ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 28 — 1992

Lagged Dependent Variables Distributed Lags and Autoregressive Residuals

Stephen POLLOCK *

ABSTRACT. – Whenever the specification of a dynamic regression relationship is in doubt, we should think of adopting a rational transfer-function model with separate parameters in the systematic part and the disturbance part. Some of the models which are commonly used in applied econometrics can give rise to very misleading estimates when the two parts of the true regression relationship have different dynamic properties.

Variables endogènes retardées, retards échelonnés et erreurs autorégressives

RÉSUMÉ. – Quand la spécification d'une relation de régression dynamique est en doute, il est convenable d'adopter un modèle avec une fonction de transfert rationnelle qui donne des paramètres distinctifs à la partie systématique et à la partie des erreurs. On démontre que de nombreux modèles qui sont souvent utilisés dans l'économétrie appliquée peuvent engendrer des valeurs estimées très trompeuses quand les deux parties de la vraie relation de régression comportent des propriétés dynamiques qui sont différentes.

---

* D. S. G. POLLOCK: Department of Economics, Queen Mary College, University of London, Mile End Road, London E1 4NS.

This paper was written while I was enjoying the hospitality of the Faculty of Actuarial Science and Econometrics of the University of Amsterdam. The Netherlands Organisation for Scientific Research (NWO) were generous in giving me financial support.

1 Introduction: Casual Modelling

Economic theory gives only weak indications of the dynamic structure of econometric relationships; and the specification of a dynamic regression equation is often determined in a casual way.

The simplest way, and the most common way, of setting a regression equation in motion is to place the lagged value of the dependent variable on the RHS of the equation in the company of the explanatory variables. When the resulting equation has been fitted, one may examine various diagnostic statistics to see whether the residuals are serially correlated. If serial correlation is detected, then the usual recourse is to attribute a first-order scheme to the disturbances. This may be where the process of modelling ends; for it is not always clear what should be done next if the model persists in failing its tests.

The notion that an autoregressive disturbance scheme may be used for repairing a econometric model which has failed its tests was rejected strongly more than a decade ago in an influential article of HENDRY and MIZON. The lag polynomial entailed by the autoregressive disturbances can be depicted as a factor which is common to the separate lag polynomials operating on the dependent variable and the explanatory variables. Hendry and Mizon asserted that the presence of such a common factor should be demonstrated and not simply assumed.

A model with an autoregressive lag scheme for the dependent variable and with distributed lag schemes for the explanatory variables will be described, in this paper, as an Autoregressive Distributed-Lag Model or an ADM. An ADM in which there are also common factors in the lag schemes will be called a Common-Factor Model or a CFM.

According to Hendry and Mizon, we should begin the process of modelling by fitting an ADM with lags of a relatively high order. Then we should seek to remove excess parameters from the model by asking whether some of the parameters which are associated with high-order lags can be set to zero, and by looking for common factors in the lag polynomials.

Whilst this is an attractive methodology, it should not be imagined that it has a universal validity; and one of the purposes of this paper is to investigate cases where it is inappropriate to use an ADM. We will argue that a Rational Transfer-Function Model or RTM should often be adopted in preference to an ADM.

The essential difference between the RTM and the ADM is that the RTM uses separate parameters to model the systematic and disturbance parts of a regression relationship whereas the ADM does not. The equation of the ADM implies that the two parts of the model are related to each other in a special way. If such a relationship does not hold, then the resulting misspecification might vitiate the estimation of the parameters throughout the model. By contrast, the RTM is capable of delivering consistent estimates of the systematic parameters even when the disturbance part of the model is misspecified.

If one believes that no model can wholly describe the complex and changing nature of an economic process, then the business of fitting a model becomes a matter of providing an approximate mathematical description of the process in which a balance is struck between the criteria of goodness of fit and of parametric parsimony. If one is mindful of the scarcity of econometric data, then one may be inclined to begin with a parsimonious model and to add parameters to it only when the need for them is demonstrated by tests of misspecification.

If we could imagine that the statistical process underlying the data were characterised precisely by an RTM or an ADM, then a fitted RTM with an excess of parameters would suffer from a problem of parametric indeterminacy. This might be regarded as a further reason for beginning a process of model fitting with a parsimonious specification which is unlikely to contain an excess of parameters.

For the purpose of establishing the properties of an estimator, we are bound to postulate a statistical process underlying the data which is of a regular nature. There is no doubt that useful information about the estimator can emerge even when the postulated process is too simple to be realistic.

2 The ADM and the RTM in Comparison

The particular ADM upon which we shall base much of our analysis is described by the equation

$$(1)\qquad (1-\alpha_1 L-\alpha_2 L^2)y(t)=(\beta_0+\beta_1 L)x(t)+\varepsilon(t),$$

where $\varepsilon(t)$ is an unobservable white-noise sequence of independently and identically distributed random variables which are also independent of the elements of the observable signal sequence $x(t)$. In fulfilment of the BIBO (bounded input-bounded output) stability condition, we may assume that the roots of $\alpha(L)=1-\alpha_1 L-\alpha_2 L^2$ lie outside the unit circle.

By imposing a somewhat complicated nonlinear restriction upon the parameters of equation (1), we obtain the equation of a CFM:

$$(2)\qquad (1-\rho L)(1-\alpha L)y(t)=(1-\rho L)\beta x(t)+\varepsilon(t).$$

When both sides of (2) are divided by $1-\rho L$ and when the term $-\alpha L y(t) = -\alpha y(t-1)$ is carried over to the RHS, we get

$$(3) \qquad y(t) = \alpha y(t-1) + \beta x(t) + \eta(t),$$

where $\eta(t) = \rho\eta(t-1) + \varepsilon(t)$ represents a disturbance term which follows a first-order autoregressive scheme. This is the form which is most familiar to econometricians.

In engineering disciplines, it is more common to represent dynamic equations in a transfer-function form which shows how the signal $x(t)$ and the noise $\varepsilon(t)$ are mapped into the output $y(t)$.

The transfer-function form of equation (1) is obtained by dividing throughout by the polynomial operator $\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2$ to give

$$(4) \qquad y(t) = \frac{\beta_0 + \beta_1 L}{1 - \alpha_1 L - \alpha_2 L^2} x(t) + \frac{1}{1 - \alpha_1 L - \alpha_2 L^2} \varepsilon(t).$$

The corresponding form of the CFM is

$$(5) \qquad y(t) = \frac{\beta}{1 - \alpha L} x(t) + \frac{1}{(1 - \rho L)(1 - \alpha L)} \varepsilon(t).$$

The strong assumptions entailed by the ADM are apparent in equation (4) where the denominators of the two transfer functions are identical. This feature implies that, in the absence of a cancellation between the factors in $\beta(L) = \beta_0 + \beta_1 L$ and $\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2$, which would give us equation (5), the two parts of the model have dynamic properties which are essentially the same.

Occasionally an ADM is called for by a particular application. One example concerns the in-flight flutter testing of an aircraft (see WRIGHT). Here the signal $x(t)$ is an excitation applied, by means of the control surfaces, to one part the structure. The output $y(t)$ is a record of the vibrations transduced from another part of the structure. An analysis of the frequency spectrum of $y(t)$ should indicate whether the structure is excessively resonant at any frequencies. The noise part of the model is due to the aerodynamic buffeting which is the source of a further unobserved excitation. In fact, in this application, it would be desirable to include the same number of terms in the numerator of the noise transfer function as in the signal transfer function.

The presence of a common denominator $\alpha(L)$ in the two parts of the model in this example reflects the fact that both the signal and the noise are being mediated through the same structure. The modern-day practice in analysing mechanical vibrations is to characterise each mode of vibration within a structure in terms of the complex roots of $\alpha(L) = 0$.

It is not difficult to find examples in engineering where the ADM is inappropriate and where it must give way to a model with distinct parameters in each transfer function. Consider an old-fashioned "wireless" radio receiver equipped with valves. The signal $x(t)$ stands for the radio transmission and the systematic transfer function represents the means by which it is converted to an audible sound. The noise in this model is the thermionic interference which is caused by the heating of the radio valves (see RICE). Since its origin is separate from that of the radio signal, and since it is mediated through different parts of the circuitry of the wireless, its effect has to be modelled with a separate transfer function. The appropriate model would be a rational transfer-function model or RTM which might take the form of

$$(6) \qquad y(t)=\frac{\delta_0+\delta_1 L}{1-\gamma_1 L-\gamma_2 L^2}x(t)+\frac{\theta_0+\theta_1 L}{1-\varphi_1 L-\varphi_2 L^2}\varepsilon(t).$$

In economics, there is a familiar dichotomy which is similar, in some respects, to the one which embraces the engineering models. The dichotomy arises from the comparison of the adaptive-expectations hypotheses and the partial-adjustment hypothesis.

The adaptive-expectations hypothesis, of which Friedman's permanent-income hypothesis is an example, gives rise to an RTM with a white-noise disturbance process. The equation, in its simplest form, is

$$(7) \qquad y(t)=\frac{\delta}{1-\gamma L}x(t)+\varepsilon(t).$$

According to FRIEDMAN, income receipts have a lingering effect upon consumption which is explained, indirectly, in terms of their effect upon permanent income. The effects of the disturbances, on the other hand, are transitory. They are forgotten after one period. Thus, with reference to his permanent income, the consumer navigates on a strict course in the face of the disturbances; and it is as if he were applying a firm hand to a small boat in a choppy sea.

The partial adjustment hypothesis, which provides an alternative model for the process of consumption, gives rise to an ADM; and, in its simplest form, the equation is

$$(8) \qquad (1-\alpha L)y(t)=\beta x(t)+\varepsilon(t).$$

The notion underlying this model is that it is costly for a consumer to change his behaviour. Therefore his adaptation to a new level of income may be a sluggish one; and his tendency will be to adhere to established habits of consumption. Such habits are influenced by the disturbances; and past disturbances will have the same lingering effect as past receipts of income. Thus the consumer navigates a course which could be compared to that of a big boat in a rough sea steered by a loose hand.

We might imagine an economy in which some consumers act in the manner depicted by the permanent-income hypothesis whilst others act according to the partial-adjustment hypothesis. We might also imagine that the coefficient α of partial adjustment differs from the parameter γ of the expectations mechanism. Then we should derive an aggregate consumption function which is more adequately represented by the RTM of (6) than by either of the constituent models. Moreover, if we were doubtful of the nature of the consumption behaviour, then we should favour the RTM on the grounds that its specification does not prejudge the issue.

In the ensuing sections, we shall show that, if an ADM is used when an RTM is appropriate, then, in some circumstances, a very distorted picture can be given of the dynamic structure of a economic relationship. The examples which will be used to demonstrate the point are necessarily of a stylised nature.

3. The Fitting of a Parsimonious ADM

Let us imagine that the true model, which accurately represents the processes generating our data, is a simple RTM in the form of

$$y(t)=\frac{\delta}{1-\gamma L}x(t)+\frac{1}{1-\varphi L}\varepsilon(t), \tag{9}$$

where $|\gamma|$, $|\varphi|<1$ to ensure stability. For comparison with the ADM of equation (1), we might write this as

$$\{1-(\gamma+\varphi)L+\gamma\varphi L^2\}y(t)=(\delta-\delta\varphi L)x(t)+(1-\gamma L)\varepsilon(t). \tag{10}$$

The latter is a structured version of the equation

$$(1-\alpha_1 L-\alpha_2 L)y(t)=(\beta_0+\beta_1 L)x(t)+(1-\mu L)\varepsilon(t) \tag{11}$$

of a General Temporal Model or a GTM, which is also described as an ARMAX model.

To complete the description of $y(t)$, we also need to specify how the signal $x(t)$ is generated. We shall assume, for simplicity, that $x(t)$ comes from a stationary first-order autoregressive process described by the equation

$$(1-\pi L)x(t)=\xi(t), \tag{12}$$

where $\xi(t)$ is a white-noise process, and where $|\pi|<1$. On occasion, we shall set $\pi=0$.

Our first task will be to analyse the effects of fitting a parsimonious ADM in the form of

$$(1-\alpha L)y(t)=(\beta_0+\beta_1 L)x(t)+e(t). \tag{13}$$

This is a specialised version of the ADM of (1) which arises when $\alpha_2=0$.

The usual criterion for fitting such a model is to minimise the sum of squares of the residuals which form the sequence $e(t)$. Since we know that $y(t)$ and $x(t)$ are actually generated by these equations, we can substitute (9) and (12) into (13). After some rearrangement, we get the following expression for the residual sequence:

$$(14) \quad e(t)=(1-\alpha L)\left\{\frac{\delta}{1-\gamma L}-\frac{\beta_0+\beta_1 L}{1-\alpha L}\right\}\frac{1}{1-\pi L}\xi(t)+\frac{1-\alpha L}{1-\varphi L}\varepsilon(t)$$
$$=p(L)\xi(t)+q(L)\varepsilon(t).$$

Here $p(L)=\{p_0+p_1 L+p_2 L^2+\ldots\}$ and $q(L)=\{1+q_1 L+q_2 L^2+\ldots\}$ are the infinite series which come from expanding the rational functions of the lag operator L associated with $\xi(t)$ and $\varepsilon(t)$ respectively.

In the limit, when the sample size T becomes indefinitely large, the sum of squares of the residual sequence scaled by T^{-1} tends to $V\{e(t)\}$, which is the variance of $e(t)$. This is the consequence of the law of large numbers.

From the assumption that $\xi(t)$ and $\varepsilon(t)$ are uncorrelated white-noise processes with $V\{\varepsilon(t)\}=\sigma_\varepsilon^2$ and $V\{\xi(t)\}=\sigma_\xi^2$, it follows that

$$(15) \quad V\{e(t)\}=\sigma_\xi^2\sum p_i^2+\sigma_\varepsilon^2\sum q_i^2$$
$$=S(\alpha, \beta_0, \beta_1).$$

Our object is to find the probability limits of the estimated values of the parameters α, β_0, β_1 of the fitted equation under (13). It follows from an basic theorem, which is proved, for example, by AMEMIYA and by DOMOWITZ and WHITE, that the probability limits are simply the values which minimise the function $V\{e(t)\}$ which is the asymptotic form of the criterion function.

Using the methods which are described in the appendix, we can show that the asymptotic form of the criterion function is

$$(16) \quad S(\alpha, \beta_0, \beta_1)=\sigma_\xi^2\left\{(\delta-\beta_0)^2+\frac{W^2}{1-\pi^2}+\frac{D^2\gamma^2}{1-\gamma^2}+\frac{2DW\gamma}{1-\pi\gamma}\right\}$$
$$+\sigma_\varepsilon^2\left\{\frac{(\alpha-\varphi)^2}{1-\varphi^2}+1\right\},$$

where

$$(17) \quad W=(C-\beta_0)\pi-\beta_1,$$

$$C=\frac{\delta(\pi-\alpha)}{\pi-\gamma} \text{ and}$$

$$D=\frac{\delta(\alpha-\gamma)}{\pi-\gamma}.$$

By differentiating this with respect to β_0 and β_1 and setting the results to zero, we discover conditions from which we can deduce that

$$(18) \qquad \beta_0 = \delta \text{ and}$$
$$\beta_1 = \frac{\delta(\gamma - \alpha)}{1 - \gamma\pi}.$$

When these are substituted back into the criterion function, we obtain a concentrated function in the form of

$$(19) \qquad S(\alpha) = \sigma_\xi^2 \frac{\delta^2 \gamma^2 (\alpha - \gamma)^2}{(1-\gamma^2)(1-\pi\gamma)^2} + \sigma_\varepsilon^2 \left\{ \frac{(\alpha - \varphi)^2}{1 - \varphi^2} + 1 \right\}.$$

By differentiating $S(\alpha)$ with respect to α and setting the result to zero, we discover a condition from which we deduce that

$$(20) \qquad \alpha = \frac{\kappa\gamma + \lambda\varphi}{\kappa + \lambda}, \quad \text{where}$$
$$\kappa = \frac{\sigma_\xi^2 \delta^2 \gamma^2}{(1-\gamma^2)(1-\pi\gamma)^2} \quad \text{and} \quad \lambda = \frac{\sigma_\varepsilon^2}{1-\varphi^2}.$$

An inspection of equation (20) shows that α is formed as a convex combination of the systematic parameter γ and the disturbance parameter φ which belong to the RTM which actually generates $y(t)$. We can see that, if $\gamma = \varphi$, then we shall have $\alpha = \gamma = \varphi$, $\beta_0 = \delta$ and $\beta_1 = 0$; and so the ADM will provide consistent estimates of the parameters of the process. However, if γ and φ differ markedly in value, then the value of α will succeed in representing neither of them; and the fitted model may give a very inaccurate representation of the true process. In such cases, the value of the weights κ and λ play a crucial role in determining α.

We can recognise immediately that the value of λ is just the variance of the disturbance part of the RTM. The value of κ is closely related to the variance of the systematic part of the RTM. The latter is given, in fact, by

$$(21) \qquad V\left\{ \frac{\delta}{(1-\gamma L)(1-\pi L)} \xi(t) \right\} = \frac{\sigma_\xi^2 \delta^2 (1+\gamma\pi)}{(1-\gamma^2)(1-\gamma\pi)(1-\pi^2)}$$
$$= \kappa \frac{(1+\gamma\pi)}{\gamma^2(1-\pi^2)}.$$

The variance of the signal is given by $\sigma_x^2 = \sigma_\xi^2/(1-\pi^2)$; and it is clear that an increase in the signal-to-noise ratio $\sigma_x^2/\sigma_\varepsilon^2$ will increase the weight which is attributed to the systematic parameter γ. Also, the value of κ is seen to depend crucially on whether or not γ and π share the same sign. Thus, when $\gamma\pi \to 1$, we find that $\kappa \to \infty$; whereas, when $\gamma\pi \to -1$, we find that κ remains small whilst the variance of the systematic component tends to zero. However, it should be recognised that the value of κ is undefined at the point where $\gamma\pi = 1$. Indeed, the criterion function S is undefined at the values $\varphi = \pm 1$, $\gamma = \pm 1$ and $\pi = \pm 1$, all of which have been excluded by assumption.

We can summarise matters roughly by saying that whenever γ and φ are at odds, they are liable to engage in a struggle to preempt the value of α. The outcome of this struggle will depend upon the relative power of the systematic and disturbance parts of the RTM.

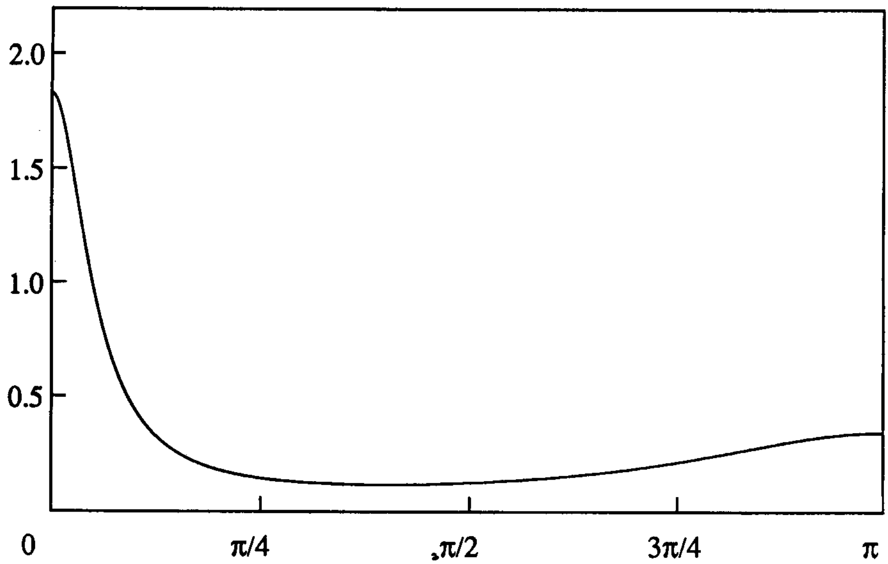

FIGURE 1

The spectral density function of the ARMA (2, 1) process $(1-0.45\,L-0.34\,L^2)\,y(t)=(1-0.39\,L)\zeta(t)$ ***when*** $V\{\zeta(t)\}=1.37$.

To illustrate these matters, let us assume that the RTM, which truly describes how $y(t)$ is generated, has the parameter values $\delta = 1$, $\gamma = 0.85$ and $\varphi = -0.4$. We shall attribute a range of values to the autoregressive parameter π which characterises the process generating the signal. However, in the case of $\pi = 0$, $V\{\xi(t)\} = V\{x(t)\} = 0.25$ and $V\{\varepsilon(t)\} = 0.75$, the dependent variable $y(t)$ follows an ARMA (2, 1) process described by the equation

$$(22) \qquad (1-0.45\,L-0.34\,L^2)\,y(t)=(1-0.39\,L)\,\zeta(t),$$

where $\zeta(t)$ is a white-noise process with $V\{\zeta(t)\}=1.37$. The spectral density function is given in Figure 1.

The parameters of this process are virtually the same as those of an ARMA (2,1) model which GRANGER and NEWBOLD have fitted to the first differences of an index compiled from "help wanted" advertisements. Some of the results from fitting the ADM to this and other processes are reported in Table 1.

Perhaps the most startling aspect of these results concerns the value of the steady-state gain or long-term multiplier of the transfer function of the systematic part of the fitted model. The true value of the multiplier is given by $\mu = \delta/(1-\gamma)$, whilst the estimated value is given by $(\beta_0 + \beta_1)/(1-\alpha)$. In all cases, the latter has a severe downward bias.

The problem with the value of the estimated multiplier is mitigated somewhat when the signal $x(t)$ has a strong positive autocorrelation, as it does when the value of π is positive and close to unity. In case D, for example, where $\pi = 0.9$, the low-frequency signal $x(t) = \xi(t)/(1-0.9\,L)$ is being strongly amplified by the lowpass filter $\beta/(1-\gamma\,L) = 1/(1-0.85\,L)$ with the effect that both the variance of the systematic component and the value of κ are large. The consequence is that value of α is tending towards that of the systematic parameter γ and away from that of the disturbance parameter φ. Even so, the value of the multiplier is seriously underestimated.

TABLE 1

The effects of fitting the model $(1-\alpha\,L)\,y(t) = (\beta_0 + \beta_1\,L)\,x(t) + e(t)$ ***when the true relationship is*** $y(t) = (1-0.85\,L)^{-1}\,x(t) + (1+0.4\,L)^{-1}\,\varepsilon(t)$ ***and*** $x(t) = (1-\pi\,L)^{-1}\,\xi(t)$ ***is a first-order autoregressive process.***

	Case A	Case B	Case C	Case D
π	−0.30	0.00	0.60	0.90
σ_x^2	0.250	0.250	0.250	0.250
σ_ξ^2	0.228	0.250	0.160	0.048
σ_ε^2	0.750	0.750	0.750	0.750
α	−0.030	0.127	0.425	0.494
β_0	1.000	1.000	1.000	1.000
β_1	0.701	0.723	0.867	1.516
S	1.163	1.338	1.671	1.747
True Variances				
Systematic	0.535	0.901	2.776	6.766
Disturbance	0.893	0.893	0.893	0.893
Sum	1.428	1.794	3.669	7.659
Estimated Variances				
Systematic	0.263	0.434	1.629	5.349
Disturbance	1.164	1.360	2.040	2.311
Multipliers				
True	6.667	6.667	6.667	6.667
Estimated	1.652	1.974	3.248	4.970

The amplification of the systematic variance, which is due to the lowpass filtering of low-frequency signals, has doubtless been responsible for saving many exercises in applied econometrics from the worst pitfalls of misspecification. However, there are also disadvantages associated with a strong positive autocorrelation of the signal. For, as π approaches unity, the moment matrix of the signal $x(t) = \xi(t)/(1-\pi\,L)$ becomes increasingly ill-conditioned with the consequence that, in finite samples, the estimates of the model's parameters become increasingly ill-determined.

In order to obtain well-determined estimates of the parameters of a transfer function, we usually require not only that the signal-to-noise ratio should be large but also that the signal itself should possess a sufficiently wide range of frequencies to excite all of the resonant modes of the system. Thus, an ideal signal for the purpose of model identification, which can sometimes be generated in an experimental environment, is one with a uniform spectral density function. In the present case, however, we find that, as $\pi \to 1$, the spectrum of the signal process $x(t)$ is increasingly concentrated in the vicinity of the zero frequency. This phenomenon is the spectral counterpart of the increasingly ill-conditioned moment matrix.

The problems of estimation which arise when $\pi = 1$ require a special treatment of the sort which has been accorded recently to systems of cointegrated variables. See, for example, ENGLE and GRANGER.

If the signal process does comprise an autoregressive operator with a unit root, then we can estimate the value of the multiplier $\mu = \delta/(1-\gamma)$ consistently by running a simple regression of $y(t)$ on $x(t)$. Given the value of μ, which is described, in this context, as a cointegrating factor, we can proceed to estimate the parameter $\theta = \gamma - 1$ by applying the technique of simple regression to the equation

$$(1-\mathrm{L})y(t) = \theta\{y(t-1) - \mu x(t)\} + \zeta(t), \tag{23}$$

which is another form of equation (9). The efficiency of the estimate of θ may be improved by taking account of the structure of the disturbance term $\zeta(t) = \{(1-\gamma\mathrm{L})/(1-\varphi\mathrm{L})\}\varepsilon(t)$.

As STOCK [12] has shown, the estimator of the multiplier μ which is obtained from the cointegrating regression benefits from the property of superconsistency whereby its convergence to the true value is at the rate of $1/\mathrm{T}$ rather than the usual rate of $1/\sqrt{\mathrm{T}}$.

4 Fitting an Extended ADM and a CFM

The distortions in our estimates can be reduced by adding extra parameters to the fitted ADM. Consider writing equation (9), which represents the true RTM, in the form of

$$(1-\varphi\mathrm{L})y(t) = \frac{\delta(1-\varphi\mathrm{L})}{1-\gamma\mathrm{L}}x(t) + \varepsilon(t). \tag{24}$$

Given that $|\gamma|<1$, there exists a series expansion of the rational function $\delta(1-\varphi L)/(1-\gamma L)$ comprising a convergent sequence of coefficients. By truncating the series, we derive an ADM in the form of

$$(25) \qquad (1-\varphi L)y(t)=(\beta_0+\beta_1 L+\ldots+\beta_k L^k)x(t)+\varepsilon(t).$$

If γ is close to zero, then such a model should provide a reasonable approximation to the RTM, even with a small value of k. However, if $|\gamma|$ is close to unity, then the number of coefficients in the distributed-lag operator $\beta(L)$ which is necessary to ensure a reasonable approximation is likely to be unacceptably large. Moreover, given the typically ill-conditioned nature of the empirical moment matrix associated with $x(t)$, the estimates of these coefficients $\beta_0, \beta_1, \ldots, \beta_k$ are liable to be ill determined.

Another way of seeking a model which fits better than the ADM of equation (13) is to add extra parameters to the polynomial $\alpha(L)$. In this section, we shall consider adding one extra parameter to $\alpha(L)$. This leads to the model of equation (4). We shall also investigate the effect of forcing a common-factor restriction upon this equation so as to obtain the equation (5). In that case, our question will be whether or not the inappropriate presence of the factor $1-\alpha L$ in the disturbance part of the CFM seriously affects the ability of this model to approximate the RTM of (9).

The residual sequence from fitting the CFM is given by

$$(26) \quad \begin{aligned} e(t)&=(1-\alpha L)(1-\rho L)\left\{\frac{\delta}{1-\gamma L}-\frac{\beta}{1-\alpha L}\right\}x(t)+\frac{(1-\alpha L)(1-\rho L)}{1-\varphi L}\varepsilon(t)\\ &=p(L)x(t)+q(L)\varepsilon(t). \end{aligned}$$

On the assumption that $x(t)=\xi(t)$ is a white-noise process, the asymptotic form of the criterion function is

$$(27) \quad \begin{aligned} S(\alpha,\beta,\rho)=\sigma_\xi^2&\left\{(\delta-\beta)^2+\{\delta(\gamma-\alpha)+\rho(\beta-\delta)\}^2+\frac{\{\delta(\gamma-\alpha)(\gamma-\rho)\}^2}{1-\gamma^2}\right\}\\ &+\sigma_\varepsilon^2\left\{1+(\varphi-\alpha-\rho)^2+\frac{\{(\varphi-\alpha)(\varphi-\rho)\}^2}{1-\varphi^2}\right\}. \end{aligned}$$

The problem of finding the values which minimise the criterion function is no longer straightforward; and there are no closed-form expressions for these values. However, we can decompose the problem into two simple problems which are linked sequentially. The first is to find the values of $\alpha=\alpha(\rho)$ and $\beta=\beta(\rho)$ which minimise the conditional function $S(\alpha, \beta|\rho)$ in which the disturbance parameter ρ is held constant. The second is to find the value of $\rho=\rho(\alpha, \beta)$ which minimises the function $S(\rho|\alpha, \beta)$ in which α and β are held constant. In the appendix, we present the normal equations which provide these various minimising values. We can attempt to find the values which minimise $S(\alpha, \beta, \rho)$ unconditionally by applying the well-know Cochrane-Orcutt iterative procedure, for which the rth iteration is specified by

(28) $$\alpha_{(r)} = \alpha\{\rho_{(r-1)}\},\ \beta_{(r)} = \beta\{\rho_{(r-1)}\},$$
$$\rho_{(r)} = \rho\{\alpha_{(r)},\ \beta_{(r)}\}.$$

OBERHOFER and KMENTA have demonstrated that the Cochrane-Orcutt procedure is bound to converge.

There is no guarantee, in general, that the function $S(\alpha, \beta, \rho)$ will have a unique minimum or that the Cochrane-Orcutt iteration will have a unique fixed point. The matter of uniqueness depends upon the precise values assumed by the RTM parameters δ, γ and φ. However, we can easily assess the number of minima by plotting the concentrated function $S(\rho) = S\{\alpha(\rho), \beta(\rho), \rho\}$ which is obtained from $S(\alpha, \beta, \rho)$ by putting the relevant estimating equations for α and β in place of these arguments.

For an illustration, we shall assume, as before, that $\delta = 1$, $\gamma = 0.85$ and $\varphi = -0.4$. We shall set $\pi = 0$ so as to make $x(t)$ a white-noise sequence. Figure 2, which is the graph of the function $S(\rho)$ for the case where $\sigma_\varepsilon^2 = 0.75$ and $\sigma_x^2 = 0.25$, reveals two minima which occur at the points $\rho = -0.664$ and $\rho = 0.689$.

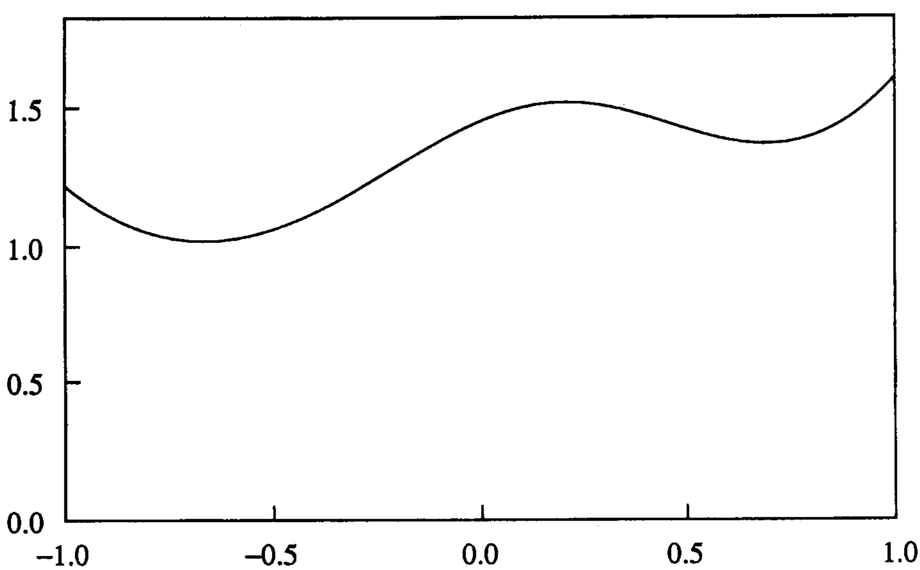

FIGURE 2

The graph of the function $S(\rho)$ ***associated with fitting the CFM*** $(1-\rho L)(1-\alpha L)\ y(t) = (1-\rho L)\beta x(t) + e(t)$ ***when the true relationship is*** $y(t) = (1-\gamma L)^{-1}\delta x(t) + (1-\varphi L)^{-1}\varepsilon(t)$, ***where*** $\gamma = 0.85$, $\varphi = -0.4$, $\delta = 1$ ***and where*** $x(t)$ ***and*** $\varepsilon(t)$ ***are white-noise process with*** $V\{x(t)\} = 0.25$ ***and*** $V\{\varepsilon(t)\} = 0.75$.

The full set of parameter values which correspond to these minima are presented in Table 2 under the headings CFM (i) and CFM (ii) respectively. Also displayed in the table are the parameter values which come from fitting the ADM of equation (1). The parameter values of the true RTM are displayed, in the appropriate locations, under the following guise:

$$\begin{aligned} &\alpha_1=\gamma+\varphi, &\quad \alpha_2=-\gamma\varphi, \\ &\beta_0=\delta, &\quad \beta_1=-\delta\varphi, \\ &\alpha=\gamma, &\quad \beta=\delta. \end{aligned} \tag{29}$$

We obtain these equalities by comparing the form of the RTM equation given under (10) with that of the ADM under (1), and by comparing the form of the RTM equation given under (9) with that of the CFM under (5).

TABLE 2

The effects of fitting the ADM model $(1-\alpha_1 L-\alpha_2 L^2)\ y(t)=(\beta_0+\beta_1)\ x(t)+e(t)$ ***and the CFM model*** $(1-\rho L)(1-\alpha L)\ y(t)=(1-\rho L)\beta x(t)+e(t)$ ***when the true relationship is*** $y(t)=(1-0.85\ L)^{-1}\ x(t)+(1+0.4\ L)^{-1}\ \varepsilon(t)$ ***and*** $x(t)$ ***is a white-noise process.***

	RTM	ADM	CFM (i)	CFM (ii)
σ_x^2	0.25	0.25	0.25	0.25
σ_ε^2	0.75	0.75	0.75	0.75
α_1	0.450	0.011	0.016	0.151
α_2	0.340	0.440	0.452	0.371
β_0	1.000	1.000	1.078	0.352
β_1	0.4	0.839	0.716	−0.242
α	0.850	–	0.680	−0.538
β	1.000	–	1.078	0.352
ρ	–	–	−0.664	0.689
S	0.750	1.012	1.017	1.369
Multipliers				
True	6.667	6.667	6.667	6.667
Estimated	–	3.348	3.374	0.229

The addition of an extra parameter to the ADM has clearly improved the estimate of the multiplier; for the value of 3.348 compares favourably with the value of 1.974 which is to be found under case B in Table 1. However, the estimated multiplier is still remote from the true value.

The first of the common-factor models, CFM (i), is surprisingly similar to the ADM. This suggests that the common-factor restrictions are liable to be accepted quite readily by statistical tests which use the misspecified ADM for the alternative hypothesis. The second of the common-factor models, CFM (ii), is vastly different, and it fails to capture any of the characteristics of the RTM which generates $y(t)$. If an investigator were in a position to compare the two common-factor models, then he would certainly reject CFM (ii) in favour of CFM (i) on the grounds that latter is associated with the lesser value of S, which is the variance of the residual sequence.

An interesting indication of the failure of the various models is provided by the frequency spectra of the residual sequences. The spectra of the ADM and of the first of the common-factor models are virtually identical, and only the former is shown. The spectrum of the residual sequence from the second of the common-factor models retains some of the characteristics of the spectrum of the $y(t)$ which was shown in Figure 1. A successful model should generate a residual sequence which has the characteristics of white noise. The frequency spectrum of white noise is flat.

Figures 3 and 4 point to the fact that, given a sufficiency of data, a careful investigator should be able to detect the misspecification via tests which compare the properties of the sequence of residuals to those of a white-noise sequence. An appropriate test which is correctly calibrated, at least in large samples is a version of the lagrange-multiplier test of BREUSCH and GODFREY which bases the null hypothesis upon the ADM of equation (1) and the alternative hypothesis upon the GTM of equation (11).

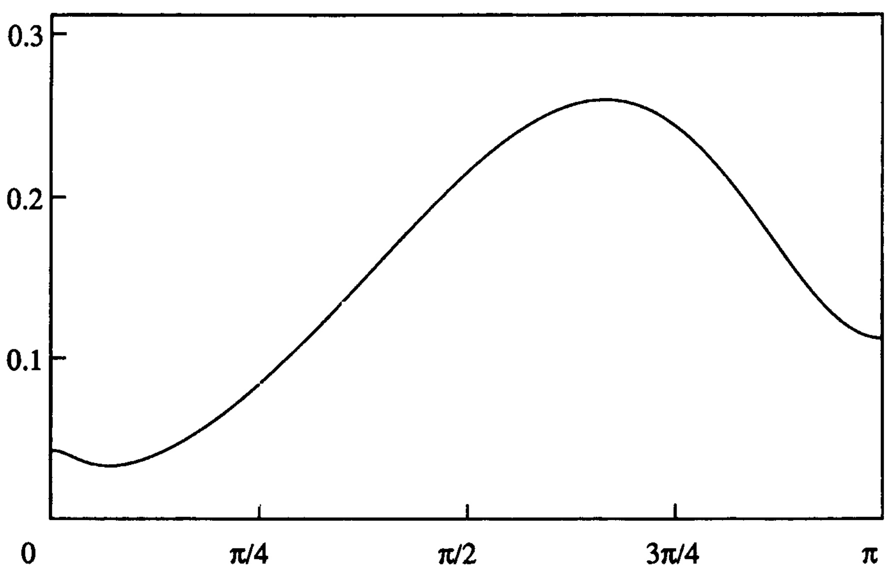

FIGURE 3

The spectrum of the residual sequence from fitting the model under ADM in Table 2.

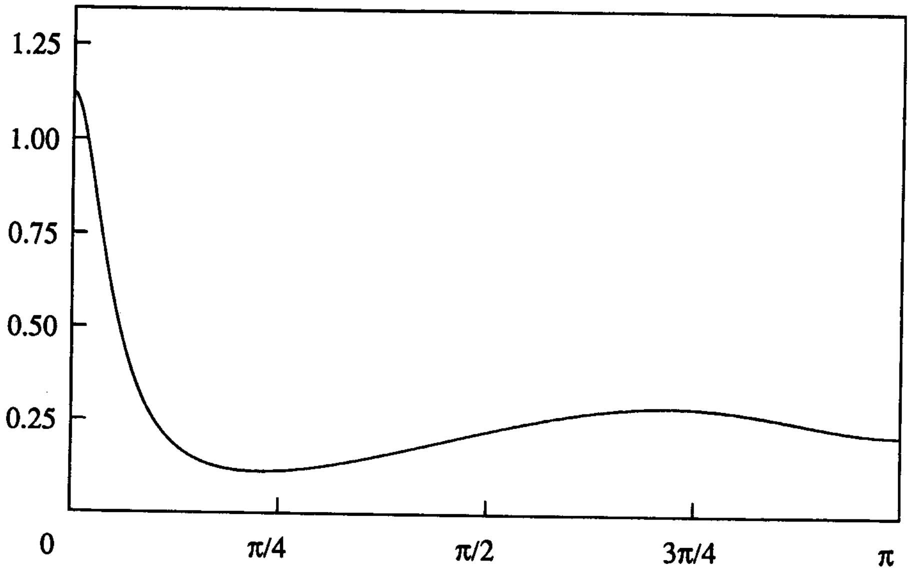

FIGURE 4

The spectrum of the residual sequence from fitting the model under CFM (ii) in Table 2.

5 The Robust Nature of the RTM

The failures which we have witnessed in the foregoing sections have resulted from the fact that the two sides of the RTM are vying for the same parameters of the fitted ADM or CFM. A way of overcoming the problem is to attribute separate sets of parameters to the systematic and disturbance parts of the fitted models. In that case, our estimates of the systematic parameters may still be viable even if we are ignorant or misinformed about the nature of the disturbance process in the true regression relationship.

Let us assume, for the sake of argument, that the true relationship is described by the common-factor model of equation (2), and let us imagine that the RTM of equation (9) is the fitted model. Then the expression for the residual sequence is

$$(30)\quad \begin{aligned} e(t) &= (1-\varphi L)\left\{\frac{\beta}{1-\alpha L} - \frac{\delta}{1-\gamma L}\right\}x(t) + \frac{1-\varphi L}{(1-\alpha L)(1-\rho L)}\varepsilon(t) \\ &= p(L)\,x(t) + q(L)\,\varepsilon(t). \end{aligned}$$

The variance of $e(t)$, which coincides with the asymptotic form of the least-squares criterion function, is given by

$$V\{e(t)\}=V\{p(L)x(t)\}+V\{q(L)\varepsilon(t)\}, \tag{31}$$

where

$$V\{q(L)\varepsilon(t)\}=\sigma_\varepsilon^2\frac{(1+\alpha\rho)(1+\varphi^2)-2\varphi(\alpha+\rho)}{(1-\alpha^2)(1-\alpha\rho)(1-\rho^2)}. \tag{32}$$

The function $V\{p(L)x(t)\}$ attains its minimum value of zero when $\delta=\beta$ and $\gamma=\alpha$. The result is the same regardless of the value of the parameter φ which belongs to the disturbance part of the RTM. The function $V\{q(L)\varepsilon(t)\}$, which contains none of the parameters from the systematic part of the RTM, attains a unique minimum value when

$$\varphi=\frac{\alpha+\rho}{1+\alpha\rho}. \tag{33}$$

The RTM fails to reflect the dynamic properties of the disturbance of the CFM; but this failure will not affect the consistency of the estimates of the systematic parameters.

It is interesting to consider the effects of using the Cochrane-Orcutt procedure to fit a CFM which is indeed correctly specified. It is quite possible that, given inappropriate starting values, the procedure will converge to a set of inconsistent estimates.

For the sake of an emphatic illustration, let us assume that parameter values in the true equation (2) are $\alpha=0.85$, $\beta=1$ and $\rho=-0.85$, and, assuming that $\varepsilon(t)$ and $x(t)$ are white noise, let us set $\sigma_\varepsilon^2=0.75$ and $\sigma_x^2=0.25$. The equation of the fitted CFM, which has the same form as equation (2), may be denoted by

$$(1-\varphi L)(1-\gamma L)=(1-\varphi L)\delta x(t)+e(t), \tag{34}$$

and the asymptotic form of the criterion function, which is the variance of $e(t)$, may be denoted by $S(\gamma, \delta, \varphi)$. The graph of the concentrated function $S(\varphi)=S\{\gamma(\varphi), \delta(\varphi), \varphi\}$ is shown in Figure 5. There are minima at the values $\varphi=-0.85$ and $\varphi=0.879$. The values of γ, δ corresponding to the first of these minima are equal respectively to those of the parameters α and β of the underlying CFM. The values which correspond to the minimum at $\varphi=0.879$ are $\gamma=-0.771$ and $\delta=0.879$.

We have seen that fitting a CFM is a hazardous business even when it does correspond to the process underlying the data. However, the CFM must be regarded as a synthetic model rather than a natural one; for it is difficult to imagine a physical or a social process which would suggest equation (2) in the first instance.

In their celebrated article, Hendry and Mizon supported the use of the CFM with the claim that it represents a convenient simplification of an ADM. In this paper, we have cast doubt on the universal applicability of the ADM; and we have suggested that, even when the ADM is appropriate, the possibility that the Cochrane-Orcutt iteration has more than one fixed point implies that the CFM is a model which should be treated with caution.

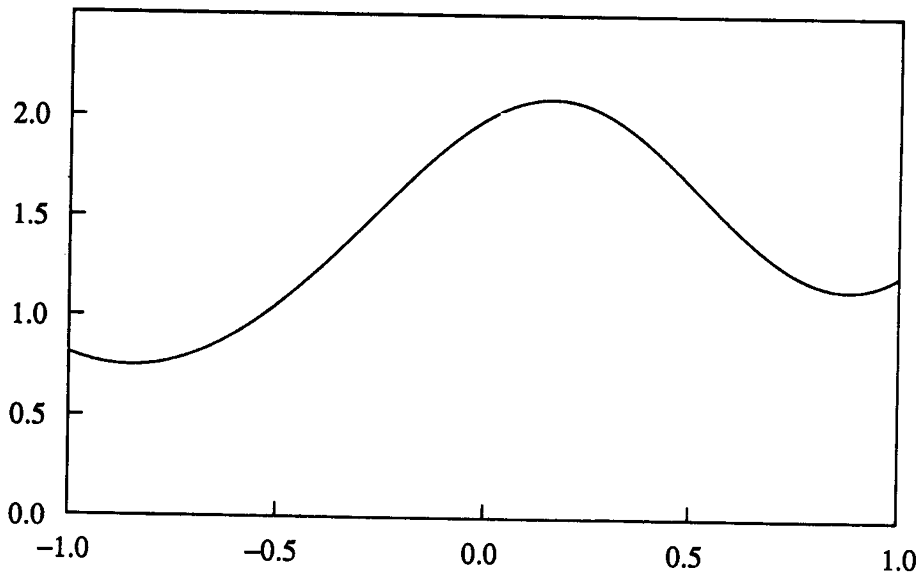

FIGURE 5

The graph of the function $S(\varphi)$ ***associated with fitting a correctly specified CFM when the true equation*** $(1-\rho L)(1-\alpha L)\,y(t)=(1-\rho L)\,\beta\,x(t)+\varepsilon(t)$ ***has*** $\rho=-0.85$, $\alpha=0.85$ ***and*** $\beta=1.0$ ***and when*** $x(t)$ ***and*** $\varepsilon(t)$ ***are white-noise processes with*** $V\{x(t)\}=0.25$ ***and*** $V\{\varepsilon(t)\}=0.75$.

The Cross-Covariances of ARMA Processes

Throughout the paper, we face the problem of evaluating the variance of a residual sequence which is expressible, in general, as a sum of ARMA processes. Therefore we have to find the autocovariances and the cross-covariances of these processes. For this purpose it is efficient to work with the corresponding covariance generating functions. Consider the processes $\alpha(L)y(t)=\beta(L)\varepsilon(t)$ and $\varphi(L)x(t)=\theta(L)\varepsilon(t)$ where $\varepsilon(t)$ is a white-noise sequence. Define

(35)
$$\alpha(z)=1+\alpha_1 z+\ldots+\alpha_p z^p=\prod_{i=1}^{p}(1-\lambda_i z),$$
$$\varphi(z)=1+\varphi_1 z+\ldots+\varphi_f z^f=\prod_{i=1}^{f}(1-\kappa_i z),$$
$$\beta(z)=1+\beta_1 z+\ldots+\beta_q z^q=\prod_{i=1}^{q}(1-\mu_i z),$$
$$\theta(z)=1+\theta_1 z+\ldots+\theta_h z^h=\prod_{i=1}^{h}(1-\nu_i z).$$

Then the generating function for cross-covariances of $y(t)$ and $x(t)$ is

(36)
$$\gamma(z)=\sigma_\varepsilon^2\frac{\beta(z^{-1})\theta(z)}{\alpha(z^{-1})\varphi(z)}.$$

The cross-covariance at lag τ of $y(t)$ and $x(t)$ is the coefficient associated with z^τ in the Laurent expansion of $\gamma(z)$. Unless $\beta(z^{-1})/\alpha(z^{-1})$ and $\theta(z)/\varphi(z)$ are both proper rational functions, it is easiest to expand the numerator and denominator separately and then to form their product.

The partial-fraction expansion of $\alpha^{-1}(z^{-1})$ is given by

(37)
$$\frac{1}{\alpha(z^{-1})}=\frac{C_1}{1-\lambda_1 z^{-1}}+\ldots+\frac{C_p}{1-\lambda_p z^{-1}},$$

where the generic coefficient is

(38)
$$C_k=\frac{\lambda_k^{p-1}}{\prod_{i\neq k}(\lambda_k-\lambda_i)}.$$

Likewise, for $\varphi^{-1}(z)$, we have

(39)
$$\frac{1}{\varphi(z)}=\frac{D_1}{1-\kappa_1 z}+\ldots+\frac{D_f}{1-\kappa_f z}.$$

It follows that the denominator of $\gamma(z)$ is

(40)
$$\frac{1}{\alpha(z^{-1})\varphi(z)}=\sum_k\sum_l \frac{C_k D_l}{(1-\lambda_k z^{-1})(1-\kappa_l z)}.$$

This expression may be evaluated using the result that

(41)
$$\frac{C_k D_l}{(1-\lambda_k z^{-1})(1-\kappa_l z)} = \frac{C_k D_l}{(1-\lambda_k \kappa_l)}\left\{\ldots+\frac{\lambda_k^2}{z^2}+\frac{\lambda_k}{z}+1+\kappa_l z+\kappa_l^2 z^2+\ldots\right\}.$$

To find an expression for the numerator of $\gamma(z)$, we use

(42)
$$\left(\sum_{i=0}^{q}\beta_i z^{-i}\right)\left(\sum_{j=0}^{h}\theta_j z^j\right)=\sum_{j=-q}^{h}\left(\sum_{k=m}^{n}\beta_k\theta_{j-k}\right)z^j,$$

where $m=\max(0, j-h)$ and $n=\min(q, j)$.

Normal Equations for the Cochrane-Orcutt Procedure

In the paper we use the Cochrane-Orcutt procedure to find the probability limits of the estimates of the parameters of the equation

(43)
$$(1-\rho L)(1-\alpha L)y(t)=(1-\rho L)\beta x(t)+e(t)$$

when the true relationship is given by

(44)
$$y(t)=\frac{\delta}{1-\gamma L}x(t)+\frac{1}{1-\varphi L}\varepsilon(t).$$

We take $x(t)$ and $\varepsilon(t)$ to be white-noise processes with $V\{x(t)\}=\sigma_x^2$ and $V\{\varepsilon(t)\}=\sigma_\varepsilon^2$ respectively.

Given the value of ρ, the values of α and β may be found by solving an equation in the form of

(45)
$$\begin{bmatrix} P & Q \\ R & S \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix}=\begin{bmatrix} V \\ W \end{bmatrix}.$$

The elements of the equation are

(46)
$$P=\sigma_x^2\delta^2\left\{1+\frac{(\gamma-\rho)^2}{1-\gamma^2}\right\}+\sigma_\varepsilon^2\delta^2\left\{1+\frac{(\varphi-\rho)^2}{1-\varphi^2}\right\},$$
$$Q=R-\sigma_x^2\delta\rho,$$
$$S=\sigma_x^2(1+\rho^2),$$

$$V = \sigma_x^2 \delta^2 \left\{ (\gamma - \rho) + \frac{\gamma (\gamma - \rho)^2}{1 - \gamma^2} \right\} + \sigma_\varepsilon^2 \left\{ (\varphi - \rho) + \frac{\varphi (\varphi - \rho)^2}{1 - \varphi^2} \right\},$$

$$W = \sigma_x^2 \{ \delta - \delta \rho (\gamma - \rho) \}.$$

Given the values of α and β, the value of ρ may be found as

(47) $$\rho = \frac{H}{G},$$

where

(48) $$G = \sigma_x^2 \delta^2 \left\{ (\delta - \beta)^2 + \frac{\delta^2 (\gamma - \alpha)^2}{1 - \gamma^2} \right\} + \sigma_\varepsilon^2 \left\{ 1 + \frac{(\varphi - \alpha)^2}{1 - \varphi^2} \right\},$$

$$H = \sigma_x^2 \delta^2 \left\{ \delta (\gamma - \alpha)(\delta - \beta) + \frac{\delta^2 \gamma (\gamma - \alpha)^2}{1 - \gamma^2} \right\} + \sigma_\varepsilon^2 \left\{ (\varphi - \alpha) + \frac{\varphi (\varphi - \alpha)^2}{1 - \varphi^2} \right\}.$$

• References

AMEMIYA, T. – *Advanced Econometrics*. Cambridge: Basil Blackwell, 1985.

BREUSCH, T. S. – « Testing for autocorrelation in dynamic linear models ». *Australian Economic Papers*, 17, (1978), pp. 334-355.

COCHRANE, D. and ORCUTT, G. H. – « Application of least-squares regressions to relationships containing autocorrelated error terms ». *Journal of the American Statistical Association*, 44, (1949), pp. 32-61.

DOMOWITZ, I. and WHITE, H. – « Misspecified Models Dependent Observations ». *Journal of Econometrics*, 20, (1982), pp. 35-58.

ENGLE, R. F. and GRANGER, C. W. J. – *Long-Run Economic Relationships: Readings in Cointegration*. Oxford: Oxford University Press, 1991.

FRIEDMAN, M. – *A Theory of the Consumption Function*. Princeton N.J.: Princeton University Press, 1957.

GODFREY, L. G. – « Testing for Higher Order Serial Correlation ın Regression equations when the Regressors Include Lagged Dependent Variables ». *Econometria*, 46, (1978) pp. 1303-1310.

GRANGER, C. W. J. and NEWBOLD, P. – *Forecasting Economic Time Series*. New York: Academic Press, 1977.

HENDRY, D. F. and MIZON, G. – « Serial Correlation as a Convenient Simplification and not a Nuisance: a Comment on a Study of the Demand for Money by the Bank of England ». *Economic Journal*, 88, (1978), pp. 549-563.

OBERHOFER, W. and KMENTA, J. – « A General Procedure for Obtaining Maximum Likelihood Estimates in Generalised Regression Models ». *Econometrica*, 42, (1974), pp. 579-590.

RICE, S. O. – « Noise in FM Receivers ». In Rosenblatt, M. (ed.), *Time Series Analysis*, pp. 395-422. New York: Wiley and John, Sons, 1963.

STOCK, J. – « The Asymptotic Properties of Least-Squares Estimators of Cointegrating Vectors ». *Econometrica*, 55, (1987), pp. 1035-1056.

WRIGHT, J. R. — « Flutter Test Analysis in the Time Domain Using a Recursive System Representation ». *Journal of Aircraft*, 11, (1974), pp. 774-777.