Arbeidsnotater

STATISTISK SENTRALBYRÅ
Dronningensgt. 16, Oslo-Dep., Oslo 1. Tlf. 41 38 20, 41 36 60

WORKING PAPERS FROM THE CENTRAL BUREAU OF STATISTICS OF NORWAY

IO 74/17 25 March 1974

TWO PAPERS ON ANALYTIC GRADUATION

by Jan M. Hoem et al.

CONTENTS

THEORETICAL AND EMPIRICAL RESULTS ON THE ANALYTIC GRADUATION OF FERTILITY CURVES

by Jan M. Hoem and Erling Berge

1. INTRODUCTION

1 A. The diagram of age-specific fertility rates for a population, based on data for a calendar period, say, will typically picture a curve which looks much like a left-skewed unimodal probability density, such as a gamma density or some of the beta densities (starting just below age 15), for instance, but with superposed fluctuations. Unless the population is very large, the diagram of the sequence of fertility rates, plotted against age, will have quite a ragged appearance. It is frequently assumed that "real fertility" would be portrayed by a smooth curve and that the irregularities of the observed curves are due to accidental circumstances. The observed fertility rates are then regarded as "raw" or primary estimates of the underlying "real" rates, and graduation is employed to get a smoother curve.

1 B. When the smooth curve is produced by fitting a nice, parametric function to the original data, we call it analytic graduation. The present paper reports some findings on the following question: Given that one has chosen a particular graduating function with a specified parametrization, which method should one select to fit the function to the observed fertility data?

Some previous research (Hoem, 1972) has shown that a modified minimum chi-square method cannot be outdone by any other weighted or unweighted least squares method or by any moment method, in the sense that asymptotically as the population size increases the modified minimum chi-square estimators have minimal variances within the class of estimators considered. This holds for the estimators of the parameters of the fitting function as well as for the estimators of the function values (i.e., of the "true" fertility rates at all ages). So far, one has not known, however, to what extent this theoretical result has practical consequences in that the numerical values of the asymptotic variances of chi-square estimators are noticeably smaller than the corresponding variance values for reasonable competitors like least squares estimators.

1 C. We have applied analytic graduation methods to a number of empirical fertility curves, and this paper reports briefly on some selected typical findings. It turns out that estimated asymptotic variances (and the corresponding estimated coefficients of variation) almost without exception are much smaller for rates graduated by least squares than for the original ungraduated rates. In every case which we have investigated, there is also some further real gain in estimated variance in going from least squares graduation to minimum chi-square graduation. On a rare occasion, the extra gain in variance is as "low" as some 10 per cent, but it frequently is higher, and it can be substantial. Thus the optimality (in terms of estimated asymptotic variances) of the modified minimum chi-square method does have considerable practical interest.

1 D. The presentation goes as follows:

In Section 2 we summarize results given in the previous paper, but restated here in a form meant to be more easily accessible. The account is largely phrased in terms of fertility graduation, and it is of course relevant to this problem, but the reader should bear in mind that the theory is by no means limited to the case of fertility. It can be applied to any type of occurrence/exposure rate.

Section 3 then contains the selected empirical results. They are based on curves of observed fertility rates, and they are relevant for such curves as well as for other curves of vital rates of a similar form, such as marriage rates, rates of migration, etc.

2. THEORETICAL RESULTS

2 A. The "raw" age-specific fertility rates in a set will have been calculated for a number of age groups, which for simplicity we shall take to be the single-year age intervals $\alpha, \alpha+1, \ldots, \beta$. Let $\hat{\lambda} = \{\hat{\lambda}_\alpha, \hat{\lambda}_{\alpha+1}, \ldots, \hat{\lambda}_\beta\}'$ denote the vector of "raw" rates, the prime signifying a transpose. These rates are to be graduated by means of some sequence of parametric functions

$$g(\Theta) = \{g_\alpha(\Theta), g_{\alpha+1}(\Theta), \ldots, g_\beta(\Theta)\}',$$

that is, one is required to select some value $\hat{\Theta}$ of the vector $\Theta = (\Theta_1, \ldots, \Theta_r)'$ such that $g(\hat{\Theta})$ fits as well as possible to the "raw" rates in $\hat{\lambda}$. In our particular empirical situation, $g_x(\Theta)$ is the value of, say, the Hadwiger function at argument x, so that

$$g_x(\Theta) = \frac{RH}{T\sqrt{\pi}}\left(\frac{T}{x-D}\right)^{3/2} \exp\left\{-H^2\left(\frac{T}{x-D} + \frac{x-D}{T} - 2\right)\right\}.$$

Other investigations (Hoem and Berge, 1974) have shown that it pays to take

$$\Theta_1 = R,\quad \Theta_2 = D+T\{(1+\tfrac{16}{9}H^4)^{1/2} - 1\}/(\tfrac{4}{3}H^2),\quad \Theta_3 = T+D, \text{ and } \Theta_4 = \tfrac{1}{2}T^2/H^2$$

as the basic parameters of the graduation. (If we regard $g_x(\Theta)$ as a function of a continuous x, $g_x(\Theta)/R$ then becomes a probability density with mode Θ_2, mean Θ_3, and variance Θ_4.)

We shall take this to mean that there is some underlying "true" vector $\lambda^0 = \{\lambda_\alpha^0, \ldots, \lambda_\beta^0\}'$ of fertility rates for the age intervals in question, and that the task at hand is to remove as well as possible the random deviation $\hat{\lambda} - \lambda^0$. For our purposes, λ^0 can be represented with sufficient accuracy by $g(\Theta^0)$, where Θ^0 is a corresponding "true" value of Θ. Then, Θ^0 is estimated by $\hat{\Theta}$, and λ^0 by $g(\hat{\Theta})$.

2 B. The fitting can be done in various ways, but the ordinary (i.a., unweighted) least squares method and the modified minimum chi-square method are immediate and prominent candidates. They consist, of course, of minimizing

$$Q_{LS}(\Theta) = N\sum_{x=\alpha}^{\beta}\{\hat{\lambda}_x - g_x(\Theta)\}^2$$

and

$$Q_{CS}(\Theta) = \sum_{x=\alpha}^{\beta}\frac{\{\hat{\lambda}_x - g_x(\Theta)\}^2}{\hat{\sigma}_x^2/N} = \sum_{x=\alpha}^{\beta}\frac{\{B_x - L_x g_x(\Theta)\}^2}{B_x}.$$

Here, N denotes the total number of women under observation, i.e., the number of women who are potential contributors to the numbers of liveborn babies counted. B_x is the number of liveborn babies with mothers at age x at the time of childbearing, and L_x is the total number of personyears observed at age x, so that $\hat{\lambda}_x = B_x/L_x$. Furthermore, $\hat{\sigma}_x^2 = N\hat{\lambda}_x/L_x$. We can regard $\hat{\sigma}_x^2/N$ as an estimator of the asymptotic variance of $\hat{\lambda}_x$. These two methods are members of a whole class of procedures which are generated in the following way:

Let $\underset{\sim}{M}$ be a square matrix of elements m_{xy} $(x, y = \alpha, \alpha+1, \ldots, \beta)$ which may (but need not) be random variables (by depending on the data). Let

$$\begin{aligned} Q_{\underset{\sim}{M}}(\underset{\sim}{\Theta}) &= N\{\underset{\sim}{\hat{\lambda}} - \underset{\sim}{g}(\underset{\sim}{\Theta})\}' \underset{\sim}{M}\{\underset{\sim}{\hat{\lambda}} - \underset{\sim}{g}(\underset{\sim}{\Theta})\} \\ &= N \sum_x \sum_y m_{xy}\{\hat{\lambda}_x - g_x(\underset{\sim}{\Theta})\}\{\hat{\lambda}_y - g_y(\underset{\sim}{\Theta})\}, \end{aligned}$$

and assume that there exists a value of $\underset{\sim}{\Theta}$, say $\underset{\sim}{\hat{\Theta}}(\underset{\sim}{M})$, which minimizes $Q_{\underset{\sim}{M}}(\underset{\sim}{\Theta})$. Then $\underset{\sim}{g}\{\underset{\sim}{\hat{\Theta}}(\underset{\sim}{M})\}$ represents one possible graduation of $\underset{\sim}{\hat{\lambda}}$.

If $\underset{\sim}{M} = \underset{\sim}{I}$, the graduation is by means of least squares.

Let $\underset{\sim}{\hat{\sigma}} = \text{diag}\,(\hat{\sigma}_\alpha^2, \hat{\sigma}_{\alpha+1}^2, \ldots, \hat{\sigma}_\beta^2)$. Then, if $\underset{\sim}{M} = \underset{\sim}{\hat{\sigma}}^{-1} = \text{diag}\,(1/\hat{\sigma}_\alpha^2, \ldots, 1/\hat{\sigma}_\beta^2)$, $\underset{\sim}{g}\{\underset{\sim}{\hat{\Theta}}(\underset{\sim}{M})\}$ represents a graduation by modified minimum chi-square.

<u>2 C</u>. In order to choose between the various possible methods of fitting $\underset{\sim}{g}(\underset{\sim}{\Theta})$ to $\underset{\sim}{\hat{\lambda}}$, we must know something about their statistical properties. A framework for a discussion of these can be found in a previous paper (Hoem, 1972), and we shall recapitulate a few of the main results stated there. These results hold under assumptions which were given in that paper, and which we shall not repeat in full here, but which we can reasonably assume to hold for the kind of practical situation which we have in mind.

We shall assume, then, that as the population size N increases, $\underset{\sim}{\hat{\lambda}}$ is asymptotically multinormally distributed with mean $\underset{\sim}{\lambda}^0$ and some covariance matrix $\underset{\sim}{\Sigma}_0/N$. [In our empirical studies of fertility, we take $\underset{\sim}{\Sigma}_0$ to be diagonal, i.e., we take the $\hat{\lambda}_x$ to be asymptotically independent.]

We shall assume also that as $N \to \infty$, $\underset{\sim}{M}$ converges in probability to some positive definite matrix $\underset{\sim}{M}_0$.

Then $\underset{\sim}{\hat{\Theta}}(\underset{\sim}{M})$ is asymptotically multinormally distributed with mean $\underset{\sim}{\Theta}^0$ and a covariance matrix $\underset{\sim}{\Sigma}/N$, which satisfies

$$(2.1) \qquad \underset{\sim}{\Sigma} = \underset{\sim}{A}\,\underset{\sim}{\Sigma}_0\,\underset{\sim}{A}',$$

with

$$(2.2) \qquad \underset{\sim}{A} = (\underset{\sim}{J}_0'\,\underset{\sim}{M}_0\,\underset{\sim}{J}_0)^{-1}\,\underset{\sim}{J}_0'\,\underset{\sim}{M}_0\,.$$

Here, $\underset{\sim}{J}_0 = \underset{\sim}{J}(\underset{\sim}{\Theta}^0)$, where

$$\underset{\sim}{J}(\underset{\sim}{\Theta}) = \begin{Bmatrix} \frac{\partial}{\partial\Theta_1} g_\alpha(\underset{\sim}{\Theta}), & \ldots, & \frac{\partial}{\partial\Theta_r} g_\alpha(\underset{\sim}{\Theta}) \\ \cdots & \cdots & \cdots \\ \frac{\partial}{\partial\Theta_1} g_\beta(\underset{\sim}{\Theta}), & \ldots, & \frac{\partial}{\partial\Theta_r} g_\beta(\underset{\sim}{\Theta}) \end{Bmatrix}.$$

Similarly, $\underset{\sim}{g}(\hat{\underset{\sim}{\theta}}(\underset{\sim}{M}))$ is asymptotically multinormally distributed with mean $\underset{\sim}{\lambda}^0 = \underset{\sim}{g}(\underset{\sim}{\theta}^0)$ and (singular) covariance matrix $\underset{\sim}{J}_0 \underset{\sim}{\Sigma} \underset{\sim}{J}_0'/N$. If, in particular, $\underset{\sim}{M}_0 = \underset{\sim}{\Sigma}_0^{-1}$, then $\underset{\sim}{\Sigma}$ is equal to

$$\underset{\sim}{\Sigma}_{0,0} = (\underset{\sim}{J}_0' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{J}_0)^{-1} .$$

This will be the case in our practical situation if we use the modified minimum chi-square method. In the general case, $\underset{\sim}{\Sigma}$ will equal $\underset{\sim}{\Sigma}_{0,0}$ if $\underset{\sim}{M} = \hat{\Sigma}_0^{-1}$, where $\hat{\underset{\sim}{\Sigma}}_0$ is a consistent estimator of $\underset{\sim}{\Sigma}_0$.

One can prove that for any M_0, $\underset{\sim}{\Sigma} - \underset{\sim}{\Sigma}_{0,0}$ will be positive semidefinite, and so will $\underset{\sim}{J}_0 \underset{\sim}{\Sigma} \underset{\sim}{J}_0' - \underset{\sim}{J}_0 \underset{\sim}{\Sigma}_{0,0} \underset{\sim}{J}_0'$. This means, among other things, that the asymptotic variance of each $\hat{\theta}_i(\underset{\sim}{M})$ will be minimized if $\underset{\sim}{M}$ is chosen such that $\underset{\sim}{M}_0 = \underset{\sim}{\Sigma}_0^{-1}$. Similarly, the asymptotic variance of each $g_x(\hat{\underset{\sim}{\theta}}(\underset{\sim}{M}))$ will be minimized by the same choice of $\underset{\sim}{M}$. Thus, in our particular empirical investigation, the use of <u>this</u> criterion of optimality will imply that one should use modified minimum chi-square graduation or some other method with the same asymptotic properties, such as the maximum likelihood method described by Hoem (1972).

One can prove also that $\underset{\sim}{\Sigma}_0 - \underset{\sim}{J}_0 \underset{\sim}{\Sigma}_{0,0} \underset{\sim}{J}_0'$ is positive semidefinite, so that chi-square graduation must result in some non-negative gain in asymptotic variance of the age-specific fertility rates.

3. SOME EMPIRICAL RESULTS

<u>3 A</u>. We have applied the theory of the previous Section to a substantial number of Norwegian fertility curves, and a comprehensive presentation of our empirical results will be given in a forthcoming Working Paper. The present Section contains a brief account of two such cases for purposes of illustration.

<u>3 B</u>. We have selected the fertility curve for the city of Oslo, 1968-71, as well as the curve for the same period for an aggregate of 13 communes on the Norwegian West Coast, viz., Kvitsøy, Bokn, Utsira, Austevoll, Sund, Øygarden, Austrheim, Fedje, Solund, Askvoll, Selje, Sande, and Giske. The observed fertility curves have been plotted in Figures 1 and 2, respectively. The curve for Oslo represents a low level of fertility and a rather symmetrical age pattern of fertility for present-day Norway. (The total fertility rate by the chi-square fit is 1.9, while the corresponding modal and mean ages at childbearing are 25.2 and 27.3, respectively.) In contrast to this, the other curve represents a high fertility level for the Norway of to-day (TFR = 3.5) and a particularly skew age-pattern. (The modal and mean ages at childbearing by the chi-square fit are 24.1 and 28.4, respectively.) These curves thus correspond to quite different patterns of fertility.

<u>3 C</u>. For each set of data, the "raw" fertility rates $\hat{\lambda}_x$ have been calculated for females for single-year age intervals, counting live offspring of both sexes. Age x corresponds to age as of December 31 in any observational year, and the contribution to L_x of any year is the arithmetic mean of the number of x-year olds at the beginning and end of the year, age being calculated as of Dec. 31 of <u>that</u> year.

The Hadwiger function of Subsection 2 A, with the parametrization $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_4)$ given there, has been fitted to the rates for ages 16 to 44. Then θ_1 corresponds to the total fertility rate, θ_2 to the modal age of childbearing, θ_3 to the corresponding mean age, and θ_4 to the variance of the age-pattern of fertility.

We have made graduations based on the method of least squares as well as separate graduations of the same data based on the modified minimum chi-square method. The graduated rates have been plotted in both Figures. Eyeball inspection judges the fit to be acceptable. As compared with the fit to other curves we have investigated, that of the two examples is about average.

The numerical calculations have been carried out on a Honeywell-Bull H6060 computer by means of a program developed by Berge (1974) on the basis of various algorithms previously published by others.

3 D. Let us denote the min.-χ^2 estimator for $\underset{\sim}{\Theta}$ by $\underset{\sim}{\hat{\Theta}}$, and the least squares estimator by $\underset{\sim}{\Theta}^*$. The corresponding estimates for the two sets of data mentioned have been listed in Table 1. In the case of the curve for Oslo, the estimates are pretty much the same for both methods. For the other curve, the two methods give somewhat different parameter estimates.

3 E. We now turn to variance estimation. We have proceeded as follows. Let $\underset{\sim}{\hat{J}} = \underset{\sim}{J}(\underset{\sim}{\hat{\Theta}})$, $\underset{\sim}{\hat{A}} = \{\underset{\sim}{\hat{J}}'(\underset{\sim}{\hat{\sigma}}/N)^{-1}\underset{\sim}{\hat{J}}\}^{-1}\underset{\sim}{\hat{J}}'(\underset{\sim}{\hat{\sigma}}/N)^{-1}$, $\underset{\sim}{A}^* = (\underset{\sim}{\hat{J}}'\underset{\sim}{\hat{J}})^{-1}\underset{\sim}{\hat{J}}'$, $\underset{\sim}{\hat{\Sigma}}/N = \underset{\sim}{\hat{A}}(\underset{\sim}{\hat{\sigma}}/N)\underset{\sim}{\hat{A}}'$, and $\underset{\sim}{\Sigma}^*/N = \underset{\sim}{A}^*(\underset{\sim}{\hat{\sigma}}/N)(\underset{\sim}{A}^*)'$.

Then, $\underset{\sim}{\hat{\Sigma}}/N$ is our estimator of the asymptotic covariance matrix of $\underset{\sim}{\hat{\Theta}}$, and $\underset{\sim}{\Sigma}^*/N$ is our estimator of the corresponding matrix for $\underset{\sim}{\Theta}^*$, on the assumption that the true age-specific fertility rates are represented adequately by the Hadwiger curve. Similarly,

$$\underset{\sim}{\hat{J}}(\underset{\sim}{\hat{\Sigma}}/N)\underset{\sim}{\hat{J}}' \quad \text{and} \quad \underset{\sim}{\hat{J}}(\underset{\sim}{\Sigma}^*/N)\underset{\sim}{\hat{J}}'$$

are our estimators of the asymptotic covariance matrices of $\underset{\sim}{g}(\underset{\sim}{\hat{\Theta}})$ and $\underset{\sim}{g}(\underset{\sim}{\Theta}^*)$, respectively.

Estimates of the asymptotic variances of the $\hat{\Theta}_i$ and Θ_i^* have been computed and listed in Table 2. The chi-square method seems impressively much better than the method of least squares for these estimands.

Corresponding variance estimates for graduated age-specific fertility rates for selected ages have been listed in Table 3, along with estimates for coefficients of variation. Column 4 shows that chi-square graduation of the "raw" rates can result in a substantial reduction in variance, by a factor of 2 to 18 in the cases reported here. Column 3 shows that there is some real gain in using the chi-square method rather than least squares for these estimands as well. For the central ages (in the twenties), the gain is some ten to twenty per cent. In the tails of the curve, the chi-square method is at least twice as good as least squares, as judged by the variance estimates. Given the much larger weight placed on the tail ages by the former method than by the latter, such a pattern is not surprising, of course.

The variance reduction achieved by using least squares estimates rather than the "raw" rates can be gauged by dividing each of the entries in column 4 by the corresponding entry in column 3. For most ages, the variance is reduced by a factor of 2 to 12. For age 16 in each of the two cases reported, however, the estimated asymptotic variance of the least squares estimate of the fertility rate exceeds that of the "raw" rate (by 15 and 48 per cent for the two curves, respectively). Such "reversals" occur occasionally at isolated ages in our data sets.

Columns 7 and 8 give similar results for the coefficient of variation.

These numerical estimates show that for this type of fertility curves, the optimality of the modified minimum chi-square method is of practical importance and is not only of theoretical interest. This conclusion is substantiated by the rest of our experience with fertility curves.

3 F. The variance estimators introduced above are functions of the chi-square estimator $\underset{\sim}{\hat{\Theta}}$. We have computed similar variance estimates based on the least squares estimator $\underset{\sim}{\Theta}^*$. The numerical results do not seem to contain much beyond what we report here.

3 G. The modified minimum chi-square value for the curve of the 13 communes is 41.9, which corresponds to the upper 1.8 percentage point of the chi-square distribution with 25 (= 29 - 4) degrees of freedom. Most of the chi-square values which we have calculated for curves of regional populations in Norway are better than this. For the larger communities, the chi-square value is invariably big, however, and our curve for Oslo is a case in point. (See line 5 of Table 1.) We take this as an indication of systematic deviations between the observed and the graduated curves. This corresponds to earlier findings by others in attempts at analytic graduation of mortality rates.

Sometimes, deviations between the observed and the graduated curve are regarded as caused at least in part by particular historical events of some substantive interest, or as an effect of cohort differences. In such a case, the graduated curve can still be useful in providing a "standard" with which the observed curve is compared to bring out systematic deviations more clearly. We regard this as an important aspect of fertility graduation (and a potentially more realistic one than the straightforward superposed random fluctuations interpretation), and we plan to discuss it in a future communication.

REFERENCES

[1] Berge, Erling (1974): "MINSYS: A computer program for the analytic graduation of vital rates." Central Bureau of Statistics of Norway, Working Paper IO 74/11, to appear (in Norwegian).

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

[3] Hoem, Jan M. and Erling Berge (1974): "Some problems in Hadwiger fertility graduation." Central Bureau of Statistics of Norway, Working Paper IO 74/5.

OBSERVED AND GRADUATED FERTILITY CURVES.

Females. Oslo and 13 Norwegian communes, 1968-71.
Live offspring of both sexes.

Figure 1: Oslo

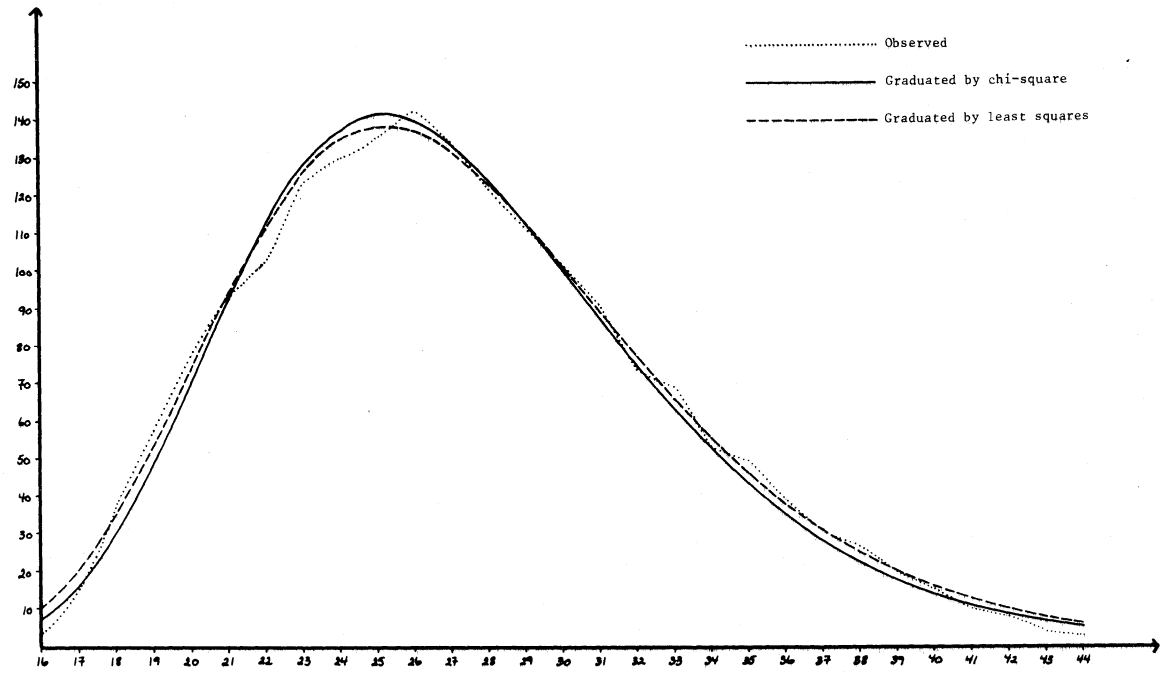

Figure 2: Aggregate of 13 communes on the Norwegian West Coast.

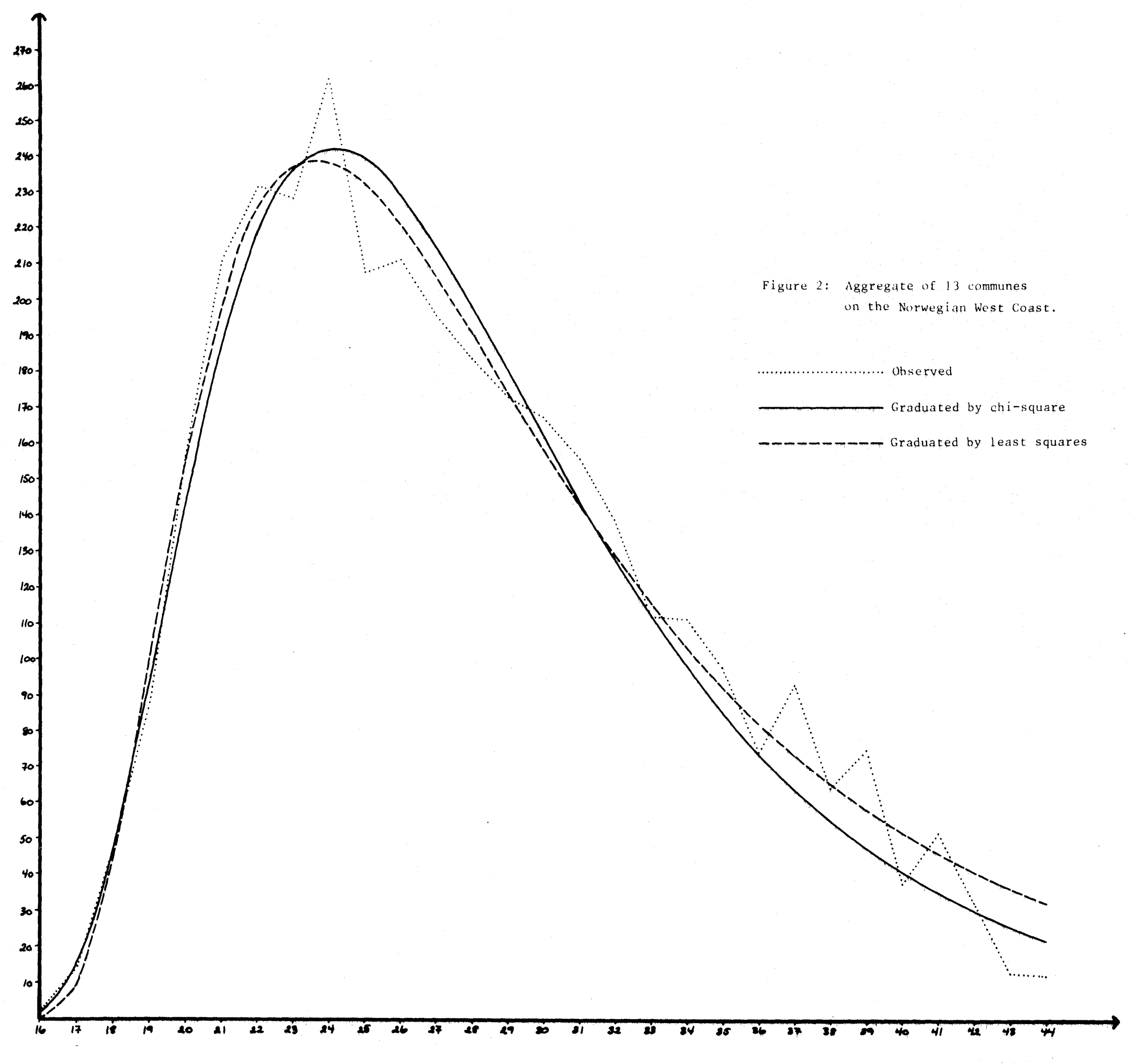

TABLE 1. ESTIMATES OF BASIC GRADUATION PARAMETERS

Estimand	Min. statistic	Oslo, 1968-71		13 communes on the West Coast of Norway, 1968-71	
		By min.-χ^2	By least squares	By min.-χ^2	By least squares
Θ_1 = R		1.88	1.93	3.50	3.72
Θ_2 = mode		25.23	25.28	24.12	23.62
Θ_3 = mean		27.28	27.40	28.44	29.70
Θ_4 = variance		32.46	35.53	54.00	83.04
	χ^2	186.1		41.9	
	LSQ		$2.7 \cdot 10^{-4}$		$42.8 \cdot 10^{-4}$

Graduation of fertility curves by means of the Hadwiger function.

Single-year age intervals for ages 16-44. Rates per 1 female, counting live offspring of both sexes.

TABLE 2. ESTIMATED ASYMPTOTIC VARIANCES OF ESTIMATORS OF BASIC GRADUATION PARAMETERS[1)]

Parameter		Oslo, 1968-71		13 Norwegian communes, 1968-71	
		Min.-χ^2 graduation[2)] (1)	Least squares graduation (NB: In per cent of (1).) (2)	Min.-χ^2 graduation[2)] (3)	Least squares graduation (NB: In per cent of (3).) (4)
Θ_1 = R		142	117	6 283	122
Θ_2 = mode		3 609	132	44 281	166
Θ_3 = mean		1 720	186	50 255	182
Θ_4 = variance		201 572	216	17 999 297	201
	ΣL_x	369 178.0		18 900.5	

1) Corresponding to min.-χ^2 estimates of Table 1.

2) Multiplied by 10^6.

TABLE 3. ESTIMATED ASYMPTOTIC VARIANCES AND COEFFICIENTS OF VARIATION[1)] OF ESTIMATORS OF FERTILITY RATES FOR SOME SELECTED AGES

Age as of Dec. 31 x	Variances				Coefficients of variation			
	$\frac{\hat{\sigma}_x^2}{N}=\frac{\hat{\lambda}_x}{L_x}$ (1)[2)]	Min.-χ^2 graduation (2)[2)]	Least squares graduation (NB: In per cent of (2).) (3)[3)]	(1) in per cent of (2) (4)[3)]	$\frac{\hat{\sigma}_x/\sqrt{N}}{\hat{\lambda}_x}$ (5)[4)]	Min.-χ^2 graduation (6)[4)]	Least squares graduation (NB: In per cent of (6).) (7)[3)]	(5) in per cent of (6) (8)[3)]
				a. Oslo, 1968-71				
16	0.360	0.186	224	194	16.01	5.87	150	273
18	3.144	0.454	175	692	4.63	2.20	132	210
20	4.836	0.795	112	608	2.83	1.26	106	224
24	6.424	1.053	112	610	1.95	0.75	106	261
28	9.775	1.198	115	816	2.56	0.88	107	292
35	5.572	0.325	152	1 713	4.77	1.33	123	358
40	1.425	0.117	225	1 215	7.88	2.50	150	315
44	0.202	0.040	238	501	18.26	4.15	154	440
				b. 13 Norwegian communes, 1968-71				
16	2.206	0.839	390	263	57.75	42.76	197	135
18	51.202	13.548	231	378	15.08	7.59	152	199
20	219.575	45.390	119	484	9.49	4.65	109	204
24	393.465	44.963	117	875	7.58	2.77	108	273
28	317.285	35.990	116	882	9.71	3.03	107	320
35	173.826	9.776	138	1 778	13.48	3.68	118	367
40	63.956	5.960	180	1 073	21.32	5.99	134	356
44	18.703	3.836	203	488	35.36	8.93	141	396

1) Corresponding to the min.-χ^2 estimates of Table 1.

2) Multiplied by 10^6, i.e., corresponding to rates per 1 000 females.

3) Calculated from figures with six effective digits.

4) Multiplied by 100.

ON THE OPTIMALITY OF MODIFIED MINIMUM CHI-SQUARE ANALYTIC GRADUATION

by Jan M. Hoem*)

ABSTRACT

It has been shown previously that the modified minimum chi-square method of analytic graduation of vital rates produces estimators with minimum variance asymptotically as the population size increases. In the present paper, it is shown that these estimators are similarly optimal according to a different criterion, viz., one where one sets up a quadratic form of deviations between the "true" underlying fertility curve and the graduated curve, and seeks to minimize the mean of the asymptotic distribution of the quadratic form.

1. INTRODUCTION

1 A. A number of methods useful for the analytic graduation of vital rates can be found in the literature. Most of these can be described as moment methods, or as weighted or unweighted least squares methods, including chi-square methods. Some previous research (Hoem, 1972) has shown that a modified minimum chi-square method cannot be outdone by any other method readily available in practice, in the sense that asymptotically as the population size increases the modified minimum chi-square estimators have minimal variances within the class of estimators considered. This holds for the estimators of the parameters of the fitting function as well as for the estimators of the function values (i.e., of the "true" fertility rates at all ages).

1 B. The author has been involved in a project where curves of age-specific fertility rates have been graduated analytically. (See, e.g., Hoem, 1974.) The diagram of such rates for a population, based on data for a calendar period, say, will typically look much like a left-skewed unimodal probability density, such as a gamma density or some of the beta densities (starting just below age 15), for instance, but with superposed fluctuations. In this connection, the author has had to give some consideration to arguments which appear to suggest that something must be wrong with a line of reasoning which results in bringing out a chi-square method as optimal. The modified minimum chi-square method gives more weight to deviations between the empirical and the fitted fertility rates at ages where fertility is low than it does to equally large absolute deviations at the ages of high fertility. This is said to be undesirable on intuitive grounds, as it is more important to get a good fit at the highly fertile ages than at the tail ends of the fertility curve. Such a fit, so the argument goes, necessitates the use of unweighted least squares estimates, or even weighted least squares estimates where the weights at the central fertile ages are larger than in the tails.

---

*) Discussions with Peter Anderson, Lucien Le Cam, and Herman Chernoff have been helpful. Erling Berge kindly checked the manuscript.

The present author believes that a different line of common-sense reasoning can throw some doubt on this type of argument and even support the modified minimum chi-square method on intuitive grounds. The point is that this method consists of minimizing the sum of squares of deviations when the latter are measured in terms of their estimated standard deviations rather than in "absolute" terms. At the less fertile ages, the standard deviations will tend to be smaller than at the highly fertile ages (given a reasonably even age distribution in the population), so that the empirical age-specific fertility rates correspondingly will be more accurate estimators for the "true" rates at the tails than at the central ages. Not to normalize by the standard deviations would then be to underestimate the importance of the less fertile ages, so it seems to me. This argument is not particular to the present application of chi-square methods, of course, nor is it new with me. Indeed, I have only picked it up on the statistical "grapevine".

1 C. More important to the present author, though, is the fact that I believe that I am able to reduce the credibility of the counter-argument on a technical basis: The formulation of the argument suggests that there is something the matter with the criterion according to which the modified minimum chi-square estimator is optimal. Perhaps one should not focus on minimizing the asymptotic variances of the estimators, but on some other criterion which measures the fit of the curve directly rather than possibly indirectly. Such a measure might take the form of a weighted sum of squares (or, more generally, a quadratic form) of deviations between the "true", underlying fertility curve which we are trying to estimate on the one hand, and the curve actually fitted to the raw fertility rates on the other hand.

It turns out that as long as the fitting criterion (or loss function) is of this type, no moment or (possibly weighted) least squares method can do better in the long run than the modified minimum chi-square method, in the sense that the loss function has an asymptotic distribution whose mean is minimized by the latter. On this criterion the modified minimum chi-square method fares at least as well as even its most obvious contender, viz., the procedure which consists of minimizing the weighted sum of squares of deviations with weights which are (estimators of) those of the fitting criterion. On this basis as well, therefore, the modified minimum chi-square procedure seems to be at least as good as any alternative that has been suggested so far.

Of course, this does not settle once and for all the question of selecting the graduation procedure. It is quite possible to suggest other criteria according to which modified minimum chi-square graduation may not be optimal, even asymptotically, and we have not investigated small-sample properties of any method. I do feel, however, that a result like the one mentioned above reinforces our trust in the usefulness of graduation by means of modified minimum chi-square.

1 D. The purpose of the present paper is to present the results mentioned above. In order to simplify our presentation, and because they do not generally give satisfactory empirical fits, moment methods will be left aside in most of the paper. For the sake of completeness, however, the theory is extended to such methods also in Subsection 3 D.

The problem arose in connection with a study of fertility rates, and our results are directly relevant to this situation, but they apply similarly to vital rates in general. Our account will be for the general case.

2. PREVIOUS RESULTS

2 A. The "raw" age-specific vital rates in a set of observations will have been calculated for a number of age groups, which for simplicity we shall take to be the single year age intervals $\alpha, \alpha+1, \ldots, \beta$. Let $\hat{\lambda} = \{\hat{\lambda}_\alpha, \ldots, \hat{\lambda}_\beta\}'$ denote the vector of "raw" rates, the prime signifying a transpose. These rates are to be graduated by means of some sequence of parametric functions

$$g(\theta) = \{g_\alpha(\theta), \ldots, g_\beta(\theta)\}',$$

that is, one is required to select some value $\hat{\theta}$ of the vector $\theta = (\theta_1, \ldots, \theta_r)'$ such that $g(\hat{\theta})$ fits as well as possible to the "raw" rates in $\hat{\lambda}$. We denote the "true" value of θ by θ^0, and the corresponding value of the underlying "true" vital rates is $\lambda^0 = g(\theta^0)$.

2 B. We shall suppose that the fitting will be done as follows. Let M be a square matrix of elements m_{xy} $(x, y = \alpha, \alpha+1, \ldots, \beta)$ which may be random variables. Let

$$Q_M(\theta) = N\{\hat{\lambda} - g(\theta)\}' M\{\hat{\lambda} - g(\theta)\},$$

where N denotes the number of individuals under observation, i.e., the number of persons who are potential contributors to the vital rates. Assume that there exists a value of θ, say $\hat{\theta}(M)$, which minimizes $Q_M(\theta)$. Then $g\{\hat{\theta}(M)\}$ represents one possible graduation of $\hat{\lambda}$.

We shall assume that as the population size N increases, M converges in probability to a positive definite matrix M_0, and $\hat{\lambda}$ is asymptotically multinormally distributed with mean λ^0 and some covariance matrix Σ_0/N. [In empirical studies, we usually take Σ_0 to be diagonal, but we will not make use of this here.] Let $\hat{\Sigma}_0$ be a consistent estimator of Σ_0 as $N \to \infty$. We shall call $\hat{\theta}(\hat{\Sigma}_0^{-1})$ the <u>modified minimum chi-square estimator</u> of θ.

2 C. The following results have been proved by Hoem (1972) under some natural assumptions which we shall not repeat here, but which we shall assume to hold.

Let

$$J(\theta) = \begin{bmatrix} \frac{\partial}{\partial\theta_1} g_\alpha(\theta), & \ldots, & \frac{\partial}{\partial\theta_r} g_\alpha(\theta) \\ \cdot & \cdots & \cdot \\ \frac{\partial}{\partial\theta_1} g_\beta(\theta), & \ldots, & \frac{\partial}{\partial\theta_r} g_\beta(\theta) \end{bmatrix},$$

let $J_0 = J(\theta^0)$,

(2.1) $$A = (J_0' M_0 J_0)^{-1} J_0' M_0 ,$$

and

(2.2) $$\Sigma = A \Sigma_0 A' .$$

Then $\hat{\theta}(M)$ is asymptotically multinormally distributed with mean θ^0 and covariance matrix Σ/N. Similarly, $g\{\hat{\theta}(M)\}$ is asymptotically multinormal with mean λ^0 and (singular) covariance matrix

$$J_0 \Sigma J_0'/N .$$

2 D. If, in particular, $\underset{\sim}{M}_0 = \underset{\sim}{\Sigma}_0^{-1}$, then $\underset{\sim}{\Sigma}$ is equal to

$$\underset{\sim}{\Sigma}_{0,0} = (\underset{\sim}{J}_0' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{J}_0)^{-1}.$$

For any $\underset{\sim}{M}_0$, $\underset{\sim}{\Sigma}-\underset{\sim}{\Sigma}_{0,0}$ will be positive semidefinite, and so will $\underset{\sim}{J}_0 \underset{\sim}{\Sigma} \underset{\sim}{J}_0' - \underset{\sim}{J}_0 \underset{\sim}{\Sigma}_{0,0} \underset{\sim}{J}_0'$.

[This is Theorem 4 in Hoem (1972). Part (iii) of the proof there rests on the fact that $\underset{\sim}{\Sigma}$ is of the form given in (2.2), where A is any matrix such that $\underset{\sim}{A} \underset{\sim}{J}_0 = \underset{\sim}{I}$. Note that the covariance matrix given in (7.11) there is also of the same form, so that the proof of Theorem 6 there and part (iii) of the proof of Theorem 4 are essentially two versions of the same proof.]

3. OPTIMALITY OF MODIFIED MINIMUM CHI-SQUARE GRADUATION ACCORDING TO A LEAST QUADRATIC FORM CRITERION

3 A. As we mentioned in Subsection 1 C already, we have had to consider an argument to the effect that the modified minimum chi-square method should be avoided because it gives too much weight to deviations at ages x where $\hat{\lambda}_x$ is small. The present anthor believes that the discussion in the Subsection mentioned at least weakens this argument. Nevertheless, let us accept that some people (perhaps we ourselves) would like to attach weights to the various age groups different from those implied by the choice of $\underset{\sim}{M} = \hat{\underset{\sim}{\Sigma}}_0^{-1}$. Let us assume that somebody specifies weights $w_\alpha, w_{\alpha+1}, \dots, w_\beta$, and that he or she claims that one wants to measure how well the graduated values $g_\alpha(\hat{\underset{\sim}{\theta}}), \dots, g_\beta(\hat{\underset{\sim}{\theta}})$ fit the true values $\lambda_\alpha^0, \dots, \lambda_\beta^0$ by calculating

$$Q^0(\hat{\underset{\sim}{\theta}}) = N\sum_x w_x\{\lambda_x^0 - g_x(\hat{\underset{\sim}{\theta}})\}^2.$$

Let us assume, furthermore, that it is claimed that $\hat{\underset{\sim}{\theta}}$ should be selected in a way such that in the long run, $Q^0(\hat{\underset{\sim}{\theta}})$ is as small as possible in the mean. Let us permit our "somebody" (who may be ourselves) to also specify weights w_{xy} to products $\{\lambda_x^0 - g_x(\hat{\underset{\sim}{\theta}})\}\{\lambda_y^0 - g_y(\hat{\underset{\sim}{\theta}})\}$ of discrepancies at different ages $(x \neq y)$, so that his or her measure of goodness of fit becomes

$$Q^0_{\underset{\sim}{W}}(\hat{\underset{\sim}{\theta}}) = N\sum_x \sum_y w_{xy}\{\lambda_x^0 - g_x(\hat{\underset{\sim}{\theta}})\}\{\lambda_y^0 - g_y(\hat{\underset{\sim}{\theta}})\} = N\{\underset{\sim}{\lambda}^0 - \underset{\sim}{g}(\hat{\underset{\sim}{\theta}})\}'\underset{\sim}{W}\{\underset{\sim}{\lambda}^0 - \underset{\sim}{g}(\hat{\underset{\sim}{\theta}})\},$$

where $\underset{\sim}{W}$ is the matrix of the w_{xy}, with $w_{xx} = w_x$. In statistical parlance, we take $Q^0_{\underset{\sim}{W}}(\hat{\underset{\sim}{\theta}})$ to be the loss function applicable, and we want to minimize its asymptotic expected value. We shall assume that $\underset{\sim}{W}$ is positive definite.

We shall permit $\underset{\sim}{W}$ to possibly depend on the parameters, but in that case, we shall assume that it can be estimated by some matrix $\hat{\underset{\sim}{W}}$ which converges in probability to $\underset{\sim}{W}$ as $N \to \infty$. (If $\underset{\sim}{W}$ is known beforehand, as it will be if one takes $\underset{\sim}{W} = \underset{\sim}{I}$, for instance, we define $\hat{\underset{\sim}{W}} = \underset{\sim}{W}$.)

3 B. Since the λ_x^0 are not known, of course, $\hat{\underset{\sim}{\theta}}$ cannot be found by simply minimizing $Q^0_{\underset{\sim}{W}}(\underset{\sim}{\theta})$, say. It seems plausible, however, that it would be a good idea to minimize

$$Q_{\hat{\underset{\sim}{W}}}(\underset{\sim}{\theta}) = N\{\hat{\underset{\sim}{\lambda}} - \underset{\sim}{g}(\underset{\sim}{\theta})\}'\hat{\underset{\sim}{W}}\{\hat{\underset{\sim}{\lambda}} - \underset{\sim}{g}(\underset{\sim}{\theta})\},$$

i.e., to use $\hat{\underset{\sim}{\theta}}(\hat{\underset{\sim}{W}})$ as the estimator for $\underset{\sim}{\theta}^0$. One might believe that this would minimize the asymptotic mean of $Q^0_{\underset{\sim}{W}}\{\hat{\underset{\sim}{\theta}}(\underset{\sim}{M})\}$, in the sense that as $N \to \infty$, the asymptotic mean of $Q^0_{\underset{\sim}{W}}\{\hat{\underset{\sim}{\theta}}(\hat{\underset{\sim}{W}})\}$ would be smaller than the asymptotic mean of $Q^0_{\underset{\sim}{W}}\{\hat{\underset{\sim}{\theta}}(\underset{\sim}{M})\}$ for any positive definite $\underset{\sim}{M}$, and then certainly for $\underset{\sim}{M} = \hat{\underset{\sim}{\Sigma}}_0^{-1}$. It may be surprising, then, at least at first sight, to learn that quite on the contrary, the asymptotic mean of $Q^0_{\underset{\sim}{W}}\{\hat{\underset{\sim}{\theta}}(\underset{\sim}{M})\}$ is minimized by choosing $\underset{\sim}{M} = \hat{\underset{\sim}{\Sigma}}_0^{-1}$.

This implies that instead of using $\hat{\underset{\sim}{\theta}}(\hat{\underset{\sim}{W}})$ or any other $\hat{\underset{\sim}{\theta}}(\underset{\sim}{M})$, one should use $\hat{\underset{\sim}{\theta}}(\hat{\underset{\sim}{\Sigma}}_0^{-1})$ (or some method with the same asymptotic properties) if one wishes to optimize one's graduation procedure according to this criterion. Thus, modified minimum chi-square graduation cannot be outdone by any other method of the given class by this criterion either.

3 C. The above result is contained in the following theorem.

Theorem: The asymptotic mean of $Q_{W}^{0}\{\hat{\Theta}(M)\}$ equals

$$d(\Sigma) = \mathrm{tr}(F\,\Sigma), \tag{3.1}$$

with $F = J_0'W\,J_0$ and Σ as in (2.2). Here

$$d(\Sigma_{0,0}) \leq d(\Sigma) \tag{3.2}$$

for all such Σ .

Proof: We note that the asymptotic distribution (as $N \to \infty$) of

$$N^{1/2}\{g(\hat{\Theta}(M)) - \lambda^0\}$$

is the same as that of

$$N^{1/2}J_0\{\hat{\Theta}(M) - \Theta^0\} .$$

Thus, the asymptotic distribution of $Q_{W}^{0}\{\hat{\Theta}(M)\}$ is the same as that of

$$N\{\hat{\Theta}(M) - \Theta^0\}'\; F\{\hat{\Theta}(M) - \Theta^0\} ,$$

with $F = (f_{ij}) = J_0'WJ_0$. Now, if X is a random vector of the same dimension as F, with mean 0 and covariance matrix $T = (\tau_{ij})$, then

$$E(X'FX) = E(\sum_i \sum_j f_{ij}X_iX_j) = \sum_i \sum_j f_{ij}\tau_{ij} = \mathrm{tr}(FT) .$$

Therefore, the asymptotic mean of $Q_{W}^{0}\{\hat{\Theta}(M)\}$ equals $d(\Sigma) = \mathrm{tr}(F\,\Sigma)$, with Σ given in (2.2). This proves (3.1).

Let $\mathcal{T}$ be the class of covariance matrices generated by varying M_0 in (2.1). Then $\Sigma - \Sigma_{0,0}$ is positive semidefinite for all $\Sigma \in \mathcal{T}$. Therefore, the left inequality of (3.2) follows from the following lemma. $\square$

Lemma: Let $\mathcal{T}$ be a class of positive definite $n{\times}n$ matrices, and assume that there exists a positive definite $n{\times}n$ matrix T_0 such that $T - T_0$ is positive semidefinite for all $T \in \mathcal{T}$. Let F be any positive semidefinite $n{\times}n$ matrix, and let $d(T) = \mathrm{tr}(FT)$. Then

$$d(T_0) \leq d(T) \quad \text{for all } T \in \mathcal{T} ,$$

no matter how F is chosen.

Proof of Lemma: Since F is positive semidefinite, there exists an $n{\times}n$ matrix G such that $F = G'G$. Then $d(T) = \mathrm{tr}(G'GT) = \mathrm{tr}(GTG')$. Since $T - T_0$ is positive semidefinite, so is $G(T - T_0)G'$. The trace of a positive semidefinite matrix equals the sum of its latent roots and is therefore non-negative. Thus $d(T) - d(T_0) = \mathrm{tr}\{G(T-T_0)G'\} \geq 0$. $\square$

3 D. In the account above, we have only discussed analytic graduation methods which consist of minimizing some quadratic form. A different class of methods is generated by using some kind of moment method estimator for Θ . For an account of this class, we refer to Section 7 in Hoem (1972). In our practical work, we have not found the empirical fits provided by moment methods to fertility curves generally satisfactory and have therefore left them aside so far. By extending $\mathcal{T}$ in the proof of Theorem 1 to contain also the covariance matrices Σ of the form given for moment methods

in (7.11) in Hoem (1972), and by using

$$\underset{\sim}{A} = (\underset{\sim}{M}\underset{\sim}{J}_0)^{-1}\underset{\sim}{M}$$

instead of (2.1), we can easily extend Theorem 1 to cover this class as well, however.

Even though graduation by the method of moments will frequently be unsatisfactory, such estimators do have some practical interest. Graduation through minimization of a quadratic form or through likelihood maximization typically must be done by means of an iterative numerical procedure. Moment estimators can then provide starting values for the iterations.

REFERENCES

[1] Hoem, Jan M. (1972): "On the statistical theory of analytic graduation." Proc. Sixth Berkeley Symp. Math. Statist. Prob., $\underset{\sim}{1}$: 569-600.

[2] Hoem, Jan M. and Erling Berge (1974): "Theoretical and empirical results on the analytic graduation of fertility curves." Pages 2 to 10 in this Working Paper.