CALIFORNIA INSTITUTE OF TECHNOLOGY
Division of the Humanities and Social Sciences
Pasadena, California 91125

TWO NOTES ON DISTRIBUTED LAGS, PREDICTION, AND SIGNAL EXTRACTION

David M. Grether

Social Science Working Paper
Number 123
May 1976

TWO NOTES ON DISTRIBUTED LAGS, PREDICTION, AND SIGNAL EXTRACTION

David M. Grether

NOTE ONE: SPECIFICATION

I. INTRODUCTION

A wide variety of economic models include as explanatory variables either expectational variables or variables representing the result of some decision-making process. The first category includes both expectations about the future values of variables, e.g., next period's sales, the level of unemployment two quarters ahead, etc. and other subjective variables such as permanent income or the "normal" level of prices and interest rates. Examples of the second type are "desired" capital stock, planned production or inventory accumulation and so on.

Since data on expectations or specific decisions are frequently unavailable, these models are often made empirically testable by specifying the way in which the expectational or choice variables are related to observable quantities. As is well known, this specification often leads to a distributed lag model such as the familiar adaptive expectations model or the stock adjustment model. In this note we consider models of the form

(1) $$y_t = a + bx_t^* + u_t$$

where x_t^* is one of the types of variables mentioned above. We show that in a large variety of cases (1) reduces to a distributed lag model in which the lag distribution is a rational distributed lag [5]. Specifically, it is shown that rational distributed lags arise when the exogenous variable (or its p^{th} difference) has a mixed autoregressive, moving average representation and x_t^* is chosen to minimize the expected value of a quadratic objective function. For example, x_t^* could be the least squares forecast of x_{t+j} made at time t. It is also shown that the orders of the polynomials in the lag operator depend in a simple way upon the structure of the exogenous variable and upon the nature of the optimization problem.

In what follows it is assumed that all stochastic processes are zero mean covariance stationary processes with autoregressive representations.[1/] We use the following notational conventions:

(i) $$g_{yx}(z) = \sum_{k=-\infty}^{\infty} E(y_t x_{t-k}) z^k$$

(ii) If $H(z) = \sum_{-\infty}^{\infty} h_i z^i$ is the Laurent expansion of a function which converges in an annulus containing the unit circle, then $[H(z)]_+ = \sum_{0}^{\infty} h_i z_i$, and $[H(z)]_- = \sum_{-\infty}^{-1} h_i z_i$.

(iii) $$L^k \cdot x_t \equiv x_{t-k}$$

II. SIGNAL EXTRACTION AND PREDICTION

Let

(2) $$w_t = x_t + \eta_t$$

$$x_t = \frac{N(L)}{D(L)} \varepsilon_t$$

$$\eta_t = \frac{R(L)}{S(L)} \zeta_t$$

where N( ), D( ), R( ), and S( ) are polynomials of degree n, d, r, and s, respectively, and $\{\varepsilon_t\}$, $\{\zeta_t\}$ are mutually uncorrelated white noise sequences. Further, let T( ) be a polynomial of degree max (n + s, d + r) satisfying:

$$\text{(3)}\qquad \sigma^2 T(z)T(z^{-1}) = \sigma_\varepsilon^2 N(z)S(z)N(z^{-1})S(z^{-1}) + \sigma_\zeta^2 R(z)R(z^{-1})D(z^{-1})D(z)$$

with the roots of T( ) lying outside the unit circle and σ^2 chosen so that $t_o = 1$.

Theorem (Signal Extraction): Let $\hat{x}_{t+v,t}$ be the least squares estimate of x_{t+v} made at time t based upon observations on w_s, $s \leq t$. Then

$$\hat{x}_{t+v,t} = \gamma(L)\, w_t$$

$$\gamma(z) = \frac{S(z)N_v(z)}{T(z)}$$

where N_v( ) is a polynomial of order max (n - v, d - 1, 0).[2/]

Proof. From (2) we have

$$g_{wx}(z) = g_{xx}(z) = \sigma_\varepsilon^2 \frac{N(z)N(z^{-1})}{D(z)D(z^{-1})}$$

$$g_{ww}(z) = \sigma^2 \frac{T(z)T(z^{-1})}{D(z)S(z)D(z^{-1})S(z^{-1})}.$$

Thus by [9],

$$\text{(4)}\qquad \gamma(z) = \frac{D(z)S(z)}{\sigma^2 T(z)} \left[\frac{\sigma_\varepsilon^2 N(z)N(z^{-1})S(z^{-1})D(z^{-1})}{D(z)D(z^{-1})T(z^{-1})z^v}\right]_+$$

$$= \frac{D(z)S(z)}{\sigma^2 T(z)} \left[\frac{\sigma_\varepsilon^2 N(z)N(z^{-1})S(z^{-1})}{T(z^{-1})D(z)z^v}\right]_+ \quad {}^{3/}$$

The expression under the $[\ ]_+$ operator can be evaluated using theorem 1 on page 93 of [9]. To see this, note that

$$\left[\frac{\sigma_\varepsilon^2 N(z)N(z^{-1})S(z^{-1})}{T(z^{-1})D(z)z^v}\right]_+ = \frac{\left[\dfrac{\sigma_\varepsilon^2 N(z)N(z^{-1})S(z^{-1})}{T(z^{-1})z^v}\right]_+}{D(z)} + \left[\frac{\left[\dfrac{\sigma_\varepsilon^2 N(z)N(z^{-1})S(z^{-1})}{T(z^{-1})z^v}\right]_-}{D(z)}\right]_+$$

The first term is clearly of the form $A_1(z) / D(z)$ where A_1( ) is of order max (n - v, 0). To obtain the second term we may expand $\frac{1}{D(z)}$ by partial fractions and apply Whittle's theorem to each term in in the resulting sum. On recombining terms, the second expression is of the form $A_2(z) / D(z)$ where A_2( ) is of order d - 1.

Q. E. D.

Corollary (Prediction): Let $x_t = \frac{N(L)}{D(L)} \varepsilon_t$ where $\{\varepsilon_t\}$ is white noise and N( ) and D( ) are polynomials of degree n and d respectively. The least squares forecast of x_{t+v} made at time t based upon observation on x_s, $s \leq t$ is given by

$$\hat{x}_{t+v} = \frac{N_v(L)}{N(L)} x_t$$

where N_v( ) is a polynomial of degree max (n - v, d - 1, 0).

In this case $g_{xx}(z) = g_{ww}(z)$, and the same proof works with $T(z) \equiv N(z)$, and $S(z) \equiv 1$.

If in the structural model (1), x_t^* is the least squares forecast of some covariance stationary process, either based on past observations on the process itself or upon observations with (serially correlated) measurement error, the model becomes

$$(1)' \qquad y_t = a + b\gamma(L)x_t + u_t.$$

Except for the case of finite order autoregressions observed without error, these forecasts in general depend upon the entire past history of the x series. The preceding result shows that for arbitrary autoregressive, moving average processes, the lag distribution is rational and the orders (or at least upper bounds on them) may be obtained by examining the properties of the observed exogenous variable.

Models employing expectations about future levels of observable economic variables are sufficiently common that citing examples seems unnecessary. For some examples explicitly using expectations about unobserved components of economic time series, see [6] and the references there.

III. OTHER APPLICATIONS

Suppose that the decision problem is not forecasting or estimating a noise corrupted signal, but, instead, it is to optimize an objective function which depends upon the future values of a time series or upon some unobserved component of a time series. It is well known that if the objective function is suitably restricted, the unknown variables may be replaced by their conditional expectations and the solution obtained in terms of the certainty equivalents [8]. Replacing these conditional expectations by the optimum forecasts or extractions will then lead to a

distributed lag model. As before the order of the lag operators will depend relatively simply upon the characteristics of the process being forecast and upon the nature of the objective function. While this does provide a generalization of the results of the previous section, we emphasize at the outset that the approach has some severe limitations. First, it is restricted to problems in which the objective function is quadratic which rules out many, perhaps most, interesting applications. Also, the restriction to considering only linear decision rules or normal processes ought to be reemphasized.

Consider the following generalization of the prediction problem:

$$(5) \qquad \min_{\gamma} E\{((A(L)\gamma(L) + C(L))x_t)^2\}$$

$$\gamma(z) = \sum_{i=k}^{\infty} \gamma_i z^i, \qquad g_{xx}(z) = \sigma^2 B(z)B(z^{-1}).$$

The solution is

$$(6) \qquad \gamma(z) = -\frac{1}{B(z)A(z)}\left[B(z)C(z)\right]_k$$

$$= -\frac{z^k}{B(z)A(z)}\left[\frac{B(z)C(z)}{z^k}\right]_+ .\ \underline{4/}$$

If $B(z) = N(z)/D(z)$, then the corollary above gives the order of

$$\frac{D(z)}{N(z)C(z)}\left[\frac{C(z)N(z)}{D(z)z^k}\right]_+ = \frac{N_k(z)}{C(z)N(z)},$$

where the order of $N_k(\ )$ is $\max(c + n - k,\ d - 1,\ 0)$. Thus one can easily determine the order of $\gamma(\ )$.

If instead of (4) we wish to minimize the sum of several such terms, e.g.,

$$E\left\{\sum_{i=1}^{\rho} \lambda_i[(A_i(L)\gamma(L)+C_i(L))x_t]^2\right\},$$

then the solution is

$$\text{(7)} \qquad \gamma(z) = -\frac{z^k}{A(z)\Sigma(z)}\left[\frac{B(z)\sum_{i=1}^{\rho}\lambda_i C_i(z)A_i(z^{-1})}{A(z^{-1})z^k}\right]_+ ,$$

where $A(z)A(z^{-1}) = \sum_{i=1}^{\rho}\lambda_i A_i(z)A_i(z^{-1})$ and $A(z)$ has its roots outside the unit circle.

As an example consider the case of a firm which produces to stock; i.e., holds inventories. Assume that the firm's costs in period t are given by

$$\text{(8)} \qquad C_t = \lambda_1(P_t - P_{t-1})^2 + \lambda_2(I_t - \alpha S_t)^2$$

where

P_t = production in period t

S_t = sales in period t

I_t = inventories at the end of period t

and

$$P_t = S_t + I_t - I_{t-1}$$ [5]

We also assume that the firm must choose the level of production for period t before the amount of sales for that period is known. Given the accounting identity between production, sales and the change in inventories it suffices to determine either inventory holdings or the rate of production. While formally it makes little difference, it seems more natural to assume that it is the rate of production which is decided upon rather than the level of inventories. This makes any discrepancy between expected and actual sales show up as unplanned inventory accumulation rather than in unanticipated fluctuations in the rate of production.

So let

$$\text{(9)} \qquad P_t = \gamma(L)S_t = \sum_{i=1}^{\infty}\gamma_i S_{t-i}.$$

The expected costs for period t are given by

$$\text{(10)} \qquad E(C_t) = E\left\{\lambda_1(P_t - P_{t-1})^2 + \lambda_2(I_t - \alpha S_t)^2\right\}.$$

If the firm behaves so as to minimize

$V = E\left\{\sum_{v=0}^{\infty}\rho^v C_{t+v}\right\}$, the problem is

$$\text{(11)} \qquad \min_{\{P_{t+v}\}} E\left\{\frac{\lambda_1(P_t - P_{t-1})^2 + \lambda_2(I_t - \alpha S_t)^2}{1 - \rho L^{-1}}\right\}$$

where $P_t = \gamma(L)S_t$.

Consider, first, the case in which the firm acts to minimize costs in period t. Assuming that the current level of inventories is simply the sum of past differences between production and sales (i.e., initial inventories I_0 equals zero), the problem reduces to

$$\min_{\gamma} E\left\{\lambda_1((1-L)\gamma(L)S_t)^2 + \lambda_2\left[\left(\frac{\gamma(L)-1}{1-\delta L} - \alpha\right)S_t\right]^2\right\} \quad |\delta| < 1$$ [6]

This problem is clearly of the type just discussed with

$$A_1(L) = 1 - L$$

$$C_1(L) = 0$$

$$A_2(L) = \frac{1}{1 - \delta L}$$

$$C_2(L) = -\frac{1 + \alpha(1 - \delta L)}{1 - \delta L}$$

The solution is given by

(12) $$\gamma(z) = \frac{z(1-\delta z)}{R(z)B(z)}\left[\frac{\lambda_2\left(1+\alpha(1-\delta z)\right)B(z)}{(1-\delta z)R(z^{-1})z}\right]_+$$

where $$R(z)R(z^{-1}) = \lambda_1(1-z)(1-\delta z)(1-z^{-1})(1-\delta z^{-1}) + \lambda_2$$

$$g_{SS}(z) = \sigma^2 B(z)B(z^{-1}).$$

If the sales series is a mixed autoregressive, moving average process, say

$$S_t = \frac{N(L)}{D(L)}\varepsilon_t$$

where N( ) and D( ) have orders n and d, respectively, then it is easily seen that

(13) $$\gamma(z) = \frac{zP(z)}{N(z)R(z)}$$

where P( ) is of order max (n, d, 0). If in the structural equation (1) x_t^* is the planned level of production for time period t, then as in the previous examples the estimating equation is a rational lag distribution.7/

If the firm attempts to minimize V rather than expected cost in period t (or the average cost per period), then because of the form of the objective function, the certainty equivalence principle applies. That is, the one may choose future levels of production to minimize V in which all unknown future variables (in this problem the S_{t+j}'s) are replaced by their conditional expectations as of time t - 1. Thus assuming normality or alternatively restricting ourselves to linear forecasting rules, the future sales may be replaced by their least squares forecasts and P_{t+j} chosen to minimize

$$\hat{V} = \sum_{i=0}^{\infty} \rho^v \left[\lambda_1(P_{t+v}-P_{t+v-1})^2 + \lambda_2(I_{t+v} - \alpha\hat{S}_{t+v,\ t-1})\right]$$

subject to: $P_{t+v} = \hat{S}_{t+v,t-k} + I_{t+v} - I_{t+v-1}$.

The solution (see Appendix or [2] for details) will obviously be a distributed lag between production and expected future sales. If these latter variables are expressed as linear combinations of current and past sales, one ends up with a distributed lag between current (and past) production and the current and lagged values of sales.

FOOTNOTES FOR NOTE ONE

1. The results presented can be extended to processes whose p^{th} differences are as stated, but the less general case is assumed for ease of exposition. See [9, Ch. 8, esp. pp. 92-96]. Also, without the additional assumption that the processes are Gaussian, the solutions presented need to be interpreted as optimal only in the class of linear rules.

2. Thomas McCoy has pointed out that certain kinds of coefficient restrictions can reduce the order of $N_v(\ )$. Without a priori knowledge of such restrictions, one would have to allow for lags of the order indicated.

3. [9, equation 3.7.2] gives $\gamma(z) = \frac{1}{B(z)}\left[g_{yx}(z)/B(z^{-1})\right]_+$ where $g_{xx}(z) = \sigma^2 B(z)B(z^{-1})$ and B(z) has all its zeroes outside the unit.

4. See [9, pp. 118-122, esp. equation 10.5.11].

5. This problem is similar to, but simpler than, those treated by [1], [3], and [4]. It is presented as an example for expository purposes only and is not intended to be very realistic. One could allow for deterministic components in the series and add linear terms to the cost function without adding any essential complications. See [9, section 10.6].

6. For convenience we have added the constant δ. Once the solution is obtained we consider the limiting case when δ tends to unity (from below).

7. Other examples are given in [7].

REFERENCES FOR NOTE ONE

[1] Childs, Gerald L.: Unfilled Orders and Inventories. Amsterdam: North Holland Publishing Company, 1967.

[2] Grether, David M.: "Distributed Lags, Prediction, and Signal Extraction," Cowles Foundation Discussion Paper No. 279, Yale University, 1969.

[3] Hay, G. A.: "Production, Price, and Inventory Control," American Economic Review, 60, (1970), 531-545.

[4] Holt, C. C., F. Modigliani, J. F. Muth, and H. Simon: Planning Production, Inventories, and Work Force. Englewood Cliffs, New Jersey: Prentice Hall, 1960.

[5] Jorgenson, D. W.: "Rational Distributed Lag Functions," Econometrica, 34 (1966).

[6] Nerlove, Marc: "Distributed Lags and Unobserved Components of Economic Time Series," pp. 127-169 in W. Fellner, et al., Ten Economic Studies in the Tradition of Irving Fisher. New York: John Wiley, 1967.

[7] __________: "Lags in Economic Behavior," Econometrica, 40 (1972), 221-251.

[8] Theil, H.: Optimal Decision Rules for Government and Industry. Amsterdam: North Holland Publishing Company, 1964.

[9] Whittle, P.: Prediction and Regulation by Linear Least Squares Methods. London: The English Universities Press, 1963.

APPENDIX TO NOTE ONE

Assume that a firm's costs in period t are given by

(A.1) $$C_t = \lambda_1(\hat{P}_t - \hat{P}_{t-1})^2 + \lambda_2(\hat{I}_t - a - \alpha\hat{S}_t)^2 + \lambda_3\hat{P}_t + \lambda_4\hat{I}_t$$

where

$\hat{P}_t$ = production in period t

$\hat{S}_t$ = sales in period t

$\hat{I}_t$ = inventories at the end of the t^{th} period

$\hat{P}_t = \hat{S}_t + \hat{I}_t - \hat{I}_{t-1}$

At time t - 1 the firm is assumed to choose the level of production for period t in order to minimize

(A.2) $$V = E\{\sum_{v=0}^{\infty} \rho^v C_{t+v}\} \qquad 0 \leq \rho \leq 1.$$

Let

$$\hat{S}_t = \overline{S} + S_t \qquad E(S_t) = 0 \qquad \overline{S} \text{ a constant.}$$

(A.3) $$\hat{I}_t = \overline{I} + I_t$$
$$\hat{P}_t = \overline{P} + P_t, \qquad \overline{P} = \overline{S}.$$

Then (A.2) may be written as

(A.4) $$V = E\{\sum_{v=0}^{\infty} \rho^v(\lambda_1(P_{t+v} - P_{t-1+v})^2 + \lambda_2(I_{t+v} - \alpha S_{t+v})^2\}$$
$$+ \sum_{v=0}^{\infty} \rho^v(\lambda_3\overline{P} + \lambda_4\overline{I} + \lambda_2(\overline{I} - a - \alpha\overline{S})).$$

The second term is easily minimized, so the problem reduces to

$$\min E\{\sum_{v=0}^{\infty} \rho(\lambda(P_{t+v} - P_{t+v-1})^2 + (1-\lambda)(I_{t+v} - \alpha S_{t+v})^2\}$$

where $P_t = S_t + I_t - I_{t-1}$ and all series have mean zero. Substituting

$$I_t = \frac{P_t - S_t}{1 - L} \qquad \text{(A.5)}$$

into (A. 4) the problem becomes

$$\min_{\text{wrt. } \{P_{t+v}\}} E\{\sum_{v=0}^{\infty} \rho^v(\lambda(P_{t+v} - P_{t+v-1})^2 + (1-\lambda)\left(\frac{P_{t+v} - S_{t+v}}{1 - L} - \alpha S_{t+v}\right)^2\} .$$

$$0 \leq \lambda \leq 1.$$

Differentiating the objective function with respect to P_{t+v} gives the following first order conditions:

$$\frac{(\lambda(1-L)^2(1-\rho L^{-1})^2 + (1-\lambda))P_{t+v} - (1-\lambda)(1+\alpha(1-L)\hat{S}_{t+v,\,t-1})}{(1-L)(1-\rho L^{-1})} = 0 \qquad \text{(A.6)}$$

$$v = 0, 1, 2, \ldots$$

In (A.6) S_{t+v} has been replaced by its least squares forecast $\hat{S}_{t+v,\,t-1}$.

Now

$$\lambda(1-z)^2(1-\rho z^{-1})^2 + (1-\lambda) = \lambda z^{-2}(z-\beta)(z-\bar{\beta})(z - \frac{\rho}{\beta})(z - \frac{\rho}{\bar{\beta}}) \qquad \text{(A.7)}$$

$$= \lambda|\beta|^2(1 - \frac{1}{\beta}z)(1 - \frac{1}{\bar{\beta}}z)(1 - \frac{\rho}{\beta}z^{-1})(1 - \frac{\rho}{\bar{\beta}}z^{-1})$$

$$= R(z)Q(z^{-1})$$

where

$$R(z) = \sqrt{\lambda}|\beta|(1 - \frac{1}{\beta}z)(1 - \frac{1}{\bar{\beta}}z)$$

$$Q(z^{-1}) = \sqrt{\lambda}|\beta|(1 - \frac{\rho}{\beta}z^{-1})(1 - \frac{\rho}{\bar{\beta}}z^{-1}) .$$

By inspection it is seen that if β is a root of (A.7), then $\frac{\rho}{\beta}$ is also a root. Next it will be shown that there is always a root β with $|\beta| > 1$.

If z_0 is a root of (A.7) then ($\lambda \neq 0$),

$$(1 - z_0)(1 - \rho z_0^{-1}) = \pm i\sqrt{\frac{1-\lambda}{\lambda}} \qquad \text{(A.8)}$$

or

$$z = \frac{\pm\ 1 + \rho - i\sqrt{\frac{1-\lambda}{\lambda}} \pm \sqrt{(1 + \rho - i\sqrt{\frac{1-\lambda}{\lambda}})^2 - 4\rho}}{2} .$$

From (A.7) it is easily seen that for any λ in (0, 1) there must be at least one root outside the unit circle if $\rho = 0$ or if $\rho = 1$. Further, it is clear that for any value of ρ in (0, 1) there can be no roots on the unit circle. But for any fixed λ the roots are bounded continuous functions of ρ, so there must be at least one root outside the unit circle, i.e., $|\beta| > 1$.

The preceding argument implies that (A.6) may be written as:

$$\frac{R(L)Q(L^{-1})P_{t+v} - (1-\lambda)(1 + \alpha(1-L))\hat{S}_{t+v,\,t-1}}{(1-L)(1-\rho L^{-1})} = 0 \qquad \text{(A.9)}$$

$$v = 0, 1, 2, \ldots$$

where the roots of $R(z)$ lie outside the unit circle, and the roots of $Q(z^{-1})$ lie inside the unit circle.

Now

$$\text{(A.10)}\quad \frac{R(L)O(L^{-1})P_{t+v} - (1-\lambda)(1+\alpha(1-L))\hat{S}_{t+v,t-1}}{(1-L)(1-\rho L^{-1})}$$

$$= \frac{R(L)O(L^{-1})P_{t+v} - (1-\lambda)(1+\alpha(1-L))\hat{S}_{t+v,t-1}}{1-L}$$

$$+ \rho\left[\frac{R(z)O(L^{-1})P_{t+v+1} - (1-\lambda)(1+\alpha(1-L))\hat{S}_{t+v+1,t-1}}{(1-L)(1-\rho L^{-1})}\right].$$

But the second term is zero so one may write

$$\text{(A.11)}\quad \frac{R(L)P_{t+v}}{1-L} - \frac{(1-\lambda)(1+\alpha(1-L))\tilde{S}_{t+v,t-1}}{1-L} = 0$$

where $\tilde{S}_{t+v,t-1} = \frac{\hat{S}_{t+v,t-1}}{Q(L^{-1})}$ is a weighted average of all sales expected in future periods.

Now

$$\text{(A.12)}\quad \frac{(1+\alpha(1-L))\tilde{S}_{t+v,t-1}}{1-L} = \text{const.} + (1+\alpha)\tilde{S}_{t+v,t-1}$$

$$+ \tilde{S}_{t+v-1,t-1} + \tilde{S}_{t+v-2,t-1} + \cdots$$

and

$$\frac{(1+\alpha(1+L))\tilde{S}_{t+v-1,t-2}}{1-L} = \text{const.} + (1+\alpha)\tilde{S}_{t+v-1,t-2}$$

$$+ \tilde{S}_{t+v-2,t-2} + \tilde{S}_{t+v-3,t-2} + \cdots$$

so that

$$\text{(A.13)}\quad R(L)P_{t+v} = (1-\lambda)\Big[(1+\alpha)\tilde{S}_{t+v,t-1} - \alpha\tilde{S}_{t+v-1,t-2}$$

$$+ \sum_{j=1}^{\infty}\left(\tilde{S}_{t+v-1-j,t-1} - \tilde{S}_{t+v-1-j,t-2}\right)\Big]$$

The last steps in the derivation of (A.13) assumed that the firm has been running optimally in previous periods. If one were to estimate the parameters of such a model one would typically have to use data on firms (or more likely industries) over a relatively short period of time. In general there will be no reason to assume that the optimization began with the first observation the investigator has; rather it seems more natural to assume a long (but unobserved) history of optimal running and that the initial conditions are sufficiently far in the past so that their influence on the observed behavior is negligible.

To gain some understanding of the preceding equation, consider the case in which changing the level of production is costless. If $\lambda = 0$, $R(L)O(L^{-1}) = 1$ so $\tilde{S}_{t+v,t-1} = \hat{S}_{t+v,t-1}$ and the solution is

$$\text{(A.14)}\quad P_t = \alpha(\hat{S}_{t,t-1} - \hat{S}_{t-1,t-2}) + \hat{S}_{t,t-1} + (S_{t-1} - \hat{S}_{t-1,t-2}).$$

In this case production is simply set equal to expected sales plus the expected change in the desired level of inventories plus a correction factor for the error in forecasting the sales of the current period.

If one assumes that $S_t = B(L)\varepsilon_t$; that is, S_t is a non-deterministic covariance stationary time series,

$$R(L)P_t = \gamma(L)S_{t-1}$$

(A.15) where

$$\gamma(z) = \lim_{\delta \to 1} \frac{1 - \delta z}{B(z)} \left[\frac{B(z)(1 - \lambda)(1 + \alpha(1 - \delta z))}{(1 - \delta z) Q(z^{-1}) z} \right]_+ .$$

<u>Proof:</u>

(A.16) $$\frac{1 - \delta z}{B(z)} \left[\frac{B(z)(1 - \lambda)(1 + \alpha(1 - \delta z))}{(1 - \delta z) Q(z^{-1}) z} \right]_+ =$$

$$\frac{(1 - \lambda)\alpha(1 - \delta z)}{B(z)} \left[\frac{B(z)}{Q(z^{-1}) z} \right]_+ + \frac{(1 - \lambda)(1 - \delta z)}{B(z)} \left[\frac{B(z)}{(1 - \delta z) Q(z^{-1}) z} \right]_+ .$$

Now $\frac{1}{B(z)} \left[\frac{B(z)}{Q(z^{-1}) z} \right]_+$ is the generating function of the operator that estimates $\frac{S_t}{Q(L^{-1})}$. Thus as δ goes to one the first term will give $(1 - \lambda)\alpha(\tilde{S}_{t,t-1} - \tilde{S}_{t-1,t-2})$. Applying Whittle's theorem to the second term gives

(A.17) $$\frac{(1 - \lambda)(1 - \delta z)}{B(z)} \left[\frac{B(z)}{(1 - \delta z) Q(z^{-1}) z} \right]_+ = \frac{(1 - \lambda)}{B(z)} \left[\frac{B(z)}{Q(z^{-1}) z} \right]_+ + (1 - \lambda) \frac{\left(\left[\frac{B(z)}{Q(z^{-1}) z} \right]_- \right) \Big|_{z = \frac{1}{\delta}}}{B(z)} .$$

The first term clearly gives $(1 - \lambda)\tilde{S}_{t,t-1}$ so all that remains to be shown is that the second term leads to $\sum_j (\tilde{S}_{t-j-1,t-1} - \tilde{S}_{t-j,t-2})$.

Now

(A.18) $$\tilde{S}_{t-j-1,t-1} = \sum_{k=0}^{\infty} q_k \hat{S}_{t-j-1+k,t-1} \quad \text{where}$$

$$\frac{1}{Q(z^{-1})} = \sum_{k=0}^{\infty} q_k z^{-k}$$

and

(A.19) $$\frac{1}{B(z)} \left[\frac{B(z)}{z^{k-j}} \right]_+ - \frac{z}{B(z)} \left[\frac{B(z)}{z^{k-j+1}} \right]_+ = \begin{cases} 0 & k < j \\ \frac{b_{k-j}}{B(z)} & k \geq j . \end{cases}$$

So the generating function for

(A.20) $$\sum_{j=0}^{\infty} (\tilde{S}_{t-j-1,t-1} - \tilde{S}_{t-j-1,t-2}) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} q_k (\hat{S}_{t-j+k-1,t-1} - \hat{S}_{t-j+k-1,t-2})$$

is given by

(A.21) $$\frac{1}{B(z)} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} q_{k+j} b_k = \frac{\left[\frac{B(z)}{Q(z^{-1}) z} \right]_- \Big| z = 1}{B(z)} .$$

Now suppose that S_t is a moving average, autoregressive process, e.g. let

(A.22) $$B(z) = \frac{C(z)}{D(z)}$$

where C( ) and D( ) are polynomials of order c and d respectively. Then

(A.23) $$\frac{\gamma(z)}{1 - \lambda} = \lim_{\delta \to 1} \frac{(1 - \delta z) D(z)}{C(z)} \left[\frac{C(z)(1 + \alpha(1 - \delta z))}{D(z) Q(z^{-1}) z (1 - \delta z)} \right]_+$$

Letting

$$S(z) = \frac{C(z)(1 + \alpha(1 - \delta z))}{Q(z^{-1})z} \qquad \text{(A.24)}$$

It is not hard to show that

$$\frac{\gamma(z)}{1 - \lambda} = \frac{[S(z)]_+}{C(z)} + \frac{T(z)}{C(z)} \text{ where } T(z) \text{ is of order d.} \qquad \text{(A.25)}$$

$[S(z)]_+$ is obviously of order c so

$$\gamma(z) = (1 - \lambda)\frac{V(z)}{C(z)} \text{ where } V(z) \text{ is of order } \max[c, d, 0]. \qquad \text{(A.26)}$$

This gives

$$R(L)C(L)P_t = V(L)S_{t-1} \text{ a rational lag distribution.} \qquad \text{(A.27)}$$

NOTE TWO: ESTIMATION

Econometricians are well aware of the difficulties inherent in estimating distributed lag relations. Insufficient degrees of freedom and the often associated problem of multicollinearity make precise estimation of arbitrary lag distributions an essentially hopeless task. A common response to these problems is to use an approximation which depends upon a relatively small number of parameters; e.g., a rational distributed lag model. Though this type of approximation reduces the problem to one involving a finite number of parameters, several serious difficulties remain. For one thing, the parameters of the distribution typically appear in a highly nonlinear fashion. Though rational lag distributions may be transformed so that the parameters enter in a more linear way, this transformation will almost surely result in disturbances which are serially correlated. An additional problem is that the orders of the polynomials in the lag operator are often not known *a priori*. Quite apart from computational difficulties, this latter problem can result in identification problems if the disturbances in the equation to be estimated are themselves autocorrelated.

In the previous note we have shown that in a variety of situations one can derive lag distributions that are rational. This does not alter the computational problems due to the nonlinearities or serially correlated disturbances, but it does provide information about the order of the lag operators. In all cases, the order of the lag operators depends upon the covariance structure of the exogenous variables and the particular decision problem, e.g., the forecast period. Since data on the exogenous variable are available, one could determine an appropriate parametric model for this time series using the methods discussed in, for example, Box and Jenkins [1]. If one is willing to assume that the economic agent correctly perceives the structure of the time series, then this allows one to make use of the parametric model in estimating the structural relation.

If the situation is one of least squares forecasting or extraction with the forecast interval known, then there are two obvious alternatives. First, one could estimate simultaneously the parameters of the structural equation and the parameters of the auto-regressive moving average representation of the exogenous variable. Alternatively, one could adopt computationally simpler but less efficient procedures which should produce at least consistent estimates of all the parameters. The example below is intended to illustrate these approaches.

Suppose the structural equation is

(1) $$y_t = a + bx_t^* + z_t'c + u_t \qquad t = 1, 2, \ldots, T$$

where x_t^* is a least squares forecast or extraction based upon observations of an exogenous variable x_t, z_t is a vector of other exogenous variables, and $\{u_t\}$ are independent normally distributed random variables with mean zero and variance σ_u^2. We assume that preliminary analysis has "identified" the model for the exogenous variable and that initial estimates of the parameters of that model have been obtained. For purposes of illustration only assume that the exogenous variable is a second order auto-regression.

(2) $$x_t = \rho x_{t-1} + \delta x_{t-2} + \epsilon_t \qquad \{\epsilon_t\} \text{ independent } N(0, \sigma_\epsilon^2).$$

We take x_t^* to be the predicted value of x_{t+k} (or the predicted average, total, etc. over some future period) based upon information up to time t. Thus in this case we have

(3) $$x_t^* = \theta x_t + \phi x_{t-1}$$

where $\theta = \theta(\rho, \delta)$
and $\phi = \phi(\rho, \delta)$

are in general nonlinear functions of ρ and δ, the exact forms depending upon the length of the forecast interval. Notice that the assumption that x_t is an auto-regressive process makes the length of the lag distribution in (3) independent of the prediction period though, of course, the coefficients vary. We assume that the interval of prediction is known so that the functions $\theta(\ ,\ )$ and $\phi(\ ,\ )$ are known <u>a priori</u>.

Under these assumptions the log likelihood function is given by

(4) $$\begin{aligned} \ln L = {} & \left[-\frac{T}{2}\ln\sigma_u^2 - \frac{1}{2\sigma_u^2}\sum_t (y_t - b\theta x_t - b\phi x_{t-1} - z_t'c)^2\right] \\ & + \left[-\frac{T}{2}\ln\sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2}\sum_t (x_t - \rho x_{t-1} - \delta x_{t-2})^2\right] \\ = {} & \ln L_1 + \ln L_2 . \end{aligned}$$

For notational convenience we assume that the observation periods for all series are identical and we treat the initial values x_0, x_1, as constants. Also we have suppressed the intercept term a, and take all variables as deviations from population means.

Treating the two terms in the likelihood function separately allows for consistent estimates of ρ, δ, σ_ϵ^2, $b\theta$, $b\phi$, c, and σ_u^2. The parameter b is unidentified in this approach, but from the estimates of

ρ and δ and the <u>a priori</u> knowledge of θ and ϕ, b can also be consistently estimated. Of course, in finite samples $\widehat{b\theta}/\theta(\hat{\rho},\hat{\delta})$ and $\widehat{b\phi}/\phi(\hat{\rho},\hat{\delta})$ will not be equal so that this approach leads to multiple (consistent) estimates of b. This difficulty is easily resolved by noting that

$$(5)\quad \ln L_1 = -\frac{T}{2}\ln\sigma_u^2 - \frac{1}{2\sigma_u^2}\sum_t\left(y_t - b(\theta(\rho,\delta)x_t + \phi(\rho,\delta)x_{t-1}) + z_t'c\right)^2$$

$$= -\frac{T}{2}\ln\sigma_u^2 - \frac{1}{2\sigma_u^2}\sum_t\left(y_t - bx_t^*(\rho,\delta) - z_t'c\right)^2$$

Using the estimates of ρ and δ obtained from maximizing $\ln L_2$, the remaining parameters can be estimated by maximizing $\ln L_1$ using the estimated values of ρ and δ in computing $\hat{x}_t^*$.

The above procedure while producing consistent estimates of all parameters is somewhat unsatisfactory on two grounds. First, the estimates are likely to be inefficient as none of the information in $\ln L_1$ relevant to the determination of ρ and δ has been used. In general,

$$-\frac{1}{T}\frac{\partial^2\ln L}{\partial b\,\partial\rho} \quad\text{and}\quad -\frac{1}{T}\frac{\partial^2\ln L}{\partial b\,\partial\delta}$$

do not converge to zero, so the information matrix is not block diagonal which implies that estimate of b is also inefficient. The second problem is more a matter of practice than of principle, viz.: in practice the standard errors used in making inferences about the parameters of the structural equations are calculated <u>conditional</u> upon the values of ρ and δ actually used. Thus they will be incorrect, even in large samples. One could avoid this by estimating the information matrix for the full set of parameters, but this is rarely done except when obtaining maximum likelihood estimates of all the parameters.

Let

$$(6)\quad \begin{pmatrix}\rho_1\\ \delta_1\end{pmatrix} = \begin{bmatrix}\Sigma x_{t-1}^2 & \Sigma x_{t-1}x_{t-2}\\ \Sigma x_{t-1}x_{t-2} & \Sigma x_{t-2}^2\end{bmatrix}^{-1}\begin{bmatrix}\Sigma x_t x_{t-1}\\ \Sigma x_t x_{t-2}\end{bmatrix}$$

be the estimates of ρ, δ obtained by maximizing $\ln L_2$. Also, let $\hat{\theta}_0$ and $\hat{\phi}_0$ be the estimates of θ and ϕ obtained by maximizing $\ln L_1$ <u>given</u> the value of b. The first-order conditions for maximizing ln L are as follows.

$$(7)\quad \begin{aligned}
&\text{(i)} && \frac{\partial\ln L}{\partial b} = \frac{1}{\hat{\sigma}_u^2}\sum_t(y_t - \hat{b}\hat{\theta}x_t - \hat{b}\hat{\phi}x_{t-1} - z_t'\hat{c})^2(\hat{\theta}x_t + \hat{\phi}x_{t-1}) = 0\\
&\text{(ii)} && \frac{\partial\ln L}{\partial c_i} = \frac{1}{2\hat{\sigma}_u^2}\Sigma(y_t - \hat{b}\hat{\theta}x_t - \hat{b}\hat{\phi}x_{t-1} - z_t'\hat{c})z_{ti} = 0 \qquad i = 1,\dots,K\\
&\text{(iii)} && \frac{\partial\ln L}{\partial\sigma_u^2} = -\frac{T}{2}\frac{1}{\hat{\sigma}_u^2} + \frac{1}{2\hat{\sigma}_u^4}\Sigma(y_t - \hat{b}\hat{\theta}x_t - \hat{b}\hat{\phi}x_{t-1})^2 = 0\\
&\text{(iv)} && \frac{\partial\ln L}{\partial\rho} = \frac{1}{\hat{\sigma}_u^2}\Sigma(y_t - \hat{b}\hat{\theta}x_t - \hat{b}\hat{\phi}x_{t-1})(\hat{b}\hat{\theta}_\rho x_t + \hat{b}\hat{\phi}_\rho x_{t-1})\\
&&& \quad + \frac{1}{\hat{\sigma}_\epsilon^2}\Sigma(x_t - \hat{\rho}x_{t-1} + \hat{\delta}x_{t-2})x_{t-1} = 0\\
&\text{(v)} && \frac{\partial\ln L}{\partial\delta} = \frac{1}{\hat{\sigma}_u^2}\Sigma(y_t - \hat{b}\hat{\theta}x_t - \hat{b}\hat{\phi}x_{t-1} - z_t'\hat{c})(\hat{b}\hat{\theta}_\delta x_t + \hat{b}\hat{\phi}_\delta x_{t-1})\\
&&& \quad + \frac{1}{\hat{\sigma}_\epsilon^2}\Sigma(x_t - \hat{\rho}x_{t-1} - \hat{\delta}x_{t-2})x_{t-2} = 0\\
&\text{(vi)} && \frac{\partial\ln L}{\partial\sigma_\epsilon^2} = -\frac{T}{2}\frac{1}{\hat{\sigma}_\epsilon^2} + \frac{1}{2\hat{\sigma}_\epsilon^4}\Sigma(x_t - \hat{\rho}x_{t-1} - \hat{\delta}x_{t-2})^2 = 0
\end{aligned}$$

Now from the first three equations above, it is clear that b and c are simply estimated by a least squares regression with $x^*(\hat{\rho},\hat{\delta})$ and z_t as independent variables, and that σ_u^2 is estimated from the residual sum of squares in the usual fashion. In order to keep the algebra as simple as possible, we now assume that there are no other exogenous variables in the model.

Expand θ and ϕ about the point ρ_0, δ_0 defined by

$$(8) \quad \begin{cases} \theta_0 = \theta(\rho_0, \delta_0) \\ \phi_0 = \phi(\rho_0, \delta_0) \end{cases}$$

giving

$$(9) \quad \theta(\hat{\rho}, \hat{\delta}) = \theta_0 + \theta_{\rho_0}(\hat{\rho} - \rho_0) + \theta_{\delta_0}(\hat{\delta} - \delta_0) + R_\theta .$$

$$\phi(\hat{\rho}, \hat{\delta}) = \phi_0 + \phi_{\rho_0}(\hat{\rho} - \rho_0) + \phi_{\delta_0}(\hat{\delta} - \delta_0) + R_\phi$$

with partial derivatives evaluated at the point (δ_0, ρ_0). Substituting (9) into (iv) and (v) above gives

$$(10) \quad \frac{\partial \ln L}{\partial \rho} = \frac{1}{\hat{\sigma}_u^2} \Sigma ((\rho_0 - \hat{\rho}) w_t + (\delta_0 - \hat{\delta}) v_t) w_t + \frac{1}{\hat{\sigma}_\varepsilon^2} \Sigma (x_t - \hat{\rho} x_{t-1} - \hat{\delta} x_{t-2}) x_{t-1} + \text{remainder terms} = 0$$

$$\frac{\partial \ln L}{\partial \delta} = \frac{1}{\hat{\sigma}_u^2} \Sigma ((\rho_0 - \hat{\rho}) w_t + (\delta_0 - \hat{\delta}) v_t) v_t + \frac{1}{\hat{\sigma}_\varepsilon^2} \Sigma (x_t - \hat{\rho} x_{t-1} - \hat{\delta} x_{t-2}) x_{t-2} + \text{remainder terms} = 0$$

where

$$w_t = \theta_\rho b x_t + \phi_\rho b x_{t-1}$$

$$v_t = \theta_\delta b x_t + \phi_\delta b x_{t-1}$$

and we have used the fact that

$$(11) \quad \Sigma (y_t - b\theta_0 x_t - b\phi_0 x_{t-1}) w_t = \Sigma (y_t - b\theta_0 x_t - b\phi_0 x_{t-1}) v_t = 0$$

as residuals from least squares regressions are orthogonal to the regressors. Rearranging terms and ignoring the remainder gives

$$(12) \quad \begin{pmatrix} \frac{1}{T} \frac{\partial \ln L}{\partial \rho} \\ \frac{1}{T} \frac{\partial \ln L}{\partial \delta} \end{pmatrix} = \frac{1}{\hat{\sigma}_u^2} \begin{bmatrix} \hat{\sigma}_w^2 & \hat{\sigma}_{wv} \\ \hat{\sigma}_{wv} & \hat{\sigma}_v^2 \end{bmatrix} \begin{pmatrix} \rho_0 - \hat{\rho} \\ \delta_0 - \hat{\delta} \end{pmatrix}$$

$$+ \frac{1}{\hat{\sigma}_\varepsilon^2} \begin{bmatrix} \hat{\sigma}_x^2 & \hat{\sigma}(1)_{xx} \\ \hat{\sigma}(1)_{xx} & \hat{\sigma}_x^2 \end{bmatrix} \begin{pmatrix} \rho_1 - \hat{\rho} \\ \delta_1 - \hat{\delta} \end{pmatrix} = 0$$

Solving for $\hat{\rho}$ and $\hat{\delta}$ we get

$$(13) \quad \begin{pmatrix} \hat{\rho} \\ \hat{\delta} \end{pmatrix} = \left[\frac{1}{\hat{\sigma}_u^2} \begin{pmatrix} \hat{\sigma}_w^2 & \hat{\sigma}_{wv} \\ \hat{\sigma}_{wv} & \hat{\sigma}_v^2 \end{pmatrix} + \frac{1}{\hat{\sigma}_\varepsilon^2} \begin{pmatrix} \hat{\sigma}_x^2 & \hat{\sigma}(1)_{xx} \\ \hat{\sigma}(1)_{xx} & \hat{\sigma}_x^2 \end{pmatrix} \right]^{-1}$$

$$\left[\frac{1}{\hat{\sigma}_u^2} \begin{pmatrix} \hat{\sigma}_w^2 & \hat{\sigma}_{wv} \\ \hat{\sigma}_{wv} & \hat{\sigma}_u^2 \end{pmatrix} \begin{pmatrix} \rho_0 \\ \delta_0 \end{pmatrix} + \frac{1}{\hat{\sigma}_\varepsilon^2} \begin{pmatrix} \hat{\sigma}_x^2 & \hat{\sigma}(1)_{xx} \\ \hat{\sigma}(1)_{xx} & \hat{\sigma}_x^2 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \delta_1 \end{pmatrix} \right]$$

Thus we see that the maximum likelihood estimate of ρ and δ are simply weighted averages of two estimates, one obtained from maximizing $\ln L_2$ with respect to ρ and δ, and the one obtained from maximizing $\ln L_1$ with respect to ρ and δ <u>conditional</u> on a value of b. In fact, equation (13) is simply an example of the standard procedure for combining two estimates to obtain a single more efficient estimator by taking a weighted average of the two estimates the weights being proportional to the precision of the estimates.

The preceding discussion suggests the following procedure:

(1) Obtain estimates of ρ_1, δ_1, and σ_ϵ^2 from the x_t series.

(2) Regress y_t on $\theta(\rho_1, \delta_1)x_t + \phi(\rho_1, \delta_1)x_{t-1}$ to obtain initial estimates of b and σ_u^2.

(3) Regress y_t on x_t and x_{t-1} to obtain estimates of $b\theta_0$ and $b\phi_0$, which combined with the initial estimate of b provides estimates of θ_0 and ϕ_0

(4) Using equation(13) obtain new estimates of ρ and δ.

(5) Repeat step 2 to obtain a new estimate of b.

(6) Using the estimates of b, ρ, and δ, calculate new estimates of σ_ϵ^2 and σ_u^2.

The assumption that x_t is a second order auto-regression has three rather different effects on the preceding discussion. First, it clearly simplifies the discussion as well as the computations in step 1, 3 and 4, though in substance the treatment for auto-regressive moving average processes would be the same. Second, assuming that the order of the auto-regression process is greater than one makes the full model overidentified and thus requires iteration via equation (13). Note that if x_t is known to be a first-order auto-regression, then the model is exactly identified and all the parameters are estimated from steps 1 and 2. Finally, it was assumed that the structure of the x series was known or obtained from analyzing the data on x_t only. In fact, of course, the specifications of the structural equation can also be used to aid in "identifying" the model for x_t. In this case, for instance, including more lagged values of $\{x_t\}$ in the structural equation should not lead to a significant increase in explanatory power.

In general if one wanted a computationally simpler procedure, one could stop with step 2 which should provide consistent estimates of all parameters. Alternatively one could use the model of the x_t series only to determine the order of the lag operators and then estimate the structural equation with or without imposing whatever restrictions there are among the parameters. In this case the estimating equation is of the form

$$(14) \qquad y_t = a + b\gamma(L)x_t + u_t = a + b\frac{N(L)}{D(L)}x_t + u_t$$

so that b is not in general identified without some normalization rule such as $\gamma(1) = \Sigma n_j / \Sigma d_j = 1$. If the series being forecast or the signal being extracted is of the type

$$(15) \qquad (1 - \theta L)^p x_t = z_t, \qquad |\theta| < 1$$

where z_t is a covariance stationary process, then, as Whittle [2] has proved, $\gamma(\theta^{-1}) = \theta^\nu$. So if the p^{th} difference of x is covariance stationary, $\gamma(1)$ will be exactly one. This condition will not necessarily be exactly satisfied for economic time series; hence, one would have to adopt the normalization rule that $\gamma(1) = 1$ in order to determine b uniquely.

If the forecast period is not known, then the first method suggested above is obviously not available. While in this case the order of the lag distribution is not exactly known, note that the order of the denominator of the rational distribution is known and there is a known upper bound for the order of the numerator. Though there is a loss of efficiency, as before it should be possible to obtain at least consistent estimates of the parameters.

If x_t^* is planned production or some similar variable, then it would appear that only the order of the lag distribution (and some nonlinear restrictions on the parameters) can be determined. The difficulty is that in these cases the coefficients of the lag distribution are not determined solely by the auto-covariance function of x_t but also depend upon parameters of say the firm's cost functions. Whether or not all the parameters are identified would depend upon the structure of the entire model. If they are not, one might proceed along the lines of the latter alternative suggested above.

REFERENCES FOR NOTE TWO

[1] Box, G. E. P. and G. M. Jenkins: Time Series Analysis, Forecasting and Control. San Francisco: Holden-Day, 1970.

[2] Whittle, P.: Prediction and Regulation by Linear Least Squares Methods. London: The English Universities Press, 1963.