Arbeidsnotater

STATISTISK SENTRALBYRÅ

Working Papers from the Central Bureau of Statistics of Norway

IO 70/14 Oslo, 3.September 1970

THE SHIFTED HADWIGER FERTILITY FUNCTION

By

E. Gilje and L. Yntema

Contents

1. Introduction

The age-specific (female) fertility rates f(y) are usually defined by applying the conventional definition of the birth rate to the sub-population consisting of y-aged females. Without further notification only live-born girls are taken into account; in the cases where both live-born girls and boys are considered the rates may be indicated as "total". We distinguish between the above "gross" rates and the "net" rates p(y)f(y), where p(y) refers to the probability that a new-born girl will be alive at age y. Apart from proportionality factors these interdependent rates often show only small differences in a given situation.

In comparing different situations we may, of course, find more important deviations between the observed sets $f_o(y)$. However, in nearly all cases $f_o(y)$ is found to be a unimodal function with slightly varying skewness. Hence it seems justified to look for a mathematical function, which for appropriate values of its parameters may provide us with satisfactory graduations of the observed sets in a large number of situations. If such a model can be found, it will serve many purposes. It will be useful for additional specification, completion and correction of scanty population statistics. Extrapolating its (time-dependent) parameters we may use it in population projections; and in a similar way it may be used with actuarial calculations in social insurance (children's allowances, orphan pensions). It may also provide us with general and explicit solutions of the functional equations arising in demometric analysis.

Actuarial applications of the Hadwiger model discussed in this article have been studied by Yntema (1955, 1956).

2. The Hadwiger fertility model

The choice from the numerous functions, which might serve the above purpose, is restricted by demanding something more than general applicability. Thus we may require that the model function is mathematically tractable, containing only a few parameters with immediately clear meanings; to obtain rough estimations of these, preferably only a limited amount of calculations should be involved. Obviously these requirements are not entirely independent.

They are, to a large extent, fulfilled by the fertility model

$$f_H(y) = \frac{RH}{T\sqrt{\pi}}\left(\frac{T}{y}\right)^{3/2} \exp\left[-H^2\left(\frac{T}{y} + \frac{y}{T} - 2\right)\right] \qquad \begin{matrix} (y \geq 0) \\ (R,T,H > 0) \end{matrix} \qquad (1)$$

originally suggested by Hadwiger (1940). Integrating we find

$$\int_0^\infty f_H(y)\,dy = R \qquad \int_0^\infty y\, f_H(y)\,dy = RT$$

and thus the parameters R and T represents the gross reproduction rate and the mean age at childbearing, respectively. Using the observed fertility rates, the corresponding moments are

$$\hat{R} = \Sigma t_o(y) \quad \text{and} \quad \hat{R}\hat{T} = \Sigma y t_o(y) \qquad (2)$$

which give us estimators for R and T.

The parameter H, however, has no similar "meaning". Aiming at a good fit in the more important central part of the childbearing period we have refrained from making use of 2nd moments. Instead one can demand that $f_o(\tilde{T}) = f_H(\hat{T})$ where $\tilde{T}$ is the integer value obtained by rounding off $\hat{T}$. Solving this equation by inserting $\hat{R}$ for R and $\hat{T}$ for T in (1), we get as a first estimator for H:

$$\hat{H}_o = \frac{\hat{T}\sqrt{\pi}}{\hat{R}} f_o(\tilde{T}) \qquad (3)$$

Here it should be remarked that a rough graphical graduation of the central values of $f_o(y)$ may be useful before estimating T, if the observed fertility rates show a very rugged curve.

Obviously this method for obtaining rapid estimates of R, T and H, using a desk calculator only, remain useful also if a computer is available. Then, however, better estimates can be obtained by minimizing for instance $\Sigma\ (f_o - f_H)^2$. (In the following these two methods will be indicated by DSK and LSM respectively.)

3. The shifted Hadwiger function

Until resently the Hadwiger model in this version has given a good fit to the data, even if only DSK-estimates were used. See for instance the estimates for Oslo 1966 in table 1.

In discussing the fertility rates of Hungary 1956/1962 it was, however, demonstrated by Tekse (1967) that the Hadwiger model in spite of its seemingly general usefulness sometimes fails to produce an acceptable graduation. This was confirmes by Gilje (1969) in a subsequent investigation of local fertility rates in Norway.

This failure is easily understood, if we differentiate (1) to find the mode M_H. Then we get the equation $H = [3\, M_H \cdot T/2 \cdot (T^2 - M_H^2)]^{\frac{1}{2}}$. A new estimator for H is obtained from this by inserting $\hat{T}$ for T and $\hat{M}$ for M_H where $\hat{M}$ is the age for which the set $f_o(y)$ reaches its maximum value (see also the commentarys in section 4 about estimating M):

$$\hat{H}_1 = \left[\frac{3\,\hat{M}\,\hat{T}}{2(\hat{T}^2 - \hat{M}^2)}\right]^{\frac{1}{2}} \qquad (5)$$

Hence at least

$$\hat{M}/\hat{T} < 1 \qquad (6)$$

is required, but apart from that a good graduation is only to be expected if the estimators (3) and (5) give about equal results. Therefore the values $\hat{R}_o$, $\hat{T}_o$, $f_o(\tilde{T}_o)$ and $\hat{M}_o$ must approximately satisfy the relation which is found by setting $\hat{H}_o = \hat{H}_1$. Writing

$$\hat{a} = \frac{4}{3}\pi \left[\frac{\hat{T}\ f\ (\tilde{T}_o)}{\hat{R}}\right]^2 \qquad (7)$$

we thus obtain instead of (6) the sharper condition

$$\frac{\hat{M}}{\hat{T}} = \frac{\sqrt{1+a^2} - 1}{a} \simeq 1 - \frac{1}{\hat{a}} \qquad (8)$$

The more this relation is violated, the less satisfactory graduation of the observed fertility rates can be expected using the original Hadwiger model. This is illustrated in table 1 in the cases of Hungary and Japan.

In order to make the model more flexible the present writers introduced, independently, a fourth parameter d by shifting the origin of the age-axis. (See Gilje (1969) and Yntema (1969).) Accordingly the new model is defined by

$$f_{H',d}(y) = \frac{R'H'}{T'\sqrt{\pi}}\left(\frac{T'}{y-d}\right)^{3/2} \exp\left[-H'^2\left(\frac{T'}{y-d} + \frac{y-d}{T'} - 2\right)\right]; \qquad (9)$$

$$(y \geq d);\ (R', T', H' > 0),$$

and while using a computer-program the LSM-estimation of the four parameters needs no further explanation. In order to obtain DSK-estimators we write $y' = y-d$ and graduate $f_1(y') = f_o(y)$ using (1) with parameters R', T', H'.

From (2) and (3) we obtain

$$\left.\begin{aligned} \hat{R}' &= \Sigma f_1(y') = \Sigma f_o(y) = \hat{R} \\ \hat{R}'\hat{T}' &= \Sigma y' f_1(y') = \Sigma(y-d) f_o(y) = \hat{R}(\hat{T}-d) \\ \hat{H}_o' &= \hat{T}'\sqrt{\pi} f_1(\tilde{T}') \,/\, \hat{R}' = (\hat{T}-d)\sqrt{\pi} f_o(\tilde{T}) \,/\, \hat{R} \end{aligned}\right\} \quad (10)$$

where $\hat{R}'$, $\hat{T}'$ and $\hat{H}_o'$ are estimators for R', T' and H_o', respectively. $\tilde{T}'$ is the integer value obtained by rounding of $\hat{T}'$.

In (10) the fourth parameter, d, is assumed known. To find an estimator, $\hat{d}$, for this we use (8) which in the new situation takes the form

$$\frac{\hat{M}'}{\hat{T}'} = 1 - \frac{1}{\hat{a}'} \quad (11)$$

with $\hat{M}' = \hat{M}-\hat{d}$, $\hat{T}' = \hat{T}-\hat{d}$ and $a' = (1-\hat{d}/\hat{T})^2 \cdot a$ according to (7). Solving this with regard to d we get the DSK-estimator

$$\hat{d} = \frac{1-1/\hat{a} - \hat{M}/\hat{T}}{1-\hat{M}/\hat{T}}\,\hat{T} \quad (12)$$

(hence, if (8) exactly holds, $\hat{d} = 0$ as it should be).

4. Some examples

In table 1 we have given estimates for the parameters in these fertility models for some different situations. The estimates indicated by DSK are calculated according to (2) and (3) in the non-shifted cases, and according to (10) and (12) in the shifted cases. The LSM-estimates have been found by minimizing $\Sigma(f_1(y') - f_{H',d}(y))^2$ with regard to the four unknown parameters in (9). M/T and 1-1/a refer to (8). We have used the sum of squares of deviation between the observed values and the estimated values as a measure of the goodness of fit.

Observe that the non-shifted DSK-estimates of H are all except in the case of Japan, very close to 3. We have found this same property in several other cases not included in this paper.

The estimates by (12) of the shift-parameter, d seem to be good in the case of Hungary but using this estimator in the four remaining cases we get somewhat confusing results. The data for Rotterdam 1937 for instance lead to a $\hat{d} = -28{,}3$ which gives a wholly unsuitable function.

An examination of the observed fertility rates gives us a clue to where the trouble-maker may be found. (See figs. 1 and 2.) The observed mode $\hat{M}$ is not where one should expect it to be, or rather where one wants it to be. As (12) is very sensitive to variations in $\hat{M}$, a maximum observed fertility rate for an "unexpected" age leads to a $\hat{d}$ making no sense. In the Rotterdam-case for instance we have found $\hat{M} = 28$. If we instead had used $\hat{M} = 26$ this would have lead to a $\hat{d}$ - 4.7 which is very close to the LSM-estimate of d. Similar adjustments of $\hat{M}$ can be made for the two Norwegian cases as well, leading to reasonable shifts in (9).

We have examplified this problem in full in the case of Japan. Here the observed $\hat{M} = 25$ leads to a $\hat{d} = 16.00$ which is too large compared to the shift obtained after using the LSM (see table 1). It is in fact so large that it had been better not using a shift at all according to the respective sums of squares of deviation. In fig. 2 we have drawn this DSK-curve together with the LSM-curve and the observed rates. It is quite obvious that the observed rate for the age 25 is irregular. Just looking at these rates one should expect the mode to come somewhat later. Thus, setting $\hat{M} = 26$, we see from both table 1 and 2 that the fit is substantially better.

As well as finding the LSM-estimates minimizing $\Sigma(f_1(y') - f_{H,d}(y))^2$ we also minimized the sum of absolute differences, $\Sigma \mid f_1(y') - f_{H,d}(y) \mid$, with regard to the unknown parameters. This technique gave approximately the same estimates as the LSM, and we have therefore not included these in the examplifications.

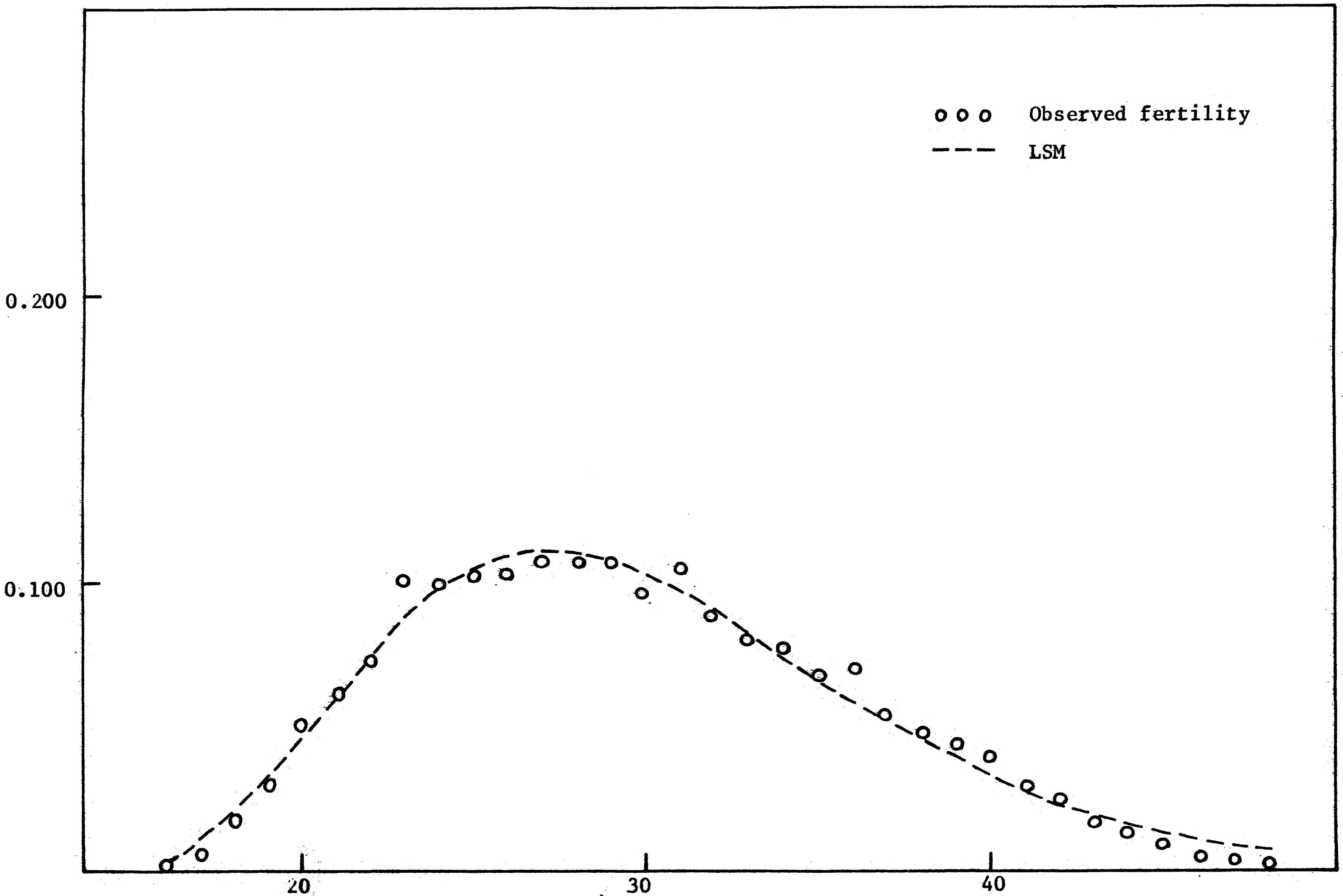

Fig. 1. Observed and smoothed fertility. Rotterdam 1937

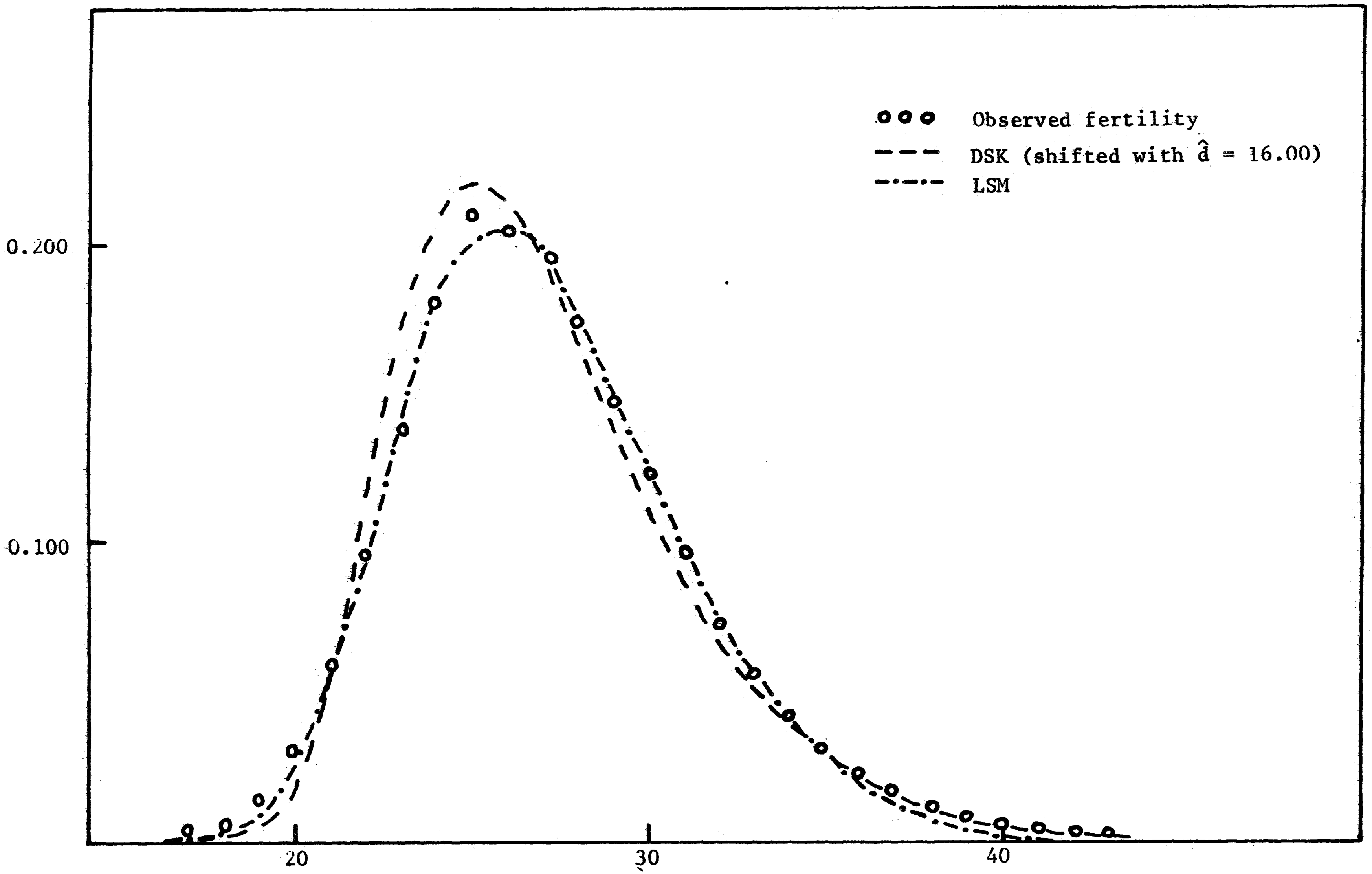

Fig. 2. Observed and smoothed fertility. Japan 1963

Table 1. Estimated parameters for some different situations

Parameters	R	T	H	d	M/T	1-1/a	Sum of squares of deviation, mult. by 10^6
Rotterdam 1937 (total, gross)							
DSK (non-shifted)	1.81	29.5	3.0	-	0.949	0.900	1 201
LSM	1.87	26.1	2.5	4.04	-	-	1 048
Hungary 1961 (total, gross)							
DSK (non-shifted)	1.92	25.1	3.0	-	0.837	0.916	6 096
DSK (shifted)	1.92	12.9	1.5	12.17	0.837	0.916	824
LSM	1.96	12.6	1.4	13.05	-	-	168
Japan 1963 (total, gross)							
DSK (non-shifted)	1.97	27.2	4.7	-	0.919	0.966	1 846
DSK (shifted using the observed $\hat{M}$ = 25)	1.97	11.2	1.9	16.00	0.919	0.966	2 590
DSK (shifted using $\hat{M}$ = 26)	1.97	21.0	3.7	6.18	0.956	0.966	706
LSM	1.95	16.0	2.8	11.23	-	-	218
Oslo 1966 (total, gross)							
DSK (non-shifted)	2.00	26.5	3.2	-	0.943	0.922	3 261
LSM	2.07	26.8	2.9	0.10	-	-	2 324
Norway 1966 (total, gross)							
DSK (non-shifted)	2.83	26.7	2.8	-	0.936	0.906	4 020
LSM	2.96	18.2	1.8	9.26	-	-	1 468

Table 2. Observed and estimated fertility rates for Japan 1963 (total, gross) multiplied by 10^3

Age	Observed rates	Shifted DSK-graduation with $\hat{d}$ = 16.00	Shifted DSK-graduation with $\hat{d}$ = 6.18	LSM-graduation
17	2	0	1	0
18	5	0	4	1
19	13	1	13	8
20	28	12	32	25
21	58	49	61	57
22	96	108	98	100
23	140	167	138	145
24	183	207	172	182
25	212	223	196	203
26	207	217	204	208
27	196	197	199	197
28	175	170	182	177
29	149	141	158	151
30	125	114	131	124
31	97	90	104	98
32	74	69	80	76
33	56	53	60	57
34	43	40	43	42
35	31	30	31	30
36	23	22	21	21
37	17	16	14	15
38	12	12	10	10
39	9	9	6	7
40	6	6	4	5
41	5	5	3	3
42	3	3	2	2
43	2	2	1	1
44	1	2	1	1
45	1	1	0	1
46	0	1	0	0
47	0	1	0	0
48	0	0	0	0
49	0	0	0	0

Source: Yamaguchi (1965).

5. Restrictions

After these apparently satisfactory results we were tempted to test the validity of the shifted model in less recent situations. Therefore we studied the fertility rates of Rotterdam 1870, 1890 and 1909 given by Angenot (1966). In this way certain restrictions were found which, however, seem of little importance in contemporary situations.

In view of their irregular behaviour we may feel suspicious about the reliability of the observed rates for 1870, as shown by Fig. 3, but in spite of this it is clear that the observed $\hat{M}$ is somewhere near age 35. The mean fertile age $\hat{T}$, however, is found to be 32.5, which means that (11) cannot be fulfilled by any shift $\hat{d}$. This is reflected by a "meaningless" LSM-estimate for d, viz. -526 and therefore also for the other parameters. The goodness of fit, though, was remarkably good with a sum of squares of deviation amounting to $14002 \cdot 10^{-6}$ (see table 3 for comparison). Who would think of trying estimates like this using the DSK-method, however? Thus, in cases (nowadays very unusual) where $\hat{M} > \hat{T}$, we may, perhaps, obtain "normal" estimates in the Hadwiger-model if the direction of the age-axis is reversed.

We have therefore replaced the age-parameter y in (9) by t = 67-y. The reversed origin of the axis, 67, was chosen out of convenience. Another choice will lead to a different shift, $\hat{d}$, and therefore also different $\hat{T}'$ and $\hat{M}'$, but the resulting estimated fertility rates are of course invariant to where this origin is placed.

Since a good fit was, in view of the observed scatter, not to be expected, the results are not entirely unsatisfactory, as can be seen from table 3 and fig. 3.

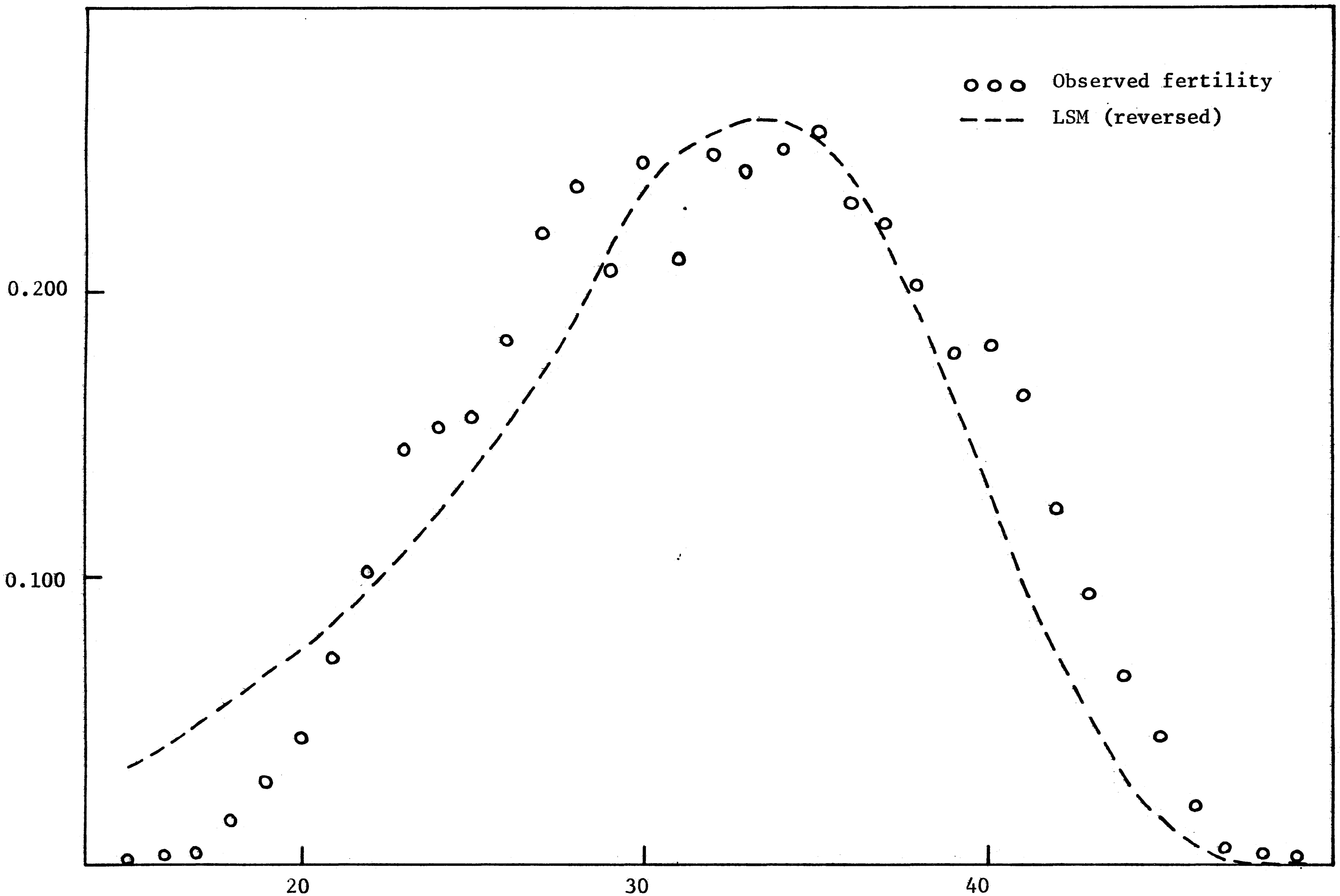

Fig. 3. Observed and smoothed fertility. Rotterdam 1870

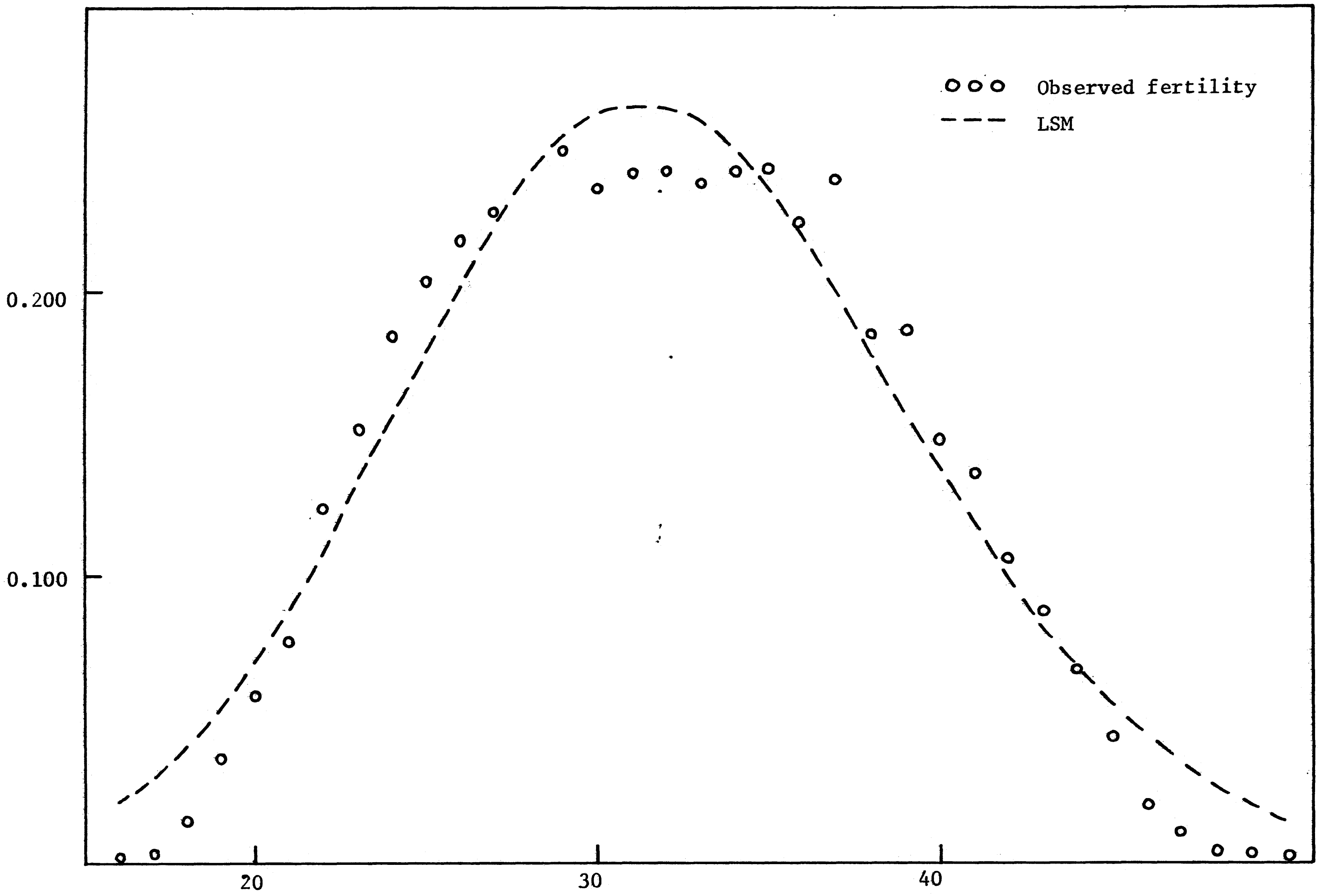

Fig. 4. Observed and smoothed fertility. Rotterdam 1890

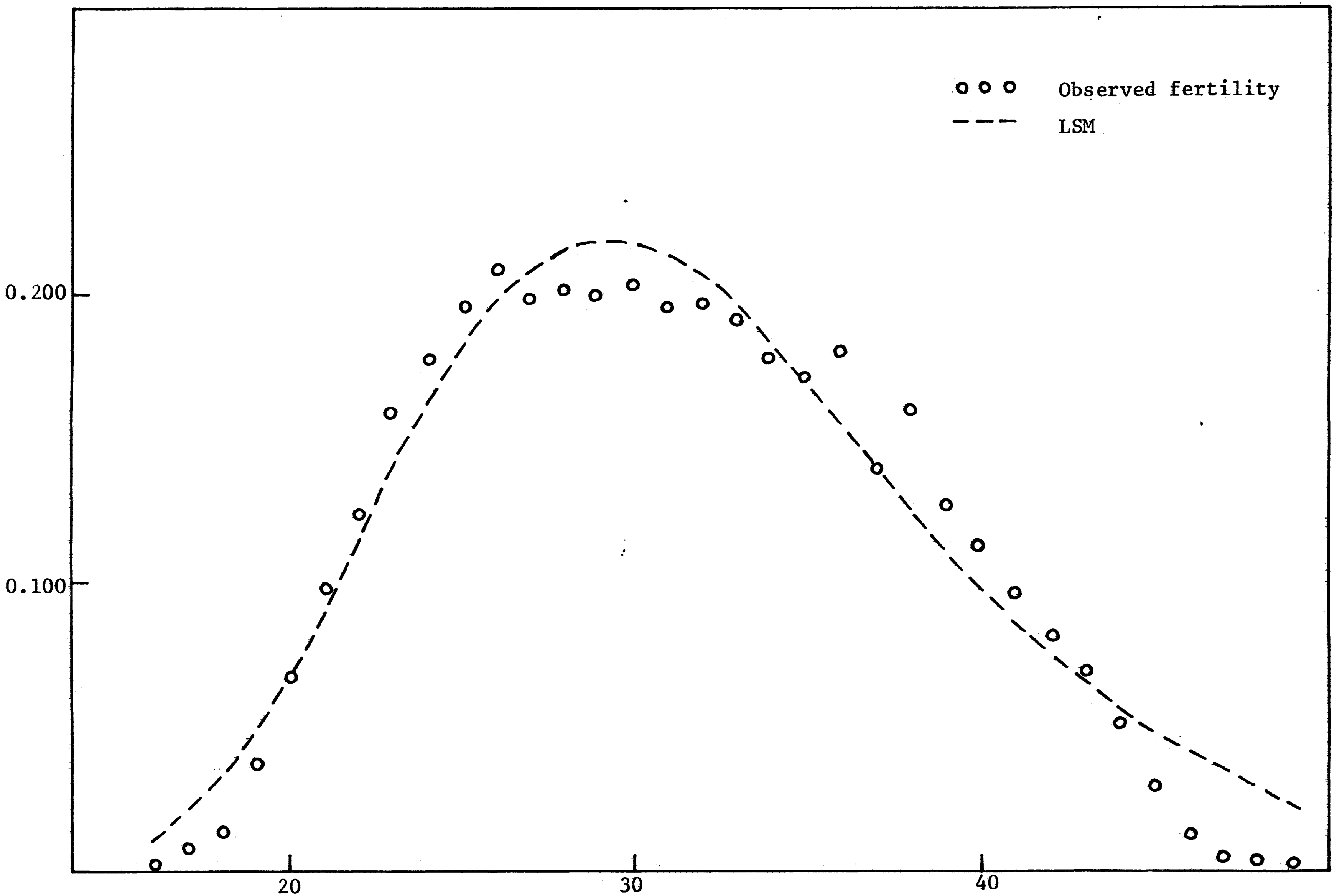

Fig. 5. Observed and smoothed fertility. Rotterdam 1909

From figs. 3, 4, and 5 we see that the same problem as mentioned in section 4 arise when we want to determine the mode. If we try to imagine that the observed rates have already been smoothed, though, it seems that the above relation between mode and mean gradually changes into the opposite direction. In 1890 the mode is approximately equal to the mean which should lead to a large shift according to (12). This was also obtained by the LSM. In this situation it is difficult to decide wether the age-axis should be reversed or not. A reversal would probably, however, make little difference for the estimated parameters. In table 3 only the non-shifted DSK-estimates are given as a comparison to the LSM-estimates.

Table 3. Estimated parameters for Rotterdam (total, gross)

Parameters	1870[1]		1890		1909	
	DSK	LSM	DSK	LSM	DSK	LSM
d	-	8.94	-	-81.43	-	-0.70
R	4.533	4.740	4.698	4.880	3.869	4.074
T	34.50	27.09	31.88	113.44	31.09	32.60
H	3.332	2.383	2.913	10.945	2.850	2.900
Sum of squares of deviation, mult. by 10^6	-	21 959	-	12 354	-	8 966

1) In (9) the age-parameter is replaced by $t = 67-y$ in this example.

In the last case, Rotterdam 1909, the inequality, $\hat{M} < \hat{T}$, seems to be fulfilled again. Still, the sum of squares of deviation is remarkably higher in this case than in any of the cases in table 1. Some of this can be explained by the fact that these sums must be more or less propotional to the reproduction rate R. If we divide the sums of squares of deviation into $\hat{R}$ in the cases of Oslo (the least satisfactory result in table 1) and of Rotterdam 1909, however, we get $1123 \cdot 10^{-6}$ and $2201 \cdot 10^{-6}$ respectively. The scatter of the observed rates does not seem to justify a difference like this, and we have to search for the explanation elsewhere.

We had the impression that good results could not be expected because the observed rates run rather flatly in the central part of the childbearing age interval. By differentiating (1) twice we obtain for the mode M of $f_H(y)$:

$$f_H''(M) = -\frac{3}{2}\,\frac{T^2+M^2}{T^2-M^2}\,\frac{1}{M^2}\,f_H(M) \qquad (13)$$

which have a large absolute value for $T \approx M$. Consequently small differences between mode and mean of the observed rates are inconsistent with a very flat curve if the Hadwiger model is to be applied. It is therefore impossible to obtain a good fit using this model if the observed rates are both symmetric and flat around the mean.

6. References

Angenot, L.H.J. 1966. De vruchtbaarheid van de vrouwenbevolking te Rotterdam in 1870. Rotterdams Jaarboekje, 1966, 249- ?

Gilje, E. 1969. Fitting curves to age-specific fertility rates: Some examples. Statistisk Tidskrift 7, 118-134.

Hadwiger, H. 1940. Eine analytische Reproduktionsfunktion für biologische Gesamtheiten. Skand. Akt. Tidskr. 23, 101-113.

Tekse, K. 1967. On demographic models of age-specific fertility rates. Statistisk Tidskrift 5, 189-207.

Yamaguchi, Kiichi 1965. Standardized vital rates, reproduction rates and intrinsic vital rates in Japan: 1963. The Journal of Population Problems (Jinko Mondai Kenkyu 96, 36 - 48.

Yntema, L. 1955. Approximate valuation for orphans' pensions. Het Verzekeringsarchief 32, 71- ?

\- - - 1956. An approximation of the family structure. Proc. of the First Internat. Conf. of Social Security Actuaries and Statisticians, Brussel 3, 351- ?

\- - - 1969. On Hadwiger's fertility function. Statistisk Tidskrift 7, 113-117.