Parameter estimates depend both on the source model and on the fitted model: An example

CHARLES E. COLLYER
University of Rhode Island, Kingston, Rhode Island

Distributions of estimates of the threshold for mental rotation under two fitted models were collected from computer simulations of three source models and from human data. The estimates depend both on the source of the data and on the fitted model used for parameter estimation. The concept of a source model's "fertility" is illustrated by a characteristic of the human data shared by one of the source models.

Here is a statement with which virtually everyone will agree: the outcome of a measurement depends both on the properties of the thing being measured and on the procedure being used. This brief paper simply reiterates this idea, within the context of fitting simple mathematical models to data from experiments.

As stated in the title, parameter estimates depend both on the source model and on the fitted model. A *fitted model*, such as simple linear regression, is a set of operations that can be applied to empirical data and that results in a structure of predicted data. Fitted models are based on a specific theory of where the data came from; however, the set of operations should not be confused with the theory to which they relate. For example, simple linear regression is one of several ways of fitting a straight line: thus it is one of several fitted models relating to the same underlying theory about a straight line. *Parameter estimates*, such as the estimates of slope and intercept in simple linear regression, are measurements that can be interpreted in terms of the theory underlying the fitted model. Like other measurements, they may have more or less reliability and validity.

A *source model* is a process capable of generating empirical data. For example, the equation $y = a+bx$, together with some familiar rules of substitution and arithmetic, is a simple linear source model capable of generating values of y. Notice that this equation, which defines a source model, is not the same as the set of operations defining simple linear regression, a fitted model. The distinction between source and fitted models, even when both embody a common idea, such as the idea of a linear relationship, is useful for determining how to test and compare alternative theories.

One goal of research is to discover nature's source models, the processes that generate observed data. Fitted models are often compared in order to choose a candidate source model. For example, Fitted Model A and Fitted Model B can both be applied to a set of data; if Model A fits better than Model B, the researcher may de-

cide that a source version of Model A is more likely to have generated the data. One problem with this decision is that a model's goodness of fit depends on the complexity of the fitted model as well as on which source model actually generated the data. It is particularly likely that a complex fitted model will fit better than its competitors, even when the data source corresponds to one of the simpler models. This fact is familiar to users of multiple regression, log-linear, and related fitted-model methods.

The properties of both source and fitted models can be studied using Monte Carlo methods. In general, the goodness of fit of a fitted model tends to increase as the complexity of the fitted model increases, and it tends to decrease as the complexity of the source model increases (Collyer, 1985, 1986). In one area of modeling, multidimensional scaling, the interactive influences of source model and fitted model on goodness of fit have been termed the *elbow effect* (see Wagenaar & Padmos, 1971).

Goodness of fit is only one aspect of model performance; parameter estimates derived from model fitting represent another potential area of investigation. The purpose of the present study is to show how the estimates of one theoretically important parameter in an earlier analysis (Collyer, 1985) depended both on the source model and on the fitted model.

The parameter of interest in this demonstration is the threshold for mental rotation. In the earlier analysis, two threshold models, one with three free parameters and one with four, were compared with a two-parameter linear model of the mental rotation response time function. In the present analysis, source versions of all three models were used to generate data. Distributions of the hypothetical threshold for mental rotation were compiled by applying fitted versions of both threshold models to each set of data. Data from an experiment with human subjects (Rossi & Collyer, 1986) were analyzed similarly and compared with the results of the simulations.

METHOD

Each of the source models was specified by an equation giving the hypothesized structure of a correct response time. Model I was a two-parameter linear model:

Address correspondence to Charles E. Collyer, Department of Psychology, University of Rhode Island, Kingston, RI 02881-0808.

$$RT = 14.4A + 1034 + e, \qquad 0 < A < 180,$$

where A denotes the angular disparity of the two stimulus figures in degrees, and e is a Gaussian random error term with an expected value of zero and a standard deviation determined by the standard error of estimate in human data. Model II was a three-parameter threshold model:

$$RT = 1118 + e, \qquad 0 \leq A < 11.9;$$
$$RT = 13.7A + 1118 + e, \qquad 11.9 \leq A \leq 180.$$

Model III was a four-parameter threshold model:

$$RT = 1114 + e, \qquad 0 \leq A < 22.4;$$
$$RT = 11.0A + 1499 + e, \qquad 22.4 \leq A \leq 180.$$

The parameter values in these source models were determined by averaging the corresponding fitted-model parameter estimates from the human data. Thus each source model ostensibly simulates a typical subject operating according to that model's interpretation of the original experiment.

The techniques used for simulating mental rotation response times from the source models and for fitting models to the data were the same as in the earlier goodness-of-fit analysis of these models (Collyer, 1985). Data were obtained from 500 simulated subjects under each source model.

A description of the human experiment was given by Rossi and Collyer (1986). Briefly, this experiment consisted of a mental rotation task, in which there were many small angular disparities of the stimulus figures. Data were obtained from 20 subjects.

RESULTS AND DISCUSSION

Figure 1 shows the distributions of threshold estimates, with one graph for each data source and two distribution curves per graph (one for the three-parameter and one for the four-parameter fitted model). Cumulative distributions were used, so that the lateral position of the curve reflects the mean of the distribution and the slope of the curve reflects the variance.

Comparison of Fitted Models

In general, the threshold estimates from the four-parameter fitted model were greater than the estimates from the three-parameter fitted model. Note that in
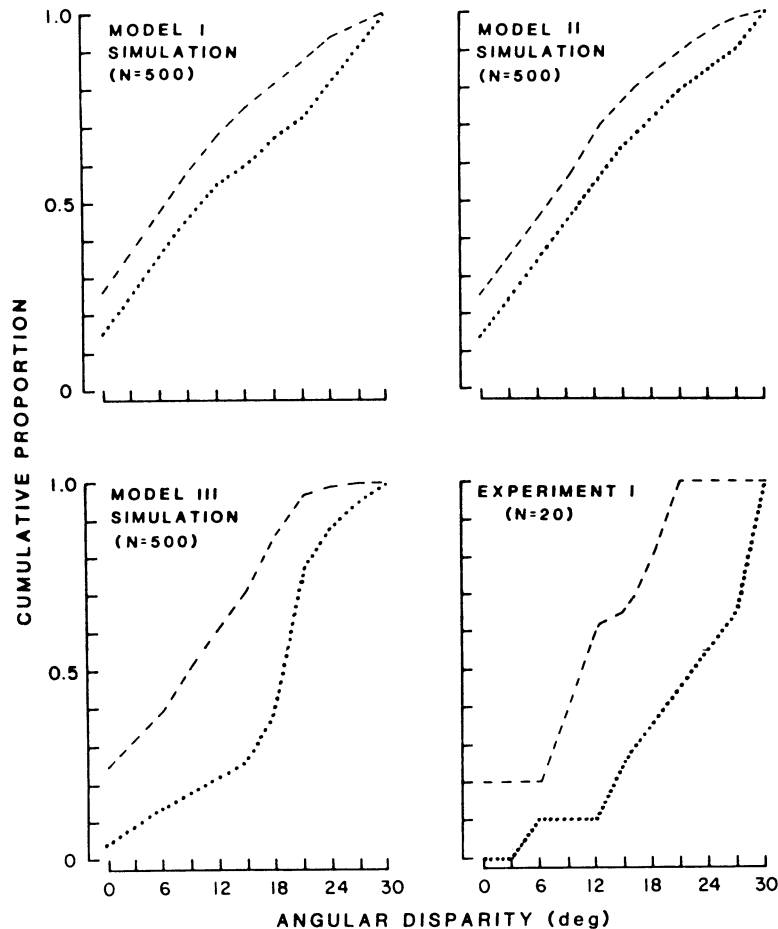

Figure 1. Cumulative distributions of estimates of the threshold for mental rotation. The dashed curves represent estimates obtained using the three-parameter fitted model, and the dotted curves represent those obtained using the four-parameter model. Response times of simulated subjects were generated under Source Models I, II, and III, which are described in the text. Response times were also collected from human subjects in an earlier study, here termed Experiment 1.

Figure 1, the curve for the four-parameter fitted model is consistently positioned to the right of the other curve. Thus, the outcome of an attempt to measure the threshold for mental rotation depends on which fitting operations—that is, which fitted model—the researcher selects.

Why should the threshold estimate from the more complex fitted model be consistently greater? The answer is probably related to the way in which parameters constrain each other in the simpler, three-parameter threshold model. In the simpler model, the regression intercept for suprathreshold response times is used to predict the subthreshold response time, so the model's goodness of fit tends to be best when the threshold is close to zero and the few subthreshold times are close to the intercept. In the more complex, four-parameter model, the subthreshold response time (which is the new fourth parameter) is estimated separately from the suprathreshold intercept, so that this constraint is relaxed. As a result, the threshold, as estimated by the four-parameter model, tends to seek a new, usually higher, value.

Influence of the Source Model

The threshold estimates under each fitted model were influenced by the source model in some cases but not in others. For the three-parameter fitted model (left curve in each graph), the mean threshold was fairly constant: it was 9.9° under Model I, 9.8° under Model II, and 10.2° under Model III. For the four-parameter fitted model (right curves), the mean threshold was 13.6° under Model I, 13.1° under Model II, and 18.3° under Model III.

Thus, threshold estimates increased with increases in the complexity of the fitted model, but they did not show a monotonic effect of source model complexity.

The pattern of threshold estimates was influenced by the source model. Data generated from simulations of Model I or Model II resulted in distributions of threshold estimates from the two fitted models that were quite close together, as shown by the small horizontal displacement of the curves in the upper two panels of Figure 1. However, data generated from simulations of Model III resulted in more distinct distributions, as shown by the larger horizontal displacement of the curves in the lower left panel of Figure 1. This larger separation of the threshold distributions was also characteristic of the human data (Experiment 1), as shown in the lower right panel of Figure 1.

The Concept of Fertility

The *fertility* of a source model can be defined as its ability to reproduce characteristics of human data. Thus, the four-parameter source model is fertile with respect to the distinctness of the threshold distributions obtained from the four-parameter and three-parameter fitted models. Fertility of a source model is not the same as goodness of fit, which applies to fitted models. For example, among the models under consideration here, there is no apprecia-

ble gain in fertility of the three-parameter source model over the two-parameter source model; however, the three-parameter fitted model has uniformly better goodness of fit than the two-parameter fitted model (Collyer, 1985).

Fertility is probably a necessary, but not a sufficient, condition for identifying the source model underlying human performance in a task. That is, the correct model would have to be fertile, but one can conceivably have two or more competing source models that are equally fertile with respect to the same (or different) aspects of performance. It is important to develop methods for choosing among such source models (a problem that arises in areas where modeling is done using computer simulation, such as artificial intelligence), just as it is important to develop methods for choosing among fitted models (a traditional problem in statistics).

Estimation Bias

The source threshold value of Model II (11.9) was underestimated by the corresponding three-parameter fitted model by about 18%. The source threshold value of Model III (22.4) also was underestimated by the corresponding four-parameter fitted model by about 18%. (Estimation bias—that is, under- or overestimation of a parameter estimate relative to its source value—is a fitted-model property.)

CONCLUSION

The word *model* has too many meanings for its use to be always clear. There are mathematical models of various kinds, animal models, role models, modeling as theorizing, modeling as imitation, computer models, model projects, box models, organizational models, prototype models, and model children. These varied meanings provide many opportunities for confusion. In particular, confusion between source and fitted models can easily arise, because both kinds of model are used in the same domain, that is, in theorizing about processes that explain data.

Both a source model and a fitted model may relate to the same theory, yet have different properties. Earlier analyses have shown how the goodness of fit of fitted models depends on both the complexity of the fitted model and the complexity of the source model, for both conventional model fitting (Collyer, 1985) and for cross validation (Collyer, 1986). The concept of a fitted model's sensitivity to variation in the source model was introduced in the earlier work, and it was suggested that sensitivity be demonstrated before using a fitted model to help identify source models.

The present paper has shown how the distribution of a parameter estimate likewise depends on both the source model and the fitted model. Estimation bias was viewed here as another fitted-model property. The concept of a source model's fertility was also introduced. An indicator of fertility is needed for each important aspect of the

performance to be simulated. It was suggested that fertility is a necessary, but not a sufficient, criterion for choosing a source model as a theory.

REFERENCES

COLLYER, C. E. (1985). Comparing strong and weak models by fitting them to computer-generated data. *Perception & Psychophysics*, **38**, 476-481.

COLLYER, C. E. (1986). Goodness-of-fit patterns in a computer cross-validation procedure comparing a linear and a threshold model. *Behavior Research Methods, Instruments, & Computers*, **18**, 618-622.

ROSSI, J. S., & COLLYER, C. E. (1986). Is there a threshold for mental rotation? *Bulletin of the Psychonomic Society*, **24**, 1-3.

WAGENAAR, W. A., & PADMOS, P. (1971). Quantitative interpretation of stress in Kruskal's multidimensional scaling technique. *British Journal of Mathematical & Statistical Psychology*, **24**, 101-110.