D R A F T--To Be Revised

(1)

THEORIES AND FORECASTING IN INTERNATIONAL RELATIONS:

THE ROLE OF VALIDATION EFFORTS

by

Charles F. Hermann
Warren R. Phillips
Stuart J. Thorson

The Ohio State University

DDC
RECEIVED
JAN 9 1978
D

THEORIES AND FORECASTING IN INTERNATIONAL RELATIONS:

THE ROLE OF VALIDATION EFFORTS[1]

INTRODUCTION

Periodically, international relations scholars are urged to cast the results of their studies in terms of forecasts or expectations about the future. The reason seems clear enough. At some future point the forecasts can be compared against actual occurrences and based on the degree of confirmation the original research (or the researcher) can be evaluated. Moreover, if the forecasts have concerned the near future, the investigator can presumably use inadequate forecasts to revise his reasoning. New estimates of the future can be made and subsequently checked in a cyclical manner to produce successive approximations that hopefully achieve a continuously improved fit between forecast and subsequent observation. What is more, if the forecast obtains acceptance, it becomes the basis for prescriptive action. Humans thus participate consciously in shaping their future and engage in self-fulfilling or self-denying forecasting. ("If certain occurrences will happen, we need to undertake the actions to promote, obstruct or take advantage of them.") Perhaps few proponents of greater forecasting in international relations would state their case in such unqualified terms, but the above description appears to capture the core of such arguments. The argument has

much merit. A forecast that is stated in such a way as to permit its verification against the unfolding future provides one type of criterion for validity.[2]

The difficulties arise in moving from these simple statements of aspiration to the development of insights and procedures that can be applied in research. At the point of actually validating forecasts a host of philisophical and practical questions arise. What is it that the forecast represents? Or put a different way, assuming that a forecast could be validated, what does it mean? How does purpose affect the validation of a forecast? What validation procedures can be employed? What about inconsistencies between the results of forecasts and other means of validating a theory? How can one confidently know (and measure) the future reference system when one sees it? These questions tip off the reader to the conclusions to be found at the end of this chapter. Using forecasts as a validation procedure is much more complex and the results less certain than appears at first glance. Nevertheless, it is an important, if insufficient, operation for improving our knowledge of international relations. For that reason, the following pages seek to provide some initial exploration of the issues posed by these questions and where possible to suggest some possible procedures.

---

2 See C.F. Hermann, "Validation Problems in Games and Simulations with Special Reference to Models of International Politics," Behavioral Science 12 (May 1967), 216-231.

THEORY AS THE GENERATOR OF FORECASTS

Assume that we momentarily set aside the problems of determining
how a forecast is valid, one question that remains is what do we know
when we have a validated forecast? In such circumstances, we would
know that a particular estimate made at some prior time has been con-
firmed to some degree by subsequent developments. This confirmation
of forecasts can be variously referred to as validation, goodness of
fit, verisimilitude, isomorphism, verification, or accuracy. Beyond
this information about the relationship between the forecast and actual
events, however, we frequently want to infer something about the means
and the source by which the forecast was generated. More specifically,
we might normally wish to infer something about the ability of that
source to generate other forecasts. ("Carl was correct in anticipating
the outcome of this week's soccer game, but will his judgment be as good
for next week's match?") In this simple example, the inference is about
the ability of an individual to make a forecast. Unless he was making
an ungrounded guess, the forecaster performed some calculations that
formed the basis of his estimate. As long as they remained unarticulated
we know very little about the mental images or models that generated the
forecast. Policy makers also have mental images which they use in esti-
mating the future. For example, an expert on the Soviet Union probably
has mental models of how political decisions are made in that country.
He could use these images in evaluating the alternative future policies
that the USSR might take on a given issue. Similarly, scholars also
use mental models or images which delineate the problems they should
attack and the likely approaches to delineating forecasts in a particular

substantive domain. These mental images are frequently relied upon;
unfortunately there are major problems associated with this form of
forecasting with respect to establishing the validity of its source.
Different researchers have different mental images, each dealing with
a wide range of overlapping substantive interests, and each frequently
inconsistent with the others. We are faced with difficulties in knowing
which images are applicable in a specific case, and because the relation-
ships in each image are not explicitly and clearly identified. The
sources of contradiction may not be obvious, because the relationships
in each image are not clearly defined. The lack of explicitness in
mental images makes it difficult to communicate the assumptions upon
which any forecasts are based. In cases in which disputes about alter-
native outcomes actually are recognized, unidentified assumptions implicit
in the mental images that researchers hold frequently are the cause of
these differences. Perhaps more importantly in long range projections,
it is difficult to manipulate the variables in mental images in order
to assess the various impacts of individual changes that could operate
on the initial conditions. Thus, the complexity of social phenomena
makes it extremely difficult to move from a vague set of assumptions
about the world through the dynamic consequences resulting from these
assumptions to various forecasting alternatives.

As many chapters in this book make clear there are numerous ways of
generating forecasts. The unexplicated mental images in the minds of
one or more individuals are only one such means but they are a frequent
one in international relations. They deserve attention not only because
of their frequency but because they illustrate a basic problem. When
one or more forecasts are used as a means of validating the utility of

an explanatory source for subsequent forecasts and explanations, the components of the forecasting system and their logical relationship to one another must be explicit. Otherwise, what can be inferred about the validity of any future performances of the system will be quite limited. In short, we assert that in order to use forecast validation as a means for inferring the future predictive capability of the source, the source should have the characteristics of a deductive theory. Such a requirement certainly limits the range of sources that can be subjected to validity estimates through forecasts. Nevertheless, the requirement of a deductive theory as the source of forecasts seems appropriate, if our validity studies must take into account the following considerations:

(1) Forecasts are used to estimate the utility of the source for future forecasts.

(2) It is necessary to establish the parameters or boundaries beyond which the source may decline sharply with respect to the accuracy of its forecasts.

(3) The forecast concerns a dynamic reference system that is suspected of containing some components that can assume a substantial range of values which in turn may yield quite variant outcomes.

We believe these are conditions that frequently confront the international relations scholar who evaluates the validity of forecasts.

Before proceeding further, it would be desirable to offer some definitions of the basic terms we have been using. A _deductive theory_ is stipulated as a set of sentences which is closed under deduction, that is, the set contains any sentence that is logically implied by

any other sentences in the set. Generally the sentences in a theory are asserted to be true (of some world).

A _forecast_ is generally thought to be a statement made at one time about the state of some world at some future time. Thus the theories to be considered for forecasting must be dynamic theories in the sense that the value (state) of some variables are related to values of other variables at other points in time.

More precisely, consider a theory about some world consisting of state variables $(x_1, x_2...x_n)$. We want our theory to contain sentences relating at least some of these state variables to previous states of the system. In physics, for example, these sentences are often expressed in differential equations of the form:

$$\frac{dx}{dt} = f_i (x_1, x_2...x_n)$$

An example of a theory of this type drawn from the international relations literature would be the theory of arms races developed by Richardson.[3] Here again differential equations are used to relate a nation's level of defense at one time to system states at previous times.

A second example might be the world simulation described by Forrester.[4] The sentences are in the DYNAMO language and levels of variables at one time are related to levels at previous times. This time, the statements are in difference equations form.

---

[3] See L.F. Richardson, _Arms and Insecurity_ (Pittsburgh:Boxwood Press, 1960) and L.F. Richardson, _Statistics of Deadly Quarrels_ (Pittsburgh: Boxwood Press, 1960).

[4] J.W. Forrester, _World Dynamics_ (Cambridge: Wright-Allen, 1971)

In principle, a theory need not be expressed in an artificial
language (such as DYNAMO or differential equations) to be a member of
the class being discussed. Theories expressed in a natural language,
such as English, may also satisfy the above conditions. It might be
argued, for example, that Galtung's "rank theory" meets the criteria
set out above.[5] A problem with most natural language theories (including
Galtung's) is that it is very difficult to unambiguously identify the
objects and relations being discussed.

Specifically excluded from the analysis that follows will be means
of generating forecasts which are not "dynamic" theories of the sort
identified above. Thus, trend and cyclical analysis that simply project
prior patterns without any antecendent explanations are excluded. So
too are the development of speculative or plausible scenarios, Delphi
techniques, and the various devices associated with assessing the validity
of measures (as for example in the psychological test and measurement
literature). All have a role in forecasting in international relations.
But evaluation of the validity of the forecast from such sources has
limited utility for theory development.

Now that the class of theories to be discussed has been identified,
it is appropriate to specify the concept of validity which will be
employed in this chapter. In discussing a concept such as validity it
is important to distinguish between semantic and methodological question
of how it becomes known whether a particular theory is, in fact, valid.
Answers to the methodological question would seem to presume adequate
answers to the semantic one. Therefore, the first task will be to
explicate what will be meant in this chapter when validity is predicted

[7] J. Galtung, "A Structural Theory of Aggression," Journal of Peace
Research 2 (1964) pp. 95-119.

of a theory. A theory--a set of sentences in some language--is valid if it does what it purports to do. Thus, as is noted by Forrester[6] and Hermann[7] the question of validity is inextricably intertwined with the purpose to which (in this case) a forecasting system will be put. A number of possible purposes and criteria of validity appropriate to these purposes will be treated subsequently. However, we can now state the semantic conception of validity being employed in this chapter. A theory, T, is valid with respect to purpose, P, to the extent T achieves P. Relating validity to purpose, is, of course, compatible with an extremely pragmatic view of theory evaluation. This compatibility, however, does not require that we adopt such a pragmatic view. One might argue, for example, that the purpose of a scientific theory is to generate (or be capable of generating) true sentences.[8] Thus, the test of validity of a scientific theory is whether the sentences comprising the theory (as well as those logically implied by these sentences) are true. That is to say, for a scientist taking this position to assert that T is a valid theory is equivalent to his asserting that the sentences comprising T are true. Note again that this semantic definition of validity does not entail any particular methodological position as to how a particular theory is known to be valid (i.e., known to consist of true sentences). For example, it might be argued that the goal of

---

[6] J.W. Forrester, Industrial Dynamics (Cambridge: MIT Press, 1961).

[7] C.F. Hermann, "Validation Problems in Games and Simulations."

[8] See K.R. Popper, Conjectures and Refutations: The Growth of Scientific Knowledge, Harper Torchbooks (1965) pp. 223 ff.

science is to construct true theories (i.e., theories whose sentences are true) and yet still argue that it can never be known whether any particular sentence is in fact true and therefore be some sort of falsificationist rather than a verificationist.

The important point here is that the validity of a theory is contingent upon its purpose(s) and therefore it makes little sense to inquire of the validity of a theory without inquiring as to its purpose(s). Purpose is just one of the factors that affect the relationship between a forecast and the theory used to generate it. The most important of these issues must be considered in greater detail.

SOME CONSIDERATIONS AFFECTING THE RELATIONSHIP BETWEEN THEORY AND FORECAST

Let us take a brief review. The theories of interest in this chapter must generate forecasts, that is, statements concerning changes in the values of objects at different points in time. We contend in this chapter that the question of forecast validity is actually one of using the forecast to assess the validity of the theory that generated the predictions. The assertion that under certain conditions a particular pattern of events will occur during some future period of time suggests an obvious criterion for establishing validity of the theory. If the specified conditions transpired, did the projected pattern occur as predicted? The accuracy of forecasts is certainly an essential feature of the validation effort, but a number of issues must be taken into account in evaluating the relationship between a theory and its forecasts.

As we noted at the end of the previous section, no discussion of the factors that affect the interpretation of the relationship between forecast validity and the theory which generated it would be complete without consideration of the purpose the user intends to make of both the theory and the forecasts. Any interpretation of the accuracy of a forecast as an indicator of the adequacy of a theory must be evaluated in terms of the purposes of the user. As purposes vary so does the degree of tolerance in goodness of fit between forecasts and observed patterns of events. In fact, the user's purpose should determine whether inferences about the theory from confirmed forecasts are of major importance. Elsewhere some distinctive purposes of simulations (one type of theory) have been described together with their implications for validity. Among the purposes mentioned were (a) the discovery of alternatives, (b) the evaluation of alternative outcomes, (c) prediction, (d) instruction, (e) construction of hypotheses and theory, and, (f) the exploration of non-existent universes. For the present, however, we need only establish that the user's purpose will make a difference. For example, if the user seeks explanation for why certain events transpire, then the confirmed forecast may be of minimal value in assessing a theory's adequacy. It is quite possible for a theory involving a number of stochastic processes to yield accurate forecasts about a closed system without providing much insight into why the observed pattern occurs when it does. With respect to the degree of accuracy in forecasting, numerous illustrations come to mind. A scholar developing a theory which estimates the rate of inter-action between nations of opposing military alliances given various levels of interstate conflict in the international system may find

support for his theory in a goodness of fit ratio that remains quite
modest. On the other hand, a theory that estimated the number of ICBM
launches that could be built by either the Soviet Union or the United
States without detection by the other side would have to have a much
better predictive capability if it were to be used as the basis for
signing, or not signing, an arms limitation agreement. In assessing
the degree of accuracy necessary for the user's purpose, one criterion
must be the alternative available for forecasting. In statistical
tests, forecast performance is often compared to chance, but that may
not be the relevant standard in a particular case.

Another issue we must address is probability as opposed to deter-
minism in the theory. Suppose we have a theory which leads to the
following assertion: If nations of the world are ranked according
to military and economic capability, the first-ranked nation will
always initiate war with the second-ranked nation, if--and only if--
the latter's rate of growth in both military and economic capability
relative to the first-ranked nation will lead to a reversal of ranks
within five years. Such a statement can be contrasted with one which
concludes that the first-ranked nation is more likely to initiate war
against the second if its projected economic and military growth rate
will cause it to overtake the first-ranked nation within five years.
The first statement claims to contain all the conditions that are
necessary to produce the projected outcome and that the outcome occurs
every time the conditions are met. The second assertion contends only
that the specified conditions increase the likelihood of the outcome.
Although the example may seem a bit far-fetched, some theories can

generate forecasts which are held to be completely determined by the
configuration of specified conditions; whereas others are probabilistic
theories, the most sophisticated of which may be able to estimate the
probability associated with different possible outcomes.[9] When the
theory's specified prior conditions are not related in a deterministic
fashion to the estimated outcome, a forecasting exercise cannot provide
insight into the theory's degree of validity without consideration of
the impact of exogenous variables. Moreover, even in the case of the
deterministic theory, the lack of congruity between forecast and out-
come may lead no further than to recasting the relationship in probabi-
listic terms.

A deterministic theory yields a set of expected values in some
future state but makes no provision for the outcome if the expected
values do not occur. It is as if our theory projected the rate of
decent of a ball of a certain mass down an inclined plane having an
angle that is a certain number of degrees from horizontal, but taking
no account of friction resulting the air density, the surface of the
plane and ball, etc. Or, consider the example of theory that projects
that a certain rate of economic development in a less developed country
will begin, at a given point, to generate a certain amount of capital.
These theories neglect what happens if the forecasts are not fulfilled--
the amount of friction drastically slows the ball or internal revolution

---

[9]The distinction between the projected outcomes from probabilistic as
compared to deterministic theories overlaps somewhat with Choucri's
distinction between predictions and forecasts. We maintain, however,
that a deterministic theory could still produce a forecast in Choucri's
sense of the term. See her discussion in Chapter 1.

slows capital formation. If the distribution of outcomes around the

projected one involve only gradual deviations, we still might give the

theory "high marks" even if slight errors occur. If the distribution

of outcomes surrounding the one that is forecasted falls off sharply,

then a deterministic theory poses severe problems--particularly if

the forecasted outcome is regarded as desirable and those around it

appear undesirable. Thus, for example, instead of capital formation

a country experiences revolution. Therefore, although forecasts

of a deterministic theory may more readily be tested for their validity,

inaccuracies may be more difficult to interpret (i.e., how far off is

the actual outcome?) and pose serious difficulties for some purposes

(e.g., policy analysis).

There is a counterpart in the reference system to the deterministic-

probabilistic characteristics of theories. We must consider the actual

distribution of the forecasted events in international relations. Are

the occurrences considered unique and non-current or are they repeated

regularly? Examples of the former include the death of Mao or the

acquisition of nuclear weapons by Japan. Whereas the latter include

such things as changes in political leadership of a country or the rate

of diffusion of a technology. If the phenomena that are the subject

of the theory reoccur in the reference system, we need to take into

account the frequency of their appearance. Are they frequent occurrences--

such as diplomatic exchanges or trade negotiations--or relatively less

frequent--such as inter-state wars or global economic depressions?

Suppose that a theory forecasts the probability of the outbreak of war

under certain conditions is .75 and in subsequent actuality the conditions

are fulfilled but no war occurs. Over a series of such forecasts we could establish whether the forecasts correspond to events three-fourths of the time, provided that the class of predicted events occurred with sufficient regularity together with the set of conditions specified in the theory. Then we would have a situation comparable to that used in weather forecasts of precipitation. ("The probability of rain in the next 24 hours is 80 percent—or more precisely, the probability of precipitation is 80 percent under conditions such as those that are expected to prevail in this locality during the next 24 hours.") Unfortunately, there are numerous events in international relations that do not occur with the frequency with which rain falls on many parts of the earth. Thus, we have a situation in which a theory can predict a pattern of occurrences which do not occur in the real world with sufficient regularity to assess with confidence for forecasts.

One thoughtful critic has charged that in his previous writing on the subject, the first author has failed to consider that an error in forecasting (or other criterion for validating a model) can result from a misinterpretation of the reference system—or "real world"—rather than from an inadequate model.[10] The charge highlights another problem in the inferential relationship between forecasts and theory. When an incongruity exists between forecasts and subsequent developments, one might ask whether it results from the theory—let us call it theory X—that led to the forecasts that is unsatisfactory or the theory—designated

_____

[10] See Charles A. Powell, "Validity in Complex Experimentation," Experimental Studies in Politics (1970).

theory Y--used to observe and interpret the reference system? When an astronomer calculates from deflections in the movement of other bodies in our solar system that a previously undetected planet should be observable at a certain point in space and none is found, is the astronomer's theory of the missing planet wrong or should we re-examine the theory of optics or the theory for locating other objects in space relative to the earth? If a simulation forecasts a certain pattern of national economic growth which is not substantiated in subsequent economic activity as measured by the Gross National Product, do we re-examine the simulation or the indicator of actual economic performance?

Certainly, a committed scientist ought to consider all such avenues in cases of unconfirmed forecasts. It ought to be possible for him to develop a strategy for determining which explanation for the lack of a confirmed forecast he should pursue first. (Has the theory of optics been substantiated independently in other tests? Does the present test use GNP in ways the measure has not previously been used?) Given the relative newness of simulations in international relations and the restricted presentation that exists in any simulation, it is easy to conclude that inaccurate forecasts are indicative of inadequate simulations. Perhaps, such inferences are too easy. Our conceptualizations and observation techniques in international relations have seldom been confirmed in a systematic fashion. In a given area of international relations there may be no definition of the key concepts, no explicit statement of assumptions, and very elastic measures of observation. Under such circumstances, the scholar must be acutely sensitive to the possibility that his means for verifying the forecasts require careful examination.

Although it is always desirable to check the theories of observation
and interpretation used in confirming forecasts, the tendency to do so
is greater the more discrepancy occurs between forecast and subsequent
events. Another type of problem arises in instances in which the goodness
of fit between forecast and events seems substantial. How confidently
can we infer from such verisimilitude to the theory assumed to have
accounted for the observed developments? There is the possibility that
the correspondence of events and forecasts is the product of a spurious
correlation, coincidence, or an overdetermined event. The appearance of
a substantial goodness of fit that actually results from fortuity should
be eliminated by repeated forecasting attempts that would reveal the
coincidence as random error. Repeated tests should also reveal those
situations which are overdetermined--that is, outcomes that result from
any of several different factors and all of which happen to be present
in a given instance. Across a variety of forecast occasions, some of
the relevant exogenous conditions may not occur, and those accounted
for in the theory will be responsible for the observed result. Some-
what more troublesome is the systematic error in the form of a spurious
correlation. Although repeated forecast efforts may reveal the presence
of this problem, one can put the theory in an operational form--or
simulation--and conduct sensitivity tests to determine the effects of
individual components on the outcome when other elements are held
constant.

The reference to sensitivity testing as a means of checking on
spurious correlations that might explain a high degree of accuracy in
a forecast openly makes a point applicable to all the issues discussed

in this section. In order to clarify these problems that can affect
the assumed relationship between a forecast and the theory that generated
it, we must examine directly the theory. For spurious correlations, we
want to conduct sensitivity tests on the theory. To determine the
implications for forecasting of the user's purpose, we need to examine
the theory for its correspondence with such purposes. If we have a
deterministic theory, we need to identify with special care the exogenous
variables not contained in the theory that could alter the forecast.
Should the theory predict rare events in the reference system, we need
to establish estimates of our confidence in the theory independently of
its forecasts of those infrequent occurrences. (We will return to this
point in the discussion of plausibility in the next section.) Again, in
deciding between errors in theories that generate forecasts and errors
in theories involved in assessing the actual occurrences in international
politics, we must move outside the forecasts themselves. In short,
issues that can affect our inferences about theory which are made from
confirmed forecasts, require us to deal directly with the source. This
observation is one reason why we contend that validity of more than the
forecast itself requires that the source of the forecast be an explicit
theory. Unless the source of the forecast reveals its components and
their relationships, resolution of the issues discussed in this section
often becomes impossible.

VALIDATING THE FORECASTS

Assuming that we want to make inferences about the future predictive
capability of the source of a forecast (a theory) and that we can manage

the kinds of difficulties outlined in the previous section, the task remains of determining the accuracy or goodness of fit between the forecast and subsequent events. After all it is from this degree of congruence that we move to inferences about the theory that generated the forecast. In this section we consider two aspects of validating forecasts--plausibility and empirical verification.

Although validation is often thought of as exclusively an empirical exercise, at the time a forecast is made it attempt to describe future events for which we have no immediate empirical capability for validating. Because this is the case, and because the careful validation of forecasts can often be expensive in time and money, we ought to satisfy ourselves that such an effort is justified. Of course, this justification depends in part on the user's purpose. It should also depend on the plausibility of the forecasts, that is, the contextual constraints which must not be exceeded if forecasts are to be taken seriously. We might begin by considering the caution of Newell and Simon who observe:

> The plausibility of a fundamental hypothesis about the
> world is almost always time-dependent. Hypotheses are
> seldom plausible when they are new and have not yet
> been widely accepted. If empirical evidence supports
> a hypothesis increasingly, and if the hypothesis suc-
> ceeds in providing explanations for a significant range
> of phenomena it becomes more and more plausible.[11]

---

[11] A. Newell and H.A. Simon, Human Problem Solving (Englewood Cliffs, N.J.: Prentice Hall, 1972), p. 19.

This psychological relationship between "plausibility" and "empirical success" mitigates against using plausibility as a sole criterion for validating forecasts. Nevertheless, as Kanter and Thorson[12] note, we would not advocate important policy changes in our actions if the theoretically predicted consequences were not at least plausible. Because this is the case, plausibility is likely to be a necessary although certainly not a sufficient condition for evaluating the validity of a forecast. This is especially true when our forecast assumes policy relevance.

One method of estimating plausibility is to consult with people who deal with the empirical domain being projected. Policy planners, for example, often have expectations about the phenomena with which they operate routinely and they make informal judgments regarding the probable consequences of actions. The evaluation of these experts offers a valuable source of information. Indeed, this is likely to be an area of the policy maker's comparative advantage with which social scientists interested in making policy inputs will have to pay more attention in the future.

Another method of testing plausibility is to see whether the forecast violates any logical constraints. Occasionally, a theory which generates plausible forecasts when the values of variables are held to expected or previous levels, yields absurd results if certain values exceed "normal" levels. For example, education planners argued for a theory which predicted exponential enrollment growth. Predictions from

---

[12]A. Kanter and S.J. Thorson, "The Weapons Procurement Process: Choosing Among Competing Theories," Public Policy 20 (Fall 1972).

the theory seemed to fit the data very well until about 1969. After
that point the model predicted exceedingly larger student enrollments.
By the year 2050 the number of U.S. college students was predicted to
exceed the total predicted population of the United States.[13] Systems
stressing of this kind is frequently ignored becasue the theory makes
quite plausible predictions in shorter time frames or for more normal
ranges of events. A "quick and dirty" sensitivity test may reveal that
much of the process about which a theory forecasts is not yet understood.

Turning to the empirical aspects of validation, one of the important
questions concerns how much of a theory need be included in the statistical
attempts at verification. In complex theories with a large number of
variables, one possible strategy is to treat the theory in subdivisions
with forecasts from each module. Obviously in large, rich theories it
would be desirbale from both a financial aspect as well as a logical
aspect to test subsections independently. Computer costs reach astronomical
levels when the number of variables and interrelationships becomes large.
In addition, it becomes increasingly difficult to identify the reason
for errors in forecasts when using numerous variables. This problem is
especially acute when we have reason to believe that the independent
variables are not lineraly independent of each other.[14] Ando, Fisher

_____

[13] This example is described more fully in A. Kanter and S.J. Thorson,
"The Weapons Procurement Process."

[14] See N.R. Draper and H. Smith, Applied Regression Analysis (New York:
Wiley, 1968), and M. Ezekiel and K.A. Fox, Methods of Correlational
Regression Analysis (New York: Wiley, 1959).

and Simon[15], however, have demonstrated that if we are dealing with linear systems in our theory and our system is completely decomposable (that is, the variance to be accounted for is explanable by the variables in each decomposed subset), we will not do an injustice to our theory by validating each of the subsections independently. They proceed to show that it is more frequently the case that the subsystems are only partially decomposable (most, but not all, variance is explanable by variables within the subset). In such cases the subsystems can be treated independently only over short periods of time. Over long periods of time interaction between subsystems becomes dominant. Thus in longer range forecasting it is generally an unwise strategy to attempt to break a theory into more manageable subsets having fewer variables. This conclusion is similar to that of George who suggests that, at least for policy-making, theories with more variables may have greater utility.[16]

The number of statistical techniques potentially useful in testing the validity of forecasts is extremely large.[17] Most of them require additional assumptions not required in cross-sectional analysis, however. For example, if we want to determine the relative importance of particular

---

[15]A. Ando, F.M. Fisher, and H.A. Simon, eds., _Essays on the Structure of Social Science Models_ (Cambridge: MIT Press, 1963).

[16]A. George, "Introduction," in A.L. George, D.K. Hall and W.R. Simons The Limits of Coercive Diplomacy (Boston: Little Brown, 1971), p. xvi.

[17]For a discussion of specific tests, see T.H. Naylor, ed., _Computer Simulation Experiments with Models of Economic Systems_ (New York: Wiley, 1971).

independent variables using normal variance accounting techniques, generally the ordinary least squares is not an appropriate technique for testing the significance of each variable. Hibbs[18] states that if auto correlation occurs in our disturbance terms, ordinary least squares leads to a serious overestimation of the impact of independent variables. This impact can be subdivided into two particular classes. In the first case, when there are no lag variables in the analysis, the overestimation effects do not influence the prediction of the regression coefficient but they do affect the importance of the T test or the multiple R^2. In the second case where lag variables are included in the analysis, not only are the above affects noticed, but the actual level of the regression coefficients is influenced in such a way that usually the non-lagged variables' importance is decreased and the lag variable's importance is increased. These increases and decreases can be of a magnitude of three to four hundred percent.

Another factor in the validation of forecasts from a theory is the need for consistence in the level of aggregation employed in the theory, the forecast, and the test data. If, for instance, the unit of time employed in our forecast is the foreign policy act but the data are aggregated into monthly or yearly units, the identification of the true explanatory variables is difficult.[19] The reason is that the across

---

[18]D.A. Hibbs, "Problems of Statistical Estimation and Causal Inference in Dynamic, Time-Series Regression Models." Paper prepared for delivery at the 1972 meetings of the American Political Science Association, Washington, D.C.

[19]J. Johnston, Econometric Methods (New York: McGraw-Hill, 1972), especially Chapter 12.

time fluctuations considered important in making forecasts would be lost or obscured in the larger units of analysis.[20] That particularly novel results can be achieved without due consideration of the theoretical implications for choosing different time frames or making differing assumptions about the auto regressive affects of error is certainly not a new finding. Yule[21] demonstrated that varying the lags in one's data can produce contradictory expectations.

The importance of the Ando and Fisher theory,[22] the summary of auto correlative effects by Hibbs[23] and the levels of analysis problem is that particular care must be taken when one begins the statistical validation of forecasts. It is important to keep in mind that we cannot simply rely on statistical analysis free from theoretical concerns to derive a validated forecast.

[20]G.M. Orcutt, H.W. Watts, and J.P. Edwards, "Data Aggregation and Information Loss," American Economic Review 68 (September 1968).

[21]G.U. Yule, "Why Do We Sometimes Get Nonsense Correlations Between Time--Series?" Journal of the Royal Statistical Society, LXXXIX (January 1926).

[22]A. Ando, F.M. Fisher, and H.A. Simon, Essays on the Structure of Social Science Models.

[23]D.A. Hibbs, "Problems of Statistical Estimation."