Volume Title: Annals of Economic and Social Measurement, Volume 6, number 4

Volume Author/Editor: NBER

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm77-4

Publication Date: October 1977

Chapter Title: Optimal Experimental Design for Dynamic Econometric Models

Chapter Author: Elizabeth Chase Macrae

Chapter URL: http://www.nber.org/chapters/c10524

Chapter pages in book: (p. 399 - 405)

Annals of Economic and Social Measurement, 6/4, 1977

OPTIMAL EXPERIMENTAL DESIGN FOR DYNAMIC ECONOMETRIC MODELS

BY ELIZABETH CHASE MACRAE

A methodology for designing time series experiments is developed through the use of stochastic control theory. One implication that can be drawn is that with less initial information it may be better to postpone most of the information gathering activity, until the results of earlier periods are available to help in designing a more reliable and cost-effective experiment.

1. INTRODUCTION

The growing interest in controlled social experimentation has led economists to devote more attention to the appropriate design of such experiments. In general, the formulation of the design problem involves a trade off between the maximization of information gained by the experiment and the minimization of costs, both to the experimenter and possibly to the subjects of the experiment. When the model under consideration is a classical static regression model, the analysis of experimental design is straightforward and has been discussed by Watts and Conlisk [5]. If, however, the model is dynamic and if time-series data are to be collected then the analysis becomes much more complex.

The purpose of this paper is to use stochastic control theory to develop a methodology for designing time-series experiments. The basic approach to stochastic optimization by MacRae [4] is extended to include a valuation of the stock of information at the termination of the experiment. The experimental design is then derived as a sequence of plans in which the information that becomes available in each period is used to update and refine the design for the remainder of the experiment.

2. PROBLEM STATEMENT

Assume that model under consideration has the form

$$x_k = \mathcal{A}x_{k-1} + \mathcal{B}u_k + \mathcal{C}z_k + \epsilon_k, \quad k = 1, 2, \ldots \tag{2.1}$$

where u_k is a design vector which may be chosen by the experimenter in period k, x_k is a vector of endogenous variables, and z_k is a vector of exogenous variables. Matrices $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are the unknown parameter matrices to be estimated and ϵ_k is a vector of random disturbances, independent over time, with zero mean and variance matrix Ω.

The experimenter is also faced with a function, J_1, which incorporates not only the monetary costs of conducting the experiment but also any social costs or benefits that accrue to the subjects of the experiment. This cost function over the N periods of the experiment is assumed to be quadratic in form:

$$J_1 = \sum_{k=1}^{N} \frac{1}{2} x_k' Q_k x_k + \frac{1}{2} u_k' R_k u_k + s_k' x_k + t_k' u_k, \tag{2.2}$$

where Q_k, R_k, s_k, and t_k are fixed matrices and vectors. If the exogenous variables play a role in the cost function, they are subsumed by the Q, R, s, and t coefficients.

The final element of the experimenter's problem is a measure, J_2, of the accuracy of the parameter estimates at the end of the experiment. Letting Γ_N be the variance-covariance matrix of the estimated parameters as of period N, a natural choice for J_2 would be some scalar function of Γ_N^{-1}, the information matrix. Thus,

$$J_2 = L\{\Gamma_N^{-1}\}, \tag{2.3}$$

where the function L is, for example, a determinant or weighted trace.

The problem facing the experimenter is to determine a sequence of vectors, $u_1, u_2, \ldots, u_N$, so as to minimize J_1 and maximize J_2. This may be handled in three ways. First, the experimenter may choose to minimize costs subject to attaining some given level of information. Second, he may maximize the information gained, subject to some upper bound on costs. Finally, he may minimize a weighted sum of J_1 and $-J_2$. Since, by appropriate manipulation of the weights on J_1 and J_2, solutions can be obtained which are equivalent to the first and second approaches above (the weights taking on the role of Lagrangean multipliers), only the third method will be dealt with explicitly in this paper.

As will become apparent in the next section of the paper, the experimenter must start with prior guesses, A_0, B_0, C_0, at the values of the unknown coefficient matrices, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, as well as a prior value for the inter-equation noise variance, Ω. This prior information is the same as that required for design of experiments in the static structural equation case discussed by Conlisk [1]. In addition, experimentation in a time-series model requires a prior variance-covariance matrix, Σ_0, which measures the uncertainty associated with the prior parameter values, A_0, B_0, C_0. At the beginning of the experiment, the experimenter calculates a series of control vectors, u_1, $u_2, \ldots, u_N$, utilizing his prior guesses. As the observed results of the first period become available, the experimenter revises his guesses or estimates of the unknown parameters, and recalculates the optimum values for the remaining control variables, $u_2, u_3, \ldots, u_N$. Thus as the experiment progresses, more and more information becomes available and the initial guesses at the parameter values may be replaced by better estimates, which in turn are used to update the design of the remainder of the experiment.

3. Solution

The mathematical problem facing the experimenter in each period is to minimize an objective function

$$J = \lambda_1 E\{J_1\} + \lambda_2 J_2 \tag{3.1}$$

which is a weighted sum of the expected cost, J_1, and the information gain, J_2. The minimization is carried out subject to the constraint imposed by the model.

$$x_k = \mathbf{A}x_{k-1} + \mathbf{B}u_k + \mathbf{C}z_k + \epsilon_k \equiv \mathbf{D}w_k + \epsilon_k, \; k = 1, \ldots, N, \tag{3.2}$$

where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are matrices of random variables used to model the uncertainty regarding the constant but unknown parameters, $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$, and where D and w_k are defined as $[A, B, C]$ and $[x'_{k-1}, u'_k, z'_k]'$ respectively. The experimenter's prior guesses at the unknown parameter values, A_0, B_0, and C_0, will be taken as the prior means of the random matrix $\mathbf{D}$ and his guess at Γ_0 will be used as the prior variance-covariance matrix of $\mathbf{D}$.

There is in general no way of obtaining an exact solution to the above stochastic optimization problem except through numerical techniques. Moreover, for problems of any reasonable magnitude, numerical solutions are simply not feasible, and some sort of approximate solution must be developed. The solution to be used here is a straight-forward extension of that presented in [4], in which the random matrix $\mathbf{D}$ is replaced by a sequence of independent random matrices $\mathbf{D}_0, \mathbf{D}_1, \ldots, \mathbf{D}_{N-1}$, whose means are all equal to the prior mean of D (i.e., equal to the experimenter's guess, $D_0 = [A_0, B_0, C_0]$) and whose variances reflect the growing amount of information that is expected to become available in each period of the experiment. The rationale behind this approximation is discussed in detail in the above-mentioned paper.

The variance matrices are related to each other by the equation,

$$\Gamma_k^{-1} = \Gamma_{k-1}^{-1} + \Omega^{-1} \otimes E\{w_k w'_k\}, \tag{3.3}$$

where Γ_k is the variance-covariance matrix of the elements of D_k (arranged by rows). If it were not for the expected value operator on the right-hand side, the above equation would describe the change in the variance of ordinary least squares estimates of the unknown parameters as additional observations become available. As it stands, however, equation (3.3) may be interpreted as measuring the anticipated growth in the stock of information over the course of the experiment.

Before going on to describe the solution to the experiment design problem under the approach just discussed, it is useful to clarify the interpretation of the two components of the objective function, J_1 (the cost) and J_2 (the information gain). In the first period of the experiment, a set of design vectors will be determined for the first and all subsequent periods. Only the first of these design vectors (u_1) will actually be implemented, of course, but it is necessary to make tentative plans as to what will be done later in the experiment in order to make an optimal choice of what is to be done in the first period. However, in view of the fact that the future cannot be predicted precisely, the tentative design vectors will not be calculated as fixed numbers, but as functions of variables which will be observed later in the experiment. In other words, the set of design vectors, $u_1, \ldots, u_N$, will not be calculated as a set of explicit values, but rather as a set of strategy rules or contingency plans. This means, of course, that the cost of carrying out the tentative experiment design cannot be calculated exactly at the beginning of the experiment, nor is it possible to calculate exactly what gain in information will result. The procedure to be used here is to use the expected cost of the tentative plan in place of the actual (but unpredictable) cost, and to use the final information matrix Γ_N^{-1} (as defined by (3.3)) as the argument in the information gain measure, J_2. These two conventions have been incorporated in (3.1).

The solution to the experimental design problem under the assumptions discussed above may be obtained in a manner similar to that used in [4]. The objective function in that paper corresponds to $E\{J_1\}$ in (3.1); the objective function shown in (3.1) simply has that term multiplied by the scalar λ_1 and an additional term involving the terminal stock of information, Γ_N, and the scalar λ_2. Neither of these changes affects the derivation of the solution in any substantial way.

The optimal set of vectors, or strategies, $u_1, u_2, \ldots, u_N$, is given by the following system of solution equations:

$$u_k = -H_k^{-1}(F_k x_{k-1} + f_k) \qquad k = 1, \ldots, N \tag{3.4}$$

where, for $k = 1, \ldots, N$,

$$H_k = B'K_k B + K_k \circledast \Gamma_{k-1}^{BB} - \Omega^{-1} \circledast M_k^{BB} + \lambda_1 R_k, \tag{3.5}$$

$$F_k = B'K_k A + K_k \circledast \Gamma_{k-1}^{BA} - \Omega^{-1} \circledast M_k^{BA} \tag{3.6}$$

and

$$f_k = (B'K_k C + K_k \circledast \Gamma_{k-1}^{BC} - \Omega^{-1} \circledast M_k^{BC})z_k + B'g_k + \lambda_1 t_k. \tag{3.7}$$

Matrices A, B and C are equal to the experimenter's initial guesses A_0, B_0 and C_0 (the subscript 0 is omitted for clarity), and the superscripts on

Γ refer to particular elements of the full covariance matrix. The matrix Γ^{BA}_{k-1}, for example, contains those elements of Γ_{k-1} which are covariances between elements of **B** and elements of **A**. Matrices K_k, M_k and the vectors g_k are defined recursively, for $k = N, N-1, \ldots, 1$, by

$$K_{k-1} = \lambda_1 Q_{k-1} + A'K_k A + K_k \circledast \Gamma^{AA}_{k-1} - \Omega^{-1} \circledast M^{AA}_k - F_k' H_k^{-1} F_k, \tag{3.8}$$

$$g_{k-1} = \lambda_1 s_{k-1} + A' g_k + (A'K_k C + K_k \circledast \Gamma^{AC}_{k-1} - \Omega^{-1} \circledast M^{AC}_k) z_k - F_k H_k^{-1} f_k, \tag{3.9}$$

and

$$M_{k-1} = M_k + \Gamma_{k-1}(K_k \circledast E\{w_k w_k'\})\Gamma_{k-1}, \tag{3.10}$$

$$\begin{aligned} K_N &= \lambda_1 Q_N \\ g_N &= \lambda_1 s_N \\ M_N &= -\lambda_2(\partial L\{\Gamma_N^{-1}\}/\partial \Gamma_N^{-1}). \end{aligned} \tag{3.11}$$

The symbols $\otimes$ and $\circledast$ stand for the Kronecker product and star product[1] respectively.

If the design problem were specified in terms of minimizing a weighted sum of cost and information gain, then explicit values would be assigned to the two weighting parameters, λ_1 and λ_2, and the system of equations (3.3) to (3.11) would be solved iteratively to give the design vector u_1 which is to be implemented in the first period, and the tentative strategy rules, $u_2, \ldots, u_N$, for the remaining periods.

If the design problem were originally stated in terms of maximizing information gain for a given cost, then the above system of equations would be augmented by the additional constraint

$$E\{J_1\} \leq \text{maximum allowable cost}, \tag{3.12}$$

the parameter λ_2 would be set equal to 1, and λ_1, which now plays the role of Lagrangean multiplier for constraint (3.12), would be determined by the system of equations. It will generally be the case that additional expenditures on the experiment will yield additional information, so that (3.12) will almost always be satisfied by equality.

If the constraint is on the information gain, then the extra equation

[1] The star product of an m by n matrix A and a mp by nq matrix B is a p by q matrix C, $C = A \circledast B$, defined by $C = \sum_{ij} a_{ij} B_{ij}$, where a_{ij} is the ijth element of A and B_{ij} is the ijth submatrix of B. The B_{ij} are all of dimension p by q. A more complete description of the star product may be found in MacRae [3], along with techniques for calculating the matrix derivative found in (3.11) above.

becomes

(3.12′) $J_2 \geq$ minimum required information.

parameter λ_1 is set to 1, and λ_2 becomes the Lagrangean multiplier to be determined by the system of equations. In general, (3.12′) will also be satisfied by equality except in the unrealistic situation where the initial information is more than is wanted at the end of the experiment.

4. Analysis

The set of equations which defines the tentative design vectors involve three Lagrangean multipliers, λ_1, λ_2, and the matrices M_k. These may all be interpreted as the marginal gain in the objective function of relaxing the associated constraint. Thus for example, λ_1, which is associated with the cost constraint (3.12), measures the marginal value in units of information of having an additional dollar allocated to the experiment. Similarly, if the design problem is specified with an information constraint such as (3.12′), then λ_2 shows how many dollars could be saved by a marginal reduction in the amount of information required at the end of the experiment.

The interpretation of matrices M_k is somewhat less obvious. They were introduced into the problem as Lagrangean multipliers for the variance-update constraints (3.3), and as such may be interpreted as the imputed price of the stock of information, Γ_k^{-1}, in each period k. As equation (3.11) states, the value of having more information in the last period of the experiment is exactly equal to the marginal contribution of Γ_N^{-1} to the objective function. Whatever is learned during the last period has no additional value to the experimenter since it cannot be used to improve the experiment design in the earlier periods. As can be seen from equation (3.10), the matrices M_k grow in value the nearer k is to the first period. This simply indicates that additional information is of more value early in the course of the experiment where it contributes not only to the terminal stock of information, but also permits a more finely tuned experimental design.

Matrices M_k appear in the strategy rules only in conjunction with Ω^{-1}, the inverse of the variance of the basic model. The effect of a larger M_k is generally to make the design vector more radical so as to increase the information level more quickly. This effect is modified, however, if the system of equations is noisy (i.e., if Ω^{-1} is small), for then it is not clear that actively manipulating the design vector would result in an information gain which is worth the cost.

Paradoxically, in a dynamic model, less initial information (i.e., smaller Γ_0^{-1}) does not necessarily make it optimal to do more active experimentation to learn more in the earlier periods. The reason is that the potential cost of carrying out the experiment is increased if less is known about how the model behaves. Thus it may actually be better to adopt a rather conservative experimental design in the earlier periods and postpone most of the information gathering activity until such time as a more reliable and cost-efficient experiment may be designed, using some of the results of earlier periods.

Federal Energy Agency

Submitted: March 1977
Revised: June 1977

REFERENCES

[1] Conlisk, John. "Experimental Design in Econometrics: The Simultaneous Equations Problem."

[2] Federov, V. V., *Theory of Optimal Experiments* Academic Press, New York, (1972).

[3] MacRae, Elizabeth Chase, "Matrix Derivatives with an Application to an Adaptive Linear Decision Problem." *The Annals of Statistics*, (1974).

[4] MacRae, Elizabeth Chase, "An Adaptive Learning Rule for Multiperiod Decision Problems." *Econometrica* (Sept. Nov., 1975).

[5] Watts, Harold and John Conlisk, "A Model of Optimizing Experimental Designs for Estimating Response Surfaces." *American Statistical Proceedings, Social Statistics Section, 1969.*