Volume Title: Annals of Economic and Social Measurement, Volume 5, number 3

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm76-3

Publication Date: July 1976

Chapter Title: Econometric Policy Model Construction: The Post-Bayesian Approach

Chapter Author: Arnold M. Faden, Gordon C. Rausser

Chapter URL: http://www.nber.org/chapters/c10485

Chapter pages in book: (p. 349 - 363)

Annals of Economic and Social Measurement, 5/3, 1976

ECONOMETRIC POLICY MODEL CONSTRUCTION: THE POST-BAYESIAN APPROACH‡

BY ARNOLD M. FADEN* AND GORDON C. RAUSSER†

The recent bayesian revival constitutes a searching critique of orthodox statistical procedures, but is itself not free of difficulties. Its prescription imposes in general a crushing computational burden if taken literally. In practice, even avowed bayesians resort to drastic simplifications (e.g., conjugate distributions), and researchers in general seem to deviate considerably from this procedure.

The "post-bayesian" approach takes formal account of this need for simplification. Specifically, in making cognitive judgments one balances the cost of inaccuracy *against the cost of* complexity *of the various alternatives. Here "cognitive judgments" include the entire realm of statistical inference—selection of models, testing of hypotheses, estimation, prediction, etc. "Inaccuracy" refers not to deviations from the true state of nature as in conventional decision theory, but to deviations from one's personal probability distribution as justified by prior assessments and available information. That is, one deliberately distorts one's assessments for the sake of tractability incurring a (hopefully) small inaccuracy cost for a large reduction in complexity cost.*

We apply these ideas here to some problems of prediction and control, the trade-offs being the complexity cost of including more predictor or control variables vs. the inaccuracy cost of missing the true or target values. The analyses are compared to the straight bayesian approach of Lindley, who covers a similar range of problems.

The applications of econometrics to policy questions have grown dramatically in recent decades. Sophisticated techniques have evolved for the estimation of parameters, system identification, testing of models, setting of objective functions, incorporation of new data, etc.[1] And yet there are certain doubts about the validity of much of this literature—doubts not merely about minor points but about the very foundations themselves.

More specifically, the criteria used in economic policy applications are generally borrowed from statistical theory in a fairly uncritical manner—mostly from conventional statistics, as in the use of maximum likelihood methods, confidence intervals, significance tests and the like but also (especially in the control literature) from bayesian statistics. Now these criteria are themselves under attack. The recent bayesian revival constitutes a serious challenge to the validity of most conventional methods; on the other hand, bayesian methods themselves have certain shortcomings. In particular, a rigorous bayesian would need superhuman abilities—a perfect and infinite memory, perfect deductive powers, including faultless and instantaneous calculating ability, and the ability to understand questions of arbitrary complexity.

Our basic approach is that the entire process of specifying, estimating, testing and applying models is itself an economic activity which should be judged by economic criteria, viz., the costs and benefits associated with alternative ways of

* Associate Professor of Economics, Iowa State University.

† Professor of Economics and Statistics, Iowa State University.

‡ The authors wish to thank two anonymous referees for valuable comments.

[1] Many of the early developments in this field of investigation are summarized in Tinbergen [1956], Theil [1964] and Fox, et al., [1973] while more recent developments are available in Aoki [1967, 1975], Chow [1975], Heal [1973].

organizing the research. This approach results in a strategy which involves the balancing between two costs: the cost of *complexity* and the cost of *inaccuracy* due to abstraction or distortion. It turns out that the approach yields criteria which differ substantially from both bayesian and conventional prescriptions.[2]

The post-bayesian approach attempts not only to provide a superior prescriptive theory than the bayesian or any conventional approach, but also a superior descriptive theory. The novelty of the approach involves the explicit introduction of complexity costs. The incorporation of this notion brings us closer to the way scientists actually do behave. Hence, by coming closer to the structure of costs and benefits we can presumably provide a sounder guide for actual practice.

1.1 *Complexity.* One dimension of the cost benefit structure of alternative research strategies emanates from complexity. An intractable model is useless except as an educational instrument for forging more tractable models. And even tractable models differ considerably in complexity. Cost here may take the form of money, time, resources or effort used in developing models, storing and retrieving information, adapting them to various applications, solving them, and communicating their results to others. Some illustrative examples: *ceteris paribus*, linear models are simpler than non-linear, deterministic models are simpler than stochastic models, equilibrium models are simpler than dynamic models, lumped parameter models are simpler than distributed parameter models; in general, complexity rises with the number of free parameters.

To indicate how one would actually assess complexity costs, consider the problem of alternative regression models aimed, say, at predicting a certain variable of interest. Complexity will rise with the inclusion of every new explanatory variable, and it is reasonable to assume that complexity cost is a function of the number of explanatory variables. But which function? Certain aspects of cost go up linearly with the number of variables; e.g., tabulating the data; some go up quadratically, e.g., printing the covariance matrix; some go up cubically, e.g., inverting the moment matrix. These are not the only costs, but they suggest that a cubic polynomial may be one possible representation of complexity cost.

1.2 *Inaccuracy.* The other important aspect of model construction is accuracy. The more accurate a model is, the more benefit accrues from employing it to resolve various policy issues. Or, turned around, there is a cost associated with inaccuracy. But what is inaccuracy, and how does one measure its cost? We take the bayesian prescription as the ideal of perfect accuracy. That is, if (X_i), $i \in I$, is the family of random variables in which one is interested, the perfectly accurate model would be the probability distribution over this family which is justified by the available evidence and one's prior beliefs.[3] As a rule, however, such a distribution would be completely intractable, and so one resorts to simplifying approximations.

The cost of an inaccurate model depends on how it is used. That is, for models that are used as guides in making decisions, inaccuracy tends to degrade the

[2] For a more detailed examination of the foundations of this approach, see Faden and Rausser [1975]. In this paper, the approach is also applied to a number of traditional statistical problems—estimating (or testing) the mean of a normal distribution, point estimation in general, simple hypothesis testing, and optimal roundoff.

[3] A random variable refers to some unknown quantity and does not necessarily involve the idea of repeatability. This accords with the bayesian outlook and also with the language of stochastic processes.

quality of the decision. This implies that to assess the costs of inaccuracy, one must embed the model in a more complete policy framework. There are several ways of making this embedding, each generally leading to a different inaccuracy cost function. Thus, there is no absolute "metric" for inaccuracy.

Our contention is that the entire realm of statistical procedures should be reconstructed in terms of the framework set out above. Questions of estimation, hypothesis testing and the like should be answered by selecting the model which minimizes the total cost of complexity plus inaccuracy. As previously noted, the results of such an approach differ sharply from the recommendations of conventional statistical procedures.

In this paper, we explore some of these results in the context of both prediction and control. One surprising result is that the conventional dichotomy between estimation and testing seems to dissolve.[4] Specifically, the problem is formulated as one of estimation, but the solution is qualitatively what would arise from the problem of deciding which regression coefficients are significantly different from zero or which control variables should be set to zero. The source of this outcome is the discontinuity of the complexity cost function.

Our aim in the following examples is to find the structure that an optimal model would have, taking account of both complexity and inaccuracy costs. To make this approach operational would require the specification of an explicit complexity cost function. We have not attempted this.

1.3 *Costs of Computing vs. Costs of Using Models.* A further problem in implementing this approach is the very cost of finding the optimal model. In the following examples the search problem reduces to a combinatorial one, possibly of rather large size. Clearly, it will not do for the "cost of assessing the costs" to exceed the gain from model improvement. In this connection the following points should be noted.

First, the trade-off between complexity and inaccuracy operates for methods as well as for models, including methods of searching for a good model. Thus, non-exhaustive searches leading to generally sub-optimal models may be justified by the saving in searching costs. Specifically, the following examples involve a search over the integers $K_1 = 0, 1, \ldots, K$, K_1 being the number of non-zero regression coefficients or active controls, and where the figure of merit for each integer may involve a complex computation. A heuristic procedure that suggests itself is to search for a "local" optimum, i.e., an integer that does better than its N nearest neighbors, where N is small compared with K. The best N is then itself the subject of a "higher-order" search.

Second, the more important and general the problem to which the model is addressed, the greater the level of complexity which is justified, both in the model itself and in the methods used in searching for and constructing it. Thus, an exhaustive search might be justified for a model which is to be used over and over, but not for a "one-shot" model.

2. Illustrative Prediction and Control Applications

2.1 *Optimal Prediction.* Consider a regression model $Y = X\beta + U$ in which the objective is to estimate β so as to predict Y accurately. Inaccuracy loss is

[4] This phenomenon also occurs in several of the models investigated in Faden and Rausser [1975].

quadratic in the prediction error. Complexity loss goes up with the number of nonzero components of vector $\hat{\beta}$.

Though framed as an estimation problem, this can also be thought of as "testing the significance of the components of β." The structure of the test is, however, quite different from the usual one, as might be expected since the *truth* of $\beta_i = 0$ is not really the question at issue.

As usual, random variables $Y(T, 1)$ and $X(T, K)$ of rank K are observed; scalar $\tilde{Y}$ and $\tilde{X}(1, K)$ are not yet observed, $\beta(K, 1)$ and scalar σ^2 are unknown.

Of the following assumptions, (1) is the normal regression model, and (2) a weak version of the same. (3) is the formal expression of "complete ignorance" concerning β. (4) may be thought of as the bayesian counterpart of "estimating the second-moment matrix of $\tilde{X}$ by the average second-moment of the observations X". It will generally not be satisfied exactly even when $\tilde{X}$ and the rows of X are independently and identically distributed, but will be a good approximation in this case for large T.[5]

Assume:

(1) $$P(Y|\tilde{X}, X, \beta, \sigma^2) \sim N(X\beta, \sigma^2 I_T);$$

(2) $$E(\tilde{Y}|\tilde{X}, X, \beta, Y) = \tilde{X}\beta;$$

(3) $$P(\beta|\tilde{X}, X, \sigma^2) \propto 1;$$

(4) $$E(\tilde{X}'\tilde{X}|X, Y) = \frac{X'X}{T};$$

and the variance terms in the following proof are all finite.

Let $\hat{\beta}$ be the estimator of β, and let $e = \tilde{Y} - \tilde{X}\hat{\beta}$ be the forecasting error. Then,

Theorem 1: $E(e^2|X, Y) = \dfrac{\hat{\beta}'X'X\hat{\beta}}{T} - \dfrac{2\hat{\beta}'X'Y}{T}$ + terms not involving $\hat{\beta}$.

Proof:

(5) $$P(\beta|\tilde{X}, X, Y, \sigma^2) \propto P(\beta, Y|\tilde{X}, X, \sigma^2) \propto \exp[-(X\beta - Y)'(X\beta - Y)/2\sigma^2]$$

which yields

(6) $$P(\beta|\tilde{X}, X, Y, \sigma^2) \sim N[(X'X)^{-1}X'Y, \sigma^2(X'X)^{-1}].$$

Let $W = \tilde{X}\beta$; then

(7) $$P(W|\tilde{X}, X, Y, \sigma^2) \sim N[\tilde{X}(X'X)^{-1}X'Y, \sigma^2\tilde{X}(X'X)^{-1}\tilde{X}'].$$

Hence,

(8) $$E(W|\tilde{X}, X, Y, \sigma^2) = \tilde{X}(X'X)^{-1}X'Y.$$

[5] In the univariate case ($K = 1$), assumption (4) is satisfied exactly if $\tilde{X}$ and the components of X are independent with common density function

$$p(x|h) = hx \exp(-hx^2/2), \qquad x > 0,$$

the parameter h being itself uniformly distributed over the positive halfline.

But also

$$E(\tilde{Y}|\tilde{X}, X, Y, \beta) = \tilde{X}\beta = W, \tag{9}$$

hence,

$$E(\tilde{Y}|\tilde{X}, X, Y) = \tilde{X}(X'X)^{-1}X'Y. \tag{10}$$

This yields

$$E(e|\tilde{X}, X, Y) = \tilde{X}[(X'X)^{-1}X'Y - \hat{\beta}] \tag{11}$$

so that

$$E(e^2|\tilde{X}, X, Y) = [(X'X)^{-1}X'Y - \hat{\beta}]'\tilde{X}'\tilde{X}[(X'X)^{-1}X'Y - \hat{\beta}] + \text{variance}(e|\tilde{X}, X, Y). \tag{12}$$

Now, variance $(e|\tilde{X}, X, Y) =$ variance $(\tilde{Y}|\tilde{X}, X, Y)$ does not involve $\hat{\beta}$. Taking expectations conditional on X, Y, we obtain

$$E(e^2|X, Y) = [(X'X)^{-1}X'Y - \hat{\beta}]'\frac{X'X}{T}[(X'X)^{-1}X'Y - \hat{\beta}] + \text{res} \tag{13}$$

$$= \frac{\hat{\beta}'X'X\hat{\beta}}{T} - \frac{2\hat{\beta}'X'Y}{T} + \text{res},$$

where res does not involve $\hat{\beta}$. QED.

The optimal estimator $\hat{\beta}$ (which may depend on the observations) is that minimizing the sum of two terms, the forecasting loss $E(e^2|X, Y)$, and the loss from allowing the complication $\hat{\beta}_i \neq 0$.

Theorem 2: The optimal $\hat{\beta}$ has one of the following 2^K forms: partition $X = (X_1X_2)$ and correspondingly $\hat{\beta} = \begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{pmatrix}$, set $\hat{\beta}_2 = 0$ and $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y$[6].

Proof. Given $\hat{\beta}$, let $\hat{\beta}_1$ be the nonzero components of $\hat{\beta}$. From Theorem 1, the inaccuracy loss is then

$$\frac{\hat{\beta}_1'X_1'X_1\hat{\beta}_1}{T} - \frac{2\hat{\beta}_1'X_1'Y}{T} + \text{res}. \tag{14}$$

As $\hat{\beta}_1$ varies, complexity cost does not rise, so $\hat{\beta}_1$ should be chosen to minimize inaccuracy loss, and this occurs at

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y. \quad \text{QED.} \tag{15}$$

[6] A referee has redone this analysis with assumption (4) weakened to $E(\tilde{X}'\tilde{X}|X, Y) = W$, a general positive definite matrix. Partitioning W conformably to $(X_1\ X_2)$ he obtains

$$\hat{\beta}_1 = [I_{K_1}\ W_{11}^{-1}W_{12}](X'X)^{-1}X'Y$$

in place of (15), K_1 being the number of columns in X_1.
Also, (16) should be replaced by

$$R_1^2 = \frac{Y'X(X'X)^{-1}\begin{bmatrix}W_{11} & W_{12} \\ W_{21} & W_{21}W_{11}^{-1}W_{12}\end{bmatrix}(X'X)^{-1}X'Y}{Y'Y}.$$

Note that this result is equal to the posterior mean of β_1, conditional on $\beta_2 = 0$.

Thus, the problem is reduced to a combinatorial one. Assume $Y'Y \neq 0$, and define[7]

$$R_1^2 = \frac{Y'X_1(X_1'X_1)^{-1}X_1'Y}{Y'Y}. \tag{16}$$

Theorem 3: The optimal solution $\hat{\beta}$ has the following structure. If there are K_1 nonzero terms, then the given nonzero subvector is that maximizing R_1^2 over all possible subvectors of size K_1.

Proof: Fixing the number of nonzero components fixes complexity cost, hence the optimal subvector is that minimizing inaccuracy cost. Substituting the value $\hat{\beta} = (X_1'X_1)^{-1}X_1'Y$ yields

$$E(e^2|X, Y) = -\frac{Y'X_1(X_1'X_1)^{-1}X_1'Y}{T} + \text{res.}$$

The minimizer of this is the maximizer of

$$\frac{Y'X_1(X_1'X_1)^{-1}X_1'Y}{T} = \frac{Y'Y}{T}R_1^2. \qquad \text{QED.}$$

2.2 *Optimal Control.*[8] Consider first the case of one-stage control. This again is a regression model with vector components $Y(T, 1)$, $X(T, K)$ of rank K, $\beta(K, 1)$, $\tilde{Y}$ scalar, $\tilde{X}(1, K)$ as above. The interpretation of $\tilde{Y}$ and $\tilde{X}$ is, however, quite different; $\tilde{X}$ is now a control vector, subject to choice. Choosing $\tilde{X}$ yields a value $\tilde{X}\beta + \tilde{U}$. As before, β and scalar $\tilde{U}$ are unknown, distributed independently of each other and of $\tilde{X}$. The object is to hit a target value $\tilde{Y}$. $\tilde{Y}$ itself may not be known exactly and has a distribution which is independent of β and of $\tilde{X}$.

The loss from missing the target is assumed quadratic:

$$(\tilde{X}\beta + \tilde{U} - \tilde{Y})^2. \tag{17}$$

In addition, there is a complexity loss which increases with the number of active controls. Specifically, partition $\tilde{X}$ and β correspondingly into active $\tilde{X}_1$, β_1, and passive $\tilde{X}_2$, β_2 parts, so that

$$\tilde{X}\beta = \tilde{X}_1\beta_1 + \tilde{X}_2\beta_2. \tag{18}$$

In what follows, $\tilde{X}_2\beta_2$ is set equal to zero.[9] The complexity loss then depends on K_1, the number of components in X_1 or β_1.

Let $b_1 = E\beta_1$, $\Omega_1 = E(\beta_1\beta_1')$, and let V_1 be the covariance matrix of β_1 so that $\Omega_1 = V_1 + b_1b_1'$. (V_1 and Ω_1 are assumed to be invertible.)

[7] R_1^2 coincides with the ordinary coefficient of determination if $\bar{Y} = 0$, and if either X has a constant column or $\bar{X} = 0$ ($-$ indicates the average of the T observations).

[8] This model is similar to that of Prescott [1971, 1972]. For an exposition, see Zellner [1971, Chap. XI].

[9] This may be interpreted either as accepting a null hypothesis $\beta_2 = 0$, or as setting the controls $\tilde{X}_2$ to zero. Zero may be thought of as "status-quo" setting, any alteration of which incurs an overhead cost in use.

Theorem 4: The optimal control setting has the following structure. $b_1'\Omega_1^{-1}b_1$ is maximized over all possible active subvectors having K_1 components. The optimal setting is

$$\tilde{X}_1 = (E\tilde{Y} - E\tilde{U})b_1'\Omega_1^{-1}.$$

Proof. Fixing the number of active components fixes complexity loss, so $\tilde{X}_1$ should be chosen to minimize inaccuracy loss, which is

$$E(\tilde{X}_1\beta_1 + \tilde{U} - \tilde{Y})^2 = \tilde{X}_1\Omega_1\tilde{X}_1' + 2\tilde{X}_1 b_1(E\tilde{U} - E\tilde{Y}) + E(\tilde{U} - \tilde{Y})^2. \tag{19}$$

The minimizer of this expression is

$$\tilde{X}_1 = (E\tilde{Y} - E\tilde{U})b_1'\Omega_1^{-1}, \tag{20}$$

and the minimum value is

$$-(E\tilde{Y} - E\tilde{U})^2 b_1'\Omega_1^{-1}b_1 + E(\tilde{U} - \tilde{Y})^2. \tag{21}$$

Hence the optimal active subset is the one maximizing $b_1'\Omega_1^{-1}b_1$. QED.

Now, suppose that information concerning β has been obtained from previous observations of X, Y in the regression relation

$$Y = X\beta + U. \tag{22}$$

(Here, the vector Y does not represent the preceding target values but instead the attained values resulting from the settings of X.) We assume $Y'MY > 0$, where $M = I_T - X(X'X)^{-1}X'$; $T > K + 2$; the standard linear model with (unknown) precision h,

$$P(Y|\tilde{X}, X, \beta) \sim N(X\beta, I_T/h); \tag{23}$$

and a noninformative prior (Jeffreys [1961]), i.e.,

$$P(\beta, h|\tilde{X}, X) \propto 1/h. \tag{24}$$

The posterior distribution of β, conditional on $\tilde{X}$, X, Y, is then a multivariate t with mean $b = (X'X)^{-1}X'Y$ and covariance matrix

$$V = Y'MY(X'X)^{-1}/(T - K - 2). \tag{25}$$

(Zellner [1971, p. 67, 383]).

Now, for any partition $(\tilde{X}_1\ \tilde{X}_2)$ of $\tilde{X}$ into active and passive components, partition X conformably into $(X_1\ X_2)$, and let

$$M_2 = I_T - X_2(X_2'X_2)^{-1}X_2'. \tag{26}$$

The following result then restates Theorem 4 in terms of X_1, X_2, and Y.

Theorem 5: The optimal control setting has the following structure: The expression

$$Y'M_2X_1(X_1'M_2X_1)^{-1}X_1'M_2Y \tag{27}$$

is maximized over all possible partitions $(X_1\ X_2)$ of X, where X_1 contains K_1 columns. The optimal setting is

$$\tilde{X}_1 = \frac{(E\tilde{Y} - E\tilde{U})Y'M_2X_1}{[Y'MY/(T-K-2)] + Y'M_2X_1(X_1'M_2X_1)^{-1}X_1'M_2Y}. \tag{28}$$

Proof:

$$(X'X)^{-1} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} (X_1'M_2X_1)^{-1} & -(X_1'M_2X_1)^{-1}X_1'X_2(X_2'X_2)^{-1} \\ \cdots & \cdots \end{bmatrix} \tag{29}$$

hence

$$V_1 = Y'MY(X_1'M_2X_1)^{-1}/(T-K-2), \tag{30}$$

and

$$b_1 = (X_1'M_2X_1)^{-1}[X_1' - X_1'X_2(X_2'X_2)^{-1}X_2']Y = (X_1'M_2X_1)^{-1}X_1'M_2Y. \tag{31}$$

Now,

$$\Omega_1^{-1} = [V_1 + b_1b_1']^{-1} = V_1^{-1} - \frac{V_1^{-1}b_1b_1'V_1^{-1}}{1 + b_1'V_1^{-1}b_1}; \tag{32}$$

thus,

$$b_1'\Omega_1^{-1} = \frac{b_1'V_1^{-1}}{1 + b_1'V_1^{-1}b_1}, \tag{33}$$

and

$$b_1'\Omega_1^{-1}b_1 = \frac{b_1'V_1^{-1}b_1}{1 + b_1'V_1^{-1}b_1}. \tag{34}$$

Hence, maximization of $b_1'\Omega_1^{-1}b_1$ is equivalent to maximization of

$$b_1'V_1^{-1}b_1 = \frac{Y'M_2X_1(X_1'M_2X_1)^{-1}X_1'M_2Y}{Y'MY/(T-K-2)}. \tag{35}$$

Since $Y'MY > 0$ this proves the first statement of the theorem. Finally, the optimal setting

$$\tilde{X}_1 = \frac{(E\tilde{Y} - E\tilde{U})b_1'V_1^{-1}}{1 + b_1'V_1^{-1}b_1}, \tag{28a}$$

and substitution yields the last statement. QED.

One special case may be noted. If the columns of X are orthogonal, then $M_2X_1 = X_1$, and the optimal active subvector is the one in which

$$Y'X_1(X_1'X_1)^{-1}X_1'Y \tag{36}$$

is a maximum, which is the same criterion as in the prediction model above.

Turning to N-stage control some rather obvious results are immediate from Theorems 4 and 5. Let X_t be the control setting at stage t, $t = 1, \ldots, N$, and consider first the case in which the settings on X_t must all be announced at the

beginning of the first stage or planning period and remain unaltered thereafter. Under these circumstances, information provided by Y_t in determining the optimal setting for X_{t+i}, $i>0$, cannot be employed and thus the choice of X_t will not affect the optimal determination for subsequent controls, i.e. X_{t+i}, $i>0$. Hence, this N-stage problem reduces to N nonrecursive one-period problems.

A few interesting implications of the N-stage control problem as specified here, however, can be obtained. In particular, the effects of variations in the distributions of target values or disturbances on X_t and K_{1t} across stages of the planning horizon can be ascertained. To examine these issues we shall let the total cost be the sum of the costs incurred at each stage. Furthermore, complexity cost at stage t is presumed an increasing function, C, of the number of active control components, K_{1t}, utilized at that stage, the function itself being invariant in time. Inaccuracy cost at stage t is again the squared error at that stage. Thus, we wish to minimize the expectation of

$$\sum_{t=1}^{N} (\tilde{Y}_t - X_t\beta - \tilde{U}_t)^2 + \sum_{t=1}^{N} C(K_{1t}). \tag{37}$$

Here $\tilde{Y}_t$ and $\tilde{U}_t$ are the (scalar) target values and disturbances at stage t. These will in general be unknown and are allowed to change from stage to stage. X_t is the $(1, K)$ control vector for stage t, the number of non-zero components of X_t being of course K_{1t}. β is independent of $\tilde{Y}_1, \ldots, \tilde{Y}_N, \tilde{U}_1 \ldots \tilde{U}_N$, and the joint distribution of these variables is unaffected by the choice of any $X_1, \ldots, X_N$.

Now suppose that the control vectors $X_1, \ldots, X_N$ must be chosen in advance, and suppose there is a unique optimal solution $X_1^0, \ldots, X_N^0$. Then we have the following results.

Theorem 6: Let t', t'' be any two stages.

(i) If

$$E\tilde{Y}_{t'} - E\tilde{U}_{t'} = E\tilde{Y}_{t''} - E\tilde{U}_{t''},$$

then

$$X_{t'}^0 = X_{t''}^0;$$

(ii) If

$$|E\tilde{Y}_{t'} - E\tilde{U}_{t'}| \geq |E\tilde{Y}_{t''} - E\tilde{U}_{t''}|,$$

then

$$K_{1t'} \geq K_{1t''}.$$

That is, the number of active components at t' is not less than the number of active components at t''.

(iii) If $K_{1t'} = K_{1t''}$, then $X_{t'}^0$ and $X_{t''}^0$ are proportional to each other.

Proof: (i) The expected cost at stage t is

$$X_t \Omega X_t' + 2X_t b(E\tilde{U}_t - E\tilde{Y}_t) + E(\tilde{U}_t - \tilde{Y}_t)^2 + C(K_{1t}). \tag{38}$$

where $\Omega = E(\beta\beta')$ and $b = E\beta$. If $E\tilde{U}_{t'} - E\tilde{Y}_{t'} = E\tilde{U}_{t''} - E\tilde{Y}_{t''}$, then the expressions (38) for t' and t'' differ from each other only in a term not involving the control vector. Hence a given X^0 minimizes the t'-expression iff it minimizes the expression for t''. By uniqueness of solution, $X_{t'}^0 = X_{t''}^0$.

(ii) Let X_{1t}, $X_{1t''}$ be the active subvectors of $X_{t'}^0$, $X_{t''}^0$, respectively; $\beta_{1t'}$, $\beta_{1t''}$ be the corresponding subvectors of β; $b_{1t'} = E\beta_{1t'}$; $b_{1t''} = E\beta_{1t''}$; $\Omega_{1t'} = E(\beta_{1t'}, \beta'_{1t'})$; and $\Omega_{1t''} = E(\beta_{1t''}, \beta'_{1t''})$.

We shall assume that $K_{1t'} < K_{1t''}$ and reach a contradiction. Throughout we assume that $|E\tilde{Y}_{t'} - E\tilde{U}_{t'}| > 0$, since otherwise (ii) is trivially correct. The costs incurred by $X_{t'}^0$ at stage t' are, from (21),

$$-(E\tilde{Y}_{t'} - E\tilde{U}_{t'})^2 b'_{1t'}\Omega_{1t'}^{-1}b_{1t'} + E(\tilde{U}_{t'} - \hat{Y}_{t'})^2 + C(K_{1t'}). \tag{39}$$

Since $X_{t'}^*$ is optimal, (39) cannot exceed the total costs incurred by any other vector, in particular by the vector whose active components $(E\tilde{Y}_{t'} - E\tilde{U}_{t'})b'_{1t'}\Omega_{1t''}^{-1} = Z$, say, where these components match those of $X_{t''}^0$. The inaccuracy cost incurred by this vector is

$$E(Z\beta_{1t''} + \tilde{U}_{t'} - \hat{Y}_{t'})^2 = -(E\tilde{Y}_{t'} - E\tilde{U}_{t'})^2 b'_{1t'}\Omega_{1t''}^{-1}b_{1t''} + E(\tilde{Y}_{t'} - \hat{U}_{t'})^2 \tag{40}$$

Hence we get the inequality

$$-(E\tilde{Y}_{t'} - E\tilde{U}_{t'})^2 b'_{1t'}\Omega_{1t'}^{-1}b_{t'} + C(K_{1t'}) \leq -(E\tilde{Y}_{t'} - E\tilde{U}_{t'})^2 b_{1t''}\Omega_{1t''}^{-1}b_{1t''} + C(K_{1t''}). \tag{41}$$

In fact, this is a strict inequality, since $Z \neq X_{1t''}^*$, and the solutions are unique.

A similar argument applied to $X_{t''}^0$ yields (41) with t' and t'' interchanged throughout. Adding (41) to this latter inequality, and simplifying, we obtain

$$0 < [(E\tilde{Y}_{t'} - E\tilde{U}_{t'})^2 - (E\tilde{Y}_{t''} - E\tilde{U}_{t''})^2][b'_{1t'}\Omega_{1t'}^{-1}b_{1t'} - b'_{1t''}\Omega_{1t''}^{-1}b_{1t''}]. \tag{42}$$

The first bracketed expression must be positive by assumption, hence

$$b'_{1t'}\Omega_{1t'}^{-1}b_{1t'} > b'_{1t''}\Omega_{1t''}^{-1}b_{1t''}. \tag{43}$$

But this contradicts optimality of $X_{t''}^0$. For, by minimizing over vectors with $K_{1t'} < K_{1t''}$ active components one could reduce complexity cost and, (by equation 21), inaccuracy cost as well. This proves that $K_{1t'} \geq K_{1t''}$.

(iii) The optimal active subset at t' is the one maximizing $b'_1\Omega_1^{-1}b_1$ over all subsets with $K_{1t'}$ components by Theorem 4. By uniqueness of solution there is just one such maximizing subset. Since $K_{1t'} = K_{1t''}$, the optimal active subset at t'' solves the same maximization problem, hence the active subsets coincide. The active subvectors themselves are proportional to $b'_1\Omega_1^{-1}$ by (20), hence proportional to each other. Q.E.D.

The intuitive explanation for part (ii) of Theorem 6 is that a larger absolute $E\tilde{Y}_t - E\tilde{U}_t$ makes inaccuracy cost more sensitive to a change in control. Hence, it may pay to accept more complications to obtain a better "fix" on the target. Part (iii) shows that the number of active components actually determines the optimal control vector, up to a scale factor.

If information from previous stages is employed to determine current controls but not conversely, then we have the case of passive learning or sequential

updating (Rausser and Freebairn [1974]). Here, the solution does take account of information in previous stages to arrive the control setting for the t-th stage. Letting $Y(T, 1)$ and $X(T, K)$ be as above, viz. the data available in advance; $X_\tau(1, K)$ the control setting at stage τ; Y_τ (scalar) the realization (not the target) at stage τ, $\tau = 1 \ldots, t-1$; $Y_t^{*\prime} = [Y' Y_1' \ldots Y_{t-1}']$; $X_t^{*\prime} = [X' X_1' \ldots X_{t-1}']$; $M_{2t}^* = I_{T+t-1} - X_{2t}^*[X_{2t}^{*\prime} X_{2t}^*]^{-1} X_{2t}^{*\prime}$; and $M_t^* = I_{T+t-1} - X_t^*[X_t^{*\prime} X_t^*]^{-1} X_t^{*\prime}$ then we have

Theorem 7: The optimal control setting has the following structure for the sequential updating problem: In each stage t, the expression

$$Y_t^{*\prime} M_{2t}^* X_{1t}^* (X_{1t}^{*\prime} M_{2t}^* X_{1t}^*)^{-1} X_{1t}^{*\prime} M_{2t}^* Y_t^* \tag{44}$$

is maximized over all possible partitions (X_1^*, X_{2t}^*) of X_t^*, $t = 1, \ldots, N$, where X_{1t}^* contains K_{1t} columns. The optimal settings are

$$\tilde{X}_{1t}^* = \frac{(E\tilde{Y}_t - E\tilde{U}_t) Y_t^{*\prime} M_{2t}^* X_{1t}^*}{[Y_t^{*\prime} M_t^* Y_t^*/(T+t-K-3)] + Y_t^* M_{2t}^* X_{1t}^* (X_{1t}^{*\prime} M_{2t}^* X_{1t}^*)^{-1} X_{1t}^{*\prime} M_{2t}^* Y_t^*} \tag{45}$$

Proof. Just substitute X_t^*, Y_t^* for X, Y along with $T+t-1$ for T in Theorem 5. QED.

Note that under Theorems 5, 6 and 7 the first stage controls are equivalent; differences arise only with respect to the second and subsequent stages. Furthermore, neither Theorems 6 or 7 admit any influence of subsequent controls on the determination of current control settings, i.e. the determination of current controls is made without taking account of how these settings affect subsequent stage control settings. To account for this influence, an adaptive control approach which explicitly recognizes the experimental design aspects of the problem is required. We are presently investigating this approach in context of the post-bayesian framework advanced in this paper.

2.3 *Lindley's Approach*. These examples invite comparison with the results of Lindley [1968]. He also considers both a prediction and a control problem. In each case there is a quadratic loss from missing the true or target value of the dependent variable and an additional cost depending on the variables selected for observation or control.

In Lindley's control problem (1968, pp. 46–53], the controls are random variables, the uncontrolled variables taking values stochastically according to a distribution conditioned on the selected values of the controlled variables. In our example, of course, the "uncontrolled" variables are kept at their status-quo value of zero. Not surprisingly, the two analyses diverge completely in their recommended selection of controls and settings.

In Lindley's prediction problem [1968, pp. 33–46], $\tilde{X}$ and the rows of X are ultimately assumed to be i.i.d., multinormal, with common random parameters having a Fisher-Cornish prior. (By contrast, (4) is the only distributional assumption we make on the $\tilde{X}$ or X variables). The real contrast, however, is not so much in the different stochastic assumptions, as in the cost structure of the models, which illustrates neatly the distinction between the *bayesian* and *post-bayesian* approaches. Our second cost component refers to the complexity of the model, which can be reduced by dropping the terms $\tilde{X}_2\beta_2$—that is, actually distorting the model in a most unbayesian manner. Lindley's cost, on the other hand, is the cost

of *observing* variables; the analysis is of the standard "preposterior" form [Raiffa–Schlaifer, 1961], balancing this cost against the expected benefit of making the observation.

The essential distinction is brought out in a comment made by Lindley in the discussion following his paper. Considering the case of polynomial regression ($\tilde{x}_j = \tilde{x}^j$, $j = 0, 1, 2, \ldots$), where the cost of observing *all* the components $\tilde{x}_j$ is scarcely greater than the cost of observing *one* of them, Lindley states: "My Bayesian solution would fit a polynomial of degree $n-1$ to n points .... This is absurd .... In practice I would fit a low degree polynomial, but I do not know why, or at least not in any way that I can express precisely. The example is a useful test case for Bayesian methods." [Lindley, 1968, p. 66]. Our reply, of course, is that to justify this common practice one must go beyond Lindley's bayesian approach and consider explicitly the cost of complexity in selecting models.

Finally, note that observation costs may be subsumed under complexity costs: if one excludes a variable from a model, one need not observe it. Thus, Lindley's costs should be incorporated as a contributor to the complexity cost function.[10]

3. Concluding Comments

In the construction and use of econometric models for various purposes, we have argued that conventional procedures are lacking. Their limitations emanate from their failure to explicitly recognize complexity costs and thus the need to balance these costs with the cost of inaccuracy due to abstraction or distortion. The incorporation of these costs leads to what we have characterized as the post-bayesian approach and requires a reexamination of model construction procedures. The result of such an examination is hopefully not only a better prescriptive theory than the bayesian or any conventional approach, but a superior descriptive theory as well.

The main practical thrust of the post-bayesian approach for problems of prediction and control is that it provides a correct formal apparatus for accomplishing what researchers are now doing either on a purely intuitive basis or with the aid of tests which are inappropriate. Somehow a selection of "significant" explanatory variables or "appropriate" control variables must be made from a pool of such variables which is indefinitely large, and the proper estimates or settings made on the variables selected. The post-bayesian approach makes this selection in a systematic way involving the weighing of alternative costs, avoiding the inappropriate tests inherited from conventional statistics.

In practice, of course, we do not generally have accurate estimates of complexity and inaccuracy costs and thus post-bayesian procedures must often be implemented with crude estimates of such costs. Nevertheless, for the illustrative applications considered in this paper, it was possible to employ very crude estimates of these costs to motivate procedures which proved superior to conventional treatments.

[10] It is interesting to note that Lindley's prediction model implies the conclusions of our Theorems 2 and 3 despite its rather different assumptions [1968, p. 42]. Any modification of either model would in general destroy this coincidence. See footnote 6 above, for example.

We intend to go well beyond the prediction and control applications advanced in section 2. More specifically, in addition to our previous work (Faden and Rausser [1975]) and the examples presented in Section 2, we are presently examining the implications of the approach for dealing with aggregation, selection among alternative functional forms, specification and estimation of distributed lag relationships, pooling of cross-section-time-series data, regime changes, determination of the number of classes in discriminant analysis, and the construction of autoregressive-moving average processes. Moreover, some preliminary results are now available on a comparison of the post-bayesian approach with conventional stepwise regression routines (Dahm, et. al. [1975]). It is clear from these results that rather substantial differences are obtained when costs of complexity are explicitly recognized.

Other potential applications where the approach would prove valuable are not difficult to isolate. These applications might be classified under one of three categories, (i) specification of econometric models, (ii) selection of estimation methods, and (iii) selection of policy or control solution methods. The first category covers such questions as whether a model for a particular system should be specified as a set of stepwise recursive, block recursive, or simultaneous equations. The second addresses issues such as the selection of estimation procedures for dynamic, stochastic models which do not admit estimators with determinable small sample properties. Since most economic policy problems require the formulation of a rational, multiperiod decision problem under conditions of imperfect information (an adaptive control problem) for which no analytical control solution is available (Rausser and Freebairn [1974]), the third category is concerned with the selection of "approximate" solution procedures which involve some alterations of the original structure of the problem. The severity of these alterations will depend upon the combined cost of complexity and inaccuracy. The specification of these costs and use of the resulting post-bayesian procedures will allow researchers to determine the optimal degree of approximation to adaptive or dual control problems.

Iowa State University

References

[1] M. Aoki, *Optimization of Stochastic Systems* (Academic Press, 1967).
[2] M. Aoki, *Dynamic Economic Theory and Control* (American Elsevier, 1975).
[3] G. C. Chow, *Analysis and Control of Dynamic Economic Systems* (Wiley, 1975).
[4] F. Dahm, A. M. Faden, and G. C. Rausser, "Explanatory Variable Selection: Stepwise Regression Versus Post-Bayesian Analysis," Unpublished mimeo, Department of Economics, Iowa State University, 1975.
[5] A. M. Faden and G. C. Rausser, "Post-Bayesian Statistical Inference," mimeo (submitted for publication), Department of Economics, Iowa State University, 1975.
[6] K. A. Fox, J. K. Sengupta, and E. Thorbecke, *The Theory of Quantitative Economic Policy* (American Elsevier, second ed., 1973).
[7] G. M. Heal, *The Theory of Economic Planning* (American Elsevier, 1973).
[8] H. Jeffreys, *Theory of Probability* (Clarendon Press, 3rd ed., 1961).
[9] D. V. Lindley, "The Choice of Variables in Multiple Regression," *Journal of the Royal Statistical Society*, Series B, 30:31–53 (discussion 54–66), No. 1, 1968.
[10] E. C. Prescott, "Adaptive Decision Rules for Macroeconomic Planning," *Western Economic Journal*, 9:369–378, 1971.

[11] E. C. Prescott, "The Multi-Period Control Problem Under Uncertainty," *Econometrica*, 40:1043–1058, November 1972.
[12] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*. (Harvard Graduate School of Business Administration, Boston, 1961).
[13] G. C. Rausser and J. W. Freebairn, "Approximate Adaptive Control Solutions to U.S. Beef Trade Policy," *Annals of Economic and Social Measurement*, 3:177–203, January 1974.
[14] H. Theil, *Optimal Decision Rules for Government and Industry*, (Rand-McNally, 1964).
[15] J. Tinbergen, *Economic Policy: Principles and Design* (North-Holland, 1956).
[16] A. Zellner, *An Introduction to Bayesian Inference in Econometrics*, (Wiley, 1971).

National Bureau of Economic Research is a private, nonprofit organization founded in 1920 as a center for independent and impartial economic research. In the belief that such research can contribute significantly to the sound treatment of economic problems, it has sought to conduct its activities under conditions that safeguard the scientific nature of the findings and that help make them authoritative and acceptable to persons of different interests and opinions. The National Bureau is supported in part by grants from philanthropic foundations, business associations, and various government agencies for particular studies; in part by investment income on capital-sum grants; and in part by unrestricted contributions and subscriptions from companies, banks, labor organizations, foundations, and individuals.

Further information regarding contributions can be obtained from the office of Harvey J. McMains, Vice President, National Bureau of Economic Research, 261 Madison Avenue, New York, N.Y. 10016.

The purpose of the *Annals of Economic and Social Measurement* is to promote the exchange of ideas in quantitative research. The journal will contain articles generated in the research program of the National Bureau of Economic Research or relevant to this program, mainly but not solely dealing with data generation, computer applications, and methodological problems. Participants in workshops held by the NBER'S *Conference on the Computer in Economic and Social Research* and researchers affiliated with the NBER'S *Computer Research Center for Economics and Management Science* will be encouraged to contribute papers.

We hope that the *Annals* will serve a catalytic role in improving techniques for economic and social measurement. Although research methodology will be stressed, descriptions of computer programs, information retrieval techniques, and data sets will also be published. Such an emphasis means that scientists can benefit from work in fields outside their immediate range of research. The objective is to provide rapid and wide dissemination of advances in the state of the art—resulting, it is hoped, in more effective use of the computer in quantitative analysis.