AMSTER

7/89

FACULTY

~~INSTITUTE~~ OF ACTUARIAL SCIENCE & ECONOMETRICS

Report AE 7/89

THE MISSPECIFICATION OF DYNAMIC REGRESSION MODELS

D.S.G. Pollock

University of Amsterdam

THE MISSPECIFICATION OF DYNAMIC REGRESSION MODELS

by

D.S.G. POLLOCK* AND EVANGELIA PITTA

This paper investigates some of the hazards of estimating dynamic regression relationships with misspecified models. It is argued that we should adopt models which attribute separate sets of parameters to the systematic and disturbance parts of the regression relationship. The models which are commonly adopted in applied econometrics attempt to capture the basic dynamic properties of the two parts of the relationship with the same set of parameters. If the properties of these two parts differ, then they are bound to be misrepresented by such models which can give rise to grossly misleading estimates.

* *The paper was completed while Stephen Pollock was enjoying the hospitality of the Faculty of Actuarial Science and Econometrics of the University of Amsterdam. He wishes 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

THE MISSPECIFICATION OF DYNAMIC REGRESSION MODELS

D.S.G. POLLOCK AND EVANGELIA PITTA

1. INTRODUCTION

This paper investigates the consequences of using an inappropriate model in estimating a dynamic regression relationship. We imagine that the model is aimed at explaining the observable sequence $y(t)$ in terms of another observable sequence $x(t)$ and an unobservable white-noise sequence $\varepsilon(t)$. Our object is to examine the nature of the estimates which are derived when the model is incapable of accommodating the relationship which actually exists between these sequences.

Our approach will be to infer the properties of the misspecified estimators from the expected value of the criterion function which they are designed to minimise. The criterion function converges to its expected value as the sample size increases; and the values which minimise the asymptotic form of the expected function are the probability limits of the estimators. Only a moderate use of the computer is required in finding these values; and, in some simple cases, they may be found directly by analytic methods.

In the first section of this paper, we derive the asymptotic form of the expected value of the least-squares criterion function for a general class of dynamic regression models. In the remaining sections, we use this function to make inferences about the asymptotic behaviour of the misspecified estimators in a range of circumstances which are intended to resemble conditions encountered by econometricians in their empirical work. In these circumstances, the dynamic relationships amongst the temporal sequences which are indicated by the misspecified estimators can differ from the true relationships to an extent which is startling. The problems are compounded by the fact that there is sometimes more than one solution to the estimating equations.

The analysis of misspecified dynamic models has already received considerable attention in the econometric literature; and we should give a brief account of some the important contributions to the topic.

First, we should mention Hendry [7] who proposed a methodology for determining the behaviour of inconsistent instrumental-variables estimators applied to a dynamic system of simultaneous equations. In this system, the jointly dependent variables reenter the structural equations after a one-period lag, whilst the structural disturbances are subject to a first-order autoregressive scheme.

Hendry's intention was to devise a method for determining the consequences of selecting inappropriate instruments which are correlated with the disturbances, as well as the consequences of ignoring the dynamic structure of the disturbances. He also endeavoured to find the second moments of the limiting normal distributions of the misspecified estimates.

Maasoumi and Phillips [13] have discovered flaws in Hendry's analysis of the limiting distributions of the misspecified estimators. They have corrected the analysis; but, in doing so, they have show that, when properly applied, the recommended methodology can be very laborious, even when the estimators have simple closed forms which generate unique values. In particular, the correct formulae for the asymptotic variance, which will depend upon the fourth-order moments of the processes which generate the input variables, will generally entail many more terms than were included by Hendry.

There has also been some doubt as to whether the limiting distributions of the misspecified estimators can provide adequate approximations to the distributions of estimators calculated from small samples. Part of the problem here is that the conditions of stationarity and invertibility, which must be obeyed by the relevant parts of the time-series models, impose bounds or "truncations"on the sampling distributions which are not reflected in the approximating distributions.

Because of these difficulties, it does not seem too modest to concentrate on the problem of analysing the probability limits of the misspecified estimators to the exclusion of the other aspects of their distributions. Indeed, it seems more important to reveal the grosser defects of the estimators that to concentrate upon the finer points of their distributional properties.

Another notable contribution to the econometric literature is an article by Cragg [4] which develops inferential procedures for dynamic regression models which are valid even when the precise form of the serial correlation or the heteroskedasticity which affects the disturbance process is unknown. Cragg's procedures depend upon the instrumental-variables method of estimation which has also been pursued, in the context of engineering studies, by Young [30], Jakeman et al. [8] and Söderström and Stoica [22] amongst others.

Finally, we should mention the work by White [25], [26], [27], [28], and of Domowitz and White [5] which provides a very general framework for analysing the properties of the maximum-likelihood estimators and least-squares estimators of misspecified models, and for constructing tests of misspecification. By following their suggestions, it is possible to construct tests of the specification of the systematic part of a dynamic regression model which are valid, asymptotically, even when the structure of the disturbance part is unknown or is misspecified. Thus Domowitz and White provide a natural setting in which to view the results of Cragg. However, only the limiting distributions of the test statistics are available; and it is reasonable to doubt their adequacy as approximations to small-sample distributions.

The contributions which we have mentioned have helped to establish a sound theoretical framework for analysing the behaviour of the estimators of misspecified models and for developing tests of misspecification. However, there remains a dearth of qualitative results concerning the practical effects of misspecifying the dynamic structure of a regression model; and such matters are the primary concern of the present paper.

The results of this paper depend, partly, upon an analysis of the spectral properties of the systematic and disturbance parts of the fitted regression models. To the extent that it deals with the frequency-domain aspects of dynamic modelling, the paper has affinities with the work of Whittle [29], Walker [24], Bloomfield [3] and Kabaila and Goodwin [9], all of which represent contributions to the general literature of time-series analysis.

2. THE MODELS

We shall use generating functions or z-transforms to represent our models. The General Temporal Model or GTM is represented by

$$\alpha(z)y(z) = \beta(z)x(z) + \mu(z)\varepsilon(z), \tag{1}$$

where $y(z)$ and $x(z)$ are the z-transforms of the observable sequences $y(t)$ and $x(t)$, and $\varepsilon(z)$ relates to the disturbance sequence $\varepsilon(t)$. The polynomials $\alpha(z) = 1+\alpha_1 z+\cdots+\alpha_a z^a$, $\beta(z) = \beta_0+\beta_1 z+\cdots+\beta_b z^b$ and $\mu(z) = 1+\mu_1 z+\cdots+\mu_m z^m$, whose degrees are a, b and m respectively, contain the parameters of the model. The rational form of equation (1) is

$$y(z) = \frac{\beta(z)}{\alpha(z)}x(z) + \frac{\mu(z)}{\alpha(z)}\varepsilon(z). \tag{2}$$

For this model to be viable, the roots of the polynomial equations $\alpha(z) = 0$ and $\mu(z) = 0$ must lie outside the unit circle. This is to ensure that the coefficients of the series expansions of $\alpha^{-1}(z)$ and $\mu^{-1}(z)$ form convergent sequences.

The Rational Transfer-Function Model or RTM is represented by

$$y(z) = \frac{\delta(z)}{\gamma(z)}x(z) + \frac{\theta(z)}{\phi(z)}\varepsilon(z), \tag{3}$$

or, alternatively, by

$$\gamma(z)\phi(z)y(z) = \phi(z)\delta(z)x(z) + \gamma(z)\theta(z)\varepsilon(z). \tag{4}$$

We shall denote the degrees of the polynomials $\gamma(z), \delta(z), \phi(z)$ and $\theta(z)$ by g, d, f and h respectively. The leading coefficients of $\gamma(z), \phi(z)$ and $\theta(z)$ are set to unity. The roots of the equations $\gamma(z) = 0, \phi(z) = 0$ and $\theta(z) = 0$ must lie outside the unit circle.

We shall assume that the signal $x(t)$ is generated by an autoregressive process so that

$$\pi(z)x(z) = \xi(z) \tag{5}$$

with $\xi(t)$ as a white-noise process with a zero mean. We can accommodate an autoregressive moving-average process in this specification by allowing the order of the polynomial $\pi(z)$ to be infinite.

3. THE RESIDUAL SEQUENCE

Let us imagine that the true model is an RTM and that the fitted model is a GTM. Then the generating function of the residual sequence of the fitted model, which is the sequence of one-step-ahead prediction errors in the terminology of Ljung [12], is given by

$$e(z) = \frac{\alpha(z)}{\mu(z)}y(z) - \frac{\beta(z)}{\mu(z)}x(z). \tag{6}$$

On substituting for $y(z)$ from (3) and $x(z)$ from (5), we get

$$\begin{aligned} e(z) &= \frac{\alpha(z)}{\mu(z)}\left\{\frac{\delta(z)}{\gamma(z)} - \frac{\beta(z)}{\alpha(z)}\right\}\frac{\xi(z)}{\pi(z)} + \frac{\alpha(z)\theta(z)}{\mu(z)\phi(z)}\varepsilon(z) \\ &= p(z)\xi(z) + q(z)\varepsilon(z). \end{aligned} \tag{7}$$

Notice that the leading coefficient of the expansion of the rational function $q(z)$ is unity on account of the normalisation of the leading coefficients of each of its factors. If $\beta(z)/\alpha(z)$ and $\mu(z)/\alpha(z)$ were to attain the values of $\delta(z)/\gamma(z)$ and $\theta(z)/\phi(z)$ respectively, then we should have $e(z) = \varepsilon(z)$.

In order to write an expression for the residual sequence $e(t)$ instead of one for its generating function $e(z)$, we can use the lag operator L defined by the equation $Lx(t) = x(t-1)$. Then we have $e(t) = p(L)\xi(t) + q(L)\varepsilon(t)$.

On the assumption that the white-noise processes $\xi(t)$ and $\varepsilon(t)$ are mutually independent, the covariance generating function for $e(t)$ is

$$C(z) = \sigma_\xi^2 p(z)p(z^{-1}) + \sigma_\varepsilon^2 q(z)q(z^{-1}), \tag{8}$$

where $\sigma_\xi^2 = V\{\xi(t)\}$ and $\sigma_\varepsilon^2 = V\{\varepsilon(t)\}$ are the variances of the processes. The variance of the residual sequence $e(t)$ is given by the function

$$\begin{aligned} S &= S_p + S_q \\ &= \sigma_\xi^2 \sum p_j^2 + \sigma_\varepsilon^2 \sum q_j^2, \end{aligned} \tag{9}$$

where the elements p_j and q_j are coefficients from the series expansions of $p(z)$ and $q(z)$ respectively, and where $q_0 = 1$.

The spectral density function, or "spectrum", of the residual sequence, which is a function of the frequency variable ω over the interval $[-\pi, \pi]$, is obtained by setting $z = e^{-i\omega}$ in $C(z)$ and scaling the result by $1/2\pi$. Since $C(z) = C(z^{-1})$, the spectrum can be characterised completely in terms of its values over the interval $[0, \pi]$. The variance of the residual sequence is expressible as the integral of its spectrum over its entire domain:

$$S = \frac{1}{2\pi} \int_{\omega=-\pi}^{\pi} C(e^{-i\omega}) d\omega. \tag{10}$$

We shall be able to gain some insight into the nature of the misspecified estimates by examining the form of the function $C(e^{-i\omega})$. In particular, we can assess the extent to which the estimates misrepresent the true relationship by gauging the extent to which $C(e^{-i\omega})/2\pi$ differs from the spectrum of the white-noise process $\varepsilon(t)$, which is given by the constant function $f(\omega) = \sigma_\varepsilon^2/2\pi$.

4. LEAST-SQUARES FITTING

Given a sample of T elements of $x(t)$ and $y(t)$ running from $t = 0$ to $t = T - 1$, the least-squares criterion for fitting the GTM is to minimise the function

$$S_T(p, q) = \frac{1}{T} \sum_{t=0}^{T-1} e_t^2. \tag{11}$$

In finite samples, the precise value of this function depends upon how we represent the presample elements $\varepsilon_{-1}, \ldots, \varepsilon_{-m}$, $y_{-1}, \ldots, y_{-a}$, and $x_{-1}, \ldots, x_{-b}$. It is reasonable to attribute to these elements their expected values which are zeros.

In the appendix, we use the law of large numbers to show that, as $T \to \infty$, the function S_T tends in probability, uniformly for all points in the parameter set, to the function S which gives the variance of the residual sequence. It follows from a fundamental theorem, which is proved by Amemiya [1, Theorem 4.1.1] and by Domowitz and White [5, Theorem 2.2] amongst others, that the values which minimise S_T will tend in probability to those which minimise S. Thus we can find the probability limits of our estimators by finding the values which minimise the asymptotic form of the criterion function. When the latter are equal to the parameters of the true model, we can say that these parameters are estimated consistently.

The convergence of S_T to S is the same regardless of how we represent the presample elements. Also, the normal maximum-likelihood estimators are asymptotically equivalent to the least-squares estimators. This well-known result relieves us of any need to make a separate analysis of the maximum-likelihood estimators.

For the fitting of the GTM by least squares to result in the consistent estimation of the parameters of the true RTM, it must be possible for the GTM polynomials of equation (1) to attain the values of the corresponding RTM polynomials of equation (4). Therefore the degrees of $\alpha(z)$, $\beta(z)$ and $\mu(z)$ must be at least as great as those of $\gamma(z)\phi(z)$, $\phi(z)\delta(z)$ and $\gamma(z)\theta(z)$ respectively; and so it is necessary that

$$a \geq g + f, \quad b \geq f + d \quad \text{and} \quad m \geq g + h. \tag{12}$$

However, there must be at least one equality here; for, otherwise, there will be a redundant factor common to $\alpha(z)$, $\beta(z)$ and $\mu(z)$ which will prevent them from being uniquely determined. To demonstrate the consistency of the least-squares estimation when (12) holds and there are no redundant factors, we need only confirm that S will achieve its absolute minimum whenever $\beta(z)/\alpha(z)$ and $\mu(z)/\alpha(z)$ attain the values of $\delta(z)/\gamma(z)$ and $\theta(z)/\phi(z)$ respectively. In that case, we will have $p(z) = 0$, $q(z) = q_0 = 1$ and hence $S = \sigma_\varepsilon^2$.

In this paper, we are interested primarily in circumstances where the inequalities in (12) are not fulfilled and where, consequently, the parameters of the GTM fitted by least squares cannot converge upon the parameters of the true RTM. We can readily infer, from the form of the criterion function, that, when any of the conditions of (12) are violated, the entire set of GTM estimates will be affected. This is so even when it is only the degree $\mu(z)$ which is mistaken. However, if it is only the specification of the disturbance process which is in doubt, then we may be able to estimate the remaining parameters of the RTM consistently by using a robust instrumental-variables procedure of the sort analysed by Cragg [4]. However, even with the enhancements proposed by Cragg, the instrumental-variables estimator is liable to be inefficient.

The effects of misspecification are liable to be less serious if we fit an RTM of too few parameters by least squares than if we fit a GTM of too few. Let us imagine that, in place of the GTM of equation (2), our fitted model is an RTM denoted by

$$y(z) = \frac{\beta(z)}{\alpha(z)}x(z) + \frac{\mu(z)}{\rho(z)}e(z), \tag{13}$$

where $\rho(z)$ stands for a polynomial of degree r with $\rho_0 = 1$ as its leading coefficient. For the consistent estimation of the parameters of the true model of (3), we now require that $a \geq g$ and $b \geq d$ with at least one equality and that $r \geq f$ and $m \geq h$ with at least one equality.

When we fit the RTM of (13), the expression for the residual sequence is

$$\begin{aligned} e(z) &= \frac{\rho(z)}{\mu(z)}\left\{\frac{\delta(z)}{\gamma(z)} - \frac{\beta(z)}{\alpha(z)}\right\}\frac{\xi(z)}{\pi(z)} + \frac{\rho(z)\theta(z)}{\mu(z)\phi(z)}\varepsilon(z) \\ &= p(z)\xi(z) + q(z)\varepsilon(z). \end{aligned} \tag{14}$$

From this, we can see that any misspecification which understates the degrees of $\mu(z)$ and $\rho(z)$ in the disturbance part of the model will not affect the consistency of the estimates of the parameters of the systematic part when the degrees of $\alpha(z)$ and $\beta(z)$ are correctly specified. In that case, the value of the term $S_p = \sigma_\xi^2 \sum p_j^2$ within the expression for $S = S_p + S_q$ will be zero in consequence of $\alpha(z)$ and $\beta(z)$ assuming the values of $\gamma(z)$ and $\delta(z)$ respectively, whilst the value of the term $S_q = \sigma_\varepsilon^2 \sum q_j^2$ will be minimised independently.

This result is due to the fact that, in the RTM, the parameters of the two parts of the model are distinct. A familiar instance of it is that, in a classical linear regression model, a misspecification of the covariance properties of the disturbance vector may affect the efficiency of the least-squares estimation but will not affect its consistency.

The specification of the RTM, which uses separate rational functions to model the transfer functions of the systematic and disturbance parts of the system, accords well with the view, which is taken by Kabaila and Goodwin [9] and by Walker [24] amongst others, that the estimation of time-series models is essentially a matter of functional approximation.

In the ensuing sections of this paper, we shall concentrate on analysing the effects of fitting a misspecified GTM. The GTM is used in econometrics far more often than is the RTM. We shall use the RTM for our model of the true relationship because it affords a more parsimonious parametrisation than does the GTM. Also, if we should wish to assume that the true model is a GTM, then we can easily represent it as a special case of the RTM.

5. A FITTED MODEL WITH A LAGGED DEPENDENT VARIABLE

The dynamic regression model which is most commonly used in applied econometrics has the form of

$$\alpha(z)y(z) = \beta(z)x(z) + e(z). \tag{15}$$

This is a special case of the GTM where $\mu(z) = 1$ has a degree of zero. The popularity of the model is explained by the fact that it can be estimated easily using the method of ordinary least-squares regression. Its tractability is also reflected in the relative ease with which we can analyse the consequences of using it incorrectly in place of an RTM to estimate a relationship which exists amongst the sequences $y(t)$, $x(t)$ and $\varepsilon(t)$.

We can gain some basic insights into the likely effects of what may be the most common kinds of misspecification by examining some simple cases which are directly amenable to an algebraic analysis.

To begin, let us consider the case where the true model is an RTM in the form of

$$y(z) = \frac{\delta}{1+\gamma z}x(z) + \frac{1}{1+\phi z}\varepsilon(z), \tag{16}$$

and let us imagine that the fitted model takes the form of

$$(1+\alpha z)y(z) = \beta x(z) + e(z). \tag{17}$$

This can be regarded as a degenerate case of the GTM wherein $\beta(z) = \beta$ and $\mu(z) = 1$ are both of zero degree.

On the assumption that $x(t) = \xi(t)$ is a white-noise process, the generating functions $p(z)$ and $q(z)$ of (7) become

$$\begin{aligned} p(z) &= \frac{\delta(1+\alpha z)}{1+\gamma z} - \beta \qquad \text{and} \\ q(z) &= \frac{1+\alpha z}{1+\phi z}. \end{aligned} \tag{18}$$

The asymptotic form of the criterion function may be derived by finding the variances of $p(L)\xi(t)$ and $q(L)\varepsilon(t)$ using the formula for the variance of an ARMA(1, 1) process which is given, for example, by Pandit and Wu [15]. It can also be found directly from the expressions for the sums of the squares of the coefficients of the power series expansions of $p(z)$ and $q(z)$. Thus it may be shown that

$$S(\alpha,\beta) = \sigma_x^2\left\{\delta^2\frac{(\alpha-\gamma)^2}{1-\gamma^2} + (\delta-\beta)^2\right\} + \sigma_\varepsilon^2\left\{\frac{(\alpha-\phi)^2}{1-\phi^2}+1\right\}. \tag{19}$$

By setting the derivatives of the function to zero, we find that the minimising values, which are also the probability limits of the estimators of the misspecified model, are given by

$$\begin{aligned} \bar{\beta} &= \delta \qquad \text{and} \\ \bar{\alpha} &= \frac{\kappa\gamma+\lambda\phi}{\kappa+\lambda}, \qquad \text{where} \\ \kappa &= \frac{\sigma_x^2\delta^2}{1-\gamma^2} \qquad \text{and} \qquad \lambda = \frac{\sigma_\varepsilon^2}{1-\phi^2}. \end{aligned} \tag{20}$$

The least-squares estimator of the parameters of equation (17) has a closed form. Therefore it is possible to derive the probability limits directly from the estimating equation as well as from the criterion function as we have done. The advantage of the method which we have used is that is practicable even when the estimating equations are of an intractable nonlinear form.

The value of the probability limit $\bar{\alpha}$ is seen to be a convex combination of the values of γ and ϕ. The weights κ and λ of the combination are simply the variances of the systematic part and the disturbance part of the RTM in (16). This result becomes intelligible when we put the equation of the GTM in the same form as the equation of the RTM by dividing it throughout by $(1+\alpha z)$; for then it can be seen that a single parameter α of the GTM is being called upon to fulfil the functions of two parameters γ and ϕ of the RTM. If the latter have similar values, then they might both be well approximated by $\bar{\alpha}$. However, in many econometric contexts, there may be good reason to believe that the properties of the systematic and disturbance parts of the model will be dissimilar.

To illustrate some of the hazards of misspecification, let us assume that the parameters of the true RTM under (16) take the values of $\delta = 1$, $\gamma = -0.85$ and $\phi = 0.85$. According to these assumptions, the systematic part of the model embodies a low-pass filter whilst the disturbance part embodies a high-pass filter. If we wish to think of the RTM as a model of human behaviour, then we may imagine a person in an environment beset by high-frequency noise who reacts sluggishly to the information conveyed by the signal $x(t)$. By varying the relative sizes of σ_x^2 and σ_ε^2, we may vary the signal-to-noise ratio.

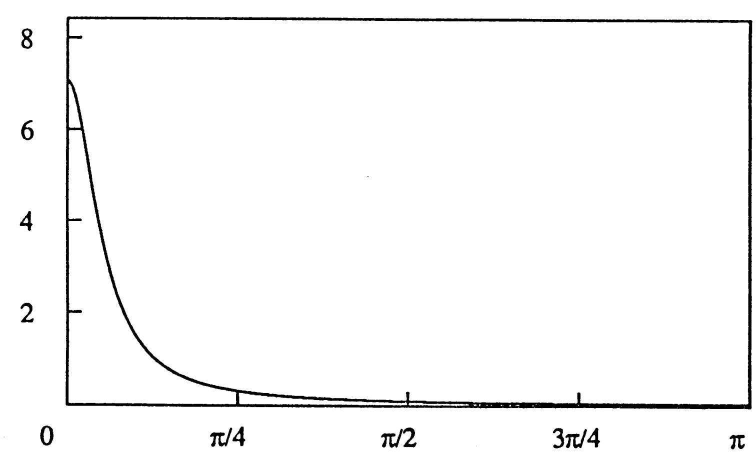

Figure 1. The graph of the frequency response function of the low-pass filter $\gamma^{-1}(L) = (1-0.85L)^{-1}$. The frequency response function of the high-pass filter $\phi^{-1}(L) = (1+0.85L)^{-1}$ is the mirror image of the graph above.

Figure 1 shows the frequency response function for the filter $\gamma^{-1}(L) = (1-0.85L)^{-1}$ which is the transfer function of the systematic part of the model. Thus the graph shows the value of the function $\gamma^{-1}(z)\gamma^{-1}(z^{-1})/2\pi$ for $z = e^{-i\omega}$ when ω is in the interval $[0, \pi]$. This gives us the spectrum of the systematic part of the RTM for the case where $V\{x(t)\} = 1$. The frequency response function of the filter $\phi^{-1}(L) = (1+0.85L)^{-1}$ associated with the stochastic part of the RTM is the mirror image of the graph. Thus the spectrum of $y(t)$, which is formed from the superposition of the spectra of the two parts of the model, has a U-shaped appearance. The relative heights of the branches of the U vary with the signal-to-noise ratio.

Table 1 shows the effect of fitting the misspecified GTM in circumstances with differing signal-to-noise ratios. Its most notable feature is the wide variation in the value of $\bar{\alpha}$; and it is clear that, in all the cases, this probability limit conveys a very misleading impression of the dynamic properties of the relationship between $x(t)$ and $y(t)$.

Table 1. The effects of fitting the model $(1+\alpha z)y(z) = \beta x(z) + e(z)$ when the true relationship is $y(z) = (1-0.85z)^{-1}x(z) + (1+0.85z)^{-1}\varepsilon(z)$ and $x(t)$ is white noise.

	Case A	Case B	Case C
σ_x^2	0.25	0.5	0.75
σ_ε^2	0.75	0.5	0.25
$\bar{\alpha}$	0.425	0.0	−0.425
$\bar{\beta}$	1.0	1.0	1.0
S	2.703	3.104	2.203
True Variances			
Systematic	0.901	1.802	2.703
Disturbance	2.703	1.802	0.901
Sum	3.604	3.604	3.604
Estimated Variances			
Systematic	0.305	0.5	0.915
Disturbance	3.298	3.104	2.688

The table also shows how the variance of $y(t)$ is attributed to the systematic and the disturbance parts of the model. For the true RTM, the two components of the variance are given by

$$V\left\{\frac{\delta}{1+\gamma L}x(t)\right\} = \frac{\delta^2\sigma_x^2}{1-\gamma^2} \quad \text{and} \quad V\left\{\frac{1}{1+\phi L}\varepsilon(t)\right\} = \frac{\sigma_\varepsilon^2}{1-\phi^2}; \tag{21}$$

whereas, for the fitted GTM, they are given by

$$V\left\{\frac{\beta}{1+\alpha L}x(t)\right\} = \frac{\beta^2\sigma_x^2}{1-\alpha^2} \quad \text{and} \quad V\left\{\frac{1}{1+\alpha L}e(t)\right\} = \frac{S}{1-\alpha^2}. \tag{22}$$

It can be shown algebraically that the sum of the variance components of the fitted GTM in (22) must be identical to the sum of the variance components of the true RTM in (21). The table shows that the variance of the systematic component is liable to be underestimated significantly when a misspecified model is fitted by ordinary least-squares regression.

The significant aspect of these results, from the point of view of their econometric interpretation, is the extent to which the dynamic properties of the transfer functions are misrepresented by the probability limits of the fitted models. In all cases, the gain of the systematic transfer function (ie. the long-term multiplier) and the median lag of its impulse response (ie. the time lags of the adjustment process) are seriously underestimated.

6. THE MODEL WHEN THE SIGNAL IS A SERIALLY CORRELATED SEQUENCE

The transparent nature of the results in Table 1 is due in part to our assumption that $x(t)$ is a white-noise sequence. It would be more realistic to assume that $x(t)$ is generated by a stationary stochastic process. Therefore, let us imagine that it is generated by a first-order autoregressive process represented by

$$(1+\pi z)x(z) = \xi(z). \tag{23}$$

In that case, the generating function $p(z)$ of (18) is replaced by

$$p(z) = \frac{\delta(1+\alpha z)}{(1+\gamma z)(1+\pi z)} - \frac{\beta}{(1+\pi z)}. \tag{24}$$

The asymptotic form of the expected criterion function is given by

$$S(\alpha,\beta) = \sigma_\xi^2\left\{\frac{(C-\beta)^2}{1-\pi^2} + \frac{D^2}{1-\gamma^2} + \frac{2(C-\beta)D}{1-\pi\gamma}\right\} + \sigma_\varepsilon^2\left\{\frac{(\alpha-\phi)^2}{1-\phi^2}+1\right\},$$
$$\text{where}\quad C = \frac{\delta(\pi-\alpha)}{\pi-\gamma} \quad\text{and}\quad D = \frac{\delta(\alpha-\gamma)}{\pi-\gamma}. \tag{25}$$

Differentiating this with respect to β and setting the result to zero gives us the condition from which we can deduce that the minimising value satisfies

$$\beta(\alpha) = \frac{\delta(1-\alpha\pi)}{1-\gamma\pi}. \tag{26}$$

When this is substituted into the criterion function, we obtain a concentrated expression in the form of

$$S(\alpha) = \sigma_\xi^2\frac{\delta^2(\alpha-\gamma)^2}{(1-\gamma^2)(1-\pi\gamma)^2} + \sigma_\varepsilon^2\left\{\frac{(\alpha-\phi)^2}{1-\phi^2}+1\right\}. \tag{27}$$

This is to be compared with the concentrated function which comes from setting $\beta = \delta$ in (19). By differentiating $S(\alpha)$ with respect to α and setting the result to zero, we obtain a condition from which we deduce that the probability limit for the least-squares value of α is

$$\bar{\alpha} = \frac{\kappa\gamma + \lambda\phi}{\kappa + \lambda}, \quad \text{where}$$
$$\kappa = \frac{\sigma_\xi^2 \delta^2}{(1-\gamma^2)(1-\pi\gamma)^2} \quad \text{and} \quad \lambda = \frac{\sigma_\varepsilon^2}{1-\phi^2}. \tag{28}$$

These results contain our previous results as special cases. Again, we see that $\bar{\alpha}$ is a convex combination of γ and ϕ. However, κ no longer represents the variance of the systematic part of the RTM which is given, in fact, by $\kappa(1+\pi\gamma)$. It now transpires that, if the parameter π of the autoregressive process generating $x(t)$ has the same sign as the systematic parameter γ, then additional weight will be given to the γ in forming $\bar{\alpha}$. If the signs of γ and π are opposite, then the weight given to γ will be reduced. These results are illustrated in Table 2 which shows the effect that various values of π have upon the probability limits $\bar{\beta}$ and $\bar{\alpha}$.

Table 2. The effects of fitting the model $(1+\alpha z)y(t) = \beta x(z) + e(z)$ when the true relationship is $y(z) = (1-0.85z)^{-1}x(z) + (1+0.85z)^{-1}\varepsilon(z)$ and $x(t) = (1+\pi L)^{-1}\xi(t)$ is a first-order autoregressive process.

	Case D	Case E	Case F	Case G
σ_x^2	0.75	0.75	0.75	0.75
σ_ξ^2	0.6625	0.6625	0.48	0.1425
σ_ε^2	0.25	0.25	0.25	0.25
π	0.3	−0.3	−0.6	−0.9
$\bar{\alpha}$	−0.228	−0.563	−0.661	−0.700
$\bar{\beta}$	0.851	1.116	1.231	1.575
S	1.901	2.414	2.564	2.624
True Variances				
Systematic	1.604	4.553	8.329	20.299
Disturbance	0.901	0.901	0.901	0.901
Sum	2.505	5.454	9.230	21.200
Estimated Variances				
Systematic	0.500	1.921	4.675	16.058
Disturbance	2.005	3.533	4.554	5.142

In our examples, the positive autocorrelation of $x(t)$ mitigates the distortion of the estimate of the systematic part of the model which is caused by the misspecification. Nevertheless, the distortion is considerable even in Case G where the coefficient of the AR(1) process generating $x(t)$ is as large as $\pi = -0.9$. For, in that case, the probability limit $\bar{\alpha} = -0.70$ implies a median lag in the impulse response of the systematic part of only 0.94 periods, whilst the true parameter $\gamma = -0.85$ implies a median lag of 3.265 periods

7. A MODEL WITH A MOVING-AVERAGE DISTURBANCE

The two misspecified models which we have already investigated both embody the assumption that the disturbance part is a white-noise sequence. Such models, which can be estimated easily by ordinary least-squares regression, are predominant in applied econometrics. Nevertheless it is fairly common nowadays to encounter models with moving-average disturbances. Examples are to be found in the papers by Pagan and Nicholls [14], Trivedi [23] and Zellner and Palm [31]. The monograph by Pesaran and Slater [16] gives details of a method by which such models may be estimated; and it includes a listing of a FORTRAN computer programme which implements the method.

It seems appropriate, therefore, to investigate the effect of fitting a model in the form of

$$(1+\alpha z)y(z) = (\beta_0 + \beta_1 z)x(z) + (1+\mu z)e(z) \tag{29}$$

when the true model is the RTM of (16) and $x(t) = \xi(t)$ is white noise. The generating functions $p(z)$ and $q(z)$ in this case are given by

$$\begin{aligned} p(z) &= \frac{\delta(1+\alpha z)}{(1+\gamma z)(1+\mu z)} - \frac{\beta_0+\beta_1 z}{1+\mu z} \quad \text{and} \\ q(z) &= \frac{1+\alpha z}{(1+\mu z)(1+\phi z)}; \end{aligned} \tag{30}$$

and the asymptotic form of the criterion function becomes

$$\begin{aligned} S(\alpha,\beta_0,\beta_1,\mu) = \sigma_x^2 &\left\{ (G+H-\beta_0)^2 + \frac{J^2}{1-\mu^2} + \frac{J^2\gamma^2}{1-\gamma^2} + \frac{2HJ\gamma}{1-\mu\gamma} \right\} \\ &+ \sigma_\varepsilon^2 \frac{(1+\mu\phi)(1+\alpha^2) - 2\alpha(\mu+\phi)}{(1-\mu^2)(1-\phi^2)(1-\mu\phi)}, \\ \text{where} \quad G = \frac{\delta(\mu-\alpha)}{\mu-\gamma}, \quad & H = \frac{\delta(\alpha-\gamma)}{\mu-\gamma} \quad \text{and} \quad J = (G-\beta_0)\mu + \beta_1. \end{aligned} \tag{31}$$

We should recognise that, in so far as the factor $1/(1-\mu^2)$ stands for a convergent series, our expression for S is admissible only for values of μ which lie in the open interval $(-1,1)$ and which, thereby, satisfy the condition of invertibility. In the event of $|\mu| > 1$, the expression remains meaningful only if μ is replaced by $1/\mu$.

By differentiating S with respect to β_0 and β_1 and setting the results to zero, we get conditions from which we can deduce that the minimising values obey the equations

$$\beta_0 = \delta \quad \text{and} \quad \beta_1(\alpha,\mu) = \frac{\delta(\alpha-\gamma)}{1-\gamma\mu}. \tag{32}$$

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

$$S(\alpha,\mu) = \sigma_x^2 \frac{\delta^2\gamma^2(\alpha-\gamma)^2}{(1-\gamma^2)(1-\mu\gamma)^2} + \sigma_\varepsilon^2 \frac{(1+\mu\phi)(1+\alpha^2) - 2\alpha(\mu+\phi)}{(1-\mu^2)(1-\phi^2)(1-\mu\phi)}. \tag{33}$$

The value of this function is undefined at the points $\mu = -1$ and $\mu = 1$ which represent the boundaries of the admissible region for μ. It is also undefined at the points $\mu = \gamma^{-1}$ and $\mu = \phi^{-1}$ which lie outside the admissible region on the assumption that the parameters of the RTM are bounded by the conditions $|\gamma|$, $|\phi| < 1$.

By differentiating the function with respect to α and setting the result to zero, we obtain a condition from which we find that, for a given value of μ, the minimising value of α is

$$\begin{aligned} \alpha(\mu) &= \frac{\kappa\gamma + \lambda(\mu+\phi)}{\kappa + \lambda(1+\mu\phi)}, \quad \text{where} \\ \kappa &= \frac{\sigma_x^2\delta^2\gamma^2}{(1-\gamma^2)(1-\mu\gamma)^2} \quad \text{and} \\ \lambda &= \frac{\sigma_\varepsilon^2}{(1-\mu^2)(1-\phi^2)(1-\mu\phi)}. \end{aligned} \tag{34}$$

We can see that $\alpha \to 1$ as $\mu \to 1$ and that $\alpha \to -1$ as $\mu \to -1$. It follows from equation (32) that $\beta_1 \to \delta$ as $\mu \to 1$ and that $\beta_1 \to -\delta$ as $\mu \to -1$. Thus, when $|\mu|$ is close to unity, the factors in z within the transfer functions of the fitted model $y(z) = \{(\beta_0 + \beta_1 z)/(1+\alpha z)\}x(z) + \{(1+\mu z)/(1+\alpha z)\}e(z)$ tend to cancel leaving an equation, $y(z) = \delta x(z) + e(z)$, from which the dynamic properties have vanished.

On differentiating the function $S(\alpha,\mu)$ with respect to μ and setting the result to zero, we obtain a condition from which, for a given value of α, we might attempt to find the minimising value of μ. However, since the equation which determines this value of μ is exceedingly nonlinear, the task of minimising $S(\alpha,\mu)$ is more easily accomplished by numerical methods.

By substituting the expression for $\alpha = \alpha(\mu)$ into the concentrated function under (33), we obtain a function $S(\mu) = S\{\alpha(\mu), \mu\}$ which has μ as its sole argument. Thus the minimand of $S(\mu)$ can be found simply by searching the interval $(-1 < \mu < 1)$ in small steps; and values of α, and β_0, β_1 which are associated with this minimand can then be recovered from equations (34) and (32) respectively.

We shall assume, as before, that the RTM of (16) which describes the true relationship between $y(t)$, $x(t)$ and $\varepsilon(t)$ has the parameters $\delta = 1$, $\gamma = -0.85$ and $\phi = 0.85$. Figure 2 gives the graph of the function $S(\mu)$ for the case where $\sigma_\varepsilon^2 = \sigma_x^2 = 0.5$.

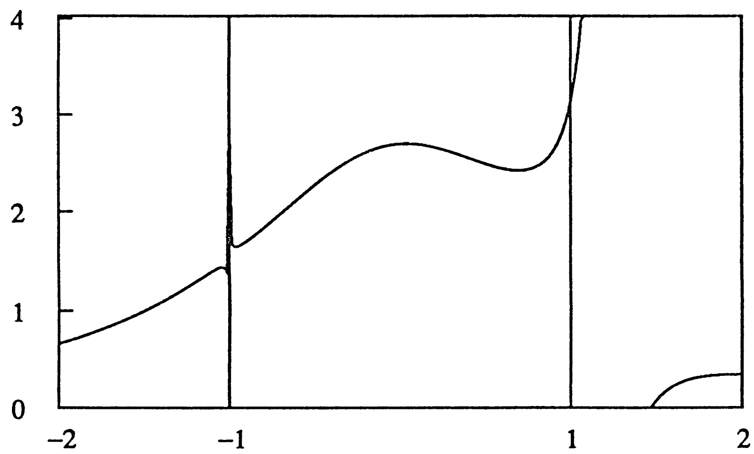

Figure 2. The graph of the function $S(\mu)$ for the case where $\sigma_\varepsilon^2 = \sigma_x^2 = 0.5$.

The fact that the function $S(\mu)$ is undefined at the points $\mu = 1$ and $\mu = \gamma^{-1}$ is not perceptible from its graph. However, there are evident discontinuities in the regions of the values $\mu = -1$ and $\mu = \phi^{-1}$. The most notable feature of the graph is the fact that the function has two minima within the admissible interval $(-1 < \mu < 1)$. Table 3 gives the values of these minima together with the associated values of $\bar{\alpha}$, $\bar{\beta}_0$ and $\bar{\beta}_1$ which are the probability limits of the least-squares estimates. It also gives values for the case where $\sigma_\varepsilon^2 = 0.25$ and $\sigma_x^2 = 0.75$.

The nature of the estimates in the case L(i), which corresponds to a global minimum of the criterion function, is best revealed by writing the fitted GTM in the form of an RTM:

$$y(z) \doteq \frac{1 + 0.07z}{1 - 0.84z} x(z) + \frac{1 - 0.96z}{1 - 0.84z} e(z). \tag{35}$$

Table 3. The effects of fitting the model $(1+\alpha z)y(z) = (\beta_0 + \beta_1 z)x(z) + (1+\mu z)e(z)$ when the true relationship is $y(z) = (1-0.85z)^{-1}x(z) + (1+0.85z)^{-1}\varepsilon(z)$ and $x(t) = \xi(t)$ is white noise.

	Case L(i)	Case L(ii)	Case M(i)	Case M(ii)
σ_x^2	0.5	0.5	0.75	0.75
σ_ε^2	0.5	0.5	0.25	0.25
$\bar{\mu}$	−0.964	0.693	−0.966	0.815
$\bar{\alpha}$	−0.837	0.904	−0.846	0.903
$\bar{\beta}_0$	1.0	1.0	1.0	1.0
$\bar{\beta}_1$	0.073	1.103	0.024	1.035
S	1.636	2.412	0.820	2.671
True Variances				
Systematic	1.802	1.802	2.703	2.703
Disturbance	1.802	1.802	0.901	0.901
Sum	3.604	3.604	3.604	3.604
Estimated Variances				
Systematic	1.880	0.609	2.742	0.822
Disturbance	1.723	2.995	0.862	2.782
Sum	3.604	3.604	3.604	3.604

This shows that the systematic part of the GTM is a reasonable estimate of the systematic part of the true RTM. However, small movements of μ in the neighbourhood of the minimising value are accompanied by large movements in the values of the other arguments of the function. Therefore, we would expect the actual estimates of these values, which would be obtained from finite statistical samples, to have large sampling variances.

The major deficiency of the fitted model of L(i) is its representation of the disturbance part. Here the operator $1+\bar{\mu}z$ in the numerator is virtually cancelled by the operator $1+\bar{\alpha}z$ in the denominator; which suggests, quite wrongly, that the disturbance process resembles white noise.

The estimates in the case L(ii), which correspond to the other minimum of the criterion function, give rise to a very different model:

$$y(z) = \frac{1+1.10z}{1+0.90z}x(z) + \frac{1+0.69z}{1+0.90z}e(z). \tag{36}$$

Here the numerators and the denominators have a tendency to offset each other in both the systematic part and the disturbance part of the model; and so the estimated filters bear little resemblance to the corresponding filters of the true RTM. It could be said, however, that the disturbance part of this model is more sucessful in approximating the true RTM than is the systematic part.

The clearest indication of the misspecification is provided by the sequence of residuals. Their generating function is

$$e(z) = \frac{1+\alpha z}{1+\mu z}\left\{\frac{1+\delta z}{1+\gamma z} - \frac{1+\beta z}{1+\alpha z}\right\}x(z) + \frac{(1+\alpha z)(1+\theta z)}{(1+\mu z)(1+\phi z)}\varepsilon(z); \tag{37}$$

and this is just a specialisation of equation (7) with $\pi(z) = 1$ and $\xi(z) = x(z)$. We are reminded by this equation that, if the values of the fitted parameters were to coincide with those of the true parameters, so as to make $\alpha = \gamma = \phi$, $\beta = \delta$ and $\mu = \theta$, then the residual sequence $e(t)$ would coincide with the white-noise sequence $\varepsilon(t)$.

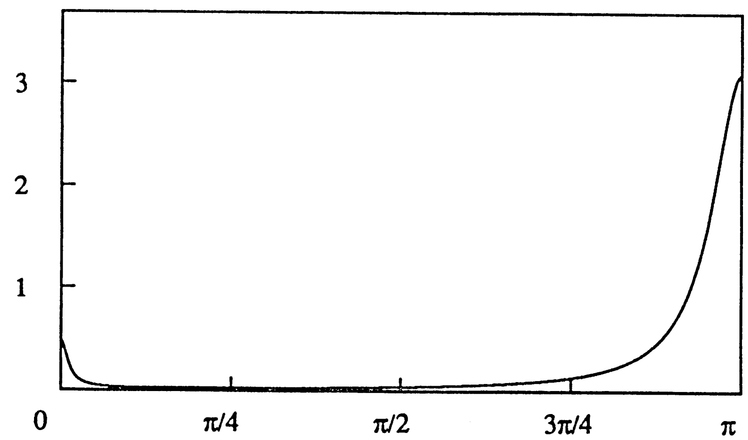

Figure 3. The spectrum of the residual sequence from fitting the GTM under L(i).

The residual sequence, which is the sum of two independent ARMA processes, is an ARMA process in its own right. In the case of the model L(i) it can be represented by

$$e(z) = \frac{1 - 1.67z + 0.69z^2}{1 - 0.96z - 0.72z^2 + 0.69z^3}\zeta(z), \tag{38}$$

where $\zeta(t)$ is a white-noise process with $V\{\zeta(t)\} = 0.511$. The corresponding spectral distribution is shown in Figure 3. When this is compared with the U-shaped spectrum of $y(t)$, we recognise that the fitted model L(i) has successfully accounted for the low-frequency components which are due to the systematic part of the RTM. Its failure to account for the high-frequency components, which are due to the disturbance part of the RTM, is evident from their presence in the residual spectrum.

In the case of the model L(ii), the residual sequence follows a process which can be represented by

$$e(z) = \frac{1 + 0.81z - 0.085z^2}{1 + 0.69z - 0.72z^2 - 0.50z^3}\zeta(z), \tag{39}$$

where $\zeta(t)$ is a white-noise process with $V\{\zeta(t)\} = 1.468$. The spectral distribution is shown in Figure 4. Here the failure of the model is manifest; for the residual spectrum contains much of the power of the systematic and the disturbance parts of the RTM which ought to have been captured by model. It has clearly inherited some of the U shape of the spectrum of $y(t)$. However, there is more power in the low-frequency components than in the high-frequency components. This confirms our assertion that the disturbance part of the model L(ii) is more successful in approximating the true RTM than is the systematic part.

An attempt at fitting the GTM of (29) when the true relationship is described by the RTM of (16) can lead to either of two quite different failures. It not too fanciful to say that, on this occasion, the tension which exists between the two parts of the RTM in their struggle to preempt the operator $\alpha(z)$ has resulted in a rupture instead of a compromise.

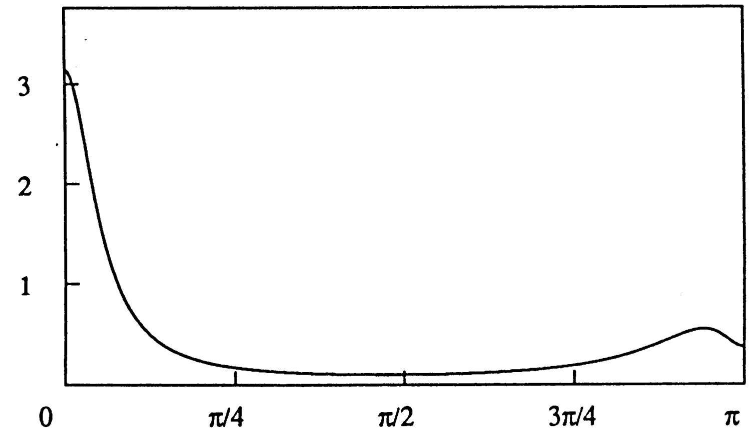

Figure 4. The spectrum of the residual sequence from fitting the GTM under L(ii).

The spectra in Figures 3 and 4 are quite different from that of a white-noise process. Therefore we might be sanguine about our ability to detect misspecification through diagnostic tests based on the least-squares residuals from fitting the models to finite sets of data.

The empirical counterpart of the spectrum is the periodogram. However, unless it is smoothed, the periodogram has a highly volatile appearance which makes it difficult to interpret visually. A better way of presenting the information is to plot the cumulated periodogram.

The cumulated spectrum of a white-noise process rises in a straight line; and, when the cumulated periodogram of an empirical sequence diverges significantly from this line, we may suspect that there is serial correlation. Significance limits, which are based on the Kolmogorov–Smirnoff statistic, have been given by Bartlett [2, p. 318]. These limits can also be used in an approximate way for assessing the significance of the correlation in the least-squares regression residuals.

Durbin [6] has developed significance bounds which are appropriate to the cumulated periodogram of the residuals of a classical regression model; but their validity for finite samples does not extend to the case of a model with lagged dependent variables.

The advantage of tests based on the cumulated periodogram is that they may provide a useful indication of the nature of the misspecification as well as a rough indication of its extent.

It is straightforward to derive tests for particular kinds of misspecification by following the Lagrange multiplier procedure of Rao [20] and Silvey [21]. Details of such tests for the RTM and the GTM have been given by Poskitt and Tremayne [19]. In particular, a statistic is readily accessible for testing the hypothesis that $\gamma(z) = \phi(z)$.

The significance levels for such statistics are derived from their asymptotic distributions; and, as Kiviet [10,11] has show, they are often unreliable when data samples are of a size which is typical in applied econometrics. In fact, the probability of a Type I error in these tests may vary substantially for different parameter values of the process generating the data; and this finding conflicts with the asymptotic theory which suggests that the probabilities are invariant.

8. CONCLUSIONS

Our experiments have revealed some of the dangers of fitting a GTM with too few parameters when the true model is an RTM. The problems stem from the fact that the GTM uses a common set of parameters in its attempt to capture the dynamic properties of the systematic part and the disturbance part of the regression relationship. If the properties of these two parts differ, then they are bound to be misrepresented in the fitted model which can be grossly misleading.

It might be argued that the problem could be overcome by increasing the number of the parameters in the fitted GTM. Thus we have only to replace the model in (29) by the model

$$(1+\alpha_1 z+\alpha_2 z^2)y(z)=(\beta_0+\beta_1 z)x(z)+(1+\mu z)e(z), \tag{40}$$

which incorporates an extra parameter in $\alpha(z)$, and we will succeed in consistently estimating the RTM of (16) which becomes a special case of the extended model.

To argue in this way would be to misunderstand the real problems of econometrics where any model which is practical is bound to give a simplified account of the data-generating process in terms of a limited number of parameters. Therefore, if we are to avoid the danger of confounding the properties of the two parts of a dynamic regression relationship, we must attribute a distinct set of parameters to each part; and this means using a rational transfer-function model. We should use the alternative general temporal model only if we can show that none of the dangers which we have illustrated in this paper can arise; and we are rarely in a position to do so.

Finally we may say that, if one adopts a model, such as the RTM, which is capable making adequate approximations to a wide range of the data-generating processes, then one need not be too worried about the unreliability in small samples of the usual diagnostic tests which are intended to reveal the inadequacy of a model.

REFERENCES

1. Amemiya, T. *Advanced Econometrics.* Cambridge: Basil Blackwell, 1985.
2. Bartlett, M.S. *An Introduction to Stochastic Processes,* (2nd ed.). Cambridge: Cambridge University Press, 1966.
3. Bloomfield, P. On the error of prediction of a time series. *Biometrika* 59 (1972): 501–507.
4. Cragg, J. C. Estimation and testing in time-series regression models with heteroskedastic disturbances. Journal of *Econometrics* 20 (1982): 135–157.
5. Domowitz, I. and H. White. Misspecified models with dependent observations. *Journal of Econometrics* 20 (1982): 35–58.
6. Durbin, J. Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals, *Biometrika,* 56 (1969): 1–15.
7. Hendry, D.F. The behaviour of inconsistent instrumental variable estimators in dynamic systems with autocorrelated errors. *Journal of Econometrics* 9 (1979): 295–314.
8. Jakeman, A.J., L.P. Steele and P.C. Young. Instrumental variables algorithms for multiple input systems described by multiple transfer functions. *IEEE Transactions on Systems, Man and Cybenetics.* SM 10 (1980): 593–602.

9. Kabaila, P.V. and G.C. Goodwin. On the estimation of the parameters of an optimal interpolator when the class of interpolators is restricted. *SIAM Journal of Control and Optimization* 18 (1980): 121-144.

10. Kiviet, J. F. Model selection test procedures in a single linear equation of a dynamic simultaneous system and their defects in small samples. *Journal of Econometrics* 28 (1985): 327–362.

11. Kiviet, J. F. On the rigour of some misspecification tests for modelling dynamic relationships. *Review of Economic Studies* 53 (1986): 241–261.

12. Ljung, L. On the consistency of prediction error identification methods. In R.K. Mehra and D.G. Lainiotis (eds.), *System Identification: Advances and Case Studies.* New York: Academic Press, 1976.

13. Maasoumi, E. and P.C.B. Phillips. On the behaviour of inconsistent instrumental variable estimators. *Journal of Econometrics* 19 (1982): 183–201.

14. Pagan, A.R. and D.F. Nicholls. Exact maximum likelihood estimation of regression models with finite order moving average errors. *Review of Economic Studies* 43 (1976): 383–288.

15. Pandit, S.M. and S.M. Wu. *Time Series and System Analysis with Applications.* New York: John Wiley and Sons, 1983.

16. Pesaran, M.H. and L.J. Slater. *Dynamic Regression : Theory and Algorithms.* Chichester: Ellis Horwood, 1980.

17. Pierce, D. A. Least squares estimation in the regression model with autoregressive moving average errors. *Biometrika* 58 (1971): 299–312.

18. Pollock, D.S.G. The misspecification of ARMA models. Economic Research Bureau, Queen Mary College, University of London, 1986.

19. Poskitt, D.S. and A.R. Tremayne, An approach to testing linear time series models. *The Annals of Statistics* 6 (1981): 974–986.

20. Rao, C.R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Proceedings of the Cambridge Philosophical Society* 44 (1948): 50–57.

21. Silvey, S.D. The Lagrange multiplier test. *The Annals of Mathematical Statistics* 30 (1959): 389–407.

22. Söderström, T. and P. Stoica. Comparison of some instrumental variable methods: consistency and accuracy aspects. *Automatica* 17 (1981): 101-115.

23. Trivedi, P. K. Inventory behaviour in U.K. manufacturing, 1956–67. *Review of Economic Studies* 37 (1970): 517–527.

24. Walker, A.M. Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time series. *Journal of the Australian Mathematical Society* 4 (1964): 363–384.

25. White, H. Maximum likelihood estimation of misspecified models. *Econometrica* 50 (1982): 1–25.

26. White, H. Maximum likelihood Estimation of misspecified dynamic models. In: Theo Dijkstra (ed.), *Misspecification Analysis.* Berlin: Springer-Verlag, 1984.
27. White, H. Specification testing in dynamic models. In: T.F. Bewley (ed.), Advances in Econometrics. Fifth World Congress. New York: Cambridge University Press, 1987.
28. White, H. *Estimation, Inference and Specification Analysis,* (forthcoming). New York: Cambridge University Press.
29. Whittle, P. *Hypothesis Testing in Time Series Analysis.* Uppsala: Almquist and Wicksell, 1951.
30. Young, P.C.A. Instrumental variable method for real time identification of a noisy process. *Automatica* 6 (1970): 271–287.
31. Zellner A. and F. Palm. Time series analysis and simultaneous equation econometric models. *Journal of Econometrics* 2 (1974): 17–54.

APPENDIX: THE LEAST-SQUARES CRITERION FUNCTION

Realisations of the Relationship

A set of T realisations of the relationship in (1) running from $t = 0$ to $t = T - 1$ can be represented by the equation

$$Y\alpha = X\beta + \mathcal{E}\mu. \qquad (A.1)$$

Here $Y = [y_{i-j}]$ is a $T \times T$ Toeplitz matrix in which each diagonal band contains repetitions of a single element of $y(t)$. Thus, the principal diagonal contains the initial sample element y_0 whilst successive subdiagonal bands contain the elements $y_1, \ldots, y_{T-1}$. The supradiagonal bands contain the presample elements $y_{-1}, \ldots, y_{1-T}$. The matrices $X = [x_{i-j}]$ and $\mathcal{E} = [\varepsilon_{i-j}]$, which are based on the elements of $x(t)$ and $\varepsilon(t)$ respectively, have structures which are similar to that of Y. The equation (A.1) also comprises $\alpha = [1, \alpha_1, \ldots, \alpha_a, 0, \ldots, 0]'$, $\beta = [\beta_0, \beta_1, \ldots, \beta_b, 0, \ldots, 0]'$ and $\mu = [1, \mu_1, \ldots, \alpha_m, 0, \ldots, 0]'$. These are vectors of order T containing the parameters of the model in the leading positions followed by zeros.

The corresponding T realisations of the equation (5), which indicates how the sequence $x(t)$ is generated, can be written as

$$X\pi = \xi, \qquad (A.2)$$

where $\xi = [\xi_0, \xi_1, \ldots, \xi_{T-1}]'$, which is the leading vector of the matrix $\Xi = [\xi_{i-j}]$, contains elements from the white-noise process $\xi(t)$.

We shall simplify our analysis by assuming that the presample elements of $\varepsilon(t)$ and $\xi(t)$ are all zeros, which implies that the presample elements of $y(t)$ and $x(t)$ are also zeros. A more elegant assumption would be to regard $\varepsilon(t)$

and $\xi(t)$ as stationary stochastic processes defined over the entire set of positive and negative integers. In that case, we would need to demonstrate that the presample elements play a role within the criterion function which becomes negligible as the sample size T increases. Such a demonstration has been provided by Pollock [18] in a companion paper for the case of the autoregressive moving-average model. Pierce [17] has also dealt with the matter in a paper on the subject of the estimation of the RTM. By assuming that the presample elements are zeros, we can avoid having to repeat the arguments, and we can remove some otiose complications from our analysis.

When all the presample elements are zeros, the matrices X, Y, $\mathcal{E}$ and Ξ assume a lower-triangular form. In this form, they are completely characterised by their leading vectors which are given by $x = Xe_1$, $y = Ye_1$, $\varepsilon = \mathcal{E}e_1$ and $\xi = \Xi e_1$, where e_1 is the leading vector of the identity matrix of order T. To express the fact that the matrix Ξ, for example, is completely determined by the vector ξ, we may write $\Xi = \Xi(\xi)$. We shall define a set of analogous triangular matrices based on the parameter vectors; and we shall write these as $A = A(\alpha)$, $B = B(\beta)$, $M = M(\mu)$ and $\Pi = \Pi(\pi)$.

Since banded lower-triangular matrices commute in multiplication, we find that $Y\alpha = (YA)e_1 = (AY)e_1 = Ay$, and, likewise, that $X\beta = Bx$, $\mathcal{E}\mu = M\varepsilon$ and $X\pi = \Pi x$. It follows that we can rewrite equation (A.1) in the equivalent form of

$$Ay = Bx + M\varepsilon. \tag{A.3}$$

Equation (A.2) can be rewritten likewise as

$$\Pi x = \xi. \tag{A.4}$$

The Expected Criterion Function

Now let us imagine that the true model is an RTM. If this model is to be represented by equation (A.3), then the matrices must factorise to give $A = \Gamma\Phi$, $B = \Phi\Delta$ and $M = \Gamma\Theta$, where $\Gamma = \Gamma(\gamma)$, $\Phi = \Phi(\phi)$, $\Delta = \Delta(\delta)$ and $\Theta = \Theta(\theta)$ are all banded lower-triangular matrices. It follows that

$$\begin{aligned} y &= A^{-1}Bx + A^{-1}M\varepsilon \\ &= \Gamma^{-1}\Delta\Pi^{-1}\xi + \Phi^{-1}\Theta\varepsilon. \end{aligned} \tag{A.5}$$

Imagine that the fitted model is a GTM of the form $Ay = Bx + Me$. Then the residual vector e is given, in terms of the parameters of both the fitted model and the true model, by

$$
\begin{aligned}
e &= M^{-1}Ay - M^{-1}Bx \\
&= M^{-1}A\{(\Gamma^{-1}\Delta - A^{-1}B)\Pi^{-1}\xi - \Phi^{-1}\Theta\varepsilon\} \\
&= P\xi + Q\varepsilon,
\end{aligned}
\tag{A.6}
$$

where $P = M^{-1}A(\Gamma^{-1}\Delta - A^{-1}B)\Pi^{-1}$ and $Q = -M^{-1}A\Phi^{-1}\Theta$ are both banded lower-triangular matrices. The sum of squares of the residuals, which constitutes the criterion function of least-squares estimation, is given by

$$
\begin{aligned}
e'e &= \xi'P'P\xi + \varepsilon'Q'Q\varepsilon + 2\xi'P'Q\varepsilon \\
&= p'\Xi'\Xi p + q'\mathcal{E}'\mathcal{E}q + 2p'\Xi'\mathcal{E}q,
\end{aligned}
\tag{A.7}
$$

where $p = Pe_1$ and $q = Qe_1$.

If we assume that $\xi(t)$ is a white-noise process which is independent of $\varepsilon(t)$, then we have

$$
E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 I \quad , \quad E(\xi\xi') = \sigma_\xi^2 I \quad \text{and} \quad E(\varepsilon\xi') = 0; \tag{A.8}
$$

and it follows that the expected value of the criterion function is

$$
E(e'e) = \sigma_\xi^2 \text{Trace}\{P'P\} + \sigma_\varepsilon^2 \text{Trace}\{Q'Q\}. \tag{A.9}
$$

The Convergence of the Criterion Function

It can be shown easily that, if it is scaled by the factor T^{-1}, the expected criterion function converges to a limiting form as T increases. The scaled function can be written as

$$
\frac{1}{T}E(e'e) = \sigma_\xi^2 \sum_{j=0}^{T-1} p_j^2 \frac{T-j}{T} + \sigma_\varepsilon^2 \sum_{j=0}^{T-1} q_j^2 \frac{T-j}{T}. \tag{A.10}
$$

As $T \to \infty$, the terms on the RHS of this expression, which are the Cesaro sums of the convergent sequences $\{q_0^2, q_1^2, q_2^2, \ldots\}$ and $\{p_0^2, p_1^2, p_2^2, \ldots\}$, must themselves converge to the values of $S_q/\sigma_\varepsilon^2 = \lim(T \to \infty)\sum q_j^2$ and $S_p/\sigma_\xi^2 = \lim(T \to \infty)\sum p_j^2$ respectively. These are the components of the expression $S = S_p + S_q$ of (9) which gives the variance of the residual sequence.

It is also straightforward to show that, as T increases, the criterion function in (A.7), scaled by T^{-1}, converges to the same limiting value as does its expectation. Thus, by invoking the law of large numbers, we can establish that the probability limit of the jth diagonal element of $\Xi'\Xi/T$ is

$$
\text{plim}(T \to \infty) \sum_{t=0}^{T-j} T^{-1}\xi_t^2 = \sigma_\xi^2, \tag{A.11}
$$

and that the probability limit of an element in an off-diagonal position is zero. It follows that $\text{plim}(p'\Xi'\Xi p/T) = S_p$. Likewise, we find that $\text{plim}(q'\mathcal{E}'\mathcal{E}q/T) = S_q$ and that $\text{plim}(p'\Xi'\mathcal{E}q/T) = 0$.

Given that the criterion function and its expected value converge to the same limiting form, it follows—according to a fundamental result which is proved by Amemiya [1, Theorem 4.1.1] and by Domowitz and White [5, Theorem 2.2] amongst others—that the values which minimise the asymptotic form of the expected criterion function must also be the probability limits of the estimates.