FACULTY ~~INSTITUTE~~ OF ACTUARIAL SCIENCE & ECONOMETRICS

Report AE 6/89

LAGGED DEPENDENT VARIABLES DISTRIBUTED LAGS AND AUTOREGRESSIVE RESIDUALS

D.S.G. Pollock

University of Amsterdam

LAGGED DEPENDENT VARIABLES DISTRIBUTED LAGS AND AUTOREGRESSIVE RESIDUALS

by

D.S.G. POLLOCK*

Whenever the specification of a dynamic regression relationship is in doubt, we should adopt 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.

* *This paper was written while I was enjoying the hospitality of the Faculty of Actuarial Science and Econometrics of the University of Amsterdam. I also wish to acknowledge the financial support of the Netherlands Organisation for Scientific Research (NWO).*

D.S.G. Pollock
Department of Economics
Queen Mary College
University of London
Mile End Road
London E1 4 NS

Tel : 01-980-4811 Ext 3552

LAGGED DEPENDENT VARIABLES DISTRIBUTED LAGS AND AUTOREGRESSIVE RESIDUALS

by

D.S.G. POLLOCK

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 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 [5]. 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. A model of this sort in which there are also common factors in the lag schemes will be called a COMFAC ADM.

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.

This prescription would be valid in general if the ADM were capable of accommodating a wide range of dynamic relationships without ever using an exorbitant number of parameters. However, the ADM does not have such capabilities, and therefore this method of deriving a dynamic specification should not be used without a particular justification.

If one believes that no model can wholly describe the complex and changing nature of an economic process, and if one is mindful of scarcity of econometric data, then one is inclined to look for a model which is both parsimonious in its parametrisation and flexible in its functional form. In this paper, we shall argue that the Rational Transfer-Function Model or RTM is superior in this respect to a parsimonious ADM and that it ought always to be adopted in preference when the dynamic specification is in doubt.

We shall attempt to substantiate this assertion by examining the defects of the ADM and the COMFAC ADM when they entail a misspecification of the dynamic structure of the disturbance part of the relationship. 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 very specific way. If this relationship does not hold, then the resulting misspecification will 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.

THE AUTOREGRESSIVE DISTRIBUTED-LAG MODEL AND THE RATIONAL TRANSFER-FUNCTION MODEL.

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

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

By imposing a somewhat complicated nonlinear restriction upon the parameters of this equation, we obtain the equation of a COMFAC ADM:

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

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

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

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

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

The corresponding form of the COMFAC ADM is

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

The very 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 $\alpha(L)$, 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 [8]). 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 wholly 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 [7]). 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

$$y(t) = \frac{\delta_0 + \delta_1 L}{1-\gamma_1 L - \gamma_2 L^2}x(t) + \frac{\theta_0 + \theta_1 L}{1-\phi_1 L - \phi_2 L^2}\varepsilon(t). \tag{6}$$

We must ask ourselves which of these alternative models, the ADM and the RTM, is appropriate to the typical econometric relationship. The opinion which we advance here is that there has to be great certainty in the appropriateness of the ADM before it is used in econometric modelling. To use it inappropriately is to invite a very distorted picture of the dynamic structure of an economic relationship. This we shall show in the ensuing sections.

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

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

For comparison with the ADM of equation (1), we might write this as

$$\{1-(\gamma+\phi)L+\gamma\phi L^2\}y(t) = (\delta-\delta\phi L)x(t) + (1-\gamma L)\varepsilon(t). \tag{8}$$

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 first-order autoregressive process described by the equation

$$(1-\pi L)x(t) = \xi(t), \tag{9}$$

where $\xi(t)$ is a white-noise process. 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

$$(1-\alpha L)y(t) = (\beta_0 + \beta_1 L)x(t) + e(t). \tag{10}$$

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 (7) and (9) into (10). After some rearrangement, we get the following expression for the residual sequence:

$$\begin{aligned} e(t) &= (1-\alpha L)\left\{\frac{\delta}{1-\gamma L} - \frac{\beta_0+\beta_1 L}{1-\alpha L}\right\}\frac{1}{1-\pi L}\xi(t) + \frac{1-\alpha L}{1-\phi L}\varepsilon(t) \\ &= p(L)\xi(t) + q(L)\varepsilon(t). \end{aligned} \tag{11}$$

Here $p(L) = \{p_0 + p_1 L + p_2 L^2 + \cdots\}$ and $q(L) = \{1 + q_1 L + q_2 L^2 + \cdots\}$ are the infinite-order polynomials in the lag operator L which come from expanding the rational polynomials 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

$$\begin{aligned} V\{e(t)\} &= \sigma_\xi^2 \sum p_i^2 + \sigma_\varepsilon^2 \sum q_i^2 \\ &= S(\alpha, \beta_0, \beta_1). \end{aligned} \tag{12}$$

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 (10). It follows from an basic theorem, which is proved, for example, by Amemiya [1] and by Domowitz and White [3], 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

$$\begin{aligned} S(\alpha, \beta_0, \beta_1) = \sigma_\xi^2 &\left\{ (\delta - \beta_0)^2 + \frac{W^2}{1-\pi^2} + \frac{D^2\gamma^2}{1-\gamma^2} + \frac{2DW\gamma}{1-\pi\gamma} \right\} \\ &+ \sigma_\varepsilon^2 \left\{ \frac{(\alpha-\phi)^2}{1-\phi^2} + 1 \right\}, \end{aligned} \tag{13}$$

where

$$\begin{aligned} W &= (C - \beta_0)\pi + \beta_1, \\ C &= \frac{\delta(\pi - \alpha)}{\pi - \gamma} \qquad \text{and} \\ D &= \frac{\delta(\alpha - \gamma)}{\pi - \gamma}. \end{aligned} \tag{14}$$

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

$$\begin{aligned} \beta_0 &= \delta \qquad \text{and} \\ \beta_1 &= \frac{\delta(\gamma - \alpha)}{1 - \gamma\pi}. \end{aligned} \tag{15}$$

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

$$S(\alpha) = \sigma_\xi^2 \frac{\delta^2\gamma^2(\alpha-\gamma)^2}{(1-\gamma^2)(1-\pi\gamma)^2} + \sigma_\varepsilon^2 \left\{ \frac{(\alpha-\phi)^2}{1-\phi^2} + 1 \right\}. \tag{16}$$

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

$$\begin{aligned}\alpha &= \frac{\kappa\gamma+\lambda\phi}{\kappa+\lambda}, \qquad \text{where} \\ \kappa &= \frac{\sigma_\xi^2\delta^2\gamma^2}{(1-\gamma^2)(1-\pi\gamma)^2} \qquad \text{and} \qquad \lambda = \frac{\sigma_\varepsilon^2}{1-\phi^2}.\end{aligned} \tag{17}$$

An inspection of equation (17) 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 $\gamma = \phi$, then we shall have $\alpha = \gamma = \phi$, $\beta_0 = \delta$ 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

$$\begin{aligned}V\left\{\frac{\delta}{(1-\gamma L)(1-\pi L)}\xi(t)\right\} &= \frac{\sigma_\xi^2\delta^2(1+\gamma\pi)}{(1-\gamma^2)(1-\gamma\pi)(1-\pi^2)} \\ &= \kappa\frac{(1+\gamma\pi)}{\gamma^2(1-\pi^2)}.\end{aligned} \tag{18}$$

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 $\gamma\pi \longrightarrow 1$, we find that $\kappa \longrightarrow \infty$; whereas, when $\gamma\pi \longrightarrow -1$, we find that κ remains small whilst the variance of the systematic component tends to zero.

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.

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 $\phi = -0.4$. 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

$$(1 - 0.45L - 0.34L^2)y(t) = (1 - 0.39)\zeta(t), \tag{19}$$

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

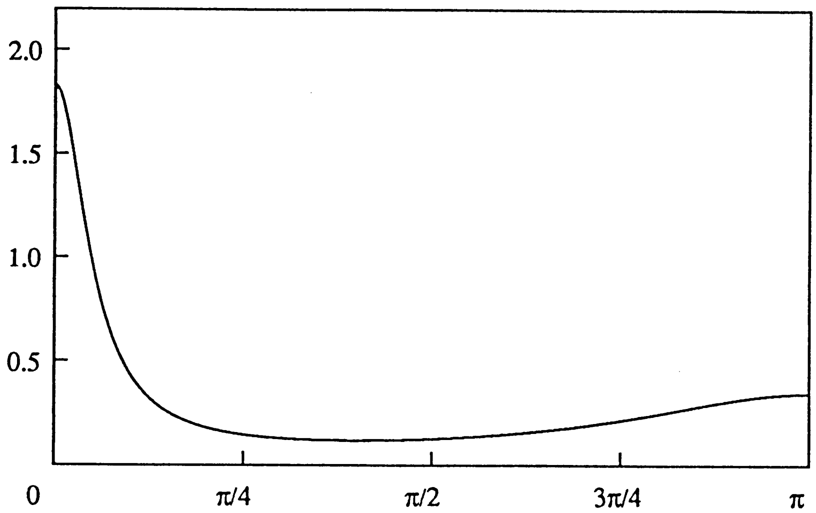

Figure 1. The spectral density function of the ARMA(2, 1) process $(1 - 0.45L - 0.34L^2)y(t) = (1 - 0.39L)\zeta(t)$ when $V\{\zeta(t)\} = 1.37$.

The parameters of this process are virtually the same as those of an ARMA(2,1) model which Granger and Newbold [4, p.86] 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. In all cases this has a severe downwards 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 $x(t) = \xi(t)/(1 - 0.9)$ is being strongly amplified by the lowpass filter $\beta/(1 - \gamma L) = 1/(1 - 0.85)$ 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 $y(t) = (1-0.85L)^{-1}x(t) + (1+0.4L)^{-1}\varepsilon(t)$ 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	0.228	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

FITTING AN EXTENDED ADM AND A COMFAC ADM

The distortions in our estimates can be reduced somewhat by adding extra parameters to the fitted model. In this section, we shall add an extra parameter to the polynomial $\alpha(L)$ of the ADM. This leads us to the model of equation (1). At the same time, we shall investigate the effect of forcing a COMFAC restriction upon this equation so as to obtain the equation (2).

The residual sequence from fitting the COMFAC ADM is given by

$$\begin{aligned} e(t) &= (1-\alpha L)(1-\rho L)\left\{\frac{\delta}{1-\gamma L} - \frac{\beta}{1-\alpha L}\right\}x(t) + \frac{(1-\alpha L)(1-\rho L)}{1-\phi L}\varepsilon(t) \\ &= p(L)x(t) + q(L)\varepsilon(t). \end{aligned} \tag{20}$$

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

$$S(\alpha,\beta,\rho) = \sigma_\xi^2\left\{(\delta-\beta)^2 + \{\delta(\gamma-\alpha)+\rho(\beta-\delta)\}^2 + \frac{\{\delta(\gamma-\alpha)(\gamma-\rho)\}^2}{1-\gamma^2}\right\} + \sigma_\varepsilon^2\left\{1+(\phi-\alpha-\rho)^2 + \frac{\{(\phi-\alpha)(\phi-\rho)\}^2}{1-\phi^2}\right\}. \tag{21}$$

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 [2], for which the rth iteration is specified by

$$\begin{aligned} \alpha_{(r)} &= \alpha\{\rho_{(r-1)}\}, \qquad \beta_{(r)} = \beta\{\rho_{(r-1)}\}, \\ \rho_{(r)} &= \rho\{\alpha_{(r)},\beta_{(r)}\}. \end{aligned} \tag{22}$$

Oberhofer and Kmenta [6] 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 β and α in place of these arguments.

For an illustration, we shall assume, as before, that $\delta = 1$, $\gamma = 0.85$ and $\phi = -0.4$. 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 $\rho = 0.689$.

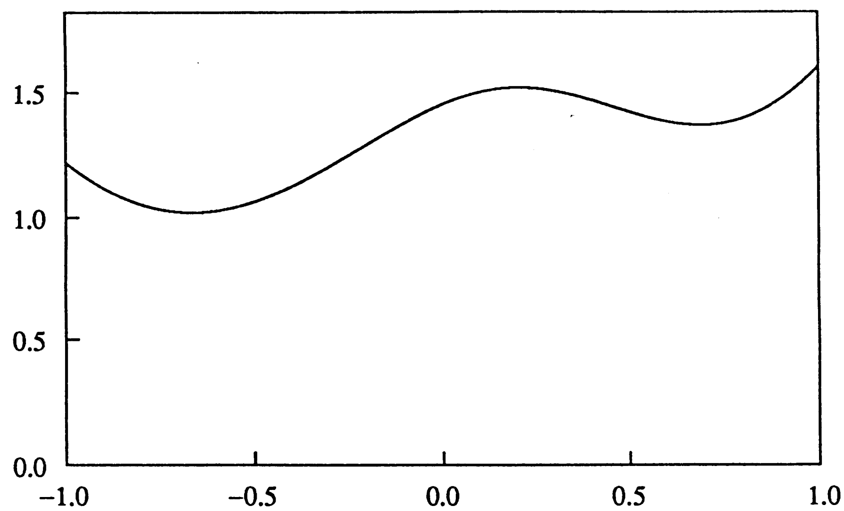

Figure 2. The graph of the function $S(\rho)$ associated with fitting the COMFAC model $(1-\rho L)(1-\alpha L)y(t) = (1-\rho L)\beta x(t) + e(t)$ when the true relationship is $y(t) = (1-\gamma L)^{-1}\delta x(t) + (1-\phi L)^{-1}\varepsilon(t)$, where $\gamma = 0.85$, $\phi = -0.4$, $\delta = 1$ and where $x(t)$ and $\varepsilon(t)$ are white-noise processes 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 COMFAC(i) and COMFAC(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:

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

We obtain these equalities by comparing the form of the RTM equation given under (8) with that of the ADM under (1), and by comparing the form of the RTM equation given under (7) with that of the COMFAC ADM under (5).

The addition of an extra parameter to the ADM has clearly improved the estimate of the multiplier; for the value of 3.348 compares favourable 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.

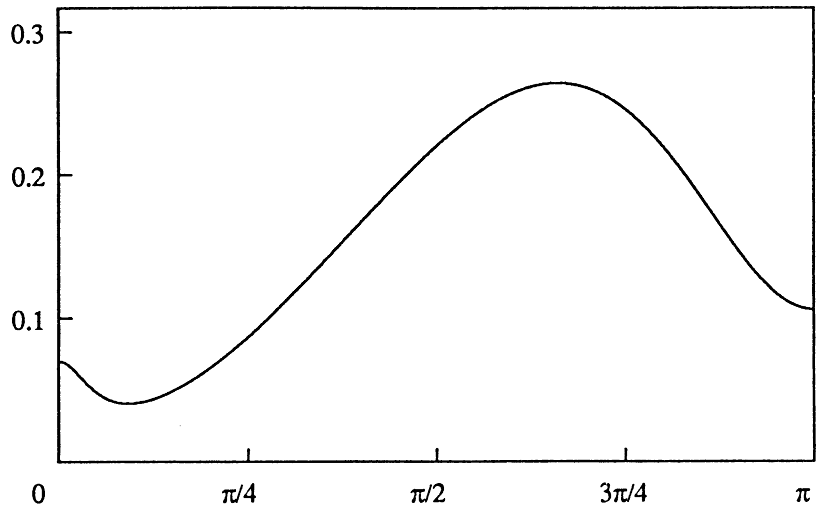

Figure 3. The spectrum of the residual sequence from fitting the model under COMFAC(i).

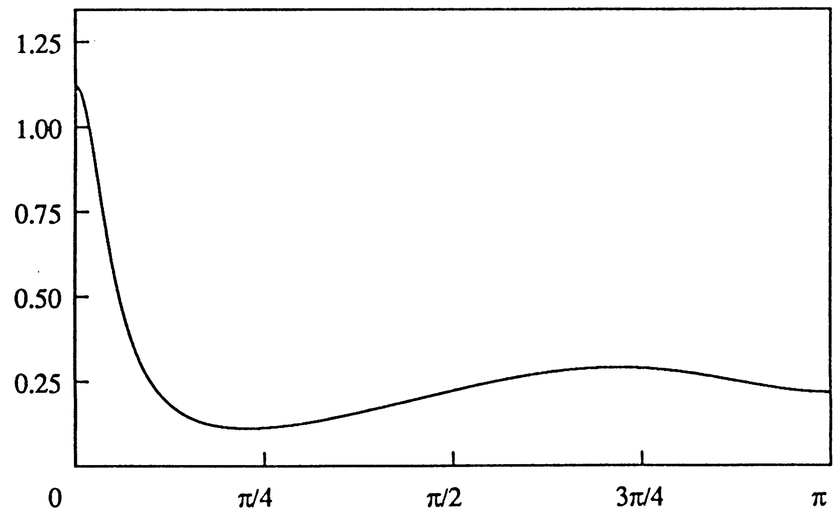

Figure 4. The spectrum of the residual sequence from fitting the model under COMFAC(ii).

The first of the COMFAC models is surprisingly similar to the ADM. This suggests that the COMFAC restrictions are liable to be accepted quite readily by statistical tests which use the misspecified ADM for the alternative hypothesis. The second of the COMFAC models is vastly different, and it fails to capture any of the characteristics of the RTM which generates $y(t)$.

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

	RTM	ADM	COMFAC(i)	COMFAC(ii)
σ_x^2	0.75	0.75	0.75	0.75
σ_ε^2	0.25	0.25	0.25	0.25
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

An interesting indication of this failure is provided by the frequency spectrum of the residual sequence given in Figure 4. This has 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.

CONCLUSIONS

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 COMFAC ADM. The only way in which we can overcome 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 COMFAC model of equation (2), and let us imagine that the RTM of equation (7) is the fitted model. Then the expression for the residual sequence is

$$\begin{aligned} e(t) &= (1-\phi L)\left\{\frac{\beta}{1-\alpha L} - \frac{\delta}{1-\gamma L}\right\} x(t) + \frac{1-\phi L}{(1-\alpha L)(1-\rho L)}\varepsilon(t) \\ &= p(L)x(t) + q(L)\varepsilon(t). \end{aligned} \tag{23}$$

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

$$V\{e(t)\} = V\{p(L)x(t)\} + V\{q(L)\varepsilon(t)\}, \tag{24}$$

where

$$V\{q(L)\varepsilon(t)\} = \frac{(1-\alpha\rho L)(1-\phi^2) - 2\phi(\alpha+\rho)}{(1-\alpha^2 L)(1-\alpha\rho L)(1-\rho^2 L)}. \tag{25}$$

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

$$\phi = \frac{\alpha+\rho}{1-\alpha\rho}. \tag{26}$$

The RTM fails to reflect the dynamic properties of the disturbance of the COMFAC model; 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 COMFAC model which is indeed correctly specified. It is quite possible that, given an inappropriate starting value for ρ, 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 (1) 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 graph of the concentrated function $S(\rho) = S\{\alpha(\rho), \beta(\rho)\}$ is shown in Figure 5. There are minima at the values $\rho = -0.85$ and $\rho = 0.879$. The values of α, β corresponding to the first of these minima are the ones belonging to the underlying COMFAC model. Those which correspond to the minimum at $\rho = 0.879$ are $\alpha = -0.771$ and $\beta = 0.879$.

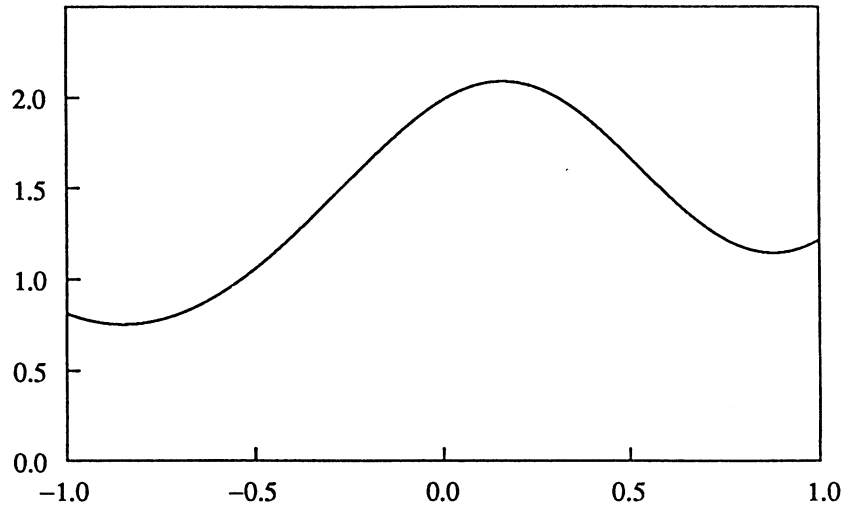

Figure 5. The graph of the function $S(\rho)$ associated with fitting the correctly specified COMFAC model $(1-\rho L)(1-\alpha L)y(t) = (1-\rho L)\beta x(t) + e(t)$ when the true relationship has $\rho = -0.85$, $\alpha = 0.85$ and $\beta = 1.0$ and when $x(t)$ and $\varepsilon(t)$ are white-noise processes with $V\{x(t)\} = 0.25$ and $V\{\varepsilon(t)\} = 0.75$.

We have seen that fitting a COMFAC model is a hazardous business even when it does correspond to the process underlying the data. However, the COMFAC ADM 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 [5] supported the use of the COMFAC ADM with the claim that it represents a convenient simplification of an ADM. In this paper, we have cast doubt on the general applicability of the ADM; and we have suggested that, even if the ADM were appropriate in a particular instance, the possibility that the Cochrane–Orcutt iteration has more than one fixed point implies that the COMFAC ADM is an inconvenient specialisation which we would be loathe to pursue.

REFERENCES

1. Amemiya, T. *Advanced Econometrics.* Cambridge: Basil Blackwell, 1985.
2. Cochrane, D. and G.H. Orcutt. Application of least-squares regressions to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* 44 (1949): 32–61.
3. Domowitz, I. and H. White. Misspecified models with dependent observations. *Journal of Econometrics* 20 (1982): 35–58.

4. Granger, C.W.J. and P. Newbold. *Forecasting Economic Time Series.* New York: Academic Press, 1977.
5. Hendry, D.F. and G. Mizon. 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): 549–563.
6. Oberhofer, W. and J. Kmenta. A general procedure for obtaining maximum likelihood estimates in generalised regression models. *Econometrica* 42 (1974): 579–590.
7. Rice, S.O. Noise in FM Receivers. In M. Rosenblatt (ed), *Time Series Analysis,* pp. 395–422. New York: John Wiley and Sons, 1963.
8. Wright, J.R. Flutter test analysis in the time domain using a recursive system representation. *Journal of Aircraft* 11 (1974): 774–777.

APPENDIX

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 $\phi(L)x(t) = \theta(L)\varepsilon(t)$. Define

$$\begin{aligned}
\alpha(z) &= 1 + \alpha_1 z + \cdots + \alpha_p z^p = \prod_{i=1}^{p}(1 - \lambda_i z),\\
\phi(z) &= 1 + \phi_1 z + \cdots + \phi_f z^f = \prod_{i=1}^{f}(1 - \kappa_i z),\\
\beta(z) &= 1 + \beta_1 z + \cdots + \beta_q z^q = \prod_{i=1}^{q}(1 - \mu_i z),\\
\theta(z) &= 1 + \theta_1 z + \cdots + \theta_h z^h = \prod_{i=1}^{h}(1 - \nu_i z).
\end{aligned} \qquad (A.1)$$

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

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

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 $\theta(z)/\phi(z)$

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

$$\frac{1}{\alpha(z^{-1})} = \frac{C_1}{1-\lambda_1 z^{-1}} + \cdots + \frac{C_p}{1-\lambda_p z^{-1}}, \tag{A.3}$$

where the generic coefficient is

$$C_k = \frac{\lambda_k^{p-1}}{\prod_{i\neq k}(\lambda_k - \lambda_i)}. \tag{A.4}$$

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

$$\frac{1}{\phi(z)} = \frac{D_1}{1-\kappa_1 z} + \cdots + \frac{D_f}{1-\kappa_f z}. \tag{A.5}$$

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

$$\frac{1}{\alpha(z^{-1})\phi(z)} = \sum_k \sum_l \frac{C_k D_l}{(1-\lambda_k z^{-1})(1-\kappa_l z)}. \tag{A.6}$$

This expression may be evaluated using the result that

$$\frac{C_k D_l}{(1-\lambda_k z^{-1})(1-\kappa_l z)} = \frac{C_k D_l}{(1-\lambda_k \kappa_l)}\left\{\cdots + \frac{\lambda_k^2}{z^2} + \frac{\lambda_k}{z} + 1 + \kappa_l z + \kappa_l^2 z^2 + \cdots\right\} \tag{A.7}$$

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

$$\left(\sum_{i=0}^{q} \beta_i z^{-i}\right)\left(\sum_{j=0}^{h} \theta_j z^j\right) = \sum_{j=-q}^{h}\left(\sum_{k=m}^{n} \beta_k \theta_{j-k}\right) z^j, \tag{A.8}$$

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

$$(1-\rho L)(1-\alpha L)y(t) = (1-\rho L)\beta x(t) + e(t) \tag{A.9}$$

when the true relationship is given by

$$y(t) = \frac{\delta}{1-\gamma L}x(t) + \frac{1}{1-\phi L}\varepsilon(t). \tag{A.10}$$

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

$$\begin{bmatrix} P & Q \\ R & S \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} V \\ W \end{bmatrix}. \tag{A.11}$$

The elements of the equation are

$$\begin{aligned} P &= \sigma_x^2\delta^2\left\{1 + \frac{(\gamma-\rho)^2}{1-\gamma^2}\right\} + \sigma_\varepsilon^2\delta^2\left\{1 + \frac{(\phi-\rho)^2}{1-\phi^2}\right\}, \\ Q &= R = -\sigma_x^2\delta\rho, \\ S &= \sigma_x^2(1+\rho^2), \\ V &= \sigma_x^2\delta^2\left\{(\gamma-\rho) + \frac{\gamma(\gamma-\rho)^2}{1-\gamma^2}\right\} + \sigma_\varepsilon^2\left\{(\phi-\rho) + \frac{\phi(\phi-\rho)^2}{1-\phi^2}\right\}, \\ W &= \sigma_x^2\{\delta - \delta\rho(\gamma-\rho)\}. \end{aligned} \tag{A.12}$$

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

$$\rho = \frac{H}{G}, \tag{A.13}$$

where

$$\begin{aligned} G &= \sigma_x^2\delta^2\left\{(\delta-\beta)^2 + \frac{\delta^2(\gamma-\alpha)^2}{1-\gamma^2}\right\} + \sigma_\varepsilon^2\left\{1 + \frac{(\phi-\alpha)^2}{1-\phi^2}\right\}, \\ H &= \sigma_x^2\delta^2\left\{\delta(\gamma-\alpha)(\delta-\beta) + \frac{\delta^2\gamma(\gamma-\alpha)^2}{1-\gamma^2}\right\} + \sigma_\varepsilon^2\left\{(\phi-\alpha) + \frac{\phi(\phi-\alpha)^2}{1-\phi^2}\right\}. \end{aligned} \tag{A.14}$$