Testing the Validity of Kinship Microsimulation: An Update

KENNETH W. WACHTER

DEBRA BLACKWELL

EUGENE A. HAMMEL

University of California, Berkeley

20 February 1998

ABSTRACT

Computer microsimulation is capable of generating detailed reconstructions and forecasts of the numbers of living kin that members of a population have. The available detail, specificity, and consistency of vital rates for the past and the accuracy of vital rates projected for the future necessarily limit the accuracy of the kinship estimates that can be produced.

This study updates the validity tests published in the *Journal of Mathematical and Computer Modeling.* volume 26, pages 89-104, in 1997. The principal goal is an external validity test of Reeves' (1982) reconstruction and forecast of U.S. kin counts with the SOCSIM microsimulation program. The external standard is provided by estimates from the 1987-88 wave of the National Survey of Families and Households, the first detailed information on numbers and ages of kin for

the United States. We compare forecasts with estimates for average numbers of living grandchildren, living full siblings, and living half siblings as functions of age.

We find remarkably close agreement for some of the predictions, along with several instances of systematic discrepancies. The discrepancies most likely stem from errors in the forecasts rather than in the survey estimates. Interacting rather than isolable errors appear to be responsible. Our validity tests provide a basis for qualified faith in the effectiveness of kinship microsimulation.

KEY WORDS: Kinship; Forecasting; Aging; Microsimulation

----

1. Demographic Microsimulation

Microsimulation of kinship structure is perhaps the most ambitious and demanding application of computers in demography to date. It incorporates processes from all the main areas of demographic studies -- fertility, mortality, marriage formation and dissolution, migration -- and requires rate sets for all of them from the past and predictions for all of them for the future. Measuring kinship is a matter of tracking links between individuals, not just tracking the changing statuses of individuals in isolation. The focus on networks not only complicates the computer representation of the evolving system but also amplifies the importance of interactions involving the underlying processes.

From a computing standpoint, the scale of demographic kinship simulation is still small in comparison with policy simulation models like TRIM, DYNASIM, HITSM, MATH, or PRISM (see Citro and Hanushek (1991:290-298)), which are designed to study impacts of taxation, welfare, and health programs.   Those models include and update hundreds of status variables for individuals and have elaborate modules for establishing program eligibilities and various transition rates. On the other hand, that type of model typically takes for its starting current population the set of cases from an actual sample survey and draws transition rates from responses to questions in the same survey. In that respect, demographic kinship simulation is far more difficult. There are no national-level sample surveys with sufficiently complete enumeration of the ages of relevant kin to serve as a starting state for kinship simulation. The starting current population with its kinship links must be laboriously reconstructed with prior simulations starting with a population of unrelated individuals and applying historical

vital rates. The historical vital rates come not from a single survey but from a multiplicity of mutually inconsistent sources.

The field of social science microsimulation has been reviewed and assessed by a panel of the National Research Council in two volumes edited by Citro and Hanushek (1991). Demographic modelling and kinship forecasting are discussed on pages 286 to 289 of Volume I. With respect to the field as a whole, the Panel centered in on the problem of model validation and the lack of attempts at external validation of predictions from microsimulations. As computing technology has become more powerful, models have become more elaborate and less scrutable, and doubts about their believability are natural. The Panel's literature review conducted by M. L. Cohen (1991) concluded that "(1) very few microsimulation models have had *any* validation analysis ... (2) validation often leads to model improvement, and (3) there has probably never been a detailed analysis of the size and origin of the errors of a microsimulation model's outputs ... " (1991:272). The Panel called for new efforts at external validation of microsimulation forecasts.

We responded to the NRC Panel's call in an article published in the *Journal of Mathematical and Computer Modeling.* (volume 26, pages 89-104, in 1997. Here we review and revisit the arguments and findings presented there. Our study takes advantage of the fact that the first microsimulation forecasts of kin counts for the United States were made around 1980, nearly a decade before the first results from the National Survey of Families and Households (NSFH) became available. The NSFH is the first national U.S. survey with detailed kinship data for any period.   Thus the first kinship forecasts were genuine shots in the dark -- as much for reconstructing the

kinship links of 1980 as for predicting them forward in time -- and comparison of the forecasts with the NSFH estimates provides a meaningful test of the accuracy of the method.

The first microsimulation forecasts of kin counts for the United States are found in Hammel, McDaniel, and Wachter (1981) "The Kin of the Aged in the Year 2000" (which we shall refer to as KOA) and in the Ph. D. thesis of J. Reeves (1982) later summarized in Reeves (1987). Both these forecasts used the SOCSIM demographic microsimulation routines in the release described in the manual by Hammel *et al.* (1976). Kin count calculations have also been made with a number of other models, but all others in the pre-NSFH period assumed sets of vital rates unchanging over time and thus produced stable-population-theory values rather than forecasts attuned to the U.S. case with its history of changing vital rates, 1950s' baby boom, and 1970s' baby lull.

Naturally, microsimulation has advanced since 1980, and the successes and failures of the 1980 round of forecasts do not necessarily indicate what degree of accuracy can be achieved by the methods of today. However, today's reconstructions of the current population's kinship structure are all being made with the hindsight afforded by the NSFH results, so success at reconstruction from today onward no longer provides any fully blind test. Success at prediction can still be tested, but kin counts change relatively slowly, so a trenchant test of today's predictions will probably have to wait fifteen or twenty years.

In principle, perfect knowledge of the historical vital rates up to 1980 should allow perfect reconstruction of the kinship networks of the 1980 population, and the

only remaining uncertainty in forecasts for the 1987-1988 epoch of the NSFH would derive from uncertain predictions in 1980 of vital rates for 1980 to 1988. In practice, imperfections in our knowledge of historical vital rates abound, and these limit the achievable accuracy of kinship reconstruction just as much as uncertainty about the future imposes limits on the achievable accuracy of predictions.

Our imperfect knowledge involves problems of detail and specificity, heterogeneity, and consistency. Kinship forecasting demands a level of detail for vital rates that is often missing in published tables. For example, to estimate distributions of numbers of siblings the fertility rates need at least to be specific to marital status and parity, that is, number of children already borne. The National Center for Health Statistics does not now publish such rates for the United States. So these rates need to be estimated for better or worse from a *pot pourri* of other sources. Breakdowns by race are missing for crucial data.   Even when official or quasi-official numbers exist, the heterogeneity of the population's vital rates within the tabulated categories constrains the accuracy of kinship forecasts. Finally, fertility, marriage, and mortality rates derive from disparate sources with different and poorly known structures of non-sampling error. Without *ad hoc* tuning, they are not sufficiently consistent with each other to produce the historically observed growth rates and age pyramids when they are put to work in tandem within a microsimulation.

These considerations limit the accuracy of kinship forecasts, and they limit it to a hitherto unknown extent.   Our philosophy in this paper is to study the accuracy of an actually existing forecast, granting that intrinsic shortcomings of data and method combine with the limited stamina of actually existing human researchers and actually

existing computing power to produce forecasts with some -- in this case measurable -- level of error.   We follow the example of Stoto (1983) who shed light on the margins of error assignable to population forecasts by reviewing the levels of error that actual forecasters have made over time.

A comprehensive recent review of studies of kin availability including micro-simulation is given by Wolf (1994).   Part V of the volume *Family Demography* edited by Bongaarts, Burch, and Wachter (1987) presents leading practical approaches to kinship estimation.   DeVos and Palloni (1989) survey the field with an emphasis on formal models and implementations of stable population theory.   In current termi-nology, microsimulation is contrasted with macrosimulation.   In microsimulation the fundamental unit is the individual, and characteristics of individuals and links between them are updated, usually by Monte Carlo.   In macrosimulation, the fundamental unit is the group, and sizes of groups are updated, usually by matrix multiplication.   Micro-simulation studies with stable-theory assumptions (unchanging vital rates) include LeBras (1973), Menken (1985), Smith (1987), Wolf (1988) and Zhao (1994).   The most widely-cited macrosimulations of kinship are Zeng Yi (1988) and Bartlema (1989), who incorporates a microsimulation component into his macrosimulations.

Microsimulations can be checked against macrosimulations, and *vice versa.* This is a way of testing the consistency of the models with their own stated assumptions and of identifying mistakes or idiosyncrasies. Typically, such tests are carried out under the simplified conditions of unchanging vital rates.   In an extensive test of this kind, Jiang Lin (1992:34-47) compared microsimulations using SOCSIM with the macrosimulations of Zeng Yi (1988).   Vital rates (taken from cross-sectional rates for

China in 1981) were chosen in the microsimulation to be as nearly identical as possible to those for the macrosimulation. Lin found very close agreement for a wide range of kinship predictions. Such studies are revealing, but they are naturally no substitute for direct comparisons with data on real people.

Kinship microsimulations with changing demographic rates, as we have said, go back to Hammel, McDaniel, and Wachter's (1981) "Kin of the Aged in the Year 2000". The computer implementation, the SOCSIM demographic microsimulation, had been developed originally for studies of historical household structure presented in Wachter *et al.* (1978). Reeves (1982) and (1987) and Wachter (1995) use the same programs for more refined estimates of U.S. kin counts, while Hammel *et al.* (1991), and Jiang Lin (1994) take up applications to China.

2. The Computer Model

Simulations in demography differ from simulations in the physical and mathematical sciences in that the underlying probabilistic processes are quite simple but the variety of transitions and interconnections are quite complex. The SOCSIM program is a descendant of the POPSIM programs described by Rao *et al.* (1973). Its distant cousins include a pioneering program of Peter Kunstadter (1963), the REPSIM models described by Clague and Ridley (1973), and Howell's (1979) AMBUSH. A close relation is Smith's (1987) CAMSIM. SOCSIM was designed by E. A. Hammel and K. W. Wachter and implemented, in successive stages, by D. Hutchinson, H. Doughty, R. Deuel, C. Mason, and M. Feitel. The version of SOCSIM in use for the Reeves forecasts under study in the present paper was a version written in FORTRAN

running on CDC 7600 hardware, documented in Hammel *et al.* (1976). The version in present use is a Pascal conversion and extension which runs under UNIX on Sun SPARC workstations.

Demographic microsimulations are divided into "closed" and "open" models. In a closed model, marriage partners are found from among individuals already present with life histories and kin within the simulated population. In an open model, marriage partners are created at marriage, with characteristics conditional on those of the marrying population member. Closed models are much more complicated, but they permit bilateral kinship reckoning to arbitrary depth. SOCSIM is a closed model.

Demographers new to microsimulation sometimes resist the idea of closed simulation models on the grounds that real people may not be choosing their partners from any specified closed group nor even necessarily from within any limited national or regional geographical boundary.   Although closed models can be used to study closed small mating groups (cf. the study by Hammel *et al.* (1979) of incest taboos among hunters and gatherers), their use here is quite different. With populations of thousands, close-cousin pairings are too rare to matter, and for practical purposes the pool of partners is as random as in an open model.   The essential feature of closed models is that every partner has his or her own network of kin, so that kinship linkages are not truncated at marriage nodes, as they are in open models.

The internal data structure within SOCSIM consists of three flat arrays, the population roster P, the marriage roster M, and the socio-economic supplement to the population roster X. The P-array contains a row for each person who is alive at any point during the simulation. When a person is born during the simulation, a new row

is added to the P-array. The row remains after the person's death, so that the person's kin and characteristics can be retrieved. Columns of the P-array include the person's identification number, sex, the population group to which a person belongs, dates of birth and death, pointers to a person's eldest ever-born child, a person's next eldest siblings through the father and through the mother, and the person's most recent marriage in the M-array. Individual-specific multipliers for fertility are also included. The M-array contains a row for every marriage union established during the simulation, with a marriage identification number, date of marriage, date and cause of any marriage termination, pointers to the husband and wife of the marriage, and to the next previous marriage if any of husband and of wife, The X-array has the same person identifiers as the P-array and includes variables for specific applications such as wealth, income, and education.

Time is measured in integral months. The timespan of the simulation is divided into segments, generally of 5 or 10 years each, during which a given set of vital rates stored in input rate tables hold force. Transitions for members of the population include the initiation of a marriage search, childbirth, migration, and death. These transitions are called "events" and the type and waiting time for a person's next event are determined in a stochastic event competition whenever a previously-scheduled event is executed. An event competition for a person also takes place when the person is affected by death of a spouse, migration, or other events previously scheduled for other related individuals.

The model for the event competition is a standard model of competing risks. A candidate waiting time is generated for each event for which a person is eligible. The

event with the shortest waiting time wins the competition and is scheduled. The losing waiting times are discarded. The distribution of each waiting time is piecewise exponential, meaning that the hazards are piecewise constant.

Specifically, let $\mu(a)$ be the piecewise-constant hazard of a given event, say death, as a function of age a, for a person with a particular combination of characteristics, say never-married women in Group 1. Let ξ be the person's individual multiplier for that event from the P-array, a life-long person-specific risk-enhancement or frailty which is introduced to model heterogeneity in risks within the subgroups of the population. Let α be the person's age at the time of the event competition. Let

$$H(x) = (\int_{\alpha}^{x} \mu(a) \, da \,)$$

A uniform pseudo-random number U is generated by a linear congruential random number generator with period 2^{16}. Then the waiting time is given by

$$T = H^{-1} \, (\, (\, -1 \, / \, \xi \,) \log U \,)$$

with the inverse of the piece-wise linear function H evaluated by successive decrement followed by linear interpolation.

Events, once scheduled, are listed in an event queue. In the simulation, control proceeds month by month. In each month events scheduled for that month are executed (in random order). Characteristics in the P-array and X-array are updated and pointers are reset in response to the events. For instance, when a child is born, a new row is added to the P-array, characteristics are generated by various random processes, the last-born-child pointers for the parents are reset, the next-elder sibling pointers for the next-younger siblings are reset, and new events are scheduled with event competitions for mother and for child. Lengthy code is involved, but the logic is mainly an

expression of common sense.

The most elaborate part of the simulation is the marriage module. One of the competing events for people eligible to marry is the initiation of a marriage search. Those previously scheduled for marriage who have not yet been paired with a suitable partner are kept in a marriage queue. A person initiating a marriage search, searches in random order through the opposite-sex members of the marriage queue, checking each candidate against a set of criteria for acceptable age range and group member-ship. (The population is divided, if desired, into groups, which in various applications may be households, lineages, ethnic groups, social classes, etc.) In the simulations of the 1980s, the population was not divided into groups. The strictness of the marriage criteria varied according to a person's "choosiness" index. The first candidate not pre-cluded by incest taboos who meets all criteria is then selected as the marriage partner. If no one meets the criteria, the person joins the marriage queue and becomes eligible to be chosen by newly arriving searchers of the opposite sex. At the end of each seg-ment, all members of the marriage queue are removed, have new events scheduled for them, and become eligible to initiate new searches.

With populations of a thousand or more with historically realistic rates, searchers find partners quickly. In the jargon of economics, "the marriage market clears." A check on the clearance time in five simulations with populations of Reeves' size and with current best estimates of U.S. rates from 1900 to 1990 showed a mean waiting time per marriage of 5 months.

There are many special features of the model which we mention briefly. Waiting times to next birth for women are adjusted conditional on the timing and survival of

the preceding child. Fertility is parity-specific (it depends on the number of preceding births), and is also specific to woman's age (but not partner's age) and marital status. Fertility is heterogeneous; each woman has her own fertility multiplier raising or lowering her probabilities of childbirth at all ages, and a component of this multiplier can be inherited from her mother, introducing mother-daughter correlations in fertility. Divorce depends on duration of marriage. Migration is implemented as a change of group membership. Some groups are defined as groups "abroad", others as groups "at home". There are endogenous feedback mechanisms for adjusting fertility and migration to targets. Some of these capabilities, specifically migration, inherited fertility, and feedback, were not in use in the kinship simulations of the 1980s and are innovations in more recent forecasts.

3. Reeves' Simulations and the NSFH

Jaxk Reeves' forecasts of U.S. kinship were completed in June of 1981 and represent the state of the art of kinship forecasting in the United States in the era before the availability of survey-based estimates of current kin.   The chief advance in Reeves' work over the earlier KOA forecasts of Hammel *et al.* (1978) is the modelling of divorce and remarriage. As with KOA, Reeves used parity-specific fertility rates with birth-spacing adjustments and heterogeneous fertility, and a marriage market based on preferences for partner's age and heterogeneous choosiness. Childbearing rates were independent across generations. There was no non-marital childbearing in his simulations.   Marital fertility rates were augmented in proportion to non-marital fertility.

Reeves did not attempt to forecast kin for the African-American portion of the population, whose vital rates for the earlier 1900s proved difficult to reconstruct. His simulations did not incorporate migration. In deciding on an identifiable subpopulation in the NSFH against which to compare Reeves' simulated population, we choose non-Hispanic whites. This choice avoids discrepancies arising from the absence in Reeve's simulations of the immigrant Hispanic population of the 1970s and 1980s, although the resident Hispanic population does still contribute to the vital rates. Incorporation of migration is one of the advances made in kinship microsimulation since Reeves' work.

Reeves assembled and tabulated counts of kin at twenty-year intervals from 1900 to 2020. It is necessary to interpolate between Reeves' estimates for 1980 and for 2000 to obtain numbers that apply to the 1987-1988 epoch of the first wave of the NSFH. We interpolate linearly between 1980 and 2000 period values holding age constant. For our graphs, this method gives results that are practically identical to those obtained by interpolating linearly between age groups holding cohort constant.

The computer tapes and input rate sets from the original microsimulations are not extant. The rate sets for the KOA studies do survive but those of Reeves do not, so we cannot rerun his simulations and reconstruct outputs missing from printed materials. Our comparisons make use of the tabulated values in Reeves (1982) and (1987) along with three thick computer printouts which have been preserved and which Professor Reeves generously supplied. These printouts show for each simulation run for each vigintennium in cross section the number of people in each age and sex group and the mean numbers and standard deviations of their relatives of many types.

Demographic microsimulation creates an artificial populations which is designed to be equivalent to a random sample of family lines from the real population. Each individual in the real population has in principle the same probability of inclusion in the notional sample as every other, but the individuals are not sampled independently. Individuals from the same genealogical family are sampled together.

The NSFH is a national weighted probability sample of households, not of families or individuals. We are using the first wave, with interviews conducted between March 1987 and May 1988, described in Sweet, Bumpass, and Call (1988). The institutionalized elderly population is excluded from the sample, though some younger people living institutionally in barracks and college dorms are included. The survey oversampled certain groups including minority members, single-parent families, and families with stepchildren. The kinship statistics for our comparisons are those for the primary respondent. Once a household was selected into the NSFH, the primary respondent was chosen at random from all household members above the age of 19 (with a few exceptions).

Individuals in the real population thus have rather different probabilities of inclusion as primary respondent in the NSFH depending on age, coresidence, and other factors. This fact would make comparisons of statistics like age pyramid counts quite complicated. But the estimates we compare, like average living grandchildren for females of a given age, are in effect estimates conditional on respondent's age and sex, so the different sampling principles underlying the simulation and the survey are of less consequence. It may be, for instance, that elderly grandmothers living in households are underrepresented among primary respondents due to their own or their

family's reluctance. But if the numbers of grandchildren for those who did respond are not systematically different from the numbers for those who did not, this effect will not bias our results.

If we were looking at kin with high rates of coresidence, a weighting correction for coresidence would be essential for the NSFH. For example, if the NSFH included teenage primary respondents (as it does not) and if we were comparing sibling counts for teenagers, who generally coreside in households, then a teenager with many brothers and sisters would have a smaller probability of being selected as a primary respondent in the NSFH than of being present and counted in the simulated population. Reweighting would be a large issue. For the estimates for adults on which we focus, coresidence as it bears on sample inclusion is a comparatively small effect.

Reeves simulated a single reconstruction of the 1980 population and kinship universe and then investigated nine alternative demographic scenarios for the vital rates of 1980 to 2020. His reconstruction for 1980 was generated from a simulated 1900 population with stylized (stable theory) kinship links, using detailed historical rates for 1900 to 1980.

The estimates for 2020 which we evaluate are the estimates for Reeves' medium scenario called PMFMMMD, an acronym which stands for "population with medium fertility, medium mortality, and medium divorce". The simulated 1980 population numbered 2314. Five independent replications of the simulations were run. The combined-run total of people as of 1980, 11570, is roughly comparable with the number of white non-Hispanic primary respondents in the 1987-88 NSFH, 9416. Needless to say, Reeves' sample sizes and numbers of replications are far smaller than

are feasible today. For their time, they were impressive, and they entail sampling standard errors small enough in most cases to permit meaningful comparisons.

4. Comparisons

Our aim is to assess the success of the kinship forecasts of the early 1980s at matching the levels ascertained by the NSFH in 1987-88.   The NSFH is not a perfect standard, and in cases of discrepancies we shall consider whether the forecasts or the NSFH estimates are more likely to be in error. Of course the forecasts and the NSFH estimates may agree and yet both be wrong, but we have no data with which to test for that possibility.

We choose kinship statistics for our comparisons whose dependence on lateral links or multiple generations made them harder to guess or piece together from other surveys before the release of the NSFH. The three we choose are

A. The expected number of living grandchildren;

B. The expected number of living full siblings;

C. The expected number of living half-siblings.

The grandchildren studied here include only full biological grandchildren. The simulations generated step-grandchildren, but the available computer printouts do not tabulate them.   Full siblings share both biological parents. Half-siblings share one biological parent. Step siblings each have a biological parent who is the spouse of a biological parent of the other. Step-siblings were not tabulated. Regrettably, the printouts do not show distributions of numbers of grandchildren or siblings. We would

dearly like to compare whole distributions, but the printouts restrict us to average levels. Despite these limitations, the existing information reveals some interesting successes and failures.

Figure 1, 2, and 3 all have the same format. Each shows the average number of kin of a specified type as a function of age, in five-year age groups centered on the mid-point age. Reeves' predictions interpolated to January 1988 are shown as dotted lines. Estimates from the NSFH of 1987-1988 are shown as solid lines. A normal-theory 95% confidence interval for the forecasts for each age group separately is shown as a dotted vertical line. A similar confidence interval for each NSFH estimate is shown as a pair of solid vertical lines, positioned left and right of the plotted point so that they do not overlap and obscure the dotted line. The computer population can be regarded as a sample of genealogical family lines from the real population. Appropriate cross-sectional sampling variability is maintained from decade to decade by the Monte Carlo features of the simulation. Thus the confidence intervals for the forecasts have the same kind of sampling-theory interpretation as the confidence intervals for the NSFH estimates.

An intuitive visual criterion is to view a difference as significant when the intervals fail to overlap. Such a test is more stringent than a 5% significance test, even though the intervals are 95% confidence intervals, because the standard error of the difference is the root sum of squares of the standard errors rather than their simple sum. We must also remember that, formally speaking, the intervals are suited only for isolated single comparisons and not for multiple comparions. Practically speaking, however, the visual criterion is adequate, since the discrepancies that do occur are

obvious without any subtle statistical testing.

When interpreting the figures, it is essential to bear in mind that "up in age" means "back in time of birth". As we move to the right on each graph, we are viewing the relatives of cohorts born earlier and earlier in the century. U.S. vital rates changed rapidly in the course of the Twentieth Century, and these changes are reflected in the cross-sectional age profiles in the graphs. Anyone used to looking at kinship estimates from stable population theory, with its unchanging vital rates and its purely age-driven profiles, needs to be alert to the salience of cohort effects in our simulations and in true kinship counts.

Figure 1 shows the average number of living grandchildren for women as a function of age. The forecasts are remarkably close to the survey values up to age 70. Beyond age 70, the forecasts and the NSFH estimates part company.

The close agreement at younger ages is undoubtedly a triumph for microsimulation. The forecasts and the NSFH findings are essentially indistinguishable. The NSFH represents the state of existing knowledge about the levels of these kin in 1988. Reeves was able, in June of 1981, to predict these levels with the same degree of precision as the NSFH.

That is the good news. There is also bad news. Above age 70, something is clearly amiss. The forecasts rise toward 7 while the estimates level out between 4 and 5. The cohorts with the discrepancy were those born around World War I. Both the simulations and the NSFH have relatively small sample sizes for the older age groups and the confidence intervals broaden, but the discrepancies are too large and systematic to be ascribed to random error.

The later ages at the right of the graph are ages when people's children have mostly completed their own childbearing, so the pure age-driven effects that one sees in stable population theory should not be operating. The issue here is an issue of cohort-to-cohort changes. Comparing cohorts born just before and just after 1915, the cohorts' own fertility has an upward trend but the cohorts' children's fertility has a downward trend, and it requires some thought and analysis to be sure which trend should dominate.

Could the problem lie with the NSFH values? The NSFH does exclude the older institutionalized population, which the simulations include. But one would not, on the face of it, expect those with larger kinsets to be more likely to be institutionalized. It is true that averages computed from the NSFH without use of sample weights are half-a-grandchild closer to the forecasts for women 70 to 80 than the weighted averages. Conceivably the high-weight (sparsely sampled) cases are problematic in some unknown way. Only 69% of women 70 to 80 in the NSFH have any living (biological) grandchildren, and that seems intuitively like a figure that could be too low. For Figure 1, the 3% of cases who gave no response to one or more of the grandchild questions on the survey have been coded as zeros. Some 53 of them are women over age 70. Excluding these cases, however, raises the 70 to 75 average by only 0.02 and depresses the 75 to 80 average by 0.07, so the treatment of non-response is not an explanation for the discrepancy. It could of course be that there are respondents who do not know of all their grandchildren or who underreport them, although one could also imagine respondents exaggerating the number of their grandchildren or reporting step-grandchildren as biological grandchildren. All in all, it seems that the NSFH numbers could be underestimates, but it seems hard to ascribe the lion's share of the

discrepancy to problems with the NSFH estimates.

Neither the KOA forecasts that preceded Reeves nor the recent microsimulations of Wachter (1995) show the rise toward levels above 6 in average biological grandchildren after age 70 that Reeves' forecasts do. The recent simulations do show total grandchildren including step and biological grandchildren near the levels found in Figure 1 for biological grandchildren alone. But examination of Reeves' computer code confirms that step-grandchildren were excluded. On the whole, we believe that the forecasts are in error over the age of 70.

Can we identify a likely source for such a specific forecasting error? Selective mortality of the elderly with respect to numbers of children and grandchildren (after controlling for race and Hispanic status) is very unlikely to be strong enough to drive a two-child difference in averages. Mortality of children or grandchildren is too low to matter. Some of the small differences at ages below 70 may be produced by correlations in the timing of fertility between parents and their children (as a referee has suggested). Reeves did not include any such correlations in his simulations. But the discrepancies for late ages and early cohorts are too pronounced to be ascribed to such a source. Thus, errors in simulations of family sizes for relevant cohorts are to be suspected.

Graphs of living children by age from the simulations, not given here, do show a modest excess over NSFH estimates for women over 70, (2.35 vs. 2.15) and an excess for ages 30 to 45. Those excesses are at more or less the proper ages to contribute an excess of grandchildren for those over 70, but the size of the excesses is too small to be the sole explanation. Excesses in living children of the same size but in

the opposite direction for those 45 to 70 do not destroy the agreement between fore-casts and estimates for grandchildren for those under 70.

Heterogeneity in fecundability may give a few large families undue influence over the means. (We return to this issue in our discussion of living siblings.) Since Reeves' printouts do not give distributions, we cannot compare them to the distribu-tions from the NSFH. In the end, this isolated discrepancy at older ages side by side with the close agreement at younger ages remains something of a mystery, and a cau-tionary instance.

We turn now to living full siblings, a comparison shown in Figure 2. The counts include siblings of both males and females. The solid line for estimates from the NSFH only extends between the main ages of primary respondents, whereas the dotted line for the simulations extends over all age groups. Agreement is good for the youngest adult ages, but the forecasts and the observations diverge after about age 35. The forecasts overshoot the observations by about half a sibling till age 65 and then drop off more rapidly than the observations.

The premature dropoff in living siblings after age 65 is easily explained. Reeves' middle scenario for mortality from 1980 to 1990, like most mortality predic-tions of the later 1970s, were unduly pessimistic and missed the substantial improve-ments in mortality at older ages that actually took place. His ten-year survivorship from age 70 to age 80 for 1980 to 1990 was 0.63 compared to 0.81 for the U.S. white lifetable of 1985 from the *Statistical Abstract* of the U.S. Bureau of the Census (1992:78). Reeves also generated an optimistic scenario (whose outputs we do not have) which presumably tracks the NSFH estimates for living siblings more closely at

older ages.

The middle-age discrepancy is more complex. While underreporting of siblings in the NSFH is a possible source, it seems unlikely, since quite detailed questions about siblings were asked, with little non-response. The pattern of errors seems to arise not from a single cause but from several reinforcing inaccuracies. Reeves (1982: 257) himself expected that "the average number of siblings (rather than half siblings) will be overestimated for a few children..." because in his simulations out-of-wedlock births were credited to married women. This feature should have suppressed half-siblings, but, as we shall see from Figure 3, it turns out that half-siblings are also too numerous in the forecasts in these age groups.

In one way or another, modest overestimates of the sizes of families into which the babies of the inter-war years were born are producing substantial discrepancies in sibling counts. One plausible culprit is the provision for heterogeneous fecundability, which may have been too extreme and generated too long a tail in the distribution of family size. Sibling counts are particularly sensitive to the proportion of large families, since a large family contributes more potential respondents than a small one. It must be borne in mind that the refined estimates of heterogeneity in birth interval components of Wood and Weinstein (1988) were not available when these microsimulations were designed.

Our third comparison pertains to living half siblings, shown in Figure 3.   These are called "step-siblings" by Reeves (1982), but his computer code indicates that biological children of either but not both one's parents are the siblings tabulated. In both the microsimulations and the survey, the levels are smaller and the error bars are

wider relative to the levels. There are few statistically significant differences to be seen. A pattern of forecasts higher than estimates for middle-aged respondents is, however, visible in the curves. That is the same kind of pattern as we have seen for full siblings. It suggests that overestimation of family sizes in the simulations affects both full and half-siblings.   That, in turn, suggests that the modeling of divorce and remarriage in the simulations has been fairly successful, since such errors ought to have different or countervailing effects on full and half siblings.

We turn now to a discussion of some doubts that might arise concerning the successes that Reeves' kinship predictions do achieve.   Predictions can be right by accident as well as by design. Many complications of real demographic processes are omitted from the computer model. These omissions might be causing large errors in different directions that fortuitously cancel each other out.   Sensitivity testing of the sizes of omitted effects is an important concomitant of evaluation.

One omitted feature, emphasized by a referee, is correlated fertility between mothers and daughters. Reeves' simulations assume independence between genera-tions.   The SOCSIM routines are, however, programmed to allow such correlations if desired. Although Reeves' simulations cannot be duplicated exactly, broadly similar simulations can be run with and without correlated fertility. In this way, the likely size of the effect of this omission from the model can be gauged. Figure 4 shows results from such an experiment.

The dependent variable in Figure 4 is the same as in Figure 1, average living grandchildren by age of mother. Here, both biological grandchildren and step-grandchildren are shown in separate sets of curves.   The solid lines show the result of

one baseline simulation with no correlation in fertility. The four dotted lines show the results of four replications of simulations with the same positive correlation in fertility, differing from each other by Monte-Carlo variation. The predictions pertain to 1990. The circles show the observed levels for the 1987-1988 NSFH. Error bars are not shown here, but they are essentially the same (for biological grandchildren) as those in Figure 1, and none of the differences are large enough to be statistically significant at the 5% level.

The correlation in mother-daughter fertility for the simulations is, on average, 0.061, chosen to match the mean correlation for U.S. women aged 55 to 80 of 0.068 from the 1984 supplement on aging to the National Health Interview Survey given by Pullum and Wolf (1991:406). Since Pullum's and Wolf's correlations are not controlled for race, they must overstate the correlations for the non-black populations being simulated. Presumably, these correlations are not mainly family-level associations but rather effects of the decision to aggregate subpopulations with differing typical fertility levels. Our comparison probably overstates the effects on kin counts of correlated fertility *per se,* because the input rates for the simulations for the later years have not been scaled down to counteract the intrinsic tendency of "inherited" fertility levels to raise the average fertility levels of later generations. Thus an effect of higher effective fertility in later generations is being confounded with the direct effect of mother-daughter correlations. Either the higher effective fertility or the correlations themselves do seem to improve the match between predictions and observations, but not by statistically significant amounts. As far as average numbers of living grandchildren are concerned, intergenerational correlations in fertility appear to be of rather minor importance and are not a reason for doubting Reeves' successes, such as

they are.

The simulations for Figure 4 incorporate refinements that were not present in Reeves' simulations. Provisions for migration bring in a foreign-born population raised (on average) under higher-fertility conditions than the native-born population. Non-marital fertility rates have been estimated from a variety of sources and marriage and divorce rates have been improved. These refinements have eliminated in Figure 4 the discrepancy at older ages between predicted and observed numbers of average living grandchildren that is so evident in Figure 1.

With these supplemental simulations we can examine many forms of output statistics that cannot be recovered from Reeves' original simulations. For example, we can go beyond average values to whole distributions. Figure 5 shows one such comparison. It compares the observed and predicted distributions of number of living biological grandchildren for women aged 65 to 80 around 1990. The observed figures are taken from the NSFH for 1987-88, from the same data as Figures 1 to 3, with 840 women contributing to the bars. The predicted figures are from a single simulation with correlated mother-daughter fertility of the same form as those that supply the dotted lines in Figure 4, with 505 women contributing. Standard errors on the bars are on the order of one or two percentage points. The percentage without living grandchildren in this age group is 29% according to both the NSFH and the simulation. The percentage with a single grandchild is 6% according to both. The percentage with nine or more grandchildren is slightly over-predicted by the simulation, perhaps because the assumed correlation in mother-daughter fertility is too high. The difference is just statistically significant at the 5% level.

A full comparison of distributions would require a whole new paper. The one case presented (the only one so far examined) at least suggests that the efficacy of simulation in predicting kin is not entirely limited to averages of the sort we have been able to test for Reeves' original predictions.

5. Conclusions

Our comparisons between the microsimulation forecasts of the early 1980s and the NSFH kinship estimates of 1987-88 bring some good news and some bad news.

The good news:

i) Some kinship statistics -- average grandchildren below age 70, average siblings below age 40 -- were predicted with impeccable accuracy. Where the survey results are themselves more precise, the microsimulations achieve good accuracy.

ii) The random error in the simulations done in 1981 was as small as the sampling error in the 1987-88 survey.

iii) The discrepancies amount to no more than 15% to 20% of the levels, so in all cases the forecasts were at least usable.

iv) Even somewhat elusive categories like living half-siblings were predicted nearly within the limits of sampling error of the survey.

The bad news:

v) There are occasional substantial systematic discrepancies.

vi) Not surprisingly, wrong guesses about future demographic rates produce wrong numbers for kinship forecasts.

vi)   While some portion of some of the discrepancies may reflect biases in the survey, most are probably errors in the assumptions of the simulations.   These errors, however, appear to involve contributions from several sources in interaction, rather than isolable mistakes.

It is noteworthy that the most conspicuous discrepancy, that for living grandchildren of women over 70, would have shown up before the availability of the NSFH in comparisons between different microsimulations. Reeves (1982:179-182) carried out such comparisons for selected ages, but the highest age group he examined for any statistics was the group of 65 to 70-year-olds, and for grandchildren in this group the agreement with KOA is still good. The larger standard errors in all kin counts above age 70 probably induced Reeves to stop with 70 and not to place much credence or give much attention to his older-age results.

For the purpose of improving future kinship forecasts, several lessons emerge. First, simulations need better old-age mortality forecasting. In the last few years, more sophisticated stochastic approaches to mortality forecasting have been published, for example, Lee and Carter (1992). These are being used in the most recent kinship forecasts. However, only time will tell whether this greater sophistication leads to forecasts that prove true.   Second, the modeling of heterogeneity in fecundability (and presumably in marital stability, partner choice, and other components) needs more attention. As with mortality, a better empirical basis for modeling heterogeneity in fecundability has become available, and heterogeneity in other processes is also subject to study. Third, an early-warning signal for inaccurate forecasts can be provided by discrepancies between alternative microsimulation projects. Up till now, there has

been little attempt at cross-replication of forecasts by separate investigators and teams. It appears from our study that such cross-replication could be very effective at identifying problematic portions of generally credible predictions.

The kinship microsimulations of the late 1970s and early 1980s had to cope not merely with unavoidable uncertainty about the vital rates of what was then the future. They also operated in a vacuum of empirical information about the numbers of kin in the current population. Under these challenging conditions, they succeeded in anticipating much of what has now been learned about numbers of close living kin from the high-quality surveys of the last 15 years. But the forecasts also here and there go systematically awry. The numerical levels of the errors are not perhaps too serious, but the patterning of the errors is troubling and potentially limits the scope for qualitative generalization. On the other hand, there are indicators of kinds of improvements that should allow better accuracy in future attempts.

Kinship microsimulation is at the forefront of computer modeling in demography. Rapid advances in technological capabilities are being made and impressively complex models are coming into use. However, sophistication is not an end but a means. Every field can benefit from having its practitioners step back and ask whether the answers that come out of the methods succeed in matching anything in the real world. This study is such an external validity test. Some of the answers match. Some of the answers miss. We need to know, in order to put properly qualified faith in the predictions that come out of our machines.

References

Bongaarts, John, T. Burch, and K.W. Wachter, editors (1987) *Family Demography: Methods and their Applications,* The Clarendon Press, Oxford.

Citro, Constance, and E. Hanushek, Editors (1991) *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling,* two volumes, National Academy Press, Washington, D.C.

Clague, Alice, and J.C. Ridley (1973) "The Assessment of Three Methods of Estimating Births Averted", in Dyke and MacCluer (1973).

Cohen, Michael L. (1991) "Evaluations of Microsimulation Models: Literature Review", in Citro and Hanushek (1991).

DeVos, Susan, and A. Palloni (1989) "Formal Models and Methods for the Analysis of Kinship and Household Organization", *Population Index* 55:174-198.

Dyke, Bennet, and J. MacCluer, Editors (1973) *Computer Simulation in Human Population Studies,* Academic Press, New York.

Hammel, Eugene A., C. Mason, K. Wachter, F. Wang, and H. Yang (1991), "Rapid population change and kinship : the effects of unstable demographic changes on Chinese kinship networks, 1750-2250", *Consequences of Rapid Population Growth in Developing Countries* edited by G. Tapinos, D. Blanchet, and D. Horlacher, published in behalf of the United Nations, Taylor and Francis, New York, pages 243-271.

Hammel, Eugene A., K. W. Wachter, and C.K. McDaniel (1981) "The Kin of the Aged in A.D. 2000" in *Aging* , edited by S. Kiesler, J. Morgan, and V. Oppenheimer, Academic Press, New York, pages 11-40.

Howell, Nancy (1979) *Demography of the Dobe !Kung,* Academic Press, New York, Chapters 5, 11, 14,15, and 16.

Hammel, Eugene A., D. Hutchinson, K. Wachter, R. Lundy, and R. Deuel (1976) *The SOCSIM Demographic-Sociological Microsimulation Program Operating Manual* Institute of International Studies Research Monograph No. 27, University of California, Berkeley, California.

Hammel, E. A., C. K. McDaniel, and K. W. Wachter (1979) "Demographic Consequences of Incest Prohibitions: A Microsimulation Analysis" *Science* 205:972-977.

Kunstadter, Peter, R. Buhler, F. Stephan, and C. Westoff (1963) "Demographic Variability and Preferential Marriage Patterns", *American Journal of Physical Anthropology* 21: 511-519.

LeBras, Hervé (1973) "Parents, grandparents, bisaieux", *Population* 1:9-37.

Lee, Ronald D, and Larry Carter (1992) "Modeling and Forecasting the Time Series of U.S. Mortality", *Journal of the American Statistical Association* 87:659-671.

Lin, Jiang (1992) "Parity and Security: A Simulation Study of Population Aging, Kinship Network, and Old Age Security in China", Ph.D. Dissertation, Department of Demography, University of California, Berkeley, California.

Lin, Jiang (1994b) "Parity and Security: A Simulation Study of Old Age Support in Rural China", *Population and Development Review* 20:423-448.

Menken, Jane (1985) "Age and Fertility: How Late Can You Wait?", *Demography* 22:469-483.

Pullum, Thomas, and Douglas Wolf (1991) "Correlations Between Frequencies of Kin", *Demography,* 28, 391-410.

Rao, A. V., Q. Lindsey, R. Bhavsar, B. Shah, D. Horvitz, and J. Batts (1973) "The Evaluation of Four Alternative Family Planning Programs for Popland, a Less Developed Country", in Dyke and MacCluer (1973).

Reeves, Jaxk (1982) "A Statistical Analysis and Projection of the Effects of Divorce on Future U.S. Kinship Structure", Ph.D. Thesis in Statistics, University of California, Berkeley.

Reeves, Jaxk (1987) "Projection of Number of Kin", in *Family Demography* editted by J. Bongaarts, T. Burch, and K. Wachter, Clarendon Press, Oxford.

Smith, James E. (1987) "The Computer Simulation of Kin Sets and Kin Counts". in *Family Demography* editted by J. Bongaarts, T. Buch, and K. Wachter, Clarendon Press, Oxford.

Stoto, Michael (1983) "The Accuracy of Population Projections", *Journal of the American Statistical Association,* 78:13-20.

Sweet, James, L. Bumpass, and V. Call (1988) "The Design and Content of the National Survey of Families and Households", NSFH Working Paper No. 1, Center for Demography and Ecology, University of Wisconsin.

U.S. Bureau of the Census (1992) *Statistical Abstract of the United States 1992,* 112th Edition, Washington, D.C.

Wachter, Kenneth W. (1995) "2030's Seniors : Kin and Step-Kin", paper presented at the 1995 meetings of the Population Association of America, San Francisco.

Wachter, Kenneth W., with E.A. Hammel and P. Laslett (1978) *Statistical Studies of Historical Social Structure* Academic Press, New York.

Wolf, Douglas (1988) "Kinship and Family Support in Aging Societies", in *Social and Economic Consequences of Population Aging,* U.N. Department of International Economic and Social Affairs, New York.

Wolf, Douglas A. (1994) "The Elderly and their Kin: Patterns of Availability and Access", in *Demography of Aging, edited by Linda Martin and Samuel Preston, National Academy Press, Washington, D.C.*

Wolfson, Michael (1989) "Divorce, homemaker pensions, and lifecycle analysis", *Population Research and Policy Review* 8:25-54.

Wood, James, and M. Weinstein (1988) "A Model of Age-specific Fecundability",

    Population Studies 42:85-113.

Yi, Zeng (1988) "Changing Demographic Characteristics and the Family Status of

    Chinese Women", *Population Studies* 42:183-203.

Zhao, Zongwei (1994) "Demographic Conditions and Multi-generation Households in

    Chinese History. Results from Genealogical Research and Microsimulation",

    Population Studies 48:413-426.

Footnote:

Address for Correspondence:

Prof. Kenneth W. Wachter
Department of Demography, UCB
2232 Piedmont Avenue
Berkeley, CA. 94720

phone 510-642-1578; 510-642-9800

fax 510-643-8558

e-mail wachter@demog.berkeley.edu

Figure Captions

Figure 1: Average Numbers of Living Biological Grandchildren for Women in 1988 as a function of age. The dotted curve shows the simulation forecasts, the solid curve shows estimates from the 1987-88 National Survey of Families and Households. Vertical lines show separate 95% confidence intervals for the plotted points.

Figure 2: Average Numbers of Living Full Siblings for Men and Women in 1988 as a function of age in the same format as Figure 1.

Figure 3: Average Number of Living Half Siblings for Men and Women in 1988 as a function of age in the same format as Figure 1.

Figure 4: Effects of correlated mother-daughter fertility on estimates of average numbers of living grandchildren for women around 1990 by age. The solid line shows results from a simulation with no correlation between mothers' and daughters' fertility. The dotted lines show four replications from simulations with correlations around 0.061. The circles show estimates from the 1987-88 N.S.F.H.

Figure 5: The Distribution of number of living grandchildren observed in the 1987-88 NSFH for women aged 65 to 80 (solid bars) compared to that predicted in a simulation using rates described in Wachter (1995) (dotted bars). The simulation includes the same correlated fertility as in Figure 4.