THE QUALITY OF DATA AND THE CHOICE AND DESIGN OF PREDICTIVE MODELS

William Alonso

Department of City and Regional Planning and
Center for Planning and Development Research

Working Paper No. 72

An earlier and briefer version of this paper was delivered at a meeting of the Highway Research Board of the National Academy of Sciences held in June, 1967. The work was supported in large measure by a grant from the Economic Development Administration of the U. S. Department of Commerce.

December, 1967

Long chains of argument are the delight of theorists and the source of their mistrust by practical men. There is some merit in this mistrust. Imagine that we argue that if A then B, if B then C, etc. If we are 80 per cent certain of each step in the chain, from the joint probability of the steps it follows that we are less than 50 per cent certain of where we stand after four steps.[1] Thus the brilliant deductive chains of Sherlock Holmes or the young Ellery Queen, while dazzling, leave us with the feeling that they will not secure a conviction. In this paper I will raise the issue of the effect on models for prediction of errors and their propagation, and suggest some strategies for the selection and construction of models which are intended for applied work. The gist of my argument is that the use of sophisticated models is not always best in applied work, and that the design of the model must take into account the accuracy of the data on which it will be run. There exists the possibility, which should be explored, that some of our intellectually most satisfying models should be pursued as fundamental scientific research, but that simpler and more robust models should be used in practice.

Let us distinguish two types of error: error of measurement and error of specification. Error of specification arises from a misunderstanding or purposeful simplification in the model of the phenomenon we are trying to represent. A simple instance is the representation of a non-linear relation by a linear expression; another is the omission from the model of variables which have only a small effect on the aggregation of variables. Measurement errors are those that arise from inaccuracy in assessing a magnitude. If I say that a man is six foot tall, or a nation has a population of 200 million, I really mean that he is six foot give

---

[1] I first met this argument in I. C. Barnard, *The Functions of the Executive* (Cambridge: Harvard University Press, 1958). A similar argument appears somewhere in A. Marshall's work.

or take an inch, or that the population is 200 million give or take 10 million. Thus, in scientific work it is customary to indicate measurement, M, as having an error, e, attached, and we may write the height of a man, the population of a nation, or the density of a population as $M \pm e$. It is customary to use either the standard deviation or the probable error as the measure of error.[1]

A quantitative model puts together various numbers obtained by measurement, and combines them through algebraic operations. Normally we consider only the measurements and forget the error terms, and give the result of our calculations as a number without indicating its error. Too often we seem to hope that the errors in the inputs somehow cancel out as they go through the model, but in fact they do not. There exists a well-known formula for estimating the error in the output which results from the propagation of errors in the inputs. If we have

$$z = f(x_1, \ldots x_n),$$

$$e_z^2 = \sum_i f_{x_i}^2 \; e_{x_i}^2 + \sum_i \sum_j f_{x_i} \; f_{x_j} \; e_{x_i} \; e_{x_j} \; r_{ij}$$

where e_z is the error of z; f_{x_i} is the partial derivative of f with respect to x_i; e_{x_i} is the measurement error in x_i; and r_{ij} is the correlation between x_i and x_j.[2]

---

[1]The probable error is a distance from the mean such that one-half of the probability distribution lies within the mean plus or minus the probable error; in other words, there is a fifty-fifty chance that the true magnitude lies within $M \pm e$. The probable error is approximately 0.675 the standard deviation. I will not deal in this discussion with the question of an asymmetric error distribution.

[2]See E. B. Wilson, *An Introduction to Scientific Research* (New York: McGraw-Hill, 1952), pp. 272-274; L. G. Parratt, *Probability and Experimental Errors in Science* (New York: J. Wiley and Sons, 1961), pp. 110-118; A. deF. Palmer, *The Theory of Measurements* (New York: McGraw-Hill, 1930). H. Theil uses a different approach in *Applied Economic Forecasting* (Amsterdam: North-Holland Publishing Co., 1966), pp. 262 ff. He formulates the problem in terms of information theory and considers the prediction of sets of numbers.

This formula is exact when the function is linear, but an approximation when it is not. However, recent work has shown it to be a much better approximation than had been previously thought.[1] Thus, by applying it to a model, we may estimate the probable error in the result that arises from errors of measurement in the input variables.

Examination of the formula gives several simple rules of thumb for the construction of models or the selection of models, and these may be useful when, as is often the case, the investigator has several choices in the formulation of the model.

The first rule is to avoid intercorrelated variables whenever possible. The second term on the right hand side shows that the error in the dependent variable can increase very rapidly from this source.

Let us now examine the most basic algebraic operations to derive some other general rules. For simplicity, let us have $z = f(x,y)$, where $x = 10 \pm 1$ and $y = 8 \pm 1$. We will assume that x and y are mutually independent.

Addition:

$$z = x + y$$

$$18 = 10 + 8$$

$$e_z^2 = e_x^2 + e_y^2 = 1 + 1 = 2$$

$$e_z = 1.4$$

We see therefore that, in the case of addition, the absolute magnitude of the error in the dependent variable is greater than in the independent variables. On the other hand, the percentage error is smaller (8 per cent) than in the independent variables (10 and 12.5 per cent). It may be said, then, that the operation of addition is relatively benign with respect to

---

[1] J. W. Tukey, "The Propagation of Errors, Fluctuations and Tolerances: Basic Generalized Formulas," Tech. Report No. 10, Statistical Techniques Research Group, Department of Mathematics, Princeton University. This paper and its companions Tech. Reports Nos. 11 and 12 were not published, and they are now unobtainable.

the cumulation of error. With one exception,[1] it is the only operation which reduces relative error. It must be noted, however, that the size of the absolute error increases.

Subtraction:

$$z = x - y$$

$$2 = 10 - 8$$

$$e_z^2 = e_x^2 + e_y^2 = 1 + 1 = 2$$

$$e_z = 1.4$$

The deceptively simple operation of subtraction is explosive with respect to relative error, especially when the difference is small relative to the independent variables. In this case the relative error is 70 per cent.

Multiplication and Division:

$$z = xy$$

$$80 = 10(8)$$

$$e_z^2 = y^2 e_x^2 + x^2 e_y^2 = 64(1) + 100(1) = 164$$

$$e_z = 13.3$$

It can be seen that multiplication not only raises the absolute error, but also the relative error (in this case to 20 per cent). Division behaves exactly like multiplication.

Raising to a Power:

$$z = x^2$$

$$100 = 10^2$$

$$e_z^2 = (2x)^2 e_x^2 = 400(1) = 400$$

$$e_z = 20$$

---

[1] The exception is when an independent variable is raised to a power with an absolute value smaller than one, in which case both the absolute and the relative error are reduced.

Raising to a power is another explosive operation. In this case the relative error has climbed to 20 per cent. It may be thought of as multiplication of perfectly correlated variables, and thus, from the second term in the basic equation, we may expect the error to be substantially higher. However, if the variable is raised to a power between 1 and -1, both the absolute and the relative error decrease.

From these simple exercises, we can generalize a few rules of thumb for building or choosing models if choices are available:

(a) Avoid intercorrelated variables.

(b) Add where possible.

(c) If you cannot add, multiply or divide.

(d) Avoid as far as possible taking differences or raising variables to powers.

I will illustrate these rules by a "simple" model of the type we all use every day without thinking twice about it. We want to predict population in 1980, P_{80}, from the populations of 1950, P_{50}, and 1960, P_{60}. These populations were enumerated by excellent censuses, with very small errors in the order of one percent. To take arbitrary but typical numbers, let us say that $P_{50} = 100 \pm 1$ and $P_{60} = 105 \pm 1$. We will extrapolate the 1950 to 1960 rate of growth to 1980:

$$P_{80} = P_{60}\left(\frac{P_{60}}{P_{50}}\right)^2 = \frac{P_{60}^3}{P_{50}^2}$$

Of course, we are squaring the rate of growth because we are predicting for two decades. Simple application of the formula, without taking into account any correlation between the 1950 and 1960 populations, gives $P_{80} = 115.76 \pm 4.03$. The relative error in the prediction has risen to 3.5 from 1 per cent in the data. But if we ask the accuracy with which we are predicting the change in population, we obtain 10.76 ± 4.03, which represents a 38 per cent error. This error is due entirely to measurement errors and assumes that the specification of the model is perfect. That is to say, that if we know the exact rate of growth from 1950 to 1960, we would be able to predict the 1980 population exactly.

If we consider that such an extrapolation is a crude model and we say, for instance, that the use of the rate of growth has a 20 per cent specification error (that is to say, that the rate of growth per decade will be 5 ± 1 per cent from the effect of variables not considered) and that the variations from decade to decade are independent, we would have $P_{80} = 115.76 \pm 1.56$, and the change in population will be 10.76 ± 1.56. In other words, there would be a 14.5 per cent specification error if the data were perfect. The joint effect of measurement and specification errors will be 40.2 per cent on the population change.[1]

Before continuing the discussion of the strategy of model construction it may be useful to point out the usefulness of this type of analysis for determining strategies of improvement of data to minimize the compounding of measurement errors. By taking the partial derivative of the error in the dependent variable with respect to the error in an independent variable, we can get the rate of improvement to be gained from better measurement of that variable. The expression, if we disregard the correlation term, is quite simple:

$$\frac{\partial e_z}{\partial e_i} = \frac{f_{x_i}^2 \, e_i}{e_z}$$

That is to say, the marginal rate of improvement in predictive error is equal to the square of the partial of the variable times the measurement error divided by the error in the dependent variable. By use of these marginal rates of improvement divided into the cost of improving the data (e.g. by denser sampling), one can determine the best distribution of budget in data collection. Examination of the expression gives two general rules: (a) concentrate on important variables (i.e., those which affect the dependent variable significantly, as shown by a large f_{x_i}), and (b) concentrate on those with large errors.

---

[1] Out of concern for the sensibilities of those who make projections, I am not questioning the exactitude of the period of two decades, but, of course, the length of time is also a variable. When we say 1980, we really mean sometime around 1980, and thus the exponent of the rate of growth might be 2 ± 0.1 decades. The consequences of this upon the error are spectacular.

Let us illustrate this point by an example. Let us assume the following information:

$$z = xy + w$$

$$x = 100 \pm 10 \qquad y = 50 \pm 5 \qquad w = 200 \pm 50$$

Marginal cost of improving x (to 100 ± 9):	\$ 5.00
Marginal cost of improving y (to 50 ± 4):	\$ 6.00
Marginal cost of improving w (to 200 ± 49):	\$ 0.02

Using our formula we obtain

$$\frac{\partial e_z}{\partial e_x} = 35.2 \; ; \qquad \frac{\partial e_z}{\partial e_y} = 70.5 \; ; \qquad \frac{\partial e_z}{\partial e_w} = 0.0705 \; .$$

These are the marginal improvements on e_z that derive from a marginal improvement in each of the variables. To find the cost of marginal improvement in e_z, we divide these rates into the cost of marginal improvement in each of the variables, and we obtain:

marginal cost of e_z	from improvements in x:	\$ 0.142
	from improvements in y:	\$ 0.085
	from improvements in w:	\$ 0.284 .

Consequently, it would pay us to improve y <u>if</u> marginal reductions in e_z are worth 8.5 cents. It should be noted that improvement of any one variable is subject to diminishing returns, even if the cost of improvement does not rise (which it will normally do). Therefore the analysis should be repeated at small intervals. As might be expected from economic theory, the most efficient situation is that where the marginal cost of improvement in the predicted variable is the same for all variables.

* * *

Let us return now to considerations of model construction. Imagine a situation in which we have a choice of some very naive model, which we shall not describe, which has historically given us 30 per cent error, largely because of poor specification. We have a perfect specification model to predict the same phenomenon of the form $z = x_1x_2x_3x_4x_5$. If each of these variables has a 10 per cent error, the estimate of z will have an error of 22.2 per cent resulting entirely from measurement errors. We will then choose the second model.

Consider now the same situation in a developing country, where the data are poorer. The naive model performs worse because its data inputs are worse, although its specification error will be the same. Let us say that the error of the naive model is 40 per cent. We can use the perfect specification model, but now each of the variables has a 20 per cent measurement error. The second model, then, will have an error of 44.5 per cent due entirely to the compounding of measurement error. In this case we would be better off with the naive model.

The point being made is that the choice of model depends in part upon the quality of the data. The more complex the model, in the sense of having more operations of the same kind or more "explosive" operations such as raising to powers, the more the measurement errors cumulate as the data churn through their arithmetic. The gains in correctness of specification in a more complex model may be offset by the compounding of measurement errors. Although I lack the competence to demonstrate it, I am suggesting that if we tried to predict celestial phenomena by Einstein's General Theory of Relativity using data of the quality which were available to Copernicus, the predictions might be worse than if we used the Copernican theory.

To use a homier example, suppose that we had the wit to design the structure of a skyscraper, but our construction material were timber and our joints were secured with nails. The give in the joints and the members are like weakness in the data: we sometimes can design beyond the capacity of our materials. With timber one should build relatively low and wide, with steel one can build tall and narrow. We shall return to this analogy.

The proposition may be represented in a diagram, as in Figure 1. On the horizontal axis we measure the complexity of the model. I know of no good definition of complexity. The suggested definition, that one model is more complex than another if it has more operations of the same kind or if it has operations which are more explosive with regard to the compounding of errors (such as subtraction of nearly equal numbers or exponential functions) is somewhat circular, but it will have to do. If we are good model-builders, we will only complicate a model to gain advantages of specification. Thus, if our data were perfect, we could imagine a curve of specification error, e_s, which slopes downward asymptotically to the horizontal axis.[1]

On the same diagram we can draw a curve, e_m, for the prediction errors that result solely from measurement errors in our input variables. As the complexity of the model increases, the compounding of measurement error increases. Measurement error increases rapidly at first, but under most conditions, it will increase more slowly with further complications. Total predictive error, E, is the combination of these two types of error in a multiplicative relation, so that $E = [e_s^2 + e_m^2]^{1/2}$. The best point for prediction is the bottom of the total error curve.

In Figure 2 we consider two cases subject to the same specification, but the data in one case are worse than in the other. The curve for the case with poor data, e_m^*, is therefore higher than the curve e_m of the case with good data. The curve E^* of total error in the poor data case is quite naturally higher than the curve E of the good data case. But the important point is that E^* bottoms out at a lower degree of complexity than E. In other words, when accurate data are available, complex models are possible. When the data are poor, simpler models are advisable.[2]

---

[1] I am speaking of situations in which models are improved by progressive refinements, rather than by radical reformulation which may give better predictive accuracy with a simpler model. Such a radical reformulation, which may be called a Copernican advance, would result in a new e_s curve substantially below the original one. This would be the case of a radical scientific advance, and these cannot be called upon at will. I am speaking here of the marginal improvements in the formulation of models.

[2] Note that under these assumptions, as the effects of cumulation of errors and of better specification both flatten out, the curve of total error approximates the curve of measurement error and becomes relatively flat.

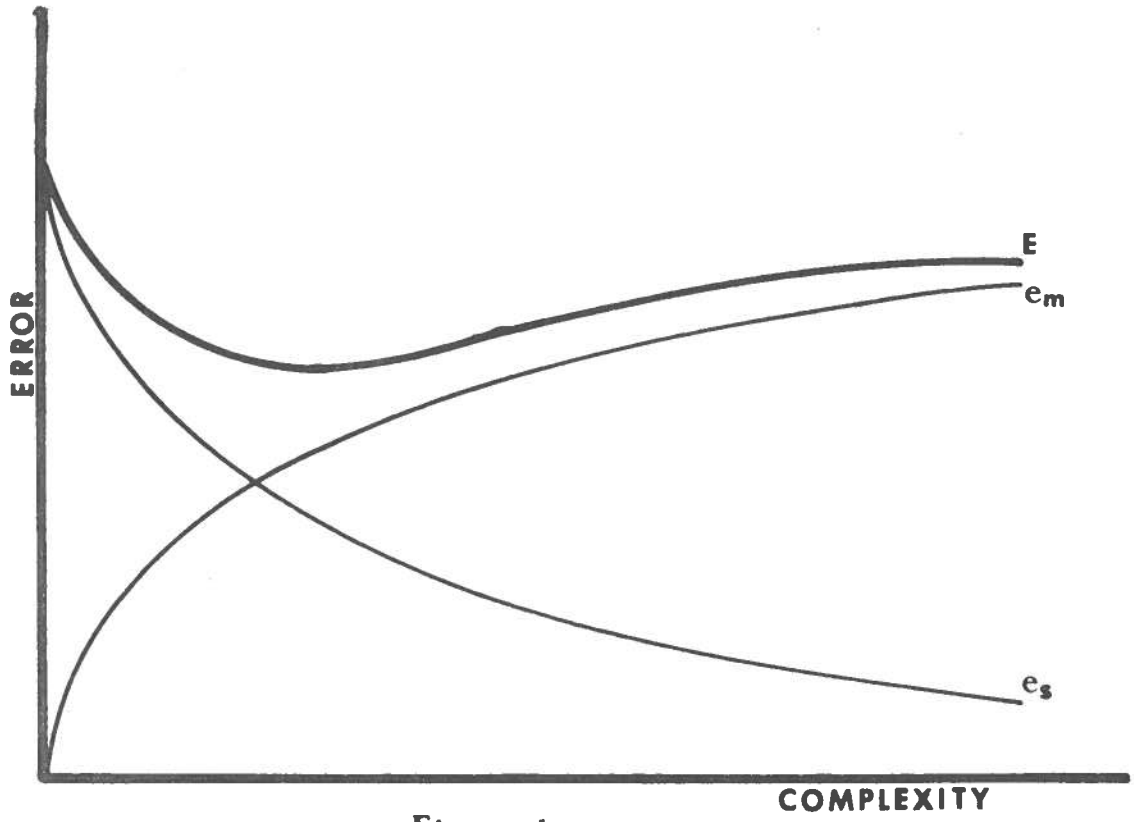

Figure 1

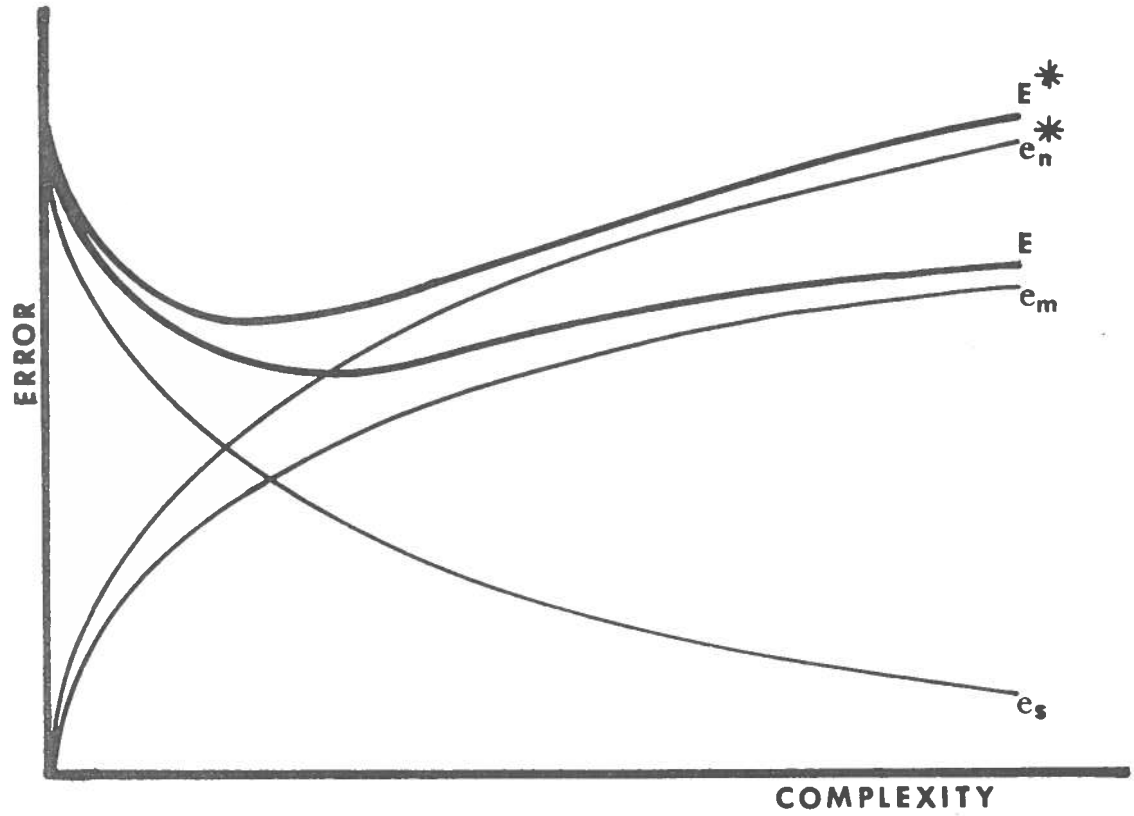

Figure 2

In this view, it is perfectly conceivable that we can devise predictive models which are beyond the capacity of the data, in the sense that, although they are more "accurate" in their specification, the quality of the data results in a deterioration of prediction. I raise the question of whether in the field of land use and traffic models we have not gone beyond the best predictors. I must stress that I do not know whether we have or have not: but we must try to find out.

Let me outline a fairly typical form of one of today's models for predicting needed changes in highway capacity. (1) We predict a change in the absolute numbers of basic employment defined by some reference to a theory of export multipliers. (2) Based on an estimated participation rate, we predict "basic" population. (3) We predict a basic-service employment ratio. (4) We predict service employment. (5) We predict "service" population. We now pass on to (6) a prediction of the location of basic employment. Based on this, (7) we predict the location of "basic" population. (8) We distribute service employment according to basic population. (9) We distribute "service" population.[1] We now have a predicted distribution of jobs and people. Based on this we predict (10) travel patterns or desire lines. (11) We subtract existing capacity from projected demand to obtain (12) needed changes in highway capacity.

Data for earlier periods are fitted, most commonly by multiple regressions, to each of the postulated relationships, and the ordinary tests of significance are applied to the relations one by one. But when the predictive phase is reached, these relations are strung together in a chain. The values of some variables such as basic employment are estimated outside of the model and plugged in. The predicted values of these variables are measurements, and their errors in estimate are measurement errors. The parameters which were calculated in the first stage of the model now become themselves variables to which we can often attach concrete significance such as land requirements per worker or propensity to travel. It should be noted that standard regression techniques give us estimates of error for the parameters. The estimates of the final variables will thus have five sources of error: (1) specification error in

---

[1] I omit here some possible iterative steps to adjust services to total population.

the period for which we have calibrated; (2) further specification error if conditions in the future differ structurally to some degree from conditions in the calibration period, so that a perfect specification of past relations does not specify perfectly for the future; (3) measurement (or predictive) errors in the exogenous variables; (4) measurement errors in the parameters (now variables) in the calibration period; and (5) measurement (or predictive) errors resulting from using past values in place of future values for these parameters/variables.[1] Predictive errors from each of these sources will compound through the operations of the model, as the dependent variables of one step in the chain become the "exogenous" inputs into the next step.

The effects of such a chain can lead to rapid deterioration of prediction. Imagine a three-step chain of regression equations, each validated with an R = 0.9. Assume further that the last four types of error are nil, so that we are dealing only with specification errors in the original relations. The result of the three-step chain will have a 34 per cent standard error of estimate from this source alone.

The general point has now been made, and gives rise to a fifth rule of model construction:

(e) Avoid as far as possible models which proceed by chains.

This general rule, in its positive aspect, says that we should proceed by models which do not build step upon step. This rule increases in importance with weaker data.

But certainly, if we have information on many variables, we want to put it to use, on the general principle that any further information will assist our prediction, and that simple models which use few variables neglect some of the information available. My suggested strategy, which I cannot illustrate by concrete example, is to build several simple models which among them use all of the data, and to make some sort of average of them. To paraphrase a scientist in a field that faces similar problems,[2]

---

[1] Sometimes these parameters are adjusted for the predictive equations, based on some other information; this may reduce this source of error, but, of course, will not eliminate it.

[2] R. Levine, "The Strategy of Model Building in Population Biology," *American Scientist* (December, 1966).

the strategy is not to build one master model of the real world, but rather a set of weak models as alternative models for the same set of phenomena. Their intersection will produce "robust theorems." As complementary models they shed light on different aspects of the same problem. In other words, an average of simple models will give predictors which are far stronger than the individual models. For instance, if we have eight variables with 10 per cent measurement errors, and we can construct four simple models of products of different pairs of the variables, each of these simple models having 40 per cent specification error, the total error in their average will be 23 per cent. If we had a single multiplicative model using all eight variables which were perfect with regard to specification, the expected error would be 28 per cent.

This strategy of netting out weak and complementary models may be called a technique of mulling over, in contrast to the deductive chains of our present models and the classic detectives of fiction. It is what most of us do in real life when faced with a difficult problem. We consider first one aspect and then another; when we have considered every aspect we can think of, we start all over again, and eventually we come to a decision.

I want to stress that I am by no means certain that our urban and regional models have reached the level of mathematical complexity where the compounding of experimental or predictive error offsets the gains in specification accuracy. I am only raising the question of whether they have. If this is the case, I am suggesting a strategy of many short, stubby models to be averaged as opposed to the present strategy of long, thin models. In terms of my earlier analogy, I am questioning whether we have arrived at the design of skyscrapers but we have only lumber for construction material. If we do, we had better build low to the ground while we improve upon our materials.

This argument has a complementary conclusion. If the data are very good, it is wasteful to use too simple models. This raises some interesting issues concerning the applicability of models generated in developed countries to situations in developing countries, where the data are invariably poorer. The use of the same model implies that its specification is acceptable in both cases. But if the specification is properly matched to the quality of data in the developed

country, the poorer data in the developing country assures us that we shall be well past the low point of the total error curve for that country, as illustrated in Figure 2. On the other hand, if the model is well suited to the quality of the data in the developing country, it is wasteful of the power of the data in the developed country. The use of the same model for reasons other than those of expediency in both situations will be justified only if we cannot think of alternative model designs.

In conclusion I want to touch lightly upon three general points which have to do with the uses of models rather than with their design. The first point is that, whether or not we have exceeded the capacity of our data in the design of models, we are surely operating in broad areas of uncertainty, and that we find errors of 50 or even 100 per cent acceptable because alternative means give even larger errors.[1] Yet the institutional context in which most of our most advanced models are constructed results in relatively short tenures by the key investigators, in the order of one to five years. In that conditions of high uncertainty place an extraordinary premium on feedback, it seems to me that the love-them-and-leave-them nature of most of our significant modeling efforts is extremely wasteful. At a time when urban planners are rebelling against the master plan and calling for continuing planning (that is, continuously revised plans), our quantitatively most advanced planning is reverting to the master plan, not out of the logic of its instruments, but out of the sociology and the institutional matrix of the investigations. It is obvious that these models should be designed and placed in their institutional context in such a way that continuing revisions and improvements can be incorporated easily and, more important, the consequences and importance of such revisions be understood by the decision-makers who are the consumers of the model.

---

[1] See O. Morgenstern, *On the Accuracy of Economic Observations* (Princeton: Princeton University Press, 1963), for a sobering discussion of the magnitudes of error in national economic statistics. Urban data may be expected to have errors of at least this magnitude. For a discussion of the actual errors in the prediction of national macro-economic variables, see Victor Zarnowitz, *An Appraisal of Short-Term Economic Forecasts*, Occasional Paper 104, National Bureau of Economic Research, 1967. The Bureau of the Census and other agencies frequently produce studies on the reliability of their data. One that has received considerable recent attention is *Measuring the Quality of Housing: An Appraisal of Census Statistics and Methods*, Working Paper No. 25, Bureau of the Census.

This matter must be stressed. A decade ago, these models were viewed primarily as predictors of the future. Somewhat later, stress was placed in their use as conditional predictors of the consequences of alternative policies, and efforts were made to incorporate into them policy variables which would permit such experimentation. Most recently, as experience has been gained, the practitioners of this craft have tended to play down the ability of the models to predict, and to stress their value as educational instruments which serve to bring to the consciousness of those who make decisions the complex interrelations among the variables, including those which can be manipulated for normative purposes. Thus, the downgrading of the importance of the numbers which emerge from the model accords with the viewpoint being advanced here. The large model may serve as a context or evolving background for a collection of more partial and overlapping quantitative models and for that vast reservoir of knowledge about the urban system which inhabits the heads of experienced men and which has yet to find its way into formal models.

In justice, it must be noted that some of this takes place now during the relative privacy of the period in which the model is calibrated.[1] Commonly, in the early runs the model will produce some outrageous results, and the modellers will use their necromantic powers to have the black boxes give reasonable results. Although these false starts are little advertised, the corrections constitute a combining or averaging of models. For how would we know what is outrageous or what is reasonable except by appeal to other models of the future, even if some of these are implicit or intuitive? Rather than treating this process as an embarrassing occurrence during the model's infancy, it should be treated as a continuing source of strength and enrichment to be carried into adulthood.

A second general point has to do with models as instruments for decision. Our models give point estimates for the variables which we are predicting. In this paper I have suggested that these estimates should include estimates of predictive error. The crude techniques I have employed have dealt entirely with probable error or standard deviation.

---

[1]This is the term commonly applied by traffic and land-use technicians to the obtaining of parameters to fit data for earlier periods.

But the purpose of the study of error is not to burnish our scientific conscience, but to assist in the making of decisions. If either the penalties of error or the probabilities of error are asymmetric about the most probable estimate, in all likelihood our best action will not be addressed to the most probable estimate. This may be illustrated concretely. Suppose that our estimate of traffic on a new roadway is a central value with a symmetric probability distribution about it. If we guess too high, the cost is the waste of a bit more land bought and a bit more surfacing laid. If we guess too low, the costs of widening the road later are far higher, for we are forced to buy out development by the side of the road which the road itself has induced, to widen bridges, rebuild cloverleafs, etc. Where the costs of error are asymmetrical, the best action will be off the most probable value in the direction in which the error has higher cost.[1] Similarly, skewness in the error distribution about our central estimate should lead us to take action for quantities other than the point prediction. Thus far, builders of models of urban traffic and land use have been content to predict a central value. The challenge is to pass from this to more reasoned recommendations for concrete action. This is undoubtedly very difficult to do, but it needs doing.

The third point is one of which I am statistically less certain, although the institutional sociology of it is clear. The most important studies in this field have cost several millions of dollars. The techniques they have employed are generally pioneering in that they have tried new things. But they have found themselves in a difficult position. As pioneering studies, they went into the unknown, where there is a high possibility of failure. As professional agents, they were in fact charged with using an existing and generalized body of knowledge upon a concrete situation. After having spent some millions of dollars, they could not afford to say that the experiment did not work. I will submit that after spending tens of millions of dollars in studies of this sort, the reportage on what we have found out has been minimal. A few journal articles and a few handfuls of agency reports which are generally unclear is all

---

[1] I am indebted to Leon Cole's as yet unpublished work for this insight.

we have. Seldom do we find clear and self-examining evaluations of the work.[1]

It seems that the vast expense required for these studies places them beyond what the sources of funds are currently disposed to spend for basic research, and labels them as planning costs for applied work, where a handful of million are acceptable in the face of investments in infrastructure in the order of magnitude of one billion. Yet, to those with an interest in this subject, the promise of one after another of these multi-million-dollar studies has not been fulfilled. It is not that strong statistical findings have been lacking: it is my impression that there is a wealth of regularities. What is lacking is a dispassionate report on findings and failures from which scholars in this field, including those in the project, can test and evolve new understanding of the phenomena with which we are dealing and techniques to deal with them. Researchers are being put in the very difficult position of being both practitioners and innovators. As practitioners, they are called upon to use techniques that have a high probability of success; in effect, to apply known and proven methods. But in this field most of our methods are still in their infancy, still in the process of discovery. Innovative or scientific work is by definition an exploration beyond what is presently known, and any one probe will have low probability of success. The societal logic of support for scientific work is that the rare successes tend to have very high pay-offs. The institutional context of these studies blurs this distinction under the pressing need to decide how to spend vast quantities of money in urban infrastructure, and thus hampers the openness of method, the candidness of reportage, and the freedom of discussion of these important studies. This represents a dreadful waste, as errors are repeated and successes are not followed up. Although there have been significant advances, they have not matched the possibilities.

In conclusion, I would like to advance, with considerable hesitation, a statistical argument for a distinction between models for fundamental research and models for applied work. Consider a case in which we use the same model design for both purposes. In the research model we

---

[1]The most significant exception of which I am aware is Ira Lowry's, *A Model of Metropolis* (Santa Monica, Calif.: The RAND Corporation, 1964).

are asking what are the relations among the measured variables, and whether they conform to what we would expect from various theories and prior empirical work. We may regard the parameters we obtain not as variables in their own right, but as relations among the variables we have measured.[1] But, as we have discussed, if we are using the model for prediction, all of our numbers become variables. Further, as variables they have a larger error when they are predicted for a future state of the system, and the model itself, as mentioned, may have a larger specification error with regard to a future state. From these considerations, it would seem that a model that seeks to increase our understanding by asking how certain variables relate to each other is in a sense less subject to some of the sources of error than the identical model design used to predict the future.

This point may be arguable both in statistical terms and in terms of the philosophy of science, in which it is often held that the purpose of all scientific work is prediction. This may be countered by pointing to much of the good scientific work which classifies, describes efficiently, generalizes, merely checks that things are as we expect them to be, and in other ways improves our comprehension of nature. Such work will often result in better prediction, not by its direct use, but by shedding light on some facets of the structure we are considering, while the prediction itself proceeds in the fashion which I have called mulling over. But if there is merit to the statistical argument, it follows that, for a given quality of data, the scientific model is more tolerant of complexity of formulation than an applied model. If this is true--and the alert reader will note that this is a deductive chain--it follows that we should have research groups in universities and other centers working on complex models, while operational agencies would be working with simpler and safer models.

---

[1] This, of course, is relative, for we often "measure" a variable as a relation among variables which have been measured directly. The question of what is direct measurement is a difficult one.

CONCLUDING NOTE ON MODELS WITH NEGATIVE FEEDBACK

At a presentation of an earlier version of this paper, Britton Harris raised the question of whether the problem of cumulation of error would apply with equal force to models which possessed negative feedback. This is an intriguing suggestion, although it is hard to determine what constitutes negative feedback. In many cases the underlying theory may have this feature, but the computational model does not. In other cases, the feedback is no more than a series of dampening mechanisms to foster conservative predictions. Among these mechanisms are the use of rigid constraints which will not yield, or soft ones which yield grudgingly. Of course, the mathematical form of some models, based on systems of equations, provides for a particular form of negative feedback. Recent models have frequently been based on algorithms for numerical estimation of the hill-climbing type, in which the investigator uses feedback from each successive estimation to proceed to the next. Yet this feedback, used to find the solution of the model, must not be confused with feedback within the model itself, but rather results from our inability to solve the model analytically. If we could solve it analytically, the arguments presented in the body of this paper would apply. When we use a technique of numerical approximation, we add to the errors of specification and compounded errors of measurement a further error resulting from the remaining inaccuracy of our last numerical approximation to the analytic solution.

Of course, a form of feedback is involved when consistency checks are used, often by the use of alternative models. This reflects the position argued in this paper. Lastly, unambiguous feedback occurs when models are kept a long time and incorporate corrections as new information becomes available.