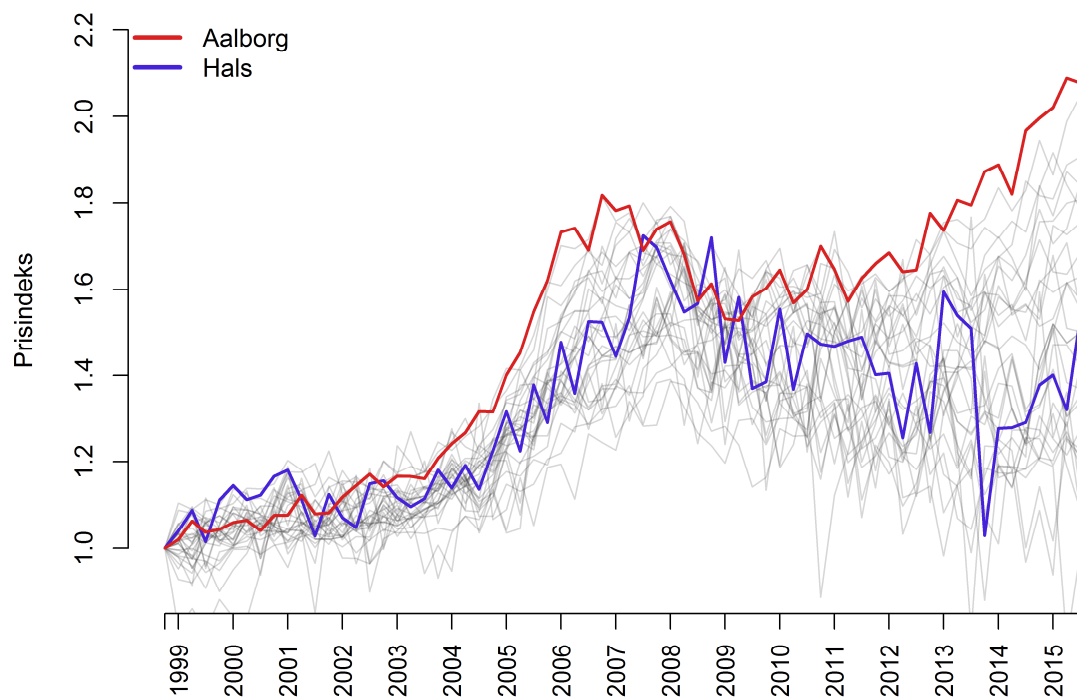


Udglatning af GWR-prisindeksene

En målsætning for GWR-prisindeksene er, at de er temporale og spartiale kontinuere. Den geografiske vægtede regression (GWR), beskrevet ovenover, sikrer geografisk spartial kontinuitet naboboligerne imellem. Metoden sikrer dog ikke kontinuitet over tid. Figur 1, viser GWR-prisindekset for 50 tilfældigt udvalgte ejerboliger i Nordjylland før udglatning. I figuren er to ejendomme beliggende i Aalborg kommune fremhævet. Den ene bolig, den røde kurve, er en ejerlejlighed på 93 kvadratmeter i selve Aalborg by. Den anden ejendom er et parcelhus på 115 kvadratmeter i Hals, en mindre by med knap 2.500 indbyggere beliggende ved Limfjordens udmundning i Kattegat, det blå prisindeks. Disse to ejerboliger bruges efterfølgende til at eksemplificere udglatning af prisindeksene¹.

Figur 1. Prisindeks for 50 tilfældigt udvalgte ejerboliger i Nordjylland før udglatning, 1999 indekseret til 1.



*Kilde: Egne beregninger på baggrund af registerdata fra Danmarks Statistik.
Anm: Det røde prisindeks er en ejerlejlighed på 93 kvadratmeter i selve Aalborg by. Det blå prisindeks er et parcelhus på 115 kvadratmeter i Hals.*

Det ses fra Figur 1, at prisindekset fra GWR estimationen er udsat for en del tidsafhængig støj, idet den enkelte boligs prisindeks svinger meget fra kvartal til kvartal.

Udsvingene i prisindeksene opstår hovedsagligt som en konsekvens af inhomogenitet i de solgte boliger der indgår i estimationsgrundlaget for de enkelte boliger. Inhomogeniteten forklares af forskellen i den vedligeholdsmæssige stand mellem de solgte boliger, som ikke er

¹ Bemærk at det er de samme to ejendomme der er brugt i afsnit XXXTODO.

ikke observeret i data. Følgende afsnit beskriver en metode til at udglatte GWR-prisindeksene, således at prisindeksene bliver kontinuere over tid.

LOESS Smoothing

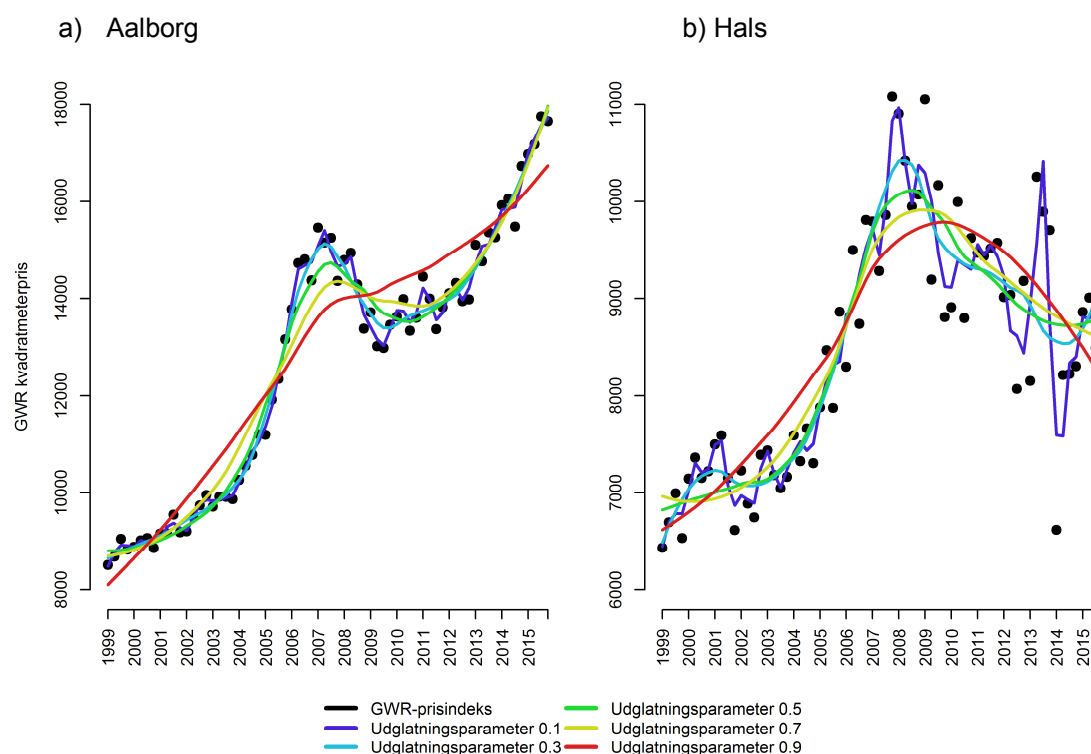
Til at udglatte og for at opnå temporal kontinuitet af prisindeksene bruges LOESS metoden, Cleveland 1979. Metoden er en ikke-parametrisk metode, som for ethvert punkt af et givet GWR-prisindeks tilpasser et polynomium med baggrund i en delmængde af nabopunkterne som forklarende variabler. Polynomiet tilpasses via weighted least squares, hvor vægtene er afstanden fra nabopunkterne til det betragtede punkt, og er beregnet ud fra en nearest neighbor algoritme. Herved har nabopunkterne tættest på den største vægt.

Tilpasningen til hvert enkelt punkt i GWR-prisindekset sker derfor lokalt med afsæt i punkternes naboer. Til at styre størrelsen af naboområdet vælges en udglatningsparameter. Parameteren angiver hvor stor en andel af alle punkter i prisindekset, der skal medtages i tilpasningen af polynomiet til hvert enkelt punkt. Udglatningsparameteren kontrollerer på denne måde udglatningsfleksibiliteten i LOESS regressionen. For lave parameterværdier, tæt på nul, er metoden lokal fokuseret, og den resulterer derfor i minimal udglatning. Vice versa for høje værdier af udglatningsparameteren, værdier tæt på 1, her er metoden mere global og resulterer i maksimal udglatning.

Figur 2 viser LOESS regressionen med forskellige værdier af udglatningsparameteren for de to ejendomme beliggende i Aalborg kommune. De sorte punkter i Figur 2 angiver prisindekset, der kommer direkte fra GWR estimationen. Det ses, at boligen i Hals oplever mere støj i prisindekset end boligen i Aalborg by. Dette er et generelt fænomen, desto mindre byområde boligen er beliggende i desto mere støj er boligens GWR-prisindeks udsat for. Datagrundlaget til GWR-estimationen dækker et større geografisk område hvis boligen er beliggende i et tyndt befolket område. I det søgeradiusen skal være større for at opnå det ønskede antal faktiske salg der danner datagrundlag. En konsekvens af dette, er at datagrundlaget i tyndtbefolkede områder er mindre homogent end i større byer. I de større byer er søgeradiusen mindre og boligerne minder mere om hinanden.

Figur 2 viser herudover 5 forskellige udglattede versioner af prisindekset til de to boliger. Det fremgår, som nævnt at jo større værdi udglatningsparameteren er desto mere glat er prisindekset. Det fremgår også, hvor "godt" den udglattede kurve passer til GWR-prisindekset ændrer sig for forskellige værdier af udglatningsparameteren. Vælges en parameter på 0,1 ses det for begge boliger, at prisindekset stadig ikke er kontinueret. Vælges derimod en værdi på 0,9 ses det at prisindeksene bliver udglattet for meget. Den rigtige parameterværdi findes et sted mellem disse to ekstremer. Oplagt fremgår det, at prisindekset med mest støj skal udglattes mest og vice versa, det er dog ikke oplagt præcist hvor meget de skal udglattes. Generelt ses det, at der er en positiv sammenhæng mellem søgeradius, støj og udglatningsparameteren.

Figur 2. Prisindekset for ejendommene i Aalborg og Hals under forskellige værdier af udglatningsparameteren.



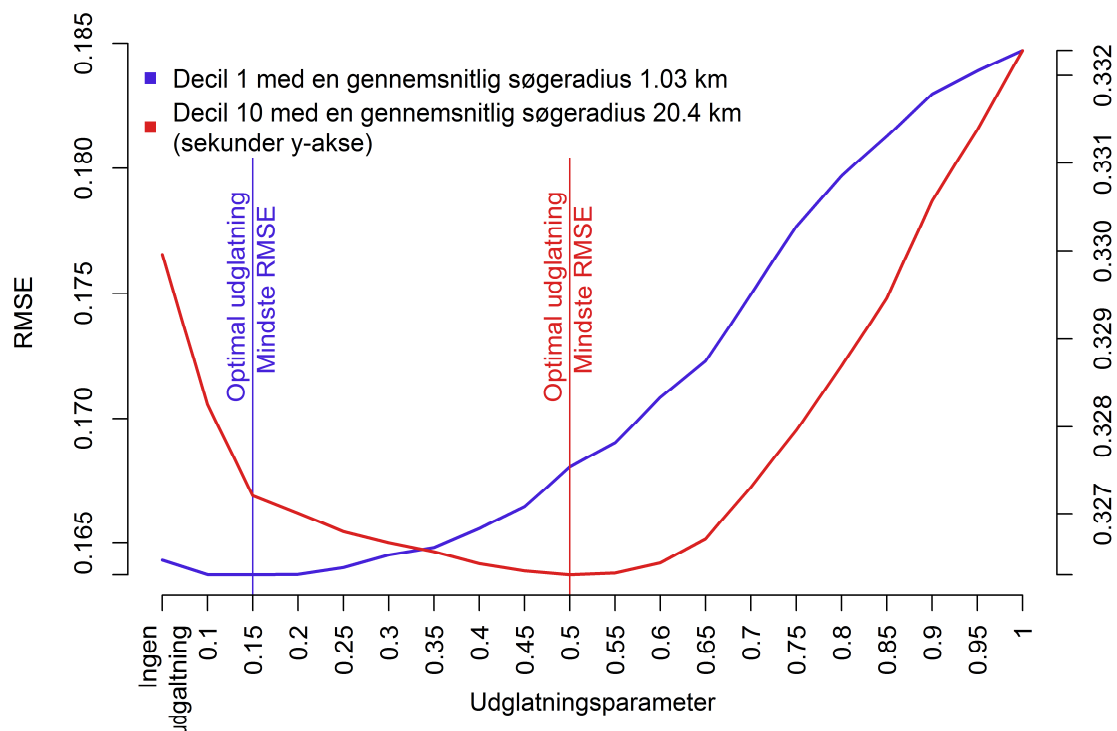
Kilde: Egne beregninger på baggrund af registerdata fra Danmarks Statistik.

Optimal udglatning

Afsnittet bestemmer for hver enkelt ejerbolig den optimale udglatning af GWR-prisindekset. Dette gøres ved kvantificere den ovenover beskrevne sammenhæng mellem søgeradius og udglatningsparameteren.

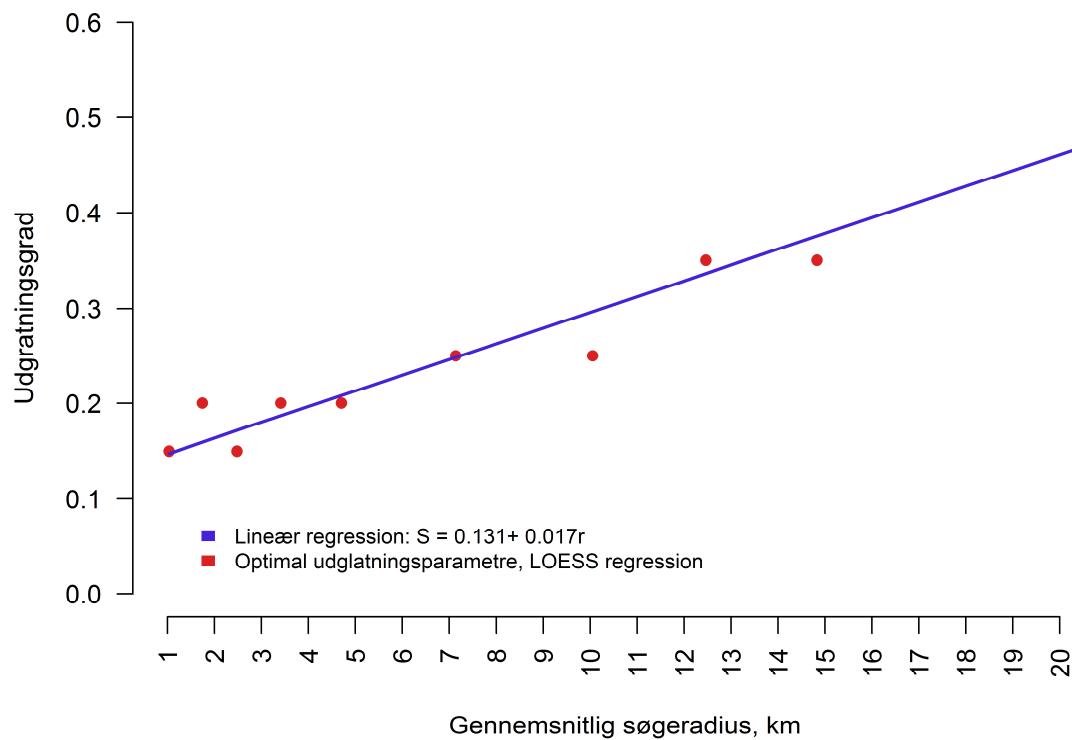
For alle faktiske solgte ejerboliger estimeres LOESS regressionen med 20 forskellige værdier af udglatningsparameteren i intervallet fra 0.05 til 1. Efterfølgende opdeles de solgte ejerboliger i deciler efter størrelsen af søgeradius. For hver decilgruppe og hver værdi af udglatningsparameteren beregnes Root-mean-square-error, RMSE, mellem de udglattede kurver og de faktiske solgte værdier. Udglatningsparameteren der giver den mindst RMSE, vælges som den optimale værdi til ejerboliger med denne søgeradius. Figur 3 viser sammenhængen mellem RMSE og udglatningsparameteren for den første og sidste decil. Fra Figur 3 ses det, at den optimale udglatningsparameter er 0.15 for den første decilgruppe hvor den gennemsnitlige søgeradius er på 1.03 km. For den sidste decilgruppe, hvor den gennemsnitlige søgeradius er 20.4 km, er den optimale udglatningsparameter 0.5.

Figur 3. RMSE og udglatningsparameteren for decil 1 og decil 10.



Kilde: Egne beregninger på baggrund af registerdata fra Danmarks Statistik.

Figur 4. Sammenhæng mellem udglatningsgrad og søgeradius.



Kilde: Egne beregninger på baggrund af registerdata fra Danmarks Statistik.

Den generelle sammenhæng mellem den gennemsnitlige søgeradius og den optimale udglatningsparameter for hver decil er vist i Figur 4. Figuren viser også en simpel lineær regression af denne sammenhæng, hvilket ses, at være en god tilnærmelse med en justeret R^2 på 0.93 . Estimationsresultatet giver:

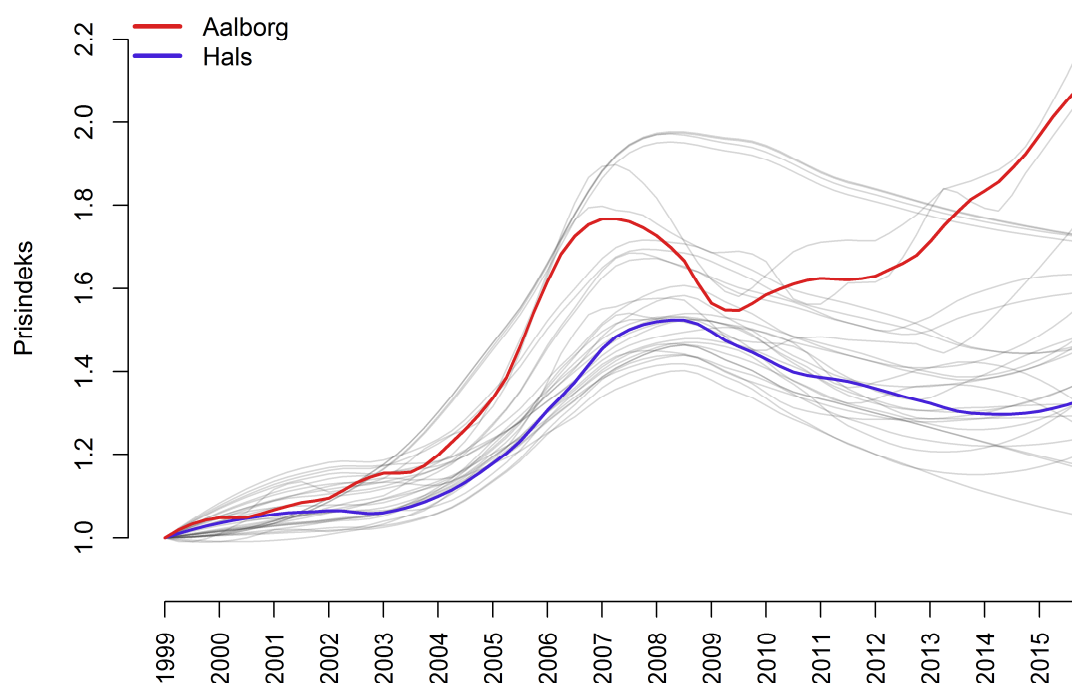
$$s = 0.131 + 0.017r$$

0.0145 0.0015

Resultatet danner regel for valget af udglatningsparameter for både solgte og ikke solgte boliger.

Resultatet af LOESS udglatningsmetoden er eksemplificeret i Figur 5. Her ses det udglattede GWR-prisindeks for de samme 50 tilfældigt udvalgte ejerboliger i Nordjylland som i Figur 1. Det ses, at prisindeksene her er kontinuerte over tid.

Figur 5. Udglattet prisindeks for 30 tilfældigt udvalgte ejerboliger i Nordjylland efter udglatning.



Kilde: Egne beregninger på baggrund af registerdata fra Danmarks Statistik.