

Usikkerhed ved opgørelse af udviklingen i boligpriser

Af Jakob Holmgaard

Da udviklingen på boligmarkedet har stor betydning for mange danskere og for den danske økonomi, er der ofte stor interesse for udviklingen i boligpriserne, som opgøres af Danmarks Statistik hvert kvartal. Ofte tillægges selv små ændringer stor betydning, selvom tallene er behæftet med usikkerhed.

Denne analyse sætter tal på den statistiske usikkerhed for det officielle kvartalsvise prisindeks for enfamiliehuse og ejerlejligheder.

Analysens hovedkonklusioner:

- Usikkerheden for den kvartalsvise prisudvikling på landsplan er +/- 1,1 procentpoint for enfamiliehuse og +/- 1,7 procentpoint for ejerlejligheder. Usikkerheden skal ses i forhold til, at prisudviklingen typisk er lille fra et kvartal til det næste.
- Usikkerheden er større for ejerlejligheder end for enfamiliehuse, fordi der sælges færre ejerlejligheder. Usikkerheden er større på regionalt niveau end på landsplan, da handlerne er færre.
- Usikkerheden er vigtig at tage med i betragtning, særligt når ændringerne fra kvartal til kvartal vurderes. Udviklingen over længere tidsperioder er i mindre grad påvirket af statistisk usikkerhed.

Udviklingen i boligpriserne er behæftet med usikkerhed

Hvert kvartal sætter Ejendomssalgstatistikken tal på prisudviklingen på ejerboligmarkedet opdelt på enfamiliehuse og ejerlejligheder. Der er ofte stor interesse for statistikken – og tit fokus på ret små udsving i tallene. I Ejendomssalgstatistikken offentliggøres der både foreløbige og endelige tal. Denne analyse vedrører alene de endelige tal, der er baseret på udtræk fra det elektroniske tinglysningsssystem¹.

Usikkerheden skyldes, at der er variation i hvilke boliger, der er handlet i et givet kvartal. Så selvom de faktiske priser på de solgte boliger er kendt, afhænger prisniveauet af de handlede boligkvalitet og antal. Det medfører en usikkerhed, som er vigtigt at tage i betragtning, når man bruger Ejendomssalgstatistikken.

Kvalitetsjustering og usikkerhedsberegning

Der er ikke stikprøveusikkerhed på de endelige tal, der ligger til grund for Ejendomsprisstatistikken, men derfor er der alligevel usikkerhed ved statistikken. Det skyldes at det ikke er de samme boliger, som sælges i forskellige kvartaler, så man kan ikke sammenligne boligpriserne i to forskellige kvartaler uden at kvalitetskorrigerer.

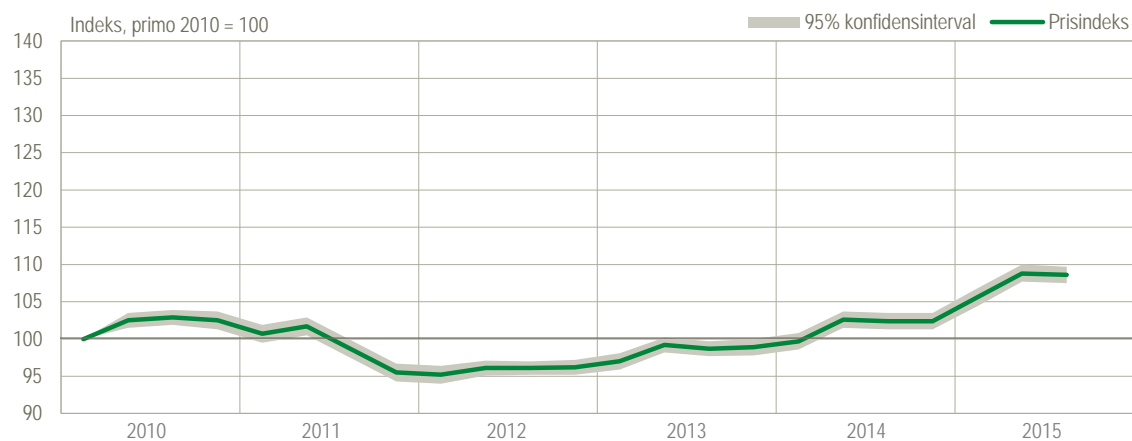
I Ejendomsstatistikken foretages kvalitetskorrektionen ved at måle prisen på en vurderingskrone, hvilket vil sige at en boligs salgspris sættes i forhold til boligens offentlige ejendomsvurdering. Denne kvotient kaldes afstandsprocenten. Afstandsprocenten har to begrænsninger som et mål for kvalitet. For det første vil to identiske boliger i samme kvalitet og sammenlignelig beliggenhed ikke nødvendigvis blive handlet til samme pris på et bestemt tidspunkt, da der ikke er tale om et perfekt gennemsigtigt marked. For eksempel kan en bolig opnå en højere pris, hvis sælgeren har ventet på den 'rigtige køber', og der er altid et element af tilfældigheder i prisfastsættelsen. For det andet er den offentlige ejendomsvurdering behæftet med usikkerhed, så to identiske boliger ikke nødvendigvis har samme ejendomsværdi (se bilag).

For at beregne den statistiske usikkerhed på prisudviklingen, anvendes *bootstrapping* (se bilag). Denne metode gør det muligt at beregne et 95% konfidensinterval, der angiver øvre og nedre grænse for usikkerheden på den kvartalsvise boligprisudvikling. Dette interval betegnes i teksten som et usikkerhedsinterval og vil i 95 pct. af tilfældene indeholde den sande værdi af prisudviklingen. Du kan læse mere om kvalitetsjustering og datagrundlag for Ejendomsstatistikken i [statistikdokumentationen](#).

Usikkerheden på den langsigtede prisudvikling er begrænset

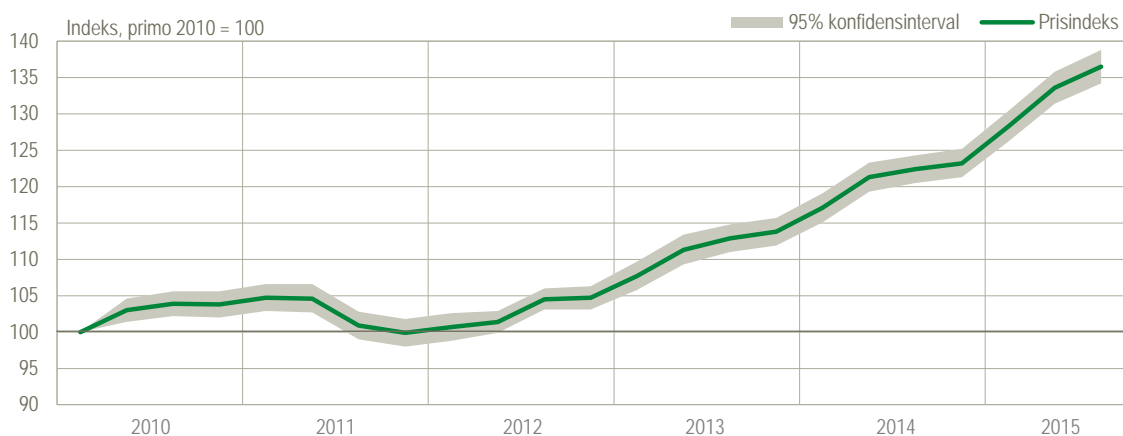
Der er usikkerhed på prisindekset for hhv. enfamiliehuse og ejerlejligheder, men ikke i et omfang, så det påvirker vurderingen af den langsigtede prisudvikling, jf. figur 1 og 2. For enfamiliehuse er bevægelserne i prisindekset henover et par år langt større end hvad, der kan tilskrives usikkerhed. Denne tendens er endnu tydeligere for ejerlejligheder, hvor den opadgående prisudvikling efter 2012 har været af et omfang, der langt overstiger usikkerheden.

Figur 1 Usikkerhedsinterval for prisindeks for enfamiliehuse på landsplan. 2010-2015



¹ Den særlige usikkerhed ved de foreløbige tal er beskrevet i [statistikdokumentationen](#).

Figur 2 Usikkerhedsinterval for prisindeks for ejerlejligheder på landsplan. 2010-2015

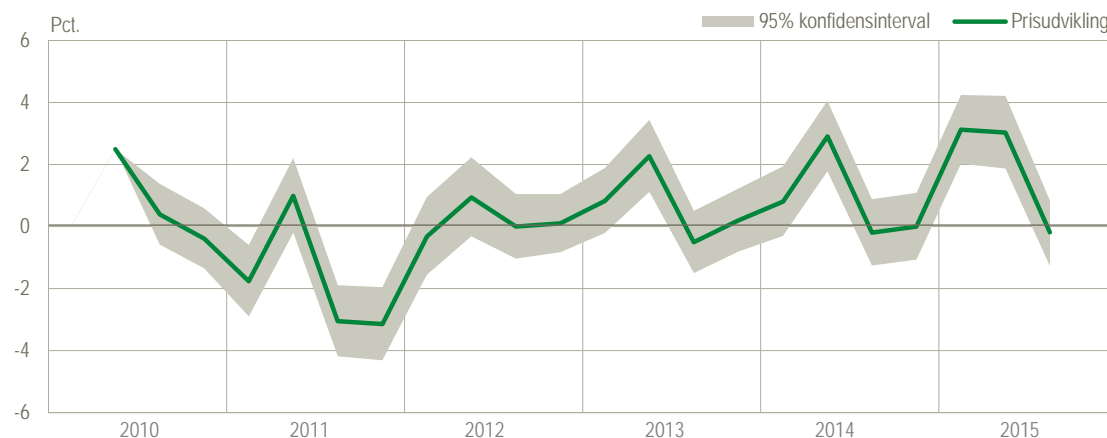


Der er usikkerhed for den kvartalsvise prisudvikling

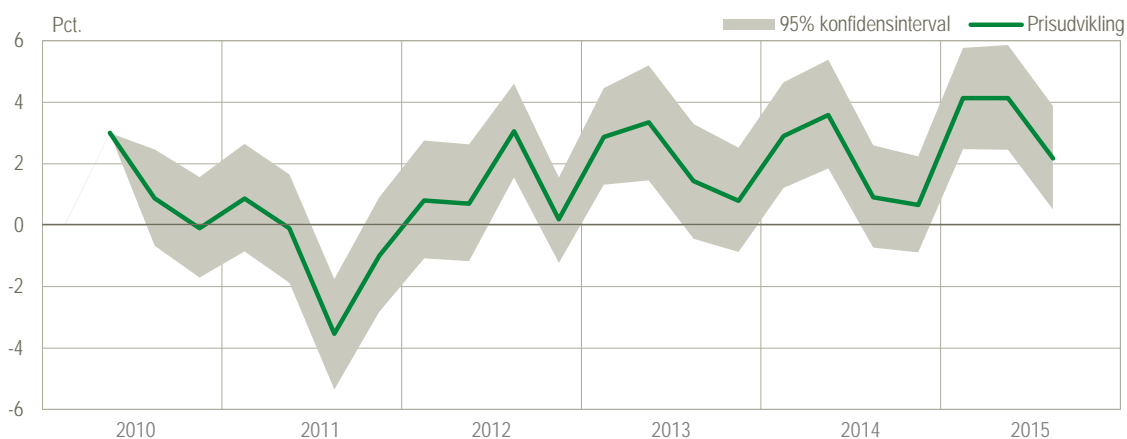
De fleste, der forsøger at tage temperaturen på boligmarkedet, vil fokusere på prisudviklingen fra kvartal til kvartal, snarere end på prisniveauet. I de seneste endelige offentliggjorte tal for 3. kvartal 2015, er prisændringen fra forrige kvartal stort set konstant for enfamiliehuse. Prisændringen blev opgjort til -0,2 pct., og usikkerhedsintervallet går fra -1,3 til 0,8 pct. Usikkerhedsintervallet indikerer at det ikke med stor sikkerhed er muligt at afgøre, om priserne for enfamiliehuse faldt eller steg i fra 2. til 3. kvartal i 2015. Dette gør sig gældende i 13 ud af 21 kvartaler, svarende til tre ud af fem kvartaler. Denne usikkerhed er en kombination af at ændringerne fra et kvartal til det næste er små, samtidig med at antallet af handler i hvert kvartal også er begrænsede.

I perioden fra 2. kvartal 2010 til 3. kvartal 2015, ligger den gennemsnitlige usikkerhed for den kvartalsvise prisudvikling på +/- 1,1 procentpoint for enfamiliehuse og +/- 1,7 procentpoint for ejerlejligheder – jf. figur 3 og 4, samt tabel 1.

Figur 3 Usikkerhedsinterval for prisudviklingen for enfamiliehuse. 2010K2 – 2015K3



Figur 4 Usikkerhedsinterval for prisudviklingen for ejerlejligheder. 2010K2 – 2015K3



Færre observationer betyder større usikkerhed ved regionale tal

Jo flere observationer, jo mindre bliver den statistiske usikkerhed. Dette betyder at usikkerheden på boligprisudviklingen er større når den opgøres på regionalt niveau sammenlignet med udviklingen på landsplan, jf. tabel 1. Prisudviklingen for ejerlejligheder udenfor Region Hovedstaden er særlig usikker, fordi antallet af ejerlejligheder og dermed antallet af handler er mindre udenfor Region Hovedstaden.

Tabel 1 Gennemsnitlig usikkerhed på prisudvikling for enfamiliehuse og ejerlejligheder

	Enfamiliehuse		Ejerlejligheder	
	usikkerhed (procentpoint)	antal	usikkerhed (procentpoint)	antal
Hele landet	1,1	7 989	1,7	2 762
Hovedstaden	2,2	1 976	2,3	1 730
Sjælland	2,4	1 251	6,1	153
Syddanmark	2,1	1 839	7,6	228
Midtjylland	1,9	1 968	3,6	461
Nordjylland	3,2	955	6,3	190

Anm: Gennemsnittet er beregnet for perioden 2010K2 – 2015K3.

Litteratur

- Beer, Michael (2007), "[Bootstrapping a Hedonic Price Index. Experience used from cars data](#)", Springer Link
- Hervig Jacobsen, Lars og Jakob Holmgaard (2016), "[Anvendelse af hedonisk boligprisindeks](#)", Symposium i anvendt statistik.
- Holmgaard, Jakob (2013), "[Revisioner af Ejendomssalgstatistikken](#)", Danmarks Statistik
- Holmgaard, Jakob (2014), "[Beregning af prisindeks for ejendomssalg](#)", Danmarks Statistik

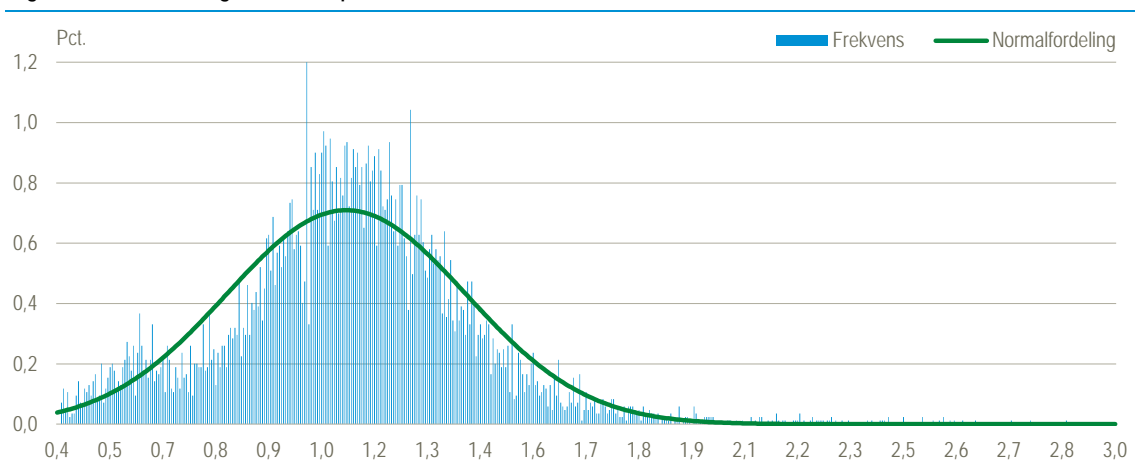
Bilag: Bootstrapping til beregning af usikkerhedsintervaller

Ved bootstrapping udtager man et stort antal forskellige stikprøver af et datasæt, der her omfatter afstandsprocenterne for de handlede boliger i et givet kvartal. Derefter beregner man den gennemsnitlige afstandsprocent for hver stikprøve. Det resulterende sæt af gennemsnitlige afstandsprocenter beskriver usikkerheden på den gennemsnitlige afstandsprocent. I den gennemførte bootstrapping-analyse er der for hvert kvartal konstrueret et passende stort antal datasæt (5.000) ved hjælp af Unrestricted Random Sampling (URS), hvilket betyder, at alle datasættene indeholder samme antal observationer som det originale datasæt, men nogle af de originale observationer er med flere gange i et datasæt, mens andre observationer ikke er med. På denne måde fås 5.000 bud på den gennemsnitlige afstandsprocent i hvert kvartal, og de kan bruges til at beregne usikkerhedsintervaller.

Denne gennemsnitlige afstandsprocent beregnes for alle 5.000 bootsamples pr. kvartal og man beregner herefter en usikkerhed på den gennemsnitlige afstandsprocent for de 5.000 bootsamples. Det bemærkes, at usikkerheden på den gennemsnitlige afstandsprocent er langt mindre end usikkerheden på den enkelte ejendoms afstandsprocent, der varierer over et stort interval. Så usikkerheden på den gennemsnitlige afstandsprocent for et kvartals handler skal ikke forveksles med den betydelige usikkerhed, der er på afstandsprocenten for de enkelte handler, hvor salgsprisen kan ligge langt fra ejendomsvurderingen.

Nedenfor er vist en figur over fordelingen af afstandsprocenter for *enkeltobservationer* på enfamiliehuse i 1. kvartal 2010. Der er tale om fejlsøgt data, så bolighandler, hvor salgspriser mindre end 40 procent eller over 300 procent af den offentlige ejendomsværdi, er frasorteret.

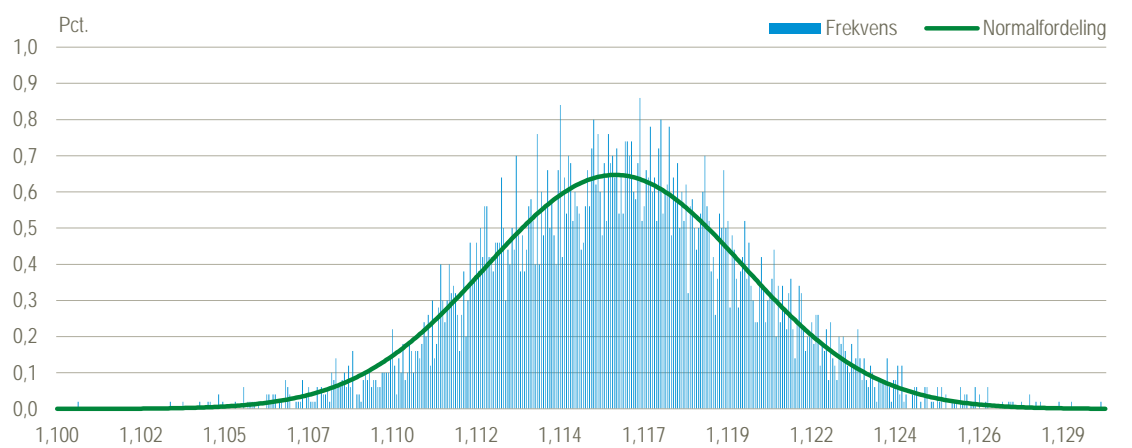
Figur 5 Fordeling af afstandsprocenter for enkeltobservationer. 2010K1



Afstandsprocenterne for de enkelte ejendomme er ikke normalfordelt, jf. at figur 5 ovenfor ikke er pænt klokkeformet men har to toppe, en lang hale mod højre og en afskåret hale mod venstre. Den afskårne hale afspejler, at afstandsprocenter under 0,4 er frasorteret.

Da prisindekset for boligpriserne gengiver udviklingen i den gennemsnitlige afstandsprocent er det ikke den store spredning i afstandsprocenten for *enkeltobservationer*, der som nævnt er illustreret i figur 5 overfor, som skal anvendes, men derimod den mere beskedne spredning i den *gennemsnitlige* afstandsprocent, der i figur 6 nedenfor er illustreret for datasættet i 1. kvartal 2010. Man kan beregne øvre og nedre grænse i usikkerhedsintervallet for prisindekset på to måder. Enten kan man aflæse 2,5 pct. og 97,5 pct. grænsen i figur 6 eller man kan beregne standardafvigelsen i figur 6's observationer og bruge +/- 1,96 gange standardafvigelsen som øvre og nedre grænse i usikkerhedsintervallet for 1. kvartal 2010. De to måder giver samme resultat for alle kvartaler, hvis de 5.000 *gennemsnitlige* afstandsprocenter er normalfordelte i alle kvartaler.

Figur 6 Fordeling af de gennemsnitlige afstandsprocenter. 2010K1



Figur 6 tyder på, at fordelingen af de gennemsnitlige afstandsprocenter er pænt klokkeformet for 1. kvartal 2010 og kan antages at være normalfordelt. Lignende figurer kan laves for andre perioder samt for ejerlejligheder.

Prisudviklingen for periode t i forhold til den forrige periode t-1 er givet ved:

$$\Delta I_{t-1:t} = I_t / I_{t-1} = \frac{afspct_t}{afspct_{t-1}} \tag{1}$$

Vi vil opfatte afstandsprocenterne i ligning (1) som stokastiske uafhængige variable. Da det er nemmere at angive variansen på en difference end variansen på en brøk, tages logaritmen til ligning (1) for at få en difference:

$$\Delta \ln I_{t-1:t} = \ln afspct_t - \ln afspct_{t-1} \tag{2}$$

Hvor $\ln I_t - \ln I_{t-1}$ er skrevet som $\Delta \ln I_{t-1:t}$.

Variansen på differencen mellem to stokastiske uafhængige variable er givet ved følgende formel:

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y) \tag{3}$$

Hvis man antager, at kovariansen er nul, reduceres ligning (3) til:

$$Var(X - Y) = Var(X) + Var(Y) \tag{4}$$

Der gælder således, at:

$$Var(\Delta \ln I_{t-1:t}) = Var(\ln afspct_t) + Var(\ln afspct_{t-1}) \tag{5}$$

Hermed kan usikkerhedsintervallerne for den kvartalsvise prisudvikling i figur 3 og 4 beregnes således:

$$Usikkerhedsinterval_t = \exp(\Delta \ln I_{t-1:t} \pm 1,96 \cdot \sqrt{\text{Var}(\Delta \ln I_{t-1:t})}) \quad (6)$$

hvor,

- \exp : Eksponentialfunktionen
- $\Delta \ln I_{t-1:t}$: Logaritmen til I_t minus logaritmen til I_{t-1}
- $\text{Var}(\Delta \ln I_{t-1:t})$: Variansen af logaritmen til prisindekset i periode t i forhold til den forrige periode $t-1$

Prisindekset i periode t med basisperiode b kan skrives som I_t/I_b :

$$\Delta I_{b:t} = I_t/I_b = \frac{afspct_t}{afspct_b} \quad (7)$$

Ved at tage logaritmen til ligning (7) fås en difference:

$$\Delta \ln I_{b:t} = \ln Afspect_t - \ln Afspect_b \quad (8)$$

Variansen til ligning (8) er givet ved:

$$\text{Var}(\Delta \ln I_{b:t}) = \text{Var}(\ln Afspect_t) + \text{Var}(\ln Afspect_b) \quad (9)$$

Hvor det antages, at kovariansen er nul.

Hermed kan usikkerhedsintervallerne for prisindeksene i figur 1 og 2 beregnes således:

$$Usikkerhedsinterval_t = \exp(\Delta \ln I_{b:t} \pm 1,96 \cdot \sqrt{\text{Var}(\Delta \ln I_{b:t})}) \quad (10)$$

hvor,

- \exp : Eksponentialfunktionen
- $\Delta \ln I_{b:t}$: Logaritmen til I_t minus logaritmen til I_b , så formel 10 gentager formel 6 med basisperioden b i stedet for foregående periode $t-1$
- $\text{Var}(\Delta \ln I_{b:t})$: Variansen af logaritmen til det kædede prisindeks i periode t med basisperiode b

Ved beregning af usikkerhedsintervaller for prisindekset er der ikke taget højde for effekten på variansen ved vurderingsskiftet.