A NOTE ON THE ANNUAL VARIABILITY OF THE CRUDE BIRTH RATE

William Seltzer and Rita S. Fand
Demographic Division, The Population Council

I. Introduction

The purpose of this paper is to investigate the extent of misinformation generated by estimating the level of the crude birth rate using data from a single year or estimating changes in this rate on the basis of data from two successive years. Briefly what we have done is assembled a few time series on the crude birth rate, computed various statistical measures that summarize aspects of the behavior of these series in time, estimated the level of random disturbance in these series using a somewhat general procedure (the variate difference technique), examined how stable these estimates of residual variability appear to be, and pointed out the implications of our results for those trying to measure or interpret short-run fertility changes.

II. Basic Data

The basic data consist of time series of annual crude birth rates in seven countries. We have used crude birth rates in this study rather than some other variable that is more closely related to the dynamics of fertility behavior for three reasons: (1) the reliance that many data users place in the crude birth rate as an important indicator of fertility behavior; (2) the ease with which time series of crude birth rates can be assembled compared with that for other fertility variables; and (3) our interest in short-run changes, where age distribution and cohort effects are minimal.

Summary statistics for the seven time series are presented in Table 1. The countries involved are: the United States (for the 61 year period 1909-69), Japan (for the 93 year period 1875-1967), England and Wales (for the 130 year period 1838-1967), Algeria (for the 80 year period 1891-1970), Sweden (for the 202 year period 1766-1967), Finland (for the same 202 year period), and Malta (for the 70 year period 1900-1969).

Together this body of data covers 834 country years of experience reflecting considerable diversity in population size and some diversity in cultural background and demographic history. However, it should be noted that the selection of countries included was arbitrary and the nature and the quality of the data used varies widely. For example, both the Algerian and the United States series are known to contain adjustments for underregistration. Despite the necessary qualifications arising from the arbitrary choice of countries and the problems of data quality, we think our results will be of interest to those trying to interpret short-run changes in fertility.

In this connection, let us point out two interesting facets of these series shown in Table 1. First, the median year to year percentage change in the crude birth rate varies from 1.0 to 5.1 among these seven countries -- in our view a rather narrow range -- and one that indicates that many annual changes in the crude birth rate will be too small to be reliably detected except by a well calibrated civil registration system. Second, note that the number of years with no change in the crude birth rate from the preceding year ranges from about 1 to 5 percent of each series. However, this is an arbitrary result depending upon the number of significant digits used in presenting the crude birth rate. In the case of Sweden if one expresses the crude birth rate as an integer rate per 1,000 (for example, 25 per 1,000) rather than using one more significant digit (such as a crude birth rate of 25.4) the number of no change years increases from 7 to 72 out of the 201 year total. The result is obvious to statisticians but has not been adequately communicated to data users.

To supplement the annual crude birth rate series from these seven countries, we also make use of ten other time series of annual crude birth rates (nine U.S. states for the 1915-1968 period, and New York City for the period 1898-1953) as well as various series of live births for England and Wales, Sweden, and New York City.

III. The Variate Difference Method

The variate difference method is a technique that permits us to estimate the level of noise of disturbance in a time series on the assumption that each term in the observed series is the sum of two unobservable components: (1) a polynomial of degree n (or less), and (2) a random disturbance element. That is, we assume in the present case that

$$b_y = f(y) + e_y \qquad [1]$$

where b_y = the (observable realized crude birth rate in some population in year y,

$f(y)$ = a polynomial in time of degree n, corresponding to the underlying trend of the crude birth rate, and

e_y = a random disturbance with $E(e) = 0$ and a constant variance, σ_e^2.

Given this rather general model, it is possible to estimate σ_e^2, the variance of the disturbance terms of the observed series.

Before proceeding further two points need to be made. The model is quite general. No specific polynomial is assumed nor does one even make a specific assumption about the degree of the polynomial involved. However, one must not mistake this generality with universality. Many series do not correspond to such a polynomial; for example, trigonometric or transcendental functions, functions that include independent variables in addition to time, and functions of a degree higher than n (although under a variety of circumstances each of these functions can be approximated by such a polynomial). Furthermore, each disturbance term is assumed to have no direct effect on any other disturbance term nor is the variance of the disturbance component permitted to vary over time. Because of these limitations the variate difference method has frequently led to unsatisfactory results [Kendall and Stuart, 1966; Anderson, 1971]. Fortunately, in the present situation it appears to lead to relatively stable estimates.

Making use of the fact that both the polynomial and disturbance elements of each term behave in opposite ways as successively higher order differences are taken, it is possible [Anderson, 1971 and Kendall and Stuart, 1966] to show that for a series of t terms that

$$\left(\frac{1}{t-n} \sum_{y=1}^{t-n} (\Delta^n b_y)^2 \right) \Big/ \begin{bmatrix} 2n \\ n \end{bmatrix} \qquad [2]$$

corresponds to an estimate of σ_e^2 using the differences of order n.

In practice, one avoids assuming knowledge one does not have by estimating the residual variance initially on the basis of first order differences, then using the second order differences, then the third order differences, and so on, stopping when the polynomial component has been suppressed and the variance estimate stabilizes. That is, one forms an estimate of σ_e^2 using equation 2 with n=1, then another estimate with n=2, and so on. If the model holds, the successive estimates of σ_e^2 should decline as long as n is less than or equal to the degree of the trend polynomial f(y); thereafter, when the order of differencing exceeds the degree of the trend polynomial the variance estimate should stabilize.

IV. Results

The results of applying the variate difference technique to the crude birth rate series of these seven countries are presented in Tables 2 and 3. Since few real time series can be expected to correspond exactly to the model given by equation 1, the estimates of the variance and coefficient of variation 1/ of the disturbance component of the series for each country do not stabilize as one takes higher order differences -- rather they tend slowly to drift. However, the median values for these seven series do stabilize after about the fourth or fifth difference, indicating a standard deviation of the residuals of about 3/4 of a point in the crude birth rate or a coefficient of variation of the residuals of somewhat over 2.5 percent.

The strategy of the balance of this section is to demonstrate the relative stability of these estimates of residual variability in the face of efforts to identify factors with which the residuals may be associated. The following factors are examined explicitly: first, the size of the population; second, the length of the time series; third, the time period covered by the series; fourth, the use of a time series of crude birth rates compared with one of live birth aggregates; and finally, the use of a time series of annual terms compared with one of monthly terms.

To summarize quickly the results of these analyses, we can find no clear pattern in the level of residual variability for any of the factors examined, except possibly when a monthly series is used rather than an annual one. In this case the pattern is clear -- the monthly series has a coefficient of variation for the disturbance term that is about 1 1/2 to 3 1/2 times as large as that for the corresponding annual series. This differential is maintained even if one attempts to control for series length and seasonality by looking at the year to year changes in the series for individual months. Our only qualification to this finding is that it is based on data for only one time series -- New York City -- and we find it is somewhat difficult to assert that this City is a typical place.

Some discussion of the factors that did not appear to be associated with the estimated level of residual variability is in order. It might be conjectured that the deviations from the polynomial trend are the result of some binomial process. If this were so, then one would expect the variance and the coefficient of variation of the residuals to be larger for countries with a small population than for those with a large one. Although far from statistically significant, the shadow of such a pattern appears to be lurking in Table 2. One finds, for example, based on estimates of residual variability using differences of order five:

Midrange Population (millions)	Countries	Median CV
Under 4	Malta, Finland	3.7
4 - 24.9	Sweden, Algeria	3.2
25 - 63.9	England and Wales	2.4
64 and over	U.S., Japan	2.4

Furthermore, the correlation coefficient between population size and the estimated variance of the residuals (again based on Δ^5) is -0.49.

In order to examine whether this pattern would emerge more clearly if other possible sources of residual variation were controlled, annual crude birth rate series for nine U.S. states covering a standard 54 year period were studied. Estimates of residual variability for these series fail to confirm that the deviations from the assumed polynomial trend are associated with population size in any simple way. 2/ This finding replicates the conclusion of Hotelling and Hotelling [1931] based on a live birth time series that the variability of the residuals is substantially greater than can be accounted for by a binomial process.

Tables 2 and 3 also suggest that the longer series for the seven countries may be more variable than the shorter ones -- with respect to both the underlying birth rates (the b_y's in equation 1) and the disturbance terms (the e_y's). However, if one compares results based on standard 54 year time periods for these same seven countries with those based on the original series, a confused and partially contradictory picture emerges:

Length of Series	CV of CBR Median	Est. residual CV (using Δ^2) Median
Varied	19.6	2.7
54 years	15.3	3.6

Although this analysis standardizes for series length and nearly so for time period covered, it includes nations at various stages of demographic development. For example, fertility in Algeria remains at high, pre-industrial levels, it has been at near-replacement levels in Sweden for some years, and it has followed a generally downwards, but widely varying, path in the United States over the past few decades. In order to see if there is some association between the level of variability and where in the demographic transition a population seems to be, data for two contrasting 54 year time periods was examined for the four longest time series (that is, Sweden Finland, England and Wales, and Japan). 3/

In both absolute and relative terms the annual crude birth rates for these four countries displayed more variability in the most recent 54 year period than in the earliest 54 year period observed in each series. This is true whether one includes Japan or excludes it from consideration because of the recency of its major fertility decline. However, the pattern is less clear for estimates of the variability of the residuals. The median estimated coefficient of variation for the most recent period is somewhat greater than for the early period (for example, 4 versus 3 percent using Δ^5). On the other hand, the estimated coefficient of variation of the residuals for the Swedish series -- one of the two longest -- was higher in the earlier period than the later one.

One may justifiably speculate about the extent to which these findings are an artifact of particular features of the demographic and social history of the populations studied. Our response can only be cautionary: seven countries, nine states, and one city is an inadequate representation of worldwide diversity and our sample in time is limited to two centuries at most. Nevertheless, it seems reasonable to us to conclude that these estimates of residual variability, because of their relative stability, can be accepted as provisional bounds to our certitude pending more extensive research or more refined analysis.

1/ Throughout this paper the estimated coefficient of variation of the residual terms is calculated by dividing the square root of the estimated variance of the residual terms by the mean of the appropriate crude birth rate or live birth time series.

2/ The correlation coefficient between population size and the estimated variance of the residuals (using Δ^5) for the nine states is -0.26; the corresponding r for the seven countries and nine states combined is -0.12. If the square root of population size is used in the correlations in place of population size, the values of r become -0.48, -0.17, and -0.08, respectively, for the seven countries, the nine states, and the two combined.

3/ Because the Japanese time series covers only 93 years, there is a 15 year overlap between the two time periods used for this country.

A. Text References

Anderson, T.W., The Statistical Analysis of Time Series. New York: John Wiley and Sons, 1971.

Hotelling, H. and F. Hotelling, "Causes of Birth Rate Fluctuations." Journal of the American Statistical Association, 26, no. 174: 135-149 (June 1931).

Kendall, M.G. and A. Stuart, The Advanced Theory of Statistics, Vol. 3, London: Griffin, 1966.

B. Data Sources

Emerson, H. and H. Hughes, Populations, Births, Notifiable Diseases, and Deaths, Assembled for New York City, N.Y.: 1866-1938 from Official Records. New York: Delamar Institute of Public Health, College of Physicians and Surgeons, Columbia University, 1941.

__________, Populations, Births, Notifiable Diseases, and Deaths Assembled for New York City, N.Y.: Supplement, 1936-1953. New York: Delamar Institute of Public Health, College of Physicians and Surgeons, Columbia University, 1955.

Mitchell, B.R. and P. Deane, Abstract of British Historical Statistics. Cambridge: University Press, 1962.

Mitchell, B.R. and H.G. Jones, Second Abstract of British Historical Statistics. Cambridge: University Press, 1971.

National Center for Health Statistics, Vital Statistics of the United States, Volume I--Natality. Rockville, Maryland: National Center for Health Statistics (formerly published by the U.S. Public Health Service; issued yearly), 1937 and later years.

__________, Vital Statistics Rates in the United States: 1940-1960. Washington, D.C.: National Center for Health Statistics, 1968.

Seers, D., "A Fertility Survey in the Maltese Islands." Population Studies, 10 (1957): 211-27, 1957.

Statistika Centralbyran, Historical Statistics of Sweden, Part I: Population 1720-1967. Stockholm: Statistika Centralbyran (second edition), 1969.

Strömmer, A., Väestöllinen Muuntuminen Suomessa (Trans: The Demographic Transition in Finland), Series A:13. Tornio: The Population Research Institute, 1969.

Tabutin, D. and J. Vallin, L'état Civil en Algerie." Paper prepared for the Colloque de Démographie Africaine, Rabat, October 1972.

Taeuber, I., The Population of Japan. Princeton: Princeton Press, 1958.

U.N. Economic and Social Council, Demographic Yearbook. New York: United Nations (issued yearly), 1958 and later years.

U.S. Bureau of the Census, Birth Statistics for the Registration Area of the United States: Annual Report. Washington, D.C.: U.S. Bureau of the Census (published yearly), 1915-30.

__________, Birth, Stillbirth, and Infant Mortality Statistics. Washington, D.C.: U.S. Bureau of the Census, 1931-6.

Table 1 Summary Information about Annual Crude Birth Rate Series for Seven Specified Countries: Various Time Periods

Item	U.S.A.	Japan	England/ Wales	Algeria	Sweden	Finland	Malta
Length of series (years)	61	93	130	80	202	202	70
First year in series	1909	1875	1838	1891	1766	1766	1900
Last year in series	1969	1967	1967	1970	1967	1967	1969
Summary Values:							
Mean	23.63	28.34	25.90	40.95	27.28	32.45	31.39
Standard Deviation	3.75	5.48	7.86	5.13	6.98	7.20	6.15
Coeff. of Variation	.1587	.1935	.3034	.1253	.2557	.2219	.1960
Highest value	30.1	36.1	36.3	52.1	37.0	43.8	41.0
Lowest value	17.5	13.8	13.9	29.3	13.7	16.6	15.8
First value	30.0	25.3	30.3	39.8	33.8	41.5	39.0
Last value	17.7	19.4	17.2	48.2	15.4	16.6	15.8
Measures of Change:1/							
Increasing years							
Number	21	37	51	46	80	87	19
Percent of all years	35.0	40.2	39.5	58.2	39.8	43.3	27.5
Decreasing years							
Number	38	51	72	32	114	111	49
Percent of all years	63.3	55.4	55.8	40.5	56.7	55.2	71.0
No change years							
Number	1	4	6	1	7	3	1
Percent of all years	1.7	4.3	4.7	1.3	3.5	1.5	1.5
Total number of runs	25	47	62	43	99	120	23
Mean length of run (years)							
All runs	2.40	1.96	2.08	1.84	2.03	1.68	3.00
Increasing runs	1.62	1.76	1.89	2.19	1.74	1.53	1.90
Decreasing runs	3.45	2.32	2.40	1.52	2.48	1.85	4.08
Runs with no change	1.00	1.00	1.20	1.00	1.00	1.00	1.00
Mean square of successive differences	0.9948	3.7849	1.1573	8.3390	2.4403	6.0276	3.5368
Median year-to-year percentage change	2.4	2.7	1.9	5.1	2.8	3.5	2.8

1/ All measures of change are based on the n-1 series of the first differences for each country.

Table 2 Variance of Annual Crude Birth Rate (CBR) Series and Estimated Residual Variance, by Order of Difference Used, for Seven Specified Countries: Various Time Periods

Item	U.S.A.	Japan	England/ Wales	Algeria	Sweden	Finland	Malta	Median	Range
Variance of annual CBR series	14.0541	30.0682	61.7429	26.3437	48.6593	51.8714	37.8268	37.8	61.7-14.1
Variance of residual terms 1/									
Estimated from Δ^1	0.4974	1.8924	0.5786	4.1695	1.2201	3.0138	1.7684	1.8	4.2-0.5
Estimated from Δ^2	0.2612	1.2882	0.4145	3.2187	0.8378	2.7780	0.8944	0.9	3.2-0.3
Estimated from Δ^3	0.1977	1.0334	0.3837	2.7376	0.6778	2.7676	0.6951	0.7	2.8-0.2
Estimated from Δ^4	0.1631	0.8848	0.3775	2.4135	0.5961	2.7733	0.6036	0.6	2.8-0.2
Estimated from Δ^5	0.1430	0.8008	0.3791	2.1661	0.5530	2.7757	0.5453	0.6	2.8-0.1
Estimated from Δ^6	0.1308	0.7552	0.3835	1.9799	0.5299	2.7728	0.5023	0.5	2.8-0.1
Estimated from Δ^7	0.1229	0.7308	0.3891	1.8417	0.5172	2.7656	0.4725	0.5	2.8-0.1
Estimated from Δ^8	0.1176	0.7175	0.3952	1.7308	0.5101	2.7558	0.4548	0.5	2.8-0.1
Estimated from Δ^9	0.1139	0.7100	0.4015	1.6337	0.5062	2.7449	0.4444	0.5	2.7-0.1
Estimated from Δ^{10}	0.1112	0.7050	0.4080	1.5524	0.5044	2.7338	0.4337	0.5	2.7-0.1

1/ Calculated using equation 2.

Table 3 Coefficient of Variation (CV) of Annual Crude Birth Rate (CBR) Series and Estimated Residual CV, by Order of Difference Used, for Seven Specified Countries: Various Time Periods

Item	U.S.A.	Japan	England/ Wales	Algeria	Sweden	Finland	Malta	Median (percent)	Range (percent)
CV of annual CBR series	0.1587	0.1935	0.3034	0.1253	0.2557	0.2219	0.1960	19.6	30.3-12.5
CV of residual terms									
Estimated from Δ^1	0.0299	0.0485	0.0294	0.0499	0.0405	0.0535	0.0424	4.2	5.4-2.9
Estimated from Δ^2	0.0216	0.0400	0.0249	0.0438	0.0336	0.0514	0.0301	3.4	5.1-2.2
Estimated from Δ^3	0.0188	0.0359	0.0239	0.0404	0.0302	0.0513	0.0266	3.0	5.1-1.9
Estimated from Δ^4	0.0171	0.0332	0.0237	0.0379	0.0283	0.0513	0.0248	2.8	5.1-1.7
Estimated from Δ^5	0.0160	0.0316	0.0238	0.0359	0.0273	0.0513	0.0235	2.7	5.1-1.6
Estimated from Δ^6	0.0153	0.0307	0.0239	0.0344	0.0267	0.0513	0.0226	2.7	5.1-1.5
Estimated from Δ^7	0.0148	0.0302	0.0241	0.0331	0.0264	0.0512	0.0219	2.6	5.1-1.5
Estimated from Δ^8	0.0145	0.0299	0.0243	0.0321	0.0262	0.0512	0.0215	2.6	5.1-1.5
Estimated from Δ^9	0.0143	0.0297	0.0245	0.0312	0.0261	0.0510	0.0212	2.6	5.1-1.4
Estimated from Δ^{10}	0.0141	0.0296	0.0247	0.0304	0.0260	0.0509	0.0210	2.6	5.1-1.4

Source: Variance estimates in Table 2 and CBR means in Table 1.