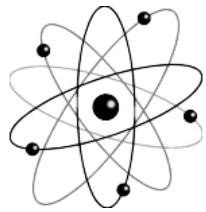
1-2001

Simulation methodologies for political scientists

Kenneth BENOIT
Singapore Management University, kbenoit@smu.edu.sg

Computational Modeling



Simulation Methodologies for Political Scientists

Kenneth Benoit
Trinity College, University of Dublin
kbenoit@tcd.ie

In this brief article I identify and characterize the variety of simulation methodologies relevant to political science research. Having identified them, I then recount a light-hearted tale of one successful application of a computer simulation to a real world problem of enormous significance. Extending the discussion to other examples, I discuss some of the main issues and benefits in simulation models for political science. Finally, I identify some software tools and additional resources for those who would like to learn more about using simulation methods for their own work.

Consider What We Already Do

The notion, and appeal, of computer simulations is simple: they enable us to build a model of a process whose workings can be examined—and more importantly controlled—as analogs to the real process being approximated. Readers of *TPM* should be very familiar with the static equivalent: presenting equations as "models" of some relationship between political variables, using parameters to characterize the relationships between these variables. The form of the equation itself—linear, exponential, logarithmic, additive, multiplicative—also imposes an assumed structure on the relationship between the variables. These mathematical equations are "models" precisely because they present a stylized representation of a more complex process, filtering out aspects of the process which it is either impossible or uninteresting to include. Whatever is not included systematically is also parameterized, typically as one or more disturbance terms, once we make tidy and fairly stringent assumptions about all these things we cannot explain. Once the model has been specified, we are then presented with a method for estimating the parameters using a statistical procedure on data collected on each variable.

Let's examine some of the global assumptions in this approach. First, in order to make it possible to estimate the parameter values, we have to impose *aggregate* assumptions on the model quantities, in particular about the distribution of the parameters, especially the disturbance parameters, and about the aggregate relationships between the variables we choose to include and exclude. We can never verify these assumptions, however, since our model will only work if they can be taken as true. When it does work the results it provides will be contingent upon, and therefore seem to reaffirm these assumptions.

A second assumption we typically make is that the observations are conditionally independent, that the values of our explanatory variables are not conditional upon the values of our dependent variable. We also assume that independent variables take on values that are independent of the values of other independent variables. (If we assume otherwise we must then alter the model further to make it effectively true.) For instance, even though we might have a model estimating strategic behavior, we assume that the units react to choice situations only and not to each other. A related assumption is the basic idea that the units being studied do not change their behavior over time: that they do not learn from their own mistakes, let alone those of others.

A final assumption is that the observations are "identically distributed," conditional of course on explanatory variables. This means that there are no differences between units we are treating as data that cannot be captured by variables and model parameters.

Of course there must also exist some practical method for obtaining reliable numerical estimates of the model parameters. There are many great quantitative models that we simply have no way of estimating in practice, although many ways exist in theory. Incredible strides in computing power have pushed this bar downward, making computation solutions possible to problems whose solutions were previously impractical. This revolution in computing power is highly relevant to the remaining discussion of simulation models, to which I now turn. Actually, to which I almost now turn, since I think a brief illustration might highlight the contrast in approaches to solving a problem where simulation might provide an answer.

Illustration: The Bus Problem

Lest the preceding comments start you thinking that one of my pseudonyms might be "Mr. Perestroika," let me point out that first, my goal was to critically examine the models we are most familiar with, so that they could be contrasted to simulation methodologies; and second, that some of my best friends are statistical methodologists. One of these friends, in fact, had an experience that illustrates a lot about the potential of simulation to crack tough modeling problems.

This friend of mine used to take the Mass Ave bus from Arlington to Cambridge (MA) every morning, and was frequently irritated to find himself arriving at the bus stop just as the bus pulled away. Instead of scheduling stops at fixed times, the bus schedule indicated only that from 7-9 am there would be buses on this route every 10 minutes. A statistician friend suggested that the mean waiting time would be minimized if a person trying to catch the bus would appear at random times, rather than at a fixed time every day, say at 8am. How, thought my friend, could he verify that this was true? First, we could model the bus being at the stop as a random variable, making assumptions about the regularity of its frequency and the duration of its stop. We could also model the arrival of the would-be passenger. We would then need to figure out a function minimizing the expected wait of the passenger for the next bus after he arrives at the bus stop. Now while proving this result analytically would be extremely easy for any of us, the proof was elusive for this friend of mine.

So after many frustrating and unsuccessful hours spent attempting to prove this result, my friend decided to change his approach. Why not conduct experiments to discover the result instead? So for the next thirty working days, he showed up to the bus stop—the one just next to the small grocery store with the green sign—at precisely 8:00 am, and recorded in his handheld computing device the time he spent waiting for the bus to arrive. Then for the next thirty days after that he did the same but randomly chose to arrive between 8:00 and 8:10, by rolling a 10-sided die which his roommate had said was from a game with the unlikely name of "Dungeons and Dragons."

At the end of the two experimental periods, this friend—let's call him "Ben"—compared the averages from the two experiments and tried to draw a conclusion. But he immediately noted some problems. First, on one or two mornings he suspected that his timing of arrival was late by about one minute. Second, he noticed that on one day he had waited for more than 12 minutes, which was not supposed to happen according to the bus schedule. Finally, during the second month of the experiment it had often snowed, slowing down traffic and apparently stretching the bus intervals. Dismayed by the lack of control over his experiment, Ben rejected his experimental results and decided to try yet another approach.

Thinking about the problems associated with his experiments, my friend realized that the essence of the problem—the efficacy of attempting to "time" the bus stop versus just showing up unplanned—was unrelated to random events like weather and oversleeping that can contaminate an experiment. So he decided to conduct a *simulated* experiment. Fortunately, Ben was an experienced computer programmer and had just purchased a new 486DX/100 running Slackware Linux 3.0. Programming his computer with crude agents for the passenger and the bus, he simulated an arrival of the bus at random times within the 10-minute interval and had the passenger arrive each morning at a fixed time. Running this 1,000 times, he recorded the waiting time for each round of the experiment. The same was done having the passenger also randomly arrive, recording the waiting times with this behavior 1,000 times. By comparing the distributions of the waiting times from the two experiments, he was then able to satisfyingly visualize and summarize the expected waiting times given each passenger's behavior. (I leave the answer as an exercise for the reader.)

This opened a whole new world of possibilities for my friend, including venture capital. Instead of starting his own firm as I suggested, however, he became obsessed with developing more complex models related to buses, including a simulation of bus passengers as agents with different forms of seating behavior. He was interested in how to explain why some buses seemed to carry more people in a more orderly fashion than others with fewer passengers. He programmed behaviors into several types of agents. The first, we will call the Arlington passenger. This passenger, knowing that the bus will pick up many additional passengers along Mass Ave, takes the last available seat in the rear of the bus, on the window side. Another type is the Porter Square passenger, picked up midway, and this passenger likes to sit surrounded by empty seats, and therefore takes the seat at the centroid of the largest available bloc of free seats. There are other types as well, such as the elderly passenger, taking the first free seat, and the Teenage Bloc of 3-5 passengers who search for a group of adjacent seats. By programming this system as agents, my friend was able to investigate the emergent behavior of seating patterns at various levels of capacity, with different levels of groups, picked up in various sequences. As far as I know, Ben is still ABD in a top-five sociology Ph.D. program.

The moral of this story is that for some problems it may be quite natural to turn to simulations as a superior methodology than either an analytical solution or experiments. For the second application involving modeling bus passengers' seating behaviors, simulation methods were used to model complex process based on individual agents operating to sets of known rules, and thereby yield insights into the dynamics of the aggregate system emerging from the behavior of the agents. The remainder of this article discusses such applications in more detail.

Enter "Simulation Methodologies"

I suppose it's high time to actually define what I mean by simulation methodologies. First let me clarify what is not included. We have all probably heard, or used, simulations in the context of statistical estimation. This is one of the great benefits of Bayesian "simulation" and related approaches: when we do not know or cannot express a distribution, we can approximate it by sampling from component distributions whose properties are known and can be expressed. The same approach is used in numerical optimization problems, where search techniques are used to explore a distribution whose global shape is unknown. Similar techniques are employed for testing convergence of parameter estimates using other iterative methods. I will not focus on these varieties of "simulation," such as Monte Carlo simulations, except to point out that the underlying approach is very similar to using simulation models to explore complex processes. In both statistical simulation and in computational simulation of complex processes, some real system is approximated by a model, this model's operation and behavior is defined by the researcher, this behavior is then produced using a computer, and observation of these results is used to yield leverage on the real system that has been approximated.

Simulation methodology broadly refers to the building of models of the world that have both inputs and outputs. Inputs are entered by the researcher, along with behaviors and rules structuring the simulation, and outputs are observed as runs of the simulation. Simulation in this context is basically synonymous with "computer simulation," since the simulation models are constructed and run as computer programs. Simulation methodologies have many variants, and I have compiled a list of the main types below. I warn however that this is is illustrative rather than exhaustive, being targeted toward political scientists and not meant as a guide to predicting the weather, extrapolating fishery yields, or perfecting nuclear warhead designs.

- Agent models. One of my personal favorites, agent models refer to the use of self-contained programs which can control their own actions based on their own perceptions of their operating environment (Huhns and Singh 1998). At the simplest level, agents are the actors in the simulations whose characteristics and range of behaviors are defined by the researcher. The researcher can determine the shape and units of their utility functions, the process and rules by which they make decisions, and whether they learn from history or from each other.

- Evolutionary models. These are either systems or agent models that are distinguished by the ability to alter their parameters or even the basic structure of the model itself in response to learning. This broad category includes rocket-science variety methods such as genetic algorithms and artificial neural networks.

- Cellular automata models. Really just stripped down agents, simulations of these types display emergent patterns based on successive iterations of rule-following behavior of individual components on a grid. The action of each "cell" on the grid is influenced by the states of its neighbors. A classic example is Conway's Game of Life.

- Systems dynamics and related models. Generally involving complex maps resembling engineering schematics, systems dynamics models focus on macro-level outcomes based on a target system described using a system of difference and differential equations. These are used to derive the future state of the system from its present state. An example would be the WORLD2 and WORLD3 models (Meadows 1974).

- Microanalytical simulation models. These model processes by shifting attention to the micro-level agents. Microsimulation follows successive generations of individual units, hoping to predict a future state given a starting state. Examples would be simulations designed to predict the effects of policy changes on some target population, such as lowering the capital gains tax.

Just as advances in computing power have revolutionized statistical modeling, advances in computing have made many simulation applications feasible that were once possible only in theory. Indeed, because of the way in which they manage complexity and uncertainty, simulation is well-suited to investigating problems for which closed form solutions are impossible, or to better understand problems whose closed form solution is uninformative.

Benefits: Simulating is Stimulating

Returning to our discussion of the "limits" of conventional statistical models, simulations offer a number of advantages. Probably the most important insight to be gained from computational modeling is the study of *emergence*. Emergent behavior "refers to a computation or phenomenon at the macro-level that was not hard-coded at the micro-level, such as when a market computes the price where supply equals demand even though no one is trying to compute the market price" (Page 1999, 4). By recreating the process using the micro-level agents, it is hoped the computational models can both explain observed emergent behavior, and investigate speculative emergent behavior when the micro-level behaviors are altered. Rather than making assumptions about aggregate-level behaviors, we can treat this aggregate-level outcome as an emergent behavior to be tested. Simulation provides the mapping of micro-level behavior to aggregate outcomes.

Formal theorists will be very familiar with this problem. Formal theory and especially game theory is structurally very similar to setting up a special sub-class of agent simulation. The key difference is that solutions are arrived at through the technique of deductive proofs, rather than actually simulating the games or behaviors that have been formalized. But what if we were to program simulations to actually *play* the games described?

For some problems where analytic solutions are impossible, this is in fact one of the most promising avenues for formal theory to follow. Conclusions about equilibriums can be derived from observing repeated plays of the game, supplementing or even supplanting purely mathematical results. Consider trying this on your students: show them a visualization of a spatial model demonstrating a chaos theorem or cycling. The intuition provided by seeing (and manipulating) the result through simulation is generally much more effective than working through formal proofs. (And not just for students.)

The flexibility of simulations also allows us to cast aside most, if not all, of the restrictions required in statistical modeling. The units we study can evolve, they can learn from past actions, they do not have to be independent, and their behaviors can be extremely complex. By making the behavior or micro-agents stochastic, simulations are well-suited to modeling the aggregate consequences of uncertainty. By examining repeated simulations and the trajectories they take, we learn not only about the outcomes but also about the dynamics of the process itself. By having access to the rules and behaviors which the simulation comprises, we can observe the consequences on outcomes of altering these rules and behaviors. In our pure environment, we control all of the factors governing the system, not subject to any of the errors, mistakes, unforeseen problems, or human or meteorological vicissitudes which might interfere with the conduct of experiments. Because simulations give the researcher ultimate control, simulations may be far better than experiments—in addition to being cheaper, faster, easier to replicate.

Examples of Applications

If you need any more convincing that computational research offers great possibilities for political science, simply consider this: our colleagues in economics are way ahead of us in simulation methodologies. Economists use computer simulations to explore the consequences of monetary and fiscal policy, commodity pricing in agricultures, the role of savings and investment on the process of capital accumulation, discrete choice models of public transportation (such as riding the bus or the train—and possibly what time to arrive at the station), global warming as affected by tax incentives—the list is quite long. An excellent summary of simulation work in economics can be found in a NSF-commissioned report on Computational Economics (Kendrick, Bergmann, Broder, David, and Geweke 1991). Economists even have a quarterly journal, *Computational Economics*[1], devoted entirely to applications, theories, and issues related to computation and simulation modeling.

[1] `http://kapis.www.wkap.nl/kapis/CGI-BIN/WORLD/journalhome.htm?0927-7099`

Political science applications also exist in respectable and growing numbers, and I have listed a few examples showing the areas of application. Once again this is only a sampling, rather than an exhaustive list. Examples include:

- the behavior of political parties in spatial elections (Kollman, Miller, and Page 1992);
- dynamic behavior of legislators in changing parties or forming new parties between elections (Laver and Benoit 2001);
- behavior of individual states or other international actors (e.g. Axelrod 1997b; Signorino 1996);
- the diffusion of culture (Axelrod 1997b; Bednar and Page 2001);
- formation of opinions and collective judgments (Johnson 1999); and
- models of social growth and resource conflict (among other social issues) (Epstein and Axtell 1996).

Other social science-type applications include analyzing traffic patterns, women's choice of birth control, aircraft engine replacement, patent renewal, regulation of nuclear power plants, school choice, decisions to marry, and retirement behavior.

Learning More

Software. Not only are there better comprehensive references to the software tools available for the implementation of simulation models, but also my extensive philosophizing about the epistemology of simulation above has edged out any room here for such a treatment. I will nonetheless mention the extraordinary Swarm Simulation System[2]. Swarm is a toolkit of code written in Objective-C, an object-oriented programming language (similar to C++). The toolkit consists of libraries of functions, routines, and objects that can be used together to set up simulations, record output from those simulations, and produce a variety of visual representations including graphs. Swarm is oriented toward agent-based models. Some other popular tools include Stella[3] and StarLogo[4]. These and other resources are well-detailed in Gilbert and Troitzsch's *Simulation for the Social Scientist* (1999), which is also an excellent introduction into simulation methods.

Additional readings. As mentioned above, Gilbert and Troitzsch covers various forms of simulations and discusses both applications and methods. A classic work with many examples is Robert Axelrod's *The Complexity*

[2] http://www.santafe.edu/projects/swarm

[3] http://www.hps-inc.com

[4] http://el.www.media.mit.edu/groups/el/Projects/starlogo/

of Cooperation (1997a); so is Epstein and Axtell's *Growing Artificial Societies*. The other works cited in the references below are also good places to start. A wealth of information can be gleaned from the Internet, including papers in progress, software, demonstrations, poster presentations, and FAQs. A good starting point is `http://www.soc.surrey.ac.uk/research/simsoc/`. So what are you waiting for? Time to get off the bus and start analyzing some political science problems.

References

Axelrod, Robert. 1997a. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration.* Princeton: Princeton University Press.

Axelrod, Robert. 1997b. The Dissemination of Culture: A Model with Local Convergence and Global Polarization. *Journal of Conflict Resolution* 41, 203–226.

Bednar, Jenna and Scott Page. 2001, June 8. Behavioral Diversity in Game Ensembles: Game-Theoretic Foundations of Culture. University of Michigan manuscript.

Epstein, Joshua M. and Robert L. Axtell. 1996. *Growing Artificial Societies: Social Science from the Bottom Up.* Cambridge: MIT Press.

Gilbert, Nigel and Klaus G. Troitzsch. 1999. *Simulation for the Social Scientist.* Philadelphia: Open University Press.

Huhns, M. and M. P. Singh. 1998. *Reading in Agents.* San Mateo, CA: Morgan Kaufmann.

Johnson, Paul E. 1999. "Protests, Elections, and Other Forms of Political Contagion." Presented at the 1999 annual meeting of the American Political Science Association, Atlanta, GA.

Kendrick, David, Barbara Bergmann, Ivy Broder, Martin David, and John Geweke. 1991. "Research Opportunities in Computational Economics." Technical Report R-91-1, Center for Economic Research, Department of Economics, University of Texas Austin. Report of the NSF Workshop on Research Opportunities in Computational Economics Washington, D.C. March 6, 1991.

Kollman, Kenneth, John H. Miller, and Scott E. Page. 1992. "Adaptive Parties in Spatial Elections." *American Political Science Review* 86 (4), 929–937.

Laver, Michael and Kenneth Benoit. 2001. "The Dynamics of Fission and Fusion in a Party System." Trinity College, University of Dublin Manuscript.

Meadows, D. L. et. al. 1974. *Dynamics of Growth in a Finite World.* Cambridge, MA: MIT Press.

Page, Scott E. 1999. "Computational Models from A to Z." *Complexity.*

Signorino, Curtis. 1996. "International Cooperation under Uncertainty: The Effects of Symmetric and Asymmetric Noise." *Journal of Conflict Resolution* 40 (1), 152–205.

Agent-Based Modeling in Political Science

Lars-Erik Cederman[1]
Harvard University
cederman@cfia.harvard.edu

Agent-based modeling is a computational methodology that allows the analyst to create, analyze, and experiment with, artificial worlds populated by agents that interact in non-trivial ways and that constitute their own environment (for introductions, see Axelrod 1997b; Casti 1997; Epstein and Axtell 1996; Epstein 1999; Axtell 2000). In these "complex adaptive systems," computation is used to simulate agents' cognitive processes and behavior in order to explore emergent macro phenomena, i.e. structural patterns that are not reducible to, or even understandable in terms of, properties of the micro-level agents (Cederman 1997, Chap. 3). Such "bottom-up" models typically feature local and dispersed interaction rather than centralized control (Resnick 1994). Moreover, as opposed to traditional models that assume either a small number of dissimilar or numerous identical actors, agent-based models normally include large numbers of heterogeneous agents. Rather than studying equilibrium behavior, the focus is often on dynamics and transient trajectories far from equilibrium. Finally, instead of assuming the environment to be fixed, many agent-based models let the agents constitute their own endogenous environment. Given its potential to bridge the gap between conventional formal tools and qualitative theorizing of complex settings, agent-based models are therefore more usefully seen as a complement to rational-choice techniques rather than as a rival.

Agent-based approaches should be contrasted to earlier uses of simulation in the social sciences (see references Gilbert and Troitzsch 1999), including the tradition of global modeling that peaked in the 1970s (Taber and Timpone 1996a, pp. 48-49).[2] Such "equation-based" models attempt to capture macro-properties of social systems numerically (for more on the distinction, see Parunak

[1] Author's note: Many thanks to Robert Axelrod, James Fowler, Nigel Gilbert, Ian Lustick, Michael Macy, Steve Majeski, Rick Riolo, Phil Schrodt, and Michael Ward for helpful comments.

[2] There are also other types of computational approaches to social science, such as rule-based models and natural language processing derived from artificial intelligence (Bainbridge et al. 1994; Taber and Timpone 1996a), or Monte Carlo simulations, but these fall outside the purview of this review.