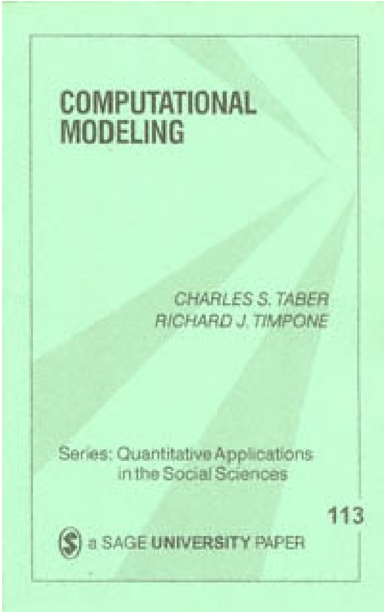

title:	Computational Modeling Sage University Papers Series. Quantitative Applications in the Social Sciences ; No. 07-113
author:	Taber, Charles S.; Timpone, Richard J.
publisher:	Sage Publications, Inc.
isbn10 | asin:	0803972709
print isbn13:	9780803972704
ebook isbn13:	9780585216805
language:	English
subject	Social sciences--Mathematical models, Social sciences--Data processing.
publication date:	1996
lcc:	H61.T237 1996eb
ddc:	003/.3
subject:	Social sciences--Mathematical models, Social sciences--Data processing.

Computational Modeling

SAGE UNIVERSITY PAPERS

Series: Quantitative Applications
in the Social Sciences

INSTRUCTIONS TO POTENTIAL CONTRIBUTORS
For guidelines on submission of a monograph proposal to this series, please write

Michael S. Lewis-Beck, Editor
Sage QASS Series
Department of Political Science
University of Iowa
Iowa City, IA 52242

Series/Number 07-113

Computational Modeling

Charles S. Taber
Richard J. Timpone

State University of New York at Stony Brook

SAGE PUBLICATIONS
International Educational and Professional Publisher
Thousand Oaks London New Delhi

For information address:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd
6 Bonhill Street
London EC2A 4PU
United Kingdom

SAGE Publications India Pvt. Ltd
M-32 Market
Greater Kailash I
New Delhi, 110 048 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Taber, Charles S.
Computational modeling / authors, Charles S. Taber and Richard J. Timpone
p. cm. (Sage university papers series. Quantitative applications
in the social sciences; no. 07-113)
Includes bibliographical references
ISBN 0-8039-7270-9 (pbk.: acid-free paper)
1. Social sciencesMathematical models. 2. Social sciencesData
processing. I. Timpone, Richard J. II. Title. III. Series
H61.T23 1996
003'.3dc20 95-50246

96 97 98 99 10 9 8 7 6 5 4 3 2 1

Sage Production Editor: Gillian Dickens

When citing a university paper, please use the proper form. Remember to cite the current Sage University Paper series title and include the proper number. One of the following formats can be adapted (depending on the style manual used).

(1) TABER, CHARLES S., and TIMPONE, R.J. (1996) *Computational Modeling*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-113. Thousand Oaks, CA: Sage.

OR

(2) Taber, C. S., & Timpone, R. J. (1996) *Computational Modeling* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-113). Thousand Oaks, CA: Sage.

Contents

Acknowledgments

A number of our colleagues contributed directly or indirectly to this monograph. Joel Kaji, Kenneth Rona, Jeffrey Segal, and Michael Young each read a working paper that became the core of Chapters 1 and 5. We are very grateful for their comments and suggestions, which ultimately convinced us to try a longer work. Stanley Feldman and Mark Schneider each answered questions that had us puzzled while we were writing the first draft. The series reviewers provided detailed suggestions and highly constructive criticism on the first draft, which led to several structural improvements. We owe a huge debt to the many friends and colleagues who helped us with earlier work. Although they remain anonymous here, they are acknowledged in our earlier computational modeling papers on foreign policy decision making (POLI), candidate evaluation (CANDI), social choice (the "dot program"), and foreign policy event interpretation (EVIN). We are particularly grateful to Philip Schrodt. Most of our published work on computational modeling, including this monograph, has benefited tremendously from the quality and quantity of his comments. The National Science Foundation supported our research directly under grants SES-9102901 and SES-9310351, and indirectly under grant DIR-9113599 to the Mershon Center Research Training Group on the Role of Cognition in Collective Political Decision Making at The Ohio State University. Finally, we owe special thanks to our families for their continuing support: Sue, Kathleen, and Christopher.

Series Editor's Introduction

Progress in the social sciences can be measured by the acquisition of more and better tools. Since the 1960s, perhaps the most important new tool has been the computer. In the Quantitative Applications in the Social Sciences monograph series, we have titles directly relevant to computer users: *Microcomputer Methods for Social Scientists, 2nd Edition* (No. 40, Schrodt), *Using Microcomputers in Research* (No. 52, Madron, Tate, & Brookshire), and *Computer-Assisted Interviewing* (No. 80, Saris). More broadly, the quantitative procedures featured in all the monographs almost always require a computer. Computers are able to do more than provide data storage and statistical manipulations. As Professors Taber and Timpone observe, computers themselves have an analytic role in what the authors call computational modeling.

Professors Taber and Timpone organize disparate methods of computational modeling into three global types of models: dynamic simulation, knowledge-based systems, and machine learning. Behind each type is a social theory, to be written as a computer program. The first approach, dynamic simulation, is the oldest and most popular. A set of equations representing part of the world, for example, arms races, has its values systematically altered by the computer. The subsequent changes in outcomes then are observed. To the extent that the model reflects reality, the alterations inform us about what might actually happen under different conditions. To illustrate dynamic simulation, the authors work through a problem in bureaucratic politics and budget setting. They go on to explain cellular automata (a subset of dynamic simulations involving spatial representations).

The second type, knowledge-based systems, is part of the artifical intelligence (AI) tradition. Taber and Timpone examine semantic networks, frame systems, hybrid systems, and expert systems. (On the last, see also in this series *Expert Systems*, No. 77, by Benfer, Brent, & Furbee.) What holds these systems together is that they store information by computer in an effort to represent human knowledge. Semantic networks treat knowledge as a web of concept nodes and links. As an example, the authors explore how interviewees in a survey formulate their responses to political questions. Most frequently, AI work is based on expert systems. Knowledge about some specific area, say statistics or mountain climbing, is organized as a collection of "if-then" rules. The rules then are drawn on to help make decisions.

Machine-learning models differ from AI in that they are not committed to representing human knowledge. A common approach here is ID3, which aims to derive classification rules from the data themselves. The authors give the example of trying to account for different enrollment levels among ten high schools.

Computational modeling is a research method, one to be learned like any other. One of its advantages is that it forces theoretical precision. At the same time, it allows incorporation of numerous variables and values. In a certain sense, it permits a unique blend of the qualitative and the quantitative. Computational modeling will increase as computers become even more accessible and powerful. This monograph provides a good introduction to the many pathways that can be taken in this growing field.

MICHAEL S. LEWIS-BECK
SERIES EDITOR

1. Introduction

Beyond Platforms and On-Ramps

Computers have opened a new world in the few decades since their creation. Across all of society, as well as in academia, computers have revolutionized our lives. Social scientists use them primarily in two ways: as platforms for programs that facilitate general tasks such as word processing and statistical analysis and, increasingly, as on-ramps for the information superhighway, which has greatly increased the speed and volume of scholarly discourse. These two areas do not exhaust the utility of computers in the social sciences: Computers offer more than platforms and on-ramps. They can serve as special-purpose analytic tools.

The social sciences have seen a rising trend in the use of computational modeling over the last three decades. For several reasons, we believe that this trend will accelerate. First, computers are far more powerful, reliable, and affordable than ever before. Contemporary personal computers based on Pentium or RISC processors have similar capabilities to the supercomputers of a few years ago. Second, because of advances in operating systems, software packages, and programming tools, computers are now much easier to use. First-generation computational modelers programmed directly in machine language, which in some cases meant that they had to physically plug and unplug circuits. Although technical skill still helps in computational modeling, programming is now far less arcane, and special software packages bring computational modeling even closer to nonprogrammers. Third, there also have been many important advances in computational modeling methods, the topic of this monograph. Specialpurpose methods, often woven from a particular theoretical framework, have been developed for classes of modeling problems. For example, decision making can be modeled using information-processing methods developed in cognitive and organizational science, and adaptation can be modeled using genetic algorithms from math and computer science. In short, computational modeling will grow in importance for social scientists because computers and computer methods are far more powerful and accessible than ever before.

The increase in computational modeling that we anticipate will depend most directly on how much practical utility social scientists see in these methods for the problems that interest them. In this monograph, we treat computational models as a single broad class of methods and discuss their merits in general terms. By contrast, social scientists (as opposed to computer scientists or engineers) generally treat computer-based methods piecemeal, not seeing them as part of a single framework. The piecemeal approach, unfortunately, diffuses recognition of the general utility of computational models and deflects consideration of their potential contribution to the growth of theoretical knowledge. It is instructive to note that the domain of statistical analysis similarly subsumes a wide variety of different tools, all of which spring from a single broad set of unifying assumptions, and that the utility of statistics for social scientists was recognized only after the epistemological framework of the behavioral revolution had taken firm hold (and after computers and software made the tools generally available). We therefore take a broader view of computational modeling than those who would apply the label only to a single method, such as dynamic simulation or artificial intelligence. In the following chapters, we will tour a wide variety of different approaches, divided into three general categories: dynamic simulation (e.g., Monte Carlo methods), knowledge-based models (semantic networks, frame systems, and rule-based systems), and machine-learning models (connectionism, ID3, and genetic algorithms). In this chapter and in the remainder of the monograph, we will focus on their similarities.

Models and Computational Models

At the risk of offending some philosophical purists (see, e.g., Brodbeck, 1969), we will use the terms "model" and "theory" rather loosely, defining a model to be a representation of a theory about some real-world phenomenon. Models are integral in the development of theoretical understanding. They fill the gap in the essential interplay between theory and data, allowing empirical regularities to guide theory (induction) and allowing theory to guide empirical analysis (deduction).

To express theoriesthat is, to build modelswe have three general languages to choose from (Ostrom, 1988). These are natural language, various dialects of mathematics (including statistics and logic), and computational symbolic processing. Although one may reasonably argue that all complete formal languages are intertranslatable, making computational symbolic processing and classical mathematics formally interchangeable (Turing, 1950), that would obscure an important practical distinction: Computational symbolic processing allows us to do many things that are only theoretically possible without the aid of a computer. Conversely, some things are done more easily mathematically, including theorem proving and anything dealing with infinities. Worse, such an argument would bog us down in painful philosophical debate, and our basic point would remain intact: Symbolic manipulation on a computer, like math or English, provides a general language for expressing scientific theories about the world.

Computational models, then, are theories rendered as computer programs. For example, McPhee (1963) developed a model of social influence on voting behavior that represents theoretical processes such as "discussion" and "learning" as symbolic relationships in a computer program. The implications of the entire theoretical structure then can be drawn by literally "running the model" and observing how it behaves. More generally, computational modeling entails developing a process theory, expressing this theory as a computer program, and simulating the theory by running the program. The seemingly distinct tools described in Chapters 2 through 4 generally follow this basic pattern of analysis. Moreover, when expressing and simulating theoretical models through the medium of computer programs, similar concerns of validation arise (Chapter 5).

Why Model Computationally?

There are many forms of analysis in the social sciences, including qualitative case studies, quantitative data analyses, and mathematical models. In our view, scientific knowledge can accumulate through careful research along any of these paths (cf. King, Keohane, & Verba, 1994). No single approach can suffice for all research problems in the social sciences. Indeed, many problems require multimethod research. Computational modeling is another approach that can be added to this list. Why do so?

A variety of advantages of formal over natural-language models have been documented (Fiorina, 1975), all of which pertain to both mathematical and computational work: (a) the definitional and conceptual precision of the model, (b) the clarity of assumptions, (c) the ease of deciding on its internal or logical validity, (d) the power of formal deduction, and (e) the relatively unambiguous communication among scientists using formal language. For Fiorina (1975, p. 138), "the major advantage of using formal models is the precision and clarity of thought which these models require, and the depth of argument which they allow." He quickly acknowledges the main limitation of mathematical models: "Doubters would argue that intellectual clarity is purchased at too dear a price, that it involves simplifying reality beyond all recognition." In subsequent paragraphs, he unhappily endorses the goal of predictive accuracy as paramount in scientific explanation: "A modeler tends to brush off criticisms that his models are hopelessly unrealistic. With no real expectation of finding the one, true explanation, he settles for one which works, i.e., predicts more accurately than anything else available." This was a reasonable response, in our view. Predictive accuracy and analytic focus cannot be sacrificed. Fortunately, however, computational methods allow scientists to model rigorously the complex process mechanisms that they may theorize.

This is the major advantage of computational over mathematical modeling: It greatly increases the level of realism one may incorporate in a formal model without sacrificing analytic focus. In research areas in which formal models become intractable, the computational approach allows one to aspire to achieve both predictive and process validity. We want to be very clear, however, that when research problems that account for process are solvable within a mathematical model, computational modeling may be unnecessary.

Some readers may be skeptical, however, believing that we are confusing descriptive accuracy with theoretical explanation. Our main argument for computational modeling, after all, is that it allows us to increase the descriptive realism of our models. Nevertheless, computational models remain models and are simplifications. We see a qualitative difference between a simplified representation of a real-world process and a model that treats the process as a black box, making no attempt at representing the real-world process. The former is a caricature with recognizable features and may facilitate theoretical explanation; the latter is not even a caricature. Most of the computational methods described in this monograph are designed to represent theoretical processes, though they vary in the "depth" of process detail they can express. For those who seek processvalid explanations, computational modeling can help with the added theoretical complexity.

Because our argument hinges on the distinction between accurate prediction and the understanding of process, a digression on this point may be helpful. Prediction focuses on the output of a theoretical model, treated as a black box, that may be compared to factors in the real world; explanation, in contrast, focuses on the model itself, illuminating the formerly black box so that we can examine the process mechanisms that produce the output. Most sciences have gone through phases corresponding to the different goals of predictive success and explanation. Dissatisfaction with pure prediction usually arises when it becomes clear that basic assumptions are false or when several quite different models accurately predict the data. At this point, controversy usually erupts between those who are interested in learning more about the actual mechanisms involved and those fully satisfied with correct predictions. Astronomers in the late 16th century, for example, fought this battle. Most Copernicans were satisfied with the highly unlikely system of interlocking epicycles that they had fashioned to account for the irregular planetary orbits they observed. After all, the system worked, in that it was reasonably accurate in predicting planetary paths and compiling navigational charts. What did it matter that the mechanisms for this calculation were "geometrical fictions" (Koestler, 1959, p. 171)? Johannes Kepler, however, could not accept this view and devoted his life to uncovering the processes actually at work in planetary motion. He sought a deeper understanding, and in retrospect, the science of astronomy could not have advanced without adopting his ethic. Predictive success was necessary but was insufficient for explanation.

Past work, across the social sciences, has demonstrated the merits of computational modeling (Hastie, 1988; Taber & Timpone, 1994a). First, although computational models, like mathematical models, enforce precision and clarity of thought, they allow for theoretical uncertainty when necessary. It can be very difficult to deal with the inevitable holes in our theoretical understanding within a tractable mathematical model. In contrast, computational models are "cheap" and highly flexible. When faced with unknown pieces of the puzzle, a computational modeler can fill the holes temporarily with random variables, or, with relatively little additional cost in time or effort, one may simulate several hypothetical subprocesses to fill the holes and examine their competing implications.

Second, computational methods are more versatile than other formal approaches. Theoretical concepts that would be very difficult to express using standard mathematics (e.g., semantic networks in cognitive models) can be