Celebrating Longitudinal Data by Making It Sing and Dance: the Role of Dynamic Microsimulation Modeling

Michael Wolfson
University of Ottawa
mwolfson@uottawa.ca

Longitudinal household survey and administrative data have become increasingly central to high quality and leading edge social science research and policy analysis. The original PSID finding that poverty, for example, was not a permanent state, but rather that people moved into and out of poverty, was seminal. Subsequent work has increasingly focused on the correlates of these kinds of transitions in fields ranging from demography to economics to epidemiology. However, the vast majority of research using longitudinal social science data tends to be one relationship at a time. For a given dependent variable – whether poverty or union formation or a heart attack – the analysis typically involves estimating a statistical relationship between this variable and a number of potential "explanatory" variables. The questions then answered are which of these independent variables have a statistically significant association with the dependent variable of interest, and if so what sign and how large. This approach is fine as far as it goes. But longitudinal data have far more potential utility. Public policy analysis frequently involves questions of the form, "what if … ?". Further, those involved in the associated analysis and debates typically put forward a "web of causality" – a story of how multiple factors interact, co-evolve, and are connected with each other in a network of mutually causal relationships. Such causal webs are often represented by a "box and arrow" diagram, where the boxes represent factors, and the arrows represent presumed causal pathways. In this context, longitudinal data analysis of the one equation at a time sort may describe the arrows pointing to one of the boxes in such a box and arrow diagram. This kind of analysis is therefore helpful, but far from providing the basis for addressing the policy analyst's "what if" question.

The response is to assemble a network of one equation at a time longitudinal data analyses into a model, specifically a longitudinal dynamic microsimulation model designed to address the policy question at hand, and typically at least a somewhat broader family of such questions. These microsimulation models are more expensive to build and maintain than one or a series of one equation at a time longitudinal data analyses. But they are far more powerful and useful. In this paper, we describe several of these models developed in Canada. In each case, we will describe the range of research or policy questions they have been designed to address, and then elucidate the critical longitudinal data sets which were used, as well as the statistical methods applied to these data sets, in order to build the model. The specific models include PopRiskS-BC = Population Risk Stratification for Breast Cancer (to determine the prospective cost-effectiveness of using genetics to stratify women for breast cancer screening according to their

assessed risk), LifePaths (for pension policy and long term care projections), and HealthPaths (to assess the relative contributions of various risk factors and socio-economic determinants to health-adjusted life expectancy).

The paper will conclude with some observations on the institutional and funding challenges of building and maintaining such models, not only in Canada, but also in several other OECD countries.