OPTIMAL CONTROL OF ECONOMETRIC MODELS
WITH AUTOCORRELATED DISTURBANCE TERMS*

Adrian Pagan

Econometric Research Program
Research Memorandum No. 156
November 1973

The research described in this paper was supported by NSF Grant GS 32003X.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey
OPTIMAL CONTROL OF ECONOMETRIC MODELS
WITH AUTOCORRELATED DISTURBANCE TERMS*

Adrian Pagan

1. Introduction

Recently a spate of articles have appeared concentrating upon the application of control theory to econometric models and economic theory, e.g., [7], [13]. Perhaps the best known discrete formulation for a stochastic system is that proposed in a series of articles by Chow [4], [5], [6], [7], which concentrates upon the linear econometric model with reduced form

\[ y_t = A y_{t-1} + C x_t + \epsilon_t \]  \hspace{1cm} (1.1)

where \( y_t \) is a \((p \times 1)\) vector, \( x_t \) a \((r \times 1)\) vector of control variables, \( A \) and \( C \) are non-stochastic matrices of

* I wish to thank both Gregory C. Chow for his perspicacious comments on earlier drafts which greatly improved the final version and the members of the Econometric Research Program Seminar at Princeton. Research was supported by a National Science Foundation Grant N.S.F. Grant No. GS 32003X and an Australian National University Post-Doctoral Travelling Fellowship. The Econometric Research Program provided a pleasant and stimulating working environment as well as access to the excellent typing of Betty Kaminski.
parameters and \( E(\varepsilon_t) = 0 \), \( E(\varepsilon_t \varepsilon_t') = V \), \( E(\varepsilon_t \varepsilon_s') = 0 \), \( s \neq t \). Given a quadratic welfare function whose expectation is to be maximized over \( T \) periods

\[
W = E \sum_{t=0}^{T} Y_t' K_t Y_t
\]

(1.2)

with \( K_t \) a symmetric, positive semi-definite matrix of weights reflecting the relative priorities of variables in the decision maker's welfare function, Chow shows that the optimal control rule obtained by maximizing (1.2) with respect to (1.1) is \( \hat{x}_t = G_t E_{t-1}(y_{t-1}) \) where \( G_t \) can be derived as the solution to the following set of equations:

\[
G_t = - (C' H_t C)^{-1} C' H_t A
\]

(1.3a)

\[
H_{t-1} = K_{t-1} + (A + C G_t)' H_t (A + C G_t)
\]

(1.3b)

\[
H_T = K_T
\]

(1.3c)

which come from a dynamic programming solution that involves the maximization of \( \hat{W}_t = E_{t-1}(y_t' H_t y_t) \) with respect to \( x_t \) for all \( t = 1, \ldots, T \), where \( E_{t-1} \) denotes that the
expectation is conditional on information prevailing in period $t-1$.\textsuperscript{1,2}

A difficulty with Chow's approach is the lack of an explicit treatment of autocorrelation in the disturbance term. He suggests that the structural equations may be adjusted with the Cochrane-Orcutt transformation if the disturbances follow an autoregressive (A.R.) format but this will increase $p$, i.e., the dimension of $y_t$, with the increase being particularly large if the fourth order processes found by Wallis [16] to be present in a number of quarterly models are encountered. Furthermore, no such simple solution is available if there are moving average (M.A.) errors entering the model; cases of this may be found in [15] and [17]. Given the potential for discovery of either form it seems apposite to extend Chow's framework to encompass both disturbance formats with the stricture that the dimension of the state variable $y_t$ should be no larger than $p$.

\textsuperscript{1}Equations (1.3a)-(1.3c) may be derived from any of Chow's articles by noting that there is no constant term in the reduced form (1.1) and that there are no target levels entering the welfare function, e.g., (1.3a) and (1.3b) correspond to (9) and (14) of [7] with $b_t = a_t = 0$.

\textsuperscript{2}Section 5 extends the algorithm of this paper to the case when $b_t, a_t \neq 0$.

Although $E_{t-1}(y_{t-1}) = y_{t-1}$ in Chow's papers this is not so in section 3.
Rather than present separate derivations for each type of disturbance it is useful to work with a representative that incorporates both, and this is defined in section 2. The general theory is then outlined in section 3 and given specific content in section 4. Section 5 modifies the resulting formulae to permit both a time-varying constant term to appear in the equations and targets for each of the variables, while section 6 discusses some complications arising in the implementation of the algorithm. Finally, section 7 provides an example based upon Klein's Model I of the U.S. economy.

2. The Autocorrelation Representation

Attention will be focused upon the \( q \)th order AR.

\[
    u_t = \phi_1 u_{t-1} + \cdots + \phi_q u_{t-q} + e_t \quad (2.1)
\]

and MA

\[
    u_t = e_t + \phi_1 e_{t-1} + \cdots + \phi_q e_{t-q} \quad (2.2)
\]

processes, where \( \phi_1, \ldots, \phi_q \) are non-stochastic matrices of parameters and \( e_t \) has the properties \( \mathbb{E}(e_t) = 0 \), \( \mathbb{E}(e_t e'_t) = \Omega \), \( \mathbb{E}(e_t e'_s) = 0 \) for \( t \neq s \). Both (2.1) and (2.2)
may be written in the form

\[ u_t = \Phi v_{t-1} + \varepsilon_t \]  \hspace{1cm} (2.3)
\[ v_t = B v_{t-1} + \xi_t \]  \hspace{1cm} (2.4)

by use of the following definitions

\[ \Phi = [\phi_1 \ldots \phi_q] \quad \text{(AR and MA)} \]  \hspace{1cm} (2.5)
\[ v'_t = [u'_t \ldots u'_{t-q+1}] \quad \text{(AR)} \]  \hspace{1cm} (2.6)
\[ v'_t = [e'_t \ldots e'_{t-q+1}] \quad \text{(MA)} \]  \hspace{1cm} (2.7)

\[ B = \begin{bmatrix}
\phi_1 & \cdots & \phi_q \\
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{bmatrix} \quad \text{(AR)} \]  \hspace{1cm} (2.8)

\[ B = \begin{bmatrix}
0 & 0 & \cdots & 0 \\
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{bmatrix} \quad \text{(MA)} \]  \hspace{1cm} (2.9)
\[ \xi_t' = [e_t' \ 0 \ldots \ 0] \quad (AR \ and \ MA) \ . \quad (2.10) \]

The structure of B assumes the crucial role, in section 4, of differentiating the control equations for each disturbance type. As it can also be shown that, if the errors were to follow an autoregressive-moving average structure \[2\], (2.3) and (2.4) would be appropriate, there is scope for a wider class of disturbances than considered explicitly in this paper.

3. General Theory

The reduced form is assumed to be

\[ Y_t = A \ Y_{t-1} + C \ x_t + u_t \quad (3.1) \]

where \( u_t \) is given in (2.3) and (2.4). If \( Y_t' = [y_t' \ v_t'] \) the relations in (2.3), (2.4) and (3.1) may be summarized in the system

\[
\begin{bmatrix}
  Y_t \\
  v_t
\end{bmatrix} =
\begin{bmatrix}
  A & \phi \\
  0 & B
\end{bmatrix}
\begin{bmatrix}
  Y_{t-1} \\
  v_{t-1}
\end{bmatrix} +
\begin{bmatrix}
  C \\
  0
\end{bmatrix}
\begin{bmatrix}
  x_t \\
  0
\end{bmatrix} +
\begin{bmatrix}
  e_t \\
  \xi_t
\end{bmatrix} \quad (3.2)
\]

or

\[ Y_t = A \ Y_{t-1} + C \ x_t + \xi_t \quad . \quad (3.3) \]
As (3.3) and (1.1) are identical, the optimal controls may be found from (1.3a) - (1.3c). Simplification of these equations begins by partitioning \( K_t \) and \( H_t \) conformably with \( y_t \) viz.

\[
K_t = \begin{bmatrix} K_t & 0 \\ 0 & 0 \end{bmatrix} \quad H_t = \begin{bmatrix} H_t & H_t^* \\ H_t^* & H_t^{**} \end{bmatrix}
\]

using the knowledge that \( H_t \) is symmetric (this following from (1.3b) and (1.3c)). Then the equivalent of (1.3a) is

\[
G_t = - \left[ \begin{bmatrix} C' & 0 \\ H_t^* & H_t^{**} \end{bmatrix} \right]^{-1} \left[ \begin{bmatrix} C' & 0 \\ H_t^* & H_t^{**} \end{bmatrix} \right]^{-1} \left[ \begin{bmatrix} C' & 0 \\ H_t^* & H_t^{**} \end{bmatrix} \right]
\]

\[
= - (C'H_tC)^{-1} \left[ C'H_tA + C'H_t^*B \right] \quad (3.4)
\]

\[
= [G_t; G_t^*] \quad (3.5)
\]

As evidenced by (3.4) it is unnecessary to evaluate all elements of \( H_t \) in order to compute the control rule contained in \( G_t \), i.e. only \( H_t \) and \( H_t^* \) are required. As these both appear in the first "row" of the partitioned matrix \( H_t \), they are obtained by multiplying the first "row" of
\((A + C G_t)'H_t\) by the matrix \((A + C G_t)\). As \((A + C G_t)\) is
\[
\begin{bmatrix}
A + CG_t & \phi + CG^* \\
0 & B
\end{bmatrix},
\]
the first "row" of the product \((A + C G_t)'H_t\) becomes
\([D'H_t; D'H^*_t]\) where \(D = A + CG_t\). Performing the multiplication with the matrix \((A + C G_t)\) yields
\[D'H_t(A + CG_t); D'H_t(\phi + CG^*_t) + D'H^*_t B\]
or, in the recursive format of (1.3b),
\[
H_{t-1} = K_{t-1} + (A + CG_t)'H_t(A + CG_t) \quad (3.7)
\]
\[H^*_t = (A + CG_t)'[H_t(\phi + CG^*_t) + H^*_t B] \quad (3.8)\]
\[H_T = K_T, H^*_T = 0 . \quad (3.9)\]

The expressions for \(G_t\) from (3.4) and \(H_t\) in (3.7) reveal that the optimal controller \(\hat{x}_t = G_t y_{t-1} + G^*_t E_{t-1} v_{t-1}\) is composed of two terms; one accounting for the reduced form dynamics \((G_t y_{t-1})\) and one \((G^*_t E_{t-1} v_{t-1})\) for the disturbance.
dynamics. Given such a division it is not surprising to find that the former is identical to the control rule found in Chow's articles.

4. Special Theory

To illustrate the methodology only the \( q \)th order AR is treated in detail. Using (2.5), (2.6) and (2.8) with a partitioning of \( H \) as \([H_1, t: : : : H_q, t] \) allows \( G_t^* \) in (3.4) to be written as

\[
- (C'H_t C)^{-1} \begin{bmatrix} C'H_{t+1}^\phi & \cdots & C'H_q^\phi \\ C'H_1 & \cdots & C'H_q \end{bmatrix} \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_q \end{bmatrix}
\]

which, if \( G_t^* = [G_1, t: : : : G_q, t] \), simplifies to

\[
G_j, t = -(C'H_t C)^{-1} (C'H_{t+1}^\phi_j + C'H_1^\phi + C'H_{j+1}^\phi) \quad j = 1, \ldots, q-1
\]

(4.2)

\[
G_q, t = -(C'H_t C)^{-1} (C'H_q^\phi + C'H_1^\phi) 
\]

(4.3)

The recursive equation (3.8) may be decomposed by substituting the following relations
\[
(\phi + C G_t^*) = (\phi_1: \ldots: \phi_q) + (C G_{1,t}: \ldots: C G_{q,t})
\]
\[
= (\phi_1 + C G_{1,t}: \ldots: \phi_q + C G_{q,t})
\]
\[
H_t^* B = [H_{1,t}: \ldots: H_{q,t}] \begin{bmatrix}
\phi_1 & \ldots & \phi_q \\
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
0 & 0 & \ldots & 1
\end{bmatrix}
\]
\[
= (H_{1,t}\phi_1 + H_{2,t}: H_{1,t}\phi_2 + H_{3,t}: \ldots: H_{1,t}\phi_q)
\]
to yield
\[
H_{j,t-1} = (A + C G_t^*)' [H_t(\phi_j + C G_{j,t}) + H_{1,t}\phi_j + H_{j+1,t}] \quad j=1, \ldots, q-1
\] (4.4)
\[
H_{q,t-1} = (A + C G_t^*)' [H_t(\phi_q + C G_{q,t}) + H_{1,t}\phi_q] \quad (4.5)
\]
\[
H_{j,t} = 0 \quad j=1, \ldots, q \quad . \quad (4.6)
\]

Gathering the results from (3.4), (3.7), (3.9), (4.2), (4.3), (4.4), (4.5) and (4.6) leaves the optimal control equations for the \(q^\text{th}\) order AR as
\[
\hat{x}_t = G_t y_{t-1} + \sum_{j=1}^{q} G_{j,t} E_{t-1}(u_{t-j}) \quad (4.7a)
\]
\[ G_t = - (C' H_t C)^{-1} C' H_t A \quad (4.7b) \]

\[ G_{j,t} = - (C' H_t C)^{-1} (C' H_t \phi_j + C' H_1, t \phi_j + C' H_{j+1}, t) \quad j=1, \ldots, (q-1) \quad (4.7c) \]

\[ G_{q,t} = - (C' H_t C)^{-1} (C' H_t \phi_q + C' H_1, t \phi_q) \quad (4.7d) \]

\[ H_{t-1} = K_{t-1} + (A + C G_t)' H_t (A + C G_t) \quad (4.7e) \]

\[ H_{j,t-1} = (A + C G_t)' [H_t (\phi_j + C G_j, t) + H_1, t \phi_j + H_{j+1}, t] \quad j=1, \ldots, (q-1) \quad (4.7f) \]

\[ H_{q,t-1} = (A + C G_t)' [H_t (\phi_q + C G_q, t) + H_1, t \phi_q) \quad (4.7g) \]

\[ H_T = K_T, \ H_{j,T} = 0 \quad j=1, \ldots, q \quad (4.7h) \]

The equations corresponding to (4.7) for the MA case are identical except that \( H_{1,t} \) is set equal to zero throughout, e.g., (4.7c) would become

\[ G_{j,t} = - (C' H_t C)^{-1} (C' H_t \phi_j + C' H_{j+1}, t) \quad j=1, \ldots, (q-1). \]

Furthermore, in contrast to the AR case where

\[ E_{t-1}(v_{t-1}) = v_{t-1}, \ \text{i.e., the values } u_{t-1}, \ldots, u_{t-q} \ \text{contained in } v_{t-1} \ \text{are completely known at } t-1, \ \text{the components of} \]
\( E_{t-1}(v_{t-1}) \) for an MA must be generated. To do this, set 
\( z_t = v_{t-1} \) and observe that (2.3) and (2.4) constitute a state space representation from which \( \hat{v}_{t-1} = E_{t-1}(v_{t-1}) = E_{t-1}(z_t) = \hat{z}_t \), i.e., the one-step predictor of \( z_t \), must be extracted. From the theory underlying the Kalman-Bucy filter \([10]\) this is given by \( \hat{z}_t = B \hat{z}_{t-1} + K(u_{t-1} - \Phi \hat{z}_{t-1}) \), where \( K' = [1\ 0 \ldots 0] \), which results in a recursive relationship for the computation of \( E_{t-1}(v_{t-1}) \).

Three comments may be made about (4.7) and its MA version. Firstly, the absence of \( H_{1,t} \) for the MA case stems from the structure of \( B \), i.e., the first row is zero for a MA but has \( \phi_j \) \((j=1,\ldots,q)\) as \( j^{th} \) element for an AR. Secondly, there is a simple structure to (4.7) that allows for easy computer programming and enables the same program to be utilized for either disturbance format -- a symmetry that has been commented upon and exploited elsewhere in estimation \([11]\). Thirdly, the order of the matrices in (4.7) is never higher than that characteristic of the model if it had no autocorrelation; an original objective fulfilled.
5. Generalization of the Algorithm

It is now time to generalize the control problem to allow for specific targets in the welfare function and a constant term in the reduced form. There is little difficulty in so doing. In the expanded problem it is necessary to minimize

\[ W = E \left[ \sum_{t=0}^{T} (Y_t - a_t) \cdot K_t (Y_t - a_t) \right] \]  \hspace{1cm} (5.1)

subject to \( Y_t = A Y_{t-1} + C X_t + b_t + \varepsilon_t \).

When \( Y_t \) is the augmented vector of section 3 the vectors \( a_t \) and \( b_t \) will be \( a'_t = [a'_t \ 0] \), \( b'_t = [b'_t \ 0] \) where the correspondence is obvious. The optimal controller for (5.1) is now \( \hat{X}_t = G_t Y_t + q_t \) (Chow [7, eq. (8)]) where

\[ q_t = - (C' H_t C)^{-1} C' (H_t b_t - h_t) \]  \hspace{1cm} (5.2)

\[ h_{t-1} = k_{t-1} + (A + C G_t)' (h_t - H_t b_t) \]  \hspace{1cm} (5.3)

\[ k_t = K_t a_t \]  \hspace{1cm} (5.4)

\[ h_T = k_T \]  \hspace{1cm} (5.5)

and all other matrices are as in (1.3a) - (1.3c) (eq. (5.3) is
Substituting \( h_t' = [h_t' \ h_t'^*] \), \( g_t' = [g_t' \ g_t'^*] \) and the partitioned form of \( H_t \) into (5.2) gives

\[
\begin{bmatrix}
  g_t' \\
  g_t'^*
\end{bmatrix} = - (C' H_t C)^{-1} (C' 0)
\begin{bmatrix}
  (H_t \ H_t^*) \\
  (H_t^* \ H_t^**) \\
  (0) \\
  (0)
\end{bmatrix}
\begin{bmatrix}
  b_t \\
  0
\end{bmatrix}
- \begin{bmatrix}
  h_t' \\
  h_t'^*
\end{bmatrix}
\]

\[
= - (C' H_t C)^{-1}
\begin{bmatrix}
  (C'0) \\
  (0)
\end{bmatrix}
\begin{bmatrix}
  H_t b_t - h_t \\
  H_t^* b_t - h_t^*
\end{bmatrix}
\]

\[
\therefore \quad g_t = - (C' H_t C)^{-1} C' (H_t b_t - h_t) \quad . \quad (5.6)
\]

Similarly, for (5.3),

\[
\begin{bmatrix}
  h_{t-1} \\
  h_{t-1}^*
\end{bmatrix} = \begin{bmatrix}
  k_{t-1} \\
  0
\end{bmatrix} + \begin{bmatrix}
  (A + C G_t)' \\
  (\phi + C G_t^*)'
\end{bmatrix}
\begin{bmatrix}
  O \\
  B'
\end{bmatrix}
\begin{bmatrix}
  (h_t) \\
  (h_t^*)
\end{bmatrix}
\begin{bmatrix}
  - (H_t b_t) \\
  - (H_t^* b_t)
\end{bmatrix}
\]

As (5.6) reveals that only \( h_t \) is required for the control computations, the appropriate recursion is

\[
h_{t-1} = k_{t-1} + (A + C G_t)' (h_t - H_t b_t) \quad . \quad (5.7)
\]
(5.6) and (5.7) are identical to the adjustment that would be necessary if there were a constant term and/or targets but no autocorrelation. The influence of these equations on the controller is expressed by modifying (4.7a) to

\[ \hat{x}_t = G_t y_{t-1} + \sum_{j=1}^{q} G_{j,t} E_{t-1}(u_{t-j}) + g_t \]  

This completes the development of the optimal control law for the linear-quadratic problem with autocorrelation, and the benefits from viewing it in this framework may be enumerated.

(i) The dimensionality of the problem has been retained at \( p \), i.e., it corresponds to control with no autocorrelation.

(ii) By a suitable partitioning of the \( H_t \) matrix, it has emerged that not all elements of this matrix need be computed to derive the requisite law: a result that has obvious benefits in terms of the reduced number of multiplications that must be performed.

(iii) Because we have merely manipulated Chow's equations it is possible to extend the current results to other situations that he has examined. In particular, if \( A, C \) and \( \phi \) are stochastic, one can employ the equations in [8] in an exactly analogous manner to the above to obtain the appropriate simplifications.

An Appendix considers the derivation of the expected welfare loss associated with the optimal control rule.
6. **Implementation of the Algorithm**

Under this heading we consider the conversion of a structural equation model (see [9, p. 297] for nomenclature)

\[ \Gamma y(t) + B x(t) + U(t) = 0 \]  \hspace{1cm} (6.1)

to a reduced form \( y(t) = \Pi x(t) + V(t) \) with an error format either of the type

\[ V(t) = \phi_1 V(t-1) + \ldots + \phi_q V(t-q) + E(t) \]  \hspace{1cm} (AR)  \hspace{1cm} (6.2)

or

\[ V(t) = E(t) + \phi_1 E(t-1) + \ldots + \phi_q E(t-q) \]  \hspace{1cm} (MA)  \hspace{1cm} (6.3)

where \( E(t) \) has zero mean, finite contemporaneous covariance matrix \( \Omega \) and no serial correlation. Once one of (6.2) or (6.3) is arrived at control can proceed with the algorithm described earlier.

(i) **AR Case**

Assume there is a \( q^{th} \) order AR in the structure i.e.,

\[ U(t) = A_1 U(t-1) + \ldots + A_q U(t-q) + \tilde{E}(t). \]  \hspace{1cm} (6.4)
Pre-multiplying (6.4) by $-\Gamma^{-1}$ gives

$$-\Gamma^{-1} U(t) = -\Gamma^{-1} A_1 U(t-1) - \ldots - \Gamma^{-1} A_q U(t-q) - \Gamma^{-1} \tilde{E}(t) \quad (6.5)$$

i.e.,

$$V(t) = \phi_1 V(t-1) + \ldots + \phi_q V(t-q) + E(t) \quad (6.6)$$

where $V(t) = -\Gamma^{-1} U(t)$, $\phi_j(t) = \Gamma^{-1} A_j \Gamma$ and $E(t) = -\Gamma^{-1} \tilde{E}(t)$. Now (6.6) has the requisite structure.

(ii) **MA Case**

Assume there is a $q^{th}$ order MA in the structure, i.e.,

$$U(t) = \tilde{E}(t) + A_1 \tilde{E}(t-1) + \ldots + A_q \tilde{E}(t-q) \quad (6.7)$$

Pre-multiplying (6.7) by $-\Gamma^{-1}$ gives

$$-\Gamma^{-1} U(t) = -\Gamma^{-1} \tilde{E}(t) - \Gamma^{-1} A_1 \tilde{E}(t-1) - \ldots - \Gamma^{-1} A_q \tilde{E}(t-q) \quad (6.8)$$

or

$$V(t) = E(t) + \phi_1 E(t-1) + \ldots + \phi_q E(t-q) \quad (6.9)$$

where $V(t) = -\Gamma^{-1} U(t)$, $\phi_j = \Gamma^{-1} A_j \Gamma$ and $E(t) = -\Gamma^{-1} \tilde{E}(t)$. Now (6.9) has the required format.
7. **An Example**

The chosen example involves the anti-depression experiments performed with Klein's model by Theil [14]. This model has the following eight equations (where the nomenclature and definitions may be found in [9, p. 303]):

\[
    C_t = b_1 P_t + b_2 W_t + b_3 P_{t-1} + b_4
\]

\[
    I_t = b_5 P_t + b_6 P_{t-1} + b_7 K_{t-1} + b_8
\]

\[
    W^* = b_9 E_t + b_{10} E_{t-1} + b_{11} A_t + b_{12}
\]

\[
    P_t = Y_t - W_t
\]

\[
    E_t = Y_t + T_t - W^*
\]

\[
    Y_t = C_t + T_t - T_t + G_t
\]

\[
    K_t = K_{t-1} + I_t
\]

\[
    W_t = W^* + W^{**}
\]

Theil considers optimal policies for the years 1933-1936 with the strategy being, generally speaking, to restore the economy to the 1929 activity levels by 1936. Expressed
in numerical terms this implies the ideal values given in Table 1 (see [14, p. 79]), except that the desired level of wages for 1936 was computed by applying the 1929 ratio of wages to income to the target income level for 1936 -- with the latter coming from substitution of the values of Table 1 into \( Y_t = C_t + I_t - T_t + G_t \) -- and then, following Theil, linearly interpolating values for 1933-5. The division of Table 1 is into the targets of consumption (C), investment (I) and total wage bill (W) and the instruments, taxes (T), government expenditure (G) and the government wage bill (W**).

\[
\begin{array}{cccccc}
\text{TABLE 1} \\
\hline
\text{Desired Values for Target Variables and Instruments} & 1933 & 1934 & 1935 & 1936 \\
C & 49.69 & 53.78 & 57.88 & 61.97 \\
I & -3.10 & 0.00 & 3.10 & 6.20 \\
W & 37.86 & 41.42 & 44.98 & 48.53 \\
W** & 5.038 & 5.254 & 5.469 & 5.685 \\
T & 7.396 & 7.635 & 7.874 & 8.113 \\
\hline
\end{array}
\]
Full Information Maximum likelihood estimates of the parameters of Klein's Model I under the assumptions of no autocorrelation and a first order moving average specification are in Table 2. An algorithm, described in [3], was used to generate estimates for the MA specification, under the assumption that pre-period values of the disturbances were zero.3

With a welfare function giving equal weight to each of the variables in Table 1 and zero to all others, the four-period feedback control matrices, when there is first order M.A. autocorrelation, are in Table 3. Table 4 contains the expected values of the target and control variables for 1933-6 assuming that the optimal control rule is in use (computed from equation (9) of the Appendix). A comparison of Table 4 with the desired values of Table 1 reveals two outstanding features:

(a) a consistent failure to attain the wage target.

(b) the strange pattern of investment behaviour, probably induced by the negative effect of current profits but positive effect of lagged profits.

These points suggest that the welfare function would need to be modified to give more weight to wages and

3 My thanks to Ray Byron for providing his program. Extensive tampering with it on my part makes it even more urgent that the usual caveat on responsibility be noted.
TABLE 2

F.I.M.L. Estimates of Klein's Model I

| b_1   | -0.232 | -0.019 |
| b_2   | 0.802  | 0.779  |
| b_3   | 0.386  | 0.267  |
| b_4   | 18.34  | 17.341 |
| b_5   | -0.801 | -0.514 |
| b_6   | 1.052  | 0.996  |
| b_7   | -0.148 | -0.182 |
| b_8   | 27.26  | 30.162 |
| b_9   | 0.234  | 0.294  |
| b_{10} | 0.285 | 0.254 |
| b_{11} | 0.235 | 0.201 |
| b_{12} | 5.79  | 3.979  |
| MA(1) |        | 0.224  |
| MA(2) |        | 0.104  |
| MA(3) |        | 0.132  |
| Likelihood | 1.782 | 0.537 |

investment if the control exercise was a serious one. However, as Klein's Model I is generally treated as a mere workhorse for experimentation, such a recomendation should not be taken too seriously.

To provide some index of the gains from recognizing autocorrelation patterns in the controller the expected welfare loss was computed twice; once with \( G_t^* \) set equal to zero and
### TABLE 3

**Optimal Control Matrices**

<table>
<thead>
<tr>
<th></th>
<th>$T_t$</th>
<th>$W_{t}^{**}$</th>
<th>$G_{t}$</th>
<th></th>
<th>$T_t$</th>
<th>$W_{t}^{**}$</th>
<th>$G_{t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{t-1}$</td>
<td>0.0126</td>
<td>-0.1447</td>
<td>-0.4526</td>
<td></td>
<td>0.0108</td>
<td>-0.1449</td>
<td>-0.4510</td>
</tr>
<tr>
<td>$E_{t-1}$</td>
<td>-0.0919</td>
<td>-0.1015</td>
<td>-0.1044</td>
<td></td>
<td>-0.0885</td>
<td>-0.1014</td>
<td>-0.1058</td>
</tr>
<tr>
<td>$K_{t-1}$</td>
<td>-0.0885</td>
<td>-0.0076</td>
<td>0.1273</td>
<td></td>
<td>-0.0685</td>
<td>-0.0071</td>
<td>0.1200</td>
</tr>
<tr>
<td>$e_{1}(t-1)$</td>
<td>0.0366</td>
<td>-0.1029</td>
<td>-0.1326</td>
<td></td>
<td>0.0343</td>
<td>-0.1029</td>
<td>-0.1317</td>
</tr>
<tr>
<td>$e_{2}(t-1)$</td>
<td>-0.0032</td>
<td>-0.0023</td>
<td>-0.0308</td>
<td></td>
<td>-0.0031</td>
<td>-0.0023</td>
<td>-0.0307</td>
</tr>
<tr>
<td>$e_{3}(t-1)$</td>
<td>-0.0478</td>
<td>-0.0527</td>
<td>-0.0543</td>
<td></td>
<td>-0.0460</td>
<td>-0.0527</td>
<td>-0.0550</td>
</tr>
<tr>
<td>$e_{4}(t-1)$</td>
<td>-0.0010</td>
<td>-0.0031</td>
<td>-0.0183</td>
<td></td>
<td>-0.0010</td>
<td>-0.0031</td>
<td>-0.0183</td>
</tr>
<tr>
<td>$e_{5}(t-1)$</td>
<td>0.0140</td>
<td>0.0155</td>
<td>0.0160</td>
<td></td>
<td>0.0135</td>
<td>0.0155</td>
<td>0.0162</td>
</tr>
<tr>
<td>$e_{8}(t-1)$</td>
<td>-0.0285</td>
<td>0.0801</td>
<td>0.1033</td>
<td></td>
<td>-0.0267</td>
<td>0.0802</td>
<td>0.1026</td>
</tr>
<tr>
<td>constant</td>
<td>-0.4326</td>
<td>12.421</td>
<td>28.312</td>
<td></td>
<td>4.2603</td>
<td>13.659</td>
<td>24.638</td>
</tr>
</tbody>
</table>

**1935**

<table>
<thead>
<tr>
<th></th>
<th>$T_t$</th>
<th>$W_{t}^{**}$</th>
<th>$G_{t}$</th>
<th></th>
<th>$T_t$</th>
<th>$W_{t}^{**}$</th>
<th>$G_{t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{t-1}$</td>
<td>-0.0910</td>
<td>-0.1557</td>
<td>-0.3790</td>
<td></td>
<td>-0.2469</td>
<td>-0.2394</td>
<td>0.0188</td>
</tr>
<tr>
<td>$E_{t-1}$</td>
<td>-0.0760</td>
<td>-0.1002</td>
<td>-0.1138</td>
<td></td>
<td>-0.0704</td>
<td>-0.1034</td>
<td>-0.1016</td>
</tr>
<tr>
<td>$K_{t-1}$</td>
<td>-0.0084</td>
<td>-0.0016</td>
<td>0.0815</td>
<td></td>
<td>0.0439</td>
<td>0.0183</td>
<td>-0.0171</td>
</tr>
<tr>
<td>$e_{1}(t-1)$</td>
<td>0.0174</td>
<td>-0.1046</td>
<td>-0.1202</td>
<td></td>
<td>-0.0054</td>
<td>-0.1167</td>
<td>-0.0629</td>
</tr>
<tr>
<td>$e_{2}(t-1)$</td>
<td>-0.0117</td>
<td>-0.0032</td>
<td>-0.0246</td>
<td></td>
<td>-0.0251</td>
<td>-0.0105</td>
<td>0.0098</td>
</tr>
<tr>
<td>$e_{3}(t-1)$</td>
<td>-0.0395</td>
<td>-0.0521</td>
<td>-0.0592</td>
<td></td>
<td>-0.0366</td>
<td>-0.0537</td>
<td>-0.0528</td>
</tr>
<tr>
<td>$e_{4}(t-1)$</td>
<td>-0.0057</td>
<td>-0.0037</td>
<td>-0.0149</td>
<td></td>
<td>-0.0130</td>
<td>-0.0076</td>
<td>0.0038</td>
</tr>
<tr>
<td>$e_{5}(t-1)$</td>
<td>0.1161</td>
<td>0.0153</td>
<td>0.0174</td>
<td></td>
<td>0.0108</td>
<td>0.0158</td>
<td>0.0155</td>
</tr>
<tr>
<td>$e_{8}(t-1)$</td>
<td>-0.1353</td>
<td>0.0815</td>
<td>0.0936</td>
<td></td>
<td>0.0042</td>
<td>0.0909</td>
<td>0.0490</td>
</tr>
</tbody>
</table>
TABLE 4

Expected Values of the Target
and Endogenous Variables Under Control

<table>
<thead>
<tr>
<th></th>
<th>1933</th>
<th>1934</th>
<th>1935</th>
<th>1936</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>47.75</td>
<td>53.05</td>
<td>58.20</td>
<td>61.93</td>
</tr>
<tr>
<td>I</td>
<td>-7.854</td>
<td>-1.231</td>
<td>2.808</td>
<td>5.024</td>
</tr>
<tr>
<td>W</td>
<td>31.28</td>
<td>36.39</td>
<td>41.62</td>
<td>44.69</td>
</tr>
<tr>
<td>W**</td>
<td>4.981</td>
<td>4.736</td>
<td>4.518</td>
<td>5.578</td>
</tr>
<tr>
<td>T</td>
<td>5.991</td>
<td>6.519</td>
<td>7.588</td>
<td>9.049</td>
</tr>
<tr>
<td>G</td>
<td>17.55</td>
<td>15.78</td>
<td>14.49</td>
<td>12.35</td>
</tr>
</tbody>
</table>

once with $G_t^*$ equal to the values of Table 3. The losses were 385.13 and 383.98 respectively, signifying only a small advantage to the employment of the autocorrelation-modified controller -- a not unexpected result given the small estimated M.A. parameters.

8. Conclusion

A method has been described for the optimal control of linear econometric models with a quadratic performance criterion when there is autocorrelation of either the auto-regressive or moving average type. By augmenting the state vector it was possible to convert the system to one with no autocorrelation and hence to apply some well-known theorems. The advantage of such a methodology is its potential for application to a wider array of models than considered explicitly in this paper.
APPENDIX

Computation of the Expected Welfare Loss Under Control

Given that a control rule has been established, it is of interest to compute the expected welfare loss under such a rule. This is most easily illustrated in the A.R. case where the task is to evaluate

$$E \sum_{t=1}^{T} (y_t - a_t)'^\top K_t(y_t - a_t)$$

subject to

$$y_t = A y_{t-1} + C x_t + b_t + \phi v_{t-1} + e_t$$

$$v_t = B v_{t-1} + \epsilon_t$$

and $E(e_t' e_t') = \Omega$, $E(\epsilon_t' \epsilon_t') = \begin{bmatrix} \Omega & 0 \\ 0 & 0 \end{bmatrix} = R$ (since from (2.4) and (2.5) $\epsilon_t' = [e_t', 0]$).

(1) can be re-written as

$$E \sum_{t=1}^{T} \text{tr} K_t[E(y_t y_t') - 2 a_t E(y_t') + a_t a_t'] .$$

The control rule is

$$x_t = G_t y_{t-1} + G_t^* v_{t-1} + g_t$$

which leaves the equation for $y_t$, under control, as

$$y_t = (A + C G_t) y_{t-1} + (C G_t^* + \phi) v_{t-1} + (C g_t + b_t) + e_t$$
or
\[ y_t = D_t y_{t-1} + F_t y_{t-1} + d_t + e_t \]  \hspace{1cm} (5)

If \( V_t = E(y_t y_t') \), \( V^*_t = E(v_t y_t') \), \( V^{**}_t = E(v_t v_t') \), \( P_t = E(y_t) \) and \( P^*_t = E(v_t) \), the expected welfare loss is
\[ \sum_{t=1}^{T} \text{tr} \left( K_t[V_t - 2a_t P_t' + a_t a_t'] \right) \]

where \( V_t \) and \( P_t \) are the solutions of
\[ V_t = D_t V_{t-1} D_t' + 2F_t V^*_t - d_t - 2F_t P^*_{t-1} F_t' + d_t d_t' + \Omega \]  \hspace{1cm} (6)

\[ V^*_t = B V^*_t - d_t + 2B P^*_{t-1} - d_t B P^*_t + \Omega \]  \hspace{1cm} (7)

\[ V^{**}_t = B V^{**}_t - B P^*_t + \Omega \]  \hspace{1cm} (8)

\[ P_t = D_t P_{t-1} + F_t P^*_t + d_t \]  \hspace{1cm} (9)

\[ P^*_t = B P^*_t \]  \hspace{1cm} (10)

Each of these equations emerges by utilizing (3) and (5) to get the various expectations, eg. \( P_t \) is formed by taking the expectation of (5). Unfortunately, the M.A. situation is more complex since the control rule is now linear in the expected value of \( v(t-1) \) and not \( v(t-1) \), so that (5) is not
strictly appropriate, i.e., it should be \( y_t = D_t y_{t-1} + C G_t^* \hat{v}_{t-1}^+ + \phi v_{t-1} + d_t + e_t \) where \( \hat{v}_{t-1} = E_{t-1}(v_{t-1}) \). When
\( \tilde{v}_{t-1} = \hat{v}_{t-1} - v_{t-1} \) it becomes \( y_t = D_t y_{t-1} + F_t \tilde{v}_{t-1} + C G_t^* \tilde{v}_{t-1} + d_t + e_t \). From the properties of the Kalman-Bucy filter [9] \( E(\tilde{v}_{t-1}) = 0 \) so that (9) and (10) hold, but the variance of \( \tilde{v}_{t-1} \) is not zero and therefore (6) - (8) do not.

Can the use of (6) - (10) be justified for a M.A.? The answer is in the affirmative if \( e(0), \ldots, e(-q+1) \) are known with certainty as \( e(t) \) will then be computed exactly from
\( e(t) = u(t) - \sum_{j=1}^{q} \phi_j e(t-j) \) for any \( t \) in the control period. Because estimates of \( e(0), \ldots, e(-q+1), A, \phi \) and \( C \) are likely to be obtained jointly by F.I.M.L. estimation, it seems logical to treat all estimates in the same fashion. In particular, if \( A, \phi \) and \( C \) are regarded as non-stochastic, treating the estimated initial conditions as the correct values would seem justified. However, if the stochastic nature of \( A, B \) and \( C \) is recognized -- as in [8] -- new versions of (6)-(10) will need to be derived.
REFERENCES


