*The Economic Journal*, **98** (*Conference* 1988), 189–205
*Printed in Great Britain*

# DYNAMIC SPECIFICATION, THE LONG-RUN AND THE ESTIMATION OF TRANSFORMED REGRESSION MODELS

*M. R. Wickens and T. S. Breusch*

The purpose of this paper is to re-examine the question of dynamic specification when interest focuses mainly on the long-run properties of a model. In a recent survey of dynamic specification in econometrics, Hendry *et al.* (1984) listed nine different types of models that can be obtained through imposing coefficient restrictions on the general autoregressive-distributed lag equation

$$y_t = \sum_{i=1}^{m} \alpha_i y_{t-i} + \sum_{i=0}^{n} \beta_i x_{t-i} + e_t, \tag{1}$$

where $x_t$ is weakly exogenous for $t = 1, \ldots, T$ and $e_t$ is i.i.d. $(0, \sigma^2)$. The main aim of this typology was to emphasise alternative short-run dynamic structures. Often, however, the principal interest is in the long-run behaviour of a model because this is what economic theory usually has most to say about, and because, as a result, tests of the theory tend to focus on its long-run properties.

In Section I we examine alternative ways of writing a single dynamic equation, paying particular attention to formulations which provide point estimates of the long-run multipliers. We compare these formulations with the currently highly favoured error correction model (ECM) and note a number of disadvantages of using the ECM. These results are then generalised to the case of a multiple equation system and to the long-run structural model.

Prior to discussing the estimation of these models in Section III, we consider in Section II the more general problem of estimating linearly transformed and re-normalised regression equations. These are equations where the variables of the original regression model (including the dependent variable) have been re-combined linearly and the equation re-normalised on the new dependent variable. In such models the new regressors may be a linear function of the original dependent variable and hence be correlated with the new error term, while the new dependent variables could be a function of the original regressors and hence be uncorrelated with the new error term. The coefficients of the new regressors will be a non-linear function of the original regression coefficients. We show that estimates of the new coefficients obtained by substituting OLS estimates of the original coefficients in this nonlinear function are identical to those obtained by estimating the new model using the instrumental variable estimator with the original matrix of regressors as the instrumental variables. These results are used in Section III to obtain an estimator for our reformulated dynamic equations. This estimator is then compared with simultaneous equation estimation methods.

In Sections I–III it is assumed that all of the variables are stationary. Some implications of having non-stationary variables are discussed in Section IV. It is found that consistent estimates of the long-run multipliers of non-stationary variables can be obtained even if all short-run dynamics are omitted. It is shown that there is a connection between this result and that of Engle and Granger (1987) who find that consistent estimates of the co-integrating vector can be obtained by OLS estimates involving the co-integrated series alone. Engle and Granger believe that there is a strong relationship between co-integration theory and the ECM. We argue that this is because they interpret the ECM differently, and that there is not a strong relationship between co-integration theory and the conventional ECM. In fact Engle and Granger's version of the ECM can be interpreted as just another transformation of the original dynamic model. We also argue, therefore, that their two-step estimation procedure is unnecessary and that the estimation methods discussed in this paper are more convenient. In our concluding section we propose an alternative modelling strategy for dynamic equations that will give better estimates of the long run in finite samples.

We do not claim that the results in this paper are always new, but where they are not we would wish to argue that by drawing upon various disparate strands of the existing literature we are able to throw new light on a neglected, though important aspect of dynamic specification. We also hope that it helps clarify some of the implications of the current literature.

## I. ALTERNATIVE DYNAMIC FORMULATIONS

### (a) *Single Equations*

When interest centres on the long-run properties of a dynamic model it will often be preferable to find an alternative way of writing the model to equation (1). The standard approach is, however, to estimate (1) directly and then to estimate the long-run multiplier $\theta$ using

$$\theta = \lambda \sum_{i=0}^{n} \beta_i, \tag{2}$$

where

$$\lambda = 1 \Big/ \left(1 - \sum_{i=1}^{m} \alpha_i\right). \tag{3}$$

Thus additional computations are required to obtain both the estimate of $\theta$ and its variance. A more convenient approach would be to re-write the model in such a way that a point estimate of $\theta$ and its variance can be obtained directly without the need for further calculations. The transform proposed by Bewley (1979) can be modified for this purpose in several different ways. Each has other advantages and these may be used to select a particular version of the transformation.

*Alternative Formulations of Equation* (1)

(i) The first alternative formulation is obtained by subtracting $(\sum_{i=1}^{m}\alpha_i)\, y_t$ from each side of (1), re-arranging the $x$'s and re-normalising to give

$$y_t = -\lambda \sum_{i=1}^{m} \alpha_i (y_t - y_{t-i}) + \lambda \left( \sum_{i=0}^{n} \beta_i \right) x_t - \lambda \sum_{i=1}^{n} \beta_i (x_t - x_{t-i}) + \lambda e_t. \tag{4}$$

Thus the coefficient of $x_t$ is the long-run multiplier $\theta$ defined by equation (2). Estimating (4) directly would therefore give a point estimate of $\theta$ and its standard error. The characteristic of (4) that distinguishes it from other transformations is that the coefficients of $\Delta_i y_t = y_t - y_{t-i}$ and $\Delta_i x_t = x_t - x_{t-i}$ are proportional to the original distributed lag coefficients of (1) and hence the shape of the distributed lag functions of (1) can still be deduced from (4).

(ii) The other formulations are obtained by noting that the terms in $\Delta_i y_t$ and $\Delta_i x_t$ can be combined linearly in an infinite number of ways without affecting the term in $x_t$. The second formulation involves only first differences and is

$$y_t = -\lambda \sum_{i=0}^{m-1} \left( \sum_{j=i+1}^{m} \alpha_j \right) \Delta y_{t-i} + \lambda \left( \sum_{i=0}^{n} \beta_i \right) x_t - \lambda \sum_{i=0}^{n-1} \left( \sum_{j=i+1}^{n} \beta_j \right) \Delta x_{t-i} + \lambda e_t. \tag{5}$$

(iii) A third expression involves terms obtained by differencing $i$ times:

$$y_t = \sum_{i=1}^{m} \phi_i \Delta^i y_t + \theta x_t + \sum_{i=1}^{n} \gamma_i \Delta^i x_t + \lambda e_t, \tag{6}$$

where $\boldsymbol{\phi} = \lambda \mathbf{P}_k \boldsymbol{\alpha}^*$, $\boldsymbol{\gamma} = \lambda \mathbf{P}_k \boldsymbol{\beta}$, $\boldsymbol{\alpha}^{*\prime} = (-1, \boldsymbol{\alpha}')$, $\boldsymbol{\phi}' = (-1, \phi_1, \ldots, \phi_k)$, $\boldsymbol{\gamma}' = (\theta, \gamma_1, \ldots, \gamma_k)$

$$\mathbf{P}_k = \begin{bmatrix} 1 & 1 & 1 & 1 & \ldots & 1 \\ 0 & -1 & -2 & -3 & \ldots & -k \\ \cdot & 0 & 1 & 3 & & \cdot \\ \cdot & & 0 & -1 & & \cdot \\ \cdot & & & 0 & & \cdot \\ 0 & & & & & (-1)^k \end{bmatrix}$$

and $k = \max(m, n)$. If $m \neq n$ then additional zero terms are included in either $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ or $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to make $m = n = k$.

Equation (6) has several uses. It provides estimates of steady-state paths which are not in static equilibrium. For example, if $y$ and $x$ are logarithms and $\Delta y_t = g_y$, $\Delta x_t = g_x$ (non-zero constants) then along the steady-state growth path we have $\Delta^i y_t = \Delta^i x_t = 0$ for $i > 1$ and hence

$$y = (\phi_1 g_y + \gamma_1 g_x) + \theta x. \tag{7}$$

Thus estimating (6) is convenient for evaluating (7).

Another situation where (6) is useful is when the steady state solution for $y_t$ involves both levels and rates of growth of $x_t$. For example, in the long-run demand for money function the logarithm of nominal money balances is often expressed as a function of the steady state rate of inflation as well as the logarithm of the price level. An unrestricted version in the form of (1) might be

$$m_t = \sum_{i=1} \alpha_i m_{t-i} + \sum_{i=0} \beta_i p_{t-i} + \sum_{i=0} c_i y_{t-i} + \sum_{i=0} d_i r_{t-i} + e_t, \tag{8}$$

where $m$ = nominal money, $p$ = prices, $y$ = income (all in logs) and $r$ = the interest rate. The long-run solution is

$$m = \theta p + \phi_1 \Delta p + (\lambda \sum_{i=0} c_i)\, y + (\lambda \sum_{i=0} d_i)\, r, \tag{9}$$

where $\theta$ and $\phi_1$ are as defined previously. Thus re-writing (8) in the form given by (6) would immediately provide estimates of (9). A further use of (6) for forecasting will be discussed in Section V.

(iv) A fourth formulation which can be used to compute the average lag in the response of $y$ to $x$ is

$$y_t = -\lambda \sum_{i=1}^{m} i\alpha_i \Delta y_t + \lambda \sum_{i=2}^{m} \alpha_i (i\Delta y_t - \Delta_i y_t) + \lambda \left( \sum_{i=0}^{n} \beta_i \right) x_t - \lambda \sum_{i=1}^{n} i\beta_i \Delta x_t + \lambda \sum_{i=2}^{n} \beta_i (i\Delta x_t - \Delta_i x_t) + \lambda e_t \tag{10}$$

If $\mathbf{A}(L) = 1 - \sum_{i=1}^{m} \alpha_i L^i$ and $\mathbf{B}(L) = \sum_{i=0}^{n} \beta_i L^i$ then equation (1) can be rewritten as

$$\mathbf{y}_t = \mathbf{C}(L)\,\mathbf{x}_t + \mathbf{A}(L)^{-1}\mathbf{e}_t \tag{11}$$

where $\mathbf{C}(L) = \mathbf{B}(L)/\mathbf{A}(L)$. The average lag for $\mathbf{C}(L)$ is

$$\left.\frac{d \log \mathbf{C}(L)}{dL}\right|_{L=1} = \left[\frac{d \log \mathbf{B}(L)}{dL} - \frac{d \log \mathbf{A}(L)}{dL}\right]_{L=1} = \frac{\sum_{i=1}^{n} i\beta_i}{\sum_{i=0}^{n} \beta_i} + \frac{\sum_{i=1}^{m} i\alpha_i}{1 - \sum_{i=1}^{m} \alpha_i}. \tag{12}$$

If the coefficients of $\Delta y_t$ and $\Delta x_t$ in (10) are denoted $\rho$ and $\nu$, respectively, then the average lag $\mu$ is

$$\mu = -(\rho + \nu/\theta). \tag{13}$$

This can also be estimated directly by a two-stage procedure. First estimate (10) and then estimate

$$y_t + \frac{\hat{\nu}}{\hat{\theta}} \Delta y_t = \left(\rho + \frac{\nu}{\theta}\right) \Delta y_t + \lambda \sum_{i=2}^{m} \alpha_i (i\Delta y_t - \Delta_i y_t) + \theta x_t + \nu \Delta x_t + \sum_{i=2}^{n} \beta_i (i\Delta x_t - \Delta_i x_t) + \lambda e_t \tag{14}$$

where the dependent variable of (14) is constructed using the estimates $\hat{\nu}$ and $\hat{\theta}$ of $\nu$ and $\theta$ obtained from the first stage. The coefficient of $\Delta y_t$ in (14) is then $-\mu$. It should be noted that for the average lag to make much sense we require that the coefficients of $\mathbf{C}(\mathrm{L})$, namely $c_i$ ($i = 0, 1, \ldots$), be non-negative. Also we assume that $\mathbf{A}(\mathrm{L})$ is invertible.

*Standard Formulations*

These alternative formulations can be compared with other models in the literature.

(i) *Partial adjustment model.* This is obtained from (1) if $m = 1$ and $n = 0$ when

$$y_t = \alpha_1 y_{t-1} + \beta_0 x_t + e_t. \tag{15}$$

Re-writing this in the form of (4) we have

$$\begin{aligned} y_t &= \frac{-\alpha_1}{1-\alpha_1}\Delta y_t + \frac{\beta_0}{1-\alpha_1}x_t + \frac{e_t}{1-\alpha_1} \\ &= \phi_1 \Delta y_t + \theta x_t + u_t, \end{aligned} \tag{16}$$

which is a special case of (6) and provides point estimates of both the long-run multiplier and the average lag.

(ii) *Error correction model (ECM).* The ECM has been popularised by David Hendry in numerous papers, see for example, Davidson *et al.* (1978) and Hendry *et al.* (1984). Also see Currie (1981) for a discussion of the dynamic properties of the ECM. The purpose of the ECM is to focus on the short-run dynamics while making them consistent with the long-run solution. The ECM imposes on (1) the restriction

$$\sum_{i=1}^{m} \alpha_i + \sum_{i=0}^{n} \beta_i = 1. \tag{17}$$

For $m = n = 1$ we can write the ECM as

$$\Delta y_t = \beta_0 \Delta x_t - (1-\alpha_1)(y_{t-1} - x_{t-1}) + e_t. \tag{18}$$

The term $y_{t-1} - x_{t-1}$ is the error which is to be corrected. It also contains the (static) long-run solution which in effect has been separated from the short-run dynamics. Thus the ECM imposes the restriction that the long-run multiplier $\theta$ equals unity. If this restriction was removed then (18) could be written in extended ECM form as

$$\Delta y_t = \beta_0 \Delta x_t - (1-\alpha_1)(y_{t-1} - \theta x_{t-1}) + e_t. \tag{19}$$

Equation (19) is, of course, just another way of re-writing (1) without imposing any coefficient restrictions. But it is not a particularly convenient form for estimation – especially of $\theta$ – hence the attraction of setting $\theta$ equal to unity. Usually, however, this is not a price worth paying for the convenience of estimation and for separating the long-run solution from the short-run dynamics. If we want an estimate of the long-run as well as the short-run dynamics it would be better to use (4) or (6) which, for this case, becomes

$$y_t = \phi_1 \Delta y_t + \theta x_t - \gamma_1 \Delta x_t + u_t. \tag{20}$$

Another problem is the initial specification of the short-run dynamics. For general $m$ and $n$ and $k = \max(m, n)$ the ECM is written (see Hendry *et al.*, 1984)

$$\Delta y_t = -\sum_{i=1}^{k-1}\left(1-\sum_{j=1}^{i}\alpha_j\right)\Delta y_{t-1}+\sum_{i=0}^{k-1}\sum_{j=0}^{i}\beta_j\,\Delta x_{t-i}-\left(1-\sum_{i=1}^{k}\alpha_i\right)(y_{t-k}-x_{t-k})+e_t. \quad (21)$$

Three points can be made about (21)

(i) It is usual with the ECM to set $m = n\,(=k)$. If $m \neq n$, then, in effect, we must add coefficients which are zero to the shorter of the two parameter sets to make the two equal in length.

(ii) The error term which captures the static long-run solution usually enters with the maximum lag of $k$. We could, however, introduce instead the error term at another lag. Lag 1 or min $(m, n)$ are alternatives.

(iii) It is necessary to choose a value of $k$ at the outset, but this choice will often be arbitrary. Furthermore, due to the need for the model to have a (static) long-run solution, we would normally expect $y_{t-k}-x_{t-k}$ to be significant even though the choice of $k$ may be incorrect. Tests for serially correlated residuals may reveal misspecification due to $k$ being too small. But it may be more difficult to detect $k$ being too large because the coefficients of the other terms will be significant in part in automatic compensation for the mis-specification of $k$. The appropriate test for the choice of $k$ too large is not therefore the significance of $y_{t-k}-x_{t-k}$ or of $\Delta y_{t-k+1}$, $\Delta x_{t-k+1}$ etc., but whether the coefficients of $\Delta y_{t-k+1}$ and $\Delta x_{t-k+1}$ differ from those of $\Delta y_{t-k+2}$ and $\Delta x_{t-k+2}$, respectively. If they do not then $k$ can be reduced to $k-1$ etc. According to (21) there should be no change in the coefficients of a higher order lagged difference in (21) if that lag does not enter (1). This procedure can also be used to test whether $m = n$ by looking at $\Delta y_{t-k+1}$ and $\Delta x_{t-k+1}$ individually instead of jointly as above. A danger for the ECM, therefore, is that it may be overparameterised. This problem does not arise in equation (4) because the coefficients of the difference terms in (4) are proportional to their values in (1). Conventional tests for the mis-specification of the short-run dynamics can be carried out in (4) more or less in the usual way. We shall return to this point in Section III when we consider methods of estimating of (4).

### (*b*) *Multiple Equation Systems*

It is possible to extend the above results to a system of equations. Consider the simultaneous equation system

$$\sum_{i=0}^{m}\mathbf{Y}_{t-i}\mathbf{A}_i = \sum_{i=0}^{n}\mathbf{X}_{t-i}\mathbf{B}_i+\mathbf{E}_t, \quad (22)$$

where $\mathbf{Y}_t$ is a $1\times p$ vector of endogenous variables $\mathbf{X}_t$ is a $1\times q$ vector of endogenous variables and $\mathbf{E}_t$ is a $1\times p$ vector of disturbances which are assumed to be i.i.d. $(\mathbf{0}, \boldsymbol{\Sigma})$. The reduced form of (22) defines a multiple equation system

$$\mathbf{Y}_t = \sum_{i=1}^{m} \mathbf{Y}_{t-i}\mathbf{C}_i + \sum_{i=0}^{n} \mathbf{X}_{t-i}\mathbf{D}_i + \mathbf{U}_t \tag{23}$$

where $\mathbf{C}_i = -\mathbf{A}_i\mathbf{A}_0^{-1}$, $\mathbf{D}_i = \mathbf{B}_i\mathbf{A}_0^{-1}$ and $\mathbf{U}_t = \mathbf{E}_t\mathbf{A}_0^{-1}$. The matrix of long-run multipliers is

$$\boldsymbol{\Phi} = \left(\sum_{i=0}^{n} \mathbf{D}_i\right)\left(\mathbf{I} - \sum_{i=1}^{m} \mathbf{C}_i\right)^{-1} \tag{24}$$

$$= \left(\sum_{i=0}^{n} \mathbf{B}_i\right)\left(\sum_{i=0}^{m} \mathbf{A}_i\right)^{-1}. \tag{25}$$

The covariance matrix of a consistent estimator of $\boldsymbol{\Phi}$ is given in Wickens (1978).

In order to estimate $\boldsymbol{\Phi}$ directly we can re-write (23) analogously to (4) as

$$\mathbf{Y}_t = \sum_{i=1}^{m} \Delta_i \mathbf{Y}_t \mathbf{F}_i + \mathbf{X}_t\boldsymbol{\Phi} + \sum_{i=1}^{n} \Delta_i \mathbf{X}_t \mathbf{G}_i + \mathbf{V}_t \tag{26}$$

where $\mathbf{F}_i = \mathbf{C}_i\mathbf{H}$, $\mathbf{G}_i = \mathbf{D}_i\mathbf{H}$, $\mathbf{V}_t = \mathbf{U}_t\mathbf{H}$ and $\mathbf{H} = (\mathbf{I} - \sum_{i=1}^{m}\mathbf{C}_i)^{-1}$. The estimation method of (26) will be discussed in Section III. It should be noted, however, that in general $\mathbf{F}_i$ and $\mathbf{G}_i$ will be constrained due to the identifying restrictions on (22).

We derived equation (23) from the structural form (22). Often, however, in empirical studies we find the model initially specified as a multiple equation system like (23) rather than a structural system. Such a model is an example of a vector ARMAX system. In these cases the coefficients of (23) are usually unrestricted. If interest centres on the long-run multiplier matrix then equation (26) is again a convenient alternative formulation and, moreover, the coefficients of (26) would also be unrestricted. A multiple equation version of the ECM has been proposed by Davidson (1983). The minimal restrictions this imposes on (23) are that $m = n$ and $\sum_{i=1}^{m}\mathbf{C}_i + \sum_{i=0}^{m}\mathbf{D}_i = \mathbf{I}$. It is also possible to add further restrictions to either the short-run dynamics or to the error correction term. In general the deviation from equilibrium of every endogenous variable will affect the dynamic behaviour of each endogenous variable. Structural information might be used to limit the number of exogenous variables entering the long-run solution of a particular endogenous variable, but it is difficult to see how it could be used to restrict the short-run dynamics, or even exclude any particular deviation from equilibrium of an endogenous variable from an equation. The criticisms made earlier in connection with the single equation ECM will also apply to systems.

### (c) *The long-run structural system*

So far our discussion of equation systems has focussed on the long-run multipliers of the system. Interest may also be attached to the long-run structural system. Assuming that the coefficient of the $i$th endogenous variable in the $i$th structural equation is normalised, the long-run structure of (22) may be written as

$$\mathbf{Y}_t\left(\sum_{i=0}^{m}\mathbf{A}_i\right)\mathbf{A}^* = \mathbf{X}_t\left(\sum_{i=0}^{n}\mathbf{B}_i\right)\mathbf{A}^*, \tag{27}$$

where $\mathbf{A}^{*-1}$ is a diagonal matrix formed from the diagonal elements of $\sum_{i=0}^{m}\mathbf{A}_i$. We can re-write (22) to obtain reformulations corresponding to equations (4)–(6). For example the formulation corresponding to equation (4) is

$$\mathbf{Y}_t\left(\sum_{i=0}^{m}\mathbf{A}_i\right)\mathbf{A}^* + \sum_{i=1}^{m}\Delta_i\mathbf{Y}_t\mathbf{A}_i\mathbf{A}^* = \mathbf{X}_t\left(\sum_{i=0}^{n}\mathbf{B}_i\right)\mathbf{A}^* + \sum_{i=1}^{n}\Delta_i\mathbf{X}_t\mathbf{B}_i\mathbf{A}^* + \mathbf{E}_t\mathbf{A}^*. \tag{28}$$

Thus the long-run structure can be estimated directly from estimates of (28).

Similar reformulations are possible for sub-systems or for single structural equations. In the latter case the results of Section I ($a$) can be applied directly if we re-write the single structural equation like equation (1) with $y_t$ as the normalised endogenous variable.

## II. THE ESTIMATION OF LINEARLY TRANSFORMED AND RE-NORMALISED REGRESSION EQUATIONS

The transformed models of Section I are a particular case of a more general situation which is the following. Consider the standard regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{29}$$

where $\mathbf{y}$ is a $T \times 1$ vector of observations on the dependent variable, $\mathbf{X}$ is $T \times k$ matrix of pre-determined variables with $\text{p lim}\, T^{-1}\mathbf{X}'\mathbf{X} = M$ and $\mathbf{e}$ is NID $(\mathbf{0}, \boldsymbol{\Sigma})$. Let $\mathbf{A}$ be a size $(k+1)$ non-singular transformation matrix of constants and define $\mathbf{y}^*$ and $\mathbf{X}^*$ to satisfy

$$(\mathbf{y}^*\,\mathbf{X}^*) = (\mathbf{y}\;\mathbf{X})\,\mathbf{A}. \tag{30}$$

We can therefore re-write (29) as the re-normalised model

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\gamma} + \mathbf{u}, \tag{31}$$

where $\boldsymbol{\gamma}$ is obtained from

$$\begin{bmatrix} 1 \\ -\boldsymbol{\beta} \end{bmatrix} = \theta\mathbf{A}\begin{bmatrix} 1 \\ -\boldsymbol{\gamma} \end{bmatrix}, \tag{32}$$

$\theta = a^{11} - \mathbf{A}^{12}\boldsymbol{\beta}$ is the normalising constant, $\mathbf{e} = \theta\mathbf{u}$, and, partitioning conformably with $(\mathbf{y}\;\mathbf{X})$,

$$\mathbf{A} = \begin{bmatrix} a_{11} & \mathbf{A}_{12} \\ a_{21} & \mathbf{A}_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \begin{bmatrix} a^{11} & a^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}.$$

It is assumed that interest centres on obtaining an estimate of $\boldsymbol{\gamma}$ or an element of $\boldsymbol{\gamma}$. The standard procedure is to estimate $\boldsymbol{\beta}$ by OLS from (29) and then solve

(32) to obtain an estimate of $\boldsymbol{\gamma}$. Let $\hat{\boldsymbol{\gamma}}_{\mathrm{OLS}}$ denote the resulting estimator of $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ the OLS estimator of $\boldsymbol{\beta}$ then

$$\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_{\mathrm{OLS}} \end{bmatrix} = \hat{\theta}^{-1}\mathbf{A}^{-1}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} \end{bmatrix}, \tag{33}$$

where $\hat{\theta} = a^{11} - \mathbf{A}^{12}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$. Estimating $\boldsymbol{\gamma}$ directly from (31) by OLS will not, in general, produce a consistent estimator because $\mathbf{X}^*$ is not asymptotically uncorrelated with $\mathbf{u}$ due to $\mathbf{X}^*$ being a function of $\mathbf{y}$. A consistent estimator of $\boldsymbol{\gamma}$ can be obtained by using instead the ordinary instrumental variable (IV) estimator with $\mathbf{X}$ as the matrix of instruments. The resulting estimator $\hat{\boldsymbol{\gamma}}_{\mathrm{IV}}$ is obtained from

$$\mathbf{X}'(\mathbf{y}^*\mathbf{X}^*)\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_{\mathrm{IV}} \end{bmatrix} = 0. \tag{34}$$

It can be shown that $\hat{\boldsymbol{\gamma}}_{\mathrm{IV}} = \hat{\boldsymbol{\gamma}}_{\mathrm{OLS}}$ as follows.

Equation (34) can be re-written

$$\mathbf{X}'(\mathbf{y}\ \mathbf{X})\,\mathbf{A}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_{\mathrm{IV}} \end{bmatrix} = \mathbf{0} \tag{35}$$

and $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ satisfies

$$\mathbf{X}'(\mathbf{y}\ \mathbf{X})\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} \end{bmatrix} = \mathbf{0}. \tag{36}$$

Multiplying both sides of (35) by $\hat{\theta}$ and then comparing with (36) we find that

$$\hat{\theta}\mathbf{A}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_{\mathrm{IV}} \end{bmatrix} = \begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} \end{bmatrix}. \tag{37}$$

It follows from (34) and (37) that $\hat{\boldsymbol{\gamma}}_{\mathrm{IV}} = \hat{\boldsymbol{\gamma}}_{\mathrm{OLS}}$.

Not only are the two estimators identical, we can show that their estimated variance-covariance matrices will also be the same. A consistent estimator of the variance-covariance matrix of $\hat{\boldsymbol{\gamma}}_{\mathrm{IV}}$ is given by

$$\hat{\mathbf{V}}_{\hat{\gamma}_{\mathrm{IV}}} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}[\mathbf{X}^{*\prime}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}^*]^{-1} \tag{38}$$

where
$$\hat{\mathbf{u}} = [\mathbf{y}^*\mathbf{X}^*]\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_{\mathrm{IV}} \end{bmatrix}$$

are the IV residuals. Since

$$\begin{aligned}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}^* &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{y}\ \mathbf{X}]\begin{bmatrix} \mathbf{A}_{12} \\ \mathbf{A}_{22} \end{bmatrix} \\ &= [\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}\ \mathbf{I}]\begin{bmatrix} \mathbf{A}_{12} \\ \mathbf{A}_{22} \end{bmatrix} \\ &= \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}\mathbf{A}_{12} + \mathbf{A}_{22},\end{aligned}$$

we can write (38) as

$$\hat{\mathbf{V}}_{\hat{\gamma}_{\mathrm{IV}}} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}[(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}\mathbf{A}_{12} + \mathbf{A}_{22})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}\mathbf{A}_{12} + \mathbf{A}_{22})]^{-1}. \tag{39}$$

The variance of $\hat{\boldsymbol{\gamma}}_{\text{OLS}}$ is obtained from the fact that $\hat{\boldsymbol{\gamma}}_{\text{OLS}} = \boldsymbol{\gamma}(\hat{\boldsymbol{\beta}}_{\text{OLS}})$ and so

$$\mathbf{V}_{\hat{\gamma}_{\text{OLS}}} = \frac{\delta \boldsymbol{\gamma}\,(\hat{\boldsymbol{\beta}}_{\text{OLS}})}{\delta \boldsymbol{\beta}} \mathbf{V}_{\hat{\beta}_{\text{OLS}}} \frac{\delta \boldsymbol{\gamma}\,(\hat{\boldsymbol{\beta}}_{\text{OLS}})'}{\delta \boldsymbol{\beta}}$$

$$= \left[\frac{\delta \boldsymbol{\beta}\,(\hat{\boldsymbol{\beta}}_{\text{OLS}})'}{\delta \boldsymbol{\gamma}} \mathbf{V}^{-1}_{\hat{\beta}_{\text{OLS}}} \frac{\delta \boldsymbol{\beta}\,(\hat{\boldsymbol{\beta}}_{\text{OLS}})}{\delta \boldsymbol{\gamma}}\right]^{-1}. \tag{40}$$

From (32),

$$\mathbf{A}^{21} - \mathbf{A}^{22}\boldsymbol{\beta} = -\theta\boldsymbol{\gamma}. \tag{41}$$

Hence differentiating (41)

$$-\mathbf{A}^{22}\frac{\delta \boldsymbol{\beta}}{\delta \boldsymbol{\gamma}} = -\theta \mathbf{I} - \boldsymbol{\gamma}\frac{\delta \theta}{\delta \boldsymbol{\beta}}\frac{\delta \boldsymbol{\beta}}{\delta \boldsymbol{\gamma}},$$

yielding

$$\frac{\delta \boldsymbol{\beta}}{\delta \boldsymbol{\gamma}} = \theta(\mathbf{A}^{22} + \gamma \mathbf{A}^{12})^{-1}. \tag{42}$$

Equation (32) also implies that

$$[\boldsymbol{\beta}\ \mathbf{I}]\,\mathbf{A}\begin{bmatrix} 1 \\ -\boldsymbol{\gamma} \end{bmatrix} = 0,$$

from which we can solve for $\boldsymbol{\gamma}$ to give

$$\boldsymbol{\gamma} = (\boldsymbol{\beta}\mathbf{A}_{12} + \mathbf{A}_{22})^{-1}\,(\boldsymbol{\beta} a_{11} + \mathbf{A}_{21}).$$

Substituting for $\boldsymbol{\gamma}$ in (42) and using $\mathbf{AA}^{-1} = \mathbf{I}$ we have

$$\frac{\delta \boldsymbol{\beta}}{\delta \boldsymbol{\gamma}} = \theta[\mathbf{A}_{22}\,\mathbf{A}^{22} + \mathbf{A}_{21}\,\mathbf{A}^{12} + \boldsymbol{\beta}(a_{11}\,\mathbf{A}^{12} + \mathbf{A}_{12}\,\mathbf{A}^{22})]^{-1}\,(\boldsymbol{\beta}\mathbf{A}_{12} + \mathbf{A}_{22})$$

$$= \theta(\boldsymbol{\beta}\mathbf{A}_{12} + \mathbf{A}_{22}). \tag{43}$$

From (43) and $\hat{\mathbf{e}} = \hat{\theta}\hat{\mathbf{u}}$, or

$$\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T} = \hat{\theta}^2\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T},$$

equation (40) becomes

$$\hat{\mathbf{V}}_{\hat{\gamma}_{\text{OLS}}} = \left[\hat{\theta}(\hat{\boldsymbol{\beta}}_{\text{OLS}}\,\mathbf{A}_{12} + \mathbf{A}_{22})'\hat{\theta}^{-2}\left(\frac{\hat{\mathbf{u}}'\mathbf{u}}{T}\right)^{-1}\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}_{\text{OLS}}\,\mathbf{A}_{12} + \mathbf{A}_{22})\,\hat{\theta}\right]^{-1}$$

$$= \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}\,[(\hat{\boldsymbol{\beta}}_{\text{OLS}}\,\mathbf{A}_{12} + \mathbf{A}_{22})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}_{\text{OLS}}\,\mathbf{A}_{12} + \mathbf{A}_{22})]^{-1}. \tag{44}$$

Equation (44) is identical to (29), which completes the proof that $\hat{\mathbf{V}}_{\hat{\gamma}_{\text{IV}}} = \hat{\mathbf{V}}_{\hat{\gamma}_{\text{OLS}}}$.

Thus whether we estimate $\boldsymbol{\gamma}$ by IV on the transformed equation (31) and use the estimate of its covariance matrix given by the IV programme, or first estimate the original model by OLS and then solve for $\hat{\boldsymbol{\gamma}}$ and its covariance matrix using the OLS estimates, is immaterial as far as the numerical values obtained. The choice of estimation method can therefore be made on the basis of which is more computationally convenient.

We can go further than this. Because $\mathbf{e} = \theta\mathbf{u}$ it follows that the usual residual based tests which are invariant to $\theta$ will give the same results.

## III. THE ESTIMATION OF THE TRANSFORMED DYNAMIC MODEL

### (a) *Single Equation Estimation*

(i) *Using Re-Normalised Regression Theory*

The results of Section II can be applied to the estimation of the transformed dynamic models considered in Section I since equation (4) can be interpreted as a re-normalised regression model. Thus

$$[y_t \vdots \mathbf{X}_t] \equiv [\mathbf{y}_t \vdots \mathbf{y}_{t-1} \dots \mathbf{y}_{t-m}\, \mathbf{x}_t\, \mathbf{x}_{t-1} \dots \mathbf{x}_{t-n}]$$

$$[\mathbf{y}_t^* \vdots \mathbf{X}_t^*] \equiv [\mathbf{y}_t \vdots \Delta_1 \mathbf{y}_t \dots \Delta_m \mathbf{y}_t\, \mathbf{x}_t\, \Delta_1 \mathbf{x}_t \dots \Delta_n \mathbf{x}_t]$$

and

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 \\ 0 & -1 & & 0 & 0 & -1 & & 0 \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & -1 & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & -1 \end{bmatrix}$$

where $\mathbf{A}$ assumes $n > m$. It follows that the IV estimator of (4) with instruments given by the explanatory variables of (1) – it is important to have neither any more nor any less instruments – will give exactly the same estimates of all the coefficients of (4), including the long-run multiplier $\theta$, that would be obtained by first estimating (1) by OLS and then solving for these coefficients; $\hat{\theta}$ is solved from (2). Moreover, the estimated standard errors will also be identical, as will the test statistics for serially correlated errors, etc.

(ii) *Using Simultaneous Equation Estimation Theory*

Another way to approach the problem of estimating (6) is to view this equation as a single exactly identified structural equation. The system can be completed either with the $m$ identities $\Delta_i y_t = y_t - y_{t-i}$, or by the reduced form equations for $\Delta_i y_t$ obtained through substituting into the identities for $y_t$ using (1). We can now use the result that the two stage least squares estimator of an exactly identified structural equation is identical to the indirect least squares estimator. The reduced form equation for $y_t$ is just equation (1) and this has the same explanatory variables as the reduced form equations for $\Delta_i y_t$. Hence the 2SLS estimator is the same as the IV estimator proposed above. The indirect least squares estimator is derived from OLS estimates of the reduced form coefficients. Again this is equivalent to estimating $\theta$ from OLS estimates of (1).

*(b) Multiple Equation System Estimation*

If equations (23) and hence (26) contain coefficient restrictions, as they almost certainly will if they are derived from a structural system like (22), then the results of Section II cannot be used. We shall therefore describe another approach which can. There are, however, cases where the results of Section II are applicable. We shall comment on these.

The objective is to find a consistent and reasonably efficient way of estimating equation (26) directly. Although equation (23) can be estimated consistently and efficiently by using OLS and taking into account the coefficient restrictions, this is not true for equation (26) because it contains terms in $\Delta_i \mathbf{Y}_t$ on the right-hand side. These will be correlated asymptotically with the disturbance term $\mathbf{V}_t$ making OLS inconsistent. Equation (26) may be interpreted as a subsystem of equations. The system can be completed by adding equations for the $\Delta_j \mathbf{Y}_t$ ($j = 1, \ldots, m$). We can choose for this purpose the reduced forms of $\Delta_j \mathbf{Y}_t$. These are

$$\Delta_j \mathbf{Y}_t = \sum_{i=1}^{m} \mathbf{Y}_{t-i}(\mathbf{C}_i - \delta_{ij}\mathbf{I}) + \sum_{i=0}^{n} \mathbf{X}_{t-i}\mathbf{D}_i + \mathbf{U}_t, \tag{45}$$

where $\delta_{ij} = 1$ for $i = j$ and 0 otherwise. The complete system can now be estimated by FIML (or 3SLS). But this is equivalent to using sub-system LIML (or subsystem 3SLS) on (26), see Phillips and Wickens (1978, Question 6.18) or Godfrey and Wickens (1982). These estimators will be asymptotically efficient provided the restrictions implied by the structural form are taken into account.

In the special case where there are no restrictions on the coefficients of equation (23), or if the restrictions are ignored, then these subsystem estimators are equivalent to using LIML (or 2SLS) equation by equation on (26). Furthermore, in this case we can apply the results of Section II.

To see this we re-write (23) in matrix form as

$$\mathbf{Y} = \sum_{i=1}^{m} \mathbf{Y}_{-i}\mathbf{C}_i + \sum_{i=0}^{n} \mathbf{X}_{-i}\mathbf{D}_i + \mathbf{U}, \tag{46}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{Y}_T \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{X}_t \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{U}_T \end{bmatrix}$$

and $\mathbf{Y}_{-i}$, $\mathbf{X}_{-i}$ are the $i$th lags of $\mathbf{Y}$ and $\mathbf{X}$ respectively. If we define

$$\mathbf{Z} = [\mathbf{Y}_{-i} \ldots \mathbf{Y}_{-m}\ \mathbf{X}\ \mathbf{X}_{-1} \ldots \mathbf{X}_n],$$

in where the $t$th row of $\mathbf{Z}$ is $\mathbf{Z}_t = [\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_{t-m}, \mathbf{X}_t, \mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-n}]$ and

$$\mathbf{H} = [\mathbf{C}_1' \ldots \mathbf{C}_m'\ \mathbf{D}_0'\ \mathbf{D}_1' \ldots \mathbf{D}_n'],$$

then (46) can be re-written

$$\mathbf{Y} = \mathbf{ZH} + \mathbf{U}$$

and the $i$th equation of (46) can be written

$$\mathbf{y}_i = \mathbf{Z}\boldsymbol{\delta}_i + \mathbf{u}_i \quad (i = 1, \ldots, p),$$

where $\mathbf{y}_i$, $\mathbf{u}_i$, $\boldsymbol{\delta}_i$ are the $i$th columns of $\mathbf{Y}$, $\mathbf{U}$ and $\mathbf{H}$ respectively. If we define $\underset{\sim}{\mathbf{y}}' = (\mathbf{y}_1', \ldots, \mathbf{y}_p')$, $\underset{\sim}{\mathbf{u}} = (\mathbf{u}_1', \ldots, \mathbf{u}_p')$, $\underset{\sim}{\boldsymbol{\delta}}' = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_p')$ and $\underset{\sim}{\mathbf{Z}} = (\mathbf{I}_p \otimes \mathbf{Z})$ we can write (46) as

$$\underset{\sim}{\mathbf{y}} = \underset{\sim}{\mathbf{Z}}\underset{\sim}{\boldsymbol{\delta}} + \underset{\sim}{\mathbf{u}}. \tag{47}$$

Equation (26) can be written in a similar way as

$$\underset{\sim}{\mathbf{y}}^* = \underset{\sim}{\mathbf{Z}}^*\underset{\sim}{\boldsymbol{\delta}}^* + \underset{\sim}{\mathbf{v}}, \tag{48}$$

where the only substantial change is in the construction of $\underset{\sim}{\mathbf{Z}}^*$. This is based on

$$\underset{\sim}{\mathbf{Z}}_t^* = [\Delta_1 \mathbf{Y}_t \ldots \Delta_m \mathbf{Y}_t\, \mathbf{X}_t\, \Delta_1 \mathbf{X}_t \ldots \Delta_n \mathbf{X}_t].$$

We have therefore re-written equation (23) as (47) and this takes the form of equation (29). Equation (26) has been re-written as (48) and this corresponds to (31). In other words, there exists a matrix $\mathbf{A}$ which satisfies

$$[\underset{\sim}{\mathbf{y}}^*\, \underset{\sim}{\mathbf{Z}}^*] = [\underset{\sim}{\mathbf{y}}\, \underset{\sim}{\mathbf{Z}}]\, \underset{\sim}{A}. \tag{49}$$

Equation (49) corresponds to (30). Equation (47) has no restrictions on $\boldsymbol{\delta}$ and can be estimated efficiently using OLS. Equation (48) can be estimated efficiently using the IV estimator with instrument matrix $\underset{\sim}{\mathbf{Z}}$. From the results of Section II, these estimates will be identical to estimating (47) and then solving for $\boldsymbol{\delta}^*$.

### (*c*) *The Estimation of the Long-run Structure*

Asymptotically efficient estimates of the coefficients of (28) can be obtained by using once more sub-system LIML or sub-system 3SLS. The resulting estimates will not in general be numerically identical to those obtained by first estimating (22) and then re-writing the estimated system in the form of (28). We pursue this point further in relation to a single structural equation.

Consistent estimates of the long-run of a single structural equation can be obtained by first consistently estimating the original equation using, for example, an instrumental variable estimator, and solving for the long-run coefficients. Alternatively, the re-formulated equation can be estimated directly by instrumental variables. The two sets of estimates of the long-run coefficients will be the same if the number of instruments used in each case is the same as the number of coefficients to be estimated. For example, if the original structural equation is exactly identified and 2SLS is used, the two sets of estimates will be identical. But if the equation is over-identified, then the 2SLS estimates will be different.

The proof of this result is related to the earlier proof of the equivalence of solved OLS and IV estimates of the long-run multipliers. Reinterpreting equation (29) as a single structural equation and denoting the matrix of instruments as $\mathbf{Z}$, the IV estimator of (29) is obtained by solving

$$\mathbf{X}'\mathbf{P}_z\,[\mathbf{y}\ \mathbf{X}]\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{\text{IV}} \end{bmatrix} = \mathbf{0}, \tag{50}$$

where $\mathbf{P}_z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. For (31) we solve

$$\mathbf{X}'\mathbf{P}_z\,[\mathbf{y}^*\ \mathbf{x}^*]\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_{\text{IV}} \end{bmatrix} = \mathbf{0}. \tag{51}$$

Both estimators will provide consistent estimators of the long-run structural coefficients $\boldsymbol{\theta}$, but these estimators will be numerically the same only if $\mathbf{Z}$ has the same number of columns as $\mathbf{X}$ and $\mathbf{X}^*$. The two estimators $\hat{\boldsymbol{\beta}}_{\text{IV}}$ and $\hat{\boldsymbol{\gamma}}_{\text{IV}}$ can then be obtained from

$$\mathbf{Z}'[\mathbf{y}\ \mathbf{X}]\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{\text{IV}} \end{bmatrix} = \mathbf{0} \tag{52}$$

$$\mathbf{Z}'[\mathbf{y}^*\ \mathbf{X}^*]\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_{\text{IV}} \end{bmatrix} = \mathbf{0}. \tag{53}$$

In this case we are simply replacing the pre-multiplying matrix $\mathbf{X}'$ in (34)–(36) by $\mathbf{Z}'$ which permits the equivalence result to carry through. Although we could impose the equivalence result on over-identified equations by discarding instruments there would be a loss of efficiency.

## IV. DYNAMIC SPECIFICATION AND NON-STATIONARY VARIABLES

So far we have assumed that all of the variables are stationary. Recent research on regression with non-stationary variables has important implications for dynamic specification and the estimation of the long run. A non-stationary stochastic variable may become stationary after the removal of a non-stochastic trend or after differencing, i.e. be trend of difference stationary. In each case it will often be possible to derive consistent estimates of some (or all) of the long-run multipliers even though the model is dynamically mis-specified. This result is related to the concept of the co-integrating vector Granger (1983; 1986), Engle and Granger (1987).

To illustrate, if in equation (1) $x_t$ is generated by the mixed deterministic non-stationary trend and stationary stochastic process

$$x_t = \mu t + \epsilon_t - \gamma\epsilon_{t-1}, \tag{54}$$

where $\epsilon_t$ is an iid $(0, \sigma^2_\epsilon)$ variable with $\text{E}(e_t\, \epsilon_s) = 0$ for $t$ and $s$ and the stochastic component is invertible, then it is possible to obtain a consistent estimate of $\theta$ using OLS on

$$y_t = \theta x_t + u_t, \tag{55}$$

where from (1)

$$u_t = e_t + \sum_{i=1}^{m} \alpha_i y_{t-i} + \sum_{i=1}^{n} \beta_i x_{t-i}. \tag{56}$$

This estimator of $\theta$ will not in general be efficient; it is necessary to include the lags in $y$ and $x$ to achieve full efficiency. A similar 'super-consistency' result exists for variables that are difference stationary, see Stock (1984).

An implication of this result is that it is possible to misspecify the dynamic structure of models with non-stationary variables by omitting higher order lags from (1) – or in the case of the alternative formulations by omitting differenced terms – without affecting the consistency of the estimates of the long-run multipliers associated with the non-stationary variables. The contributions of the differenced variables to $y_t$ are negligible asymptotically compared with those of the levels variables. In general such an estimator of $\theta$ will be biased in finite samples and this bias may not be negligible. See Stock (1984) for evidence of substantial biases and Banerjee *et al.* (1986) for evidence where the biases are not large. Co-integration theory deals with closely related issues, Engle and Granger (1987). If each element of a vector time series $\mathbf{x}_t$ achieves stationarity after differencing, and if a linear combination $\boldsymbol{\alpha}'\mathbf{x}_t$ is stationary, then $\mathbf{x}_t$ is said to be co-integrated with co-integrating vector $\boldsymbol{\alpha}$. Engle and Granger interpret $\boldsymbol{\alpha}'\mathbf{x}_t = 0$ as the long-run equilibrium relationship between the elements of $\mathbf{x}_t$. They note that having arbitrarily normalised one element of $\boldsymbol{\alpha}$, the remaining elements can be consistently estimated by OLS with the dependent variable chosen to correspond to the normalised element. Thus Engle and Granger's co-integrating regression is identical to estimating the long-run multipliers from a model misspecified through the omission of short-run dynamics.

Engle and Granger also argue that co-integrated series can be represented by an ECM and that the ECM can be estimated in two steps. In the first step the co-integrating vector is estimated by OLS and in the second step the full ECM is estimated, again by OLS, with the lagged residual of the co-integrating regression substituted for the error which is being corrected. Engle and Granger's use of the term error correction model appears to be different from that of Hendry since they do not impose the restriction that $\Sigma\alpha_i + \Sigma\beta_i = 1$. Above we have referred to this type of model as an extended ECM and we noted that it can be interpreted as just another transformation of (1) – though generalised here to have more than one exogenous variable. Because the ECM restriction is not imposed there does not seem much point in using a two step estimation procedure. Moreover, if interest attaches to the coefficients of the equilibrium relationship it would be easier to estimate in one step one of the transformations of (1) described in Section I. The asymptotic distribution of the estimates would be the same. It should be noted, however, that general results on the asymptotic distribution of coefficients regression models with difference stationary variables are not yet available. It may be conjectured from the preliminary results of Gourieroux *et al.* (1987) that for the regression coefficients $\hat{\boldsymbol{\beta}}$ of the general linear model with lags, $T^{-\frac{1}{2}}\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ will converge to a normal distribution, implying that the usual approximations to the finite sample distribution of $\hat{\boldsymbol{\beta}}$ can be used. See also Phillips (1985*a*, *b*, 1986*a*, *b*) and Perron and Phillips (1986). Perhaps of more importance, it is probable that the finite sample biases of the long-run estimates will be reduced by not omitting the short-run dynamics.

## V. CONCLUSIONS

In this paper we have shown how it is possible to re-formulate dynamic models in such a way that we can easily obtain point estimates of long-run multipliers. We have argued that this re-formulation will often be preferable to other models, including the error correction model. Identical estimates of the coefficients of the transformed model are obtained using functions of the OLS estimates of the original model and instrumental variables on the transformed equations. Virtually all of the conventional error diagnostic tests can be applied with the same results to either the original model or to the transformed model.

We have discussed the consequences of omitting the short-run dynamics entirely, or of misspecifying them. We argued that for stationary series such misspecification results in inconsistent estimates, but for non-stationary series a consistent estimator of the long-run multipliers can still be obtained. The latter result is another way of expressing Engle and Granger's finding that a consistent estimator of the co-integrating vector can be obtained from a regression involving only the co-integrating series. Engle and Granger relate co-integrated series to the ECM but because of their different interpretation, their ECM is just another alternative dynamic formulation of (1) and not a conventional ECM.

Finally, in the light of our results, we propose the following strategy for dynamic modelling. First estimate the long-run multipliers using one of the alternative formulations. For non-stationary series it may not be important to specify the length of the lag correctly. For stationary series it may be best to over parameterise by including too many lags. The conventional tests of misspecification can be carried out on this equation. If the equation passes these tests we may be reasonably confident that the long-run estimates have good statistical properties. If interest centres on the long-run multipliers we can stop at this point. Any subsequent re-estimation, for example to achieve parsimony in the short-run dynamics, should yield very similar long-run estimates. If they are different then this is evidence that invalid restrictions have been imposed.

*Queen's University and University of Southampton*
*Australian National University*

### REFERENCES

Banerjee, A., Dolado, J. J., Hendry, D. F. and Smith, S. W. (1986). 'Exploring equilibrium relationships in econometrics through static models: some Monte Carlo evidence.' *Oxford Bulletin of Economics and Statistics*, vol. 48, pp. 253–77.

Bewley, R. A. (1979). 'The direct estimation of the equilibrium response in a linear model.' *Economics Letters*, vol. 3.

Currie, D. A. (1981). 'Some long-run features of dynamic time-series models.' ECONOMIC JOURNAL, vol. 91, pp. 704–15.

Davidson, J. E. H. (1983). 'Error correction systems.' ICERD Econometrics Discussion Paper 83–70. London School of Economics.

——, Hendry, D. F., Srba, F. and Yeo, S. (1978). 'Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom.' ECONOMIC JOURNAL, vol. 88, pp. 661–92.

Engle, R. F. and Granger, C. W. J. (1987). 'Cointegration and error correction: representation, estimation and testing.' *Econometrica*, vol. 55, pp. 251–76.

Godfrey, L. G. and Wickens, M. R. (1982). 'A simple derivation of the limited information maximum likelihood estimator.' *Economics Letters*, vol. 10, pp. 277–83.

Gourieroux, C., Maurel, F. and Monfort, A. (1987). 'Regression and non-stationarity.' Mimeo.

Granger, C. W. J. (1983). 'Co-integrated variables and error-correction models.' UCSD Discussion Paper.

—— (1986). 'Developments in the study of co-integrated economic variables.' Applied Economics Discussion Paper No. 3, University of Oxford.

Hendry, D. F. (1979). 'Predictive failure and econometric modelling in macroeconomics: the transactions demand for money.' *Economic Modelling* (ed. P. Ormerod), pp. 217–42. Heinemann.

——, Pagan, A. R. and Sargan, J. D. (1984). 'Dynamic specification,' chapter 18 in *Handbook of Econometrics*, Vol. II (ed. Z. Griliches and M. Intriligator). North Holland.

—— and Richard, J.-F. (1983). 'The econometric analysis of economic time series.' *International Statistical Review*, vol. 51, pp. 111–63.

Perron, P. and Phillips, P. C. B. (1986). 'Testing for a unit root against trend alternatives.' Cowles Foundation Discussion Paper No. ?, Yale University.

Phillips, P. C. B. (1985*a*). 'Time series regression with unit roots.' Cowles Foundation Discussion Paper No. 740, Yale University.

—— (1985*b*). 'Understanding spurious regressions in econometrics.' Cowles Foundation Discussion Paper No. 757, Yale University.

—— (1986*a*). 'Regression theory for near-integrated time series.' Cowles Foundation Discussion Paper No. 781, Yale University.

—— (1986*b*). 'Towards a unified asymptotic theory for autoregression.' Cowles Foundation Discussion Paper No. 782, Yale University.

—— and Wickens, M. R. (1978). *Exercises in Econometrics*. Philip Allan.

Salmon, M. (1982). 'Error correction mechanisms.' ECONOMIC JOURNAL, vol. 92, pp. 615–29.

Stock, J. H. (1984). 'Asymptotic properties of least squares estimators of cointegrating sectors.' Manuscript, Harvard University.

Wickens, M. R. (1978). 'An additional criterion in model selection.' University of Essex Discussion Paper.