INTRODUCTION TO ACTUARIAL MODELING

James C. Hickman*

1. MODELS

Because of differences in our education and experiences, each of us probably conjures up a somewhat different concept when models are mentioned. Not only do we define models differently, but also in practice we use different methods for constructing models and for testing their degree of realism. Despite these differences in definition and applications, a strong case can be made that the selection, estimation, and verification of models is the essence of science. Models are essential tools of thought and our principal vehicles for scientific communication.

Jewell (1980) wrote the most comprehensive review of models in insurance. In his remarkable essay, Jewell starts with a definition:

> A model is a set of verifiable mathematical relationships or logical procedures which is used to represent observed, measurable real-world phenomena, to communicate alternative hypotheses about the cause of the phenomena, and to predict future behavior of the phenomena for the purpose of decision making.

Using Jewell's definition, the conference sponsored by the Society of Actuaries with the title "Actuarial and Financial Modeling: Toward a New Science," held at Georgia State University on December 16 and 17, 1996, was directed at the core of actuarial science. The only discontinuity between the conference title and Jewell's definition is the phrase "Toward a New Science," because, according to the definition, modeling has been at the center of actuarial science and, indeed, of all the sciences from the beginning. Perhaps "New Classes of Models" or "A Broader Domain for Models" would have been a less dramatic but more realistic subtitle.

Models are approximations to reality. We can use them for play or for learning, as in model airplanes, trains, and dollhouses. We can use them for adult purposes, for example, in the case of model airplanes used in wind tunnels to study flight characteristics or model bridges used to test the safety of a planned structure. The models used by actuaries are typically expressed as systems of equations and logical statements rather than diminutive structures, but they are approximations nonetheless.

Actuaries, along with other scientists, face the same threat as Pygmalion. In Greek mythology the sculptor Pygmalion fell in love with his statue. The econometrician Henri Thiel said, "Models are to be used but never believed." There will always remain a distinction between a model and the reality it represents. A scientist who becomes emotionally attached to the model will miss deeper insights into reality.

The distinctions between reality and models can be sharpened by introducing additional examples. Models are used to represent systems. The best example is the solar system and the long history of models that have been used to represent it. The astonishing success of celestial mechanics in predicting the behavior of the solar system has set a standard of predictability that is impossible for models of more open systems to attain. Our standards for models might not be so high if the center of the solar system were twin stars rather than a single dominant sun.

At the other extreme in size is the atom and the model for it, under constant revision, constructed in physics. For much of physics, no one would consider an atom a closed system, and all of chemistry is devoted to the development of models for the interactions among atoms.

The systems that most scientists study are open; that is, forces from outside can alter the system. The universe can be described as the singular closed system. Actuaries should not grieve that their models are incomplete because the systems that they model are open.

Jewell's definition of a model seems to limit the scope of models to "mathematical relationships on logical procedures." The previous examples of model airplanes, trains, and bridges demonstrate that physical or analogue models can be useful. Physical models are not common in actuarial science, although the plumbing diagram used by Trowbridge and Farr (1976) to explain pension funding could have been constructed.

2. THE PROCESS

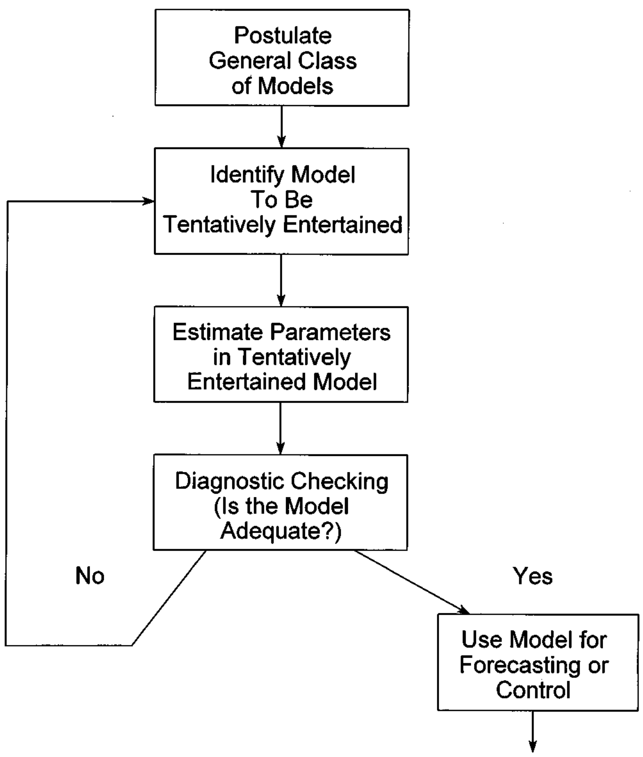
There is no standard process for conducting science, and consequently there is no standard process for

*James C. Hickman, F.S.A., A.C.A.S., Ph.D., is Emeritus Professor and Dean, University of Wisconsin, School of Business, 975 University Avenue, Madison, WI 53706.

modeling. Nevertheless, to provide a framework for our discussion we need a model of modeling. We will adopt a modeling process chart used by Box and Jenkins (1970) to establish a systematic process for constructing time series and other statistical models (see Figure 1). In applied statistics, the general class of models in this interactive approach might be the class of autoregressive integrated moving average time series models or the class of linear regression models. In mechanics, the general class might be linear differential equations.

FIGURE 1

STAGES IN THE ITERATIVE APPROACH TO MODEL BUILDING



A model is tentatively identified by using a priori information or data analysis with emphasis on graphical displays. Statistical models are estimated by using data and methods such as least squares or maximum likelihood. Differential equation models also require the estimation of parameters. For example, the gravitational constant must be estimated from data in simple models for falling bodies.

Models are checked by examining deviations between observed results and model predictions. It is most informative if the checking can be done with data not used in the estimation stage. Checking is the key step in the process and can lead to the tentative acceptance and use of the model or a return to the identification stage. Model checking is also the stage in which new hypotheses and classes of models are born. One enormous advantage of statistical models is the battery of tests of model deviations or residuals that are available and the existence of probability indices for measuring the importance of apparently large residuals.

In the checking stage, there is the option of applying the precept that terms, concepts, and assumptions should not be multiplied beyond necessity. This precept is called Occum's razor, for the 14th-century philosopher. The operation of the razor is quite explicit in statistical modeling, in which the modeler starts with a simple model and adds parameters, assumptions, and concepts as needed. The process stops when checking indicates that deviations behave like independent random phenomena and do not contain additional modeling information. To continue modeling is to risk building on sand.

In statistics Occum's razor is called the principle of *parsimony*.

When studying a modeling process, our iterative approach requires answers to four questions:

1. What is the general class of models? No one can master the class of all models.
2. How is a particular member of the general class identified? Revelation can shortcut the identification stage, but its use should be acknowledged.
3. How are the model parameters estimated? If the parameters are not based on data, that fact should be acknowledged.
4. How is the model checked with reality? The next stage in the process is determined by information revealed by the answer to this question.

3. ACTUARIAL MODELS

Several of the models presented at the conference can be classified as natural extensions of the individual risk theory model that has been under development for three centuries. It is instructive to list the components that have entered individual risk models in approximately the time order, from earliest to most recent, that they entered:

(a) Mortality
(b) Interest
(c) Expenses
(d) Withdrawals
(e) New entrants
(f) Mid-course correction such as dividends, premium adjustments, or capital infusions

(g) Exhibiting distributions of results rather than a single point value
(h) Asset cash flows
(i) Interrelationships among the variables
(j) Placing the model within a macro-model that relates to interest rates, expenses, and new entrants.

This path of development has resulted in ever more comprehensive and realistic models for life insurance or pension operations, but in no sense has the path led to an ultimate self-contained model. Within this historic view, we are not compelled to use all the components in every model. For example, benefit reserves use only components (a) and (b), though benefit reserves remain useful even if not completely realistic.

Progress in building comprehensive models within the general class of individual risk models was faster than is sometimes assumed. For example, a paper by Dow (1975) described the work of the mathematician Colin Maclaurin in connection with the organization of the Scottish Ministers' Widows' Fund in 1743. Maclaurin's projection model required assumptions about mortality, interest, new entrants, marriage rate, and the number of children.

Rather than extending the list of components incorporated into a model, at times it may be necessary to start again. The collective risk model arose from dissatisfaction with the static nature of the earlier individual risk model.

4. NEW CLASSES OF MODELS

Instead of summarizing the papers presented at the conference, it seems preferable initially to comment on the general classes of models that were used and that, at least at this time, are somewhat novel in actuarial science.

4.1 Nonlinear Dynamics

Nonlinear dynamical models are not new. The chaotic behavior of nonlinear systems was well-known to Poincare (1854–1912) and his students. Nevertheless, in recent years the study of nonlinear dynamical systems has accelerated both within mathematics and the empirical sciences. The appeal of the popular title "chaos theory" has encouraged the trend. Tilley (1989) provided an introduction to the topic, and Craighead (1994) supplied a case study of applying the techniques of nonlinear dynamics to an economic time series of interest to actuaries. Craighead's approach replicated the stages in the model-building process described in Section 2. This is true despite the fact that the class of models and the tools for identification, estimation, and checking are different from those used in applied statistics.

Richard H. Day presents an example of deductive model-building that culminates in a nonlinear dynamical model ("Complex Dynamics Market Mediation and Stock Price Behavior," p. 6). He starts with a simple and appealing model for a market in which mediators, or go-betweens, facilitate the operation of the market by supplying demanders from their inventories at announced prices and then replenish their inventories by purchasing from suppliers again at announced prices. Day then specializes the model to approximate a market for equities. Plausible parameter values are selected, and the model is used to simulate the operations of an equity market. The resulting series of simulated market prices has many of the observed characteristics of real equity markets. Roberts (1959) used a similar demonstration to support a random walk model of equity prices. Roberts' demonstration used realizations of independent and normally distributed random variables to produce series that had characteristics that are used in technical stock market analysis.

Day leads the reader through an example of the classical deductive or mathematical approach to model-building. He starts with plausible assumptions and rather simple functions as building blocks. The result is a nonlinear dynamical model that produces time series of deterministically generated equity prices that appear realistic. The final diagnostic checking of the process that I described in Section 2 is not done.

The ability of both Roberts, using random methods in 1959, and Day, using a deterministic nonlinear dynamical system in 1996, to produce simulated time series of equity prices that appear realistic is a warning signal to modelers. It indicates our incomplete understanding of the words "deterministic" and "probabilistic" and the crude nature of our identification tools.

4.2 Neural Networks

Paul Gorman presented a comprehensive case study at the conference that employed a very general class of models and a variety of identification and estimation tools. The objective of his study was to develop a model to assist in the selection of credit card applicants so that management can balance profitability and risk. The usual objective of credit risk studies has been to minimize the probability of default.

Gorman's paper is not contained in this collection, but at least one of the classes of new models that he adopted has appeared in the actuarial and insurance literature. Therefore, neural networks deserve a place on a list of new classes of actuarial models.

Neural networks have received a great deal of attention in recent years in applications in which classification is the objective. Fitting neural networks is very computer-intensive. The algorithm that constructs and fits the model is designed to mimic the operation of the human brain. Brockett et al. (1994) provided an insurance application of neural networks in which the objective was to generate early warnings of insurance company insolvencies.

There are important connections between neural networks and statistical regression models. These connections, and a valuable reference list, were provided by Warner and Misra (1996). They indicated quite clearly the advantages and disadvantages of each class of models.

Regression models require the selection of a functional form of the response function but yield a reward by permitting statistical tests of individual coefficients. Neural networks require fewer restrictions on the functional relations among the variables, but exact a price for this freedom by concealing the exact estimated functional relationship and by a paucity of checking tools.

The overriding question in this comparison is, What is the role of Occum's razor, the principle of parsimony, in a computer age? Maybe it is no longer desirable to be parsimonious.

4.3 Asymptotic Order Statistics

In his discussion (p. 21) of Day's paper Irwin Vanderhoof provides an articulate criticism of the use of stationary (unchanging over time) statistical models in modeling financial data. He presents powerful criticisms based on large deviations between observations and predicted results. Papers by Becker (1991) and Klein (1993) provided data analysis that support Vanderhoof's criticisms.

Vanderhoof also mentions the possible use of the three distributions that arise in studying the distributions of the largest- and smallest-order statistics derived from a random sample as the sample size increases. These three distributions have appeared in actuarial literature when the application is mortality measurement. The motivation for their use in mortality studies is the idea that a structure or organism is composed of a large number of systems, each of which has a time-until-failure random variable. Furthermore, these several random variables are mutually independent. Only the smallest of the realizations of these several random variables can be observed. A paper by Brillinger (1961) and its discussions provided an introduction to these ideas.

5. IMPLEMENTATION

Angus Macdonald describes the history and current state of modeling activity in insurance in his paper ("Current Actuarial Modeling Practice and Related Issues and Question," p. 24). The paper contains an outline of a comprehensive insurance company model that he has developed. The model is appropriately denoted MO for model office.

Macdonald makes the superb point that the improvement of insurance models does not necessarily come by adding detail within the existing structure. Real progress may require a new set of general models. The incorporation of new classes of models will require that actuaries cut through the barriers that have separated actuarial thought from developments in related fields. The most persuasive current example is provided by the necessity of incorporating ideas from financial economics into actuarial models.

Stephen Strommen, in his discussion (p. 35) of Macdonald's paper and in his individual contribution ("An Object-Oriented Design for Dynamic Simulation Models," p. 38), discusses a number of model implementation issues. He argues that computation tools are not just minor technicalities but in fact are at the heart of the implementation problem. First there is a business problem; then a theoretical framework for considering the problem; and then computation tools for using the resulting model to examine the problem.

The complexity of current models has created new professional responsibility issues for actuaries. Strommen and Macdonald both examine these issues. The discussions center on communication issues in describing complex models to users and the responsibility of producers of models to their customers. The resolution of these issues includes the traditional response to most challenges—education—and the possibility of professional standards for models and organized peer review of models.

Strommen goes further and illustrates a method called *object-oriented design* as a means for organizing a company's dynamic simulation modeling process. The resulting model is traditional, but the steps required to build the logical structure and to create a computer program follow a systematic design.

Modelers in the 1990s can continue to profit by reading the report by Wooddy et al. (1981). The report describes the development of a very comprehensive life insurance company model designed to assist in the determination of contingency margins for reserve calculations for financial accounting purposes.

6. SUMMARY

The current state of modeling in insurance was examined at the conference. Several new general classes of models were illustrated. The conclusion was reached that progress will come not by refining existing models, but by breaking barriers and incorporating new classes of models and their associated estimation and checking methods. Some of these new classes were identified.

REFERENCES

BECKER, D.N. 1991. "Statistical Tests of the Lognormal Distribution as a Basis for Interest Rate Changes," *Transactions of the Society of Actuaries* XLVIII: 7–57.

BOX, G.E.P., AND JENKINS, G.M. 1970. *Time Series Analysis: Forecasting and Control*. San Francisco, Calif.: Holden-Day.

BROCKETT, P.L., COOPER, W.W., GOLDEN L.L., AND PITAKONG, V. 1994. "A Neural Network for Obtaining an Early Warning of Insurer Insolvency," *Journal of Risk and Insurance* 61: 402–24.

BRILLINGER, D.R. 1961. "A Justification of Some Common Laws of Mortality," *Transactions of the Society of Actuaries* XII: 116–19.

CRAIGHEAD, S. 1994. "Chaotic Analysis on U.S. Treasury Interest Rates," *ARCH* 1994.1: 281–318.

DOW, J.B. 1975. "Early Actuarial Work in Eighteenth-Century Scotland," *Transactions, Faculty of Actuaries* 33, part 3: 193–229.

JEWELL, W.S. 1980. "Models in Insurance: Paradigms, Puzzles, Communications and Revolution," *Transactions, 21st International Congress of Actuaries*, Vol. S: 87–141.

KLEIN, G.E. 1993. "The Sensitivity of Cash-Flow Analysis to the Choice of Statistical Model for Interest Rate Changes," *Transactions of the Society of Actuaries* XLV: 79–124.

ROBERTS, H.V. 1959. "Stock-Market 'Patterns' and Financial Analysis: Methodological Suggestions," *Journal of Finance* 14: 1–10.

TILLEY, J.A. 1989. "Whither Actuarial Science? Borrowing Concepts and Mathematics from Other Disciplines," *1989 Centennial Celebration Proceedings* 2: 523–42.

TROWBRIDGE, C.L., AND FARR, C.E. 1976. *The Theory and Practice of Pension Funding*. Homewood, Ill.: R.D. Irwin, Inc.

WARNER, B., AND MISRA, M. 1996. "Understanding Neural Networks as Statistical Tools," *American Statistician* 50: 284–93.

WOODDY, J.W. ET AL. 1981. *Adverse Deviation*. Schaumburg, Ill.: Society of Actuaries.