Comments on Granger-Newbold's "Time Series Approach to Econometric Model Building" and Sargent-Sims' "Business Cycle Modeling Without Pretending to Have Too Much *A Priori* Economic Theory"

by David F. Hendry†

Granger-Newbold

I. Introduction

It is certainly impossible to disagree with Granger and Newbold [45]‡ on the desirability of incorporating the best features of time series analysis into econometric practice. Moreover, many of Granger and Newbold's criticisms of econometric modeling are obviously valid and should be implemented as rapidly as possible in the conduct of empirical research. Indeed, econometricians have voiced similar criticisms of current methodology, and the additional support of Granger-Newbold will hopefully increase the probability of changing the state of the art in an appropriate direction.

Nevertheless, one must not misconstrue the *solution* to such criticisms since it would be equally unproductive to adopt *only* a time series approach which neglected econometric methods. Also, since econometrics has derived benefit from a time series based critique, it would seem useful to undertake the converse in the hope of facilitating further interaction. Thus, I will first consider whether more appropriate econometric methodology than *OLS* with only Durbin-Watson statistic diagnostic testing can help resolve the "nonsense regressions" problem and then discuss certain difficulties which can arise in selecting appropriate univariate

†This research was undertaken in part while the author was visiting at the Cowles Foundation, Yale University, supported by grants from the National Science and Ford Foundations. He is grateful to Kenneth F. Wallis for valuable comments on an earlier draft of the paper and to Stephen M. Goldfeld, Cheng Hsiao, Charles F. Manski, and Thomas J. Rothenberg for helpful discussions about the material herein.

‡ Numbers in [ ] correspond to reference list, p. 219.

ARMA models for stationary data series. Next, the role of differencing will be examined, and finally, the systems case will be briefly considered.

At the outset, however, Granger and Newbold are incorrect in suggesting that econometrics has not been much concerned with lag structures and/or with attempting to overcome *ad hoc*-ery in theories which incorporate lags.† Further, contrary to an impression also supported by Naylor *et al.* [101], econometric estimation is not and never has been synonymous with *simultaneous* equations estimation. Considerable attention has been devoted to estimating models with autocorrelated residuals, including both autoregressive schemes‡ and moving average representations.§ I certainly agree that all too often autocorrelation has been treated as a "nuisance" and that usually only low-order processes have been examined, but this is not universal.‖ These comments are not intended as an apologia; rather they are an attempt to stress that the "two philosophies" are not as distinct as Granger and Newbold initially suggest.

II. Autocorrelation Transforms Applied to Nonstationary Data Series

The results in Tables I and II of Granger-Newbold# provide two salutory warnings:

- Not to misinterpret the "significance" of regression coefficients based on highly autoregressive data when equations have high R^2 but low Durbin-Watson statistic values.
- Mechanical use of the Cochrane-Orcutt transformation to "correct" autocorrelation will not solve the problem of spurious significance when residual autocorrelation actually represents misspecified dynamics.††

These criticisms are certainly valid, but may be labeled as applying to "poor average" practice. Thus, a more appropriate procedure would include the test for the validity of the autoregressive transform proposed by Sargan [128] and described again below. Further, this test provides a constructive diagnostic statistic for one important aspect of dynamic specification.‡‡ Indeed, such a procedure has interesting implications when applied to the "spurious regressions" problem discussed by Granger-Newbold.

†See, among others, Kyock [78], Nerlove [103], Fisher [27], Jorgenson [69], and Dhrymes [19].

‡See Cochrane and Orcutt [15], Sargan [127], and Fisk [28].

§See Klein [73], Phillips [117], and Trivedi [148].

‖For counter examples to each of these statements, see Sargan [128] and Wallis [8] — note that the former also provides the test statistic for second-order autocorrelated residuals which Granger-Newbold request; but see Durbin [20] for a slight correction.

#See pp. 12-13 in this volume.

††Compare Prais [122].

‡‡See Hendry [56].

Consider their equation

$$(1) \quad Y_t = \beta_1 + \beta_2 X_t + e_t$$

where Y_t and X_t are independent $IMA(1,1)$ processes. Direct *OLS* estimation of (1) produces the results Granger-Newbold show in their Table I. If the (false) assumption is made that the low Durbin-Watson values arise from the process

$$(2) \quad e_t = \rho e_{t-1} + v_t \qquad |\rho| < 1$$

then applying the Cochrane-Orcutt procedure generates the outcomes in Table II. The autocorrelation transform reduces but does not completely remove the spurious significance problem.

However (1) and (2) are equivalent to

$$(3) \quad Y_t = \beta_1(1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + \rho Y_{t-1} + v_t$$

which is a restricted version of

$$(4) \quad Y_t = \gamma_1 + \gamma_2 X_t + \gamma_3 X_{t-1} + \gamma_4 Y_{t-1} + w_t.$$

Sargan [128] proposed testing the validity of the autoregressive restriction in (3) using the result that if k valid restrictions are imposed, then twice the natural logarithm of the likelihood ratio is asymptotically distributed as χ^2_k. If v_t in (3) is normally distributed (as it is in Granger and Newbold's study) then

$$(5) \quad z_1 = T\, ln(\Sigma\, \hat{v}_t^2 / \Sigma\, \hat{w}_t^2) \underset{A}{\sim} \chi^2_k$$

when (4) *is* the unrestricted version of (3). On the other hand, if the apparent autocorrelation in (1) arises because it is a misspecified approximation to (4), the latter will produce a better fit than (3) and hence too large a value for z_1. It would be interesting to know how often (5) led to the rejection of (3) against (4) at a conventional significance level for χ^2_1 in Granger and Newbold's *ARIMA* model. I hazard the guesses that z_1 would reject (3) reasonably frequently relative to the number of cases of spurious significance in Table II and also that in (4), X_t and X_{t-1} would now rarely have a significant effect – certainly their individual coefficients could not generally be significant due to collinearity and this finding should prompt a further revision of the dynamic specification.

Precisely how one proceeds at this stage will depend on the status of the theory relating Y_t to X_t. If the study is simply an exploratory empirical modeling of the data set, then it would seem reasonable to repeat the analysis described above with (4) constituting the new baseline; alternatively, one could recommence the exercise using a specialization of (4) obtained

by deleting regressors with 't' values smaller in absolute value than some positive number c (for example, c might be unity). Either way, providing the refitting is conducted allowing for w_t to be autocorrelated,† then it is hard to see why an approximately correct model could not be detected even for the paradigm used by Granger-Newbold. Clearly, these suggestions could be checked in a simulation study similar to that reported by Granger-Newbold.

As with any iterative model revision approach (including Box-Jenkins methods), preliminary test biases may be a serious problem.‡ Thus, Peck [114] provides simulation evidence that (on a median absolute error criterion) for almost all values of the autoregressive error parameter ρ in (2), the maximum likelihood estimator allowing for autocorrelation like (2) is superior to most other estimators based on preliminary tests of the significance of ρ for most significance levels (except, of course, automatic rejection of $Ho{:}\rho{=}0$). On the other hand, this supports the principle (suggested above) that refitting should always allow for potential autocorrelation and that (2) is a simple and inexpensive process to estimate using modern computing equipment.§

Nevertheless, (4) is still an incorrect approximation to the *ARIMA* process

$$(6) \quad Y_t = Y_{t-1} + \eta_t + b^*\eta_{t-1}$$

which actually generated the data, and even if the X_t regressors were deleted, (4) would continue to remain inappropriate if the autocorrelation in w_t was assumed to be autoregressive. This remark raises the issue of the consequences of using incorrect approximations to *ARMA* or *ARIMA* processes, and since Granger and Newbold have described allowance for autoregressive errors like (2) as "naive," the next section will consider such incorrect approximations in some detail for simple *ARMA* models.

III. Mis-specification Analysis of an *ARMA* (1,1) Model

To illustrate the issues involved, it seems adequate to use an asymptotic analysis and examine the simplest case in which the data are generated by a stationary and invertible *ARMA*(1,1) model:

$$(7) \quad Y_t = \delta Y_{t-1} + \epsilon_t + a\epsilon_{t-1}$$

where $\epsilon_t \sim NI(0,\sigma_\epsilon^2)$ and $|\delta|, |a| < 1$.

Firstly, consider a situation in which "preliminary identification" as

†Compare Hendry [56].

‡See, for example, Bock *et al.* [8] and for a discussion of sequential testing procedures for autocorrelation see Mizon [95].

§Higher-order autoregressive schemes are also relatively easy to estimate.

proposed by Box and Jenkins [9] suggests the overly parsimonious *ARMA* (1,0) approximation to (7) given by

(8) $Y_t = \lambda Y_{t-1} + u_t$

where u_t is incorrectly assumed to be white noise.

Then

(9) $plim\ \hat{\lambda} = plim\ T^{-1}\Sigma Y_{t-1} Y_t / T^{-1}\Sigma Y_{t-1}^2 = \delta + \alpha^* = \lambda_p$

(the subscript p denotes the *plim* of the estimator of the parameter) where $\alpha^* = \phi(1-\delta^2)/(1+2\delta\phi)$ and $\phi = \alpha/(1+\alpha^2)$ (the first order error autocorrelation coefficient).

Table A records numerical values of the inconsistency in $\hat{\lambda}$ as an estimator of δ at various values of α for $\delta=0$ (an *ARMA* (1,0) approximation to an *ARMA* (0,1) process) and $\delta = 0.8$. And $\hat{\lambda}$ is of course a consistent estimator of the first order autocorrelation coefficient of Y_t.

More important is the question of detecting such a mis-specification using "diagnostic" tests based on residual autocorrelations. The residuals from (8) are given by

(10) $\hat{u}_t = Y_t - \hat{\lambda} Y_{t-1} = (\delta - \hat{\lambda}) Y_{t-1} + \epsilon_t + \alpha\epsilon_{t-1}$

so that the residual variance is $\hat{\sigma}_u^2 = T^{-1}\Sigma\hat{u}_t^2$ and

(11) $plim\ \hat{\sigma}_u^2 = \sigma_{up}^2 = \sigma_\epsilon^2(1+\alpha(\alpha-\alpha^*))$

implying that

(12) $\sigma_u^2 = \sigma_\epsilon^2(1+\alpha^2) \geq \sigma_{up}^2 \geq \sigma_\epsilon^2.$

The residual autocorrelations (denoted r_g) have *plims* given by

(13) $r_{1p} = plim(\Sigma\, \hat{u}_t\hat{u}_{t-1} / \Sigma\, \hat{u}_{t-1}^2) = (\alpha-\alpha^*)(1-\alpha\alpha^*)/(1+\alpha(\alpha-\alpha^*))$

(14) $r_{2p} = -\alpha^*(\alpha+\delta+\alpha\delta(\alpha-\alpha^*))/(1+\alpha(\alpha-\alpha^*))$

and $r_{gp} = \delta r_{g-1,p}$, $g \geq 3$. Numerical values of $\sigma_{up}^2/\sigma_\epsilon^2$, r_{1p} and r_{2p} are also shown in Table A for $\delta=0$ and $\delta=0.8$.

Table A†

Plims of the Parameter Estimates of an *ARMA* (1,0) Approximation to an *ARMA* (1,1) Process

	α	−1.	− .8	− .5	0	.5	.8	1.
	λ_p	− .50	− .49	− .40	0	.40	.49	.50
$\delta=0$	$\sigma_{up}^2/\sigma_\epsilon^2$	1.50	1.25	1.05	1	1.05	1.25	1.50
	r_{1p}	− .17	− .15	− .08	0	.08	.15	.17
	r_{2p}	− .33	− .31	− .19	0	− .19	− .31	− .33
	λ_p	− .10	0	.40	.8	.89	.90	.90
$\delta=.8$	$\sigma_{up}^2/\sigma_\epsilon^2$	1.10	1	1.05	1	1.21	1.56	1.90
	r_{1p}	− .01	0	− .08	0	.32	.41	.43
	r_{2p}	− .17	0	+ .11	0	− .11	− .13	− .13

† The figures in Tables A, B, C, and D are based on calculations rounded for ease of computation. ($\alpha=\pm 1$ is an approximation to $|\alpha| = .999\ldots$, where $|\phi|$ is a maximum.)

The most dramatic feature of the results is that if $\alpha\delta < 0$, λ_p is highly inconsistent for δ, σ_{up}^2 is close to σ_ϵ^2, and both r_{1p} and r_{2p} are close to zero. This is clearly connected to the fact that if $\alpha=-\delta$, (7) has a common root and reduces to $y_t = \epsilon_t$.

While these results do not directly clarify whether "diagnostic tests" based on residual autocorrelations could detect the above mis-specification with reasonable certainty in large samples, we can obtain useful evidence as follows. Consider a likelihood-ratio based test of $\alpha=0$ (a comparison of an *ARMA* (1,0) with an *ARMA* (1,1) which would correspond to "overfitting" for the most likely mis-specification.† This would be based on the assumption that

$$(15) \quad z_2 = T\ln(\hat{\sigma}_u^2/\hat{\sigma}_\epsilon^2) \underset{A}{\sim} \chi^2(1,0) \text{ on } Ho{:}\alpha=0$$

where $\chi^2(K, T\mu)$ is a non-central chi-square with K degrees of freedom and non-centrality parameter $T\mu$. The test consists of rejecting Ho if $z_2 \geq \chi_a^2(1,0) = c$ where $\chi_a^2(1,0)$ is the $100(1-a)$ percentage point of the $\chi^2(1,0)$ distribution.

On general grounds, one would expect z_2 to have good asymptotic properties: consistent, asymptotically unbiased, high asymptotic relative efficiency and asymptotically most powerful in the class of tests whose asymptotic power depends only on the non-centrality parameter.‡ When Ho is false and $\alpha \neq 0$, then for a sequence of alternatives $\alpha = \frac{2}{\sqrt{T}}$

†See Prothero and Wallis [124].

‡See, for example, Kendall and Stuart [72, chapters 24 and 25] and Lehmann [80, chapter 7].

$$z_2 \tilde{A} \chi^2(1, T\mu)$$

where $\mu = ln(\sigma_{up}^2/\sigma_\epsilon^2) = ln(1+a(a-a^*))$ [$\approx a(a-a^*) \geq 0$ for small $a(a-a^*)$]. The approximate power of z_2 against a sequence of alternative values of $a \neq 0$ for various finite T can be evaluated if $\chi^2(K, T\mu)$ is approximated as proportional to a central chi-square with the same first two moments; namely, $h\chi^2(K^*,0)$ where

(16) $h = 1 + T\mu/(K+T\mu)$ and $K^* = K + (T\mu)^2/(K+2T\mu)$.

Then a rough approximation to the power function of z_2 is given by

$$(17) \quad P_T(z_2/a) = \int_{h^*}^{\infty} d\chi^2(K^*,0)$$

where $h^* = c(K+T\mu)/(K+2T\mu)$ and $K^* = h^*(K+T\mu)/c$, when the nominal and large sample size of the test is a with $c = \chi_a^2(K,0)$ (although a is only an approximation to the actual size of the test for finite T). The above procedure is discussed in Kendall and Stuart [72, chapter 24]. Note that $K=1$ for z_2 defined by (15). Since $a=0$ implies than $\mu=0$ and, hence, that $h^*=c$, $K^*=1$, it is readily verified from (17) that $P_T(z_2/0) = a$. Moreover, for large T, $h^* \to .5c$ and $K^* \approx .5(1+T\mu)$ so that $P_T(z_2/a) \to 1$ if $a \neq 0$ or $-\delta$, and hence, the test is confirmed as consistent. It is also clear that the power defined by (17) increases monotonically with $T\mu$, but if K^* is non-integer, then interpolation between neighboring chi-square values must be used in computing (17).

Table B records values of $P = P_T(z_2/a)$ at $T=40$, 80 and $a=.05$ (so $c=3.84$) for the same range of a and δ as used in Table A; for simplicity, K^* was rounded to the nearest integer as shown in parentheses.

Table B
Approximate Powers of the Likelihood Ratio Test z_2

	$T \backslash a$	-1.0	-0.8	-0.5	0	0.5	0.8	1.0
$\delta=0$	40 K^*	10.8(11)	5.8(6)	1.8 (2)	1	—	—	—
	40 P	1.0	0.9	0.3	0.05	—	—	—
	80 K^*	20.8(21)	10.8(11)	2.8(3)	1	—	—	—
	80 P	1.0	1.0	0.5	0.05	—	—	—
$\delta=.8$	40 K^*	3.1(3)	1	1.8(2)	1	4.8(5)	12.0(12)	18.8(19)
	40 P	0.5	0.05	0.3	0.05	0.8	1.0	1.0
	80 K^*	4.8(5)	1	2.8(3)	1	8.8(9)	23.2(23)	36.8(37)
	80 P	0.8	0.05	0.5	0.05	1.0	1.0	1.0
$\delta=0$	40 $P(z_3)$	1.0	1.0	0.9	0.05	—	—	—

Thus, given the large number of approximations involved in an already asymptotic analysis, the power is recorded to one significant figure only. Nevertheless, several points are clear. Firstly, when $\delta=0$, the power drops very sharply for $|a| < .8$, and even at $T=80$, when $a=\pm.5$, P is only about 1/2. Secondly, when $\delta=.8$, P is small for $a<.5$. And thirdly, to keep the relative magnitude of these values of $P_T(z_2/a)$ in perspective, the last line of Table B records the correspondingly computed powers of the test of $a=0$ in $Y_t = \epsilon_t + a\epsilon_{t-1}$ when $T=40$ and where an *ARMA* (1,0) is *not* first (incorrectly) fitted; that is

$$(18) \quad z_3 = Tln(\hat{\sigma}_Y^2/\hat{\sigma}_\epsilon^2) \underset{A}{\sim} \chi^2(1, Tln(1+a^2)).$$

Thus, $P_{40}(z_3/a)$ is 0.9 or larger for $|a| \geq 1/2$.

One concludes that for samples of a size usual in empirical economics, it is far from easy to detect the incorrect approximation of an *ARMA* (1,1) process by an *ARMA* (1,0) model for a considerable range of values of a and δ *even when the correct form of the alternative hypothesis is guessed.* Smaller Type II errors would obviously occur if a much larger size of test was used (for example, $a=.2$), and this would seem reasonable given the costs of under *versus* over specification. On the other hand, Box and Jenkins [9] have stressed the need for choosing parsimonious models in view of the identification problem in *ARMA* models having common roots like $a=-\delta$. Again, the above results should be checked by an appropriate Monte Carlo study before being accorded higher status than illustrative (see below), but they may represent one factor explaining the low orders of lag polynomials for directly estimated *ARMA* models compared to derived final forms.†

In practice, diagnostic tests rarely seem to be based on likelihood-ratio arguments in time series analysis, presumably because the form of a reasonable alternative hypothesis is difficult to specify.‡ A more usual form of test is represented by the portmanteau statistic proposed by Box and Pierce [11] which is based on the assumption that if $a=0$ in (7) then

$$(19) \quad Q = T \sum_{j=1}^{J} \hat{r}_j^2 \underset{A}{\sim} \chi^2(J-1,0)$$

for some (large) number $J>1$. The distribution of Q for $a\neq 0$ does not seem to have been derived and is not easily obtained. However, the difficulty is *not* the same as that concerning the distribution of the Durbin-Watson (d) statistic when it is incorrectly applied to test for

†See Prothero and Wallis [124].

‡See, however, Sargan [128] and Durbin [20].

residual autocorrelation in equations like (7). Nerlove and Wallis [105] demonstrated that when the null hypothesis of no error autocorrelation was false, the d test was biased towards its H_0 plim of 2. Nevertheless, Durbin [20] showed that the *size* of the d test was incorrect when used in dynamic models and that if the size were appropriately recalibrated, then the asymptotic power of the resulting test (Durbin's h-statistic) would be the same as that of the corresponding likelihood-ratio based test.

Here, Q is claimed to have the correct size under H_0,† at least to the accuracy of the approximations involved. However, Q is based on estimates of *error* autocorrelations which are biased towards zero, and (as far as I can ascertain) it does not appear to have a chi-square distribution when H_0 is false. Nevertheless, for comparison with z_2 it seemed useful to consider a statistic $z_4 \tilde{A} \chi^2(J-1, T\mu^0)$ whose power for a given value of J and T depended only on

$$\mu^0 = \sum_{j=1}^{J} r_{jp}^2 .$$

At the same time, z_4 is also meant to represent a diagnostic test which incorporates the idea of seeking evidence of model inadequacy from "looking at the whole correlogram."

Based on the same approximations as used in calculating the rough power of z_2, Table C reports values of $P_T(z_4/\alpha)$ when $\delta=0$, $a=.05$, and $c^0 = \chi_a^2(J-1,0)$ for two values of both J and T, where

$$(20) \qquad P_T(z_4/\alpha) = \int_{h^0}^{\infty} d\chi^2(K^0,0)$$

with $h^0 = c^0(J-1+T\mu^0)/(J-1+2T\mu^0)$, and $K^0 = h^0(J-1+T\mu^0)/c^0$. From (13) and (14), $\mu^0 = \alpha^4(1+\phi^2)/(1+\alpha^2+\alpha^4)^2$ and so $\mu^0 \leq \mu$; consequently $P_T(z_4/\alpha) \leq P_T(z_2/\alpha)$. For any finite value of J, $P_T(z_4/0) = a$ from (20) and $P_T(z_4/\alpha) \to 1$ as $T\to\infty$ for $\alpha\neq 0$. Against the alternative of an $ARMA$ (1,1) model, z_4 must lose power as J is increased. Table C reflects all of these tendencies.

Table C
Approximate Power of the z_4 Test When $\delta=0$

J	T	α	1.	.8	.5	0.
2	40	P	0.7	0.5	0.3	.05
	80	P	0.9	0.9	0.5	.05
8	40	P	0.4	0.2	0.1	.05
	80	P	0.7	0.6	0.3	.05

†See Box and Pierce [11].

Thus, to the extent that the distribution Q depends on μ^0 for its power, it is even less likely to detect such a mis-specification. Finally, for values for r_{jp} recorded in Table A hardly inspire confidence in the ability of an "eyeball" test of simply inspecting the residual correlogram to reveal the incorrect approximation.

To immediately counter any doubts about the relevance of asymptotic analysis in finite samples, it is easy to establish from the results in Hendry and Harrison [60] that

$$\hat{\lambda} = \lambda_p + \{(T\sigma_Y^2)^{-1}\Sigma Y_{t-1} u_{pt}\} + 0(1/T) = \lambda_p + \tilde{\lambda} + (1/T)$$

where

$$u_{pt} = Y_t - \lambda_p Y_{t-1} = \epsilon_t + \alpha\epsilon_{t-1} - \alpha^* Y_{t-1}$$

and

$$\sigma_Y^2 = E(Y_t^2) = \sigma_\epsilon^2(1+2\alpha\delta+\alpha^2)/(1-\delta^2).$$

Thus, $\tilde{\lambda}$ is an asymptotic approximation to $\hat{\lambda}$ accurate to $0(1/T)$ with $E(\hat{\lambda})=0$ and $plim\ \sqrt{T}(\hat{\lambda}-\lambda_p-\tilde{\lambda}) = 0$ so that $(\hat{\lambda}-\lambda_p)$ and $\tilde{\lambda}$ are asymptotically equivalent.

Similarly,

$$\hat{r}_j = (T\sigma_{up}^2)^{-1}\ \Sigma u_{pt}\ u_{pt-j} + 0(1/T) = \tilde{r}_j + 0(1/T)$$

where $E(\tilde{r}_j) = r_{jp}$ and

$$\hat{\sigma}_u^2 = T^{-1}\Sigma u_{pt}^2 + 0(1/T) = \tilde{\sigma}_u^2 + 0(1/T)$$

with $E(\tilde{\sigma}_u^2) = \sigma_{up}^2$ so that the numerical *plim* results stated above actually hold in finite samples to $0(1/T)$ (if the required moments exist). Further, in terms of *efficient* Monte Carlo methods, $\tilde{\lambda}$, $\tilde{r}_j$ and $\tilde{\sigma}^2{}_u$ are appropriate control variables should one wish to attempt to disconfirm the above analysis by simulation. Indeed the asymptotic distributions of $\hat{\lambda}$, $\hat{r}_j$, etc., can be derived by using the equivalent $\tilde{\lambda}$, $\tilde{r}_j$.

On the other hand, the approximate test sizes and powers quoted earlier probably constitute a less reliable guide to actual finite sample behavior. However, some encouragement can be derived from the simple case analyzed in the Appendix.

Higher order autoregressive approximations to *ARMA* (1,1) processes yield similar results but are more tedious to derive analytically. It is of interest, however, if (4) is to be estimated allowing w_t to be an *ARMA* (1,0) process, to consider the *ARMA* (2,0) choice

$$(21) \qquad Y_t = \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + v_t$$

where λ_1 and λ_2 are to be estimated by least squares. Such an equation is not identifiably different from

$$(22) \quad Y_t = \delta Y_{t-1} + w_t \text{ where } w_t = \phi w_{t-1} + v_t$$

since (22) implies that

$$(23) \quad Y_t = (\delta+\phi) Y_{t-1} - \delta\phi Y_{t-2} + v_t$$

which is an unrestricted two parameter model like (21). This prompts the different interpretation of the approximation as involving an *ARMA* (1,0) in place of an *ARMA* (0,1) process for the *error term* of the *ARMA* (1,0) equation (8). On this convenient interpretation

$$v_t = \epsilon_t + \alpha^2\phi\epsilon_{t-1} - \alpha\phi\epsilon_{t-2} \text{ with } \sigma_v^2 = \sigma_\epsilon^2(1+\alpha^3\phi)$$

(that the choice of parameters in (23) is arbitrary does not matter since v_t in (25) is unique; however, if $\delta=0$, then $Y_t = \alpha Y_{t-1} - \alpha^2 Y_{t-2} + \epsilon_t - \alpha^3\epsilon_{t-3}$ is easier to work with). Thus

$$(24) \quad \begin{pmatrix}\lambda_{1p}\\ \lambda_{2p}\end{pmatrix} = \begin{pmatrix}\delta+\phi\\ -\delta\phi\end{pmatrix} + \frac{\alpha^*\phi}{(1-\lambda_p^2)}\begin{pmatrix}\alpha^*\\ \delta\lambda_p - 1\end{pmatrix} = \begin{pmatrix}\delta+\phi\\ -\delta\phi\end{pmatrix} + \alpha^*\begin{pmatrix}h_1\\ h_2\end{pmatrix}$$

[if $\delta=0$, then $(\lambda_{1p}; \lambda_{2p}) = (\phi/(1-\phi^2) ; -\phi^2/(1-\phi^2)) = (\lambda_p + r_{1p}; r_{2p})$]. The residuals, using the asymptotic approximation immediately, are given by

$$(25) \quad \begin{aligned} v_t &= Y_t - \lambda_{1p} Y_{t-1} - \lambda_{2p} Y_{t-2} \\ &= v_t - \alpha^*(h_1 Y_{t-1} + h_2 Y_{t-2}) \end{aligned}$$

so that

$$(26) \quad \sigma_{vp}^2 = \sigma_v^2 + \alpha\alpha^*\sigma_\epsilon^2[h_1^2 + h_2^2 + 2h_1h_2\lambda_p + 2\phi(h_2 + \delta h_1)]$$

(when $\delta=0$, this becomes $\sigma_{vp}^2 = \sigma_\epsilon^2(1+\alpha^4\phi^2/(1-\phi^2))$, and hence we have the analytic check that $\sigma_{up}^2 \geq \sigma_{vp}^2 \geq \sigma_\epsilon^2$). Since the parameterization is unwieldly, Table D records numerical values for λ_{1p}, λ_{2p}, and $\sigma_{vp}^2/\sigma_\epsilon^2$ for $\delta = (0,.8)$ and various values of α.

The non-nesting of the *ARMA* (2,0) and *ARMA* (1,1) models precludes any likelihood ratio test but it is clear from Table D that the likelihood values of the two models are similar over a wide range of values of α.

The formulae for the residual autocorrelations (denoted ρ_g in this *ARMA* (2,0) case) are unenlightening except that now $\rho_{gp} = \delta\rho_{g-1,p}$, $g \geq 4$ and if $\delta=0$, the residual correlogram peaks at

Table D
Plims of Parameter Estimates in an *ARMA* (2,0) Approximation to an *ARMA* (1,1) Process

	α	−1.	−.8	−.5	0	.5	.8	1.
	λ_{1p}	−.67	−.64	−.48	0	.48	.64	.76
	λ_{2p}	−.33	−.31	−.19	0	−.19	−.31	−.33
$\delta=0$	$\sigma^2_{vp}/\sigma^2_\epsilon$	1.33	1.13	1.01	1.00	1.01	1.13	1.33
	ρ_{1p}	−.08	−.07	−.02	0	.02	.07	.08
	ρ_{2p}	−.17	−.14	−.05	0	−.05	−.14	−.17
	ρ_{3p}	−.25	−.22	−.10	0	.10	.22	.25
	λ_{1p}	−.11	0	.28	.8	1.22	1.32	1.33
$\delta=.8$	λ_{2p}	−.09	0	.29	0	−.38	−.46	−.47
	$\sigma^2_{vp}/\sigma^2_\epsilon$	1.09	1.00	1.04	1.00	1.05	1.25	1.50

$\rho_{3,p} = \alpha\phi^2/[(1-\phi^2)(1+\alpha\phi(\alpha^2+\lambda_{2p}))]$. The numerical values of ρ_{gp} in Table D for the case $\delta=0$ again reveal the difficulty of detecting the inappropriate form by *residual diagnostics*.

The above analysis has a number of implications. Firstly, it puts a premium on choosing roughly the right approximation during the "preliminary identification" stage whatever diagnostic technique is to be used to check the specification. Thus, the overall statistical properties of the complete time series fitting procedure seem to merit further study for samples of the size relevant to economics (that is, when T is around 50 or less). Secondly, the 'Y_t' variable in (7) can be interpreted as *the residual from an estimated econometric relationship* of the form

$$(27) \qquad \underset{\sim}{Y}^* = \underset{\sim}{X}\beta + \underset{\sim}{\varepsilon}$$

so that $\underset{\sim}{Y} = (\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}')\underset{\sim}{Y}^*$.

Now the choice of *ARMA* (1,0) or *ARMA* (2,0) is one of approximating the residual autocorrelation by an autoregressive process rather than fitting the assumed more appropriate form *ARMA* (0,1) or *ARMA* (1,1). The above results, therefore, indicate that there is little to choose between these approximations in terms of goodness of fit or final residual autocorrelation. This provides analytic corroboration for the simulation findings of Hendry and Trivedi [61] that the correct *order* of error autocorrelation is more important than the *form* (*AR* or *MA*). In terms of the bias or mean squared error of $\hat{\beta}$ in (27), an estimated *ARMA* (1,0) approximation to an *ARMA* (0,1) for Y was as good as its approximation to a correct *ARMA* (1,0) in several dynamic models, (and vice-versa for *ARMA* (0,1)). In view of these findings, it seems difficult to chastise econometricians for choosing the computationally cheaper *autoregressive* (as opposed to *ARMA*) approximation for error autocorrelation, although, as noted earlier, I

concur with Granger and Newbold in criticizing the use of low order processes only. Moreover, the crude comparison of the powers of z_2 and z_4 suggests that there may be a distinct gain from using tests for reasonable alternative specifications based on likelihood-ratio considerations rather than just residual autocorrelations (unless the latter tests are corrected to coincide with the former). Finally, the situation $\alpha\delta<0$ poses the most serious detection problems (or the least ill consequences from the inappropriate specification) and such a parameter combination seems to occur reasonably often for data series ostensibly "reduced to stationarity" by differencing. This comment leads directly to the next important topic.

IV. Differencing Economic Time Series

Granger and Newbold recommend the use of first differences to achieve (more) stationary series and so avoid the problems discussed by Granger and Newbold [43]. There certainly do exist cases where differencing can be helpful, but it is not a universal panacea (no mechanical procedure ever is in econometrics). This occurs because there are two distinct interpretations of a difference transformation *to an equation.*

Operator Form. In this mode, $\Delta = (1-B)$ (where B is the lag operator) is considered to operate on an equation (taking (4) as an example), transforming

$$(28) \quad Y_t = \gamma_1 + \gamma_2 X_t + \gamma_3 X_{t-1} + \gamma_4 Y_{t-1} + w_t$$

to

$$(29) \quad \Delta Y = \gamma_2 \Delta X_t + \gamma_3 \Delta X_{t-1} + \gamma_4 \Delta Y_{t-1} + \Delta w_t.$$

An intercept in (29) would correspond to a trend term in (28), and the autocorrelation properties of the error term are completely altered since w_t is white noise if (and only if) w_t is a random walk.

Restriction Form. An equation in first differences can also be obtained from (28) by imposing the parameter restrictions that $\gamma_2 = -\gamma_3$ and $\gamma_4 = 1$ which yields

$$(30) \quad \Delta Y_t = \gamma_1 + \gamma_2 \Delta X_t + w_t.$$

In this case, if the restrictions are valid, the interpretation of both the intercept and the error term are *unaltered* and (30) implies the exclusion of ΔX_{t-1} and ΔY_{t-1} as compared with (29). Equally important is that if (30) *is* the true data generation process for some series Y_t, X_t such that ΔY_t, ΔX_t and w_t are stationary and w_t is white noise, then so must be the error w_t on the equation (28) in levels.† Further, the validity of the difference restriction seems to be testable since, on the null hypothesis, w_t is stationary and the variables X_{t-1} and Y_{t-1} added to (30) should have

†See Granger and Newbold [43].

zero coefficients; indeed the test is just one of their joint significance when included in (30). In this context (29) becomes an incorrect specification (falsely including ΔX_{t-1} and ΔY_{t-1} and excluding the intercept) and has a moving-average error with a coefficient of -1. However, new problems of "spurious significance" should not afflict Granger and Newbold's procedures unless these are carelessly applied, since always proceeding from the most parsimonious case should ensure detecting (30) before (29) is reached. It is also worth repeating that the original problem (of obtaining "nonsense" results if $\gamma_2 = 0$ but Y_t is regressed on X_t without including Y_{t-1}) still lurks in the background.

Nevertheless, in econometric terms the problem with differencing is not over, since both (29) and (30) have unacceptable features as *universally valid* types of formulation for economic systems. Let $y_t = ln\ Y_t$ and $x_t = ln\ X_t$ when Y and X are, say, consumption and income. Then (30) has *no* equilibrium solution (or zero if $\gamma_1 = 0$), and the time paths that y_t can describe are independent of the states of disequilibrium existing in the period prior to observation. Since there are more ways of obtaining stationarity than differencing, the choice of which transformation to adopt should be based on economic theory considerations. Marginal adjustments do suggest differencing, and Fisher [27] has argued in favor of difference formulations for short-run analysis because they avoid the need to specify the long-run behavior of the process under study.† On the other hand, long-run unit elasticities suggest the use of ratios. Reconsider (28) written as

$$(31) \quad \Delta y_t = \gamma_1 + \gamma_2 \Delta x_t + \gamma_5 (y_{t-1} - x_{t-1}) + w_t$$

$[y_{t-1} - x_{t-1} = ln(Y_{t-1}/X_{t-1})]$. This imposes the weaker (but still testable) restriction that $1 - \gamma_4 = \gamma_2 + \gamma_3 = -\gamma_5 > 0$ and produces an equation such that

- The mean lag in levels is long but finite, and most of the adjustment occurs in the first period.
- The long-run elasticity of Y with respect to X is unity, but the propensity varies with the growth rate of X.
- The time path of y_t is dependent on previous disequilibrium states.
- The equation has sensible steady state and equilibrium properties.

Such a formulation is excluded if *only* differenced variables are considered. A model of the form (31) helps explain the very low long run consumption-income elasticities reported by Wall *et al.* [156] who estimate equations like (30) (excluding all levels variables) using bivariate modeling techniques based on a Box-Jenkins transfer function approach. Moreover, the Q statistic failed to indicate the existence of this mis-specification.‡

†Compare Hendry and Anderson [59].

‡ For a more detailed discussion and an application to the consumption-income relationship, see Davidson and Hendry [18].

V. Simultaneous Equations Systems

The above analysis has deliberately ignored the problem of contemporaneous feedbacks affecting the relationship under study in order to isolate some specific difficulties which I felt merited attention. Simultaneity is well known to introduce a host of additional problems including nonunique identification of structural relations (in the econometricians' sense), bias and inconsistency in structural parameters estimated by least squares procedures (which can create spuriously significant relationships as well as camouflaging the very existence of the inconsistencies), bias in residual autocorrelations towards white noise and so on.† Thus, Teräsvirta [146] has demonstrated that the relevant variant of the Q statistic (see equation (19) above) does not have a chi-square distribution in systems with feedback. It seems as difficult to justify the use of Box-Jenkins transfer functions for *structural* estimation‡ as it is to justify two-stage least squares estimation of dynamic models with autocorrelated errors;§ both inappropriate applications are likely to yield seriously biased estimates.

The eight-step method proposed by Granger and Newbold is, of course, designed to avoid the immediately preceding objections.‖ On the one hand, the discussion in previous sections suggest that there are problems involved in correctly identifying appropriate univariate *ARMA* processes, that diagnostic tests may lack power to detect certain inadequacies, and that differencing is not fully valid in an interesting class of dynamic models. The impact of such difficulties for bivariate modeling exercises seems to deserve investigation. However, if the results are only intended for forecasting purposes, these problems in no way proscribe the use of Granger and Newbold's approach as a *supplement* to other methods. (Although its labor intensity might!) Specifically, such forecasts could provide a formidable opponent for econometric systems to try and outperform, and the pooling of forecasts from these rather different sources seems to merit greater empirical application.#

VI. Conclusion

A thorough understanding of time series analysis is an essential component of the intellectual tool-kit of econometricians studying time series data — but it remains only one of many ingredients. Granger and Newbold have helpfully pointed out problems of potential relevance to that non-negligible proportion of applied work which neglects time series considerations. One symptom of that neglect may be the less than totally impressive

†See, for example, Hendry [57].

‡Compare Wall *et al.* [156].

§See Hendry and Harrison [60].

‖Also compare Wall and Westcott [157].

#See, for example, Nelson [102].

forecasting record of macro-econometric systems.†

My major doubts concern how to remedy such a situation. This paper has investigated the possibility that time series methods may be plagued by difficulties similar to those which trouble econometric model builders, although the consequences of mis-specifying *ARIMA* models are obviously rather different from the consequences of mis-specifying econometric systems. Some practical evidence has accumulated, however, since there have been several recent empirical studies comparing time series models with econometric equivalents.‡ Such studies also serve to highlight how similar the "two philosophies" are by interpreting *ARIMA* processes as representing the (unrestricted) final forms of the same systems for which the econometrician seeks estimates of the structural equations. For short data series, valid structural information from economic theory must be of considerable importance for estimation efficiency and both Trivedi [149] and Prothero and Wallis [124] find that the econometric model fits rather better over the sample period. Unfortunately, such evidence is not incompatible with the econometric system also forecasting less well.

Nevertheless, I suspect that generalizations of econometric techniques specifically designed for multivariate time series will prove more useful in the medium term than any purely data based approach. Such developments are feasible and are applicable to simultaneous equations systems of more than two variables.§ Moreover, the modeling can be closely linked to (and test) the relevant economic theory, hopefully producing the joint outputs of understanding and reasonable forecasts.

Sargent-Sims

The Sargent and Sims generalization of factor analysis to the frequency domain is a most useful addition to the range of techniques currently available for investigating time series data. Existing time domain methods comprise a variety of specializations of vector autoregressive-moving average (*ARMA*) representations and of econometric systems, with occasional use (in economics) of factor analysis. In the frequency domain there are spectral analysis, spectral estimation of econometric models, and now the Sargent-Sims "specfac" analysis.

All of these methods can be conveniently interpreted as seeking to provide parsimonious (but different) parameterizations of the matrix of cross spectra of the endogenous variables (denoted by $Sy(\omega)$ for $\omega \epsilon (0,\pi)$). Thus, a (vector) *ARMA* process assumes that $Sy(\omega)$ can be closely

†See Naylor *et al.* [101].

‡See *inter alia* Zellner and Palm [167], Trivedi [149], and Prothero and Wallis [124].

§See Hendry and Anderson [59].

approximated by a (matrix) rational function; spectral analysis assumes constancy of the spectrum within certain neighboring frequency band groups and variation between groups (that is, a step-function form); and an econometric model assumes some specialization of an *ARMA* process for the error (often a constant over all frequency bands) and parameterizes $Sy(\omega)$ from this and the observed spectral matrix of the "exogenous" variables. In their unobservable k-index model, Sargent and Sims postulate that $2\pi Sy(\omega) = H_1(\omega) + H_2(\omega)$ where $H_1(\omega)$ is symmetric, positive semi-definite of rank k, and $H_2(\omega)$ is diagonal. Assuming that y is normally distributed, maximum likelihood estimates of H_1 and H_2 can be obtained together with information about the relative likelihoods of various values of k.

In practice, each data series is first separately prewhitened by fitting a fifth order autoregressive process with constant and trend, and r frequency bands are used so that when a value of $k > 1$ is chosen it is not obvious how parsimonious the resulting description is. Nevertheless, it is clear that new insights into the coherence structure of a data set can be achieved by using "specfac," and the development of an operational algorithm must be warmly welcomed.

However, I must question the *motivation* underlying the Sargent-Sims approach. Certainly, "important statistical regularities are missed by large scale models." But this does not imply that one must seek substitutes for conventional econometric methods, as opposed to producing complements (which is how I prefer to interpret the role of the papers presented by Granger-Newbold and Sargent-Sims), together with some (perhaps substantial) revision and re-emphasis within econometrics.

On the one hand, the claimed proliferation of non-nested theories suggests that researchers are not building successively on previous results by rejecting these in the course of establishing more general models (which yet remain as parsimonious as the data will tolerate). This demands a *more* rigorous application of methodology, not its abandonment. Too often "parsimony" in an econometric model relates to unwarranted and unnecessary assumptions about properties of the error process. When little is known concerning an *auxiliary* hypothesis, every effort should be made to avoid assumptions whose failure will produce serious inconsistencies. Conversely, one can often fruitfully exploit assumptions whose invalidity only produces inefficiency, unless this means a very large increase in variance. For example, when estimating the structural form of a simultaneous system one can *arbitrarily* restrict the error covariance matrix and/or the reduced form parameters without affecting consistency.† Considerable simulation evidence comparing "full" and "limited" information estimators suggests that even the efficiency loss of doing this is small. Compare such a situation to that of estimating a dynamic equation with an *ignored* autocorrelated error; doing so can produce

†See the "estimator generating formula" in Hendry [58].

large inconsistencies as well as serious inefficiencies in estimation and forecasting. A further example is that an omitted first order autoregressive error can be approximated by including one lag longer in every variable in the equation with consistent but inefficient results, whereas the converse is not true, in general.

A closely related point is that the particular parameterization adopted for a behavioral relationship should often be selected on the criterion of generating near orthogonal regressors to reflect agents' use of independent sources of information relevant to their decisions.† This has the added advantages of circumventing collinearity problems and of minimizing any potential biases arising from (unsuspected) omitted variables. There remain many other constructive ways of directly improving econometric practice. Thus, it does not seem useful to concentrate on obtaining possibly spurious "improved estimation efficiency" by using, say, instrumental variables (IVs) based on solved reduced forms if the cost is neglecting autocorrelation and dynamic specification problems.

These comments assume that econometrics should strive to remain quantitative economics and not become just statistical modeling. The latter certainly provides an alternative answer to the proliferation of economic theories — by eschewing them totally. But this has the consequence that there is no paradigm to guide new research or integrate and summarize achieved findings, and hence little progress results. Sargent and Sims note that the possible conclusions from "specfac" need not corroborate all conceivable economic theories; however, the deliberate avoidance of much *a priori* input also means that it is unlikely to lead to the exclusion of many theories either.

Overall, I agree with the thrust of the criticisms which both Granger-Newbold and Sargent-Sims make of existing methods. Nevertheless, I am not persuaded that *new* methods of business cycle research are the optimal response for correcting inadequacies in econometric systems as against a (major?) redirection of attention towards solving problems the existence and seriousness of which are well known.

†See Davidson and Hendry [18].

Appendix to Granger-Newbold Comments
The Power of the Likelihood Ratio Test for Autocorrelation in a Linear Model†

The usual likelihood ratio based test of $H_0:\rho=0$ in (2) is given by

$$\text{(A1)} \quad z_5 = T\, ln(\Sigma\, \hat{e}_t^2/\Sigma\, \hat{v}_t^2) \;\tilde{A}\; \chi^2(1, T\mu^*)$$

where $\mu^* = -ln(1-\rho^2)$. The approximate power of z_5 can be calculated using the approach discussed for z_2 in section 3 of the text on substituting μ^* for μ. If the regression model is (1) where X_t is a fixed regressor, then $P_T(z_5|\rho)$ is independent of the sign of ρ and of the autocorrelation properties of X_t. Moreover, in such a model the maximum likelihood estimator $\hat{\rho}$ of ρ is distributed independently of $(\hat{\beta}_1, \hat{\beta}_2)$ and has the same asymptotic distribution as the residual autocorrelation coefficient $\tilde{\rho}$

$$\text{(A2)} \quad \tilde{\rho} = \Sigma\, \tilde{e}_t \tilde{e}_{t-1} / \Sigma\, \tilde{e}_{t-1}^2$$

where $\tilde{e}_t$ are the *OLS* residuals in (1). Since z_5 is asymptotically equivalent to conducting a 't' test on $\hat{\rho}$ for $\rho=0$ and the Durbin-Watson bounds test $d \approx 2(1-\hat{\rho})$, we expect z_5 and d also to be asymptotically equivalent. The exact power of d was recently obtained by L'Esperance and Taylor [81] for a model like (1) + (2) (where v_t ~ $NI(0,1)$, $X_t = \lambda X_{t-1} + u_t$ and u_t ~ $NI(0,1)$) and shown to depend on the sign of ρ and (as would be anticipated from the nature of the bounds test) on λ. To reflect these differences from z_5 we denote the two power functions by

$$p_1 = P_T(z_5/|\rho|) \text{ and } p_2 = P_T(d/\rho,\lambda).$$

L'Esperance and Taylor investigated a parameter space defined by $(\beta_1,\beta_2) = (1.0,0.8)$, $\lambda = (.1,.7)$, $T = (15,30,50)$, $a = (.05,.01)$, and $\rho = -.9 + .2n$ $(n=0,1,..,9)$. For one non-constant regressor, $a=.05$ and $T \geq 30$, the bounds of the d-test are reasonably close, and hence, the relationship between p_1 and p_2 in this situation provides a check on the accuracy of the approximations involved in calculating the former. Since a linear equation provides the same functional form for powers and type II errors

$$\text{(A3)}\; p_{2i} = \theta_1 p_{1i} + \theta_2 \lambda_i + \theta_3/T_i + \theta_4 D_{1i} + \theta_5 D_{2i} + \theta_6 + f_i \;(i=1,..,40)$$

seemed a sensible choice where

†I am grateful to Gerry Dugay for assistance with the calculations.

$$D_{1i} = \begin{cases} 1 \text{ if } \rho > 0 \\ 0 \text{ otherwise} \end{cases} \qquad \text{and } D_{2i} = \begin{cases} 1 \text{ if } |\rho| < .5 \\ 0 \text{ otherwise.} \end{cases}$$

Thus a "good" approximation would be defined by

$$\theta_i = 1,\ \theta_j = 0 \ (j = 2, \ldots, 6)$$

and a high R^2 or small σ_f. Estimation yielded

$$\text{(A4)} \quad \underset{}{p_{2i}} = \underset{(.01)}{1.01 p_{1i}} - \underset{(.01)}{.03 D_{1i}} + \underset{(.01)}{.15 D_{2i}} \qquad R^2 = .983,\ \hat{\sigma}_f = .042$$

(coefficient standard errors in parentheses). The regression of p_{2i} on p_{1i} alone had an $R^2 = .909$; conversely adding λ_i, T_i and a constant to (A4) produced an R^2 of .984 and $\hat{\theta}_2$, $\hat{\theta}_3$ and $\hat{\theta}_6$ were jointly and individually insignificant. The hypothesis that $\theta_i = 1$ cannot be rejected at a 40 percent significance level; there is a small dependence of p_2 on the sign of ρ but no significant variation with λ; $p_2 > p_1$ at small values of ρ; and the approximations involved are no better at the larger sample size ($T = 50$). The following table illustrates the values of p_1, p_2 (from L'Esperance and Taylor) and $\hat{p}_2$ (predicted from (A4)) for $T = 50$, $\lambda = .7$, $a = .05$, and $\rho > 0$.

Table A—Appendix

$\backslash\rho$	.1	.3	.5	.7	.9
p_1	.10	.56	.98	1.0	1.0
p_2	.19	.65	.95	.99	.99
$\hat{p}_2$	.22	.68	.96	.98	.98

An equivalent accuracy for Tables B and C in the text would seem quite acceptable, subject to the qualification that low powers may be underestimated. Conversely, if the p_1 calculations are reasonably accurate, these results do not support the conjecture of Durbin and Watson [21] that the likelihood ratio test may yield a higher power than d.