

# Confessions of a Microsimulator

## Problems in Modeling the Demography of Kinship

Steven Ruggles  
*Department of History*  
*University of Minnesota*

*I could not, without effort, constrain myself to the task of either recalling, or constructing into a regular narrative, the whole burthen of horrors which lies upon my brain.*

—Thomas DeQuincey, *Confessions of an Opium Eater*

Ever since Peter Laslett and John Harrison (1963) discovered that multigenerational households were rare in preindustrial northwestern Europe, historians and demographers have been trying to estimate the effects of preindustrial demographic conditions on the potential for multigenerational households. Starting with back-of-the-envelope calculations by Ansley Coale (1965), David Glass (1966), and E.A. Wrigley (1969), the techniques for assessing the effects of demography on historical kinship patterns have become more and more elaborate, culminating in complex demographic microsimulation models consisting of many thousands of lines of computer code (Wachter, Hammel, and Laslett 1978; Ruggles 1987; Smith 1987).

All demographic models incorporate simplifying assumptions, and most investigators are well aware of the resulting potential for error. This article argues that the errors in the historical demographic models of kinship are not just misspecifications that tend to cancel one another out; instead, the assumptions common to all existing models lead to errors of consistent direction that tend to cumulate.

In real populations, members of the same kin group tend to share many of the same characteristics—what one might term a *demographic family resemblance*. Family members typically belong to the same ethnic group, class, and religion and are likely to reside in the same region. For these reasons, members of the same kin group usually resemble one another in demographic behavior more than they resemble persons randomly selected from the population as a whole. For example, a kin group in which the members are poor will usually experience systematically higher mortality than a kin group with wealthy members.

By contrast, models of kinship assume that most demographic events occurring within a kin group are inde-

pendent of one another. That is, the characteristics of one member of a group of kin are assumed to be entirely uncorrelated with the characteristics of other members of the kin group. I call this the Whopper Assumption.<sup>1</sup>

Because of the Whopper Assumption, models of kinship will produce less variation in the frequency of kin of any particular type than would occur in a real population. Our models will generally underestimate both the proportion of people with many kin and the proportion of people with few kin. As I will show, the magnitude of error is potentially large. In addition, the Whopper Assumption can affect the expected number of kin in a population.

What follows is presented in the terms of demographic microsimulation models—currently the technique of choice for analyzing the demography of the family and kinship—but most of my comments would apply equally to other kinds of kinship models. In brief, microsimulation involves creating a hypothetical population by randomly assigning demographic events—births, deaths, and marriages—to individuals on the basis of predetermined probabilities. One can keep track of the family relationships within the simulated population and thus derive estimates of the frequencies of living kin.<sup>2</sup> The microsimulation approach is especially appropriate for estimating the frequency distribution of kin instead of just mean numbers of kin (Ruggles 1990). The difficulty is that the estimates of distributions are systematically biased.

### The Fertility Problem

The correlation between the fertility of mothers and that of their daughters is well known (see, for example, Pearson and Lee 1899; Huestis and Maxwell 1932; Berent 1953; Kantner and Potter 1954; Duncan et al. 1965; Henderson 1969; Johnson and Stokes 1976; Anderton et al. 1987; Pullum and Wolf 1991); it seems likely that the fertility of other kin is also correlated. Much of this correlation results from differentials in fertility between population subgroups. For example, in the United States in

1900, fertility behavior varied dramatically according to family background: blacks, farmers, southerners, Eastern and Southern European immigrants, and much of the working class were characterized by high fertility, whereas low fertility was the norm for the bourgeoisie, families with high education, and residents of the urban Northeast (King and Ruggles 1990; Guest 1982).

Beyond the fertility correlations within population subgroups, there are also fertility correlations introduced at the level of the family. For example, both traditional and modern contraceptive knowledge was probably often transmitted by the family. To the extent that fecundity is genetic it would also contribute to fertility correlations within kin groups.

For all these reasons, in real populations some kin groups have systematically high fertility, and others have systematically low fertility. In microsimulation models, however, births are assigned to women without reference to the fertility experience of other members of the kin group. Simulations therefore show less variation between kin groups in total fertility than is found in real populations. Thus, in a real population a woman with high fertility is likely to come from a family with many children, and her children are likely to have high fertility. Such a woman typically ends up with large numbers of grandchildren. Conversely, a woman with low fertility in a real population is likely to have few grandchildren. Because no such fertility correlations exist in simulated populations, we expect such models to yield unrealistically homogeneous overall distributions of grandchildren.

As simulation studies have shown, the frequency of kin such as aunts, uncles, nephews, nieces, and grandchildren in a population is quite sensitive to the level of fertility (Ruggles 1987, Appendix D). The more distant the relationship, the greater the potential impact of fertility. For example, the number of great-grandchildren an individual has is influenced not only by the fertility of grandchil-

dren but also by the fertility of children, since the latter determines the number of grandchildren.<sup>3</sup>

Table 1 represents an attempt to assess the impact of correlated fertility on the frequency distribution of grandchildren. The figures shown were generated by the MOMSIM kinship microsimulation model (described in Ruggles 1987). All three runs shown employ demographic parameters designed to mimic the behavior of the U.S. population around 1900.<sup>4</sup> The analysis was restricted to very old women (aged 80–84) with at least one ever-married child, in order to maximize the potential for grandchildren.

The first column shows the standard output of the model: like all other models of kinship, it assumes that the fertility of the younger generation is entirely uncorrelated with that of their mothers. In the second column, I have introduced a small correlation between number of siblings and number of children ( $r^2 = .0132$ ), and in the third column, there is a larger correlation ( $r^2 = .0565$ ).<sup>5</sup> The magnitude of intergenerational fertility correlation observed in real populations typically falls within this range.<sup>6</sup>

The average number of grandchildren per woman, shown in the first row of table 1, is unaffected by the introduction of fertility correlations; the small differences shown between the three runs are the product of the random element of microsimulation. But the standard deviations—shown in the second row—are different, especially for women nearing the end of their childbearing years: the population with no correlations is significantly more homogeneous than the others. Even introducing a small correlation leads to an increase in the standard deviation of grandchildren of about 12 percent; the high correlation raises the standard deviation by 26 percent. The last two rows show the percentage of women with no grandchildren and with twenty or more grandchildren in each of the simulated populations. Analysts of kinship are

**TABLE 1**  
Measures of Grandchildren under Alternate Assumptions of  
Intergenerational Fertility Correlations for Women Aged 80–84 with  
One or More Ever-Married Children (Simulations, U.S. 1900)

Association of fertility with number of siblings	None ( $R^2 < .0001$ )	Low ( $R^2 = .0132$ )	High ( $R^2 = .0565$ )
Mean number of grandchildren	10.78	10.84	10.72
Standard deviation of grandchildren	8.32	9.32	10.50
Percentage without grandchildren	4.37	7.18	10.24
Percentage with 20+ grandchildren	12.97	15.03	16.91
<i>N</i>	2,105	2,089	2,099

often most interested in estimating the percentage of persons without any living kin of a given type. From this point of view, the most relevant statistic is the percentage of women without any grandchildren. The percentage of women with grandchildren was more than twice as great in the simulation incorporating high fertility correlation than in the simulation without any fertility correlation. If even very small correlations in fertility across generations exist in real populations, then our models will systematically overestimate the proportion of women with grandchildren.

This example illustrates the potential impact of only one facet of the Whopper Assumption. Completed fertility is not the only demographic factor that can influence the distribution of surviving grandchildren in a population. Child mortality, age at marriage, proportion who marry, widowhood, remarriage, child spacing, and age at cessation of childbearing are all potential influences. All three runs shown in table 1 assume that these factors are entirely uncorrelated across generations. A truly realistic model that incorporated correlations for all sorts of demographic behavior would doubtless reveal considerably more dramatic effects on the distribution of grandchildren.

#### The Mortality Problem

The effect of correlated mortality probabilities between members of the same kin group is similar to the effect of correlated fertility.<sup>7</sup> Kin groups with consistently high death rates tend to have few living kin of any particular type, and kin groups with low mortality tend to have high kin counts. In populations whose mortality is largely a result of contagious diseases, one expects the correlations of mortality within kin groups to be particularly high. Because simulation models do not take such correlations into account, they yield a more homogeneous overall distribution of kin than is found in a real population.

We can observe the effects of the Whopper Assumption of mortality through use of the surviving-children variable in the 1900 U.S. census. This variable can be used to calculate the distribution of number of children dying. The same distribution can be estimated by simulation. Once again, I used the MOMSIM model of kinship to simulate the distribution of surviving children in 1900. As in other simulation models, the death of each child was assigned without reference to the mortality experience of the rest of the family.

Table 2 compares the mortality experience of children in the simulated population with that of the observed population. For this example, I restricted the analysis to families in which the mother was 55 to 84 years old; younger women were excluded so that a substantial proportion of the children would be dead, and older women were excluded because internal evidence suggests a high

**TABLE 2**  
Measures of Child Survival for Women Aged 55-84  
with Five Children-Ever-Born (Simulated and  
Observed Populations, U.S. 1900)

	Simulation	Census
Percentage of children dead	31.6	31.5
Mean number of children dead	1.58	1.58
Standard deviation	1.02	1.36
Distribution of number of children surviving (percentages)		
5	14.4	26.3
4	35.0	27.4
3	32.8	21.8
2	13.9	15.0
1	3.7	6.2
0	0.1	3.4
N	6,650	1,770

degree of misreporting of children-ever-born and children surviving among women of advanced years. In addition, I limited the population to women with five children-ever-born, to avoid problems associated with correlation between parity and child mortality.<sup>8</sup> The age distribution of the simulated mothers was constrained to be identical to that in the observed population. As shown in the first two rows of table 2, the model performs well on expected number of children dying. The distribution of surviving children is given in the lower part of the table. The expected finding—that the Whopper Assumption leads to greater homogeneity of mortality experience in the simulated families than in the census families—is borne out by these figures. The percentage of women with all their children surviving is almost twice as great in the census population as in the simulated population, whereas the percentage with all their children dead is about thirty times greater in the census population than the simulation.

The implication is clear: a simulation model based on the assumption that everyone in the population has equal chances of dying at a given age allocates mortality much too evenly across kin groups. As a consequence, the distribution of living kin is more homogeneous in a demographic model than in a real population. Simulations underestimate the percentage of people without any living kin of a given type and also underestimate the percentage with many living kin of a given type.

#### The Whopper Assumption and Expected Numbers of Kin

In the preceding examples, the Whopper Assumption resulted in a reduction in the variance of kin frequencies but had no effect on the mean number of kin. In fact, as

Thomas Pullum (1987) has pointed out, errors in the distribution of kin also have the potential to distort the expected numbers of kin, depending on our point of reference for assessing kin frequencies. Specifically, failure to account for correlations in the demographic behavior of different members of the same kin group can bias downward the mean number of kin.

Consider the case of cousins. To estimate the number of cousins in a simulated population, one must generate at least three generations. The average number of cousins is then determined by counting nonsiblings who share the same grandparent. In effect, we estimate the distribution of grandchildren and then assess the number of cousins from the perspective of each grandchild. Thus, kin groups with many grandchildren count proportionately more heavily than kin groups with few grandchildren. Kin groups with many grandchildren tend to have more cousins than kin groups with few grandchildren. If we increase the number of kin groups with many grandchildren and the number of kin groups with few grandchildren, we increase the number of cousins because the kin groups with many cousins are then weighted more heavily. In other words, variance in the number of grandchildren in a population is positively associated with the expected number of cousins, even when the mean number of grandchildren is constant. Because the Whopper Assumption biases downward the variance in grandchildren, simulations of kinship tend to understate the expected number of cousins.

This pattern is illustrated in table 3, which shows measures of cousins calculated from the perspective of the same population of grandchildren simulated in table 1. Even though the mean number of grandchildren in table 1 was virtually the same across the three simulation runs, the mean number of cousins is highly sensitive to fertility correlations across generations. With increasing correlations, the expected number of cousins goes up.<sup>9</sup>

The expected number of siblings can also be biased by the Whopper Assumption (see data in table 4, which is a rearrangement of the data in table 2). By measuring from the perspective of the surviving children themselves, we are able to estimate the distribution of surviving siblings. Once again, the Whopper Assumption leads to an underestimate of the mean number of kin. The same kinds of errors can be expected to occur in the estimation of kin types such as aunts, uncles, nephews, nieces, and siblings-in-law.

### The Whopper Assumption and Multigenerational Families

Historical research on the family has prompted the development of several demographic models of kinship.<sup>10</sup> For the most part, these models are designed to estimate the maximum frequency of multigenerational families or stem families under preindustrial demographic conditions. The *stem family*, as I am using the term, denotes a family in which one child remains in the parental household after marriage, and any other children leave home and establish independent households when they get married. In cross-sectional data on household structure, stem families are usually identified by the presence of a child-in-law or grandchild of the eldest generation.

Despite the theoretical importance of the stem family, our evidence on family structure in preindustrial Western Europe indicates that such families were rare (Wachter, Hammel, and Laslett 1978; Ruggles 1987; see also Berkner 1972). The historical models of kinship have attempted to assess whether or not the low observed frequency of stem families can be ascribed to the preindustrial Western European pattern of late marriage and early death. For several reasons that I have elaborated elsewhere, different models have reached different and sometimes contradictory conclusions (Ruggles 1987, chapter 4). But all

TABLE 3  
Measures of Cousins under Alternate Assumptions of Intergenerational  
Fertility Correlations for Grandchildren of Women Aged 80-84  
(Simulations, U.S. 1900)

Association of fertility with number of siblings	None ( $R^2 < .0001$ )	Low ( $R^2 = .0132$ )	High ( $R^2 = .0565$ )
Mean number of cousins	12.10	13.80	16.03
Standard deviation of cousins	8.84	10.02	11.28
Percentage without cousins	7.94	6.18	4.92
Percentage with 20+ cousins	17.01	23.65	31.90
N	22,557	22,659	22,451

**TABLE 4**  
**Measures of Sibling Survival, Children of Women Aged 55-84**  
**with Five Children-Ever-Born (Simulated and**  
**Observed Populations, U.S. 1900)**

	Simulation	Census
Mean number of siblings	2.73	2.96
Standard deviation	0.92	1.04
Distribution of number of siblings surviving (percentages)		
4	21.1	38.4
3	40.9	32.0
2	28.8	19.1
1	8.1	8.8
0	1.1	1.8
N	4,549	1,212

these models—including my own—substantially underestimate the impact of demographic constraints on the maximum frequency of multigenerational families.

The maximum frequency of multigenerational families in a population is affected by several demographic factors. The most important of these is probably a mother's age at the birth of her children, or generation length. In natural fertility populations, generation length is largely a function of marriage age. The limited evidence available suggests that age at marriage can be significantly correlated across generations.<sup>11</sup> Generation length is critical because it counts twice; the length of time all three generations are alive simultaneously depends on generation length of both the eldest generation and that of their children.

The fact that two generation lengths are involved generates a prime situation in which violations of the Whopper Assumption can occur. Suppose we are looking at a population in which the members of some kin groups have systematically long generations while others have systematically short ones. The frequency of grandchildren is minimized when both parents and children have long generations. Therefore, we would expect members of kin groups with long generations to have exceptionally few grandchildren; many of them would have no grandchildren at all. Among the kin groups with short generations, on the other hand, there would be unusually large numbers of grandchildren.

In accordance with the Whopper Assumption, models of kinship assume that all individuals in the population have an equal probability of long generations, regardless of the behavior of the other members of their kin group. The odds of long generations occurring in both the parents' generation and the children's generation is lower in a simulated population than in a real population. Thus, the simulated population will contain a smaller proportion of people with exceptionally few grandchildren than

is found in the real population. Most important, the simulated population will contain fewer people at any moment who lacked grandchildren altogether. Because only one grandchild is needed to form a three-generation family, demographic models tend to overstate the overall potential for three-generation families.

In the case of preindustrial stem families, the problem of correlated generation lengths is compounded by the relationship between age at marriage and age at parental death. Since the preindustrial pattern of late marriage was at least partly a consequence of people's waiting to receive their inheritance before they married, there was apparently a correlation between marriage age and parental longevity.<sup>12</sup> Strictly speaking, this is not an example of the Whopper Assumption as I have defined the term; it is a correlation between two different demographic variables. Nevertheless, the consequences are similar. An association between marriage and parental death minimizes the average interval between marriage and parental deaths. It reduces the amount of time spent after parental deaths and before marriage, and it also reduces the amount of time spent after marriage and before parental deaths. If those who marry early tend to have parents who die young and those who marry late tend to have long-surviving parents, the overall overlap between the marriage of children and the death of their parents would be minimized.

An association between marriage and parental death could have important implications for the potential frequency of multigenerational families. In the most extreme case—if everyone in the population waited for their parents to die before they married—there could be no co-residence of parents with married children at all. All demographic models of multigenerational families assume that no correlation whatsoever exists between age at marriage and parental longevity; thus the models almost inevitably overestimate the potential for formation of stem families.<sup>13</sup>

Correlations of mortality within kin groups would increase the error further. As noted earlier, the assumption of uncorrelated mortality leads to overestimates of the frequency of surviving grandchildren. In addition, the maximum frequency of stem families is influenced by a correlation between the mortality of husbands and wives. The Whopper Assumption postulates that the death of one partner in a marriage is entirely unrelated to the death of the other. Although I know of no relevant data for the preindustrial period, ample evidence exists for a correlation of the spouses' longevity in the twentieth century; there is no reason to believe that this is a new phenomenon. A correlation between the mortality of husbands and wives is important because it only takes one living member of the elder generation to form a three-generation family. If the men and women who survived long enough to form stem families tended to be married to one another—and therefore concentrated in a minimal

number of families instead of randomly dispersed through the population—the maximum frequency of stem families would be reduced. To the extent that they occurred in combination, the long-surviving members of the eldest generation would be wasted.

This does not exhaust the potential list of effects of the Whopper Assumption on the expected frequency of stem families or multigenerational families. Correlations between kin in fertility, proportions never marrying, migration, childlessness, birth spacing, widowhood, remarriage, and age intervals between spouses could all potentially contribute to heterogeneity in the distributions of surviving parents, children-in-law, or grandchildren in real populations. Such heterogeneity would limit the opportunities to form multigenerational families.

Further research is necessary before we can evaluate how badly these violations of the Whopper Assumption distort the estimates of demographic models of stem and three-generation families. Nevertheless, these examples serve to underscore my comments about the Whopper Assumption in general. To the extent that exaggerated homogeneity of demographic behavior yields exaggerated homogeneity of kin frequencies, demographic models tend to underestimate the proportion of individuals without any kin of a given type. Because all techniques devised to date incorporate the Whopper Assumption, all of them tend to overestimate the proportion of the population with one or more available kin of a given type.

## Conclusion

The basic principle holds for all demographic characteristics: one would expect greater demographic homogeneity within a kin group than within a group of similar size comprising individuals selected randomly from the population. Since microsimulations of kinship ignore the correlations in demographic behavior within kin groups, they ordinarily understate the variance of kinship distributions; for many kin types, they also underestimate the expected number of kin.

It may be possible to redesign our models to avoid the Whopper Assumption. A realistic model would have to incorporate procedures for introducing correlations between demographic events occurring in the same kin group. One major problem, however, is that we lack historical data on the demographic experience of kin groups as a whole. It is unlikely that family reconstitution data can be pressed into service; because of high migration, complete demographic information is rarely available for multiple members of a reconstituted family (Ruggles 1992). More promising are genealogical studies of the sort being carried out by Fogel (1986) and Dupâquier (1986).

In general, those who design demographic models of kinship should be sensitive to the potential for systematic

error. Despite repeated calls for improved testing of our models against empirical data, such comparisons are still extremely rare. If models of kinship are to have any credibility, we must demonstrate their ability to produce moderately realistic distributions of kin in a variety of demographic contexts.

Models that incorporate no correction for the Whopper Assumption should be treated with great skepticism. At best, they can only provide lower-bound estimates of the proportion of persons with living kin of a particular type or of the proportion of households with multiple generations.<sup>14</sup> Historians in particular should be wary of claims that the demographic conditions of preindustrial Western Europe exercised little or no constraint on the potential for stem family structure.

## APPENDIX

### Method of Introducing Fertility Correlations

In the MOMSIM microsimulation model, the number of children born to each woman is first allocated by generating a single random number between 0 and 1 and then consulting a cumulative probability distribution of fertility. There are separate probability distributions for each of eight current ages and thirty-eight durations of marriage (for further details, see Ruggles 1987, 163–69, 174–77). Because a single random number was used to assign the number of children born, introducing intergenerational correlations was fairly straightforward. A random number was generated for each woman to represent her independent contribution to fertility. Then a second random number was generated to represent the degree of influence from the previous generation. For the “High” correlation run, the degree of influence was allowed to vary from 0 (no influence of previous generation) to 1 (fertility number determined entirely by previous generation). For the “Low” correlation run, the degree of influence varied from 0 to 0.5. The random number used to allocate fertility for each woman was then calculated as the average of her own random number and the random number used for the previous generation, weighted according to the degree of influence.

This procedure does not yield a flat distribution of random numbers between 0 and 1, and so an additional step was needed. An unadjusted correlation between the fertility of mothers and the fertility of daughters yields an upward bias in the fertility of the daughters. In real populations, because the average daughter has lower fertility than her mother, this bias is compensated for (see Preston 1976 for a discussion of this phenomenon). In addition to the upward bias, the fertility numbers obtained by this method will have too little variance, inasmuch as the weighted average produces a concentration of random numbers around the mean.

For both these reasons, it was necessary to adjust the random numbers to obtain a flat distribution between 0 and 1 for each group of mothers. Each version of the model was run twice: (1) to determine the cumulative probability distribution of fertility random numbers produced by the model, and (2) to apply that distribution to rescale the random numbers. The procedure was tested for each generation in each run, and the resulting random number distributions were flat.

The correlation in random numbers between generations was considerably higher than the resulting fertility correlations shown in table 1. The number of children born depends not only on the random number but also on current age and duration of marriage. The latter factors are influenced by age at marriage, widowhood, remarriage, mother's age, and age intervals between mothers and daughters, all of which were determined independently for each individual. My procedure can be viewed as a means of introducing correlations in the propensity to have children; the actual number of children born depends also on the individual demographic circumstances of each woman.

## NOTES

Funding for this research was provided by a McKnight-Land Grant Professorship and a Grant-in-Aid of Research from the University of Minnesota. Earlier versions of the article were presented at a conference on New Directions for Demographic History: A French-American Roundtable, New School for Social Research, New York City, 1986, and at Demographic Microsimulation, International Institute for Applied Systems Analysis, Budapest, 1987. In a previous article in this journal (Ruggles 1990), I listed the present article as forthcoming in a volume of essays to be published by Urban Institute Press, but that volume was subsequently canceled.

1. Several recent writers on the simulation of kinship—Wachter (1987), Wolf (1986), and Pullum and Wolf (1991)—implicitly or explicitly acknowledge the Whopper Assumption.
2. For general descriptions of simulation approaches, see Olinick (1978), Dyke and MacCluer (1973), Sheps (1969), Menken (1981), Santow (1978), and Hammersley and Handscomb (1964); also see Jaquard and Leridon (1973). Microsimulations of kinship include Howell (1979), Bartlema and Winkelbauer (1986), Wolf (1986), Rossi (1975), Gilbert and Hammel (1966), and the historical models cited in note 10. Other relevant applications of microsimulation are Orcutt (1961), and Horvitz (1971). The most important analytic and macrosimulation models of kinship are Lotka (1931), Goodman, Keyfitz, and Pullum (1974), their "addendum" (1975), Pullum (1982), and Bongaarts (1981). On the limitations of analytic techniques, see Sheps (1969), Barrett (1977), De Vos and Palloni (1984), and Ruggles (1990). Also see Pullum's comments (1982, 564).
3. In the case of fertility, the correlation of demographic events can occur not only at the level of kin groups but also at the level of individuals. Just as some kin groups may be characterized by consistently high fertility or consistently low fertility, so can individual women experience high or low fertility over the course of their childbearing years. Much of this consistency stems from factors that also operate at the level of kin groups, such as race, class, and ethnicity. As a consequence of each woman's individual circumstances—whether economic, biological, or cultural—there may be additional consistency of fertility behavior. The result is that real populations are quite heterogeneous with respect to fertility. Demographic models of fertility behavior have been reasonably successful in capturing this heterogeneity, but the models oriented to kinship analysis tend to be less sophisticated in this respect. If a model understates fertility heterogeneity at the individual level, the biases in kin variance can be even greater than those resulting from understatement of fertility heterogeneity at the level of kin groups. Theoretically, the Whopper Assumption can exist at the individual level for any demographic event that can occur multiple times to the same person, including widowhood, divorce, and remarriage as well as fertility. In populations where most people only marry once during their childbearing period, however, the effects of heterogeneity in marital behavior are probably small. I have chosen to focus on the effects of the Whopper Assumption at the level of kin groups rather than at the level of individuals, because the problem of fertility heterogeneity is already recognized by most of the analysts doing kinship microsimulation. King and Lutz (1988) provide a good general discussion of the concentration of parity distributions; for a variety of approaches used by microsimulations to introduce heterogeneity, see Crafts and Ireland (1975), Barrett (1971), Ruggles (1987), Venkatacharya (1971). The approaches used by Smith (1987), Wachter, Hammel, and Laslett (1978) and Bartlema and Winkelbauer (1986) are also of interest, although they probably produce somewhat less realistic parity distributions.
4. The demographic parameters include a median female age at marriage of 22.2, female life expectancy at birth of 48.3, and a total fertility rate of 3.79. To minimize the effects of random variation, more cases were created for the simulated population than existed in the actual population. For additional details, see Ruggles (1987, 111).
5. The method employed for introducing correlations is described in the Appendix.
6. In the 1910 public use sample of the U.S. federal census (Strong et

al. 1989), the correlation between the fertility of mothers aged 80–84 and that of their co-resident daughters was .221, which is roughly the same as my "High" correlation. For all daughters married fifteen years or longer, the correlation with their mother's fertility was .171, and for all co-resident mothers and ever-married daughters (including a high percentage of daughters married for only one or two years) the figure was .084. Of course, mothers residing with married daughters may not have been typical of the population as a whole. Although the association of fertility between successive generations has been studied extensively, comparison is problematic because the procedures used by different investigators vary widely. In a study of the British baronetage and peerage, for example, Pearson and Lee (1899) found correlations ranging from .42 to .223, depending on whether more than one daughter from each family and mothers with brief marriages were included.

My figures represent the correlation between the number of siblings and children born to women whose mothers were aged 80–84; given the advanced ages of the mothers, virtually all the daughters had completed their childbearing. A roughly comparable measure, albeit for a small homogeneous population, is provided by Huestis and Maxwell (1932), who found a correlation of .124 among 638 families sending children to the University of Oregon. In general, the correlations tend to be lower in studies where the daughters were still bearing children. Berent (1953), however, found a correlation of .187 between children and siblings among 1,482 patients of British gynecologists in 1946. By contrast, Kanter and Potter (1954) found a correlation of .09 among women in the Indianapolis Fertility Study; and Duncan et al. (1965) got virtually the same results from the United States Current Population Survey of 1962. Intergenerational fertility correlations in Canada, Israel, and Germany appear to be higher, ranging from about .15 to .32 (Pullum and Wolf 1991).

7. On family correlations in mortality, see Colver (1963) and Jaquard (1982); on class differences in mortality as a source of mortality correlations within kin groups, see Pamuk (1985).
8. In the United States of 1900, child mortality was strongly associated with parity (Smith 1983). In the developed world, high fertility has been associated with low economic status since the late nineteenth century; in natural fertility populations, however, the relationship is often the opposite because the wealthy have the resources needed for early marriage. Thus, assuming that high mortality is associated with low economic status, we might expect an inverse correlation between fertility and mortality in many populations. Microsimulations generally ignore the relationship between fertility and mortality. In general, a positive correlation between these two variables tends to counteract the Whopper Assumption, whereas a negative correlation aggravates it.
9. As in table 1, the standard deviation of kin is lowest when no fertility correlation is assumed. Unlike table 1, however, the percentage without any cousins is highest when there is no correlation, simply because the low mean number of cousins cancels out the distribution effect. Thus, depending on the perspective of measurement, the general rule that simulations overestimate the proportion of people without kin can be reversed; in this example, at least, the Whopper Assumption actually leads to an underestimate of the proportion without cousins.
10. The analytic models of this sort are shown in Coale (1965), Burch (1970), Glass (1966), Wrigley (1969); also see Wrigley (1978), Bradley and Mendels (1978), Mendels (1978). The principal historical microsimulation models of family and kin include SOCSIM, described in Wachter, Hammel, and Laslett (1978), Hammel and Wachter (1977), Hammel and Deuer (1977), and Hammel et al. (1976); CAMSIM, described in Laslett (1984) and J.E. Smith (1987); and MOMSIM, my own model (Ruggles 1986, 1987) and De Vos and Ruggles (1988); also see LeBras (1973).
11. In the 1910 public use sample (Strong et al. 1989), the correlation in age at marriage among mothers residing with married daughters was .1672, significant at the .01 level.
12. That a positive correlation existed between marriage age and parental longevity in the preindustrial West is now a commonplace; see, for example, Ohlin (1961), Wrigley (1978), Hajnal (1982), Berkner (1972), Goody, Thirsk, and Thompson (1976). Levine (1982) proposed the revisionist argument that there was an *inverse*

- relationship between marriage age and parental longevity in preindustrial England, but his analysis is badly flawed; see Ruggles (1987, 91).
13. One would expect that if parental longevity was the determining factor in age at marriage, exceptionally long generations would tend to be followed by short ones, because the parents would be older at the birth of their children. Although this could reduce the correlation of generation lengths, it would not enhance the possibilities for three-generation families as long as the inheritance rule was maintained.
  14. An exception to this generalization arises in the analysis of kin types unlikely to occur in combination. For example, the Whopper Assumption does not affect the proportion of the population with a living mother, because it is only possible to have one mother—and her chance of being alive depends entirely on her own characteristics. Problems can still arise, however, if we are interested in the characteristics of the reference person: for example, the proportion of the married population with a living mother could be biased by a correlation between marriage and parental longevity. But the main problems occur when we are analyzing combinations of kin. Analyses of orphanhood—such as those of Lotka (1931) or LeBras (1973), for example—involve calculating the simultaneous survival of mothers and fathers and are distorted by any demographic correlations between the two parents. The larger the group of kin we are analyzing, the greater is the potential for error. It is possible, for example, to have many grandchildren. If grandchildren in a given kin group behave homogeneously with respect to mortality, then the odds are greater that all will have died or that all will survive. The greater the detail with which we measure kin types, the fewer the problems that crop up as a result of assuming no systematic relationships between the characteristics of different members of the same kin group, because the probability that a category of kin will occur in combination is reduced if categories of kin are defined narrowly.
- The odds are generally high that an individual will have multiple living siblings, but the odds of having multiple widowed sisters between the ages of 45 and 49 are always low. If there is never (or almost never) more than one individual within a given category of kin, the Whopper Assumption is largely circumvented. Nonetheless, this strategy creates a new problem: if we divide kin types into narrow categories, the results of a simulation become too unwieldy to interpret. I have devised a strategy for analyzing the impact of demography on family structure that minimizes the impact of the Whopper Assumption by assuming that it results in constant proportional errors across different populations (see Ruggles 1986, 1987).
- ### 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* 24:467-80.
- Bartlema, J., and L. Winkelbauer. 1986. Modeling kinship with LISAP: A two-sex model of kin counts. WP-86-69. Laxenburg, Austria: International Institute for Applied Systems Analysis.
- Barrett, J. C. 1971. Use of a fertility simulation model to refine measurement techniques. *Demography* 8:481-90.
- . 1977. Criteria for choosing between analytical methods and simulation. In International Union for the Scientific Study of Population, *International Population Conference, Mexico, 1977* 1. Liège: IUSSP.
- Berent, J. 1953. Relationship between family sizes in two successive generations. *Milbank Memorial Fund Quarterly* 31:39-50.
- Berkner, L. K. 1972. The stem family and the developmental cycle of the peasant household: An eighteenth century Austrian example. *American Historical Review* 77:398-418.
- Bongaarts, J. 1982. Simulation of the family life cycle. In International Union for the Scientific Study of Population, *International Population Conference, Manila, 1981* 3. Liège: IUSSP.
- Bradley, B., and F. Mendels. 1978. Can the hypothesis of a nuclear family be tested empirically? *Population Studies* 32:381-94.
- Bumpass, L. L., and C. F. Westoff. 1970. *The later years of childbearing*. Princeton: Princeton University Press.
- Burch, T. K. 1970. Some demographic determinants of average household size: An analytic approach. *Demography* 7:61-70.
- Coale, A. J. 1965. Estimates of average size of household. In *Aspects of the analysis of family structure*, edited by M.J. Levy et al. Princeton: Princeton University Press.
- Collver, A. 1963. The family life cycle in India and the United States. *American Sociological Review* 28:86-96.
- Crafts, N. F. R., and N. J. Ireland. 1975. The role of simulation techniques in the theory and observation of family formation. *Population Studies* 29:75-95.
- De Vos, S., and A. Palloni. 1984. Formal methods and models for analyzing kinship and household organization. CDE Working Paper no. 84-30, Center for Demography and Ecology, University of Wisconsin-Madison.
- De Vos, S., and S. Ruggles. 1988. The demography of kinship and the life course. In *Life Span Development and Behavior* 8:258-81, edited by P. B. Baltes, D. L. Featherman, and R. M. Lerner.
- Duncan, O. D., R. Freedman, J. M. Coble, and D. P. Schlesinger. 1965. Marital fertility and size of family of orientation. *Demography* 2:508-15.
- Dupâquier, J. 1986. L'enquête genealogique sur la mobilité géographique et sociale en France depuis 1800. Paper presented at New Directions for Demographic History: A French-American Roundtable, New School for Social Research, New York City, Oct. 1986.
- Dyke, B., and MacCluer, J. W., eds. 1973. *Computer simulation in human population studies*. New York: Academic Press.
- Fogel, R. 1986. Causes and trends in nineteenth-century mortality. Paper presented at New Directions for Demographic History: A French-American Roundtable, New School for Social Research, New York City, Oct. 1986.
- Gilbert, J. P., and E. A. Hammel. 1966. Computer simulation and analysis of problems in kinship and social structure. *American Anthropologist* 68:71-93.
- Glass, D. V. 1966. Introduction. London inhabitants within the walls 1695. *London Record Society Publications* 2. Leicester: London Record Society.
- 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.
- . 1975. Addendum. *Theoretical Population Biology* 8:376-81.
- Goody, J., J. Thirsk, and E. P. Thompson, eds. 1976. *Family and inheritance: Rural society in Western Europe, 1200-1800*. Cambridge: Cambridge University Press.
- Guest, A. 1982. Fertility variation among the U.S. foreign stock population in 1900. *International Migration Review* 16:577-94.
- Hajnal, J. 1982. Two kinds of pre-industrial household formation systems. *Population and Development Review* 8:449-94.
- Hammel, E. A., et al. 1976. The SOCSIM Demographic-Sociological Microsimulation Program: Operating Manual. Research Series no. 27, Institute of International Studies, University of California, Berkeley.
- Hammel, E. A., and R. Z. Deuer. 1977. Five class programs. Research Series no. 33, Institute of International Studies, University of California, Berkeley.
- Hammel, E. A., and K. W. Wachter. 1977. Primogeniture and ultimogeniture: Their effects on stem family household frequencies. In *Population patterns in the past*, edited by R. Lee. New York: Academic Press.
- Hammersley, J. M., and D. C. Handscomb. 1964. Monte-Carlo methods. London: Methuen.
- Hendershot, G. E. 1969. Familial satisfaction, birth order, and fertility values. *Journal of Marriage and the Family* 31:27-33.
- Horvitz, D. G., et al. 1971. POPSIM, a demographic microsimulation model. International Union for the Scientific Study of Population, *International Population Conference, London, 1969* 1:95-106. Liège: IUSSP.
- Howell, N. 1979. *Demography of the Dobe !Kung*. New York: Academic Press.
- Huestis, R. R., and A. Maxwell. 1932. Does family size run in families? *Journal of Heredity* 23:77-79.
- Jaquard, A. 1982. Heritability of human longevity. In *Biological and social aspects of mortality and the length of life*, edited by S.H. Preston. Liège: Ordina Editions.
- Jaquard, A., and H. Leridon. 1973. Simulating human reproduction:



- How complicated should a model be? In *Computer simulation in human population studies*, edited by B. Dyke and J.W. MacCluer. New York: Academic Press.
- Johnson, J., and C. S. Stokes. 1976. Family size in successive generations: The effects of birth order, intergenerational change in life styles, and familial satisfaction. *Demography* 13:175-87.
- Kantner, J. F., and R. G. Potter. 1954. Social and psychological factors affecting fertility: XXIV, the relationship of family size in two successive generations. *Milbank Memorial Fund Quarterly* 32:994-1311.
- King, M., and W. Lutz. 1988. Beyond the 'average American family': Parity distributions and fertility concentration for U.S. cohorts born 1901-1935. Paper presented at meeting of the Population Association of America, New Orleans.
- King, M., and S. Ruggles. 1990. American immigration, fertility, and race suicide at the turn of the century. *Journal of Interdisciplinary History* 20:347-69.
- Laslett, P. 1984. The significance of the past in the study of aging. *Ageing and Society* 4.
- Laslett, P., and J. Harrison. 1963. Clayworth and Cogenhoe. In *Historical essays 1600-1750: Presented to David Ogg*, edited by H.E. Bell and R.L. Ollard. London: A & C Black.
- LeBras, H. 1973. Parens, grand-parents, biseaux. *Population* 28:9-38.
- Levine, D. 1982. 'For their own reasons': Individual marriage decisions and family life. *Journal of Family History* 7:255-64.
- Lotka, A.J. 1931. Orphanhood in relation to demographic factors: A study in population analysis. *Metron* 9:37-109.
- Mendels, F. F. 1978. Notes on the age of maternity, population growth, and family structure in the past. *Journal of Family History* 3:236-50.
- Menken, J. 1981. Models for the analysis of fertility change. International Union for the Scientific Study of Population, *International Population Conference, Manila, 1981* 3:435-47. Liège: IUSSP.
- Ohlin, G. 1961. Mortality, marriage, and growth in pre-industrial populations. *Population Studies* 14:190-97.
- Olinick, M. 1978. *An introduction to mathematical models in the social and life sciences*. New York: Addison Wesley.
- Orcutt, G. 1961. *Microanalysis of socioeconomic systems: A simulation study*. New York: Harper.
- Pamuk, E. 1985. Social class inequality in mortality from 1921 to 1972 in England and Wales. *Population Studies* 39:17-31.
- Pearson, K., and A. Lee. 1899. On the inheritance of fertility in mankind. *Philosophical Transactions of the Royal Society of London*, series A. 192:282-330.
- Preston, S. H. 1976. Family sizes of children and family sizes of women. *Demography* 13:105-14.
- Pullum, T. W. 1982. The eventual frequencies of kin in a stable population. *Demography* 19:549-65.
- . 1987. Some mathematical models of kinship and the family. In *Family demography: Methods and their application*, edited by J. Bongaarts, T. K. Burch, and K. W. Wachter. Oxford: Oxford University Press.
- Pullum, T. W., and D. A. Wolf. 1991. Correlations between frequencies of kin. *Demography* 28:391-409.
- Rossi, F. 1975. Un modello di simulazione per lo studio del ciclo di vita della famiglia. *Genus* 31:34-94.
- Ruggles, S. 1986. Availability of kin and the demography of historical family structure. *Historical Methods* 19:93-102.
- . 1987. Prolonged connections: The rise of the extended family in nineteenth century England and America. Madison: University of Wisconsin Press.
- . 1990. Family demography and family history: Problems and prospects. *Historical Methods* 23:22-30.
- . 1992. Migration, marriage, and mortality: Correcting sources of bias in English family reconstitutions. *Population Studies* 46:507-22.
- Santow, G. 1978. *A simulation approach to the study of human fertility*. Leiden: Nijhoff.
- Sheps, M. 1969. Simulation methods and the use of models in fertility analysis. International Union for the Scientific Study of Population, *International Population Conference, London, 1969* 1:53-64. Liège: IUSSP.
- Smith, D. S. 1983. Differential mortality in the United States before 1900. *Journal of Interdisciplinary History* 8:735-59.
- Smith, J. E. 1987. The computer simulation of kin sets and kin counts. In *Family demography: Methods and their application*, edited by J. Bongaarts, T. K. Burch, and K. W. Wachter. Oxford: Oxford University Press.
- Strong, M., S. H. Preston, A. R. Miller, M. Hereward, H. R. Lentzner, J. R. Seaman, and H. C. Williams. 1989. Users guide: Public use sample. 1910 United States Census of Population. Philadelphia: Population Studies Center, University of Pennsylvania.
- Venkatacharya, K. 1971. A Monte-Carlo model for the study of human fertility under varying fecundity. *Social Biology* 18:106-15.
- Wachter, K. W. 1987. Microsimulation of household cycles. In *Family demography: Methods and their application*, edited by J. Bongaarts, T.K. Burch, and K.W. Wachter. Oxford: Oxford University Press.
- Wachter, K. W., E. A. Hammel, and P. Laslett. 1978. *Statistical studies of historical social structure*. New York: Academic Press.
- Wolf, D. A. 1986. Kinship and family support in aging societies. WP-86-81. Laxenburg, Austria: International Institute for Applied Systems Analysis.
- Wrigley, E. A. 1969. *Population in history*. New York: McGraw-Hill.
- . 1978. Fertility strategy for the individual and the group. In *Historical studies of changing fertility*, edited by C. Tilly. Princeton: Princeton University Press.