Nokre praktiske røynsler med analytisk glatting

av Erling Berge[1] *og Jan M. Hoem*[2]

0. Samandrag

I samband med ein studie av regional fertilitet i Norge ca. 1970 har forfattarane freista å undersøke om Hadwiger-funksjonen kan seiast å vere systematisk betre eller dårlegare som glattingsfunksjon enn GG-funksjonen, som er basert på gammatettheta. Vi har ikkje funne noko eintydig svar. GG-funksjonen kjem best ut i svært mange høve. Hadwiger-funksjonen synest likevel å vere meir fleksibel, og mykje tyder på at han er den overlegne av dei to når fertilitetskurva er venstreskeiv og har tung høgrehale.

Vi har og vilja sjå i kva mon ulike val av øvre og nedre grense for dei aldersgruppene som er med under glattinga, gir ulike glatta fertilitetskurver. Det viser seg at dette valet kan ha stor verknad på resultatet.

1. Innleiing

1.1. Prosjektet

Forfattarane av dette notatet har gjort ein studie av regional fertilitet i Norge i åra rundt 1970. Dei 449 kommunane i landet 1/1 1971 er grupperte i 77 fertilitetsområde på grunnlag av næringsstruktur, geografisk plassering og om dei har systematisk inn- eller utflyttning. Kvart område vil typisk ha 30 000—50 000 innbyggjarar. For kvart område har vi rekna ut fødselsrater for kvar einskild alder frå 15 år til 50 år ved å aggregere data for åra 1968 til 1971. Når ein teiknar eit diagram av eit sett av slike rater med alderen langs abscisseaksen, får ein fram ei "fertilitetskurve". Kvar slik kurve har vi glatta analytisk ved å tilpasse ein Hadwiger-funksjon til henne. For nokre av dei har vi og laga ei alternativ glatting med ein GG-funksjon, som er ein glattingsfunksjon basert på gammatettheta.

Vi har freista å undersøke om Hadwiger-funksjonen kan seiast å vere systematisk betre eller dårlegare enn GG-funksjonen. Vi har og vilja sjå i kva mon ulike val av øvre og nedre grense for dei aldersgruppene som vi tek med i glattinga, gir ulike glatta fertilitetskurver. I dette notatet skal vi rapportere om praktiske røynsler vi har gjort i dette arbeidet. Dei resultata vi legg fram her, er eit representativt utval av dei vi har funne under denne delen av prosjektet. Nokre fleire detaljer finst i Berge og Hoem (1974).

Tidlegere rapportar frå andre delar av prosjektet har vi gitt i Berge (1973, 1974) og Hoem og Berge (1974 a, b).

1.2. Fertilitetskurvene

Formelen vi har nytta til å rekne ut fødselsraten for x-årige kvinner i fertilitetsområde nr. i, er

$$_i\hat{\lambda}_x = \frac{\sum_{n=1968}^{1971} {}^n_iF_x}{\sum_{n=1968}^{1971} \tfrac{1}{2}\left\{{}^{n-1}_{\ i}L_{x-1} + {}^n_iL_x\right\}} = \frac{_iF_x}{_iM_x},$$

der $x = 15, 16, \ldots, 50$ og $i = 1, 2, \ldots, 77$. Her er ${}^n_iF_x$ talet på levandefødde barn i

[1] Statistisk Sentralbyrå, Oslo.
[2] Forsikringsmatematisk laboratorium, Köpenhamn.

året n i kommunegruppe nr. i av mødrer som ved utgangen av året n vil vere x år gamle, og ${}^{n}_{i}L_x$ er talet på kvinner i alt i befolkninga i kommunegruppe nr. i som ved utgangen av året n er x år. I diagramma er ratene oppgjevne i promille.

For å teste eigenskapane til glattingsfunksjonane valde vi ut fem av områda slik at vi skulle få representert dei viktigaste kurvetypene vi har støytt på i materialet. Desse kommunegruppene har nummera 10, 32, 38, 43 og 77. Dei er samansett slik ein ser av tab. 1.

Oslo er det største av fertilitetsområda våre, rekna etter folketalet. Område nr. 77 (kommunar i Finnmark) er det minste.

Tab. 1. Kommunane i fertilitetsområda 10, 32, 38, 43 og 77

Fertilitetsområde nr.	Kommunar		Samla folketal 31/12 1970
10	0301	Oslo	482 000
32	0806	Skien	45 000
38	1004	Flekkefjord	
	1101	Eigersund	
	1102	Sandnes	50 000
43	1144	Kvitsøy	
	1145	Bokn	
	1151	Utsira	
	1244	Austevoll	
	1245	Sund	
	1259	Øygarden	
	1264	Austrheim	
	1265	Fedje	
	1412	Solund	
	1428	Askvold	
	1441	Selje	
	1514	Sande	
	1532	Giske	31 000
77	2011	Kautokeino	
	2020	Porsanger	
	2021	Karasjok	
	2025	Tana	
	2027	Nesseby	13 000

1.3. Glattingsfunksjonane

Med den tradisjonelle parametriseringa ser Hadwiger-funksjonen slik ut:

$$h_0(x;R,H,T,U) = \frac{RH}{T\sqrt{\pi}}\left(\frac{T}{x-U+T}\right)^{-3/2}$$

$$\exp\left\{-H^2\left(\frac{T}{x-U+T}+\frac{x-U+T}{T}-2\right)\right\}$$

for $x > U-T$.

Her tolkar vi parameteren R (arealet under kurva) som det samla fertilitetstalet. U tolkar vi som gjennomsnittleg fødealder. Parametrane H og T kan vi derimot ikkje gi ei substansiell demografisk tolking.

I eit tidlegare notat (Hoem og Berge, 1974 a) har vi vist at denne parametriseringa gir oss eit praktisk identifikasjonsproblem ved at store variasjonar i parametrane ikkje treng føre til meir enn små endringar i dei tilsvarande kurvene. Det ser ut til at vi kan løyse desse problema ved å erstatte parametrane H og T med to nye parametrar, S^2 og M, der S^2 er variansen i sannsynlighetsfordelinga med tettheta $h_0(.;R,H,T,U)/R$, medan M er modalverdien til funksjonen.

Den tilsvarande "klassiske" parametriseringa av GG-funksjonen er

$$g_0(x;R,b,c,d) = \frac{R}{\Gamma(b)c^b}(x-d)^{b-1}$$

$$\exp\{-(x-d)/c\} \quad \text{for } x > d,$$

$$g_0 = 0 \quad \text{for } x \leq d.$$

Her og løner det seg å reparametrisere, slik at ein brukar parametrane R, m, μ og σ^2 i glattinga, der m, μ og σ^2 er modalverdien, forventningsverdien og variansen i gammatettheta.

Vi har difor brukt glattingsfunksjonane $h(x;\Theta)$ og $g(x;\Theta)$, der h er Hadwiger-funksjonen parametrisert med $\Theta = (R, M, U, S^2)$ medan g er GG-funksjonen med $\Theta = (R, m, \mu, \sigma^2)$. Vi har samanlikna tilpassingsevna til desse to funksjonane.

1.4. Glattingsmetode

La $\{\hat{\lambda}_x: x = \alpha, \alpha+1, \ldots, \beta\}$ vere ei rekkje fødselsrater for aldrane α, $\alpha+1, \ldots, \beta$, og la $f(x;\Theta)$ vere ein valt glattingsfunksjon. Å glatte fertilitetskurva til desse ratene analytisk vil då seie at ein vel ein verdi $\hat{\Theta}$ av parameteren Θ slik at kurva av glatta verdiar $\{f(x;\hat{\Theta}): x = \alpha, \alpha+1, \ldots, \beta\}$ i ei eller anna meining høver godt med fertilitetskurva $\{\hat{\lambda}_x: x = \alpha, \alpha+1, \ldots, \beta\}$. Sidan $\hat{\Theta}$ må veljast på basis av observasjonane, reknar vi han som ein statistisk estimator.

Det finst mange måter å velje $\hat{\Theta}$ på. Hoem (1972) har likevel før vist at det kan løne seg å estimere Θ ved hjelp av ein modifisert kjikvadratminimeringsmetode. Det vil seie at ein som $\hat{\Theta}$ vel den verdien av Θ som gjer

$$\sum_{x=\alpha}^{\beta} \frac{\{\hat{\lambda}_x - f(x;\Theta)\}^2}{\hat{\lambda}_x/M_x}$$

minst mogeleg. Ein får da minimert den asymptotiske variansen av estimatoren $\hat{\Theta}$ og av $f(x;\hat{\Theta})$ for alle x (innanfor ei stor klasse av mogelege estimatorar). Vi har óg vist (Hoem og Berge, 1974 b) at denne metoden i praksis gir merkbart lågare asymptotisk varians enn om ein til dømes nyttar minste kvadraters metode og minimerer

$$\sum_{x=\alpha}^{\beta} \{\hat{\lambda}_x - f(x;\Theta)\}^2.$$

For å kunne jamføre, har vi i dette arbeidet nytta begge desse metodane.

1.5. Val av aldersklasser i glattinga

Når ein nyttar eittårige aldersintervall, reknar ein til vanleg ut fødselsratene for aldrane frå 15 til 44 år, eller frå 15 til 49 år eller 50 år, etter som det høver. La oss kalle lågaste og høgste alder for a og b.

Når ein skal glatta fertilitetskurva som svarer til ratene $\{\hat{\lambda}_x : x = a, a+1, \ldots, b\}$, treng ein naturlegvis ikkje ta med alle desse aldersklassane i reknearbeidet. Ein kan til dømes nøye seg med å ta med ratene $\{\hat{\lambda}_x: x = \alpha, \alpha+1, \ldots, \beta\}$ under glattinga, der $a \leqslant \alpha < \beta \leqslant b$. Når ein så har funne den verdien $\hat{\Theta}$ av parametrane ein vil bruke, kan ein sjølsagt likevel rekna ut dei glatta fødselsratene $f(x, \hat{\Theta})$ for alle x frå $x = a$ til $x = b$.

I nokre høve er ein heilt enkelt nøydd til å velje slike α og β som ikkje fell saman med a og b. For fertilitetsområda våre har vi brukt $a = 15$ og $b = 50$, men vi får ofte $\hat{\lambda}_x = 0$ for høge aldrar x, og nokre gonger er og $\hat{\lambda}_{15} = 0$. Skal ein bruke kjikvadratminimering, nyttar det ikkje å ta med aldrar x der $\hat{\lambda}_x = 0$, for da har ikkje formelen vi gav i avsnitt 1.3 noka meining. Med kjikvadratminimeringsmetoden likar ein i det heile ikkje få fødslar i ei aldersklasse. Eit vanleg kriterium er å krevje at $F_x > 5$ for alle dei x ein tar med. Som ein vil sjå i kapittel 2 nedanfor, har vi nytta dette kriteriet. Vi har der studert i kva mon val av α og β har merkbar verknad på glattinga.

1.6. Kvalitetskriteria

Ved rota av dei problema vi tar opp her, ligg dette spørsmålet:

Kva skal ein krevje av ei glatting for å seie at ho passar godt til data?

Når glattingsfunksjonen er gitt, er svaret altså gitt om vi går ut fra at minimal asymptotisk varians er eit dominerande krav. Estimeringa må da baserast på den modifiserte kjikvadratminimeringsmetoden (eller ein annan metode med like gode eigenskapar). Men når vi skal velje mellom ulike forslag til glattingsfunksjonar, treng vi kriterium ut over dette.

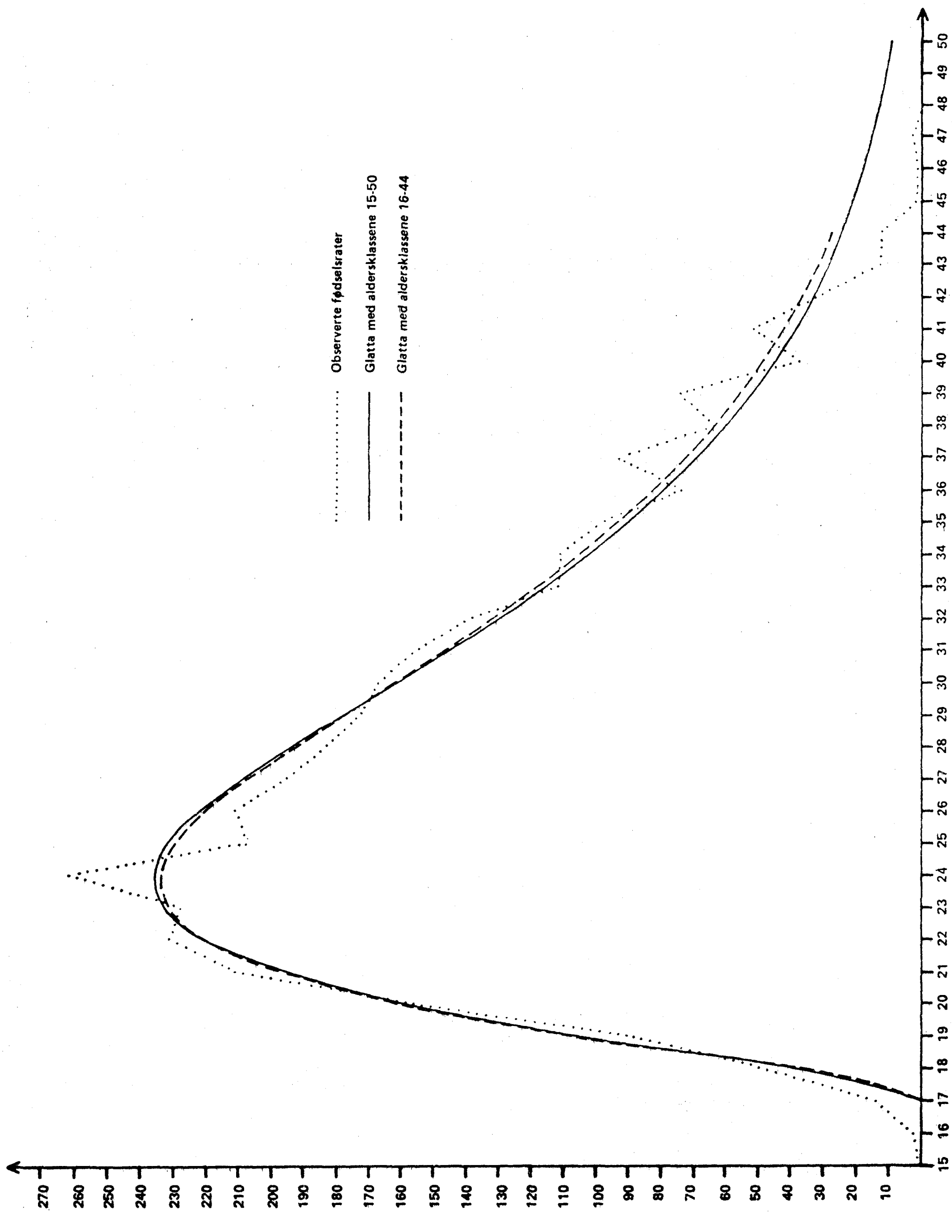

Fig. 1. Fertilitetskurva for område nr. 43 glatta med GG-funksjonen og minste kvadraters metode og med ulike aldersklasser

Eitt rimeleg slikt kriterium er at kjikvadratverdien eller kvadratavviksummen for tilpassinga skal vere så liten som mogeleg. Det kan óg vere at vi ønskjer særleg god tilpassing i dei sentrale (mest fruktbare) aldrane, til dømes frå 20 til 30 år. Vi må da sjå på kjikvadratverdi eller kvadratavviksum for tilpassinga mellom desse aldrane for seg.

Vi skal sjå kva resultat desse ideane gir i det følgjande.

1.7. Hovudkonklusjonar

(i) Det viser seg at tilpassing med GG-funksjonen gir lågare kjikvadratverdi enn om ein nyttar Hadwiger-funksjonen i tre firedelar av dei numeriske døma våre. På mange måtar synest Hadwiger-funksjonen likevel å passe betre både i dei sentrale aldersklassene og i dei aller lågaste aldersklassene vi tek med i utrekningane. Vi har og inntrykk av at Hadwiger-funksjonen stort sett høver best av dei to når fertilitetskurva er sterkt venstreskeiv og har ein "tung" hale til høgre. Dette går vi nærare inn på i kapittel 3 nedanfor.

(ii) Når glattingsfunksjonen er valgt, kan ein få store variasjonar i estimatet av Θ og i dei glatta kurvene etter kva aldersklasser ein tar med i glattinga. Dette synest å gjelde særleg når vi nyttar kjikvadratmetoden i estimeringa. Det er mindre merkbart med minste kvadraters metode. Kjikvadratmetoden vektar dei små observasjonane $\hat{\lambda}_x$ mest. (Sjå formelen i avsnitt 1.3 ovanfor.) Sidan fødselsratene $\hat{\lambda}_x$ stort sett er nær null når x kjem over 45 år, vil ein presse den høgre halen av funksjonen lenger ned mot aksen di fleire aldersklasser over 44 år ein tar med. (Noko tilsvarande gjeld visst ikkje for dei lågaste aldersklassene.) Vi drøfter dette i kapittel 2 nedanfor.

2. Val av aldersklasser i glattinga

2.1. Aldersklassene ved kjikvadratminimering

I avsnitt 1.4 ovanfor nemnde vi at ein treng å legge omtanke i kva for aldersklasser ein tar med i utrekningane. La oss "estimere" den lågaste aldersklassen, α, og den høgste, β, vi vil ta med ved kjikvadratminimering ved å sette

$$\hat{\alpha} = \min \left\{x:\hat{\lambda}_x M_x > 5 \ \& \ \hat{\lambda}_y M_y \leqslant 5 \text{ for alle } y < x\right\}$$

og

$$\hat{\beta} = \max \left\{x:\hat{\lambda}_x M_x > 5 \ \& \ \hat{\lambda}_y M_y \leqslant 5 \text{ for alle } y > x\right\}.$$

(Vi nyttar i praksis $\hat{\lambda}_x M_x$ i staden for F_x for å sleppe med å lese inn $\hat{\lambda}_x$-ane og M_x-ane i reknemaskinprogrammet.) Dette har vi gjort for alle dei 77 fertilitetsområda. I tab. 2 er frekvensfordelinga for $\hat{\alpha}$ og $\hat{\beta}$ gitt.

Tab. 2. Frekvensfordeling for "estimat" av yttergrensene for reproduktiv periode

$\hat{\alpha}$	ant. område	$\hat{\beta}$	ant. område
15	2	41	1
16	35	42	13
17	40	43	15
		44	33
		45	12
		46	3

Når ein skal nytte faste grenser for dei aldersklassene ein tar med i glattinga, viser desse tala at aldersintervallet 16—44 år vil vere eit naturleg val av α og β i Norge i dag. Denne konstateringa dekker imidlertid over at ($\hat{\alpha}$, $\hat{\beta}$) kan variere frå 15—46 i eit område til 17—41 i eit anna.

Vi må også vere klar over at $\hat{\alpha}$ og $\hat{\beta}$ ikkje berre er biologisk bestemt, men og sosialt. Særleg vil $\hat{\beta}$ variere områda imellom innan vår samfunnstype.

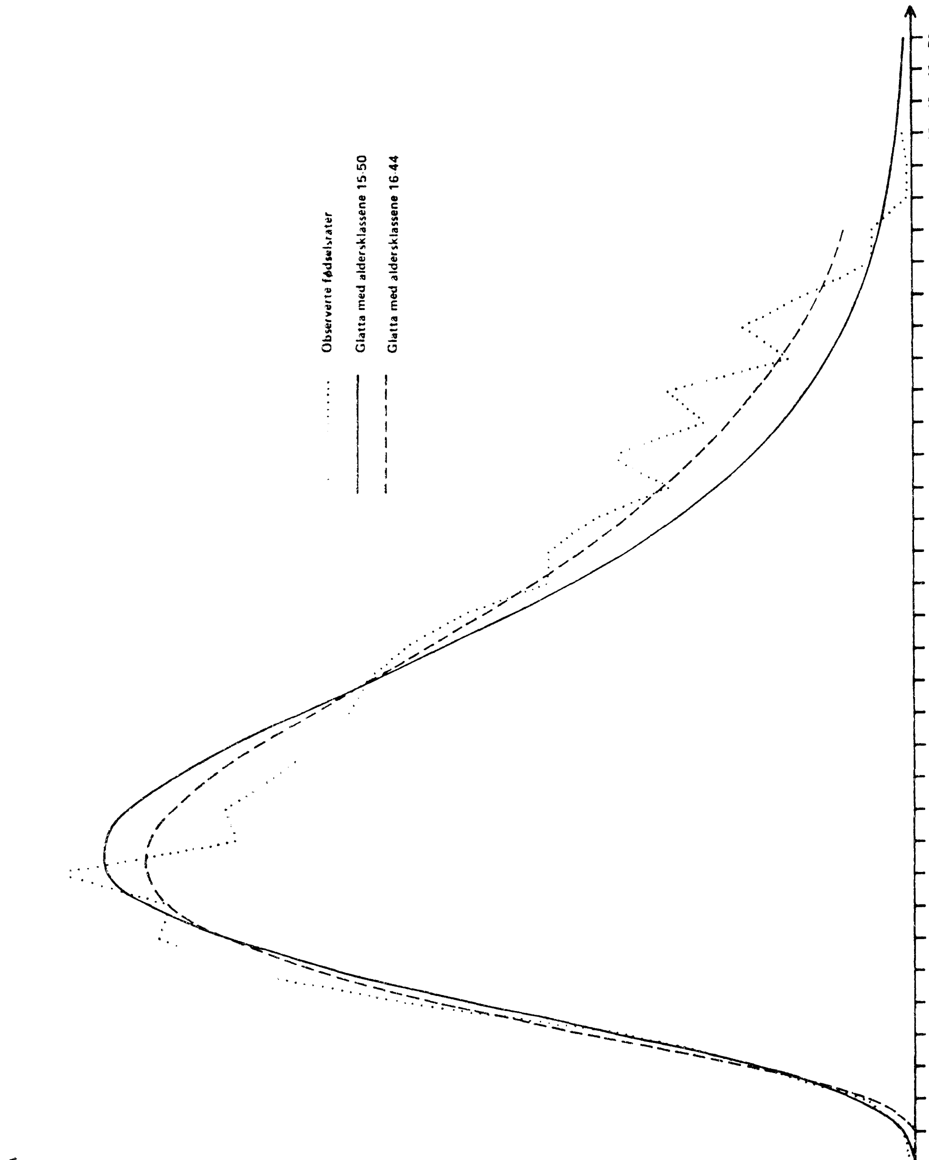

Fig. 2. Fertilitetskurva for område nr. 43 glatta med GG-funksjonen og kjikvadratminimering og med ulike aldersklasser

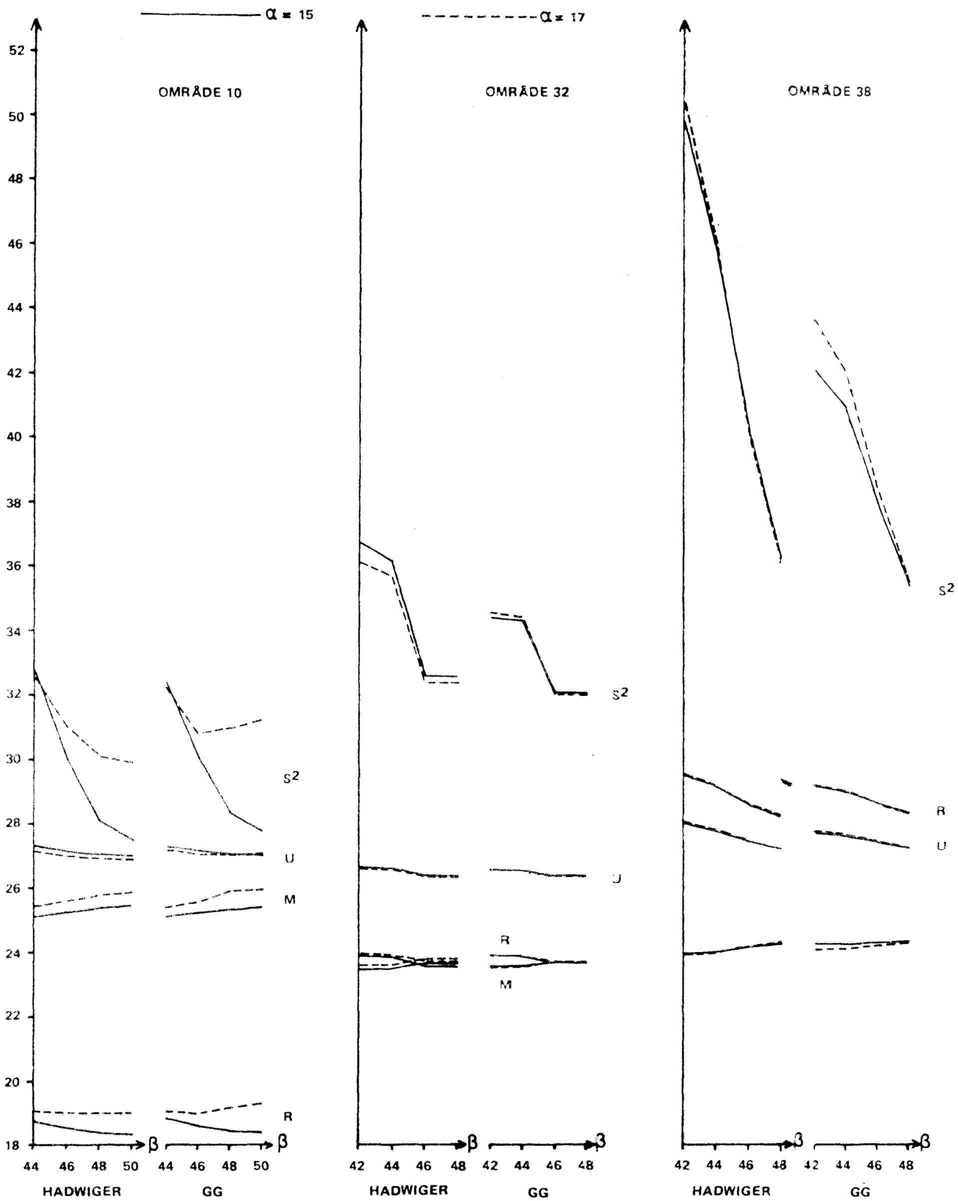

Fig. 3. Estimat av parametrane i Hadwiger- og GG-funksjonen for fertilitetsområda 10, 32 og 38 ved bruk av ulike aldersklasser

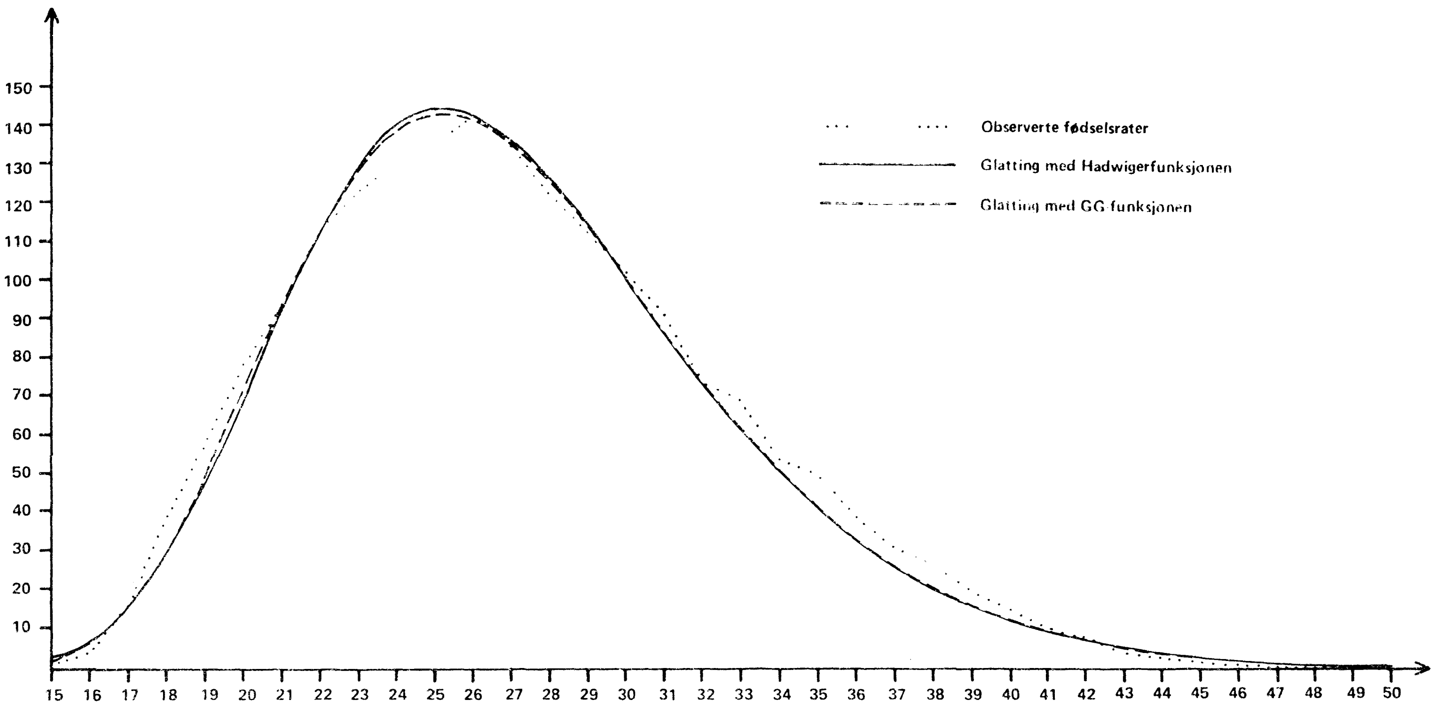

Fig. 4. Fertilitetskurva for område nr. 10 glatta ved kjikvadratminimering med aldersklassene fra $\hat{\alpha}=15$ til $\hat{\beta}=46$

2.2. *Kurve- og parameterestimat*

For å studere kva effekt ulike val av α og β har på parameterestimatet $\hat{\Theta}$ og på dei glatta kurvene, tok vi føre oss dei fem områda som er nemnde i tab. 1. Dei "estimerte" verdiane for α og β vart slik ein ser av tab. 3.

Tab. 3. $\hat{\alpha}$ *og* $\hat{\beta}$ *for fertilitetsområda 10, 32, 38, 43 og 77*

Område nr.	$\hat{\alpha}$	$\hat{\beta}$
10	15	46
32	16	44
38	17	44
43	17	44
77	17	42

For kvart av desse områda estimerte vi θ med kjikvadratminimeringsmetoden med α fast lik 15, 16 og 17 mens β varierte frå $\hat{\beta}-2$ i skritt på 2 til $\hat{\beta}+4$. Nokre av resultata er framstilte grafisk i figurane 1 til 3. I figurane 1 og 2 har vi tatt for oss fertilitetsområde 43 serskilt. Vi har plotta både den observerte fertilitetskurva og dei glatta kurvene ein får når ein tar med aldersklassene 15—50 og alternativt klassene 16—44 og nyttar GG-funksjonen i utjamninga. Diagramma for Hadwiger-funksjonen er heilt analoge. Vi har og brukt både kjikvadratminimeringsmetoden og alternativt minste kvadraters metode i glattingane. Det er tydeleg at valet av α og β har stor verknad på utsjånaden av kurvene når ein bruker kjikvadratminimeringsmetoden. Om ein nyttar minste kvadraters metode, ser ikkje valet av α og β ut til å spele så stor rolle. Dette gjeld både når ein nyttar Hadwiger-funksjonen og når ein legg GG-funksjonen til grunn. Vi får tilsvarande resultat med kurvene for dei fire andre fertilitetsområda.

I fig. 3 har vi plotta nokre parameterestimat for $\alpha = 15$ og 17 og for $\beta = \hat{\beta}-2, \hat{\beta}, \hat{\beta}+2$ og $\hat{\beta}+4$ for begge glattingsfunksjonane.

Om vi held β fast og ser på effekten av å variere α, finn vi at han for det meste er forsvinnande liten. Den stipla og den heiltrekte kurva fell omtrent saman i kvart kurvepar. For område nr. 10, der $\alpha > \hat{\alpha}$ når $\alpha = 17$, får vi likevel større utslag av endringane i α enn av endringane i β.

Ser vi på effekten av å endre β for fast α, finn vi at denne er stor for variansparameteren, men heller liten for dei andre parametrane, særleg for modal fødealder.

Parameterestimata oppfører seg om lag likt for dei to glattingsfunksjonane. Men variansparameteren i Hadwiger-funksjonen reagerer kraftigare på endringane i β enn den tilsvarande parameteren gjer i GG-funksjonen. Likevel vil variansparameteren i Hadwiger-funksjonen stort sett vere større i numerisk verdi enn variansparameteren i GG-funksjonen.

At Hadwiger-funksjonen reagerer kraftigare når ein tar med nye fødselsrater som ligg nær null, kan vi tolke som eit utslag av større fleksibilitet hos han. På den andre sida vil den større variansparameteren i Hadwiger-funksjonen føre til at vi som regel får nokså høge verdiar av $h(x;\hat{\Theta})$ for $x > \beta$, noko som i mange høve kan vere ein uheldig eigenskap ved funksjonen dersom ein treng glatta fødselsrater for desse aldrane. Når vi nyttar Hadwiger-funksjonen og treng $h(x;\hat{\Theta})$ for $x \geqslant \beta$, bør vi sikre oss at β i alle fall ikkje er mindre enn $\hat{\beta}$ sidan fødselsratene til vanleg vil bli overestimert dersom vi har $\beta < \hat{\beta}$. For $\beta > \hat{\beta}$ finn vi oftast små skilnader mellom variansparametrane i dei to funksjonane samanlikna med den skilnaden vi finn for $\beta < \hat{\beta}$.

Generelt kan vi konkludere med at dei numeriske resultata våre tyder på at vi bør sikre oss at $\alpha \leqslant \hat{\alpha}$ og $\beta \geqslant \hat{\beta}$ anten vi nyttar Hadwiger- eller GG-funksjonen.

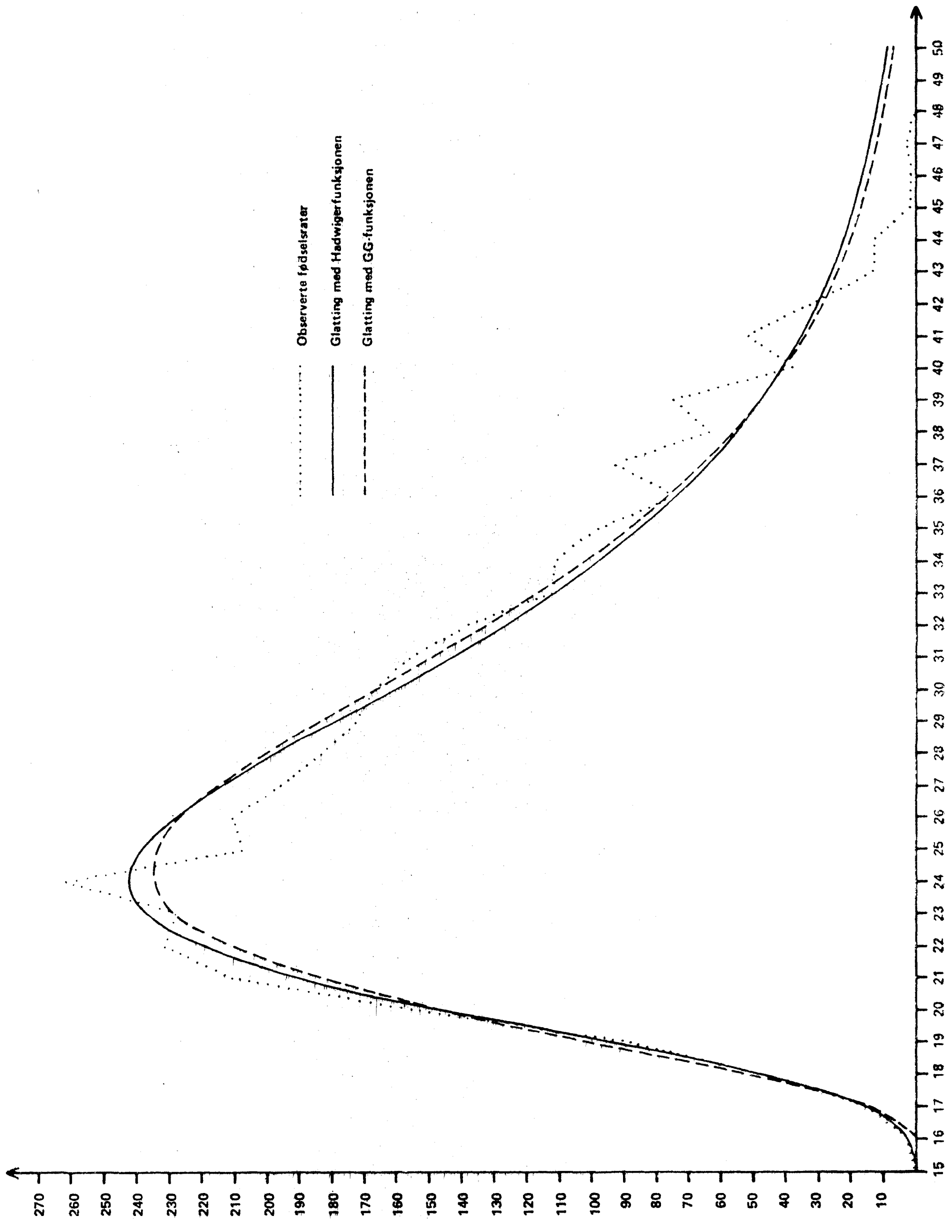

Fig. 5. Fertilitetskurva for område nr. 43 glatta ved kjikvadratminimering med aldersklassene fra $\hat{\alpha}=17$ til $\hat{\beta}=44$

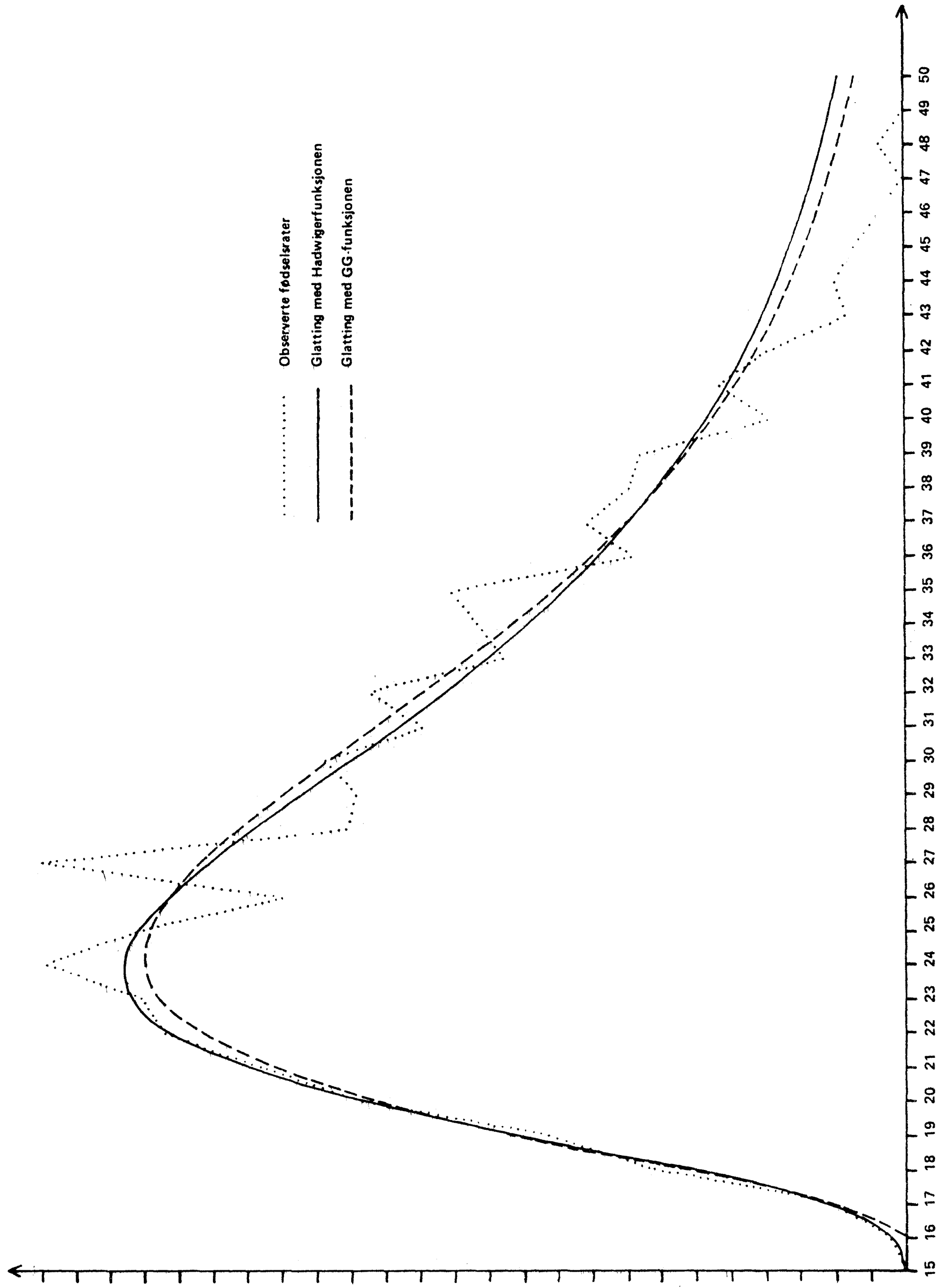

Fig. 6. Fertilitetskurver for område nr. 77 glatta ved kjikvadratminimering med aldersklassene fra $\hat{\alpha}=17$ til $\hat{\beta}=42$

Tab. 6. Kvadratavviksum og kjikvadratverdi for Hadwiger-tilpassing i prosent av same mål ved tilsvarande GG-tilpassing for utvalde fertilitetsområde [2, 3]

Område nr.		10			32			38			43			77		
Aldersintervall for estimering av Θ		Kvadrat-avviksum		Kji-kva-drat-verdi	Kvadrat-avviksum		Kji-kva-drat-verdi	Kvadrat-avviksum		Kji-kva-drat-verdi	Kvadrat-avviksum		Kji-kva-drat-verdi	Kvadrat-avviksum		Kji-kva-drat-verdi
		Alder 20–30	Alder 15–50	Alder $\alpha-\beta$	Alder 20–30	Alder 15–50	Alder $\alpha-\beta$	Alder 20–30	Alder 15–50	Alder $\alpha-\beta$	Alder 20–30	Alder 15–50	Alder $\alpha-\beta$	Alder 20–30	Alder 15–50	Alder $\alpha-\beta$
α	β															
15	40													(1)	(1)	(1)
15	42				119	126	123	79	93	84	67	100	74	91	111	89
15	44	134	136	138	120	128	124	80	93	95	91	104	100	87	106	105
15	46	137	130	127	125	132	127	86	101	111	113	115	113	93	109	115
15	48	134	123	119	125	132	127	94	110	117	114	116	113			
15	50	133	121	117												
17	40													(1)	(1)	(1)
17	42				116	119	126	84	97	97	60	100	77	90	111	100
17	44	115	117	112	117	120	126	85	98	107	87	103	108	95	107	111
17	46	116	114	108	120	122	123	94	106	116	126	116	111	102	109	115
17	48	80	99	140	120	122	123	103	112	117	127	116	111			
17	50	83	105	150												
GG-tilpassing best[4]		75 %	88 %	100 %	100 %	100 %	100 %	13 %	50 %	63 %	50 %	75 %	63 %	13 %	75 %	50 %

[1] I tre høve klarte vi ikkje å estimere parametrane til GG-funksjonen. Det skjedde i område 77 for $\beta = 40$ og $\alpha = 15$, 16 og 17.
[2] Tomme celler svarer til høve der vi ikkje har rekna ut noko.
[3] Resultata for $\alpha = 16$ er ikkje tatt med her.
[4] Talet på samanlikningar (N) er 8 heile vegen.

3.3. Numeriske samanlikningskriterium

For å få meir objektive samanlikningskriterium enn reint augnemål, har vi rekna ut verdet av tre mål til å samanlikne tilpassingane. For kvar glatting har vi rekna ut kvadratavviksummen for aldrane 20—30 år og for aldrane 15—50 år, og kjikvadratverdien for aldrane mellom α og β. For kvar Hadwiger-tilpassing har vi så rekna ut kor stort kvart av desse tre måla er i prosent av same mål ved den tilsvarande GG-tilpassinga. Eit grovt oversyn over utfallet av samanlikningane får ein av tab. 4 på side 304.

Når ein ser på kvadratavviksummen for aldrane frå 20 til 30 år, er dei to funksjonane omtrent like gode for dei døma vi har studert. Etter dei to andre kriteria er GG-funksjonen best i dei fleste høve.

3.4. Systematiske mønster

Det er viktig å finne ut om det er noko systematisk mønster i dei situasjonane der Hadwiger-funksjonen likevel er best. For å undersøke det, ser vi på resultata for dei einskilde områda. Nokre av desse resultata finn ein i tab. 6.

Det synest som om GG-funksjonen kjem betre ut av samanlikninga di større β er. Hadwiger-funksjonen er gjennomgåande best der kurvene er sterkt venstreskeive og har "tunge" høgrehaler, slik som for fertilitetsområda 38, 43 og 77. I desse områda ser det ut til at Hadwiger-funksjonen særleg er overlegen i dei sentrale aldrane.

3.5. Dei lågaste aldrane

Når ein nyttar GG-funksjonen, vil ein ofte få ei glatta fødselsrate på 0 for dei aller lågaste aldrane av di $g\ (x;\hat{\Theta}) = 0$ for $x \leqslant \hat{\mu} - \hat{\sigma}^2/(\hat{\mu} - \hat{m})$. (I tab. 5 finn ein verdien for $g\ (x;\hat{\Theta})$ for $x = 15$, 16 og 17 for alle kurvene i dette notatet.) Figurane 1, 5 og 6 og andre slike figurar tyder på at den tilpassinga ein får med GG-funksjonen blir dårlegare enn med Hadwiger-funksjonen for fødselsratene i dei aldrane der $g\ (x;\hat{\Theta}) = 0$.

Tab. 7 tyder på at dei to funksjonane er om lag jamgode i asymptotisk varians av dei glatta fødselsratene for aldrane over dei aller yngste, men for små x er Hadwiger-funksjonen best etter dette kriteriet og. (Utrekning av asymptotisk varians er forklart av Hoem og Berge, 1974 b.)

Tab. 7. Variasjonskoeffisienten for den glatta fødselsraten i Hadwiger-tilpassinga når $\alpha = \hat{\alpha}$ og $\beta = \hat{\beta}$ i prosent av variasjonskoeffisienten i den tilsvarande GG-tilpassinga

Område nr.	10	32	38	43	77
Alder					
16	89	61	5	(1)	(1)
17	98	103	78	74	81
20	99	106	101	104	108
24	103	99	100	101	99
28	101	108	105	105	107
32	104	104	106	108	105
37	101	88	99	99	98
44	103	102	102	101	103

[1] Både fødselsraten og den asymptotiske variansen er estimert til 0 under GG-tilpassinga her.

4. Sluttmerknad

I tidlegare granskingar har ein ikkje kunna legge fram avgjerande argument verken av teoretisk eller praktisk art for å velje ein gong for alle mellom Hadwiger-funksjonen og GG-funksjonen. Det ser ikkje ut til at vi kan gjere det på grunnlag av dei resultata som er lagde fram her heller. Vi har før hatt eit inntrykk av at Hadwiger-funksjonen er noko meir fleksibel enn GG-funksjonen, slik at han gir ei akseptabel tilpassing til eit breiare spektrum av fertilitetskurver. Dette inntrykket sit vi fortsatt med. Elles må ein seie at GG-funksjonen passar best til nokre fertilitetsmønster, medan Hadwiger-funksjonen har føremoner i samband med andre, særleg slike mønster som gir ei venstreskeiv kurve med ein tung høgrehale.

5. Litteraturliste

[1] Berge, Erling (1973): "Samfunnsstruktur og fruktbarhet". Upublisert hovedfagsoppgave. Universitetet i Bergen og Statistisk Sentralbyrå, Oslo.

[2] Berge, Erling (1974): "MINSYS: Eit reknemaskinprogram for analytisk glatting av befolkningsrater". Statistisk Sentralbyrå, ANO IO 74/11.

[3] Berge, Erling og Jan M. Hoem (1974): "Nokre praktiske røynsler med analytisk glatting". Statistisk Sentralbyrå, ANO IO 74/23.

[4] Hoem, Jan M. (1972): "On the statistical theory of analytic graduation". Proc. Sixth Berkeley Symp. Math. Statist. Prob., 1: 569–600.

[5] Hoem, Jan M. and Erling Berge (1974 a): "Some problems in Hadwiger fertility graduation". Statistisk Sentralbyrå, ANO IO 74/5.

[6] Hoem, Jan M. and Erling Berge (1974 b): "Theoretical and empirical results on the analytic graduation of fertility curves". Side 2–10 i Hoem et al.: "Two papers on analytic graduation". Statistisk Sentralbyrå, ANO IO 74/17.

Some practical experiences with analytic graduation

by E. Berge and J. M. Hoem

In connection with a study of regional fertility in Norway, the authors have tried to investigate whether the Hadwiger function is systematically better or worse as a graduating function than a function proportional to the density of the gamma distribution, called the GG function here. The results do not provide anything like an unambiguous answer. In a large number of cases, the GG function fits the data better than the Hadwiger function. On the other hand, the latter seems to be the more flexible one, and there is a lot of evidence that it may be superior when the fertility curve is strongly skewed to the left and has a heavy upper tail.

Once a graduating function has been selected, one may get considerable variation in the estimates of its parameters as well as in the curves fitted by varying the (single year) age classes included in the graduation process. This is particularly noticeable when the graduation is carried out by means of chi-square. It is less important (in our data) in least squares graduation. The reason is that the former method gives more weight to age classes where fertility is small. Since this is the case for ages above age 45, say, the upper tail of the graduating function will be pressed down towards the abscissa axis more decisively the more of the higher age classes one includes in the graduation. Even though fertility is low at the very early ages too, no similar effect has been seen in that tail in the present data.

Norstedts Tryckeri, Stockholm 1975