

## THE ESTIMATION OF DISTRIBUTED LAGS

BY L. R. KLEIN

KOYCK IN his recent book proposes a particular lag scheme for the purpose of studying investment behavior and similar problems in econometrics.<sup>1</sup> A general distributed lag of the form

$$(1) \quad y_t = \sum_{i=0}^{\infty} \alpha_i x_{t-i} + u_t,$$

in which  $y_t$  and  $x_t$  are observable variables of interest to the economist and  $u_t$  is a random disturbance, is clumsy and presents inherent difficulties. In the first place, the right hand sum must be truncated at a finite point allowing sufficient degrees of freedom in the statistical estimates of the parameters. Secondly, intercorrelation among the successive values of  $x_{t-i}$  often imparts a high degree of unreliability to the estimates of the individual parameters  $\alpha_i$ . Sums or other functions of the parameters may be estimated with a fair degree of precision even though individual components are quite unreliable; nevertheless for some problems we may need to use estimates of specific parameters, not the more reliably estimated functions of them. Thirdly, a substantial amount of work may be involved in estimating all the individual coefficients  $\alpha_i$ .

To get round these problems, Koyck proposes a more restrictive type of scheme

$$(2) \quad y_t = a \sum_{i=0}^{\infty} \lambda^i x_{t-i} + u_t \quad (0 \leq \lambda < 1).$$

In this system of lags the coefficients decrease geometrically. Irving Fisher put forward at one time a related scheme in which the coefficients decreased arithmetically.<sup>2</sup> Koyck's scheme has the apparent advantage that it is readily transformed into an equivalent relationship involving only three observable variables. Form the difference between the two equations:

$$\begin{aligned} y_t &= a x_t + a \lambda x_{t-1} + a \lambda^2 x_{t-2} + \dots + u_t, \\ \lambda y_{t-1} &= a \lambda x_{t-1} + a \lambda^2 x_{t-2} + \dots + \lambda u_{t-1} \end{aligned}$$

to get

$$(3) \quad y_t = a x_t + \lambda y_{t-1} + u_t - \lambda u_{t-1}.$$

This is the simplified form that he uses in his study.

<sup>1</sup> L. M. Koyck, *Distributed Lags and Investment Analysis*, (Amsterdam: North-Holland Publishing Co.) 1954.

<sup>2</sup> Irving Fisher, "Note on a Short-Cut Method for Calculating Distributed Lags," *Bulletin de L'Institut International de Statistique*, Vol. XXIX, Part 3, The Hague, 1937, pp. 323-27.

First I shall take up the statistical problem of estimating  $\alpha$  and  $\lambda$  from a sample of observations on  $y_t$  and  $x_t$ . In a final section, I shall go on to a brief general discussion of the suitability of scheme (2) or its derivative (3), for the investigation of particular problems in econometrics. An obvious procedure leading to estimates of  $\alpha$  and  $\lambda$  in (3), is the use of the ordinary method of least squares regression of  $y_t$  on  $x_t$  and  $y_{t-1}$  with  $(u_t - \lambda u_{t-1})$  treated as a composite disturbance term. Lagged values in small samples give rise to least-squares bias,<sup>3</sup> but even in large samples the straightforward least-squares treatment of (3) will, as Koyck has shown, lead to biased estimates. The bias occurs since  $u_{t-1}$ , part of the composite disturbance, is not independent of  $y_{t-1}$ , one of the "independent" variables. In addition, the composite disturbance has an automatic serial correlation even if the  $u_t$  are serially independent, i.e.,  $u_t - \lambda u_{t-1}$  is correlated with  $u_{t-1} - \lambda u_{t-2}$  because both expressions contain a mutual term in  $u_{t-1}$ .<sup>4</sup>

Koyck develops a method for estimating  $\alpha$  and  $\lambda$  without bias but makes it depend on an assumed value for  $\xi$  in

$$(4) \quad u_t = \xi u_{t-1} + e_t.$$

In other words he assumes that the original series  $u_t$  may be autocorrelated, but does not give a method for estimating  $\xi$  from the sample data.

In this paper it is proposed to show that Koyck's consistent estimate is a type of maximum likelihood estimate in the case  $\xi = 0$  (or some other known value) and to suggest a method for estimating  $\alpha$ ,  $\lambda$ , and  $\xi$  simultaneously from the sample data when  $\xi$  is not assumed to be known. In addition, we shall show how the computational steps suggested by Koyck for obtaining a consistent estimate can be simplified.

*Nonautocorrelated disturbances* ( $\xi = 0$ ). Koyck suggests that one first compute ordinary least-squares estimates of  $\alpha$  and  $\lambda$  from the regression of  $y_t$  on  $x_t$  and  $y_{t-1}$ . Call these  $a$  and  $l$ . From this regression compute the sum of squared residuals

$$\sum_{t=1}^T z_t^2 = \sum_{t=1}^T (y_t - a x_t - l y_{t-1})^2.$$

For consistent estimates of  $\alpha$  and  $\lambda$ , he proposes the two equations

$$(5) \quad \begin{aligned} \bar{a} \sum x_t^2 + \bar{\lambda} \sum y_{t-1} x_t &= \sum y_t x_t \\ \bar{a} \sum x_t y_{t-1} + \bar{\lambda} \sum y_{t-1}^2 &= \sum y_t y_{t-1} + \frac{\bar{\lambda} \sum z_t^2}{1 + l \bar{\lambda}} \end{aligned}$$

<sup>3</sup> L. Hurwicz "Least-Squares Bias in Time Series," *Statistical Inference in Dynamic Economic Models*, ed. by T. Koopmans. (N.Y.: John Wiley and Sons). 1950, pp. 365—83.

<sup>4</sup> Transformations of equations in dynamic systems may generally introduce serial correlation. See in this connection L. Hurwicz, "Stochastic Models of Economic Fluctuations," *Econometrica*, Vol. 12, 1944, pp. 114—24.

If  $\bar{a}$  were eliminated from (5), we would have a quadratic in  $\bar{\lambda}$ .

Another way of looking at equation (3) is the following:

$$(3) \quad (y_t - u_t) = \alpha x_t + \lambda(y_{t-1} - u_{t-1}).$$

In this form, we have classical equation of the linear relation with variables subject to observation error. Since  $u_t$  is, by assumption, a nonautocorrelated series, the two errors are independent. The "true" values, or "systematic" parts of the observed variables are each independent of the errors as required in the classical model as long as  $x_t$  is independent of the errors. This can be immediately seen from

$$E u_t (y_t - u_t) = E u_t \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i}.$$

In short, all the standard assumptions of the classical model are met in the case at hand. With that model it is well known that estimates of the coefficient parameters depend on the ratios of the error variances. In equation (3) this ratio is unity for a long series.

$$E u_t^2 = E u_{t-1}^2.$$

If only two variables are involved, a variance ratio of unity implies an orthogonal regression for estimation of the coefficients. In effect, we have a slight modification of that form of regression since two variables are assumed to be measured with the same error variance and a third with no error.

If the  $u_t$  follow the normal distribution, maximum likelihood estimates of  $\alpha$  and  $\lambda$  from (3) are identical with a form of least-squares estimates.<sup>5</sup> Since the two error variances are equal, we can derive least squares estimates as

$$\sum_{t=1}^T u_t^2 + \sum_{t=1}^T u_{t-1}^2 = \min.$$

$$\text{or} \quad \sum_{t=1}^T (y_t - \alpha x_t - \lambda \eta_{t-1})^2 + \sum_{t=1}^T (y_{t-1} - \eta_{t-1})^2 = \min.,$$

where

$$u_t + \eta_t = y_t.$$

The minimization is carried out with respect to each of the  $\eta_t$  (the "systematic" part of the observed variable),  $\alpha$  and  $\lambda$ . The first order conditions for minimization are

$$(6) \quad y_{t-1} - \eta_{t-1} + \lambda y_t - \lambda \alpha x_t - \lambda^2 \eta_{t-1} = 0 \quad (t = 1, 2, \dots, T),$$

<sup>5</sup> See Appendix. The "maximum likelihood" estimates discussed in the text are not *full* maximum likelihood estimates since some restrictions are ignored in the maximization process.

$$(6) \quad \sum_{t=1}^T y_t x_t - \alpha \sum_{t=1}^T x_t^2 - \lambda \sum_{t=1}^T x_t \eta_{t-1} = 0,$$

$$\sum_{t=1}^T y_t \eta_{t-1} - \alpha \sum_{t=1}^T x_t \eta_{t-1} - \lambda \sum_{t=1}^T \eta_{t-1}^2 = 0.$$

From the first of these three equations, we can express  $\eta_{t-1}$  in terms of  $y_t$ ,  $x_t$ ,  $y_{t-1}$ , and the parameters. Substitution of this expression into the other two equations and rearrangement of terms leads to

$$\lambda^2 \left( \frac{\sum x_t y_{t-1} \sum y_t x_t}{\sum x_t^2} - \sum y_t y_{t-1} \right)$$

$$+ \lambda \left[ \sum y_t^2 - \sum y_{t-1}^2 + \frac{(\sum x_t y_{t-1})^2 - (\sum y_t x_t)^2}{\sum x_t^2} \right]$$

$$+ (\sum y_t y_{t-1} - \frac{\sum x_t y_{t-1} \sum y_t x_t}{\sum x_t^2}) = 0.$$

An estimate of  $\lambda$  is obtained directly by extraction of a root from this quadratic equation, and the other coefficient is estimated from

$$\text{est } \alpha = \frac{-\text{est } \lambda \sum x_t y_{t-1} + \sum x_t y_t}{\sum x_t^2}.$$

The advantage of this computing method over Koyck's is that one does not have to make estimates in two steps, but the values obtained will be exactly the same as his, computed from (5). Without going through the algebraic manipulations we may simply remark that the quadratic in  $\bar{\lambda}$  implied by equation system (5) is exactly the same as that implied by (6).<sup>6</sup> Apart from the computational facility, we may interpret Koyck's consistent estimate as either a generalized least-squares or a maximum likelihood estimate. For given values of  $x_t$ , it can also be interpreted as the orthogonal regression of  $y_t$  on  $y_{t-1}$ .

Another way of looking at the problem, from a general point of view, is through the well known determinantal equation associated with the estimation of the linear structural equation with observational errors. To estimate the coefficients of

<sup>6</sup> To prove this, express  $\bar{\alpha}$  in terms of  $\bar{\lambda}$  in the first equation of (5); substitute into the second and collect terms. At the same time, the ordinary least-squares estimate of  $\lambda$ , written as  $l$  in (5), should be expressed in terms of the moments  $\sum x_t^2$ ,  $\sum x_t y_t$ ,  $\sum x_t y_{t-1}$ ,  $\sum y_{t-1}^2$ , and  $\sum y_t y_{t-1}$ .

$$\sum_{i=0}^n \alpha_i (y_{it} - u_{it}) = 0,$$

we form the equation

$$\det. \|\sum y_{it}y_{jt} - \delta_{ij} \mu \sigma_{ij}\| = 0$$

where  $\delta_{ij}$  is the Kronecker *della* and  $\sigma_{ij}$  is the covariance between  $u_{it}$  and  $u_{jt}$ . The characteristic vector of this equation is the estimate of the coefficients  $\alpha_i$ , but without knowing the ratios among the variances of the different  $u_{it}$ , we cannot find it. In the context of Koyck's equations, the appropriate determinant is

$$\begin{vmatrix} \sum y_t^2 - \mu\sigma^2 & \sum y_t x_t & \sum y_t y_{t-1} \\ \sum y_t x_t & \sum x_t^2 & \sum x_t y_{t-1} \\ \sum y_t y_{t-1} & \sum x_t y_{t-1} & \sum y_{t-1}^2 - \mu\sigma^2 \end{vmatrix} = 0.$$

This defines a quadratic equation in  $\mu\sigma^2$ , where  $\sigma^2$  is the common variance of  $u_t$  and  $u_{t-1}$ , the root of this quadratic and the solution vector of the associated equation system.

$$\begin{array}{rclcl} \sum y_t^2 - \mu\sigma^2 & - \alpha \sum y_t x_t & - \lambda \sum y_t y_{t-1} & = & 0 \\ \sum y_t x_t & - \alpha \sum x_t^2 & - \lambda \sum x_t y_{t-1} & = & 0 \\ \sum y_t y_{t-1} & - \alpha \sum x_t y_{t-1} & - \lambda (\sum y_{t-1}^2 - \mu\sigma^2) & = & 0 \end{array}$$

yield the same estimates as those from either equations (5) or (6). With this method of formulation, however, one can readily extend the principles to any number of variables.

*Autocorrelated disturbances* ( $\xi \neq 0$ ). Koyck suggests that nonzero autocorrelation of disturbances is usual and assumes the more complicated model involving unknown  $\xi$ , although he gives no basis for estimating  $\xi$  from the data.

If we were not confronted with the additional difficulties concerned with distributed lags, we could readily develop a procedure for estimating a linear equation with autocorrelated disturbances.<sup>7</sup> In this case it is possible to derive a polynomial in the unknown autoregressive parameter, alone. For any given value of the autoregressive parameter, it is possible to estimate the other parameters from a linear system. Thus in two steps the nonlinear system of estimation equations can be solved.

The situation combining distributed lags and autocorrelated disturbances is similar but not quite so favorable. As in the simpler case mentioned above, the coefficient parameters can be directly estimated for any given value of the autoregressive parameter. We see this immediately by remarking that the

<sup>7</sup> See L. R. Klein, *A Textbook of Econometrics* (Evanston: Row, Peterson and Co.) 1953, pp. 85—89.

equation to be estimated can be changed into one not involving autocorrelated errors by suitable transformations of variables. Thus if we have as an objective the estimation of (3) and if the disturbances follow the autoregressive law in (4), we find that (3) can be transformed to

$$(7) \quad y_t' = \alpha x_t' + \lambda y_{t-1}' + e_t - \lambda e_{t-1}.$$

The transformations are

$$\begin{aligned} y_t' &= y_t - \xi y_{t-1}, \\ x_t' &= x_t - \xi x_{t-1}. \end{aligned}$$

If the variables are subjected to the same autoregressive transformation as that followed by the disturbances we can derive an equation that has the same form as (3). If  $\xi$  were known a priori, we could make the same computations with the transformed as with the original variables for the estimation of  $\alpha$  and  $\lambda$  in the case treated previously with nonautocorrelated disturbances. The errors  $e_t$  are assumed to be nonautocorrelated.

We can see from the substitution of (4) into (3),

$$y_t = \alpha x_t + \lambda y_{t-1} + \xi u_{t-1} + e_t - \lambda u_{t-1},$$

that the special case of  $\xi = \lambda$  can be handled very simply. The least-squares regression of  $y_t$  on  $x_t$  and  $y_{t-1}$  would give consistent and efficient estimates.

The problem then is to estimate (7) for  $\xi \neq \lambda$  and  $\xi \neq 0$ . Let us assume that  $e_t$  is normally distributed and nonautocorrelated. Maximum likelihood<sup>8</sup> or generalized least-squares estimates will be obtained from

$$\sum_{t=1}^T e_t^2 + \sum_{t=1}^T e_{t-1}^2 = \min.$$

$$\text{or} \quad \sum_{t=1}^T (y_t' - \alpha x_t' - \lambda \eta_{t-1}')^2 + \sum_{t=1}^T (y_{t-1}' - \eta_{t-1}')^2 = \min.,$$

where  $e_t + \eta_t' = y_t'$ .

The minimization is carried out with respect to each of the  $\eta_t'$ ,  $\alpha$ ,  $\lambda$ , and  $\xi$ . Equations of the same form as those in (6) above will be obtained with the only alteration being that primes are placed on all the variables. Let us call such an equation system (6'). In addition there will be one more equation (minimization with respect to  $\xi$ ).

$$(8) \quad \begin{aligned} \sum_{t=1}^T [y_t - \xi y_{t-1} - \alpha (x_t - \xi x_{t-1}) - \lambda \eta_{t-1}'] (a x_{t-1} - y_{t-1}) \\ + \sum_{t=1}^T (y_{t-1} - \xi y_{t-2} - \eta_{t-1}') (-y_{t-2}) = 0. \end{aligned}$$

<sup>8</sup> See Appendix.

Equation (8) together with (6') can be solved for all the unknown parameters jointly. From similar equations developed without taking up the question of distributed lags, it was possible to derive a single polynomial in the autoregressive parameter. The form of (8) and (6') is not correspondingly simple, but a straightforward iteration process can be developed to obtain a solution. First assume a value for  $\xi$  and transform computed moments of the original variables into moments of the primed variables. Using the methods of the previous section ( $\xi = 0$ ), one can solve equation system (6') for estimates of  $a$  and  $\lambda$ . Next, with the first round estimates of  $a$  and  $\lambda$ , solve (8) for  $\xi$ . In doing this step  $\eta'_{t-1}$  will have to be eliminated from (8) by means of the first equation in (6'). Repeat the process using the first round estimate of  $\xi$ , and so on.

*Some comments on the interpretation and significance of distributed lags.* Koyck derives equation (3) from the original distributed lags form in equation (2). If the disturbances in (2) are nonautocorrelated, the disturbances in the equation (3) are autocorrelated. This approach may, however, be reversed. From an equation of the form

$$y_t = a x_t + \lambda y_{t-1} + v_t,$$

with known initial value  $y_0$ , we can derive

$$y_t = a \sum_{i=0}^{t-1} \lambda^i x_{t-i} + \sum_{i=0}^{t-1} \lambda^i v_{t-i} + \lambda^t y_0.$$

Thus if we start with the hypothesis that  $y_t$  is a linear function of  $x_t$  and  $y_{t-1}$ , subject to an additive nonautocorrelated disturbance, the derived equation of distributed lags will be autocorrelated. Between these extreme cases we have cases more general, but more difficult to handle, in which both  $u_t$  (of equation (2)) and  $v_t$  are autocorrelated.

It might be thought that if the whole history of explanatory variables,  $x_t$ , is taken into account there would not be autocorrelation in the residual variation. If this were the case, the simplest methods developed in this paper for  $\xi = 0$  would be immediately applicable. In dealing with cross-section data, the techniques for  $\xi = 0$  would seem to be appropriate. In a cross-section sample, we are concerned with mutual independence of all disturbances and lack of such independence is not, as in a time-series sample, represented by autocorrelation. The type of bias involved in Koyck's problem is, however, present if each unit in the sample is assumed to behave according to the distributed lag scheme of (2). When each individual's equation is transformed into (3), the sample estimate of the relationship should be made using the method developed here for  $\xi = 0$ . This type of model has been discussed by Prais in a study of corporate savings behavior

from a sample of individual company records.<sup>9</sup> He follows Dobrovolsky<sup>10</sup> in making lagged dividends a determinant of either current savings or dividend disbursements and shows how the latter's equation can be interpreted in terms of Koyck's distributed lags.

From an empirical point of view one is probably more likely to find low or negligible serial correlation in the computed residuals of the regression of  $y_t$  on  $x_t$  and  $y_{t-1}$ . The use of  $y_{t-1}$  as an explanatory variable will probably extract most of the serial dependence among the residuals. There is a measure of logical consistency in starting out with an equation of the form

$$y_t = \alpha x_t + \lambda y_{t-1} + v_t,$$

estimating the parameters by conventional methods on the assumption that  $v_t$  is nonautocorrelated, and then deriving the distributed lag scheme from this equation. Some authors have formulated their theories of behavior originally in terms of an equation of this type while others begin with the distributed lag of (2). Koyck, however, is alone in paying proper attention to the properties of disturbances when one passes from one form to the other.

In studies of the consumption function, Brown (with the present writer, Stone and Rowe following his lead) has used lagged consumption as an explanatory variable.<sup>11</sup> Stone and Rowe begin with the straightforward premise of lagged consumption behavior and derive Koyck's distributed lag model. They do not deal with the stochastic properties of the model and consequently do not go into the problems of bias in estimation. Friedman in a recent attempt to construct a theory of the consumption function has put forward a distributed lag scheme similar to Koyck's,<sup>12</sup> but as a continuous instead of a discrete distribution. Friedman obtains parameter estimates from a truncated distribution of lags without transforming to a simpler equation; however, his graphical residuals appear to be serially correlated. Brown's methods of estimation do not build on Koyck's premise,

<sup>9</sup> S. Prais, "Some Problems in the Econometric Analysis of Company Accounts," Paper presented at the 1956 European meeting of the Econometric Society, Aix-en-Provence.

<sup>10</sup> S. P. Dobrovolsky, *Corporate Income Retention, 1915—43* (New York: National Bureau of Economic Research), 1951.

<sup>11</sup> T. M. Brown, "Habit Persistence and Lags in Consumer Behavior," *Econometrica*, Vol. 20, 1952, pp. 355—71. L. R. Klein and A. S. Goldberger, *An Econometric Model of the United States 1929—52* (Amsterdam: North-Holland Pub. Co.) 1955. R. Stone and D. A. Rowe, "Aggregate Consumption and Investment Functions for the Household Sector Considered in the Light of British Experience," *National Ökonomisk Tidskrift*, 1956, pp. 1—32.

<sup>12</sup> M. Friedman, *A Theory of the Consumption Function*, National Bureau of Economic Research (Princeton: Princeton Univ. Press), 1957.

although his equation may be interpreted in this light. Brown estimates his equation, however, as though the disturbances are not serially correlated, and his residuals are not in fact serially correlated.

Intercorrelation among explanatory variables such as  $x_{t-1}$ ,  $x_{t-2}$ ,  $x_{t-3}$ , etc., is, as mentioned earlier, a reason advanced by Koyck for making his distribution of lags depend on only two parameters,  $\alpha$  and  $\lambda$  in

$$\alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i}.$$

While it may frequently be the case that one finds high intercorrelation among different lagged values of the explanatory variable and consequent magnification of sampling errors in individual coefficients, this is perhaps not the best criterion to consider. An  $F$ -test which compares the ratio of variances with and without some entire particular lag scheme would seem to be more appropriate in choosing a system of lags than a separate test of each coefficient individually against some null hypothesis. Koyck's distributed lag has a very special time shape which cannot be justified on purely a priori grounds against other schemes, particularly those which allow more freedom to the estimation of the coefficients of the different values of  $x_{t-i}$ . An advantage, however, of Koyck's model is that it has an infinite time span historically and need not be cut off at an arbitrary finite lag.

For purposes of the present paper, I have tried to advance the understanding of Koyck's model by only one step. In doing so, his implicit assumption that  $x_t$  is an exogenous or predetermined variable has been retained. If there were a pure lag in equation (1), i.e., if this relation took the form

$$y_t = \sum_{i=1}^{\infty} a_i x_{t-i} + u_t$$

by ruling out  $i=0$ , large sample justification for the treatment in this paper could be made. Similarly if  $x_t$  is a purely exogenous variable (independent of  $u_t$  as Koyck assumes) the procedures developed follow quite readily. The estimation of the type of distributed lag considered for systems of simultaneous stochastic equations in which we do not have  $E u_t x_t = 0$  is a separate problem.

*Institute of Statistics, Oxford University*  
*Cowles Foundation, Yale University*

## APPENDIX

MAXIMUM LIKELIHOOD ESTIMATES OF DISTRIBUTED LAGS<sup>13</sup>

In deriving maximum likelihood estimates of parameters in Koyck's distributed lag scheme

$$(2) \quad y_t = \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i} + u_t,$$

it would seem most natural to proceed directly without transforming the equation to (3). Assuming  $\xi = 0$ , the likelihood function is

$$e^L = p(u_1) p(u_2) \dots p(u_T)$$

for a sample of  $T$  observations. If  $u_t$  is normally distributed with mean zero and variance  $\sigma^2$ , the maximization of  $e^L$  is equivalent to maximization of

$$L = -T (\log \sqrt{2\pi} + \log \sigma) - \frac{1}{2\sigma^2} \sum_{t=1}^T u_t^2$$

with respect to  $\alpha$ ,  $\lambda$ , and  $\sigma$ . This, in turn, is equivalent to minimization of

$$S = \sum_{t=1}^T (y_t - \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i})^2$$

with respect to  $\alpha$  and  $\lambda$ . The first order conditions are

$$(9) \quad \sum_{t=1}^T (y_t - \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i}) \sum_{i=0}^{\infty} \lambda^i x_{t-i} = 0,$$

$$\sum_{t=1}^T (y_t - \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i}) \sum_{i=0}^{\infty} i \lambda^{i-1} x_{t-i} = 0.$$

Apart from involving infinite sums, these conditions are intractable nonlinear equations in  $\alpha$  and  $\lambda$ .

In the definition of  $\eta_t$  as the "true" or "systematic" part of  $y_t$ ,

$$y_t = \eta_t + u_t,$$

we see that  $\eta_t$  is also defined as

$$\eta_t = \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i}.$$

From this definition we can write

$$\alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i} = \alpha \sum_{i=0}^{t-1} \lambda^i x_{t-i} + \lambda^t \eta_0.$$

Hence the sum of squares to be minimized is now written as

$$S = \sum_{t=1}^T (y_t - \alpha \sum_{i=0}^{t-1} \lambda^i x_{t-i} - \lambda^t \eta_0)^2.$$

<sup>13</sup> These remarks are the result of provocative queries by R. Radner.

The first order conditions for minimization with respect to  $\alpha$ ,  $\lambda$ , and  $\eta_0$  are

$$\begin{aligned}
 & \sum_{t=1}^T (y_t - \alpha \sum_0^{t-1} \lambda^i x_{t-i} - \lambda^t \eta_0) \sum_0^{t-1} \lambda^i x_{t-i} = 0, \\
 (10) \quad & \sum_{t=1}^T (y_t - \alpha \sum_0^{t-1} \lambda^i x_{t-i} - \lambda^t \eta_0) (\alpha \sum_0^{t-1} \lambda^{i-1} x_{t-i} + t \lambda^{t-1} \eta_0) = 0, \\
 & \sum_{t=1}^T (y_t - \alpha \sum_0^{t-1} \lambda^i x_{t-i} - \lambda^t \eta_0) \lambda^t = 0.
 \end{aligned}$$

These equations are in terms of finite sums but still remain highly nonlinear in the unknown parameters. It might be possible to solve them empirically by iteration although it may be a lengthy process. For an assumed value of  $\lambda$ , the first and third equations are linear in  $\alpha$  and  $\eta_0$ . These two parameters can be estimated in the first round of approximations. With the first round estimates of  $\alpha$  and  $\eta_0$ ,  $\lambda$  can be estimated as a root of the remaining equation—a high order polynomial in  $\lambda$ . With a new estimate of  $\lambda$ , iterations of  $\alpha$  and  $\eta_0$  can proceed again, etc.

The estimates derived in the text by consideration of analogies from the theory of observation error are not maximum likelihood estimates in the sense of these implied by (10). In the first place, the double sum of squares

$$\sum_{t=1}^T u_t^2 + \sum_{t=1}^T u_{t-1}^2$$

would be derived from the likelihood function (joint normal distribution)

$$p(u_1, u_0) p(u_2, u_1) \dots p(u_T, u_{T-1}).$$

If we assume that  $\xi = 0$ , this expression becomes

$$p(u_0) [p(u_1)]^2 [p(u_2)]^2 \dots [p(u_{T-1})]^2 p(u_T);$$

if  $p(u_0) = p(u_T)$  we have the square of the ordinary likelihood function

$$e^L = p(u_1) p(u_2) \dots p(u_T).$$

The problem of end-effects in large samples is not serious; therefore this difference is not of major concern.

The straightforward procedure of minimizing

$$S = \sum_{t=1}^T (y_t - \alpha \sum_{i=0}^{t-1} \lambda^i x_{t-i} - \lambda^t \eta_0)^2$$

can be derived from another formulation which is more closely connected with that used in the text. If we define

$$\eta_t = \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i},$$

the  $\eta_t$  must satisfy the recurrence formula

$$\eta_t = \alpha x_t + \lambda \eta_{t-1}.$$

The full maximum likelihood method is, therefore, equivalent to

$$S = \sum_{t=1}^T u_t^2 = \sum_{t=1}^T (y_t - \eta_t)^2 = \min.$$

subject to

$$\eta_t = ax_t + \lambda\eta_{t-1} \quad (t = 1, 2, \dots, T).$$

If instead of using Lagrange multipliers for minimization subject to restraint, we substitute all the constraints into  $S$ , term by term, we obtain

$$\begin{aligned} S = & (y_1 - ax_1 - \lambda\eta_0)^2 + (y_2 - ax_2 - a\lambda x_1 - \lambda^2\eta_0) \\ & + (y_3 - ax_3 - a\lambda x_2 - a\lambda^2 x_1 - \lambda^3\eta_0)^2 \\ & + \dots + (y_T - ax_T - a\lambda x_{T-1} - a\lambda^2 x_{T-2} - \dots - a\lambda^{T-1} x_1 - \lambda^T \eta_0)^2 \end{aligned}$$

or

$$S = \sum_{t=1}^T (y_t - a \sum_{i=0}^{t-1} \lambda^i x_{t-i} - \lambda^t \eta_0)^2.$$

On the other hand, we did not fully substitute all the constraints when deriving the minimization equation (6) in the text. There we substituted as follows:

$$\begin{aligned} \sum_{t=1}^T u_t^2 + \sum_{t=1}^T u_{t-1}^2 &= \sum_{t=1}^T (y_t - \eta_t)^2 + \sum_{t=1}^T (y_{t-1} - \eta_{t-1})^2 \\ &= \sum_{t=1}^T (y_t - ax_t - \lambda\eta_{t-1})^2 + \sum_{t=1}^T (y_{t-1} - \eta_{t-1})^2 \end{aligned}$$

since

$$\eta_t = ax_t + \lambda\eta_{t-1}.$$

But this does not fully eliminate all the  $\eta_t$  except  $\eta_0$ .<sup>14</sup>

Two separate minimization problems may be formulated:

$$\sum_{t=1}^T (y_t - ax_t - \lambda\eta_{t-1})^2 + \sum_{t=1}^T (y_{t-1} - \eta_{t-1})^2 = \min$$

subject to

$$\eta_t = ax_t + \lambda\eta_{t-1} \quad (t = 1, 2, \dots, T)$$

or

$$\sum_{t=1}^T (y_t - ax_t - \lambda\eta_{t-1})^2 + \sum_{t=1}^T (y_{t-1} - \eta_{t-1})^2 = \min.$$

In the first, whether the method of Lagrange multipliers or direct substitution of restraints is used, the estimation equations are highly nonlinear. In the second, we have maximum likelihood equations not using all the constraints. We might call this a "limited information maximum likelihood" method and it leads to nothing more complicated than a quadratic equation in the present model.

Two observations about the nature of the "limited information maximum likelihood" estimates are revealing. (1) Although we do not have

$$\eta_t = ax_t + \lambda\eta_{t-1}$$

for each time period, we do have

$$\lambda (\eta_t - ax_t - \lambda\eta_{t-1}) = -u_{t-1} - \lambda u_t;$$

<sup>14</sup> It may be remarked that if we had two separate "systematic" variables,  $\eta_t$  and  $\zeta_t$ , not related as are  $\eta_t$  and  $\eta_{t-1}$ , the simple elimination procedure would be valid.

therefore we can say that the restraint is satisfied, on the average. (2) The last two equations of (6) imply

$$\sum_{t=1}^T (u_t - \lambda u_{t-1}) x_t = 0,$$

$$\sum_{t=1}^T (u_t - \lambda u_{t-1}) \eta_{t-1} = 0.$$

By the assumption that  $x_t$  is an exogenous variable it is independent of both  $u_t$  and  $u_{t-1}$ . Since  $y_{t-1}$  is not independent of  $u_{t-1}$ , the method proposed in the text amounts to the definition of  $\eta_{t-1}$  in such a way that it is independent of  $u_t - \lambda u_{t-1}$ .<sup>15</sup> Koyck derives his correction for consistency of estimation from just this point of view.

<sup>15</sup> The first set of equations of (6) assigns values of  $\eta_{t-1}$  that will be independent of  $u_t - \lambda u_{t-1}$ .