Annals of Economic and Social Measurement, 1/4, 1972

STOCHASTIC OPTIMIZATION IN RECURSIVE EQUATION SYSTEMS WITH RANDOM PARAMETERS WITH AN APPLICATION TO CONTROL OF THE MONEY SUPPLY*

BY H. WOODS BOWMAN AND ANNE MARIE LAPORTE†

In this paper the coefficients of a stochastic linear recursive model are presumed random. Using an expected loss (Bayesian) approach, the exact one-period solution for a single control variable, quadratic criterion function, and any number of criteria variables is derived. An application to the recursive St. Louis model indicates that a Bayesian approach leads to smaller losses and, generally, a more conservative policy than a certainty equivalence approach where the coefficients of the model are assumed known.

A promising approach to decision making with econometric models has been developed by Holt and Theil who postulate a quadratic utility (or loss) function in the criteria variables.[1] Provided the model is linear with *known* coefficients, the optimal policy is found to be one for which the criterion function is an extremum.[2] Since, in practice, the coefficients of a model are not known the technique utilizes the mean values of the coefficient estimators and for this reason it is known as the *certainty equivalence* approach.

Unfortunately this approach does not take account of the variance of the population parameters. An alternative, not yet popular, utilizes an explicit criterion function, but *presumes* that the population parameters are *unknown*. In this case, the criterion function is a function of random variables and, hence, an extremum does not exist. If the function is quadratic, however, its mathematical expectation is sometimes tractable and an extremum can be located for the expectation of the function.

This technique, known as the Bayesian or expected loss technique, has been explored by Fisher and Zellner, among others, and recently an application was made to a money multiplier model at the Federal Reserve Bank of St. Louis.[3] Fisher and Zellner derive exact results for single equation models and indicate

* Most of the work on this paper was done while Dr. Bowman was affiliated with the Federal Reserve Bank of Chicago. The authors would like to express their appreciation to Karl A. Scheld of the Federal Reserve Bank of Chicago and Arnold Zellner of the University of Chicago for their guidance and encouragement throughout the project. Calculations were performed using computer facilities at the Federal Reserve Bank of Chicago.

† Ms. Laporte is primarily responsible for the final section which treats the application. Dr. Bowman is primarily responsible for the remainder of the paper.

[1] C. C. Holt, "Linear Decision Rules for Economic Stabilization and Growth," *Quarterly Journal of Economics*, *67* (1962) and Henri Theil, *Optimal Decision Rules for Government and Industry* (Amsterdam: North-Holland, 1964), Ch. 6.

[2] Among their other virtues, quadratic functions possess unique extrema, so that a single "best" policy will always exist.

[3] Walter Fisher, "Estimation in the Linear Decision Model," *International Economic Review*, *3* (1962); Arnold Zellner, *An Introduction to Bayesian Inference in Econometrics* (New York: John Wiley, 1971), Ch. 11; Albert E. Burger, Lionel Kalish III, and Christopher T. Babb, "Money Stock Control and Its Implications for Monetary Policy," *Federal Reserve Bank of St. Louis Review*, *53* (October, 1971), pp. 6–22.

that the difference in control settings and expected loss arising from using the certainty equivalence solution instead of the Bayesian solution is rather small.

In multi-equation models, the difference can be magnified by the particular shape of the quadratic criterion function employed and by the relationships among the endogenous variables. Research in this area had been held up because exact solutions are difficult to find for multi-equation systems. One type of multi-equation system for which exact solutions can be easily found is the recursive system. This paper develops the exact form for the one period solution in the case of (1) a quadratic criterion function, (2) a single policy instrument, (3) any number of equations which are linear in the parameters and recursive in structure, and (4) any number of criteria variables.

An application of the solution is then made to a model of the U.S. economy developed at the Federal Reserve Bank of St. Louis.[4] The results indicate that the difference in expected loss can be extremely large. The magnitude of the difference is significantly affected by information contained in the data (for a given model) over which the policymaker has no control. However, if more than one criterion variable is used, the policymaker employing a certainty equivalence solution can minimize the difference by emphasizing in the criterion function the variable about which he or she is more certain.

Control in General Recursive Systems

The problem is to

(1) minimize: $EL = E(\mathbf{y} - \mathbf{a})'Q(\mathbf{y} - \mathbf{a})$, expected loss

(2) subject to: $Y = Y\Gamma + XB + U$

where these quantities are defined as follows:

- $\mathbf{y}$ is an m-element column vector of "future" observations (that is, the $T + 1$ observations) on the endogenous variables of the system, having unconditional mean value $\bar{\mathbf{y}}$;
- $\mathbf{a}$ is an m-element column vector of targets corresponding to these variables;
- Q is an $m \times m$ positive definite symmetric matrix of constants—the parameters of the loss function;[5]
- Y is a $T \times m$ matrix of observations on the endogenous variables in the system;
- Γ is an $m \times m$ upper triangular matrix of random coefficient parameters with zeros along the diagonal;
- X is a $T \times n$ matrix of observations on the predetermined variables of the system;
- B is an $n \times m$ matrix of random coefficient parameters;
- U is a $T \times m$ matrix of unobserved random error terms.

Furthermore, it is assumed that the columns of U are normally and independently distributed, such that letting

$$\mathbf{v}' \equiv (\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_m'),$$

[4] Leonall C. Andersen and Keith M. Carlson, "A Monetarist Model for Economic Stabilization," *Federal Reserve Bank of St. Louis Review*, *52* (April, 1970), pp. 7–25. Hereafter referred to as the St. Louis model.

[5] Q may be of rank $k < m$, where k is the number of criteria variables in the control problem.

then

$$Ev' = \mathbf{0}$$

$$Evv' = D(\sigma_i^2) \otimes I_T \qquad i = 1, 2, \ldots, m$$

where $\mathbf{u}_i$ represents the ith column of U and $D(\sigma_i^2)$ is a diagonal matrix with diagonal elements given by σ_i^2. The recursive nature of the system is evident from two conditions: (1) the Γ matrix is upper triangular and (2) the error terms of different equations are independently distributed.

The same model is assumed to generate "future" observations, that is

$$\text{(3)} \qquad \mathbf{y}' = \mathbf{y}'\Gamma + \mathbf{x}'B + \mathbf{u}'$$

where $\mathbf{y}'$ is an m-element row vector and $\mathbf{x}'$ is an n-element row vector representing observations on the endogenous and predetermined variables respectively in the $T + 1$ period. The m elements of the row vector $\mathbf{u}'$ are unobserved error terms in the $T + 1$ period, independently and normally distributed such that

$$E\mathbf{u} = \mathbf{0},$$

$$E\mathbf{u}\mathbf{u}' = D(\sigma_i^2) \qquad i = 1, 2, \ldots, m,$$

$$E\mathbf{u}\mathbf{u}_k' = [0] \qquad k = 1, 2, \ldots, T,$$

where $\mathbf{u}_k$ is the kth row of U.

Having defined these new quantities and placed the appropriate restrictions on the error terms, we are now in a position to consider the loss function[6]

$$\text{(4)} \qquad EL = E(\mathbf{y} - \mathbf{a})'Q(\mathbf{y} - \mathbf{a}) = E(\mathbf{y} - \bar{\mathbf{y}})'Q(\mathbf{y} - \bar{\mathbf{y}}) + (\bar{\mathbf{y}} - \mathbf{a})'Q(\bar{\mathbf{y}} - \mathbf{a}).$$

Note that the second term is the loss function associated with the certainty equivalence solution (CEL) to the control problem. It is simply a weighted sum of the deviations of the mean values of the criteria variables about their respective targets. The Bayesian approach, which is the sum of the two terms, takes into account additional loss arising from the possible failure to predict the mean values of the criteria variables accurately. The first term clearly expands into a linear combination of variances and covariances of the multivariate predictive probability density function (pdf) for $\mathbf{y}$, which are easily derivable because the form of the distribution is known.

Consider a single element of the vector $\mathbf{y}$, say y_i. The posterior pdf for this element conditional upon the other elements of $\mathbf{y}$ can be set up using a non-informative prior on the appropriate elements of B, Γ and $D(\sigma_i)$, and then integrating over these unknown quantities, thusly

$$\text{(5)} \qquad p(y_i|\mathbf{y}_{i-1}, \#) = \int p(y_i, \boldsymbol{\gamma}_i, \boldsymbol{\beta}_i, \sigma_i|\mathbf{y}_{i-1}, \#)\, d\boldsymbol{\beta}_i\, d\boldsymbol{\gamma}_i\, d\sigma_i$$

$$= \int p(y_i|\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i, \sigma_i, \mathbf{y}_{i-1}, \#)p(\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i, \sigma_i)\, d\boldsymbol{\beta}_i\, d\boldsymbol{\gamma}_i\, d\sigma_i$$

[6] Alternatively the problem could be expressed in a utility maximizing format. If so, completing the square on y yields a declining monotonic function of (4). Maximizing utility and minimizing loss are therefore equivalent statements of the problem.

where $\mathbf{y}_{i-1}$ represents the vector of endogenous variables appearing in the ith equation of the system, $\boldsymbol{\beta}_i$ is the ith column of B (the coefficients of the predetermined variables in the ith equation), γ_i represents the ith column of Γ and # represents observed data (X, Y, and $\mathbf{x}$). The first pdf in the second line above is univariate normal because of the normality assumption concerning the error terms and the second pdf is proportional to $1/\sigma_i$ because we are using non-informative prior pdf's. When the indicated integration is performed the resulting conditional predictive pdf for y_i is Student-t with mean and variance given by[7]

(6) $$\bar{y}_i^c = \mathbf{y}'\bar{\boldsymbol{\gamma}}_i + \mathbf{x}'\bar{\boldsymbol{\beta}}_i,$$

and

(7) $$v_i^c = \bar{s}_i^2(1 + \mathbf{z}_i' M_i \mathbf{z}_i)$$

respectively, where the bars in (6) represent the means of the corresponding quantities and

$$\mathbf{z}_i \equiv (\mathbf{y}_{i-1}, \mathbf{x}_i),$$

$$\bar{s}_i^2 \equiv [\nu_i/(\nu_i - 2)]s_i^2,$$

$$M_i \equiv \begin{bmatrix} Y_{i-1}'Y_{i-1} & Y_{i-1}'X_i \\ X_i'Y_{i-1} & X_i'X_i \end{bmatrix},$$

ν_i being the degrees of freedom in the ith equation, s_i^2 being the estimated OLS residual variance for the ith equation, and where the subscripts on the X and Y quantities indicate that only the variables in the ith equation are to be considered.

We may now use these quantities to find the mean and variance of the corresponding marginal distribution of y_i.

First, the mean

(8) $$\begin{aligned} \bar{y}_i &\equiv \int y_i p(y_i \mid \#)\, dy_i \\ &= \int y_i p(y_i \mid \mathbf{y}_{i-1}, \#) p(\mathbf{y}_{i-1}, \#)\, d\mathbf{y}_i \\ &= \int \bar{y}_i^c p(\mathbf{y}_{i-1}, \#)\, d\mathbf{y}_{i-1} \\ &= \bar{\mathbf{y}}'\bar{\boldsymbol{\gamma}}_i + \mathbf{x}'\bar{\boldsymbol{\beta}}_i. \end{aligned}$$

By induction, therefore,

(9) $$\bar{\mathbf{y}} = \bar{\mathbf{y}}'\bar{\Gamma} + \mathbf{x}'\bar{B},$$

and since it is easily seen that $(I_m - \bar{\Gamma})$ is non-singular,

(10) $$\bar{\mathbf{y}} = \mathbf{x}'\bar{B}(I_m - \bar{\Gamma})^{-1} \equiv \mathbf{x}'\bar{\Pi}.$$

[7] The integration is performed in Zellner, *op. cit.*, pp. 72–74.

The mean of the marginal predictive pdf for an arbitrary element of y is therefore given by

$$\bar{y}_i = \mathbf{x}'\bar{\pi}_i \tag{10a}$$

where $\bar{\pi}_i$ is the ith column of $\bar{\Pi}$.

The variance of the marginal pdf is more difficult to find but the same technique of integrating unwanted variables out of the joint distribution can still be used:

$$(11)\quad v_{ii} \equiv \int (y_i - \bar{y}_i)^2 p(y_i| \#)\, dy_i$$

$$= \int [(y_i - \bar{y}_i^c)^2 + 2y_i(\bar{y}_i^c - \bar{y}_i) + \bar{y}_i^2 - \bar{y}_i^{c2}] p(y_i|\mathbf{y}_{i-1}, \#) p(\mathbf{y}_{i-1}| \#)\, d\mathbf{y}_i.$$

All terms in the brackets will cancel except for $(y_i - \bar{y}_i)^2$. The others were added in order to make use of the expression for the variance of the conditional predictive pdf, thusly

$$v_{ii} = \int [v_i^c + (\bar{y}_i^c - \bar{y}_i)^2] p(\mathbf{y}_{i-1}| \#)\, d\mathbf{y}_{i-1},$$

$$= \int \{\bar{s}_i^2[1 + (\mathbf{y}_{i-1}, \mathbf{x}_i)' M_i (\mathbf{y}_{i-1}, \mathbf{x}_i)] + [(\mathbf{y}_{i-1} - \bar{\mathbf{y}}_{i-1})'\bar{\boldsymbol{\gamma}}_i]^2\} p(\mathbf{y}_{i-1}| \#)\, d\mathbf{y}_{i-1}.$$

The first term in brackets is a quadratic form in the elements of $\mathbf{y}_{i-1}$. The indicated integration transforms these random variables into their respective means. Using the well-known relation $Ey_k^2 = v_{kk} + \bar{y}_k^2$, this term becomes a quadratic form in the elements of $\bar{\mathbf{y}}_{i-1}$ plus a weighted sum of variances, that is

$$1 + (\bar{\mathbf{y}}_{i-1}, \mathbf{x}_i)' M_i (\bar{\mathbf{y}}_{i-1}, \mathbf{x}_i) + \sum_{j=1}^{i-1} M_i^{jj} v_{jj}$$

or

$$1 + \bar{\mathbf{z}}_i' M_i \bar{\mathbf{z}}_i + \sum_{j=1}^{i-1} M_i^{jj} v_{jj}$$

where M_i^{jj} is the jth diagonal element of M_i.

The second term in brackets is also a quadratic function and performing the indicated integration transforms it into a weighted sum of variances and covariances thusly

$$\sum_{j=1}^{i-1} \bar{\gamma}_{ij}^2 v_{jj} + 2 \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \bar{\gamma}_{ij}\bar{\gamma}_{ki} v_{jk}.$$

The covariance terms can be evaluated as follows for $j > k$

$$v_{jk} = \int (y_j - \bar{y}_j)(y_k - \bar{y}_k)p(\mathbf{y}_j|\#)\,d\mathbf{y}_j, \tag{12}$$

$$= \int (y_j - \bar{y}_j)(y_k - \bar{y}_k)p(y_j|\mathbf{y}_{j-1}, \#)p(\mathbf{y}_{j-1}|\#)\,d\mathbf{y}_j,$$

$$= \int (y_k - \bar{y}_k)(\mathbf{y}'\bar{\boldsymbol{\gamma}}_j + \mathbf{x}'\bar{\boldsymbol{\beta}}_j - \bar{y}_j)p(\mathbf{y}_{j-1}|\#)\,d\mathbf{y}_{j-1},$$

$$= \int [(y_k - \bar{y}_k)(\mathbf{y}'\bar{\boldsymbol{\gamma}}_j - \bar{\mathbf{y}}'\bar{\boldsymbol{\gamma}}_j) + (y_k - \bar{y}_k)(\bar{\mathbf{y}}'\bar{\boldsymbol{\gamma}}_j + \mathbf{x}'\bar{\boldsymbol{\beta}}_j - \bar{y}_j)]$$

$$\times\, p(\mathbf{y}_{j-1}|\#)\,d\mathbf{y}_{j-1}.$$

The first term in the brackets becomes a weighted sum of other covariance terms while the second vanishes when integration is carried out over y_k. Therefore we can simplify the above to an element of V, an upper triangular $m \times m$ matrix of covariances with variances along the diagonal. If $\mathbf{v}_i'$ is a row vector from this matrix and $\mathbf{i}'$ is the ith row vector of I_m, then

$$\mathbf{v}_i' = \mathbf{v}_i'\bar{\Gamma} + v_{ii}\mathbf{i}'. \tag{13}$$

And further letting D be an $m \times m$ diagonal matrix of variances, this generalizes to

$$V = V\bar{\Gamma} + D. \tag{14}$$

Since $(I_m - \bar{\Gamma})$ is non-singular,

$$V = D(I_m - \bar{\Gamma})^{-1}. \tag{15}$$

Thus, a covariance relation between any two endogenous variable can be expressed as a multiple of the variance associated with the variable having the lower index number. For example, v_{ij} is some multiple of v_{ii} if $i < j$, and v_{jj} if $j < i$.

Returning to the second term in brackets in the expression for the variance, (11), we can now see that it will be a linear combination of variances of endogenous variables occurring in preceding equations. In particular, if we consider the complete variance-covariance matrix $(V + V' - D)$, then the second term in brackets becomes, after integration,

$$\bar{\boldsymbol{\gamma}}_i'[D(I_m - \bar{\Gamma})^{-1} + (I_m - \bar{\Gamma}')^{-1}D - D]\bar{\boldsymbol{\gamma}}_i.$$

This cumbersome expression is of the form

$$\sum_{j=1}^{i-1} g_{ij}v_{jj}.$$

Finally, the complete expression for the variance of the endogenous variable of the ith equation is

$$v_{ii} = \bar{s}_i^2(1 + \bar{\mathbf{z}}_i'M_i\bar{\mathbf{z}}_i) + \sum_{j=1}^{i-1} (\bar{s}_i^2 M_i^{jj} + g_{ij})v_{jj}. \tag{16}$$

Before proceeding to evaluate the expected loss function, it is instructive to pause and study this expression. The first part is the variance of the ith endogenous variable conditional upon the other variables in the system being equal to their mean values. This will tend to understate the true variance, however, insofar as the mean of the *conditional* distribution of the ith endogenous variable is not equal to the mean of its marginal distribution. Any difference between these quantities will serve to increase the variance. Moreover, the effect upon the variance will be magnified by the variance of the other endogenous variables appearing in the ith equation, hence the terms in g_{ij}. The remaining terms $\bar{s}_i^2 M_i^{jj}$ will be recognized as the variance of the parameter estimates of the jth endogenous variable appearing in the ith equation. This indicates that any uncertainty introduced by random variables appearing in the ith equation is also magnified by the uncertainty associated with the corresponding coefficients.

To adapt this expression for a matrix format we let $\mathbf{d}$ be an m-element column vector consisting of the diagonal elements of D, the ith element of which is v_{ii}. Then,

$$\mathbf{d}' = \bar{\mathbf{v}}' + \mathbf{d}'\bar{G} \tag{17}$$

where we have defined the ith element of $\bar{\mathbf{v}}$ to be $\bar{s}_i^2[1 + \bar{\mathbf{z}}_i' M_i \bar{\mathbf{z}}_i]$ and the i, jth element of the $m \times m$ matrix $\bar{G}$ to be $(M_i^{jj} + g_{ij})$. Therefore,

$$\mathbf{d}' = \bar{\mathbf{v}}'(I_m - \bar{G})^{-1}. \tag{18}$$

Finally, we are ready to evaluate the loss function and to find its extremum. Recall that the loss function was given by

$$EL = E(\mathbf{y} - \bar{\mathbf{y}})'Q(\mathbf{y} - \bar{\mathbf{y}}) + (\bar{\mathbf{y}} - \mathbf{a})'Q(\bar{\mathbf{y}} - \mathbf{a}). \tag{4}$$

The expectation operator applied to the first term merely yields a weighted sum of variances and covariances. That is,

$$\text{trace}\,[Q(V + V' - D)]. \tag{19}$$

But the covariances are themselves functions of the variances,

$$\begin{aligned} &\text{trace}\,\{Q[D(I_m - \Gamma)^{-1} + (I_m - \Gamma')^{-1}D - D]\} \\ &= \text{trace}\,\{DQ[(I_m - \Gamma)^{-1} + (I_m - \Gamma')^{-1} - I_m]\} \\ &\equiv \mathbf{d}'\mathbf{q}. \end{aligned}$$

Upon substituting for $\mathbf{d}$ we have

$$\mathbf{d}'\mathbf{q} = \bar{\mathbf{v}}'(I_m - \bar{G})^{-1}\mathbf{q} \equiv \bar{\mathbf{v}}'\mathbf{w}, \qquad \text{where } \mathbf{w} \equiv (I_m - \bar{G})^{-1}\mathbf{q}. \tag{20}$$

So we see that the first term of the loss function is a weighted sum of variances of conditional pdf's given that other variables are equal to their mean values. The weights are given by the corresponding elements of the vector $(I_m - \bar{G})^{-1}\mathbf{q}$, that is w_i.

From this point it is a straightforward, but tedious, task to differentiate the loss function with respect to the policy instrument (control variable) and then solve for the loss minimizing setting for that variable, so we simply state the result without proof. However, some new quantities need to be defined:

x_1 is the control variable;
n_i is the number of predetermined variables in the ith equation;
$\mathbf{p}_i'$ is an $n_i + i - 1$ element vector equal to $\iota'(\bar{\Pi}_{i-1}, P_i)$ where
ι is an n-element vector consisting of zeros and ones, such that $\iota = d\mathbf{x}/dx_1$ and
$\bar{\Pi}_{i-1}$ is a matrix consisting of the first $i - 1$ columns of $\bar{\Pi}$ and
P_i is an $n \times n_i$ matrix consisting of zeros and ones such that $\iota' P_i = d\mathbf{x}_i/dx$, where $\mathbf{x}_i$ is a vector of predetermined variables entering in the ith equation;
$\mathbf{x}_0$ is an n-element vector of "future" values of all predetermined variables in the system with the exception of x_1, the value of which is set equal to zero;
$\mathbf{x}_{0i}$ is an n_i-element vector of all predetermined variables in the ith equation with the exception of x_1, the value of which is set equal to zero;
P_i^* is an $n \times n_i$ matrix consisting of zeros and ones such that $\mathbf{x}_{0i}' = \mathbf{x}_0' P_i^*$;
$\mathbf{p}_i^{*\prime}$ is an $n_i + i - 1$ element vector equal to $\mathbf{x}_0'(\bar{\Pi}_{i-1}, P_i^*)$.

Using these quantities, the optimal setting for the control variable is given by

$$(21) \qquad x_1^* = \frac{-\sum_{i=1}^{m} w_i \bar{s}_i^2 \mathbf{p}_i' M_i \mathbf{p}_i^* - \iota'\bar{\Pi}Q(\bar{\Pi}'\mathbf{x}_0 - \mathbf{a})}{\sum_{i=1}^{m} w_i \bar{s}_i^2 \mathbf{p}_i' M_i \mathbf{p}_i + \iota'\bar{\Pi}Q\bar{\Pi}'\iota}.$$

The other quantities, such as w_i, $\bar{s}_i^2$, M_i, Q and $\mathbf{a}$, are defined elsewhere above. We are assured that x_1^* minimizes the expected loss function because the matrix Q was assumed to be positive definite symmetric.

Had we elected to merely minimize certainty equivalence loss the terms involving sums of squares and cross-products (M_i) would not appear in the solution:

$$(22) \qquad x_1^{ce} = -\frac{\iota'\bar{\Pi}Q(\bar{\Pi}'\mathbf{x}_0 - \mathbf{a})}{\iota'\bar{\Pi}Q\bar{\Pi}'\iota}.$$

Clearly, x_1^{ce} is a linear function of x_1^*. A sufficient condition for x_1^{ce} to exceed x_1^* is that the indicated summation in the numerator of (21) be non-negative, which is always the case when there are no other predetermined variables in the system except the control variable.

The Loss Function

In the experimental example which follows, we employ two criteria variables —price changes and the unemployment rate—so it is useful at this point to consider some of the properties of a two variable loss function. Specializing (4) we have:

(23) $$CEL = q_1(\bar{y} - a_y)^2 + 2q_2(\bar{y} - a_y)(\bar{z} - a_z) + q_3(\bar{z} - a_z)^2.$$

(24) $$EL = q_1 V(y) + 2q_2 \operatorname{Cov}(y, z) + q_3 V(z) + CEL.$$

A regular minimum exists for these functions if and only if $q_1 > 0$ and $q_1 q_3 - q_2^2 > 0$.

These second order conditions imply elliptical iso-loss contours for both (23) and (24). One senses immediately, however, that only a portion of the ellipse would be economically relevant. To make the discussion concrete, consider Figure 1 and suppose that z measures the rate of change in prices and y measures the unemployment rate.

If one were to begin at point B and move clockwise, one would be trading more inflation for *more* unemployment. This bizarre implication is, of course, due to the symmetric nature of the quadratic function: undershooting a target is as undesirable as overshooting by the same amount. Those who find this symmetry unappealing may be tempted to abandon the quadratic function and search for a nonsymmetric one. But one may retain the quadratic function if one is satisfied with being confined to a limited region of the ellipse between A and A'. Within this region one always trades more inflation for less unemployment, or vice versa. Moreover, one experiences an *increasing* marginal rate of substitution—which is as it should be since deviations from either target are *undesirable*.

A set of necessary and sufficient conditions for operating in the relevant range of the ellipse are (1) the implied Phillips' curve of the system must be negatively sloped, and (2) the targets must be chosen to be below and to the left of the implied Phillips' curve. The reason for these conditions is clear when one considers that the system of regression equations acts as a constraint to minimizing expected loss.

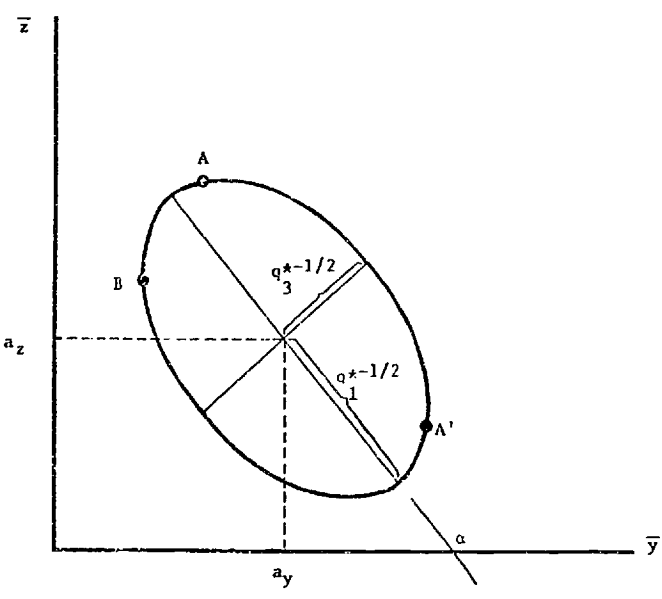


Figure 1 Elliptical iso-loss function

The optimal solution x^* is that one for which the implied Phillips' curve is tangent to the lowest possible iso-loss contour projected onto its plane. When operating with an optimal policy the policymaker's marginal rate of substitution will be equal to the slope of the implied Phillips' curve of the system.

Bearing these things in mind, we seek to reparameterize the loss function in terms of easily interpretable quantities with geometric significance. Consider first the certainty equivalence loss (23). The center of each elliptical iso-loss curve will be at the point (a_y, a_z). From analytic geometry we recognize that the loss function parameters q_1 and q_3 represent simple transformations of the projections of the elliptical axes upon the axes of the reference system, and if we rotate the ellipse about the point (a_y, a_z) until its axes are parallel to those of the reference system, q_2 vanishes and q_1 becomes q_1^*, q_3 becomes q_3^*, $\bar{y}$ becomes $\bar{y}^*$ and $\bar{z}$ becomes $\bar{z}^*$.

A rotation of the elliptical axes through an angle α can be expressed as

$$\bar{y}^* = \bar{y}\cos\alpha - \bar{z}\sin\alpha, \tag{25}$$

$$\bar{z}^* = \bar{y}\sin\alpha + \bar{z}\cos\alpha. \tag{26}$$

Substituting these transformations into the certainty equivalence loss function and using the following trigonometric identities:

$$\cos^2\alpha = \frac{1}{\tan^2\alpha + 1}, \tag{27}$$

$$\sin^2\alpha = \frac{\tan^2\alpha}{\tan^2\alpha + 1}, \tag{28}$$

$$\sin\alpha\cos\alpha = \frac{\tan\alpha}{\tan^2\alpha + 1}, \tag{29}$$

we have

$$CEL = \left[\frac{r + \tan^2\alpha}{\tan^2\alpha + 1}\right](\bar{y} - a_y)^2 + 2\left[\frac{(1 - r)\tan\alpha}{\tan^2\alpha + 1}\right] (\bar{y} - a_y)(\bar{z} - a_z) + \left[\frac{r\tan^2\alpha + 1}{\tan^2\alpha + 1}\right](\bar{z} - a_z)^2 \tag{30}$$

where $r = q_1^*/q_3^*$. This quantity is also by definition equal to $1 - e^2$ where e is the eccentricity of the ellipse, and measures the departure of the ellipse from circularity. For a circle $e = 0$ and as the ellipse becomes "tighter" $e \to 1$. Notice, however, that $e = 1$ is not possible if the loss function is to remain an ellipse.

The coefficient of $(\bar{y} - a_y)^2$ is evidently q_1, the coefficient of $(\bar{y} - a_y)(\bar{z} - a_z)$ is $2q_2$, and the coefficient of $(z - a_z)^2$ is q_3. All q's are expressed in terms of two simple quantities—the slope of the major axis of the ellipse and the ratio of the elliptical axes. Although $\tan \pi/2$ is infinite, the elements of the loss function possess finite limits as $\tan\alpha \to \infty$. Systematic variation of the parameters is facilitated by this particular parameterization because $\tan\alpha$ is a periodic function of α, and the eccentricity is bounded by zero and one.

We are now in a position to consider the effect of taking the mathematical expectation of the loss function instead of simply replacing the random variables with their respective means. To make the discussion concrete, let us consider a simple two equation system:

(31) $y_i = bx_i + u_1 \qquad i = 1, 2, \ldots, T \qquad u_i$ are $NID(0, \sigma_u^2)$

(32) $z_i = cy_i + v_i \qquad i = 1, 2, \ldots, T \qquad v_i$ are $NID(0, \sigma_v^2)$ and $E(u_i v_i) = 0$

For simplicity, consider a circular loss function where $q_2 = 0$. In this case

(33) $$EL = [V(y) + (\bar{y} - a_y)^2] + [V(z) + (\bar{z} - a_z)^2].$$

Note that the difference between this result and the certainty equivalence loss is the addition of a weighted sum of variances which are themselves expressible in terms of the variable means $\bar{y}$ and $\bar{z}$ and other quantities calculated from the data:

(34) $$V(z) = s_z^2 + [V(c) + \bar{c}^2][V(y) + \bar{y}^2] + \bar{z}(\bar{z} - 2\bar{c}\bar{y})$$

(35) $$V(y) = s_y^2 + \bar{y}^2 V(b)/\bar{b}^2.$$

However, we do not make use of the system constraint $\bar{z} = \bar{c}\bar{y}$.

Upon inserting these quantities into (33) and completing the square on $\bar{y}$ and $\bar{z}$ we have:

(36) $$EL = m(\bar{y} - a_y^*) - 2\bar{c}(\bar{y} - a_y^*)(\bar{z} - a_z^*) + 2(\bar{z} - a_z^*) + \text{constants}$$

where $m = (1 + \bar{c}^2)/t_b^2 + \bar{c}^2/t_b^2 t_c^2 + \bar{c}^2/t_c^2 + \bar{c}^2 + 1$ and the new, or virtual, targets are given by $a_y^* = a_y/m$ and $a_z^* = a_z/2$. Thus, the new iso-loss contour will be elliptically shaped, rotated with respect to the coordinate axes and centered closer to the origin. The angle of rotation and eccentricity are given by:

(37) $$\cot 2\alpha = (2 - m)/2\bar{c} \text{ and}$$

(38a) $$e^2 = 1 - (2 - \bar{c}^2)/(m + \bar{c}^2), \quad \text{if } m + \bar{c}^2 > 2 - \bar{c}^2, \quad \text{otherwise}$$

(38b) $$e^2 = 1 - (m + \bar{c}^2)/(2 - \bar{c}^2).$$

If (38a) holds, the major axis of the ellipse corresponds to the $\bar{z}$ dimension in an unrotated system, while if (38b) holds, the major axis of the ellipse corresponds to the $\bar{y}$ dimension in an unrotated system.

To illustrate this example, consider Figure 2. The system constraint $\bar{z} = \bar{c}\bar{y}$ is a straight line passing through the origin which is associated with a zero setting for the control variable. The certainty equivalence loss function is a circle with center at (a_y, a_z). The certainty equivalence solution is the point of tangency of the certainty equivalence loss function and the system constraint, and is denoted by *CEL*. The expected loss function, on the other hand, is an ellipse centered at $(a_y/m, a_z/2)$. Its tangency with the system constraint, denoted by *EL*, represents the solution to the Bayesian control problem.

The *EL* solution is a more conservative solution than the *CEL* solution because it moves the policymaker in the direction of a zero setting for the policy instrument (i.e., towards the origin).[8] The conservativeness of the policy is a function of the certainty with which the coefficients b and c are known. For example, suppose that there were a great deal of uncertainty about either b or c or both, such that $m \gg 2$, then the major axis of the ellipse would be the $\bar{z}$ axis and the angle of

[8] This conclusion may not be true if there are exogenous variables in the system which are not under control.

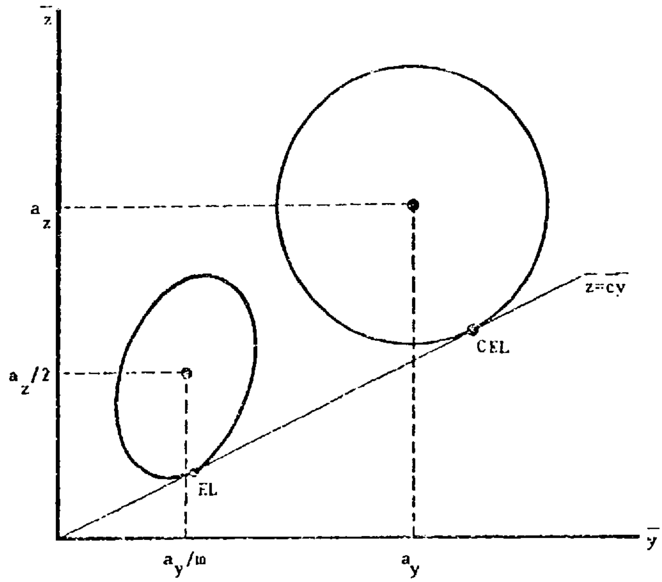


Figure 2 Comparison of certainty equivalence loss and expected loss functions

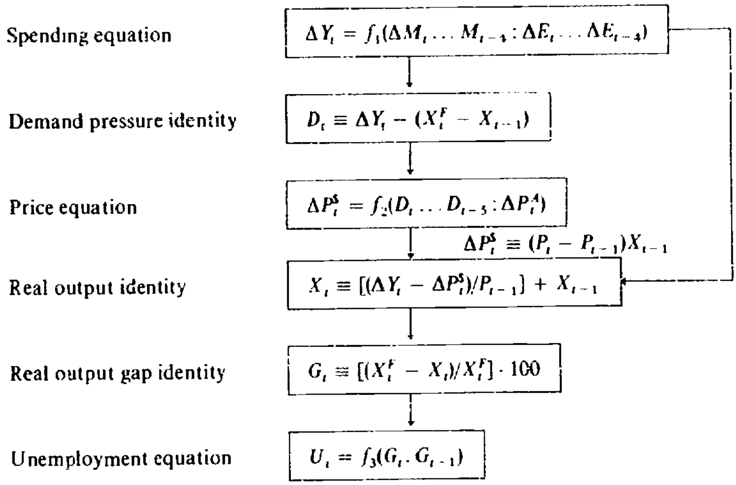
rotation would be close to zero or π. As uncertainty increases, m increases and the ellipse becomes "tighter" as the control solution assigns increasing weight to deviations of $\bar{y}$ about its virtual target.

Application of Control Solution

In the final section of this paper we apply the Bayesian control solution to variations of the St. Louis model of the U.S. economy. The model's structure can be easily seen from the flow chart on the following page. The model is fully recursive—unemployment is influenced by prices and nominal spending through a real output identity; prices, in turn, are influenced by nominal spending; nominal spending depends only upon exogenous variables, including a money variable.[9]

While this form of the model was satisfactory for estimating parameters and performing simulation experiments, it could not be used for computing the optimal value of the instrument variable. A problem arises because the identities operate on the endogenous variables of the system before they enter subsequent equations. The unemployment equation was respecified in terms of a new dependent variable, $U_t X_t^F P_{t-1}$, because the real income identity and the gap identity caused ΔY_t and ΔP_t^S to be multiplied by $(X_t^F P_{t-1})^{-1}$ in the original unemployment equation. This new unemployment variable is an approximate value of foregone production (in period $t - 1$ prices) due to underutilized resources. The measure is

[9] In addition, the necessary assumption about the error terms are made. The original St. Louis model contained additional equations for long- and short-term interest rates, but these are not considered here since the criteria variables of interest are prices and unemployment and the recursive nature of the system does not require their inclusion for either estimation or computation of the optimal solution.



Exogenous variables:

ΔM_t, change in currency plus demand deposits:
ΔE_t, change in high employment government expenditures:
X_t^F, full employment level of real output:

Endogeneous variables:

ΔY_t, change in nominal spending:
D_t, demand pressure:
ΔP_t^S, dollar change in nominal spending due to change in price deflator:
ΔP_t^A, anticipated change in prices:
X_t, real output;
G_t, gap in real output:
U_t, unemployment rate:

Figure 3 Flow Chart of the St. Louis Model

not exact, however, because the elasticity of output with respect to labor input may not be unity.

Since the Bayesian formulation of the control problem differs from the certainty equivalence formulation by taking account of the uncertainty in the model's parameters, information can be gained by comparing models with differing degrees of uncertainty in the parameters. To this end an alternative unemployment equation was selected. It is the same as the revised unemployment equation without the term introduced by the lagged real output gap. As can be seen in Table I, however, the properties of the two resulting models differ in that the model with the lagged gap term (Model-I) has much less precise coefficient estimates for the contemporaneous endogenous variables in the unemployment equation than does the model without the lagged gap term (Model-II).

To study the differences between the solutions, the targets in the criteria function were set at zero price change and zero unemployment while the shape and orientation of the quadratic criteria function were systematically varied. The two solutions were compared (1) in terms of the ratio of the expected losses generated for each particular choice of criteria function parameters and (2) in terms of the ratio of the corresponding settings for the control variable. Although the choice of

TABLE I

COEFFICIENT ESTIMATES

A. *Spending Equation*

$$\Delta Y_t = \underset{(3.56)}{2.74} + \sum_{i=0}^{4} m_i \Delta M_{t-i} + \sum_{i=0}^{4} e_i \Delta E_{t-i}$$

Sample Period: 1953/I to 1969/IV
Constraints: 4th degree polynomial
$(m_{-1} = m_5 = 0; e_{-1} = e_5 = 0)$

$m_0 =$	1.259 (2.86)	$e_0 =$	0.569 (2.66)
$m_1 =$	1.755 (7.21)	$e_1 =$	0.504 (3.91)
$m_2 =$	1.485 (3.90)	$e_2 =$	0.061 (0.32)
$m_3 =$	0.711 (3.02)	$e_3 =$	−0.440 (−3.31)
$m_4 =$	−0.045 (−0.10)	$e_4 =$	−0.611 (−2.79)

$\bar{R}^2 = 0.656$
S.E. = 3.844

B. *Price Equation*

$$\Delta P_t^s = \underset{(6.62)}{2.60} + \underset{(8.78)}{0.941}\, \Delta P_t^A + \sum_{i=0}^{5} d_i D_{t-i}$$

Sample Period: 1955/I to 1969/IV
Constraints: 2nd degree polynomial
$(d_{-1} \neq 0; d_6 = 0)$

$d_0 =$	0.0222 (2.32)	$d_3 =$	0.0146 (3.12)
$d_1 =$	0.0205 (5.95)	$d_4 =$	0.0105 (2.06)
$d_2 =$	0.0179 (6.82)	$d_5 =$	0.0056 (1.59)

$\bar{R}^2 = 0.865$
S.E. = 1.134

C. *Unemployment Equation*

1. Lagged gap term included

Sample Period: 1955/I to 1969/IV

$$\frac{U_t X_t^F P_{t-1}}{100} = \underset{(1.05)}{4.554}\left[\frac{X_t^F P_{t-1}}{100}\right] - \underset{(-0.15)}{0.0066}(\Delta Y_t - \Delta P_t^s + Y_{t-1}) + \underset{(7.34)}{0.315}\, G_{t-1}\left[\frac{X_t^F P_{t-1}}{100}\right]$$

$\bar{R}^2 = 0.915$
S.E. = 1.817

2. No lagged gap term

Sample Period: 1955/I to 1969/IV

$$\frac{U_t X_t^F P_{t-1}}{100} = \underset{(22.1)}{35.3}\left[\frac{X_t^F P_{t-1}}{100}\right] - \underset{(-19.2)}{0.314}(\Delta Y_t - \Delta P_t^s + Y_{t-1})$$

$\bar{R}^2 = 0.838$
S.E. = 2.511

(t-values in parentheses)

targets is arbitrary, zero values represent the most extreme values which satisfy the necessary and sufficient conditions for a relevant economic solution.[10]

Because the variances of the coefficients associated with contemporaneous endogenous variables are much larger in Model-I than in Model-II, we expect that the relative differences in expected loss between the certainty equivalence solution and the Bayesian solution for Model-I would be greater than the relative difference for Model-II. Further, since the control setting in the Bayesian solution tends toward zero as one becomes more uncertain, we expect the relative difference in the control settings between the certainty equivalence solution and the Bayesian solution to be greater for Model-I than for Model-II.

The results of the experiments in terms of expected loss are given in Table II. In general, they are in accord with expectations. In the case of Model-II the relative difference between the two solutions is rarely very great. In the case of Model-I the relative difference between the solutions is more sensitive to the parameters of the loss function, often increasing by an order of magnitude or more as the eccentricity of the function approaches unity. The greatest differences occur when the

[10] Friedman has suggested that the optimal rate of price change is actually negative, but this raises other issues which are beyond the scope of this paper.

TABLE II

RATIO OF EXPECTED CERTAINTY EQUIVALENCE LOSS TO EXPECTED BAYESIAN LOSS IN 1970/1 FOR ALTERNATIVE LOSS FUNCTION PARAMETERIZATIONS ASSUMING ZERO TARGETS FOR RATE OF PRICE CHANGE AND UNEMPLOYMENT RATE

A. Model-I

Angle of Rotation (radians)	Eccentricity				
	0.0 (circle)	0.75	0.87	0.97	0.99
$\pi/12$	1.001	1.000	1.000	1.000	1.000
$\pi/6$	1.001	1.092	1.188	1.512	2.258
$\pi/4$	1.001	i.354	1.815	2.925	3.759
$\pi/3$	1.001	1.672	2.752	5.288	6.837
$5\pi/12$	1.001	1.769	3.515	9.652	14.857
$\pi/2$	1.001	1.404	2.802	15.705	52.060
$7\pi/12$	1.001	1.007	1.024	1.402	859.029
$2\pi/3$	1.001	1.210	2.087	12.304	54.893
$3\pi/4$	1.001	1.469	2.569	6.599	10.292
$5\pi/6$	1.001	1.400	1.946	2.993	3.590
$11\pi/12$	1.001	1.206	1.389	1.609	1.772
$\pi(\equiv 0)$	1.001	1.055	1.100	1.150	1.328

B. Model-II

Angle of Rotation (radians)	Eccentricity				
	0.0 (circle)	0.75	0.87	0.97	0.99
$\pi/12$	1.067	1.035	1.018	1.002	1.000
$\pi/6$	1.067	1.040	1.028	1.019	1.041
$\pi/4$	1.067	1.052	1.049	1.056	1.073
$\pi/3$	1.067	1.067	1.071	1.085	1.093
$5\pi/12$	1.067	1.081	1.090	1.104	1.110
$\pi/2$	1.067	1.091	1.104	1.120	1.126
$7\pi/12$	1.067	1.098	1.115	1.135	1.143
$2\pi/3$	1.067	1.101	1.122	1.151	1.164
$3\pi/4$	1.067	1.096	1.121	1.169	1.197
$5\pi/6$	1.067	1.082	1.104	1.182	1.266
$11\pi/12$	1.067	1.061	1.063	1.125	1.573
$\pi(\equiv 0)$	1.067	1.042	1.027	1.005	1.327

angle of rotation is in the vicinity of $\pi/2$ and the eccentricity approaches unity. This particular parameterization of the criteria function puts the maximum weight on the unemployment target. Since we are very uncertain about the unemployment equation parameters in Model-I (relative to Model-II) the loss associated with using the certainty equivalence approach is much larger than the Bayesian approach which takes this uncertainty into account. However, if the loss function is circular, the certainty equivalence solution and the Bayesian solution generate very similar expected losses, regardless of the model employed.

On the basis of these results one may be tempted to conclude that the certainty equivalence solution could be used with impunity as a suitable approximation to the Bayesian solution if the loss function did not depart significantly from circularity.[11] But, according to Table III which reports the relative difference in

[11] This conclusion is, however, conditional upon the particular set of criteria variable targets chosen.

TABLE III

RATIO OF CERTAINTY EQUIVALENCE CONTROL VARIABLE SETTING TO BAYESIAN CONTROL VARIABLE SETTING IN 1970/1 FOR ALTERNATIVE LOSS FUNCTION PARAMETERIZATIONS ASSUMING ZERO TARGETS FOR RATE OF PRICE CHANGE AND UNEMPLOYMENT RATE

A. Model-I

Angle of Rotation (radians)	Eccentricity				
	0.0 (circle)	0.75	0.87	0.97	0.99
$\pi/12$	2.705	1.670	1.312	0.979	0.876
$\pi/6$	2.705	3.356	2.914	2.413	2.181
$\pi/4$	2.705	4.267	4.059	3.785	3.680
$\pi/3$	2.705	5.537	5.964	6.449	6.639
$5\pi/6$	2.705	7.105	9.012	12.373	14.239
$\pi/2$	2.705	8.274	12.785	28.088	48.233
$7\pi/12$	2.705	5.045	8.317	30.653	4328.440
$2\pi/3$	2.705	10.377	14.138	30.686	59.371
$3\pi/4$	2.705	6.970	7.733	9.503	10.616
$5\pi/6$	2.705	4.934	4.432	3.873	3.650
$11\pi/12$	2.705	3.826	2.959	2.096	1.792
$\pi(\equiv 0)$	2.705	3.506	2.497	1.634	1.354

B. Model-II

Angle of Rotation (radians)	Eccentricity				
	0.0 (circle)	0.75	0.87	0.97	0.99
$\pi/12$	1.130	1.114	1.098	1.048	0.987
$\pi/6$	1.130	1.107	1.091	1.063	1.049
$\pi/4$	1.130	1.108	1.097	1.083	1.078
$\pi/3$	1.130	1.114	1.107	1.100	1.097
$5\pi/6$	1.130	1.121	1.118	1.115	1.114
$\pi/2$	1.130	1.129	1.129	1.129	1.129
$7\pi/12$	1.130	1.138	1.142	1.145	1.146
$2\pi/3$	1.130	1.147	1.155	1.164	1.167
$3\pi/4$	1.130	1.154	1.169	1.191	1.200
$5\pi/6$	1.130	1.157	1.180	1.233	1.269
$11\pi/12$	1.130	1.149	1.173	1.285	1.577
$\tau(\equiv 0)$	1.130	1.131	1.133	1.111	1.354

control settings, the certainty equivalence solution typically calls for much larger settings for the control variable than the Bayesian solution in Model-I. Again, the largest differences occur for the loss function parameterizations emphasizing the unemployment target. Only in Model-II does there seem to be little difference in the control settings. Thus, if there is a cost to changing the control variable which has not been taken into account in formulating the control problem, use of the certainty equivalence solution instead of the Bayesian solution may impose a severe penalty upon the policymaker.

CONCLUDING REMARKS

In this paper we have found the exact setting for a single control variable in the case where the control variable is linked to several criteria variables by a stochastic linear recursive equation system and the criterion function is quadratic.

By taking the uncertainty in system parameters into account the solution differs from the certainty equivalence approach developed by other authors. By being applicable to recursive systems the solution represents an extension of the single equation solution already known. In the application developed here, the optimal setting provided a rule for the conduct of monetary policy one period at a time.

The results of the application were sensitive to the manner in which the equations were specified in each of two models. The two models, which had identical specifications of the price equation but different specifications of the unemployment rate equation, had vastly different control properties. The relative differences between the certainty equivalence approach and the Bayesian approach were generally greater for that model specification having the less precise contemporaneous coefficient estimates but better overall predictive properties. For judicious choices of loss function parameters the relative differences for expected losses were modest, but the relative differences in control settings remained large.

University of Illinois, Chicago
Federal Reserve Bank of Chicago