COWLES FOUNDATION DISCUSSION PAPER NO. 34

The Estimation of Distributed Lags

L.R. Klein

June 18, 1957

The Estimation of Distributed Lags

Koyck in his recent book proposes a particular lag scheme for the purpose of studying investment behavior and similar problems in econometrics. [1] A

1) L.M. Koyck, Distributed Lags and Investment Analysis, (Amersterdam: North-Holland Publishing Co.) 1954

general distributed lag of the form

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

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 sucessive values of x_{t-i} often impart a high degree of unreliability to the estimates of the individual parameters α_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 α_i.

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

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

In this system of lags the coefficients decrease geometrically. Irving Fisher put forward at one time a related scheme in which the coefficients decreased arithematically. [1)] Koyck's scheme has the apparent advantage

1) 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

that it is readily transformed into an equivalent relationship involving only three observable variables. Form the difference between the two equations:

$$y_t = \alpha x_t + \alpha \lambda x_{t-1} + \alpha \lambda^2 x_{t-2} + \cdots + u_t$$

$$\lambda y_{t-1} = \qquad \alpha \lambda x_{t-1} + \alpha \lambda^2 x_{t-2} + \cdots + u_{t-1}$$

to get

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

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

First I shall take up the statistical problem of estimating α and λ 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 α and λ in (3), is the use of the ordinary method of least squares regression of y_t and 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,[1] but even in large samples

1) 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

the straight-forward least-squares treatment of (3) will, as Koyck 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}.[2]

2) Transformations of equations in dynamic systems may generally intorduce serial correlation. See in this connection L. Hurwicz, "Stochastic Models of Economic Fluctuations," Econometrica, Vol. 12, 1944, pp. 114-24

Koyck develops a method for estimating α and λ without bias but makes it depend on an assumed value for ξ in

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

In other words he assumes that the original series u_t may be autocorrelated, but does not give a method for estimating ξ 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 α, λ, and ξ simultaneously from the sample data when ξ 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 α and λ 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

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

For consistent estimates of α and λ, he proposes the two equations

$$(5)\qquad \begin{aligned} \bar{\alpha}\,\Sigma\,x_t^2 + \bar{\lambda}\,\Sigma\,y_{t-1}x_t &= \Sigma\,y_t x_t \\ \bar{\alpha}\,\Sigma\,x_t y_{t-1} + \bar{\lambda}\,\Sigma\,y_{t-1}^2 &= \Sigma\,y_t y_{t-1} + \frac{\bar{\lambda}\,\Sigma\,z_t^2}{1 + \ell\,\bar{\lambda}} \end{aligned}$$

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

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

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

In this form, we have the 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 α and λ from (3) are identical with a form of least-squares estimates. [1)]

---

1) See appendix. The "maximum likelihood" estimates discussed in the text are not <u>full</u> maximum likelihood estimates since some restrictions are ignored in the maximization process.

Since the two error variances are equal, we can derive least squares estimates as

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

or

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

where

$$u_t + \eta_t = y_t \ .$$

The minimization is carried out with respect to each of the η_t (the "systematic" part of the observed variable), α and λ . The first order conditions for minimization are

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

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

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

From the first of these three equations, we can express η_{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

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

$$+ \left(\Sigma y_t y_{t-1} - \frac{\Sigma y_t x_t\, \Sigma x_t y_{t-1}}{\Sigma x_t^2}\right) = 0$$

An estimate of λ 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).[1] Apart from the computational facility, we may

1) 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 λ, written as ℓ in (5), should be expressed in 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 $\sum y_t y_{t-1}$.

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

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

we form the equation

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

where δ_{ij} is the Kronecker delta and σ_{ij} is the covariance between u_{it} and u_{jt}. The characteristic vector of this equation is the estimate of the coefficients α_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

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

This defines a quadratic equation in $\mu\sigma^2$, where σ^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.

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

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 ($\varepsilon \neq 0$). In equations with lengthy distributions of lags, the provision for autocorrelation in disturbance may be less urgent since the serial effects of the past are already accounted for in the lagged variables.

For this reason, when treating the problem of distributed lags, the case already dealt with ($\xi = 0$) may be the most important. Koyck suggests that nonzero autocorrelation is usual and assumes the more complicated model, although he gives no basis for estimating ξ from the data.

Without taking up the additional difficulties concerned with distributed lags, we can readily develop a procedure for estimating a linear equation with autocorrelated disturbances.[1] In this case it is possible to derive a polynomial

1) See L.R. Klein, A Textbook of Econometrics (Evanston: Row, Peterson and Co.) 1953, pp. 85-89.

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 is similar but not quite so favorable in the present situation combining distributed lags and autocorrelated disturbances. 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 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 the disturbances follow the autoregressive law in (4), we find that (3) can be transformed to

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

The transformations are

$$y'_t = y_t - \xi\, y_{t-1},$$

$$x'_t = x_t - \xi\, x_{t-1} .$$

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 ξ were known a priori, we could make the same computations with the transformed as with the original variables for the estimation of α and λ in the case treated previously with nonautocorrelated disturbances. The errors e_t are assumed to be nonautocorrelated.

We can see from the substutution 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

$$\xi \neq \lambda \quad \text{and} \quad \xi \neq 0 .$$

Let us assume that e_t is normally distributed and nonautocorrelated. Maximum likelihood [1)] or generalized least-squares estimates will be obtained from

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

or

$$\sum_{t=1}^{T} (y'_t - \alpha\, x'_t - \lambda\, \eta'_{t-1})^2 + \sum_{t=1}^{T} (y'_{t-1} - \eta'_{t-1})^2 = \text{min.},$$

where $e_t + \eta'_t = y'_t$.

1) See appendix.

The minimization is carried out with respect to each of the η'_t, α, λ, and ξ. 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 (6'). In addition there will be one more equation (minimization with respect to ξ).

$$(8) \qquad \sum_{t=1}^{T} [y_t - \xi y_{t-1} - \alpha (x_t - \xi x_{t-1}) - \lambda \eta'_{t-1}] (\alpha x_{t-1} - y_{t-1})$$

$$+ \sum_{t=1}^{T} (y_{t-1} - \xi y_{t-2} - \eta'_{t-1}) (- y_{t-2}) = 0 .$$

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 ξ 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 α and λ. Next, with the first round estimates of α and λ, solve (8) for ξ. In doing this step η'_{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 ξ, and so on.

<u>Some comments on the interpretation and significance of distributed lags</u>. It was mentioned above that the case of $\xi = 0$ is of great importance since it is unlikely to find significant autocorrelation in disturbances after the entire history of explanatory variables has been taken into account.

This is partly an empirical observation that computed residuals from equations in which strong lag effects have been estimated are often found not to exhibit marked serial correlation.

This is particularly the case, empirically, if lagged values of the dependent variable are included in the estimated equation. Prais, in a study of corporate savings behavior from individual company records, follows up on a device used by Dobrovolsky in making lagged dividends a determinant of either current savings or dividend disbursements.[1] One can interpret this type of equation directly in

1) 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. S.P. Dobrovolsky, Corporate Income Retention, 1915-43, (New York; National Bureau of Economic Research) 1951.

terms of the tendency towards maintenance of a stable dividend policy or more circuitously in terms of Koyck's scheme of distributed lags from which equation (3) is derived. Prais chooses the latter interpretation. If Koyck's model is to be used, one must estimate a relation between y_t, x_t, and y_{t-1} such that the disturbances have some built-in autocorrelation.[2] This ought to carry over to the

2) Dobrovolsky investigates both time-series and cross-section samples. The remarks in the text above apply to time series samples. Prais, however, is concerned with cross-section samples only, and autocorrelations of disturbances are not the relevant criterion. In a cross-section sample we are concerned with the mutual independence of all the disturbances and a lack of such independence is not, as in a time series sample, represented by autocorrelation. Least squares bias due to lags in small samples (see p. 3 above) is not particularly relevant in cross-section samples. The type of bias involved in Koyck's problem remains, however, and the methods suggested in this paper carry over to estimation in cross-section samples.

computed residuals but will not do so in most cases in which an ordinary least-squares regression of y_t on x_t and y_{t-1} is made. The use of y_{t-1} as an explanatory variable will probably extract most of the serial dependence among the residuals. Using ordinary least-squares methods, Dobrovolsky's simpler assumptions seem to be more logically consistent. On the other hand, the methods of consistent estimation suggested by Koyck and elaborated in this paper provided a technique whereby the assumptions made about the autocorrelation of disturbances may be carried over to the residuals. As mentioned, however, the case with $\xi = 0$ would seem to be the most useful and simplest model to employ.

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.[1)]

1) 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," Nationaløkonomisk Tidskrift, 1956, pp. 1-32.

This procedure could be interpreted in terms of a general theory of hysteresis or again in terms of Koyck's model. 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,[2)] but as a continuous instead of a discrete distribution. Brown's methods

2) M. Friedman, A Theory of the Consumption Function, National Bureau of Economic Research (mimeographed), 1956.

of estimation do not build on Koyck's premise, 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, α and λ in

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

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 general lag scheme would seem to be more appropriate in choosing between two systems 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

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

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.

Appendix

Maximum Likelihood Estimates of Distributed Lags[1)]

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

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

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 σ^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 α , λ , and σ . 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 α and λ . The first order conditions are

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

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

1) These remarks are the result of provocative queries by R. Radner.

Apart from involving infinite sums, these conditions are intractable nonlinear equations in α and λ .

In the definition of η_t as the "true" or "systematic" part of y_t,

$$y_t = \eta_t + u_t ,$$

we see that η_t is also defined as

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

From this definition we can write

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

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

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

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

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

(10) $$\sum_{t=1}^{T} (y_t - \alpha \sum_{0}^{t-1} \lambda^i x_{t-i} - \lambda^t \eta_o) (\alpha \sum_{0}^{t-1} i \lambda^{i-1} x_{t-i} + t\lambda^{t-1} \eta_o) = 0$$

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

These equations are in terms of finite sums but still remain highly nonlinear in the unknown parameters. It would be possible to solve them empiricially by iteration although it may be a lengthy process. For an assumed value of λ, the first and third equations are linear in α and η_o. These two parameters can be estimated in the first round of approximations. With the first round estimates of α and η_o, λ can be estimated as a root of the remaining equation - a high order polynomial in λ. With a new estimate of λ, iterations of α and η_o 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 sence 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)

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

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

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

If

$$p(u_o) = p(u_T)$$

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

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

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 η_t must satisfy the recurrence formula

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

The full maximum likelihood method is, therefore, equivalent to

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

subject to

$$\eta_t = \alpha x_t + \lambda \eta_{t-1}, \qquad t = 1,2,\ldots,T.$$

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

$$S = (y_1 - \alpha x_1 - \lambda \eta_o)^2 + (y_2 - \alpha x_2 - \alpha \lambda x_1 - \lambda^2 \eta_o) + (y_3 - \alpha x_3 - \alpha \lambda x_2 - \alpha \lambda^2 x_1 - \lambda^3 \eta_o)^2$$

$$+ \ldots + (y_T - \alpha x_T - \alpha \lambda x_{T-1} - \alpha \lambda^2 x_{T-2} - \ldots - \alpha \lambda^{T-1} x_1 - \lambda^T \eta_o)^2$$

or

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

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

$$\sum_{t=1}^{T} u_t^2 + \sum_{t=1}^{T} u_{t-1}^2 = \sum_{t=1}^{T} (y_t - \eta_t)^2 + \sum_{t=1}^{T} (y_{t-1} - \eta_{t-1})^2$$

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

since

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

But this does not fully eliminate all the η_t except η_o [1)]

---

1) It may be remarked that if we had two separate "systematic" variables, η_t and ζ_t, not related as are η_t and η_{t-1}, the simple elimination procedure would be valid.

Two separate minimization problems may be formulated.

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

subject to

$$\eta_t = \alpha x_t + \lambda \eta_{t-1} . \qquad t = 1,2,\dots,T$$

or

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

In the first, whether the method of Lagrange multipliers or direct substitution of restraints is used, the estimation equations are highly non-linear. 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

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

for each time period,we do have

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

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 variables 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 η_{t-1} in such a way that it is independent of $u_t - \lambda u_{t-1}$.[1] Koyck derives his correction for consistency of estimation

---

1) The first set of equations of (6) assigns values of η_{t-1} that will be independent of $u_t - \lambda u_{t-1}$.

---

from just this point of view.