PART I

INTRODUCTION

1

MODELING

1.1 The model-based approach

The model-based approach should be considered in the context of the objectives of any given problem. Many problems in actuarial science involve the building of a mathematical model that can be used to forecast or predict insurance costs in the future.

A model is a simplified mathematical description that is constructed based on the knowledge and experience of the actuary combined with data from the past. The data guide the actuary in selecting the form of the model as well as in calibrating unknown quantities, usually called *parameters*. The model provides a balance between simplicity and conformity to the available data.

The simplicity is measured in terms of such things as the number of unknown parameters (the fewer the simpler); the conformity to data is measured in terms of the discrepancy between the data and the model. Model selection is based on a balance between the two criteria, namely, fit and simplicity.

1.1.1 The modeling process

The modeling process is illustrated in Figure 1.1, which describes the following six stages:
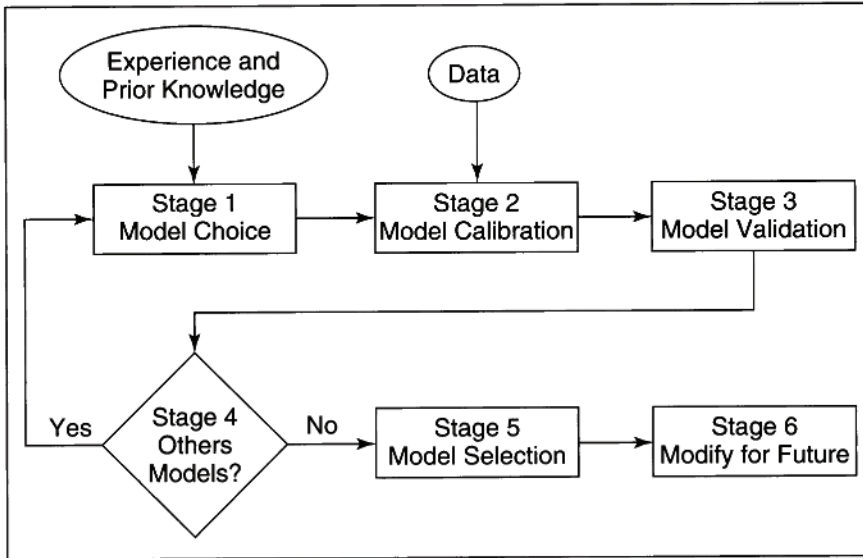

Figure 1.1   The modeling process.

Stage 1 One or more models are selected based on the analyst's prior knowledge and experience and possibly on the nature and form of available data. For example, in studies of mortality, models may contain covariate information such as age, sex, duration, policy type, medical information, and lifestyle variables. In studies of the size of insurance loss, a statistical distribution (e.g., lognormal, gamma, or Weibull) may be chosen.

Stage 2 The model is calibrated based on available data. In mortality studies, these data may be information on a set of life insurance policies. In studies of property claims, the data may be information about each of a set of actual insurance losses paid under a set of property insurance policies.

Stage 3 The fitted model is validated to determine if it adequately conforms to the data. Various diagnostic tests can be used. These may be well-known statistical tests, such as the chi-square goodness-of-fit test or the Kolmogorov–Smirnov test, or may be more qualitative in nature. The choice of test may relate directly to the ultimate purpose of the modeling exercise. In insurance-related studies, the total loss given by the fitted model is often required to equal the total loss actually experienced in the data. In insurance practice this is often referred to as *unbiasedness* of a model.

Stage 4 An opportunity is provided to consider other possible models. This is particularly useful if Stage 3 revealed that all models were inadequate. It is also possible that more than one valid model will be under consideration at this stage.

Stage 5 All valid models considered in Stages 1–4 are compared using some criteria to select between them. This may be done by using the test results previously obtained

or may be done by using another criterion. Once a winner is selected, the losers may be retained for sensitivity analyses.

Stage 6 Finally, the selected model is adapted for application to the future. This could involve adjustment of parameters to reflect anticipated inflation from the time the data were collected to the period of time to which the model will be applied.

As new data are collected or the environment changes, the six stages will need to be repeated to improve the model.

1.1.2 The modeling advantage

Determination of the advantages of using models requires us to consider the alternative: decision making based strictly upon empirical evidence. The empirical approach assumes that the future can be expected to be exactly like a sample from the past, perhaps adjusted for trends such as inflation. Consider Example 1.1.

■ **EXAMPLE 1.1**

A portfolio of group life insurance certificates consists of 1,000 employees of various ages and death benefits. Over the past five years, 14 employees died and received a total of 580,000 in benefits (adjusted for inflation because the plan relates benefits to salary). Determine the empirical estimate of next year's expected benefit payment.

The empirical estimate for next year is then 116,000 (one-fifth of the total), which would need to be further adjusted for benefit increases. The danger, of course, is that it is unlikely that the experience of the past five years accurately reflects the future of this portfolio as there can be considerable fluctuation in such short-term results. □

It seems much more reasonable to build a model, in this case a mortality table. This table would be based on the experience of many lives, not just the 1,000 in our group. With this model we not only can estimate the expected payment for next year, but we can also measure the risk involved by calculating the standard deviation of payments or, perhaps, various percentiles from the distribution of payments. This is precisely the problem covered in texts such as *Actuarial Mathematics for Life Contingent Risks* [26] and *Models for Quantifying Risk* [23].

This approach was codified by the Society of Actuaries Committee on Actuarial Principles. In the publication "Principles of Actuarial Science" [104, p. 571], Principle 3.1 states that "Actuarial risks can be stochastically modeled based on assumptions regarding the probabilities that will apply to the actuarial risk variables in the future, including assumptions regarding the future environment." The actuarial risk variables referred to are occurrence, timing, and severity—that is, the chances of a claim event, the time at which the event occurs if it does, and the cost of settling the claim.

1.2 Organization of this book

This text takes the reader through the modeling process, but not in the order presented in Section 1.1. There is a difference between how models are best applied and how they are best learned. In this text we first learn about the models and how to use them, and then we learn how to determine which model to use because it is difficult to select models in

a vacuum. Unless the analyst has a thorough knowledge of the set of available models, it is difficult to narrow the choice to the ones worth considering. With that in mind, the organization of the text is as follows:

1. Review of probability—Almost by definition, contingent events imply probability models. Chapters 2 and 3 review random variables and some of the basic calculations that may be done with such models, including moments and percentiles.

2. Understanding probability distributions—When selecting a probability model, the analyst should possess a reasonably large collection of such models. In addition, in order to make a good a priori model choice, characteristics of these models should be available. In Chapters 4–7 various distributional models are introduced and their characteristics explored. This includes both continuous and discrete distributions.

3. Coverage modifications—Insurance contracts often do not provide full payment. For example, there may be a deductible (e.g., the insurance policy does not pay the first $250) or a limit (e.g., the insurance policy does not pay more than $10,000 for any one loss event). Such modifications alter the probability distribution and affect related calculations such as moments. Chapter 8 shows how this is done.

4. Aggregate losses—To this point the models are either for the amount of a single payment or for the number of payments. Of interest when modeling a portfolio, line of business, or entire company is the total amount paid. A model that combines the probabilities concerning the number of payments and the amounts of each payment is called an *aggregate loss model*. Calculations for such models are covered in Chapter 9.

5. Review of mathematical statistics—Because most of the models being considered are probability models, techniques of mathematical statistics are needed to estimate model specifications and make choices. While Chapter 10 is not a replacement for a thorough text or course in mathematical statistics, it does contain the essential items needed later in this book.

6. Construction of empirical models—Sometimes it is appropriate to work with the empirical distribution of the data. It may be because the volume of data is sufficient or because a good portrait of the data is needed. Chapters 11 and 12 cover empirical models for the simple case of straightforward data, adjustments for truncated and censored data, and modifications suitable for large data sets, particularly those encountered in mortality studies.

7. Construction of parametric models—Often it is valuable to smooth the data and thus represent the population by a probability distribution. Chapters 13–15 provide methods for parameter estimation for the models introduced earlier. Model selection is covered in Chapter 16.

8. Adjustment of estimates—At times, further adjustment of the results is needed. When there are one or more estimates based on a small number of observations, accuracy can be improved by adding other, related observations; but care must be taken if the additional data are from a different population. Credibility methods, covered in Chapters 17–19, provide a mechanism for making the appropriate adjustment when additional data are to be included.

9. Simulation—When analytical results are difficult to obtain, simulation (use of random numbers) may provide the needed answer. A brief introduction to this technique is provided in Chapter 20.