NOTE ON STATISTICAL ANALYSIS AND MICROSIMULATION FOR STUDYING LIVING ARRANGEMENTS AND INTERGENERATIONAL TRANSFERS

*Douglas A. Wolf**

ON MODELS FOR THE ANALYSIS OF LIVING ARRANGEMENTS AND INTRA-FAMILY TRANSFERS

In the area of multivariate modelling of living arrangements and family transfers, we are inevitably led towards a desire, or need, for complexity in model specification and hence difficulty in estimation. This complexity arises because we generally wish to represent the situations of multiple actors (decision makers), for example, an older person or couple and their several children, and, possibly, the children's parents-in-law as well. Furthermore, each actor may engage in one or more of a set of multiple activities of interest, including co-residence, financial transfers, or the provision of personal-care services. The spatial proximity of members

There are no doubt numerous other instances of available data that could support unexpected analyses in domains far from their originally intended range of topics.

A difficult issue that arises in the specification of models that depict outcomes in multiple domains is that of endogeneity, or simultaneity. But, these issues can easily be misunderstood. If two outcomes are both viewed as choice variables under the control of a single actor, then it does not make sense to think of them as reciprocally causally related (i.e., that a change in A causes a change in B, while B similarly produces its own distinctive causal response in A). Instead, they should be treated as "jointly determined". This will give the statistical specification of the model the appearance of a reduced form. Variables A and B are still jointly endogenous, and presumably depend, in part, on common unmeasured variables (i.e., exhibit correlated disturbances) but do not appear as each other's regressors. For example, an area of considerable research activity at present is the question whether women's hours of paid employment and their hours of familial caregiving activity are negatively related. If a woman is viewed as the sole decision maker, and her decision is made conditional on exogenously given prices (e.g., market wages, household productivity, and costs of market substitutes for her own caregiving time), a fixed "care production technology", and fixed preferences, then the two time-use outcomes are jointly determined.

temporal sequence is the same as the causal sequence. However, actors make plans, they have expectations about the future, and they take steps today that reflect their plans about the future. Thus, there is a sense in which events in the future "cause" events in the present. Secondly, "contextual" variables are not necessarily exogenous. Multilevel modelling is, at present, a popular and rapidly developing analytic tool, but as context is virtually always location-specific, one must recognize that the inclusion of contextual variables introduces possible endogeneity bias, as actors are to some extent free to choose their location. They may choose their location so as to achieve a favourable context, for example, older persons may migrate to a service-rich area (or to their child's neighbourhood) in anticipation of future care needs. If so, the contextual variables are not exogenous. This criticism is often made; solutions to the problem are far more rare.

ON MICROSIMULATION, IN THE CONTEXT OF FAMILY/KIN NETWORKS AND INTRA-FAMILY TRANSFERS

It makes sense to turn from a discussion of model specification to microsimulation, since microsimulation must be preceded by model specification and estimation.

What is microsimulation? The essential ingredients are the use of computer-based sampling, and an analysis that is conducted at the maximally disaggregated level, that is, that of the individual (which might be a person, a couple, a firm or organization -- whatever is the fundamental analytic unit at hand). The "sampling" is, in fact, a process of making stochastic assignments of values to variables. These remarks pertain to a situation in which the "model" is a set of relationships among observed and unobserved factors (in the demographic domain, primarily); the unobserved factors are assumed to come from particular distributions; the model produces a distribution of possible values for the outcome of interest, and the computer program—the sampling algorithm—selects a particular value from that distribution. The sampling process may be repeated many times for a particular individual, and there may be many individuals (e.g., a sample, and even, perhaps, everyone in some population) to which the sampling algorithm is applied.

An interesting question is the following: is microsimulation a complement to, or an alternative to, "macro" simulation? Before addressing this question, it should be noted that the distinguishing features of networks, and the best-known work in this area is by Kenneth Wachter and his colleagues (Hammel, Wachter and McDaniel, 1981; Wachter, 1997).

Microsimulation could, in addition, be used to conduct a conventional population projection, but it is hard to imagine that anyone would seriously want to. Situations in which microsimulation does reveal its value include those characterized by (a) complex models ! for example, multiple-equation models in which multiple actors make decisions about multiple interrelated domains of behaviour, such as the intra-familial transfer situations described above; (b) situations involving interactions between individual members of a population, for example the workings of mating markets; (c) models that explicitly represent "unmeasured heterogeneity", such as the "frailty" models of human mortality developed by Vaupel and colleagues (Vaupel and Yashin, 1985; Vaupel, Manton and Stallard, 1979) or the random-mixture models of Hutterite fertility developed by Heckman and Walker (1987); (d) the analyst's wishes to quantify the various sources of uncertainty, or forecast variance, in a model. Microsimulation can also be a way to extend the range of lessons that can be learned from some types of models. For example, in a conventional linear single-equation regression setting, most of what one might want to learn from the estimated model can be learned from the coefficients themselves, or simple transformations of them. Forecasts are also easy to carry out. In contrast, a Markov renewal model of, say, labour market transitions may incorporate a set of age- and duration-dependent hazard functions for transitions among states "never worked", "working", "unemployed", and "retired". Further complexity can be introduced by distinguishing between different jobs held over the worklife. Having estimated all the parameters of such a model (even a simple one, with only a few time-invariant covariates), the analyst can draw only a limited set of conclusions about the overall life-course process from the parameters of the hazard functions themselves. But with microsimulation, the analyst is free to compute numbers that answer questions as detailed as "what are the chances that someone who entered the labour market at age 24 is in his seventh job at age 47?" and so on.

Since microsimulation is fundamentally an exercise in sampling, it is crucial that the simulator pay attention to the issue of sampling error. A run of a microsimulation computer program produces, typically, a microdata file full of randomly assigned variable values. The values might purport to represent the situation at some future date, starting from an observed starting point for some well-defined population. If a sample of equivalent size could be drawn from the actual future population, it would be possible to proceed to compute estimated standard errors for any summary statistics based on that sample data. The same should be done if the data are simulated.

It is also important to remember, as noted above, that for each individual whose future is being simulated, the "model" (embedded in the computer program) generates a probability distribution over possible values of each variable in the future, while one run of the simulation program produces one draw from this distribution for each person. These draws do not represent, on their own, the expected value of that person's variable, but rather a randomly-selected particular value of that variable. The value assigned may be far from the expected value but can still be "correct" (in a probabilistic sense). The expected value (for the person) may, in fact, not

References