealbegrepet som brukes i denne prisindeksen er bruksareal, dvs. alt areal innenfor husets yttervegger uansett etasje. Rom med skråtak regnes som måleverdig inntil 0,6 meter utenfor høyden 1,9 meter (jf. NS 3940). Bruksareal hentes fra Matrikkelen. I regresjonsmodellen omformes bruksareal til logaritmisk form. Tabell 4.2 gir en oversikt over gjennomsnittlig bruksareal og gjennomsnittlig kvadratmeterpris uten tomt fordelt i mindre grupper og totalt i tidsperioden 1. kvartal 2019 til 4. kvartal 2020. Vi ser at gjennomsnittlig bruksareal er 183 kvadratmeter og gjennomsnittlig kvadratmeterpris uten tomt er 28 226 kroner for hele utvalget.

Tabell 4.2 Oversikt over gjennomsnittlig bruksareal og gjennomsnittlig kvadratmeterpris uten tomt – Nye eneboliger. 2019K1 – 2020K4

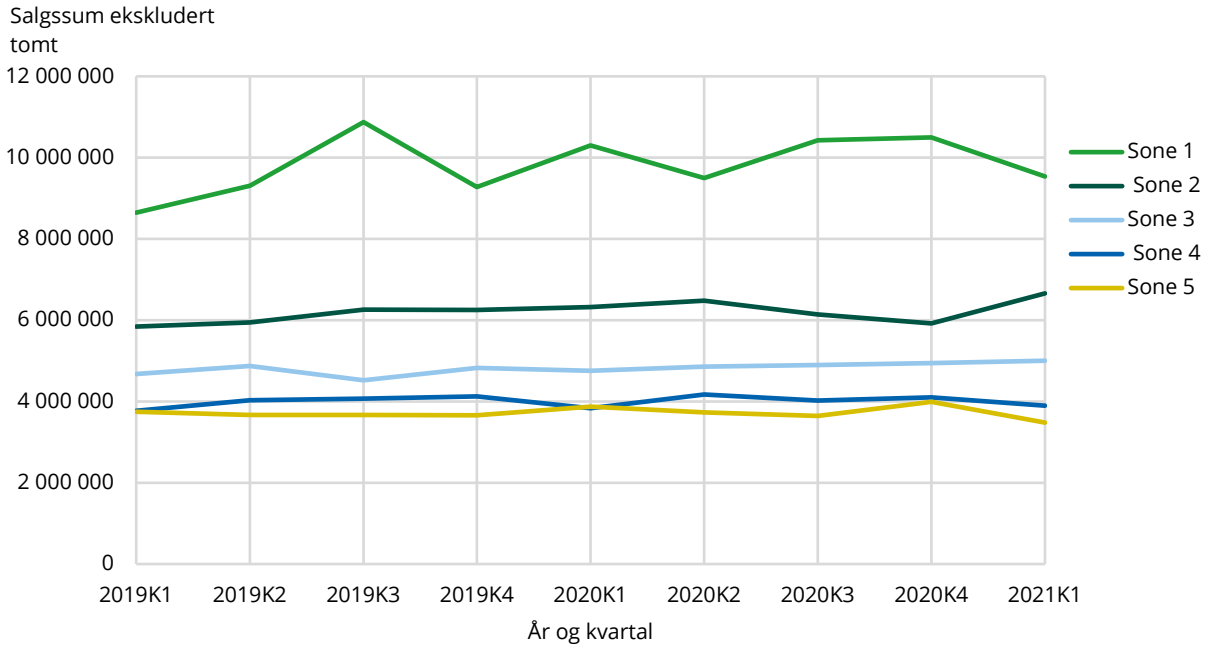
Areal	Gjen bruksareal	Gjen. kvm pris	Observasjoner
Under 100	88	31 797	57
100-149	132	27 954	605
150-199	173	28 346	965
200-249	222	28 088	541
250-299	270	28 085	195
Over 300	329	27 292	55
Totalt	183	28 226	2 418

4.3. Prissoner

Norge er her delt inn i fem prissoner med kommunekode som fordelingsnøkkel. Inndelingen er basert på gjennomsnittlig kvadratmeterpris for nyoppførte eneboliger i perioden 2004 – 2013. I soneinndelingen er det lagt vekt på at sonene er robuste med tilstrekkelig antall prisobservasjoner. Kommuner med færre enn 50 observasjoner i måleperioden er plassert i én av de to laveste prissonene basert på fylkesgjennomsnittet for alle kommuner med færre enn 50 observasjoner.

Fra 1. kvartal 2021 ble det gjort en endring i soneinndelingen. Tidligere var Norge delt inn i fire prissoner hvor Oslo og Bærum inngikk i den dyreste prissonen. Endringen nå er at Oslo og Bærum alene utgjør den dyreste prissonen, sone 1. De resterende kommunene som tilhørte sone 1 tidligere, er nå forskjøvet til sone 2 osv. Dette ble gjort fordi det viste seg at prisutviklingen i sone 1 var korrelert med andelen boliger i Oslo og Bærum i de ulike kvartalene. Det å skille ut Oslo og Bærum som en egen sone viste seg å ha en god effekt på modellen og ga en mer korrekt prisutvikling i måleperioden. Ulempen med å dele inn i ytterligere prissoner er at det blir færre observasjoner i hver sone. Figur 4.2 viser prisnivået i de ulike sonene, hvor Oslo og Bærum er skilt ut i en egen sone. Soneinndelingen er vist i vedlegget.

Figur 4.2 Salgssum ekskludert tomt i de ulike sonene



Tabell 4.3 viser gjennomsnittlig salgssum ekskludert tomt fordelt på de fem ulike sonene i perioden 1. kvartal 2019 til 4. kvartal 2020. Som forventet er den gjennomsnittlige salgssummen vesentlig lavere i sone 5 sammenlignet med sone 1. Sone 1 og sone 5 skiller seg ut med lavest andel transaksjoner.

Tabell 4.3 Gjennomsnittlig salgssum ekskludert tomt fordelt på sonene. 2019K1 – 2020K4

Sone	Gjennomsnittlig salgssum	Observasjoner
1	10 022 993	209
2	6 152 481	441
3	4 795 633	799
4	4 019 970	704
5	3 752 325	265
Hele landet	5 154 750	2 418

Tabell 4.4 viser gjennomsnittlig salgssum inkludert tomt fordelt på de ulike sonene.

Tabell 4.4 Gjennomsnittlig salgssum inkludert tomt fordelt på sonene. 2019K1 – 2020K4.

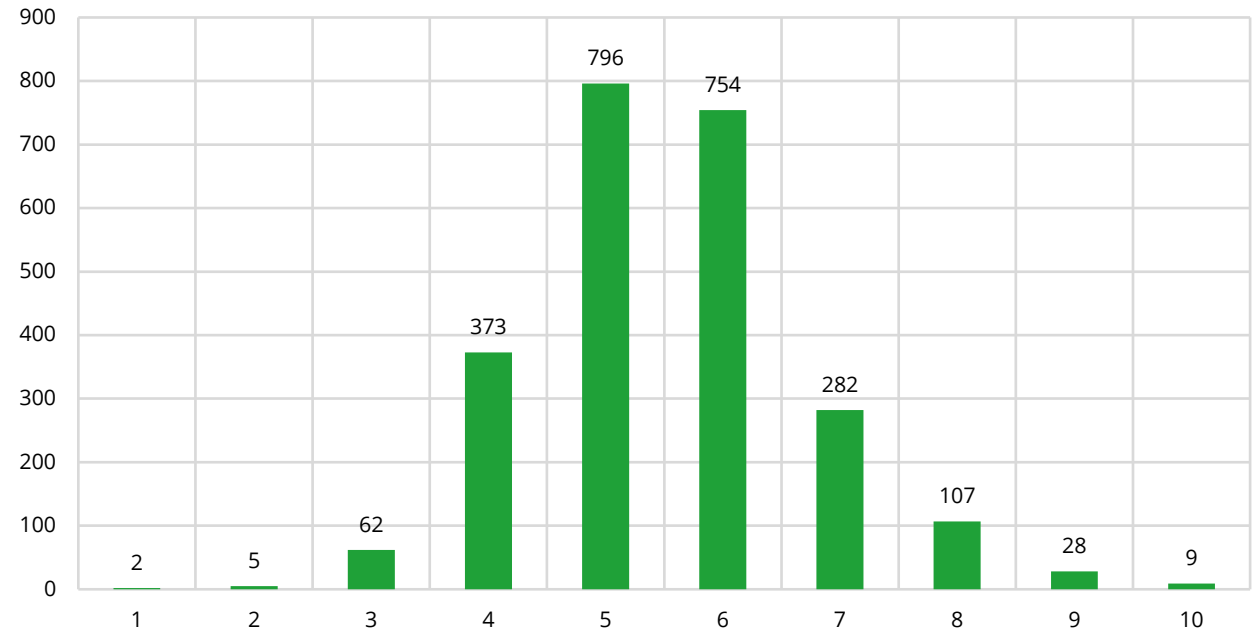
Sone	Gjennomsnittlig salgssum	Observasjoner
1	14 463 292	209
2	8 287 236	441
3	6 080 784	799
4	5 020 759	704
5	4 476 428	265
Hele landet	6 723 289	2 418

4.4. Antall rom, WC og bad

Antall rom, bad og wc som eneboligen inneholder, hentes fra matrikkelen.

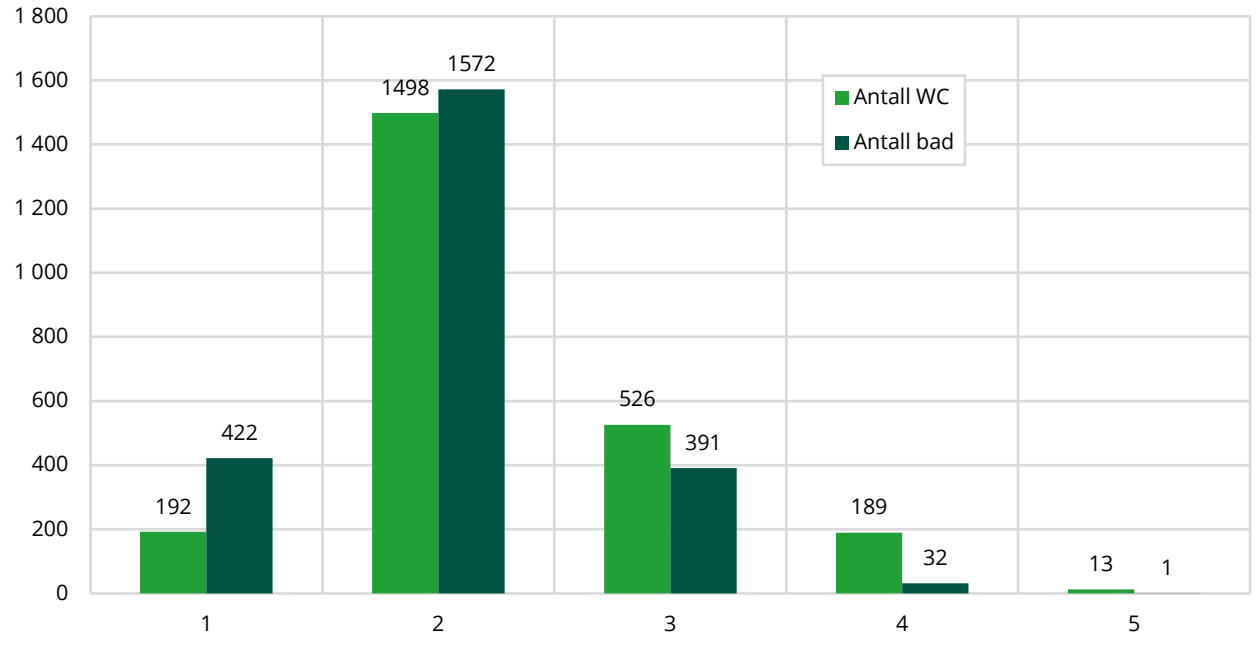
Figur 4.3 viser romfordelingen i populasjonen, klart flest nye eneboliger har 5 – 6 rom. Ingen av observasjonene manglet rom.

Figur 4.3 Romfordeling. 2019K1 – 2020K4



Figur 4.4 viser fordelingen av bad og WC i utvalget. Vi ser at normalen i populasjonen ligger på to bad og to WC.

Figur 4.4 Antall bad og WC. 2019K1 – 2020K4



Enkelte av observasjonene hadde ingen bad eller WC registrert. For å unngå å droppe disse observasjonene (som ellers har god kvalitet) fra populasjonen blir antall bad og WC imputert basert på bruksarealet til boligen. Tabell 4.5 viser gjennomsnittsstørrelsen på boligene etter antall bad og tilsvarende for WC.

Tabell 4.5 Gjennomsnittlig bruksareal fordelt på bad og WC

Antall bad	Gjennomsnittlig bruksareal	Antall WC	Gjennomsnittlig bruksareal
1	139	1	133
2	181	2	169
3	228	3	212
4	258	4	244
5	255	5	280

Basert på oversikten i tabell 4.5 ble følgende imputeringsregel brukt for de 11 observasjonene som manglet bad;

- Om bruksareal er mindre enn 145 kvadratmeter har boligen 1 bad.
- Om bruksareal er mellom 145 og 199 kvadratmeter har boligen 2 bad.
- Om bruksareal er mellom 200 og 249 kvadratmeter har boligen 3 bad.
- Om bruksareal er over 250 kvadratmeter har boligen 4 bad

Følgende imputeringsregel gjelder for de 24 observasjonene som manglet WC;

- Om bruksareal er mindre enn 135 kvadratmeter har boligen 1 WC.
- Om bruksareal er mellom 135 og 179 kvadratmeter har boligen 2 WC.
- Om bruksareal er mellom 180 og 219 kvadratmeter har boligen 3 WC.
- Om bruksareal er mellom 220 og 259 kvadratmeter har boligen 4 WC.
- Om bruksareal er over 260 kvadratmeter har boligen 5 WC.

4.5. Antall bruksenheter

Eneboliger som er registrert i matrikkelen med bygningstype «enebolig med hybelleilighet, sokkelleilighet o.l.» antas å ha en eller flere utleieleiligheter. Tabell 4.4 viser at kun 5,13 prosent av populasjonen har flere enn 1 bruksenhet. For å vurdere om eneboligen har flere bruksenheter, tas det utgangspunkt i antall bruksenheter registrert i boligen ikke bygningstypen.

Tabell 4.6 Antall bruksenheter

Bruksenheter	Antall	Prosent	Kumulativ frekvens	Kumulativ prosent
1	2294	94,87 %	2294	94,87 %
2	122	5,05 %	2416	99,92 %
3	2	0,08 %	2418	100,00 %

5. Indeksberegning

5.1. Modell og funksjonsform

SSB har brukt den hedoniske modellen for prisindeks for nye eneboliger helt siden starten i 1990. Regresjonsmodellen har blitt justert i flere tilfeller ved at nye forklaringsvariabler er tatt inn mens andre er tatt ut ettersom krav til boligene og boligbyggernes prioriteringer har endret seg. I 2015 ble også funksjonsformen i modellen endret fra lineær til log-lineær sammen med at avhengig variabel ble endret fra kvadratmeterpris til totalprisen for boligen. Ny prisfunksjon er da på formen beskrevet i formel (3.3) i kapittel 3.2.

5.2. Forklaringsvariabler

Prisindeks for nye eneboliger måler prisen for å sette opp en ny enebolig eksklusiv tomtepris. Indeksen avviker fra Eurostat sin forordning med at den ikke inkluderer tomteverdi etter ønske fra Finans Norge, som er oppdragsgiver til statistikken. Indeksen er en tilnærming til en salgsprisindeks ettersom merverdiavgift og fortjenestemargin er inkludert i prisen. I teorien inkluderes alle kostnader i denne indeksen bortsett fra tomteprisen.

Forklaringsvariablene i modellen har som nevnt blitt endret over tid, og nåværende modell gjelder fra og med første kvartal 2021. Tabell 5.1 viser variabellisten for gjeldende modell med utgangspunkt i data for 8 kvartaler i perioden 2019-2020. Soneinndelingen er såkalte dummyvariabler, dvs. de er enten 0 eller 1. For egenskaper med flere svaralternativer settes et av alternativene som referanseverdi. Sonen med høyest prisnivå er satt som referanseverdi, sone 1, som en følge får koeffisientene til de andre sonene et negativt fortegn. De resterende fortegnene på koeffisientene er positive.

Da variabelnavnene ikke alltid er selvforklarende henvises det i første kolonne til variabelbeskrivelsen i kapittel 4. «T - verdi» viser hvor signifikant variabelen er i modellen, dvs. en høy absolutt verdi indikerer høy signifikans. En tommelfingerregel tilsier at en absolutt verdi på 2 og oppover indikerer at variabelen er signifikant. Alle koeffisientene har en signifikant effekt på den endelige salgssummen, med unntak av variablene antall boliger og antall rom.

Tabell 5.1 Variabelliste

Henvisning til kap. 4	Variabel	DF	Parameter-estimat	Standard-avvik	T-verdi	Pr > |t|
	Konstant	1	12,74671	0,11044	115,42	<.0001
4.3	sone_2	1	-0,36639	0,02147	-17,06	<.0001
4.3	sone_3	1	-0,52795	0,02092	-25,24	<.0001
4.3	sone_4	1	-0,66093	0,02178	-30,35	<.0001
4.3	sone_5	1	-0,69142	0,02508	-27,57	<.0001
4.2	ln_bruks	1	0,55671	0,02513	22,15	<.0001
4.4	antall_bad	1	0,04129	0,01201	3,44	0.0006
4.4	antall_wc	1	0,05224	0,01095	4,77	<.0001
4.5	antbol	1	0,00627	0,02253	0,28	0.7810
4.4	Antall_rom	1	0,00226	0,00531	0,42	0,6709

5.3. Regresjonsberegninger

5.3.1 Prisfunksjonen

Prisfunksjonen inkluderer forklaringsvariablene i tabell 5.1 og kan skrives slik på generell form:

$$lnP_t = \beta_t^0 + \sum_{k=1}^{9} \beta_t^k z_t^k + \epsilon \qquad (5.1)$$

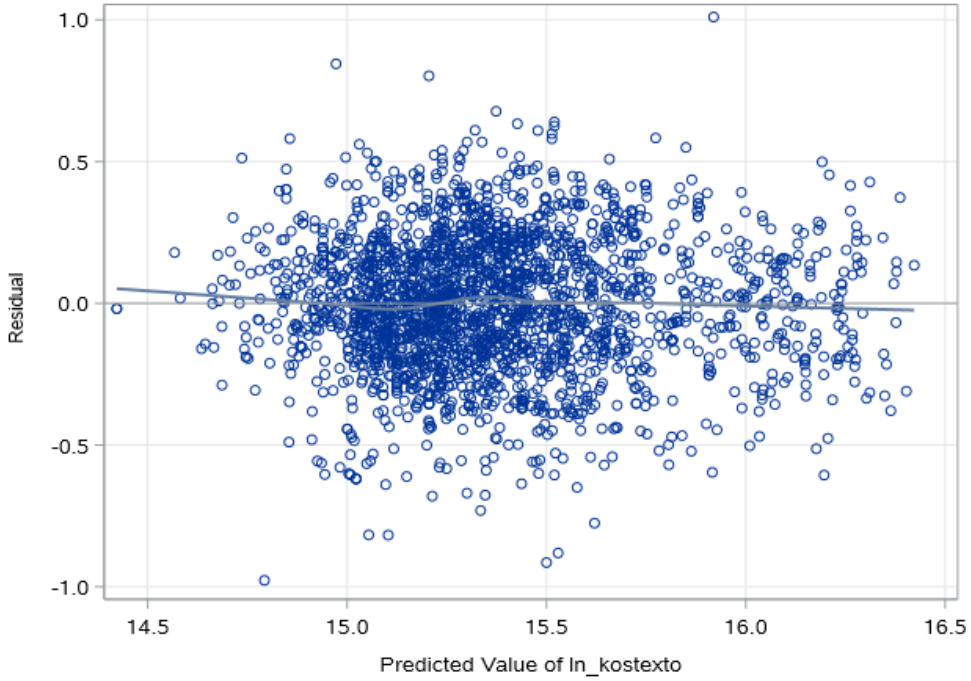
P er totalkostnaden for boligen uten tomt og β_t^0 er konstantleddet. Regresjonsresultatet ser vi i tabell 5.1 der parameterestimatet angir koeffisientverdier og standardavvik knyttet til de ulike variablene. Selv med færre forklaringsvariabler sammenlignet med tidligere modell har forklaringskraften R^2 i regresjonsmodellen økt. Forklaringskraften ligger på 0,67 mot 0,63 tidligere. Mulig årsak til den økte forklaringskraften er at vi får et mer homogent utvalg ved å ekskludere såkalte selvbyggere. En annen forklaring er at potensielle feilkilder er betydelig lavere ved registerdata sammenlignet med skjemaundersøkelser. Ved feilkilder menes for eksempel at enkelte velger å ikke svare på undersøkelsen, at enkelte rapporterer feil opplysninger osv. Av den grunn kan det antas at kvaliteten på dataene er økt, og med det også økt forklaringskraft.

5.3.2 Usikkerhet

Selv om forklaringskraften har økt med den nye modellen vil det fortsatt være prisbestemmende faktorer som ikke fanges opp. Eksempler på dette kan være faktorer ved bygningen som materialvalg og standard. Det kan også tenkes at beliggenhet har en betydning for prisene utover tomteprisene som f.eks. nærhet til kyst, avstand til skole, butikker og annen infrastruktur. Observasjoner med feilrapporterte verdier kan også virke inn på forklaringskraften.

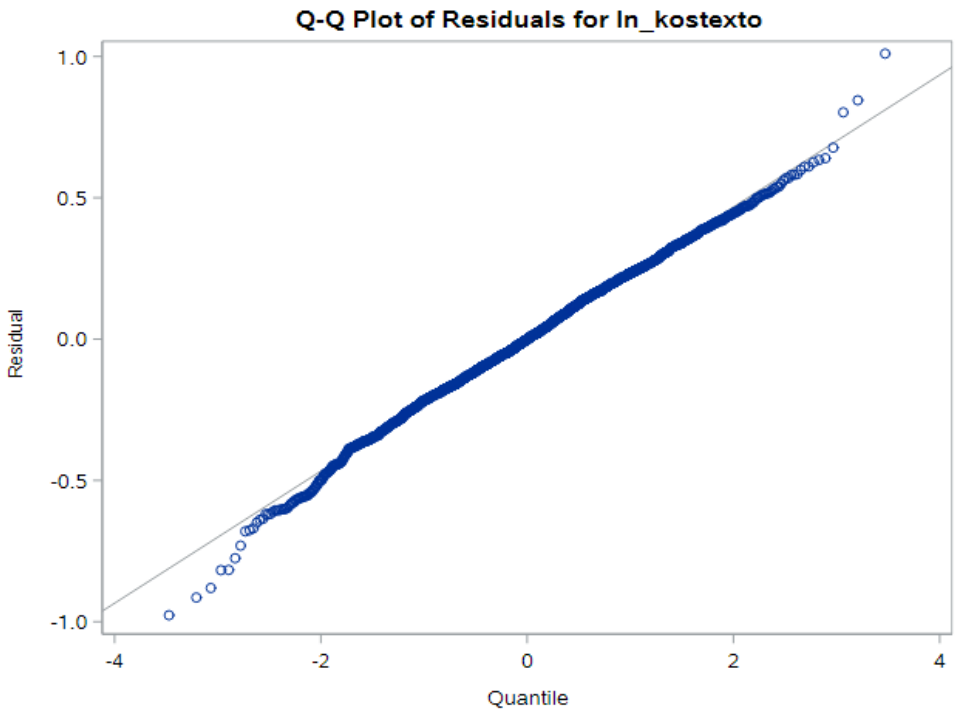
En av kontrollene for å si noe om kvaliteten til regresjonsmodellen er å se på residualene, det vil si differansen mellom observert og predikert pris. Predikert pris er den prisen vi får beregnet for hver av observasjonene ved bruk av regresjonsligningen. Residualene bør være tilnærmet normalfordelt med forventning lik 0 og konstant varians. Figurene under er hentet fra regresjonen beskrevet i tabell 5.1. Figur 5.1 viser avviket mellom predikert verdi og faktisk verdi for hver transaksjon. Variansen skal være konstant, uavhengig av pris. Om variansen øker i takt med salgsprisen eller omvendt er det tegn på heteroskedastisitet som betyr at en av antakelsene i MKM ikke holder. Figur 5.1 viser ingen sterke tegn til heteroskedastisitet og av den grunn kan vi anta at vi har konstant varians og homoskedastistitet.

Figur 5.1 Plott av residualer mot predikerte verdier.



Videre kan man lage normalplott av residualene for å se om de er normalfordelte, figur 5.2. Observasjonene bør da ligge på en linje, nærmest mulig den rette linjen som representerer normalfordelingen. Figur 5.2 viser at residualene er tilnærmet normalfordelte. Vi ser at residulene ikke er perfekt normalfordelt, men den oppfyller likevel kravene til en lineær regresjon. Det er vanlig i eiendomstransaksjoner å finne små avvik fra normalfordelingen.

Figur 5.2 Normalplott av residualer



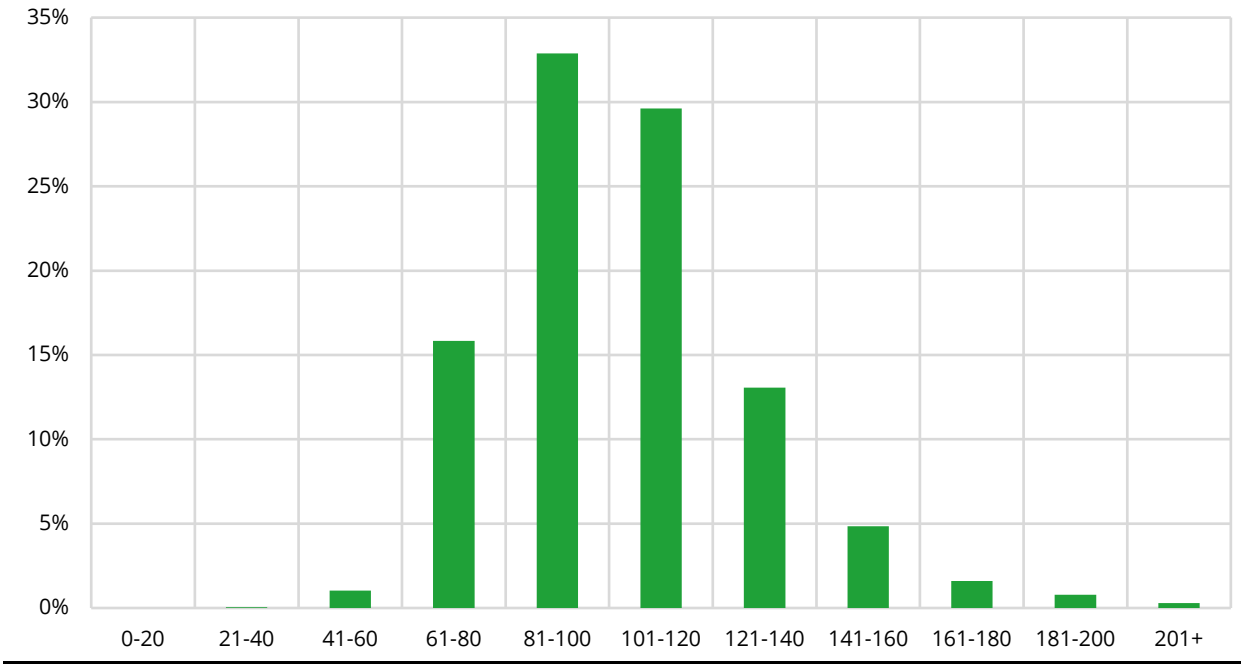
Vi kan sjekke treffsikkerheten til regresjonsmodellen ved å sammenligne estimert markedsverdi fra regresjonen mot observert omsetningsverdi. En svakhet med modellen er at den inneholder forholdsvis få forklaringsvariabler. Det er ingen forklaringsvariabler som sier noe om kvaliteten på materialene som er brukt, vi klarer ikke å plukke opp om noen velger å ha et gullbelagt baderom. Selv om to eneboliger ligger i samme kommune, har samme bruksareal og likt antall bad, wc og rom, vil det likevel kunne være store forskjeller på standard og byggekostnader. Derfor er det interessant å sjekke hvor godt modellen treffer på estimeringen. Tabell 5.2 viser den prosentvise fordelingen av forholdet mellom estimert salgspris og faktisk observert salgspris. Den samme fordelingen er vist i figur 5.3. I tabell 5.2 representerer gruppene under «Estimert/Observert» hvor mange prosent den estimerte verdien utgjør av den faktiske omsetningsverdien. Observasjoner i gruppe 81 – 100 har en estimert pris som utgjør 81 – 100 prosent av den observerte omsetningsverdien.

Ut i fra tabell 5.2 og figur 5.3 kan vi se at majoriteten, 62,34 prosent, av observasjonene ligger i gruppene 81-100 og 101-120, som vil si at de estimerte markedsverdiene ligger innenfor pluss minus 20 prosent av de observerte omsetningsverdi. Eneboligene med de største avvikene i estimert og faktisk markedspris, er i hovedsak boligene med svært høye eller lave markedspriser. Det er verdt å merke seg at selv om modellen hovedsakelig treffer godt, er hedoniske modeller i boligprisindeksen ikke først og fremst beregnet for å evaluere enkelte eiendommer, det viktigste er at den samlede eller gjennomsnittlige kvaliteten på de omsatte boligene kan estimeres tilstrekkelig (Brand 2020).

Tabell 5.2 Prosentvis fordeling av forholdet mellom den estimerte markedsprisen over den faktisk observerte markedsprisen. 2019K1 – 2020K4

Estimert / Observert	Antall	Prosent	Kumulativ prosent
0-20	0	0,00 %	0,00 %
21-40	1	0,04 %	0,04 %
41-60	25	1,03 %	1,07 %
61-80	383	15,84 %	16,91 %
81-100	795	32,88 %	49,79 %
101-120	716	29,61 %	79,40 %
121-140	316	13,07 %	92,47 %
141-160	117	4,84 %	97,31 %
161-180	39	1,61 %	98,92 %
181-200	19	0,79 %	99,71 %
201+	7	0,29 %	100,00 %

Figur 5.3 Prosentvis fordeling av forholdet mellom den estimerte markedsprisen over den faktisk observerte markedsprisen. 2019K1 – 2020K4

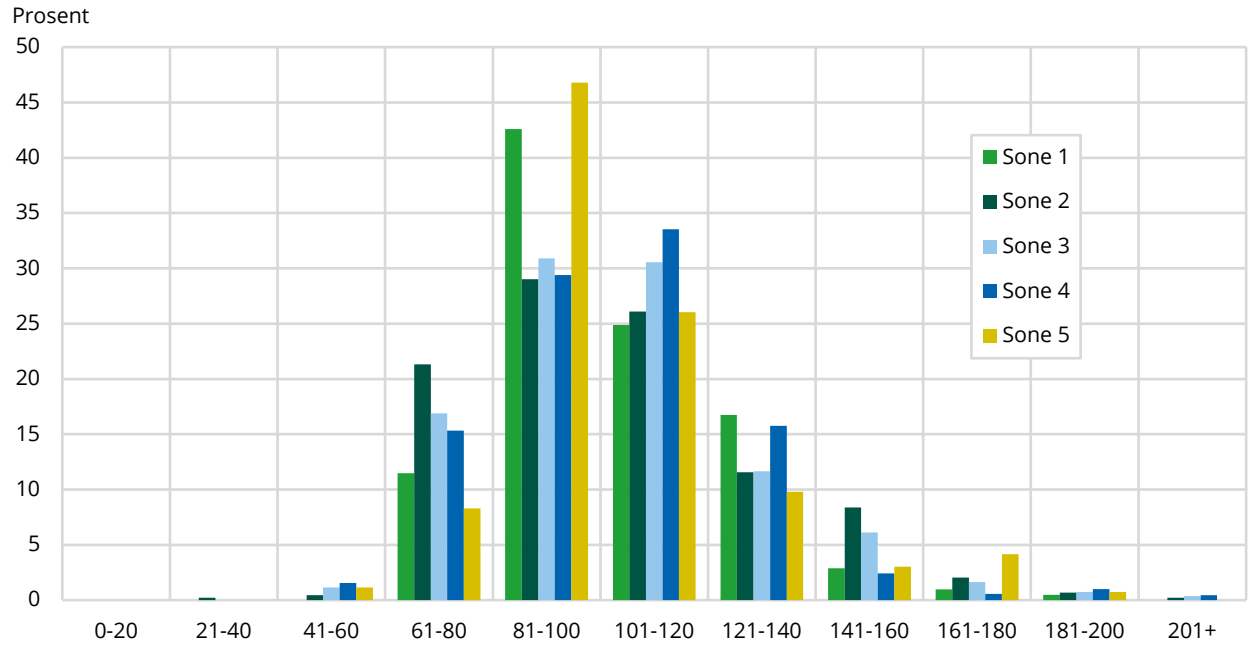


Videre undersøkes det om modellen treffer dårligere eller bedre i enkelte soner. Tabell 5.3 viser det samme som tabell 5.2, men her er det gruppert på soner.

Tabell 5.3 **Forholdet mellom estimert markedspris over faktisk observerte markedspris. Antall observasjoner i ulike kategorier, fordelt på soner. 2019K1 – 2020K4**

	Soner					
Estimert/ Observert	1	2	3	4	5	Total
20-40	0	1	0	0	0	1
41-60	0	2	9	11	3	25
61-80	24	94	135	108	22	383
81-100	89	128	247	207	124	795
101-120	52	115	244	236	69	716
121-140	35	51	93	111	26	316
141-160	6	37	49	17	8	117
161-180	2	9	13	4	11	39
181-200	1	3	6	7	2	19
201+	0	1	3	3	0	7
Total	209	441	799	704	265	2418

Figur 5.4 **Prosentvis fordeling av observasjoner etter forholdet mellom estimert markedspris over faktisk observert markedspris. Fordelt på soner. 2019K1 – 2020K4**



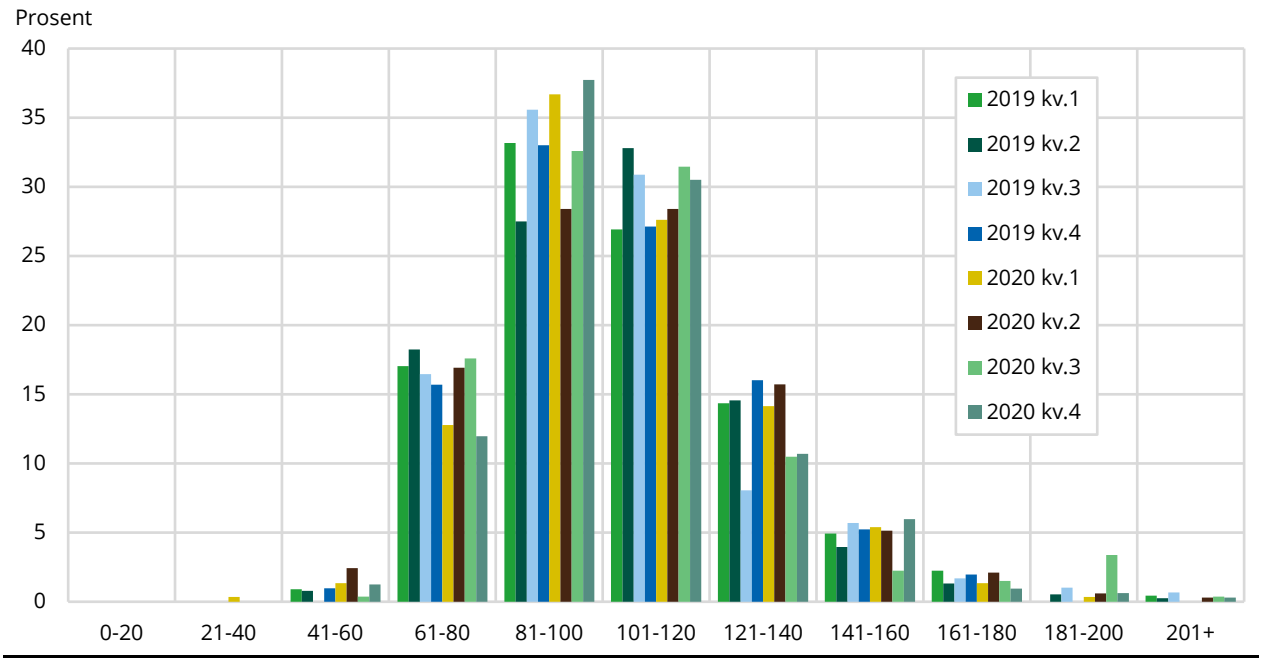
Som vi ser fra tabell 5.3 og figur 5.4 ligger majoriteten av observasjonene også her i gruppene 81-100 og 101-120 i alle de fem forskjellige sonene. Ut i fra disse resultatene er det ingenting som tyder på at modellen er mindre treffsikker i enkelte soner. Figur 5.4 viser klarere at sone 1 og sone 5 har høyere andel som treffer godt på estimert pris i prosentgruppen 81 – 100.

Den samme øvelsen ble gjort for å sjekke om det var enkelte perioder som var mindre treffsikre. Resultatet vises i tabell 5.4.

Tabell 5.4 **Forholdet mellom estimert markedspris over faktisk observert markedspris. Antall observasjoner i ulike kategorier, fordelt på kvartaler. 2019K1 – 2020K4**

	Perioder								
Estimert Observert	2019 kv.1	2019 kv.2	2019 kv.3	2019 kv.4	2020 kv.1	2020 kv.2	2020 kv.3	2020 kv.4	Total
21-40	0	0	0	0	1	0	0	0	1
41-60	2	3	0	3	4	8	1	4	25
61-80	38	69	49	48	38	56	47	38	383
81-100	74	104	106	101	109	94	87	120	795
101-120	60	124	92	83	82	94	84	97	716
121-140	32	55	24	49	42	52	28	34	316
141-160	11	15	17	16	16	17	6	19	117
161-180	5	5	5	6	4	7	4	3	39
181-200	0	2	3	0	1	2	9	2	19
201+	1	1	2	0	0	1	1	1	7
Total	223	378	298	306	297	331	267	318	2418

Figur 5.5 **Prosentvis fordeling av observasjoner etter forholdet mellom estimert markedspris over faktisk observert omsetningspris. År og kvartal. 2019K1 – 2020K4**



Resultatet fra tabell 5.4 og figur 5.5 viser samme resultat, det er ingenting som tyder på at modellen er mindre treffsikker i en spesiell periode.

5.5. Brudd i statistikken

Dette notatet beskriver endringer i rutinene for å beregne prisindeks for nye eneboliger, altså et brudd i statistikken. Det at datagrunnlaget ikke lenger inkluderer selvbyggere antas først og fremst å ha noe å si for nivået på kostnadene (f.eks. ved at selvbyggere har underestimert verdien av sin egen arbeidsinnsats). Bruddet er kjedet inn i indeksserien ved at 4. kvartal 2020 først ble beregnet med den gamle metoden, deretter med den nye. Den nye indeksen blir deretter multiplisert med forholdet mellom gammel og ny indeks (formel 3.14).

6. Indeksberegninger

6.1. Indeksformel

Vi benytter prisfunksjonen i formel (5.1) og antar at koeffisientene β^1 til β^9 er konstante over kortere tidsperioder samt at residualen ϵ forventes å ha konstant varians og forventningsverdi lik 0.

Videre brukes karakteristikk-indeksmetode beskrevet i kapittel 3.3.2. Gitt at koeffisientene er konstante over kortere tidsperioder kan indeksformelen skrives slik den er formulert i formel (3.12).

$$I_{IP} = \frac{\prod_{h=1}^{H_t} (p_{th})^{1/H_t}}{\prod_{h=1}^{H_0} (p_{0h})^{1/H_0}} \Big/ \exp\left[\Sigma_k \hat{\beta}_0^k \left(\bar{z}_t^k - \bar{z}_0^k\right)\right]$$

der P_{IP} er den implisitte Paasche indeksen mellom periode *0* og *t*, p_h er boligens totale kostnader uten tomt, β_0^k er koeffisientene fra periode 0 og $\bar{z}^k$ er gjennomsnittsverdiene for de ulike variablene. I vårt tilfelle er den avhengige variabelen p_h på logaritmisk form og den må omformes til normalform ved en eksponentialfunksjon. Formelen ovenfor kan da forenkles til:

$$I_{IP} = \frac{\exp\left(\ln\left(p_t\right)\right)}{\exp\left(\ln\left(p_0\right) + \sum_k \hat{\beta}_0^k \left(\bar{z}_t^k - \bar{z}_0^k\right)\right)} \tag{6.1}$$

6.2. Kjeding

Indeksene beregnes som kjedete indekser med årlige lenker. Basis skiftes i 1. kvartal, med foregående år som nytt basisår. Priskoeffisientene revideres hvert år basert på de to siste årenes datagrunnlag. Koeffisientene antas å være konstante i ett-årsperioden.

Kjedet indeks for andre kvartal kan generelt beskrives med formel (3.14):

$$I_{kjedet}^t = I_{ny\ base}^t \frac{I_{gammel\ base}^{t-1}}{I_{ny\ base}^{t-1}}$$

Indeksen beregnes først med gammel basis. Deretter beregnes den på nytt med ny basis. Den kjedete indeksen blir dermed den nye indeksen multiplisert med forholdet mellom gammel og ny indeks. Eventuelle oppdateringer i beregningsmetoden, som for eksempel endring i grenseverdiene, legges også til kjedingstidspunktet.

7. Tilgrensende statistikker

7.1. Byggekostnadsindeksen

Byggekostnadsindeksen for boliger er en input prisindeks og måler prisutviklingen på innsatsfaktorene i byggeproduksjonen, slik som arbeidskraft, material, transport og maskiner. Byggekostnadsindeksen for boliger avviker fra prisindeksen for nye boliger ved at den ikke omfatter viktige elementer som påvirker output-prisen, nemlig produktivitetsendringer og endringer i entreprenørens fortjenestemarginer.

7.2. Kvadratmeterpris for eneboliger

Kvadratmeterpriser for nye eneboliger er publisert siden 1999 og har som formål å vise forskjeller mellom kvadratmeterpriser for nye og brukte eneboliger. For å kunne sammenligne priser for brukte og nye eneboliger omregnes bruksarealet til p-rom slik at samme arealbegrep brukes. Videre inkluderes tomteverdien for de nye eneboligene da den inngår i prisen ved bruktboligsalg. På grunn av få observasjoner er det uvisst om denne statistikken kan videreføres etter omleggingen beskrevet i dette notatet.

7.3. Prisindeks for nye boliger

I 2015 ble det for første gang publisert en prisindeks for nye flerboliger i SSB. Denne prisindeksen er utarbeidet etter samme modell og funksjonsform som prisindeks for nye eneboliger. I 2016 ble prisindeks for nye eneboliger og nye flerboliger presentert i en felles publisering for nye boliger.

7.4. Prisindeks for boliger

I SSBs forpliktelser overfor Eurostat ligger det å levere en prisindeks for boliger med delindekser for nye og brukte boliger. Denne indeksen vektes med omsetningsvekter, dvs. den er vektet etter omsatte boliger og skiller seg derfor fra SSBs prisindeks for brukte boliger som vektes med bestandsvekter (boligbestanden). Det planlegges en nasjonal publisering av prisindeks for boliger der prisindeksen for nye boliger og prisindeksen for brukte boliger vektes sammen med bestandsvekter og dermed vil skille seg noe fra varianten vi sender til Eurostat. Hovedgrunnen til at den nye publiseringen bruker bestandsvekter er at bruktboligprisen inngår i SSBs KVARTS-modell, der formuesaspektet knyttet til boligene er sentralt. Det er derfor ønskelig at publiserte tall er sammenlignbare med framskrivningene som gjøres i denne modellen.

7.5. Videre arbeid

Ettersom vi har tilgang til tomteprisen gjennom avgiftsgrunnlagvariabelen, kan vi undersøke mulighetene for å lage en egen indeks for enebolig med tomt. Som tidligere nevnt ønsker oppdragsgiver for statistikken en indeks uten tomt, mens Eurostat ønsker en prisindeks for nye boliger som inkluderer tomt[3]. Hvorvidt de to indeksene vil utvikle seg forskjellig er uvisst, svaret ligger i hvordan tomteprisene utvikler seg i forhold til prisen på selve bygningsstrukturen. En annen mulighet kan være å utvikle en egen tomteprisindeks. I tillegg til de nybebygde tomtene som er beskrevet i dette notatet, vil man også kunne inkludere salg av ubebygde tomter i en slik indeks. Dette er områder som SSB per i dag ikke har statistikk om, men som det er stor interesse for.

[3] "...the HPI does not follow the net acquisition concept and the price of land is included in both prices and weights." (Eurostat, 2017)

Referanser

Brand, Manuel (2020): Swiss residential property price index. Quality adjustment procedures. Federal Statistical Office

Eurostat (2017): Technical manual on Owner-Occupied Housing and House Price Indices. Eurostat

Eurostat (2013): Handbook on Residential Property Prices Indices (RPPI). Eurostat Methodologies and Working papers

Haglund, Anders (2018): Prisindeks for nye eneboliger. Rapporter 2018/4. Statistisk Sentralbyrå

Lillegård, Magnar (1994): Prisindekser for boligmarkedet. Rapporter 94/7. Statistisk Sentralbyrå

Rosen, S. (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy 82.

Skatteetaten (2021): Årsrundskriv for dokumentavgift 2021.

Wass, Kurt Åge (1992): Prisindeks for nye enebolig. Rapporter 92/21. Statistisk sentralbyrå.

Wigren, R. (1986): Småhuspriserna i Sverige. Forskningsrapport från Statens Institut för Byggnadsforskning.

Vedlegg A: Prissoner

Prissone 1	Prissone 4
0301 Oslo	3001 Halden
3024 Bærum	3003 Sarpsborg
	3007 Ringerike
Prissone 2	3026 Aurskog-Høland
1103 Stavanger	3035 Eidsvoll
1124 Sola	3047 Modum
1127 Randaberg	3401 Kongsvinger
1804 Bodø	3405 Lillehammer
3022 Frogn	3411 Ringsaker
3023 Nesodden	3412 Løten
3025 Asker	3420 Elverum
3027 Rælingen	3442 Østre Toten
3029 Lørenskog	3443 Vestre Toten
3049 Lier	3802 Holmestrand
4601 Bergen	3813 Bamble
5401 Tromsø	3814 Kragerø
	4202 Grimstad
Prissone 3	4203 Arendal
1101 Egersund	4204 Kristiansand
1108 Sandnes	4213 Tvedestrand
1120 Klepp	4215 Lillesand
1121 Time	4612 Sveio
1130 Strand	4614 Stord
1133 Hjelmeland	4622 Kvam
1507 Ålesund	4624 Bjørnafjorden
1516 Ulstein	4626 Øygarden
1532 Giske	4631 Alver
1547 Aukra	4647 Sunnfjord
1554 Averøy	4649 Stad
1806 Narvik	5001 Trondheim
1813 Brønnøy	5006 Steinkjer
1841 Fauske	5007 Namsos
3002 Moss	5014 Frøya
3004 Fredrikstad	5020 Osen
3005 Drammen	5021 Oppdal
3006 Kongsberg	5022 Rennebu
3011 Hvaler	5025 Røros
3018 Våler (Østfold)	5026 Holtålen
3019 Vestby	5027 Midtre Gauldal
3020 Nordre Follo	5028 Melhus
3021 Ås	5029 Skaun
3030 Lillestrøm	5031 Malvik
3031 Nittedal	5032 Selbu
3032 Gjerdrum	5033 Tydal
3033 Ullensaker	5034 Meråker
3036 Nannestad	5035 Stjørdal
3038 Hole	5036 Frosta
3054 Lunner	5037 Levanger
3403 Hamar	5038 Verdal
3407 Gjøvik	5041 Snåase-Snåsa
3413 Stange	5042 Lierne

Prissone 3
3801 Horten
3803 Tønsberg
3804 Sandefjord
3806 Porsgrunn
3807 Skien
4602 Kinn
4625 Austevoll
4627 Askøy
4640 Sogndal
5402 Harstad
5418 Målselv

Prissone 4
1106 Haugesund
1505 Kristiansund
1149 Karmøy
1506 Molde
1515 Herøy
1520 Ørsta
1531 Sula
1535 Vestnes
1577 Volda
1579 Hustadvika
1811 Bindal
1812 Sømna
1813 Brønnøy
1815 Vega
1816 Vevelstad
1818 Herøy
1820 Alstahaug
1822 Leirfjord
1824 Vefsn
1825 Grane
1826 Hattfjelldal
1827 Dønna
1828 Nesna
1832 Hemnes
1833 Rana
1834 Lurøy
1835 Træna
1836 Rødøy
1837 Meløy
1838 Gildeskål
1839 Beiarn
1840 Saltdal
1845 Sørfold
1848 Steigen
1851 Lødingen
1853 Evenes
1856 Røst
1857 Værøy
1859 Flakstad
1860 Vestvågøy
1865 Vågan
1866 Hadsel
1867 Bø
1868 Øksnes
1870 Sortland
1871 Andøy
1874 Moskenes
1875 Hamarøy

Prissone 4
5043 Raarvihke Røyrvik
5044 Namsskogan
5045 Grong
5046 Høylandet
5047 Overhalla
5049 Flatanger
5052 Leka
5053 Inderøy
5054 Indre Fosen
5055 Heim
5056 Hitra
5057 Ørland
5058 Åfjord
5059 Orkland
5060 Nærøysund
5061 Rindal
5404 Vardø
5405 Vadsø
5406 Hammerfest
5411 Kvæfjord
5412 Tjeldsund
5413 Ibestad
5414 Gratangen
5415 Lavangen
5416 Bardu
5417 Salangen
5419 Sørreisa
5420 Dyrøy
5421 Senja
5422 Balsfjord
5423 Karlsøy
5424 Lyngen
5425 Storfjord
5426 Kåfjord
5427 Skjervøy
5428 Nordreisa
5429 Kvænangen
5430 Guovdageaidnu Kautokeino
5432 Loppa
5433 Hasvik
5434 Måsøy
5435 Nordkapp
5436 Porsanger Porsángu Porsanki
5437 Karasjohka Karasjok
5438 Lebesby
5439 Gamvik
5440 Berlevåg
5441 Deatnu Tana
5442 Unjarga Nesseby
5443 Båtsfjord
5444 Sør-Varanger

Prissone 5
Resterende kommuner

Figurregister

Tabellregister