Sutton on Marshall's Tendencies

by Carl F. Christ
The Johns Hopkins University

Professor Sutton's thought-provoking book is directed principally to the question: "Are simple mathematical models helpful in economics, or are they misleading?" His answer, baldly stated, is that if they are made right they are helpful, but if they are made wrong they are misleading.

He has in mind an ideal reader: someone who already knows from other fields how a successful theory based on formal mathematical models works, but has only recently stumbled on economics, and is skeptical about formal mathematical models here. I am not quite an ideal reader now, but I was once, because I studied and worked in physics before going into economics.

Much of Sutton's discussion is based on a type of simple mathematical model which he calls the "standard paradigm". It is a model that meets the following conditions:

---the phenomena to be explained depend on a small number of important systematic explanatory factors
---all of these factors are included in the model
---all of these factors are measurable
---any other factors are small and unsystematic, and can be treated as random noise
---the systematic part of the model has a unique equilibrium.

The only explicit mathematical-statistical example Sutton gives of such a model is the following linear equation on page 17:

$$y_i = a_1x_{i1} + a_2x_{i2} + \dots + a_nx_{in} + \eta_i$$

where y is an endogenous variable to be explained, $x_1 \dots x_n$ are systematic exogenous factors, $\eta$ is the combined effect of the small unsystematic random factors, $a_1 \dots a_n$ are parameters to be estimated, and data are available for y and $x_1 \dots x_n$ for a number of periods or places, i = 1, 2, 3, ... .

This is a reduced form equation, not a structural equation, because it contains only the single endogenous variable y. Therefore this example ignores two important problems that arise in systems of simultaneous equations: (1) their parameters need to be estimated, preferably by methods not subject to least squares bias, and (2) before parameters can be estimated, their identifiability must be established. But let that pass, for it is not the main focus of Sutton's interest.

Sutton traces the origins of the standard paradigm to Alfred Marshall's analogy between the science of the tides and economic science. It is in Book I of the third edition of Marshall's Principles of Economics (1895). In the law of the tides, gravity is the important systematic factor, and the small unsystematic factors include variations in weather. In economics, the law of demand states the "tendency" for an increase in a commodity's price to cause a decrease in the amount demanded, provided that other things are equal (hence the title of Sutton's book). Here the systematic factor is the commodity's own price; the "other things" include small unsystematic factors, and may also include other systematic factors besides the commodity's price, such as demanders' income, and the prices of related goods. Sutton also notes the important roles played by Haavelmo's Probability Approach (1944) and by Samuelson's Foundations (1947).

Marshall's Principles includes graphs (relegated to footnotes) and equations (relegated to an appendix), but does not include explicit terms for random disturbances such as the $\eta$ in the above equation, and does not include statistical estimates of parameters. Regression and correlation were proposed and developed by Galton (1885), Edgeworth (1892), and Pearson (1896). Hence I would trace the standard paradigm to them at least as much as to Marshall.

Sutton discusses Robbins' view that economic theory is the most reliable means to knowledge in economics, and contrasts it with the business cycle chroniclers' view that measurement is the best means. He sides with Marshall's and Haavelmo's view that the best way forward is through the interplay of both theory and measurement. I am in hearty agreement with this. A theory can tell empiricists what to look for. And an empirical regularity can give theorists something to try to explain.

In Chapter 2 Sutton offers stock options and auction bidding as cases where the analogy of the tides is valid, so that the true model is known and the standard paradigm applies. Here the use of the standard paradigm leads to good numerical predictions. (To these cases one might add demand studies for many commodities.)

Sutton regards the standard paradigm as the best available investigative tool for most empirical economic problems (p. 86). But he regards the analogy of the tides and the standard paradigm as misleading in some cases (p. 5), because the above-listed conditions for it are not met.

> "... the economic environments we seek to model are sometimes too messy to be be fitted into the mold of a well-behaved, complete model of the standard kind." (page 32).

> "... the search for a true model becomes futile. The problem is that there are many 'reasonable' models ... To cut through this kind of difficulty, it seems appropriate to drop any notion of the true model, and to work instead in the less restrictive setting of a class of models." (page 70).

In this approach several different models are considered, each of which is plausible for some possible situation, and none of which is the true model for all situations. It may be unknown which, if any, of these models applies to any particular situation of interest, and indeed whether any of them is the true model for any situation. Sutton seems to regard the class-of-models approach as a departure from standard econometric procedure. I prefer to regard it as normal procedure. This can be explained as follows.

Let us adopt the Cowles Commission's definition (Koopmans, 1949) of an econometric structure, namely, a set of simultaneous equations with numerically specified parameters, capable of determining the numerical values of its endogenous variables, given the numerical values of its predetermined variables and of its stochastic disturbances. Let us admit not only equations, but also inequalities, which are now commonplace in Kuhn-Tucker analysis of optimization problems. And let us define a model as a set of structures. Then, in principle, any set of structures can be considered as a model. However, the most useful models are those whose member structures have certain features in common, such as the number of relationships (whether equations or inequalities), the list of included variables, the lists of endogenous and exogenous variables, and the mathematical forms of the relationships. Such a model might be a set of relationships without numerically specified parameter-values, or what is the same thing, a set of structures that differ only in the numerical values of their parameters. Then the model selection problem is the problem of deciding what variables to try to explain, what exogenous variables to include as explanatory factors, what mathematical form to give to each relationship, what properties to assign to the stochastic disturbances, and the like. The estimation problem is the problem of using data to estimate the values of the parameters of the selected model.

Using this terminology, a class of models is itself a model. Thus Sutton's proposed retreat to a class-of-models approach is the perfectly natural choice of a less specific model when one is not able to settle on a more specific model.

The standard paradigm is clearly too restrictive for many economic problems. Some mathematical relationships may be inequalities rather than equations, as alluded to above. Perhaps for this reason, the process under study may not have a unique equilibrium. There may be many systematic factors, not just a few, and some of them may be small. Some systematic factors may not be measurable, or may be measurable only with error. Random disturbances may enter non-additively. Parameters may not be constant. The mathematical forms of relationships may change. Perhaps for some phenomena equations and inequalities are not adequate representations. Sutton deals with many of these possibilities.

In Chapter 3 Sutton takes up the task of understanding certain empirical regularities observed in the field of industrial organization. In particular, he is interested in the relation between an industry's concentration ratio and the size of its market. He calls attention to three empirical findings: (1) industries with large markets typically have lower concentration ratios than industries with small markets; (2) for any given market size there is a rather wide range of concentration ratios; and (3) for some (but not all) industries there appears to be a positive limit below which the concentration ratio does not go, no matter how large the market. For this problem Sutton rejects the standard paradigm: either there is not a unique equilibrium concentration ratio for each market size, or, if there is, it depends on unobservable factors. So Sutton invokes the class-of-models approach. In this case it leads productively to what he calls a bounds approach. That is, for each market size, there is a lower bound to the concentration ratios of industries of that size. For one group of industries the bound approaches zero as the market size grows without limit. This is plausibly explained by a convergence theorem. For another group of industries, where R&D activity or advertising is important, the bound does not approach zero, but remains positive. This is plausibly explained by a non-convergence theorem.* He notes that a bounds approach is applicable only where the unsystematic factors all operate in the same direction.

Sutton says that if there is a lower bound to the concentration ratio as a function of market size, one would expect all observed points to lie above that lower bound. (This would be true if the lower bound were nonstochastic, that is, if there were no errors of measurement or other random factors affecting the observed concentration ratios.) Hence for estimating the lower bound, it would be inappropriate to use a regression, because that would yield a curve lying above many of the observed points, rather than below all or most of them. For this case, some procedure such as the following might be used:

(1) Construct the lower envelope of the observed points
(2) Compute the regression of the concentration ratio on market size
(3) Somehow form an estimate w of the average size of measurement errors and other random factors, relative to the spread of points above the lower bound
(4) Compute the weighted average, w times the regression plus (1 - w) times the lower envelope of the observed points

If w = 0, the result is the lower envelope itself, which is either the lower bound or an overestimate of it. If w is large, the result is a mixture of the regression and the lower envelope.

In Chapter 4 Sutton discusses testing of models by means of their predictions. This is the right way to go about it, in my view. The ensuing discussion will suppose that a model has been found and estimated that conforms acceptably to the data with which it was estimated.

A common model selection and testing procedure is this: (1) specify a model based, among other things, on knowledge of the

---

* In the explanation of the noncovergence theorem on pages 80-81 there appears to be a contradiction. The issue concerns the profit of a new entrant to an industry. About this entrant Sutton writes

"... its profit is at least equal to a(k)Sy [which is positive]. The stability condition requires that such an entrant does not achieve a positive profit, whence it follows that [the concentration ratio cannot fall below a positive lower bound.]"

This is a contradiction if "profit" means the same thing in both sentences. However, the argument makes sense if "profit" in the first sentence means profit <u>before</u> subtracting fixed cost, and "profit" in the second sentence means profit <u>after</u> subtracting fixed cost.

available data (nothing wrong with that, of course), (2) split the available data into two parts, (3) use one part to estimate the model's parameters, (4) use the estimated model to make predictions about the other part of the available data, and (5) judge the model based on how good these predictions are.

In my view this does not provide a very stringent test: Before specifying the model, the model-builder has access to the very data that will be used to test the model. Therefore it is likely that the model will be specified with an eye to explaining those data. There is a good chance that a model specified in this way will fit the accidental unsystematic features of those data, which will not be found in new data, as well as the systematic enduring features that will be found in new data. Thus such a model is likely to make better forecasts of those data than of new data that were not available when the model was specified. There can be no objection to specifying a model with an eye to explaining available data. The objection is to basing a test on the same data that helped inspire the choice of the model.

A more stringent test, therefore, is to find (or wait for) new data that were not yet available when the model was being specified, use the model to make predictions of the new data, and see how good these predictions are. I recall Haavelmo saying, not entirely in jest, that we should specify our models, and then go fishing for 50 years, and when we come back, check to see how well the models describe the intervening 50 years.

This objection, to testing by means of data that were already available when the model was specified, applies particularly to time series models. Here the model-builder is likely to be familiar with many features of the data for periods from some time in the past up to and including the present. The objection is less applicable to cross section studies involving data for hundreds or thousands of individuals, families, firms, regions, or the like, because the model builder is less likely to be familiar with many features of the part of the data that is to be reserved for testing.

If a model is tested in such a stringent manner, the best that can be expected is that it conforms to the new data as well as it does to the data which inspired its specification. The more new data that a model can predict or explain, the more confidence we can have in the model. But we will never be able to prove that a model will conform to all possible data; all we can be sure of, at best, is that a model conforms to the data we have seen so far. If we find such a model, as Sutton says, we should compare its performance with that of other models.

If a model and its estimated parameters conform to the original data that inspired its specification, but not to new data, there are several possibilities.

(1) The data-generating process in the real world has not changed, either in the functional form of its relationships or in their paramater-values, but the model is incorrect in general.

(2) The parameter-values of the real data-generating process have changed, but the functional form of the relationships has not.

(3) The functional form of the relationships of the data-generating process has changed.

It is hard to tell which of these is causing the failure. For most failures, the first case seems the most likely. In this case the appropriate strategy is to revise the model and try again. Case (2) is the simplest. Here one appropriate strategy is to re-estimate the model with the new data. A more ambitious strategy is to try to specify a model general enough to describe the manner in which the parameters change. Case (3) would call for a respecified model, either a simple one to describe only the new functional forms, or a more ambitious one to describe the manner in which the functional forms change.

In conclusion, Sutton has given us a perceptive analysis of why the standard paradigm sometimes works well and sometimes doesn't, and an instructive example from industrial organization of what can be done when it doesn't. Unlike Sutton, I believe that the standard paradigm is no longer the best investigative tool we have for most economic problems. I do agree that there is no recipe for model selection. Of course there are recipes for estimation, and, as I have suggested, I believe there is a recipe for model testing.

<u>References</u>


Edgeworth, Francis Ysidro (1892) "Correlated Averages," <u>Philosophical Magazine</u>, 5th Ser., August, 1892, <u>34</u>, 190-204.

Galton, Francis (1885) "Regression Towards Mediocrity in Hereditary Stature," <u>Journal of the Anthropological Institute</u>, 1885, <u>15</u>, 246-63.

Haavelmo, Trygve (1944) <u>The Probability Approach in Econometrics</u>, <u>Econometrica</u>, <u>Supplement</u>, July, 1944, <u>12</u>.

Koopmans, Tjalling C. (1949) "Identification Problems in Economic Model Construction," <u>Econometrica</u>, April, 1949, <u>17</u>, 125-44.

Marshall, Alfred (1890, 1895, 1920) <u>Principles of Economics</u>, London, Macmillan, first, third, and eighth editions, 1890, 1895, and 1920.

Robbins, Lionel (1932) <u>An Essay on the Nature and Significance of Economic Science</u>, London, Macmillan, 1932.

Pearson, Karl (1896) Regression, Heredity, and Panmixia," <u>Philosophical Transactions of the Royal Society of London</u>, Ser. A, 1896, <u>187</u>, 253-318.

Samuelson, Paul A. (1947) <u>The Foundations of Economic Analysis</u>, Cambridge, Mass., Harvard University Press, 1947.