

WORKING PAPER

CORRELATIONS BETWEEN FREQUENCIES OF KIN

*Thomas Pullum
Douglas Wolf*

August 1989
WP-89-053

**CORRELATIONS BETWEEN
FREQUENCIES OF KIN**

*Thomas Pullum
Douglas Wolf*

August 1989
WP-89-053

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

Foreword

The small community that concerns itself with the mathematics of kinship is the intellectual progeny of A.J. Lotka, who as long ago as 1930 asked himself about the relation between mortality and orphanhood. In a famous article (Lotka, 1931) in *Metron* he developed the formula for the probability that a person of given age has a living mother or a living father. Since then formulas for kin up to first cousins have been published by LeBras (1973), Goodman et al. (1974), and others. Hammel et al. (1989) show the uses of microsimulation. Among applications of this literature are those by Louis Henry and William Brass who used the relations to infer rates of birth and child mortality, and by Noreen Goldman (1978) who inferred the rate of increase of a population from a survey of living sisters.

The phenomenon of orphanhood has virtually disappeared in developed societies, and interest has shifted to the other end of life: how many of the old have children to provide financial and moral support? Daughters of mothers that have many children themselves have many, as Pullum and Wolf show in the following pages. More complex is the question how far longevity runs in families; that is explored by Vaupel and Yashin, who find that robustness is inherited indeed, but this does not translate into any high correlation of ages at death between kin.

The present paper fits well in this research tradition. It takes advantage of the mathematical relations that descend from Lotka, and also of micro-simulation useful where the conditions are too detailed to be handled mathematically. Its central question is the correlation between kin frequencies and it caps the theory with actual data from surveys.

The paper shows how three different methods for finding the relations among kin—analytic, simulation, and survey—may be used to check and support one another.

Nathan Keyfitz
Leader, Population Program

- Goldman, Noreen (1978) Estimating the intrinsic rate of increase from the average numbers of younger and older sisters. *Demography* 15:499–508.
- Hammel, E.A., Carl Mason, Kenneth Wachter, Feng Wang, and Haiou Yang (1989) Microsimulation as a Tool in Exploring Social and Demographic Interrelationships. Working Paper #27. Berkeley: University of California.
- Le Bras, Hervé (1973) Parents, grandparents, diaeresis bisaieux. *Population* 28:9–37 (English translation Kenneth Wachter, Ed., *Quantative Studies in Historical Social Structure*, Academic Press 1977).
- Lotka, Alfred J. (1931) Orphanhood in relation to demographic factors. *Metron* (Rome) 9:37–109.

Abstract

In recent years there has developed considerable interest in the demography of kin, including the development of formal and microsimulation models of the structure and dynamics of kin networks. One issue arising in the study of kin is the reconciliation of the assumption of uncorrelated fertility within and across generations, with the finding in several actual data sets of positive correlations between frequencies of various categories of kin. This paper brings together formal models, microsimulation models, and actual data in an effort to answer questions about such correlations of kin. We first show that even with uncorrelated fertility of mothers and daughters, correlations between certain kin frequencies will arise, and provide simple expressions for the correlation which depend only on the net reproduction rate. We then consider alternative mechanisms by which fertility might be transmitted from mothers to their daughters, and illustrate the correlations achieved by various such mechanisms. Finally, we present data from several surveys which reveal substantial correlations between various kin counts. These correlations are consistent with intergenerational transmission of fertility patterns.

Acknowledgement

We have profited from several discussions with Wolfgang Lutz on the issues raised here. We are also grateful to Dr. Johannes Huinck of the Max-Planck-Institut für Bildungsforschung, Berlin, for supplying unpublished data from the 1981-1982 German Life History Survey, and to Dr. Shoshana Neuman of Bar-Ilan University for supplying unpublished data from the 1974 Israeli Labour Mobility Survey. Part of this analysis is based on Statistics Canada's microdata tape Health and Social Support, which contains data collected in the 1985 General Social Survey. All computations on these microdata were prepared at IIASA and the responsibility for the use and interpretation of these data is entirely that of the authors.

CORRELATIONS BETWEEN FREQUENCIES OF KIN

*Thomas Pullum**, *Douglas Wolf***

INTRODUCTION

The past few years have seen the development of demographic methods and data which pertain to the numbers of grand-children, siblings, and other types of kin in addition to children, who form the category of greatest interest. Some of the formal models developed have concerned the numbers of kin who would be implied in a stable population, in which birth and death schedules are fixed (Le Bras 1973; Goodman, Keyfitz, and Pullum 1974, 1975). In a stable population it is possible to estimate analytically the mean numbers of kin of specific types who would ever have been born, or would still be living, according to the age of a reference individual known as ego. Other procedures allow estimation of the distribution of the numbers who will ultimately be born into specific kin categories, using the frequency distribution of daughters (Pullum 1982, Waugh 1981). Computer simulation methods also allow the estimation of frequencies of kin with greater flexibility in assumptions than is possible with formal or mathematical approaches (Bongaarts 1987). Finally, data have been collected in recent years which permit the validation of the other procedures—in addition, of course, to other valuable uses. A variety of approaches to these themes are described in Bongaarts, Burch, and Wachter (1987), including the chapter by Pullum (1987).

Such work, including the present paper, has two main objectives. The first is to articulate the population dynamics of kinship—to specify the implications of different patterns of fertility and mortality for the network of kin. The second objective is less theoretical, and concerns the actual availability of kin for social contact and support, particularly in the older ages and particularly in contexts where reproduction is near replacement levels (Wolf 1988). Formal modelling and simulation are important for this purpose, as well as for the first one, because of a general inadequacy of empirical data and because of our interest in projections into the future.

*Population Research Center, The University of Texas, Austin, Texas 78712, USA

**The Urban Institute, 2100 M Street, NW, Washington, DC 20037, USA

A closely related and somewhat narrower issue concerns intergenerational transmission of fertility patterns, an object of research for many years. A recent contribution is the paper by Anderton et al. (1987), who also summarize the previous literature. But if the fertility levels of mothers and daughters are positively correlated, it should be true by implication that the fertility of more distant relatives is also correlated, leading us to consider more general patterns of kin correlations.

This paper will bring together formal models, computer simulation, and actual data in an effort to answer questions about correlations between the frequencies of kin of different types. If such (positive) correlations exist, then some individuals will be located in relatively sparse networks of kin, and others in relatively dense networks, with a greater range in the overall density than if there were no such correlations. These are the kinds of questions which interest us:

- Even with the restriction that all women are subject to the same birth probabilities and are independent of one another, will some categories be correlated, so that women with more kin of one type can be expected to have more kin of some other type? Which categories are related and to what degree?
- What are some plausible descriptions of inherited heterogeneity in birth probabilities (whether transmitted genetically or behaviorally)? What will be the effect of heterogeneity upon the correlations between kin categories?

We will deal solely with one-sex models. Thus, ego will be female; children will be daughters, granddaughters will be the daughters of daughters, siblings will be sisters, and so on. This simplification will not be a serious limitation. If both sexes were included, virtually none of the results would change; the number of kin in each category would be multiplied by a constant (a function of only the sex ratio and the kin category) and the magnitudes of correlations would not be affected.

The mathematical results to be presented here will be new in this context, but would be considered elementary within the general theory of branching processes.

SOME RELEVANT ISSUES

It will be helpful to describe some possible ambiguities and to propose how to resolve them. The most important issue concerns age and survivorship. Suppose, for specificity, that we asked the following question, in relation to a formal model or to a simulation or to a body of data: "Is there a correlation between the number of sisters and the number of daughters?" How is such a question to be understood?

There should be no doubt that the two counts—the number of sisters and the number of daughters—pertain to the clusters of individuals (females) with whom the sisters and daughters bear the specified relationship. In particular, this correlation should not be confused with one between daughters and granddaughters. The units of analysis are individual females whose various kin have been tallied.

It is desirable to interpret the question as referring to complete counts of sisters and daughters. This is problematic because if the two variables are defined in terms of numbers ever born, or still alive, they will change with the woman's (ego's) age. The number of kin ever born will tend to increase and the number still alive will tend to increase and then decrease. We prefer an interpretation which is somehow free of the fact that when ego is young, all of her sisters may be born but none of her daughters; and when she is old, all of her daughters may be alive but not all of her sisters.

Ego's age has another complicating property: it affects the probability that she herself is still alive. If a woman dies early, before all her daughters are born or even before all her sisters are born, should she contribute to the correlation?

These issues are illustrated in Table 1, which shows several alternative concepts of kin correlations, by age, in a simulated one-sex population of approximately 100,000 women. The population was generated using computer simulation techniques described in Wolf (1988). The assumptions underlying the simulation are in all important respects the same as those made in the analytic model developed by Goodman, Keyfitz and Pullum (1974); however the simulation approach permits the full frequency distribution of kin, by age, to be determined. In the simulated population, all births and deaths are independent, and are determined by birth and death rates which are fixed over time. The net reproduction rate in the simulated population is exactly one.¹

The table shows the lifetime profile of the correlation between number of sisters (Sis) and numbers of daughters (Dau), and between number of daughters and number of granddaughters (GDau). The effects of mortality are illustrated by including figures for both living, and ever-born, kin. Considering sisters and daughters, we see a sequence of correlations which are, in nearly every case, only trivially different from zero. Moreover there is no apparent pattern to the correlation coefficients.

In the case of daughters and granddaughters, the correlation is zero until ages 35–39, when granddaughters first begin to appear, and rises over the next seven age groups to its maximum of approximately 0.7. For living daughters and granddaughters, the correla-

¹The death rates used were those for U.S. women, 1985; the birth rates were for U.S. women in 1985, scaled down proportionally until the implied net reproduction rate became 1 (NCHS 1987).

Table 1. Mean numbers of and correlations between numbers of, living and ever-born kin, by age of ego: simulated stationary population.

| Age Group | Sample Size | Living | | | Ever-Born | | | Living | | | Ever-Born | | |
|-----------|-------------|--------|------|--------|-----------|------|--------|--------|------|-------|-----------|------|-------|
| | | Sis | Dau | Corr | Sis | Dau | Corr | Dau | GDau | Corr | Dau | GDau | Corr |
| 0-4 | 6256 | 0.63 | 0.00 | 0.000 | 0.63 | 0.00 | 0.000 | 0.00 | 0.00 | 0.000 | 0.00 | 0.00 | 0.000 |
| 5-9 | 6388 | 0.85 | 0.00 | 0.000 | 0.86 | 0.00 | 0.000 | 0.00 | 0.00 | 0.000 | 0.00 | 0.00 | 0.000 |
| 10-14 | 6325 | 0.94 | 0.00 | 0.000 | 0.95 | 0.00 | 0.000 | 0.00 | 0.00 | 0.000 | 0.00 | 0.00 | 0.000 |
| 15-19 | 6348 | 0.99 | 0.05 | -0.015 | 1.00 | 0.05 | -0.017 | 0.05 | 0.00 | 0.000 | 0.05 | 0.00 | 0.000 |
| 20-24 | 6196 | 1.00 | 0.33 | 0.009 | 1.02 | 0.34 | 0.014 | 0.33 | 0.00 | 0.000 | 0.34 | 0.00 | 0.000 |
| 25-29 | 6315 | 0.98 | 0.65 | 0.013 | 1.00 | 0.66 | 0.012 | 0.65 | 0.00 | 0.000 | 0.66 | 0.00 | 0.000 |
| 30-34 | 6232 | 0.98 | 0.86 | -0.007 | 1.00 | 0.87 | -0.008 | 0.86 | 0.00 | 0.024 | 0.87 | 0.00 | 0.023 |
| 35-39 | 6207 | 1.02 | 0.98 | -0.006 | 1.05 | 0.99 | -0.007 | 0.98 | 0.02 | 0.122 | 0.99 | 0.02 | 0.120 |
| 40-44 | 6132 | 1.00 | 1.00 | 0.006 | 1.04 | 1.02 | 0.009 | 1.00 | 0.13 | 0.326 | 1.02 | 0.13 | 0.327 |
| 45-49 | 6194 | 0.97 | 0.99 | -0.015 | 1.02 | 1.01 | -0.025 | 0.99 | 0.32 | 0.459 | 1.01 | 0.33 | 0.458 |
| 50-54 | 5935 | 0.92 | 0.99 | 0.005 | 0.99 | 1.01 | 0.001 | 0.99 | 0.56 | 0.578 | 1.01 | 0.57 | 0.575 |
| 55-59 | 5930 | 0.88 | 0.99 | 0.001 | 0.99 | 1.02 | -0.002 | 0.99 | 0.78 | 0.655 | 1.02 | 0.79 | 0.654 |
| 60-64 | 5646 | 0.85 | 0.98 | 0.006 | 1.00 | 1.01 | 0.011 | 0.98 | 0.91 | 0.675 | 1.01 | 0.92 | 0.676 |
| 65-69 | 5217 | 0.78 | 0.98 | 0.021 | 1.00 | 1.02 | 0.016 | 0.98 | 0.95 | 0.703 | 1.02 | 0.97 | 0.703 |
| 70-74 | 4665 | 0.69 | 0.97 | 0.008 | 1.00 | 1.02 | 0.009 | 0.97 | 0.99 | 0.699 | 1.02 | 1.01 | 0.706 |
| 75-79 | 4032 | 0.59 | 0.94 | -0.011 | 1.00 | 1.02 | -0.007 | 0.94 | 1.01 | 0.682 | 1.02 | 1.03 | 0.695 |
| 80-84 | 3142 | 0.46 | 0.91 | -0.000 | 0.99 | 0.99 | -0.001 | 0.91 | 0.96 | 0.693 | 0.99 | 0.98 | 0.704 |
| 85-89 | 1813 | 0.34 | 0.90 | -0.015 | 1.02 | 1.04 | -0.055 | 0.90 | 1.01 | 0.644 | 1.04 | 1.04 | 0.688 |
| 90+ | 1512 | 0.16 | 0.75 | -0.036 | 0.96 | 0.95 | -0.049 | 0.75 | 0.89 | 0.627 | 0.95 | 0.93 | 0.720 |

tion declines somewhat at older ages, as the number of living daughters drops but the number of living granddaughters does not; this decline does not, of course, appear for daughters and granddaughters ever born.

Let us consider some possible resolutions of these issues. One might propose working with all women either at or beyond the highest age of childbearing, say age 45, and counting the number of sisters ever-born and children ever-born. However, note two serious criticisms of such a choice. First, conditioning upon the survival of ego will eliminate any effect that mortality would have had upon her number of children ever-born, but it will not eliminate the effect upon sisters ever-born of possible mortality of ego's mother past the point of ego's birth. The number of sisters will be reduced in a setting of high maternal mortality. Further, this choice of cutoff age would not be appropriate for relating other categories of kin. To calculate the correlation between numbers of daughters and granddaughters, for example, age 45 would be too early to have reached the final number of granddaughters ever born (in Table 1, the mean number of granddaughters ever born continues to rise to ages 75-79).

Another alternative would be to take a random sample of all females, regardless of age, and calculate their numbers of sisters and daughters either ever born or still alive. But there would be the same problems raised above for each specific age, in addition to the arbitrary weighting effects of the age distribution.

In our view, the ideal way to state the question is in terms of genealogies, which can be modelled, simulated, or described with real data, although not easily with data from a large sample survey. The genealogy for each reference individual, ego, is constructed outwards from ego and extends far enough to include all of ego's kin ultimately born—past and future—regardless of their survival. It is convenient to assume that the data extend far past the birth of ego, regardless of how long ego survives; just how far depends upon the categories of kin which are of interest. Kin will be counted if they are ever born into a specified relationship to ego, regardless of how long they survive. It is not necessary for ego and the kin ever to have been alive simultaneously. For example, ego's sisters will consist of all daughters to ego's mother, apart from ego herself, even if such a person died before ego was born or was born after ego's death.

This definition of kin, in terms of a listing of individuals ultimately born into a complete genealogy, is attractive because it eliminates the need to choose among any of the variously objectionable controls for survivorship and age. Moreover, it corresponds to the formulation developed by Pullum (1982) which gives the full frequency distributions of kin ultimately born under the assumptions of fixed and homogeneous rates. Under the definition adopted, we need not consider age when discussing the magnitude of any corre-

lation. Note that for the simulated population represented in Table 1, the correlation between ultimate numbers of sisters and daughters is -.003 (i.e. zero); between ultimate numbers of daughters and granddaughters, 0.706.

THE BASIC MODEL

Figure 1, which appeared in Pullum (1982) but is originally due to Atkins (1974), provides a convenient representation of the links between ego and her kin. The arrows along the top show the line of direct descent to ego; there is exactly one progenitor in each category. The vertical arrows in the left column show the line of direct descent from ego. The other vertical arrows show lines of descent to and from the siblings of ego's progenitors, i.e. show ego's collateral kin. The probability, defined at birth, that any individual will eventually have exactly k daughters ever-born, is f_k . The probability generating function for daughters is $f(s) = \sum f_k s^k$. From these probabilities and assumptions it is possible to obtain the moments and in fact the complete distribution for any other kin category (i,j) . For example, the eventual expected number of daughters, the net reproduction rate, is simply $N = f'(1)$. The eventual expected number of sisters is $S = f''(1)/N$.

The assumption that these probabilities are constant over time will not be used to generate any of the usual properties of a stable population. In particular, at no point will we assume or require that the age distribution is stable, because no use is made of age. The quantities to be generated only require that the basic probabilities apply to the individuals whose fertility is involved in linking ego to the categories which are being correlated.

Correlations Between Daughters and Granddaughters

The first two categories to be examined are daughters and granddaughters. Over the universe of all possible egos, define two random variables: k_1 , the number of daughters, and k_2 , the number of granddaughters. The correlation r between these two variables will be

$$r = Cov(k_1, k_2) / \sqrt{Var(k_1) Var(k_2)}.$$

From Pullum (1982) we have $E(k_1) = N$; $E(k_2) = N^2$; $Var(k_1) = N(S+1-N)$; and $Var(k_2) = N^2(N+1)(S+1-N)$. Since $Cov(k_1, k_2) = E(k_1 k_2) - E(k_1)E(k_2)$, the problem is reduced to computing $E(k_1 k_2)$. Now

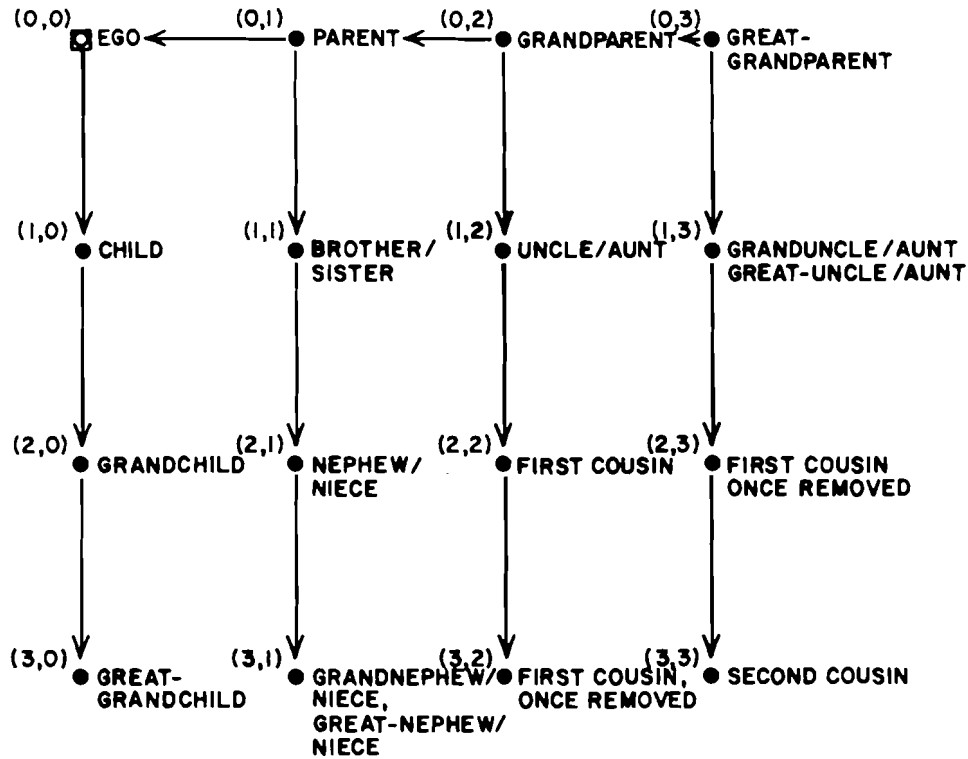


Figure 1. Two-dimensional lattice of kinship categories (i,j) for ego at $(0,0)$. Design due to Atkins (1974).

$$Pr(k_1 \cap k_2) = Pr(k_2 \cap k_1) = Pr(k_2 | k_1) Pr(k_1) .$$

It can be shown by induction (see Appendix) that $Pr(k_2 | k_1)$ is the coefficient of s^{k_2} in the expansion $f(s)^{k_1}$. Therefore,

$$\begin{aligned} E(k_1 k_2) &= \sum_{k_1} \sum_{k_2} k_1 k_2 Pr(k_2 | k_1) Pr(k_1) \\ &= \sum_{k_1, k_2} k_1 k_2 f_{k_1} \text{ (coefficient of } s^{k_2} \text{ in } f(s)^{k_1}) \\ &= \sum_{k_1} k_1 f_{k_1} \left[\sum_{k_2} k_2 \text{ (coefficient of } s^{k_2} \text{ in } f(s)^{k_1}) \right] \\ &= \sum_{k_1} k_1 f'_{k_1} \left[\frac{d}{ds} f(s)^{k_1} \Big|_{s=1} \right] \\ &= \sum_{k_1} k_1 f_{k_1} [k_1 f(s)^{k_1-1} f'(s) \Big|_{s=1}] \end{aligned}$$

$$\begin{aligned}
 &= \sum_{k_1} k_1 f_{k_1} [k_1 f(s)^{k_1-1} N] \\
 &= N \sum_{k_1} k_1^2 f_{k_1} \\
 &= N E(k_1^2) \\
 &= N [f'(1) + f''(1)] \\
 &= N^2(S + 1) .
 \end{aligned}$$

Therefore,

$$r = \frac{N^2(S + 1) - N(N^2)}{\sqrt{N(S + 1 - N) N^2(N + 1)(S + 1 - N)}} ,$$

which simplifies to

$$r = \sqrt{\frac{N}{N + 1}} .$$

Thus, in a stationary population, the correlation will be $\sqrt{1/2} = .707$. The simulated stationary population tabulated in the preceding section thus conforms to the theoretically expected result. If the population were doubling in a generation, the correlation would be $\sqrt{2/3} = .817$. Since most human populations have a current or recent level of reproduction in this range, it is possible to say that the correlation between daughters and granddaughters will be in the range of .7 to .8 even in the absence of any tendency for high or low fertility to run in families. Any such tendency would serve to raise this correlation to an even higher level.

We shall confirm this formula with a very simple example, which will also serve to make the quantities less abstract. Suppose that half of all women will have no children; half will have two. That is, $f_0 = f_2 = 0.5$; otherwise $f_k = 0$. Then the average number of daughters will be $N = 1$. Moreover, since all sibships are of size 2, every woman will have exactly one sister: $S = 1$. There are only six possible combinations of number of daughters k_1 and number of granddaughters k_2 , with probabilities $pr(k_1, k_2)$ as follows: $pr(0, 2) = pr(0, 4) = 0$; $pr(0, 0) = 0.5$; $pr(2, 0) = pr(2, 4) = 0.125$; and $pr(2, 2) = 0.25$. Summing over nonzero terms in $\sum_{k_1} \sum_{k_2} k_1 k_2 pr(k_1, k_2)$ we find $E(k_1 k_2) = 2$. Therefore we have

$E(k_1) = 1$; $E(k_2) = 1$; $Var(k_1) = 1$; $Var(k_2) = 2$; $Cov(k_1, k_2) = 2 - 1 = 1$; and $r = \sqrt{1/2} = .707$, in accordance with the general result.

Other Direct Descendants

Consider next the correlation between direct descendants in generation i and the next generation, of order $i+1$. The case considered above was for generations 1 and 2. Again let random variables k_1 and k_2 refer to the numbers of kin in each of the two generations.

The general formulas for means and variances, drawn from the basic literature on branching processes, [see Harris (1963) and Pullum (1982)] are

$$E(k_1) = N^i$$

$$E(k_2) = N^{i+1}$$

$$Var(k_1) = N^i [(N^i - 1)/(N - 1)](S + 1 - N)$$

$$Var(k_2) = N^{i+1} [(N^{i+1} - 1)/(N - 1)](N + 1)(S + 1 - N) .$$

Thus, to calculate the correlation, it is only problematic to calculate $E(k_1 k_2)$. By the same reasoning as above, using the observation that $Pr(k_2 k_1)$ is the coefficient of s^{k_2} in the expansion $f(s)^{k_1}$, it can be found that $E(k_1 k_2)$ is equal to $N E(k_1^2)$, which is the same as $N [Var(k_1) + E(k_1)^2]$. Therefore the correlation is simply

$$r = N [Var(k_1)/Var(k_2)] ,$$

which simplifies to

$$\begin{aligned} r &= \sqrt{N(N^i - 1)/(N^{i+1} - 1)} \\ &= \sqrt{(N^{i+1} - N)/(N^{i+1} - 1)} . \end{aligned}$$

In the special case of stationarity, i.e. $N = 1$, we factor a term $N - 1$ out of both the numerator and the denominator to produce the strikingly simple result that $r = \sqrt{i/(i+1)}$. The correlation rapidly approaches unity as i increases, even though in a stationary population the expected number in each generation is 1. The correlation becomes progressively stronger because of the well-known fact that the line of descent will tend either to become extinct, producing a perfect correlation between generations, or to become large.

This result is easily generalized to the correlation between any two categories of direct descendants, i_1 and i_2 , with $i_1 < i_2$. This will simply be the product of the correlations between i_1 and $i_1 + 1$, between $i_1 + 1$ and $i_1 + 2, \dots$, and between $i_2 - 1$ and i_2 , because the partial correlation between any two categories, controlling for an intermediate category, will be zero under the assumptions of the process. That is, the general formula for descendants of orders i_1 and i_2 will be

$$r_{i_1 i_2} = \sqrt{N^{i_2}(N^{i_1} - 1) / N^{i_1}(N^{i_2} - 1)} ,$$

and if $N = 1$ then

$$r_{i_1 i_2} = \sqrt{i_1 / i_2} .$$

The correlation attenuates as more distant categories are compared. It is relatively less sensitive to the value of N within a plausible range, as is seen by comparing values in the following table:

| i_1 | i_2 | N | $r_{i_1 i_2}$ |
|-------|-------|--------|---------------|
| 1 | 2 | 1.0000 | 0.7071 |
| 1 | 2 | 1.2500 | 0.7454 |
| 1 | 2 | 1.5000 | 0.7746 |
| 1 | 2 | 1.7500 | 0.7977 |
| 1 | 2 | 2.0000 | 0.8165 |
| 2 | 3 | 1.0000 | 0.8165 |
| 2 | 3 | 1.2500 | 0.8589 |
| 2 | 3 | 1.5000 | 0.8885 |
| 2 | 3 | 1.7500 | 0.9099 |
| 2 | 3 | 2.0000 | 0.9258 |
| 1 | 3 | 1.0000 | 0.5774 |
| 1 | 3 | 1.2500 | 0.6402 |
| 1 | 3 | 1.5000 | 0.6882 |
| 1 | 3 | 1.7500 | 0.7259 |
| 1 | 3 | 2.0000 | 0.7559 |

Sisters and Daughters

A correlation of great potential interest is between numbers of sisters and daughters. In a sense, this pair of categories corresponds very closely to daughters and granddaughters, but with ego shifted downward from the role of mother or grandmother to the role of one of the daughters. It might appear that the "built-in" high correlation between generations ("built-in" in the sense that it appears even when all individuals are subject

to the same basic schedule of probabilities f_k) would be seen again.

As we have described the process, however, there is no sense in which the number of daughters is conditional upon the number of sisters, or vice versa. For example, if a woman has no daughters, then she must have no granddaughters; but if she has no sisters, then we know nothing about how many daughters she has. In short, the correlation between sisters and daughters must be zero. In the simulated population depicted in Table 1, the sister-daughter correlation is (essentially) zero. If it is observed empirically to be nonzero, then we might conclude that we have evidence of genetic or behavioral inheritance of fertility.

Other Categories

There is no (nonzero) correlation between categories which are in different columns of Figure 1. The top category in each column contains one and only one individual, and there are no pathways of descent to link categories in different columns which are below the top row.

However, within any column, descent will produce correlations just as in the first column of direct descendants. As noted in Pullum (1982), for $i > 0$ and $j > 0$ the expected values and variances will depend upon i but not upon j ; that is, they vary by row but not by column.

Within any arbitrary column j , $j > 0$, let i_1 and i_2 refer to two categories, which contain k_1 and k_2 individuals, respectively. By the same reasoning as before, it can be shown that the correlation between the two counts will be

$$r_{12} = N \sqrt{[Var(k_1)/Var(k_2)]} .$$

The formula for the variance is not quite as simple as before; one general form (using k for either k_1 or k_2) is

$$Var(k) = N^{k-1}[(N^{k-1} - 1) / (N - 1)] (S + 1 - N) S + \sigma^2 N^{2k-2} ,$$

where σ^2 is the variance of the sisters' distribution:

$$\sigma^2 = f'''(1) / N + S - S^2 .$$

In particular, if the net reproduction rate is unity, then the correlation becomes

$$r_{12} = \sqrt{[(i_1 - 1)S^2 + \sigma^2] / [(i_2 - 1)S^2 + \sigma^2]} .$$

If the correlation is between sisters and any subsequent generation of their descendants (or between aunts and their descendants, etc.), i.e. if $i_1 = 1$, then there is a further simplification to

$$r_{12} = \sigma^2 / [(i_2 - 1)S^2 + \sigma^2] .$$

In the example which was presented earlier, each woman had one sister. There was no variance in the number of sisters, so all correlations involving the sisters' category would be zero, as this formula confirms. It is of interest that in this unusual situation, the correlation between any other two categories of the sisters' descendants, i_1 and i_2 , would simplify to $\sqrt{(i_1 - 1) / (i_2 - 1)}$. The correlation between nieces ($i_1 = 2$) and grand nieces ($i_2 = 3$) would be $\sqrt{1/2} = .707$ even though both of these categories would be uncorrelated with the number of sisters. (In this particular example, since ego has exactly one sister, for whom the nieces and grand nieces are her daughters and granddaughters, it is appropriate that the correlation is exactly the same as between ego's daughters and granddaughters.)

THE INHERITANCE OF REPRODUCTIVE REGIMES

So far we have considered populations in which all reproductive behavior is governed by identical probabilities, independent within genealogies. Despite this independence there are built-in correlations between the numbers of kin in certain categories, although the numbers of sisters and daughters are uncorrelated. Several studies have shown the existence of positive correlations between siblings and children, in different times and places.² We now turn to models in which such correlations can arise.

Inheritance can mean different things in this context. We will consider two quite different interpretations. In both cases there are assumed to be various different schedules of probabilities f_k , to be described as reproductive regimes, which are assigned at the point of birth. In the first case, any individual's regime is required to be the same as her mother's regime. In the second case an individual's regime is assigned on the basis of the mother's actual completed parity. We shall not consider mixtures or combinations of these two assignment procedures, but they could also be postulated.

²See, for example, Hodge and Ogawa (1986), Dansiger and Newman (1989), and several studies cited by Anderton et al. (1987).

Case 1. Strict Inheritance of Reproductive Regime

Assume that there are only two regimes, which will be labelled (), with probabilities f_k , mean number of daughters N , and mean number of sisters S ; and (*), with probabilities f_k^* , mean number of daughters N^* , and mean number of sisters S^* . We shall only consider two pairs of kin categories.

Suppose that ego and all her kin, at any remove, are subject to either regime () or to regime (*). That is, we have two classes of individuals—what an anthropologist might call moieties. One of them may have a higher NRR than the other, in which case it will increasingly dominate the other over time. In fact, the most reasonable interpretation of these two regimes would be that one of them has higher fertility and the other has lower fertility, rather than that they differed in less important ways. However, we have no need to take a long term perspective. We can assume maximum differentiation in ego's generation; that is, half of the females at the time of ego's birth are in each class. It follows that in the daughters' generation, fractions $N/(N + N^*)$ and $N^*/(N + N^*)$ will be in the two classes, and in the granddaughters' generation, the fractions will be $N^2/(N^2 + N^{*2})$ and $N^{*2}/(N^2 + N^{*2})$. If the division between the two classes is not even, then appropriate weights need to be applied here and below.

This type of heterogeneity can be regarded as one interpretation of the inheritance or inter-generational transmission of reproduction. The probabilities for a female are the same as those of her mother. But it is important to note that it is the mother's set of probabilities, rather than her actual behavior, which is transmitted. For example, a woman who is born into the class with higher fertility will transmit to her daughter(s) the predisposition to have higher fertility even if she has only one daughter herself.

Daughters and Granddaughters

We do not know which class a woman belongs to, but can nevertheless calculate the overall correlation between numbers of daughters and granddaughters.

Each expectation in the correlation, i.e. $E(k_i)$, $E(k_i^2)$ (for $k = 1,2$), and $E(k_1 k_2)$, will simply be the average of the corresponding terms for the two subpopulations. It does not follow that the variances and covariance will be simple averages, however, or will even be intermediate to their values in the separate populations. For example, the pooled variance will be (for $i = 1,2$)

$$Var_p(k_i) = [E(k_i^2) + E(k_i^{*2})]/2 - \{[E(k_i) + E(k_i^*)]/2\}^2 \quad .$$

There is no simple algebraic relationship between the correlations within the two classes,

on the one hand, and corresponding correlations in the pooled population, on the other hand.

It can be shown that if the f_k and f_k^* follow a Poisson distribution with parameters m and m^* , respectively, then the correlation between daughters and granddaughters in the pooled population will indeed lie between $\sqrt{m/(m+1)}$ and $\sqrt{m^*/(m^*+1)}$. It will take the value

$$\frac{m^2(m - m^* + 2) + m^{*2}(m^* - m + 2)}{[(m - m^*)^2 + 2m + 2m^*][(m^2 - m^{*2})^2 + 2m^2(m + 1) + 2m^{*2}(m^* + 1)]}$$

The Poisson does not provide a particularly good empirical fit for the distribution f_k , but it is a single-parameter function with the simplifying properties that N , S , and the variance are all equal to the same parameter (labelled m). If $m = 1$ and $m^* = 2$, then the correlation between daughters and granddaughters will be .707 in class (), .817 in class (*), and .808 in the pooled population, suggesting that more generally the correlation is intermediate but tends toward the one in the higher-growth subpopulation(s).

Sisters and Daughters

When the kind of heterogeneity described above exists, a correlation will emerge between the number of sisters and the number of daughters. Continue with two classes () and (*), which are of equal size in ego's generation, and assume that the class membership of ego is not known. Then the variance in the number of daughters or the variance in the number of sisters will be calculated exactly as above, using the average of the expected numbers of daughters (or sisters) and the average of the expected numbers of daughters squared (or sisters squared) as calculated within the two classes. If k_3 is the number of sisters, then the only quantity which is problematic is $E_p(k_1 k_3)$.

Within each class, we have seen earlier that the number of daughters will be independent of the number of sisters. That is, $E(k_1 k_3) = NS$ and $E^*(k_1 k_3) = N^*S^*$. Therefore

$$\begin{aligned} E_p(k_1 k_3) &= E[k_1 k_3 | \text{ego in class ()}] Pr[\text{ego in class ()}] \\ &\quad + E[k_1 k_3 | \text{ego in class (*)}] Pr[\text{ego in class (*)}] \\ &= SN(1/2) + S^*N^*(1/2) = (SN + S^*N^*)/2 \end{aligned}$$

Consequently

$$\begin{aligned} Cov_p(k_1, k_3) &= (SN + S^*N^*)/2 - [(S + S^*)/2][(N + N^*)/2] \\ &= (S - S^*)(N - N^*)/4 . \end{aligned}$$

Therefore the correlation will be non-zero (positive in any plausible context) if the two subpopulations differ in both the expected number of sisters and the expected number of daughters. This is a very general observation: if the population contains two or more homogeneous subgroups, there cannot be a correlation between numbers of sisters and of daughters unless the reproductive regimes differ across subgroups in the expected numbers of the two counts.

In the simple example of two Poisson distributions, the correlation is $(m - m^*)^2 / [(m - m^*)^2 + 2(m + m^*)]$. For parameters $m = 1$ and $m^* = 2$ the implied correlation is $1/7$ or 0.143 . For parameters $m = .5$ and $m^* = 1.5$ it is $1/5$ or 0.200 . Although this is not of the same magnitude as the correlation between daughters and granddaughters, it is large enough to be easily detected.

More general formulas could be developed for the correlation between daughters and sisters (or other pairs of categories) using more classes and even a continuous blending of reproductive regimes, but will not be presented here. The correlation will clearly be a maximum if there are two classes of equal size at the low end and at the high end, respectively, of a reproductive range. This observation, together with the simple example of a Poisson, suggest that in a developed country the type of heterogeneity discussed here can potentially produce a correlation between daughters and sisters as large as approximately 0.2 .

Case 2. Reproductive Regime Determined by Sibship Size

The above pattern of transmission could be a plausible representation of genetic factors. But to the extent that heterogeneity results from the size of the family of socialization, following the arguments developed in Anderton et al. (1987), it is more plausible that ego will be influenced by the number of children that her mother actually has, rather than by her mother's schedule of probabilities.

A desirable specification of this socialization effect might be as follows: a woman's reproductive regime will increase monotonically (as indexed by her expected fertility) according to the completed parity of her mother. A possible refinement to this would condition upon the survival of the children in the sibship, for example to age five, or would count the sibship size when ego was age fifteen, for example. Other refinements could also be imagined, depending upon the mechanisms which are postulated to underlie the transmission.

We shall say that the regime which applies to a woman is determined completely by her sibship size j , i.e. by the number of daughters her mother had ($j = 1, 2, \dots$), which will be taken to define the class of the woman. All women in class j are subject to a set of probabilities of having k children themselves ($k = 0, 1, \dots$), labelled f_{kj} . These probabilities are assumed to be fixed from one generation to the next. Define N_{1j} to be the expected completed parity of women who are born into a sibship of size j , i.e. $N_{1j} = \sum_k k f_{kj}$.

Similarly define the expected value of the squared parity of such women to be $N_{2j} = \sum_k k^2 f_{kj}$.

In generation g it is necessary to specify the proportion of the women who are in class j , labelled p_{gj} . It is reasonable to take generation $g = 0$ to be ego's generation.

Define π_{gk} to be the expected proportion of women in generation g who will have exactly k children, or equivalently, the probability that a woman who is sampled at birth—unconditional upon knowledge of her sibship size—will eventually have exactly k children. The sum $\sum_k \pi_{gk}$ must equal one in each generation g .

The class-specific probabilities must be weighted by the proportions of women to whom they apply in order to obtain the unconditional probabilities of having k children in generation g . The weights themselves are obtained by linking one generation with the next. The following two equations describe the process formally:

$$\begin{aligned} \pi_{gk} &= \sum_j f_{kj} p_{gj} && \text{for all } g \text{ and } k; \text{ and} \\ p_{gj} &= c_g \sum_k k f_{kj} p_{g-1,j} && \text{for all } g \text{ and } j. \end{aligned}$$

Here c_g is simply a constant to ensure that $\sum_k \pi_{gk} = 1$ in generation g . This condition requires that

$$c_g = 1 / \left(\sum_j p_{g-1,j} N_{1j} \right) .$$

If the conditional distributions f_{kj} are known, together with the distribution p_{gj} from some starting generation g , then with these two equations it is possible to calculate the probabilities for all categories of kin recursively.

Daughters and Granddaughters

The probability that the woman is in a sibship of size j is p_{0j} , and if she is in such a sibship, she expects N_{1j} daughters. Therefore the unconditional expected number of daughters is

$$E(k) = \sum_j p_{0j} N_{1j}$$

and the expected number of daughters squared is

$$E(k^2) = \sum_j p_{0j} N_{1j}^2 .$$

In order to avoid the proliferation of subscripts, let h represent the number of granddaughters. If a woman comes from a sibship of size j , then she has a probability f_{kj} of having k daughters. Each of these k daughters, being in a sibship of size k , will herself expect N_{1k} daughters, furnishing ego with kN_{1k} granddaughters. Therefore the expected number of granddaughters will be

$$E(h) = \sum_j p_{0j} \sum_k k N_{1k} f_{kj} .$$

The expected number of granddaughters squared is the least intuitive expectation required for the correlation between daughters and granddaughters, and is obtained by noting that $Pr(h|k)$ is the coefficient of s^h in the expansion $f_k(s)^k$, where $f_k(s)$ is the probability generating function for the daughters of females in a sibship of size k . The summation of h^2 times $Pr(h|k)$ is obtained by evaluating the sum of the first and second derivatives of $f_k(s)^k$ at $s = 1$, because $E(h^2|k) = E(h(h-1)|k) + E(h|k)$:

$$E(h^2) = \sum_j p_{0j} \sum_k k [N_{2k} + (k-1) N_{1k}^2] f_{kj} .$$

To obtain the expected value of the product of daughters and granddaughters, note that if a woman has k daughters, the expected product of daughters and granddaughters will be k times $kN_{1k} = k^2N_{1k}$. Therefore

$$E(kh) = \sum_j p_{0j} \sum_k k^2 N_{1k} f_{kj} .$$

When all sibships of all sizes j have the same reproductive regimes $f_{kj} = f_k$, it can be shown that all of these formulas, and the ones given below for sisters, are consistent with the basic model given earlier. In the basic model, the symbol S was used; it can be defined in terms of the first and second moments by $S = (N_2/N_1) - 1$.

Sisters and Daughters

Despite the inconsistency with the earlier part of this paper, it will be convenient to use the letter j as the index for sisters. The probability that ego has j sisters is simply the probability that she was born into a sibship of size $j+1$, which is $p_{0,j+1}$. Therefore the expected number of sisters is

$$E(j) = \sum_j j p_{0,j+1}$$

and the expected number of sisters squared is

$$E(j^2) = \sum_j j^2 p_{0,j+1} .$$

The remaining quantity required for the correlation between sisters and daughters is $E(jk)$. Since

$$\begin{aligned} Pr(j \text{ sisters and } k \text{ daughters}) &= Pr(k \text{ daughters} \mid j \text{ sisters}) * Pr(j \text{ sisters}) \\ &= f_{k,j+1} p_{0,j+1} , \end{aligned}$$

it follows that

$$\begin{aligned} E(jk) &= \sum_{jk} jk f_{k,j+1} p_{0,j+1} \\ &= \sum_j j p_{0,j+1} \sum_k k f_{k,j+1} \\ &= \sum_j j p_{0,j+1} N_{1,j+1} \\ &= \sum_j j N_{1,j+1} p_{0,j+1} . \end{aligned}$$

The correlation is then calculated in the usual way, from $Var(j) = E(j^2) - E(j)^2$; $Var(k) = E(k^2) - E(k)^2$, $Cov(j,k) = E(jk) - E(j)E(k)$, and $r_{jk} = Cov(j,k) / \sqrt{Var(j)Var(k)}$.

Observe that under this model the correlation is determined wholly by the following quantities:

- (1) the expected parity of females whose mothers had j daughters, N_{1j} ,
- (2) the expected squared parity of females whose mothers had j daughters, N_{2j} , and
- (3) the unconditional probability that ego is in class j , i.e. that she was born into a sibship of size j or p_{0j} .

It is thus not necessary to specify—at least in order to compute the correlation—the full distribution of conditional probabilities f_{jk} .

We now turn to the problem of specifying the plausible values of these quantities, in order to estimate—or to simulate—the magnitude of the correlation between sisters and daughters.

The chance that ego was born into a sibship of size j , or p_{0j} , will be derived for illustrative purposes from the stationary distribution of probabilities used by Pullum (1982), describing a cohort of women in the United States in 1978. Referring to their empirical probabilities of having j daughters as p_j , we shall take $p_{0j} = jp_j / \sum_j jp_j = jp_j$, since the sum $\sum_j jp_j = 1$ in this stationary distribution. The average sibship size, from ego's perspective, will be 1 plus the mean number of sisters. The relevant probability distributions are as follows:

| k or j | Probability of k daughters | Probability of sibship size j |
|------------|------------------------------|---------------------------------|
| 0 | .336 | |
| 1 | .395 | .395 |
| 2 | .213 | .426 |
| 3 | .046 | .137 |
| 4 | .008 | .032 |
| 5 | .001 | .007 |
| 6 | .000 | .002 |
| Mean | 1.000 | 1.835 |
| Variance | .815 | |

We shall assume for simplicity that there is a simple relationship between the mean and the variance of the conditional distributions of f_{kj} (conditional on j). Specifically, we assume that the variance is proportional to the mean:

$$Var(j) = N_{2j} - N_{1j}^2 = aN_{1j} \quad ,$$

so that the second moment is

$$N_{2j} = aN_{1j} + N_{1j}^2 \quad .$$

Under a Poisson distribution, a would be 1. In the observed distribution, p_j , we find $a = .815$. This value of a is taken to hold for all of the conditional distributions, so that in a sense the distributions are the same in their basic shapes but different in their means.

Finally, we specify a plausible set of mean values N_{1j} , for $j = 1, 2, \dots$. If N_{1j} were a constant, N_1 , then there would be no heterogeneity and no correlation between daughters and sisters. It is this specification of heterogeneity which will produce a (positive) correlation.

We have assumed that the mothers of ego's generation had replacement fertility, meaning that their NRR was 1; this translated to a mean sibship size of 1.835 (or a mean of .835 sisters) in ego's generation because a large proportion of women have no daughters at all. It is clear that in the implied stationary population, a woman tends to have substantially fewer daughters than siblings. Therefore, if the simulated population is not to grow, we should require that N_1 is 1 for women whose sibship size is 1.835. (Of course, the actual sibship size can only be an integer.) Consider the following function to link the generations: $N_{1j} = 1 + b(j - 1.835)$. This states that for every child by which ego's own sibship exceeds the mean of 1.835, she will herself be expected to have an additional fraction b of a child. The mean number of daughters to ego's generation will be

$$\begin{aligned} \sum_j p_{0j} N_{1j} &= \sum_j p_{0j} [1 + b(j - 1.835)] \\ &= \sum_j p_{0j} + b \sum_j (j - 1.835) p_{0j} \\ &= 1 + b \sum_j j p_{0j} - 1.835 b \sum_j p_{0j} \\ &= 1 + 1.835 b - 1.835 b = 1 \quad , \end{aligned}$$

which is 1, since

$$\sum_j p_{0j} = 1 \quad \text{and} \quad \sum_j j p_{0j} = 1.835 \quad .$$

Therefore we have no change in the overall mean from one generation to the next, despite the heterogeneity.

Using the equations given earlier and the stationary regime for the United States in 1978, we can calculate the correlation between sisters and daughters implied by different values of b . Table 2 gives the means and variances of the number of daughters, and the value of r , for women in sibship sizes $j = 1$ through 6 when b ranges from 0.0 to 1.0 in increments of .1. Observe that when $b = 0.0$ there is no differentiation.

Note that the mean fertility of only daughters—that is, those with $j = 1$ —becomes extremely low when b is large. These daughters, who are prominent in the first generation ($f_1 = .395$), are the only ones from a below-average sibship, and therefore the only

Table 2. Mean and variance of number of daughters (k), and correlation between sibship size (j) and number of daughters by sibship.

| Value of b | | Sibship size (j) | | | | | | r_{jk} |
|--------------|----------|----------------------|-------|-------|--------|--------|--------|----------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | |
| 0.0 | $E(k)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 |
| | $var(k)$ | 1.815 | 1.815 | 1.815 | 1.815 | 1.815 | 1.815 | |
| 0.1 | $E(k)$ | 0.917 | 1.017 | 1.117 | 1.217 | 1.317 | 1.417 | 0.094 |
| | $var(k)$ | 1.587 | 1.862 | 2.157 | 2.472 | 2.806 | 3.161 | |
| 0.2 | $E(k)$ | 0.833 | 1.033 | 1.233 | 1.433 | 1.633 | 1.833 | 0.186 |
| | $var(k)$ | 1.373 | 1.909 | 2.525 | 3.221 | 3.998 | 4.854 | |
| 0.3 | $E(k)$ | 0.750 | 1.050 | 1.350 | 1.650 | 1.950 | 2.250 | 0.273 |
| | $var(k)$ | 1.173 | 1.957 | 2.921 | 4.065 | 5.389 | 6.894 | |
| 0.4 | $E(k)$ | 0.666 | 1.066 | 1.466 | 1.866 | 2.266 | 2.666 | 0.354 |
| | $var(k)$ | 0.986 | 2.005 | 3.344 | 5.003 | 6.982 | 9.280 | |
| 0.5 | $E(k)$ | 0.583 | 1.083 | 1.583 | 2.083 | 2.583 | 3.083 | 0.427 |
| | $var(k)$ | 0.814 | 2.054 | 3.794 | 6.034 | 8.774 | 12.014 | |
| 0.6 | $E(k)$ | 0.499 | 1.099 | 1.699 | 2.299 | 2.899 | 3.499 | 0.493 |
| | $var(k)$ | 0.656 | 2.104 | 4.271 | 7.159 | 10.767 | 15.095 | |
| 0.7 | $E(k)$ | 0.416 | 1.116 | 1.816 | 2.516 | 3.216 | 3.916 | 0.552 |
| | $var(k)$ | 0.511 | 2.154 | 4.776 | 8.378 | 12.960 | 18.522 | |
| 0.8 | $E(k)$ | 0.332 | 1.132 | 1.932 | 2.732 | 3.532 | 4.332 | 0.603 |
| | $var(k)$ | 0.381 | 2.204 | 5.307 | 9.690 | 15.354 | 22.297 | |
| 0.9 | $E(k)$ | 0.249 | 1.149 | 2.049 | 2.949 | 3.849 | 4.749 | 0.648 |
| | $var(k)$ | 0.264 | 2.255 | 5.866 | 11.097 | 17.948 | 26.418 | |
| 1.0 | $E(k)$ | 0.165 | 1.165 | 2.165 | 3.165 | 4.165 | 5.165 | 0.687 |
| | $var(k)$ | 0.162 | 2.307 | 6.452 | 12.597 | 20.742 | 30.887 | |

ones who can balance the fertility of those from above-average sibships in our zero-growth population.

If the range between the least and most fertile classes is one daughter, i.e. the same as the level of the assumed net reproduction rate itself, then $b = .2$, implying a correlation of .1857. Both Case 1 and Case 2, although using much different mechanisms of inheritance of reproductive regime, lead us to the expectation that the correlation between sisters and daughters in a stationary population will be no higher than approximately .2, if the range across subclasses is one daughter.

SOME EMPIRICAL CORRELATIONS

We have discussed in mainly abstract terms some mechanisms which might give rise to correlations between the numbers of sisters and daughters a woman might have. We have also shown that in a population in which all individuals have independent fertility behavior, correlations will still arise between the numbers of kin along lines of descent. We now present several empirical correlations, for selected pairs of kin types, and from several countries.

As noted before, several studies have considered intergenerational transmission of fertility patterns. Generally the results have been conveyed in the form of coefficients from a multiple regression [eg. in Hodge and Ogawa (1986) or Danziger and Newman (1989)] or comparisons of group means [e.g. in Anderton et al. (1987)] rather than in the form of a zero-order, product-moment correlation coefficient. Here we present correlation coefficients, and do not restrict ourselves to sibling/offspring kin pairs.

The correlations presented here are, of course, subject to all the limitations and qualifications discussed at the outset: they are gathered from surviving members of a population at a point in time, and (with a few exceptions) pertain to counts of currently-living kin of the indicated type. Thus, the figures can not be viewed as evidence for or against any particular theory about population dynamics, but rather as illustrative and, possibly, suggestive facts to guide further investigations.

We present figures from five surveys: the Hungarian Microcensus of 1984 (Hungarian CSO, 1986); the U.S. Supplement on Aging to the National Health Interview Survey of 1984 (Kovar, 1986); Canada's 1985 General Social Survey; the 1981-1982 German Life History Survey (Tuma and Huinick, 1987); and the 1974 Israeli Labor Mobility Survey (Danziger and Newman, 1989). In the first four of these surveys, we are able to group ego by age. The types of kin considered include children, grandchildren, siblings, sisters, and daughters.

The empirical correlations are presented in Table 3. We show first the correlations between numbers of children and numbers of grandchildren, for several age-groups of older Hungarian women. In the youngest age groups (under 70) the acquisition of grandchildren is undoubtedly incomplete; in the oldest groups, problems due to selective mortality may be operating. In spite of this the observed correlations fall within a fairly narrow range. And more importantly, they are generally well above the levels expected in the independent-fertility model. Thus the numbers imply positive correlations between the fertility of mothers and daughters in Hungary.

Table 3. Selected empirical correlations between kin types; various countries.

A. Results from Hungarian microcensus of 1984; older women; children and grandchildren

| Age group | Mean number of children | Variance children | Mean number of grand-children | Variance grand-children | Correlation coefficient | n |
|-----------|-------------------------|-------------------|-------------------------------|-------------------------|-------------------------|------|
| 55-59 | 1.9307 | 1.7531 | 2.5572 | 5.8281 | 0.7431 | 7444 |
| 60-64 | 1.9870 | 2.0949 | 2.9568 | 7.1700 | 0.8070 | 7239 |
| 65-69 | 1.9442 | 2.3018 | 3.0685 | 8.3768 | 0.8255 | 3725 |
| 70-74 | 1.9007 | 2.5138 | 3.0863 | 9.0399 | 0.8337 | 5259 |
| 75-79 | 1.8434 | 2.6997 | 3.0874 | 9.8986 | 0.8349 | 3467 |
| 80-84 | 1.8339 | 2.7609 | 3.2100 | 10.7975 | 0.8039 | 1957 |
| 85-89 | 1.6773 | 2.3658 | 2.9902 | 10.2723 | 0.7859 | 815 |
| 90+ | 1.6449 | 1.9300 | 3.2757 | 11.7231 | 0.7043 | 214 |

B. Results from U.S. SOA file of 1984; older women; sisters and daughters

| Age group | Mean number of sisters | Variance sisters | Mean number of daughters | Variance daughters | Correlation coefficient | n |
|-----------|------------------------|------------------|--------------------------|--------------------|-------------------------|------|
| 55-59 | 1.7015 | 3.6942 | 1.5103 | 1.7923 | 0.0916 | 1262 |
| 60-64 | 1.6394 | 2.7817 | 1.3991 | 1.8984 | 0.0246 | 1207 |
| 65-69 | 1.5717 | 2.4859 | 1.2398 | 1.6568 | 0.0686 | 2139 |
| 70-74 | 1.4638 | 2.1810 | 1.1811 | 1.8583 | 0.0738 | 1805 |
| 75-79 | 1.2357 | 1.9569 | 1.1485 | 1.8329 | 0.0825 | 1396 |
| 80-84 | 1.0265 | 1.5542 | 1.0609 | 1.9516 | 0.0710 | 839 |
| 85-89 | 0.8375 | 1.9440 | 1.1615 | 1.5760 | 0.0886 | 420 |
| 90+ | 0.5460 | 0.8312 | 1.3664 | 2.2896 | -0.1346 | 150 |

C. Results from the 1985 Canadian GSS; older women; siblings and children

| Age Group | Mean number of siblings | Variance siblings | Mean number of children | Variance children | Correlation coefficient | n |
|-----------|-------------------------|-------------------|-------------------------|-------------------|-------------------------|-----|
| 55-59 | 4.1469 | 10.9157 | 3.2030 | 3.6694 | 0.3138 | 315 |
| 60-64 | 3.6726 | 9.0147 | 3.2887 | 5.7285 | 0.3266 | 311 |
| 65-69 | 3.5634 | 9.2108 | 3.2593 | 5.6054 | 0.2407 | 401 |
| 70-74 | 3.4654 | 7.8613 | 2.8560 | 5.7521 | 0.1874 | 507 |
| 75-79 | 3.3790 | 9.5679 | 2.3742 | 4.8865 | -0.0578 | 434 |
| 80-84 | 2.6147 | 5.4356 | 2.6742 | 5.7102 | 0.0943 | 362 |
| 85+ | 1.8872 | 4.3967 | 2.5524 | 5.7265 | 0.1663 | 357 |

D. Results from 1981-1982 German Life History Survey; women by birth cohort ; siblings and children

| Birth cohort | Mean number of siblings | Variance siblings | Mean number of children | Variance children | Correlation coefficient | n |
|--------------|-------------------------|-------------------|-------------------------|-------------------|-------------------------|-----|
| 1949-51 | 2.117 | 3.516 | 1.185 | 1.105 | 0.1519 | 729 |
| 1939-41 | 2.563 | 2.618 | 1.932 | 1.683 | 0.2127 | 727 |
| 1929-31 | 2.720 | 5.298 | 2.229 | 2.347 | 0.1552 | 709 |

E. Results from 1974 Israeli Labor Mobility Survey; currently-married women 95 and older ; siblings and children

| Mean number of siblings | Variance siblings | Mean number of children | Variance children | Correlation coefficient | n |
|-------------------------------|----------------------|-------------------------------|----------------------|----------------------------|-----|
| 3.768 | 6.062 | 3.428 | 7.089 | 0.286 | 708 |

Direct evidence on the correlation of numbers of sisters and daughters is provided for the United States in part B of Table 3. All the ego age-groups are clearly past childbearing age, but the contaminating effects of mortality (in both generations) of course remain. Although the pattern of correlations by age is mildly erratic, the correlations are (with one peculiar exception, for the rather small group of women aged 90+) distinctly positive.

Much larger correlations between mother's and ego's fertility is revealed by the Canadian, German, and Israeli data which complete Table 3. For these three countries, only the Canadian women, and the earliest cohort of German women, are beyond childbearing age. On the other hand, whereas the Hungarian, U.S., and Canadian data pertain only to currently-living kin, the German and Israeli data relate the number of siblings ever-born to ego's mother, to the number of children borne by ego. The latter are relatively uninfluenced by mortality, which may be the reason for the high observed correlations.

CONCLUSION

Our analysis, in combination with other recent empirical findings, indicates the presence of correlations between kin counts in several locations in time and space. Others [particularly Anderton et al. (1987)] have concentrated on correlations between the fertility of mothers and their daughters, and have speculated on the mechanism through which the observed correlations arise. We contribute to this literature by showing how correlations between selected pairs of kin counts can arise even when the fertility of mothers and daughters is uncorrelated. We also discuss, and illustrate with figures obtained in a microsimulation of kin patterns, the importance of age of ego and the influence of mortality on the computation of kin correlations.

Kin patterns are of interest because one's pool of available kin constitute a resource which influences other dimensions of life such as household composition and the provision of health and social-support services. Skewness in the distribution of available-kin pools, a skewness which may be due in part to correlations in reproductive behavior within and across generations, thus can be viewed as a form of unequal access to resources. These observations suggest that further analysis of correlations between kin counts, and of the reasons for these correlations, would be worthwhile.

REFERENCES

- Anderton, D.L., N.O. Tsuya, L.L. Bean and G.P. Mineau (1987) Intergenerational transmission of relative fertility and life course patterns. *Demography* 24: 467-480.
- Atkins, J.R. (1974) On the fundamental consanguineal numbers and their structural basis. *American Ethnologist* 1:1-31.
- Bongaarts, J. (1987) The projection of family composition over the life course with family status life tables. In Bongaarts, J., T. Burch, and K. Wachter (eds.), *Family Demography: Methods and their Application*. Oxford University Press, 1987.
- Bongaarts, J., T. Burch, and K. Wachter (eds.) (1987) *Family Demography: Methods and their Application*. Oxford University Press.
- Berent, J. (1953) Relationship between family sizes of two successive generations. *Milbank Memorial Fund Quarterly Bulletin* 31:39-50.
- Broberg, P. (1987) Sibling Dependencies in Branching Populations: Effects on Growth, Composition and Extinction Probability. Department of Mathematics, Gothenburg University.
- Danziger, L. and S. Newman (1989) Intergenerational effects on fertility. Theory and evidence from Israel. *Journal of Population Economics* 2:25-38.
- Duncan, O.D., R. Freedman, J.M. Coble, and D.P. Slesinger (1965) Marital fertility and size of family of orientation. *Demography* 2:508-515.
- Goodman, L.A., N. Keyfitz, and T.W. Pullum (1974) Family formation and the frequency of various kinship relationships. *Theoretical Population Biology* 5:1-27.
- Goodman, L.A., N. Keyfitz, and T.W. Pullum (1975) Addendum. *Theoretical Population Biology* 8:376-381.
- Harris, T.E. (1963) *The Theory of Branching Processes*. Englewood Cliffs, N.J.: Addison-Wesley.
- Hodge, R.W. and N. Ogawa (1986) Siblings and family size from generation to generation. *NUPRI Research Paper Series No. 29*. Population Research Institute, Nihon University, Tokyo, Japan.
- Hungarian Central Statistical Office (1986) Data of the 1984 Hungarian Microcensus. Budapest, Hungary.
- Kovar, M.G. (1986) Aging in the eighties, age 65 years and over and living alone, contacts with family, friends, and neighbors. *Advance Data from Vital and Health Statistics, no. 116*. DHHS Pub. No. (PHS) 86-1250. Public Health Service, Hyattsville, Maryland, May 9, 1986.
- Langford, C.M. and C. Wilson (1985) Is there a connection between a woman's fecundity and that of her mother? *Journal of Biosocial Science* 17:437-443.
- Le Bras, H. (1973) Parents, grands-parents, et bisaieux. *Population* 28:9-38.
- National Center for Health Statistics (1987) Advance report of final natality statistics, 1985. *Monthly Vital Statistics Report, Vol. 36, No. 4 Supp.* DHHS Pub. No. (PHS) 87-1120. Public Health Service. Hyattsville, Maryland, July 17, 1987.
- Preston, S.H. (1976) Family sizes of children and family sizes of women. *Demography* 13:105-114.
- Pullum, T.W. (1982) The eventual frequencies of kin in a stable population. *Demography* 19:549-565.

- Pullum, T.W. (1987) Some mathematical models of kinship and the family. In Bongaarts, J., T. Burch, and K. Wachter (eds.), *Family Demography: Methods and their Application*. Oxford University Press.
- Tuma, N.B. and J. Huinick (1987) Post-war fertility patterns in the Federal Republic of Germany. In Karl Ulrich Mayer and Nancy Brandon Tuma (eds.), *Applications of Event History Analysis in Life Course Research*. Materialien aus der Bildungsforschung Nr. 30. Max-Planck-Institut für Bildungsforschung, Berlin.
- Waugh, W.A. O'N. (1981) Application of the Galton-Watson process to the kin number problem. *Advances in Applied Probability* 13:631-649.
- Wolf, D.A. (1988) *Kinship and Family Support in Aging Societies*. In United Nations Department of International Economics and Social Affairs, *Economic and Social Implications of Population Aging*. New York: United Nations.

APPENDIX: Conditional probability of k granddaughters given that ego has h daughters

A result used several times in this paper is that if h and k are the numbers of daughters and granddaughters, respectively, in the homogeneous case, then the conditional probability $Pr(k | h)$ is the coefficient of s^k in the expansion $f(s)^h$. This is demonstrated in three steps.

First, recall the definition of the probability generating function as $f(s) = \sum_k f_k s^k$.

That is, the coefficient of s^k in $f(s)$ is the probability of having k daughters. Now suppose that a woman has exactly one daughter. Then this daughter serves as an alter for ego, so the probability that ego will have k granddaughters is exactly the same as the probability that the daughter will herself have k daughters, i.e. f_k . This establishes the result for $h = 1$.

As the second step, define the (unknown) probability generating function to be $h(s) = \sum_k h_k s^k$, defined by the requirement that h_k is the conditional probability $Pr(k | h)$, whatever that may be. We then note that if ego has $h+1$ daughters rather than h , then $Pr(k | h+1)$, the conditional probability of k granddaughters, will be the sum $h_k f_0 + h_{k-1} f_1 + h_{k-2} f_2 + \dots + h_0 f_k$, which is readily shown to be the coefficient of s^k in the product $h(s)f(s)$.

As a third step, we simply invoke the induction principle. If $h = 1$, we have found that $h(s)$ is $f(s)$; if $h = 2$, then $h(s)$ is $f(s)f(s) = f(s)^2$; etc.; establishing the general result.

**Recent Working Papers Produced in
IIASA's Population Program**

Copies may be obtained at a cost of US\$ 5.00 each from IIASA's
Publications Department.

- WP-88-10, *On the Concentration of Childbearing in China, 1955-1981* by W. Lutz. February 1988.
- WP-88-13, *Beyond "The Average American Family": U.S. Cohort Parity Distributions and Fertility Concentration* by M. King and W. Lutz. March 1988.
- WP-88-23, *Understanding Medical and Demographic Trends with MEDDAS* by M. Rusnak and S. Scherbov. April 1988.
- WP-88-32, *Kinship Patterns and Household Composition of the Elderly: Hungarian Women, 1984* by D. Wolf. April 1988.
- WP-88-36, *"DIAL" - A System for Modeling Multidimensional Demographic Processes* by S. Scherbov and V. Grechucha. May 1988.
- WP-88-44, *Kin Availability and the Living Arrangements of Older Unmarried Women: Canada, 1985* by D. Wolf, T. Burch, and B. Matthews. June 1988.
- WP-88-46, *Population Futures for Europe: An Analysis of Alternative Scenarios*, by D. Wolf, B. Wils, W. Lutz, and S. Scherbov. June 1988.
- WP-88-90, *Comparative analysis of Completed Parity Distributions: A Global WFS-Perspective*, by W. Lutz. October 1988.
- WP-88-104, *Future Regional Population Patterns in the Soviet Union: Scenarios to the Year 2050*, by S. Scherbov and W. Lutz. November 1988.
- WP-88-120, *AIDS and HIV Surveillance in Europe*, by M. Artzrouni and G. Heilig. December 1988.
- WP-88-124, *DDMSLT: A Computer Program for Estimating the Duration-Dependent Multistate Life Table Model*, by C. Calhoun. December 1988.
- WP-89-05, *Multi-State Analysis of Family Dynamics in Austria: Scenarios to the Year 2030*, by W. Lutz and S. Scherbov. January 1989.
- WP-89-06, *The Demographic Dimensions of Divorce: The Case of Finland*, by W. Lutz, B. Wils, and M. Nieminen. January 1989.
- WP-89-18, *Markets as Queues, with an Application to Education*, by R. Boylan. February 1989.
- WP-89-19, *Living Arrangements and Family Networks of Older Women in Italy*, by D. Wolf and A. Pinnelli. February 1989.
- WP-89-27, *Reconciling Economic and Ecological Theory on Population*, by Nathan Keyfitz. March 1989.
- WP-89-28, *Multistate Life Table with Duration-Dependence: An Application to Hungarian Female Marital History*, by Alain Belanger. April 1989.
- WP-89-34, *Recent Trends in Living Arrangements in Fourteen Industrialized Countries*, by J.-P. Gonnot and G. Vukovich. May 1989.
- WP-89-35, *Averaging Life Expectancy*, by E. Andreev, W. Lutz, and S. Scherbov. June 1989.
- WP-89-37, *Measuring Fertility Responses to Policy Measures in the German Democratic Republic*, by T. Büttner and W. Lutz. June 1989.
- WP-89-38, *BIVOPROB: A Computer Program for Maximum-Likelihood Estimation of*

- Bivariate Ordered-Probit Models for Censored Data*, by C. Calhoun. June 1989.
- WP-89-41, *Desired and Excess Fertility in Europe and the United States: Indirect Estimates from World Fertility Survey Data*, by C. Calhoun. July 1989.
- WP-89-43, *Youth Cohorts, Population Growth and Political Outcomes*, by H. Wriggins. July 1989.
- WP-89-52, *The Profile of Intecohort Increase*, by N. Keyfitz. August 1989.