THE ECONOMIC
AND SOCIAL
RESEARCH INSTITUTE
MEMORANDUM SERIES
NO.110

The Koyck Transformation

by

R.C. Geary

Confidential: Not to be quoted until the permission of the Author and the Institute is obtained.

The Koyck Transformation

The Koyck transformation is much used in the COMET system of OLS equations in time series. Let the original equation be -

(1) $$y_t = \beta_0 + \beta_1 \sum_{t'=0}^{\infty} \alpha^{t'} x_{t-t'} + \beta_2 z_t + e_t$$

$$\equiv \beta_0 + \beta_1 x_t + \beta_1 \alpha (x_{t-1} + \alpha x_{t-2} + \dots \text{ ad inf.}) + \beta_2 z_t + e_t$$

Write

(2) $$\alpha y_{t-1} = \alpha\beta_0 + \alpha\beta_1 x_{t-1} + \alpha\beta_1 (x_{t-1} + \alpha x_{t-2} + \dots \text{ ad inf.}) + \alpha\beta_2 z_{t-1} + \alpha e_{t-1}$$

Subtract (2) from (1) -

(3) $$y_t = \beta_0 (1 - \alpha) + \beta_1 (x_t - \alpha x_{t-1}) + \beta_2 (z_t - \alpha z_{t-1}) + \alpha y_{t-1} + (e_t - \alpha e_{t-1})$$

This is the transformed version. The object is to eliminate the infinite series from form (1). The transformation will be seen to introduce a lagged depvar, and lagged functions of the indvars, on the RHS. The procedure then is usually to solve (3) by OLS. There are 6 coefficients to be estimated, namely $\beta_0 (1 - \alpha)$, β_1, $-\beta_1 \alpha$ etc but there are only 4 parameters, namely $\alpha, \beta_0, \beta_1, \beta_2$. The validity of (1) as a hypothesis will be adjudged by the consistency of the estimates, allowing for sampling errors.

The formal transformation applies even if one starts with more than one infinite series on the RHS, e.g. -

(5) $$y_t = \beta_0 + \beta_1 \sum_{t'=0}^{\infty} \alpha^{t'} x_{t-t'} + \beta_2 \sum_{t'=0}^{\infty} \gamma^{t'} z_{t-t'} + e_t$$

We have to consider only the effect of the first operation, i.e. substracting αy_{t-1}, on the second Σ on the RHS. This is easily seen to be -

(6) $$\beta_2 z_t + \beta_2 (\gamma - \alpha) (z_{t-1} + \gamma z_{t-2} + \gamma^2 z_{t-3} + \dots \text{ ad inf.}),$$

the last term of which is in "geometric" form. Hence -

(7) $$y_t - \alpha y_{t-1} - \gamma (y_{t-1} - \alpha y_{t-2})$$

$$\equiv y_t - (\alpha + \gamma) y_{t-1} + \alpha\gamma y_{t-2}$$

$$= \beta_0 (1-\alpha)(1-\gamma) + \beta_1 (x_t - \overline{\alpha+\gamma}\, x_{t-1} + \alpha\gamma\, x_{t-2})$$
$$+ \beta_2 (\gamma-\alpha)(z_{t-1} - \gamma z_{t-2}) + (l_t - \overline{\alpha+\gamma}\, l_{t-1} + \alpha\gamma\, l_{t-2})$$

Both Σ's on the RHS have now been formally eliminated. In (5) there are five parameters to be determined. If (7) be regarded as an inconstrained system to be solved by OLS, there are 8 coefficients. Allowing for random error there must be 3 relationships between the coefficients of (7). Of course we could constrain the coefficients ab initio so that the estimates are absolutely in accordance with (7) (or (3)) using Lagrange procedure. We do not consider it worth while pursuing these formal solutions for the following reason.

In the OLS procedure outlined there is the fundamental theoretical objection that no account is taken of the nature of the disturbance l. For form (1) to be meaningful e_t should be assumed to to be regular (i.e. $Ee_t = 0$, $E e_t^2 = \sigma^2$ (same for all t) and $E\, e_t\, e_{t_1} = 0$, $t' \neq t$. But if this be so, and $\alpha \neq 0$, the disturbance in (3) cannot be regular, hence OLS procedure for coefficient estimation is invalid, i.e. it would result in inconsistent estimates of the coefficients:

This objection would, of course, not apply if FIML procedure were adopted for solution assuming disturbances to be normally distributed. This assumption would lead to its own practical difficulties.

To people committed to theoretical consistency who wish to use OLS, the sensible course would be to use an inconstrained version of (1), namely -

$$(1)' \quad y_t = \beta_0 + \beta_1 \sum_{t'=0}^{\infty} \alpha_{t'}\, x_{t-t'} + \beta_2 z_t + e_t,$$

the coefficients β_0, $\beta'_{t'} = \beta_1 \alpha t'$ and β_2 to be estimated by OLS the

β, being indeterminate. Though in (1)' , the Σ is formally to ∞, in practice the α (we assume $o < \alpha < 1$) coefficients tail off very rapidly, so that one or two lagged x terms will suffice. If one's theory commits one to geometrical progression terms as in (1), the α can be estimated as the geometric mean of $\beta'_0, \beta'_1 \ldots$, provided that these form a diminishing sequence, as they are likely to do.

Form (3) in an unconstrained form may also be perfectly sensible, as an initial hypothesis, namely as -

$$(3)' \quad y_t = \beta'_0 + \beta'_1 x_t + \beta''_1 x_{t-1} + \beta'_2 z_t + \beta''_2 z_{t-1} + \gamma . y_{t-1} + e'_t,$$

the coefficients being now absolutely unconstrained, the RHS containing the lagged depvar, also one lagged term (there may, of course, be more) OF each of the indvars. OLS procedure now assumes that the disturbances are normally and indepently distributed. At least the nonautoregression can be tested ex post using DW or tau. One may even find that the disturbances in both forms (1) 'and (3)' are non-autoregressed. Choice of which form to use might depend on the value of $\bar{R}^2$ or s.

Treatment here is in the simplest forms of equations, (1), (1)' etc. Generalisalation is obvious, including generalisation of the conclusions.

R.C. Geary

5 April 1976.