Microsimulation at StatsCanada

Regan Meloche - Jan 5, 2021

This post is a brief summary of the **microsimulation approach** taken by StatsCanada on several projects, as well as an exploration of how it may be relevant to a policy difference engine.

Motivation

Two key goals of simulation in general include finding explanations for observed phenomena, and predicting future phenomena. Supposing we want to simulate a policy change in order to predict its impact, we can first look at a traditional approach, sometimes called a "cell-based" approach. This involves considering variables associated with the people that the policy may target. Suppose for example, we are interested in the senior citizenship and "handed-ness" of people for a certain policy. A person is either a senior or not (say aged 65+), and they are either right-handed, left-handed, or ambidextrous. We have 2 options for senior citizenship and 3 options for the handed-ness. This gives a total of 6 different possibilities of people.

Now suppose instead of just 2 variables, we want to model 12 variables, and suppose each of those variables have 6 different options. Now we have 6^12 possibilities (or "cells"). Each time we add a new variable of interest, the number of possibilities increases exponentially. This gets even harder when we consider continuous variables. If we were to store each possibility in a database, this can get very large very quickly.

Microsimulation

When we want to model complex situations, we want to be able to model many different variables, lest we leave out some important information. It can become computationally impractical to model this complex situation by taking this cell-based approach. An alternative is, instead of modelling the different

combinations, create simulations at the “individual” level. That is, create thousands (or millions) of virtual individuals that exhibit the full spectrum of diversity inherent in the variables you want to account for (age, income level, citizenship status, height, marital status, etc.).

This is the idea behind microsimulation. A virtual world of simulated individuals is created, the changes of interest are applied, and we can observe the effect. This is useful in systems where the behaviour of individual components of the system (micro-level) is responsible for the behaviour of the aggregate system (macro). Just as liquids are made up of individual particles, and a traffic system is made up of individual cars travelling along roads, we can understand the overall behaviour and performance of these systems by looking at the micro scale, and then obtaining macro outcomes by aggregation.

Dynamic Microsimulation

Another property of a simulation is whether it is static or dynamic. A static simulation measures change without taking into account the fact that the behaviours of individuals in the system will also change as a result. A dynamic model takes this into account, and is therefore likely to be more realistic, though more complex. A dynamic microsimulation model can follow the individuals over a time period. It may be for a few months, or it may be over the course of their entire lives.

Data Considerations

A main component of a dynamic microsimulation model is a database that stores the characteristics of all the members of your population. This is dynamically updated as the simulation runs. From this data, we can output aggregated tables and files that can be further analyzed.

Contrast this with a cell-based approach, where we may only actually use a few pieces of aggregated data to build a model, such as distributions of age range,

income, etc. In a microsimulation model, we store all the data about each individual, and we can use probabilities with a degree of randomness to populate these variables.

Drawbacks

This idea of randomness highlights one of the drawbacks of microsimulation. The level of detail on the data does not relate to better predictive power, since there is this inherent randomness involved. If we are randomizing a collection of individuals in a microsimulation, then it may not necessarily correspond to the population we are attempting to make predictions for.

Another inherent issue is the complexity. Microsimulation requires high quality specific data, as well as infrastructure to store and update that data. This can create a high barrier to entry for a microsimulation approach, and may require significant investment.

Strengths and Use Cases

The strength of the microsimulation is the ability to model complex situations. We may need to forego the desire for a simple solution to a complex problem in favour of usefulness. If done properly, the microsimulation can be very useful.

An important use case is a situation where you have a heterogenous population, where there are too many combinations of characteristics to split the population into a manageable number of groups. As mentioned at the beginning, a microsimulation model is also valuable when the macro behaviour of a system is best understood by looking at the individual actors.

Existing StatsCan tools

StatsCanada has developed a number of tools that use a microsimulation approach. These can be found in the main link, but we will highlight a few here.

SPSD/M is a microsimulation model for individuals and families that uses non-confidential data sourced from surveys and administrative data. It can be used for policy analysis related to tax and transfer programs, as well as income distribution.

LifePaths is a tool that dynamically models complete lifetimes of individuals. It is no longer supported, but a successor is currently under development.

Modgen is a generic microsimulation programming language based on C++ that supports dynamic models. The underlying mechanism is hidden from the user so that they can focus on parameters, actors, and events.

Application to PDE and Next Steps

Part of Team Babel's research and development phase involves exploring existing tools that may satisfy or be related to the goals of the Policy Difference Engine. Based on the descriptions, these tools may be very relevant and will require further investigation.

Some options worth exploring, pending further research:

- Can we use the Modgen language to add value to the existing SPSD/M project in a way that aligns with the goals of the PDE? We will be getting access to the SPSD/M so we can get some hands-on experience with it to see how it fits in with the fellowship.
- Can we learn any lessons from the former LifePaths project, and can we explore the project that will be its successor?

It is emphasized that a functional microsimulation model requires sophisticated data as well as a sophisticated database. We will want to investigate further what type of data may be required for a microsimulation. If the complexity of a practical project is out of scope for the fellowship, then we can at least document some important lessons and notes for future considerations.