Columbia University

Department of Economics
Discussion Paper Series

Testing for Autocorrelation in Systems of Equations

Phoebus J. Dhrymes

Discussion Paper No.: 0506-08

Department of Economics
Columbia University
New York, NY 10027

August 2005

Testing for Autocorrelation in Systems of Equations*

PHOEBUS J. DHRYMES
Columbia University

Originated June 2003; this version August 2005

Abstract

This paper deals with the problem of testing for the presence of autocorrelation in a system of general linear models (Seemingly Unrelated Regressions, SUR) when the model is formulated as a vector autoregression (VAR) with exogenous variables. The solution presented in this paper is a generalization of the h-statistic for the single equation single parameter case given in Durbin (1970a). All derivations are based on first principles and no use is made of Durbin's original arguments.

KEY WORDS: Autocorrelation, h-test, δ-test, δ^*-test, VAR

The earlier version was presented at the Magdalinos Memorial Conference, November 20-21, 2003 in Athens, Greece. The current version will appear in the proceedings of that conference to be published by Cambridge University Press under the title The Refinement of Econometric Estimation and Test Procedures:Finite Sample and Asymptotic Analysis.

1 Introduction

Consider the model

$$y_{t\cdot} = y_{t-1\cdot}A + x_{t\cdot}B + u_{t\cdot}, \quad t = 1, 2, 3, \ldots, T, \tag{1}$$

where $y_{t\cdot}$ is an m-element row vector of dependent, and $x_{t\cdot}$ is a k-element vector of independent variables, respectively; $u_{t\cdot}$, $t = 1, 2, \ldots, T$ is the structural error vector. We assume

i. $\{u_{t\cdot} : t = 1, 2, 3, \ldots, T\}$ is a sequence of independent identically distributed (i.i.d.) random vectors with

$$E\, u_{t\cdot} = 0, \qquad \mathrm{Cov}(u_{t\cdot}) = \Sigma > 0, \tag{2}$$

defined on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$.

ii. It is further assumed that

$$\operatorname*{plim}_{T\to\infty} \frac{X'X}{T} = M_{xx} > 0, \tag{3}$$

and that the elements in X and U are mutually independent.

iii. The system of Eq. (1) is **stable**, i.e. the characteristic roots of A are less than one in absolute value.

Regarding the errors, the alternative hypothesis we entertain is

$$u_{t\cdot} = u_{t-1\cdot}R + \epsilon_{t\cdot}. \tag{4}$$

We require, for stationarity, the following assumptions:

1. The matrix R is non-singular and stable, i.e. its characteristic roots are less than one in absolute value;

2. With little loss of generality, and certainly no loss of relevance, we further assume that the matrix R is diagonalizable, i.e. it has the representation $R = P\Lambda P^{-1}$, where Λ is the (diagonal) matrix of its characteristic roots.

This problem, for the case $m = 1$, (and R a scalar) was dealt with by Durbin (1970a), (1970b). A search of widely used econometrics textbooks such as Greene (1999) and Davidson and MacKinnon (1993) discloses no mention of its generalization to VARs.

Remark 1. If one were to write down a VAR one would normally not be concerned about the behavior of the "error", since **by definition** the errors in such a system are assumed to be i.i.d. If not, in empirical applications, one simply specifies a VAR of a higher order. Notwithstanding this observation, in many applied contexts the logic of the economic model requires the presence of a specific number of lagged endogenous variables, in addition to the exogenous variables required by the specification. In such a case, the problem we are examining here may arise.

Remark 2. When the structural error, $u_{t\cdot}$, is in fact a first order autoregression, the OLS estimators for the parameters of the model in Eq. (1) are inconsistent **because of the presence of lagged endogenous variables**, which are therefore correlated with the structural error.

Thus, if we suspect that the form given in Eq. (4) may be appropriate, we may wish to test the hypothesis

$H_0 : \quad R = 0$,
as against the alternative
$H_1 : \quad R \neq 0$,

when least squares (OLS) is used to estimate the unknown parameters of Eq. (1).

2 Derivation of the Test Statistic

Writing the sample as

$$Y = Y_{-1}A + XB + U = ZC + U, \quad Z = (Y_{-1}, X), \quad C = (A', B')', \tag{5}$$

the OLS estimator of C is given by

$$\tilde{C} = (Z'Z)^{-1}Z'Y = C + (Z'Z)^{-1}Z'U. \tag{6}$$

It may be shown that, under the assumptions above, a central limit theorem (CLT) is applicable, see Dhrymes (1994), pp. 73-80, Dhrymes (1989), pp. 271 ff. and pp. 328-337. Thus, the limiting distribution of the OLS estimator may be obtained from[1]

$$\sqrt{T}(\tilde{c} - c) \sim \frac{1}{\sqrt{T}} \sum_{t=2}^{T} (I \otimes S z_{t\cdot}') u_{t\cdot}', \quad c = \mathrm{vec}(C) \quad \text{where} \tag{7}$$

$$S = \underset{T\to\infty}{\mathrm{plim}} (Z'Z/T)^{-1} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}. \tag{8}$$

As is easily seen from Eq. (7), the summands therein form a sequence of martingale difference (MD) vectors that satisfy the conditions of Proposition 21 in Dhrymes (1989) p. 337. Consequently,

$$\sqrt{T}(\tilde{c} - c) \xrightarrow{\mathrm{d}} N(0, \Sigma \otimes S). \tag{9}$$

Since

$$\sqrt{T}\mathrm{vec}(\tilde{A} - A) \sim \frac{1}{\sqrt{T}} \mathrm{vec}(S_1 Z'U) = \frac{1}{\sqrt{T}} \sum_{t=2}^{T} [I_m \otimes S_1 z_{t\cdot}'] u_{t\cdot}'. \tag{10}$$

using the same arguments as above we conclude that

$$\sqrt{T}\mathrm{vec}(\tilde{A} - A) = \sqrt{T}(\tilde{a} - a) \xrightarrow{\mathrm{d}} N(0, \Sigma \otimes S_{11}). \tag{11}$$

Let

$$\tilde{U} = Y - Z\tilde{C} = U - Y_{-1}(\tilde{A} - A) - X(\tilde{B} - B), \tag{12}$$

[1] The notation $X \sim W$ below means, in this context, that X has the same limiting distribution as W.

be the matrix of OLS residuals and consider the estimator of R

$$\tilde{R} = (\tilde{U}_{-1}'\tilde{U}_{-1})^{-1}\tilde{U}_{-1}'\tilde{U}, \quad \tilde{U} = U - Z(\tilde{C} - C). \tag{13}$$

Using Eqs. (12), (13), and omitting terms that converge to zero in probability, we may write, see Dhrymes (1989), pp. 161 ff.,

$$\sqrt{T}(\tilde{R} - R) \sim \Sigma^{-1}\frac{1}{\sqrt{T}}\left(U_{-1}'U - \frac{U_{-1}'Y_{-1}}{T}\sqrt{T}(\tilde{A} - A)\right), \tag{14}$$

either because $Z'Z/T$ converges, or because $\tilde{C}$ is consistent and has a well defined limiting distribution, or both. Moreover, using the result again, and bearing in mind that

$$\frac{1}{T}U_{-1}'Y_{-1} \xrightarrow{\mathrm{P}} \Sigma$$

we finally obtain

$$\sqrt{T}(\tilde{R} - R) \sim \Sigma^{-1}\frac{1}{\sqrt{T}}U_{-1}'U - \sqrt{T}(\tilde{A} - A). \tag{15}$$

Using Eq. (10), and giving more details, we note that

$$\sqrt{T}(\tilde{A} - A) \sim S_1\frac{1}{\sqrt{T}}Z'U, \quad S_1 = (S_{11},\ S_{12}).$$

Vectorizing, we have the expression

$$\sqrt{T}\mathrm{vec}(\tilde{R} - R) = \sqrt{T}(\tilde{r} - r) \sim \frac{1}{\sqrt{T}}\mathrm{vec}[(\Sigma^{-1}U_{-1}' - S_1Z')U] \tag{16}$$

$$= \frac{1}{\sqrt{T}}\mathrm{vec}[\sum_{t=2}^{T}(\Sigma^{-1}u_{t-1\cdot}' - S_1z_{t\cdot}')u_{t\cdot}]$$

$$= \frac{1}{\sqrt{T}}\sum_{t=2}^{T}[I_m \otimes (\Sigma^{-1}u_{t-1\cdot}' - S_1z_{t\cdot}')]u_{t\cdot}'. \tag{17}$$

The summands in the rightmost member above are recognized as a MD sequence that obeys the Lindeberg condition, as noted above. Let

$$\mathcal{A}_t = \sigma(u_{s\cdot}, s \leq t), \tag{18}$$

be the σ-algebra generated by the u's up to t, and note that

$$\frac{1}{T}E\left([I_m \otimes (\Sigma^{-1}u'_{t-1\cdot} - S_1 z'_{t\cdot})]u'_{t\cdot}u_{t\cdot}[I_m \otimes (u_{t-1\cdot}\Sigma^{-1} - z_{t\cdot}S'_1)]|\mathcal{A}_{t-1}\right) \quad (19)$$

$$= \frac{1}{T}[\Sigma \otimes (\Sigma^{-1}u'_{t-1\cdot} - S_1 z'_{t\cdot}][I_m \otimes (u_{t-1\cdot}\Sigma^{-1} - z_{t\cdot}S'_1)] \quad (20)$$

which, upon summation, obeys

$$\frac{1}{T}\sum_{t=2}^{T}[\Sigma \otimes (\Sigma^{-1}u'_{t-1\cdot} - S_1 z'_{t\cdot})][I_m \otimes (u_{t-1\cdot}\Sigma^{-1} - z_{t\cdot}S'_1)] \xrightarrow{\text{P}} \Sigma \otimes (\Sigma^{-1} - S_{11}). \quad (21)$$

Thus, one of the sufficient conditions of Proposition 21, Dhrymes (1989) p. 327, is satisfied and, consequently, under the null $R = 0$,

$$\sqrt{T}\tilde{r} \xrightarrow{\text{d}} N(0, \Sigma \otimes (\Sigma^{-1} - S_{11})). \quad (22)$$

It is interesting to note that, **if** the elements of U were **known**, we could obtain the estimator

$$\hat{R} = (U'_{-1}U_{-1})^{-1}U'_{-1}U, \quad \text{which, under the null, obeys}$$

$$\sqrt{T}\hat{r} = \frac{1}{\sqrt{T}}\sum_{t=2}^{T}(I_m \otimes (U'_{-1}U_{-1})^{-1}u^{prime}_{t-1\cdot})u'_{t\cdot} \xrightarrow{\text{d}} N(0, \Sigma \otimes \Sigma^{-1}). \quad (23)$$

We have therefore proved

Theorem 1. Given the model in Eq. (1) and the assumptions stated in i through iii, the following is true:

i. If the elements of the matrix U are known, we can estimate the unknown matrix R by means of Eq. (23), which yields a consistent estimator whose limiting distribution is $N(0, \Sigma \otimes \Sigma^{-1})$.

ii. If the elements of U are not known but are estimated by means of the matrix of the OLS residuals from the regression of Y on Z, the limiting distribution of the estimator exhibited in Eq. (6) is given by

$$\sqrt{T}(\tilde{c} - c) \xrightarrow{\text{d}} N(0, \Sigma \otimes S),$$

and thus the limiting distribution of the OLS estimator of the (matrix of) coefficients of the lagged dependent variable is given by

$$\sqrt{T}(\tilde{a} - a) \xrightarrow{\mathrm{d}} N(0, \Sigma \otimes S_{11}).$$

iii. The limiting distribution of the least squares estimator of the first order autocorrelation (matrix) exhibited in Eq. (13) is, under the **null** hypothesis $H_0 : R = 0$, given by

$$\sqrt{T}\mathrm{vec}(\tilde{R}) = \sqrt{T}\tilde{r} \xrightarrow{\mathrm{d}} N(0, \Sigma \otimes (\Sigma^{-1} - S_{11}))$$

iv. The estimator in iii above is efficient relative to the estimator in i.

Corollary 1. A test statistic, under the null, for testing the hypothesis of (first order) autocorrelation in the residuals is given by

$$\delta = T\tilde{r}'[\tilde{\Sigma}^{-1} \otimes (\tilde{\Sigma}^{-1} - \tilde{S}_{11})^{-1}]\tilde{r}) \xrightarrow{\mathrm{d}} \chi^2_{m^2}. \tag{24}$$

Proof. Evident from Theorem 1.

Remark 3. Notice that the conclusion in parts i and iii above reveal a somewhat counter-intuitive result viz. that knowing **more results in inefficiency** relative to the case where we know less; more specifically, the estimator of R assuming that U is **known is inefficient** relative to a similar estimator obtained when U **is not known, but is estimated** by means of the matrix of OLS residuals from the regression of Y on Z. As it will be made clear below, this is also true when $m = 1$ and $R = \rho$, a scalar.

Remark 4. If, in a given application, the estimated matrix $\tilde{\Sigma}^{-1} - \tilde{S}_{11}$ is **not at least positive semi-definite**, the test fails. If it is positive semi-definite but not positive definite we may use in Eq. (24) the **generalized inverse**, instead of the inverse. If the matrix itself (not only the estimated one) is positive semi-definite **but not** positive definite, the distribution is still asymptotically χ^2, but with degrees of freedom equal to the rank of $\Sigma^{-1} - S_{11}$. Notice that these matrices can easily be estimated consistently by

$$\tilde{\Sigma} = \frac{1}{T}\tilde{U}'\tilde{U}, \quad \tilde{S}_{11} = (I,0)\left(\frac{Z'Z)}{T}\right)^{-1}(I,0)'.$$

Remark 5. If the model of Eq. (1) contains more than one lag of the dependent variable, the test procedure **remains the same**, as it is clear from the derivation given above, because under the null $u_{t-1\cdot}$ is independent of $u_{t-j\cdot}$, for $j \geq 2$.

Remark 6. When the alternative is of the form

$$u_{t\cdot} = u_{t-2\cdot}R_2 + \epsilon_{t\cdot},$$

the analog of the δ-test will be applicable, but it would involve the limiting distribution of $A\sqrt{T}(\tilde{A}-A)$ instead of that of $\sqrt{T}(\tilde{A}-A)$. If the specification were $u_{t\cdot} = u_{t-3\cdot}R_3 + \epsilon_{t\cdot}$, it would involve the limiting distribution of $A^2\sqrt{T}(\tilde{A}-A)$, and so on. This is so because $(U'_{-p}Y_{-1}/T)$ converges, at least in probability, to ΣA^p.

Remark 7. Note that in the case $m = 1$, and consequently $\tilde{R} = \tilde{\rho}$, the test statistic of Eq. (23) reduces to

$$T\mathrm{vec}(\tilde{R})'[\tilde{\Sigma}^{-1} \otimes (\tilde{\Sigma}^{-1} - \tilde{S}_{11})^{-1}]\mathrm{vec}(\tilde{R}) = \frac{T\tilde{\rho}^2}{1 - \mathrm{AVar}(\tilde{a}_{11})}, \tag{25}$$

where $\mathrm{Avar}(\tilde{a}_{11})$ is the variance of the limiting distribution of the OLS estimated coefficient of the lagged dependent variable. Thus, the δ statistic reduces to the square of the h-statistic, as given by Durbin (1970a), because basically $\Sigma \otimes \Sigma^{-1}$ reduces to unity in the case $m = 1$. Thus, the case where $\Sigma^{-1} - S_{11}$ **is not at least positive semi-definite** corresponds to the case where the asymptotic variance in question is equal to or greater than 1. When this is so one should employ an alternative procedure to be derived below.

Notice, further, that in the scalar case the estimator of ρ when U is **known** converges to $N(0,1)$, while the corresponding estimator when U is **not known** but is estimated by means of the OLS residuals from the regression Y on Z converges to a $N(0, 1 - \mathrm{AVar}(\tilde{a}_{11}))$!

3 An Alternative Test when the δ Test Fails

The δ statistic occasionally yields inadmissible results. It is thus desirable to obtain another test that "always works". To this end write the model in Eq. (1) as

$$Y = ZC + U_{-1}R + E = WD + E,$$

$$E \quad = (\epsilon_{t\cdot}), \quad W = (Y_{-1}, X, U_{-1}) = (Z, U_{-1}), \quad D = (C', R')', \tag{26}$$

where we have merely made use of the alternative specification in Eq. (4). If we could observe U we would simply estimate R by OLS and then carry out a test on R as we would with any other OLS-estimated parameter. Since we cannot, we take a page out of two stage least squares procedures and estimate it by using the OLS residuals from the regression of Y on Z. The estimator thus obtained is

$$\tilde{D} = (\tilde{W}'\tilde{W})^{-1}\tilde{W}'Y = (\tilde{W}'\tilde{W})^{-1}\tilde{W}'WD + (\tilde{W}'\tilde{W})^{-1}\tilde{W}'E, \;\; \tilde{W} = (Z, \tilde{U}_{-1}). \tag{27}$$

Under the null $R = 0$, we obtain

$$(\tilde{W}'\tilde{W})^{-1}\tilde{W}'WD = \begin{bmatrix} Z'Z & Z'\tilde{U}_{-1} \\ \tilde{U}'_{-1}Z & \tilde{U}'_{-1}\tilde{U}_{-1} \end{bmatrix}^{-1} \begin{bmatrix} Z'Z & Z'U_{-1} \\ \tilde{U}'_{-1}Z & \tilde{U}'_{-1}U_{-1} \end{bmatrix} D = \begin{bmatrix} I \\ 0 \end{bmatrix} C, \tag{28}$$

so that

$$\sqrt{T}(\tilde{D} - D) \sim \frac{1}{\sqrt{T}} S^* \tilde{W}'E \sim \frac{1}{\sqrt{T}} S^* W'E, \quad S^* = \operatorname*{plim}_{T\to\infty} (\tilde{W}'\tilde{W}/T)^{-1}. \tag{29}$$

Vectorizing, we have under the null

$$\sqrt{T}\mathrm{vec}(\tilde{D} - D) \sim \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (I_m \otimes S^* w'_{t\cdot})\epsilon'_{t\cdot}. \tag{30}$$

Using the same arguments as in the derivation of δ we conclude that, under the null,

$$\sqrt{T}\mathrm{vec}(\tilde{D} - D) \xrightarrow{\mathrm{d}} N(0, \Sigma \otimes S^*), \quad S^* = \begin{bmatrix} S^*_{11} & S^*_{12} \\ S^*_{21} & S^*_{22} \end{bmatrix}, \quad \text{where} \tag{31}$$

$$S^{*-1} = = \begin{bmatrix} S^{-1} & \Sigma_* \\ \Sigma_* & \Sigma \end{bmatrix}, \quad \Sigma_* = (\Sigma, 0), \quad \text{so that} \quad S^*_{22} = (\Sigma - \Sigma S_{11}\Sigma)^{-1}$$

The preceding also implies that, under the null,

$$\tilde{D} \xrightarrow{\mathrm{P}} D. \tag{32}$$

so that D is estimated consistently and has a well defined limiting distribution.

Using the same arguments as we did in the discussion immediately following Eq. (9), we conclude that the estimator of R obeys

$$\sqrt{T}\mathrm{vec}(\tilde{R}) = \sqrt{T}\tilde{r} \xrightarrow{\mathrm{d}} N(0, \Sigma \otimes S_{22}^{*}). \tag{33}$$

We have therefore proved

Theorem 2. Under the conditions of Theorem 1, write the model of (Eq. (1), as

$$\begin{aligned} Y &= ZC + U_{-1}R + E = WD + E, \\ E &= (\epsilon_{t\cdot}), \quad W = (Y_{-1}, X, U_{-1}) = (Z, U_{-1}), \quad D = (C', R')'. \end{aligned}$$

The following statements are true

i. Regressing Y on $\tilde{W} = (Y_{-1}, X, \tilde{U}_{-1}) = (Z, \tilde{U}_{-1})$, where $\tilde{U}_{-1}$ is the lagged matrix of the OLS residuals from the regression of Y on Z, yields the estimator of Eq. (26) which, under the null, obeys

$$\sqrt{T}\mathrm{vec}(\tilde{D} - D) \sim \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (I_m \otimes S^{*} w_{t\cdot}') \epsilon_{t\cdot}'. \tag{34}$$

ii. This estimator of D is consistent and its limiting distribution is given by

$$\sqrt{T}\mathrm{vec}(\tilde{D} - D) \xrightarrow{\mathrm{d}} N(0, \Sigma \otimes S^{*}), \tag{35}$$

where S^{*} is as defined in Eq. (29).

iii. The limiting distribution of the autocorrelation matrix estimator $\tilde{R}$ is given by

$$\sqrt{T}\mathrm{vec}(\tilde{R}) = \sqrt{T}\tilde{r} \xrightarrow{\mathrm{d}} N(0, \Sigma \otimes S_{22}^{*}), \quad S_{22}^{*} = (\Sigma - \Sigma S_{11} \Sigma)^{-1}. \tag{36}$$

iv. If U were known and we employed the same procedure embedded in Theorem 2, we should get exactly the same results as in part ii, although we would have the option of estimating R directly from U, as we did in Theorem 1.

Corollary 2. A test of the null hypothesis

$H_0: \quad R = 0$
as against the alternative
$H_1: \quad R \neq 0$

may be carried out by means of the statistic

$$\delta^* = T\tilde{r}'[\tilde{\Sigma}^{-1} \otimes (\tilde{\Sigma} - \tilde{\Sigma}\tilde{S}_{11}\tilde{\Sigma})^{-1}]\tilde{r} \xrightarrow{\mathrm{d}} \chi^2_{m^2}, \tag{37}$$

where

$$\tilde{\Sigma} = \frac{1}{T}\tilde{U}'\tilde{U}, \text{ and } \tilde{S}^*_{22} \text{ is the appropriate submatrix of } (\tilde{W}'\tilde{W}/T)^{-1}.$$

Remark 8. Since under the null the process $u_{t\cdot}$ is strictly stationary, we could as well estimate

$$\tilde{\Sigma} = \frac{1}{T-1}\tilde{U}'_{-1}\tilde{U}_{-1}.$$

The loss of one observation is inconsequential if the sample is at all large.

Remark 9. Because $\tilde{S}^{*-1}_{22}$ is a (principal) submatrix of $(\tilde{W}'\tilde{W}/T)^{-1}$ and the latter must be invertible for estimators to exist, we conclude that **if estimators can be obtained in this context** $\tilde{S}^*_{22}$ is **invertible** and, thus, this test can **always** be carried out. Consequently, this should become the standard test of choice, and there is no particular reason one should employ the test based on δ. Perhaps in an era of less powerful computing capabilities the lower dimensions of the relevant regressions made δ appealing; this is no longer true, however, and consequently there is no reason to employ it, given that it does not always produce conclusive results–even if all estimators required can be obtained! The only possible reason to employ it may be deduced from Proposition 1, below.

Remark 10. The two δ tests (δ and δ^*) discussed above are not identical (or equivalent), although their (respective) test statistics have the same limiting distribution. The first version, δ, is a **conformity** test, i.e. we estimate under the null and the test asks whether the results "conform" with the null.[2] The second test, δ^*, is a likelihood ratio type of test, i.e. we estimate using the form of the alternative and we ask of the results whether they support the null. This should explain why sometimes the first does not produce a definitive answer, while the second always does.

4 Diagonal R

When the autoregression matrix R is diagonal, the situation is more complex than that of the simple Durbin context, **unless**

$$\mathrm{Cov}(\epsilon_{t\cdot}) = \mathrm{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{mm}), \tag{38}$$

in which case we are reduced to doing m h-tests *seriatim*.

We could deal with this case by simply using the results above and taking into account only the diagonal elements of the estimator of R, as defined in the previous discussion, i.e. for the diagonal specification we set[3] $\tilde{r}_D = H'\tilde{R}$ and derive the limiting distribution in the diagonal case by using he results of Theorems 1 and 2. To do so, however, entails potential loss of efficiency in that we do not fully take into account all available information and, in particular, the assertion that

$$r_{ij} = 0, \quad i \neq j.$$

Thus, we shall employ a new framework using the **fact** that R is diagonal and the covariance matrix of the structural error, ϵ, is **unrestricted**, i.e. we produce the analog of the δ and δ^*-statistics when R is diagonal but the elements of $u_{t\cdot}$ are cross correlated. Specifically, the alternative dealt with is

$$u_{t\cdot} = u_{t-1\cdot}R + \epsilon_{t\cdot}, \quad R = \mathrm{diag}(r_{11}, r_{22}, \ldots, r_{mm}), \quad \mathrm{Cov}(\epsilon_{t\cdot}) = \Sigma > 0, \tag{39}$$

[2] For reasons that are not clear to me such tests are often termed *Wald tests*.

[3] This notation will be clarified immediately below.

where

$$\Sigma = (\sigma_{ij}), \quad \sigma_{ij} \neq 0, \quad \text{for} \quad i \neq j\,.$$

If the u's could be observed, we would write the model as

$$u = V r_D + e, \qquad u = \mathrm{vec}(U), \qquad V = \mathrm{diag}(v_{\cdot 1}, v_{\cdot 2}, \ldots, v_{\cdot m}),, \tag{40}$$

where $v_{\cdot i}$ is the ith column of U_{-1}, $r_D = (r_{11}, r_{22}, \ldots, r_{mm})'$, and estimate

$$\hat{r}_D = [V'(\Sigma^{-1} \otimes I_T)V]^{-1} V'(\Sigma^{-1} \otimes I_T)u \tag{41}$$

the limiting distribution of which is given by

$$\sqrt{T}\hat{r}_D \xrightarrow{\mathrm{d}} N(0, \Omega_*^{-1}), \qquad \Omega_* = (\sigma_{ij}\sigma^{ij}). \tag{42}$$

In connection with the result above we have,

Lemma 1. The matrix

$$\Omega_* = (\sigma^{ij}\sigma_{ij})$$

is positive (semi) definite if and only if $\Sigma \otimes \Sigma^{-1}$ has this property.

Proof: Define the matrix

$$H = \mathrm{diag}(e_{\cdot 1}, e_{\cdot 2}, \ldots, e_{\cdot m}), \tag{43}$$

where $e_{\cdot i}$ is an m-element column vector all of whose element are zero save the i^{th}, which is unity. Note that H is $m^2 \times m$, of rank m and its columns are orthonormal. It may be easily verified that

$$\Omega_* = H'(\Sigma \otimes \Sigma^{-1})H. \tag{44}$$

Let α be an arbitrary m-element vector and consider

$$g = \alpha'\Omega_*\alpha = (\alpha' H)(\Sigma \otimes \Sigma^{-1})(H\alpha). \tag{45}$$

Necessity: Since $H\alpha = 0$ if an only if $\alpha = 0$ and $\Omega_* > 0$ by assumption, this implies that $\Sigma \otimes \Sigma^{-1} > 0$. Sufficiency: Suppose that $(\Sigma \otimes \Sigma^{-1}) > 0$, then for arbitrary non-null α

$$g = \alpha'\Omega_*\alpha = (\alpha' H)(\Sigma \otimes \Sigma^{-1})(H\alpha) > 0,$$

owing to the fact that if $\alpha \neq 0$, $H\alpha \neq 0$.

Remark 11. Notice that for a block matrix, say P, the operation $H'PH$ simply creates a matrix whose (ij) element is the (ij) element of the (ij) block of the matrix P. Thus, for example the (ij) block of $\Sigma \otimes \Sigma^{-1}$ is $\sigma_{ij}\Sigma^{-1}$ and the (ij) element of this block is $\sigma_{ij}\sigma^{ij}$. It is thus quite apparent that $H'(\Sigma^{-1} \otimes \Sigma)H = H'(\Sigma \otimes \Sigma^{-1})H$.

The estimator in Eq. (41) is infeasible because in fact the $u's$ are not observed. Since they are not, we may try using instead the corresponding OLS residuals from the regression of Y on Z. When we do so we have, under the null,

$$\sqrt{T}\tilde{r}_D = \left(\frac{\tilde{V}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{V}}{T}\right)^{-1} \frac{1}{\sqrt{T}}\tilde{V}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{u}. \tag{46}$$

To determine its limiting distribution we need a slightly different approach than the one employed in previous discussions. To this end note that

$$\tilde{V}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{V} = [\tilde{\sigma}^{ij}\tilde{v}'_{\cdot i}\tilde{v}_{\cdot j}], \tag{47}$$

i.e. it is an $m \times m$ matrix with typical element $\tilde{\sigma}^{ij}\tilde{v}'_{\cdot i}\tilde{v}_{\cdot i}$. Evidently,

$$\frac{1}{T}\tilde{\sigma}^{ij}\tilde{v}'_{\cdot i}\tilde{v}_{\cdot j} \xrightarrow{\mathrm{P}} \sigma_{ij}\sigma^{ij},$$

and thus

$$\frac{1}{T}\tilde{V}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{V} \xrightarrow{\mathrm{P}} \Omega_*, \tag{48}$$

as was the case when the elements of U were assumed to be known.

Recalling that

$$\tilde{v}_{\cdot i} = v_{\cdot i} - Z_{-1}(\tilde{c}_{\cdot i} - c_{\cdot i}), \quad \tilde{u}_{\cdot j} = u_{\cdot j} - Z(\tilde{c}_{\cdot j} - c_{\cdot j})$$

we have

$$\frac{1}{\sqrt{T}}\tilde{v}'_{\cdot i}\tilde{u}_{\cdot j} \sim \frac{1}{\sqrt{T}}(v_{\cdot i} - ZS_1'\sigma_{\cdot i})'u_{\cdot j} = \frac{1}{T}v^{*\prime}_{\cdot i}u_{\cdot j}, \quad v^*_{\cdot i} = v_{\cdot i} - ZS_1'\sigma_{\cdot i}. \tag{49}$$

We can now write the rightmost term of Eq. (46) as

$$\frac{1}{\sqrt{T}}\tilde{V}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{u} \sim \frac{1}{\sqrt{T}}V^{*\prime}(\Sigma^{-1} \otimes I_T)u, \quad V^* = \mathrm{diag}(v^*_{\cdot 1}, v^*_{\cdot 2}, \ldots, v^*_{\cdot m}), \tag{50}$$

the i^{th} element of the right vector above being

$$\begin{aligned}\sum_{j=1}^{m} \sigma^{ij} v^{*\prime}_{\cdot i} u_{\cdot j} &= \sum_{t=2}^{T} v^*_{ti} \sum_{j=1}^{m} \sigma^{ij} u_{tj} \\ &= \sum_{t=2}^{T} v^*_{ti} \sigma^{i\cdot} u'_{t\cdot}.\end{aligned} \tag{51}$$

Define the vector

$$\zeta^*_{(T)} = \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \zeta_t, \quad \zeta_t = \begin{bmatrix} v^*_{t1}\sigma^{1\cdot} \\ v^*_{t2}\sigma^{2\cdot} \\ \vdots \\ v^*_{tm}\sigma^{m\cdot} \end{bmatrix} u'_{t\cdot}.$$

It is easily seen that the summands above, the $zeta$'s, are a zero mean MD sequence obeying the Lindeberg condition, just as in the earlier discussion. They also obey the sufficient condition for Proposition 21, Dhrymes (1989), p. 327, because

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} E[\zeta_t \zeta'_t | \mathcal{A}_{t-1}] = (\sigma^{ij}\omega_{ij}), \quad \omega_{ij} = \sigma_{ij} - \sigma_{i\cdot} S_{11} \sigma_{\cdot j}. \tag{52}$$

Since the sequence obeys a sufficient condition for the application of a MD CLT, $\zeta^*_{(T)}$ converges in distribution, and we conclude that under the null, for the case of diagonal R,

$$\sqrt{T}\tilde{r}_D \xrightarrow{\mathrm{d}} N(0, \Phi), \quad \Phi = \Omega_*^{-1} - \Omega_*^{-1}(\sigma^{ij}\sigma_{i\cdot}S_{11}\sigma_{\cdot j})\Omega_*^{-1}, \quad \mathit{or} \tag{53}$$

$$\Phi = (\sigma^{ij}\sigma_{ij})^{-1}\Omega_1(\sigma^{ij}\sigma_{ij})^{-1}, \quad \Omega_1 = [\sigma^{ij}(\sigma_{ij} - \sigma_{i\cdot}S_{11}\sigma_{\cdot j})]. \tag{54}$$

If the matrix Ω_1 is at least positive semi-definite, we may carry out a test of the null by means of the test statistic

$$\delta_D = T\tilde{r}'_D\tilde{\Phi}^{-1}\tilde{r}_D \xrightarrow{\mathrm{d}} \chi^2_m, \quad \text{or, more generally,} \quad \delta_D \xrightarrow{\mathrm{d}} \chi^2_{\mathrm{rank}(\Omega_1)}. \tag{55}$$

We have therfore proved

Theorem 3. Under the conditions of Theorem 1, and the additional specification that $R = \mathrm{diag}(r_{11}, r_{22}, \ldots, r_{mm})$, the following statements are true:

i. If the elements of U are **known** we may estimate

$$\hat{r}_D = [V'(\Sigma^{-1} \otimes I_T)V]^{-1}V'(\Sigma^{-1} \otimes I_T)u$$

which, under the null, obeys

$$\sqrt{T}\hat{r}_D \xrightarrow{\mathrm{d}} N(0, \Omega_*^{-1}), \quad \Omega_* = (\sigma_{ij}\sigma^{ij}).$$

ii. Because we do not generally know the matrix of the structural error observations, the estimator above, though consistent, is infeasible. Instead we obtain, under the null, the feasible estimator

$$\tilde{r}_D = \left(\frac{\tilde{V}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{V}}{T}\right)^{-1} \frac{1}{T}\tilde{V}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{u} \xrightarrow{\mathrm{P}} 0, \quad \text{and also obeys}$$

$$\sqrt{T}\tilde{r}_D \sim \left(\frac{V'(\Sigma^{-1} \otimes I_T)V}{T}\right)^{-1} \frac{1}{\sqrt{T}}V^{*\prime}(\Sigma^{-1} \otimes I_T)u,$$

which has the limiting distribution

$$\sqrt{T}\tilde{r}_D \xrightarrow{\mathrm{d}} N(0, \Phi), \quad \Phi = \Omega_*^{-1} - \Omega_*^{-1}(\sigma^{ij}\sigma_{i\cdot}S_{11}\sigma_{\cdot j})\Omega_*^{-1},$$

as defined in Eqs. (42) through (54).

iii. The estimator given in ii is **efficient** relative to that given in i.

Corollary 3. A test of the null hypothesis

$H_0: \quad R = 0$
as against the alternative
$H_1: \quad R \neq 0$

may be carried out by means of the statistic

$$\delta_D = T\tilde{r}'_D\tilde{\Phi}^{-1}\tilde{r}_D \xrightarrow{\mathrm{d}} \chi^2_m,$$

or more generally

$$\delta_D = T\tilde{r}'_D\tilde{\Phi}_g\tilde{r}_D \xrightarrow{\mathrm{d}} \chi^2_{\mathrm{rank}(\Phi)},$$

where $\Phi_{22(g)}$ is the **generalized inverse** of Φ_{22}.

Remark 12. A comparison of Eqs. (42) and (53) discloses the same phenomenon noted in Remark 3 i.e. that, **knowing more** results in inefficiency relative to the case where we know less or, more specifically, **knowing** U results in an estimator which is **inefficient** relative to a similar estimator obtained when U **is not known** but is estimated by the matrix of the least squares residuals in the regression of Y on Z.

Remark 13. Notice that in the case $m = 1$, δ_D reduces to the square of the h-statistic because $\Omega_* = 1$ and $\Omega_1 = 1 - \mathrm{Avar}(\hat{a}_{11})$, as in Durbin (1970a).

If the (estimated) matrix Ω_1 is **indefinite, or negative definite**, the test above is inoperable and an alternative test may be undertaken as follows. Write (the observations on) the i^{th} equation of the model as

$$y_{\cdot i} = Y_{-1}a_{\cdot i} + Xb_{\cdot i} + r_{ii}v_{\cdot i} + \epsilon_{\cdot i} = Zc_{\cdot i} + r_{ii}v_{\cdot i} + \epsilon_{\cdot i}, \quad i = 1, 2, 3, \dots, m, \tag{56}$$

and stack them so that the observations on the entire model can be written as

$$y = (I_m \otimes Z)c + Vr_D + e, \quad r_D = (r_{11}, r_{22}, \dots, r_{mm})', \quad V = \mathrm{diag}(v_{\cdot 1}, v_{\cdot 2}, \dots, v_{\cdot m}). \tag{57}$$

Since V is not observable we use instead the columns ($\tilde{v}_{\cdot i}$) of the matrix of the OLS residuals $\tilde{U}_{-1}$, i.e.

$$\tilde{V} = \mathrm{diag}(\tilde{v}_{\cdot 1}, \tilde{v}_{\cdot 2}, \tilde{v}_{\cdot 3}, \dots, \tilde{v}_{\cdot m}), \tag{58}$$

and estimate

$$\tilde{d} = [(\tilde{W}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{W}]^{-1}\tilde{W}'(\tilde{\Sigma}^{-1} \otimes I_T)y, \quad \tilde{W} = [(I_m \otimes Z), \tilde{V}], \quad d = (c', r')'. \tag{59}$$

As in the discussion above we can show that, under the null,

$$[(\tilde{W}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{W}]^{-1}\tilde{W}'(\tilde{\Sigma}^{-1} \otimes I_T)Wd = (I, 0)'c, \tag{60}$$

so that the estimator of d and hence of r_D is consistent. We can further show that, under the null,

$$\sqrt{T}(\tilde{d} - d) \sim \Psi \frac{1}{\sqrt{T}} W'(\Sigma^{-1} \otimes I_m)e \xrightarrow{\mathrm{d}} N(0, \Psi), \tag{61}$$

where

$$\Psi^{-1} = \operatorname*{plim}_{T\to\infty} \frac{1}{T} W'(\Sigma^{-1} \otimes I_T)W. \tag{62}$$

It follows then that, under the null,

$$\sqrt{T}\tilde{r} \xrightarrow{\mathrm{d}} N(0, \Psi_{22}), \tag{63}$$

where Ψ_{22} is the $m \times m$ principal submatrix of Ψ, consisting of its last m rows and columns. Consequently, to test the null $H_0 : r_D = 0$ we may use the test statistic

$$\delta_D^* = T\tilde{r}_D'\tilde{\Psi}_{22}^{-1}\tilde{r}_D \xrightarrow{\mathrm{d}} \chi_m, \tag{64}$$

where

$$\tilde{\Psi}_{22}^{-1} = \frac{1}{T}\left(\tilde{V}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{V} - \tilde{V}'(\tilde{\Sigma}^{-1} \otimes Z(Z'Z)^{-1}Z')\tilde{V}\right), \tag{65}$$

and $\tilde{\Sigma} = \tilde{U}'\tilde{U}/T$.

If an estimator for r_D is obtainable, the matrices of Eqs. (62) and (65) will be **positive definite**, and hence invertible, so that this test is always operational in practice.

We have therefore proved

Theorem 4. Under the conditions of Theorem 1, and the additional specification that $R = \mathrm{diag}(r_{11}, r_{22}, \ldots, r_{mm})$, let the elements of U be estimated by means of the residuals of the regression of Y on Z, i.e. $\tilde{U} = Y - Z\tilde{C}$.

Using generalized least squares with estimated covariance matrix $\tilde{\Sigma} = (\tilde{U}'\tilde{U}/T)$, regress $y = \mathrm{vec}(Y)$ on $\tilde{W}$ to obtain

$$\tilde{d} = [(\tilde{W}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{W}]^{-1}\tilde{W}'(\tilde{\Sigma}^{-1} \otimes I_T)y, \quad \tilde{W} = [(I_m \otimes Z), \tilde{V}], \quad d = (c', r')'.$$

Then the following statements are true:

i. Under the null

$$\tilde{d} - d \sim \tilde{\Psi}\left(\frac{1}{T}\right)\tilde{W}'(\tilde{\Sigma}^{-1} \otimes I_T)u \xrightarrow{\mathrm{P}} 0,$$

and

$$\sqrt{T}\tilde{d} \xrightarrow{\mathrm{d}} N(0, \Psi), \quad \Psi = \operatorname*{plim}_{T\to\infty}\left(\frac{\tilde{W}'(\tilde{\Sigma} \otimes I_T)\tilde{W}}{T}\right)^{-1}.$$

ii. In addition

$$\sqrt{T}\tilde{r}_D \xrightarrow{\mathrm{d}} N(0, \Psi_{22}), \quad \Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix},$$

where $\Psi_{22} = \Omega_1^{-1}$ as the latter is defined in Eq. (54). The estimator of Ψ_{22}^{-1} is defined in Eq. (65).

Corollary 4. A test of the null hypothesis

$H_0: \quad r_D = 0$
as against the alternative
$H_1: \quad r_D \neq 0$

may be carried out by means of the statistic

$$\delta_D^* = T\tilde{r}_D'(\tilde{\Psi}_{22})^{-1}\tilde{r}_D \xrightarrow{\mathrm{d}} \chi_m^2,$$

or more generally

$$\delta_D^* = T\tilde{r}_D'\tilde{\Psi}_{(22)g}\tilde{r}_D \xrightarrow{\mathrm{d}} \chi^2_{\mathrm{rank}(\Psi_{22})}.$$

We now consider the relative efficiencies among the estimators obtained in Theorems 1 through 4. Since all estimators converge in distribution to a zero mean **normal** vector, questions of relative efficiency are completely resolved by considering differences of the covariance matrices of the respective limiting distributions.
We have

Proposition 1 The following statements are true:

i. The estimator of Theorem 1 is efficient relative to the estimator in Theorem 2.

ii. The estimator of Theorem 3 is efficient relative to the estimator of Theorem 4.

Proof. Let

$$\Delta_{12} = \Sigma \otimes F, \quad F = (\Sigma - \Sigma S_{11} \Sigma)^{-1} - (\Sigma^{-1} - S_{11}), \tag{66}$$

and note that Δ_{12} is positive (semi) definite if and only if F is. Using the result in Dhrymes (2000) p. 44, we have that

$$F = \Sigma^{-1} + (S_{11}^{-1} - \Sigma)^{-1} - (\Sigma^{-1} - S_{11}) = S_{11} + (S_{11}^{-1} - \Sigma)^{-1}. \tag{67}$$

Assuming that $\Sigma^{-1} - S_{11}$ is positive (semi) definite, so that the test of Theorem 1 is applicable, we conclude from Dhrymes (2000), Proposition 2.66 p. 89, that $(S_{11}^{-1} - \Sigma) \geq 0$, which shows that $\Delta_{12} \geq 0$, thus proving i.

To prove ii, consider

$$\Delta_{34} = \Omega_1^{-1} - \Omega_*^{-1} \Omega_1 \Omega_*^{-1}, \quad \Omega_1 = \Omega_* - G, \quad G = [\sigma^{ij}(\sigma_{i \cdot} S_{11} \sigma_{\cdot j})], \tag{68}$$

so that using again the results in Dhrymes (2000) p. 44 and p. 89, we obtain

$$\Omega_1^{-1} = \Omega_*^{-1} + \Omega_*^{-1}(G^{-1} - \Omega_*^{-1})^{-1} \Omega_*^{-1} \quad \text{and, consequently} \tag{69}$$

$$\Delta_{34} = \Omega_*^{-1}(G^{-1} - \Omega_*^{-1})^{-1} \Omega_*^{-1} + \Omega_*^{-1} G \Omega_*^{-1} \geq 0. \tag{70}$$

That G is a positive definite matrix follows from the fact that

$$G = H'(\Sigma^{-1} \otimes \Sigma S_{11} \Sigma) H.$$

q.e.d.

Remark 14. Because $r_D = H'\mathrm{vec}(R)$, it may be interesting to compare the results of Theorem 1 to those of Theorem 3, and those of Theorem 2 to those of Theorem 4. The issue is the relation between the limiting covariance matrices $H'[\Sigma \otimes (\Sigma^{-1} - S_{11})]H$ implied by Theorem 1 and Ω_*^{-1} of Theorem 3; similarly an interesting comparion would be between $H'(\Sigma \otimes S_{22}^*)H$ implied by Theorem 2, and the limiting covariance matrix Ψ_{22} of Theorem 4.

A direct comparison cannot be easily made owing to the complexity of the expressions; on the other hand, from the standard theory of restricted least squares, the **unrestricted estimator is inefficient relative to the restricted estimator**, when the restriction is valid. Consequently, estimating $\tilde{r}_D = H'\mathrm{vec}(\tilde{R})$ is inefficient relative to the direct estimator of r_D.

REFERENCES

Davidson, R and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Dhrymes, P. J. (1989), *Topics in Advanced Econometrics: Probability Foundations*, New York: Springer-Verlag.

Dhrymes, P. J. (1994), *Topics in Advanced Econometrics: vol. 2, Linear and Nonlinear Simultaneous Equations*, New York: Springer-Verlag.

Dhrymes, P. J. (2000), *Mathematics for Econometrics*, third edition, New York: Springer-Verlag.

Durbin, J. M. (1970a), "Testing for Serial Correlation in Least Squares Regression when some of the Regressors are Lagged Dependent Variables, *Econometrica*, vol. 38, pp. 410-421.

Durbin, J. M. (1970b), "An Alternative to the Bounds Test for Testing for Serial Correlation in Least-Squares Regression, *Econometrica*, vol. 38, pp. 422-429.

Greene, W. H. (1999), *Econometric Analysis*, fourth edition, New York: Prentice-Hall.