JOURNAL OF
Econometrics

ELSEVIER Journal of Econometrics 95 (2000) 157–176

www.elsevier.nl/locate/econbase

On the sensitivity of the usual t- and F-tests to covariance misspecification

Anurag N. Banerjee[a], Jan R. Magnus[b,*]

[a]*Department of Economics, Queen's University, Belfast, UK*
[b]*CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, Netherlands*

Received 1 August 1997; received in revised form 1 March 1998; accepted 1 April 1999

Abstract

We consider the standard linear regression model with all standard assumptions, except that the disturbances are not white noise, but distributed $\mathrm{N}(0, \sigma^2 \Omega(\theta))$ where $\Omega(0) = I_n$. Our interest lies in testing linear restrictions using the usual F-statistic based on OLS residuals. We are not interested in finding out whether $\theta = 0$ or not. Instead we want to find out what the effect is of possibly nonzero θ on the F-statistic itself. We propose a sensitivity statistic ϕ for this purpose, discuss its distribution, and obtain a practical and easy-to-use decision rule to decide whether the F-test is sensitive or not to covariance misspecification when θ is close to zero. Some finite and asymptotic properties of φ are studied, as well as its behaviour in the special case of an AR(1) process near the unit root. © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C12; C22; C51; C52

Keywords: Linear regression; Least squares; t-test; F-test; Autocorrelation; Sensitivity; Robustness

1. Introduction

Suppose that, in the linear regression model $y = X\beta + u$ under standard assumptions, we are interested in testing linear restrictions $R\beta = r$. We are,

* Corresponding author. Tel.: + 31-13-4663092.
E-mail address: magnus@kub.nl (J.R. Magnus)

PII: S0304-4076(99)00034-2

however, uncertain about the distribution of the disturbances. If the disturbances are white noise, then the usual F-statistic based on OLS residuals follows an F-distribution under the null hypothesis. This is the textbook case. If the disturbances are *not* white noise but are distributed $\mathrm{N}(0, \sigma^2 \Omega(\theta))$, where the structure of Ω is known (for example AR(1)) and in addition the value of θ is known, then $F(\theta)$, an extended F-statistic based on GLS residuals, can be used.

The assumption that the structure of Ω is known might have some justification, but knowledge of θ goes beyond what can be reasonably assumed. This raises two questions. First, if θ is not known (or, worse still, if the structure of Ω is not known), how should we test the restriction? Secondly, even if we could construct an acceptable test statistic, does it make any difference, that is, is the result of this test really different from the result of the usual F-test based on OLS residuals?

One way to tackle the first question is by pretesting. Nakamura and Nakamura (1978) and King and Giles (1984) look at the effect of pretesting θ in an AR(1) environment. They conclude that the problems are very serious, which may in part be caused by the fact that *all* pretesting procedures have bad properties, because of the discontinuity of the procedure, see Magnus (1999).

Several authors have attempted to answer the second question by resorting to a 'bounds test'. The simplest bounds test is to draw $F(\theta)$ as a function of θ, and to reject the restriction if $F(\theta) > c$ (say) for all θ, not reject the restriction if $F(\theta) < c$ for all θ, and draw no conclusion otherwise. This is one of Dufour's (1990, p. 483) suggestions, but as he admits 'can lead to unduly large inconclusive regions'. The approach taken by Durbin and Watson (1950, 1951) to test $\theta = 0$ can be extended to test a hypothesis about the regression coefficients, see Vinod (1976), Kiviet (1980), Vinod and Ullah (1981, Chapter 4), and Hillier and King (1987). This involves finding upper and lower bounds for the distribution of the standard t-statistic. One problem with this approach in an AR(1) environment is that the upper bound may approach infinity as θ goes to one, see Vinod (1976, p. 930), Krämer (1989). So, this method breaks down near $\theta = 1$. Dufour (1990) tries to modify the procedure by jointly estimating the confidence intervals of θ and $w'\beta - r$ (the restriction) in the particular case of AR(1) disturbances. Using the union–intersection principle he then succeeds in reducing the estimates of the bounds, using information through the Durbin–Watson test. But Dufour's approach also has a problem: When testing the hypothesis that the intercept coefficient is zero, the coefficient is not identified when $\theta = 1$.

Another approach to answer the second question is through asymptotics. Understandably, asymptotics does not work well near $\theta = 1$. Park and Mitchell (1980) show that asymptotic critical values can be very unreliable near $\theta = 1$. Even expansion methods do not give good inferences near $\theta = 1$, see Rothenberg (1984a, b, 1988).

In this paper we take a different approach, similar to the one developed in Banerjee and Magnus (1999). We do not test whether $\theta = 0$ or not. Neither do we attempt to find tighter confidence intervals for the t-statistic or to improve its asymptotic approximation using estimated GLS. Our approach is based on local sensitivity analysis and asks whether the F-statistic (t-statistic) is sensitive to deviations from the white noise assumption. One important advantage of our approach is that we do not need to know the structure of Ω; only its derivative at $\theta = 0$ is required.

In addition to the theoretical results in the paper, we propose a practical sensitivity measure ϕ, which is easy to calculate. We obtain an excellent approximation to the median of $|\phi|$ (also easy to calculate), and we propose as a 'rule of thumb' and a practical tool for econometricians, the rule that the F-statistic is 'sensitive' to covariance misspecification if $|\phi|$ is larger than its approximated median, and not sensitive otherwise.

The paper is organized as follows. Section 2 gives the set-up and notation. In Section 3 we define the sensitivity of the F-statistic, denoted ϕ, and obtain an explicit expression (Theorem 1). In Section 4 we derive three theorems about the exact and large sample behaviour of ϕ. The last of these (Theorem 4) leads to the median-based 'rule of thumb'. In Section 5 we specialize our treatment to AR(1) disturbances and discuss the behaviour of ϕ when the AR(1) parameter θ is close to one. This analysis shows that we must distinguish between three cases. In Section 6 we summarize our main findings. There are two appendices. Appendix A gives some little-known results about the product normal distribution, while Appendix B contains the proofs of the theorems.

2. Set-up and notation

We shall consider the standard linear regression model

$$y = X\beta + u, \tag{1}$$

where y is an $n \times 1$ random vector of observations, X a non-random $n \times k$ matrix of regressors ($k < n$), β a $k \times 1$ vector of unknown parameters and u an $n \times 1$ vector of random disturbances. We assume that X has full column-rank k and that u follows a normal distribution, $u \sim \mathrm{N}(0, \sigma^2 \Omega(\theta))$. Without essential loss of generality, we shall assume throughout this paper that Ω is a matrix function of a *single* parameter $\theta \in \Theta$, and that $\Omega(\theta)$ is positive definite and differentiable, at least in a neighbourhood of $\theta = 0$. We also assume that $\Omega(0) = I_n$ and that $\sigma^2 > 0$. An important role is played by the symmetric $n \times n$ matrix

$$A = \left. \frac{\mathrm{d}\Omega(\theta)}{\mathrm{d}\theta} \right|_{\theta = 0}. \tag{2}$$

We notice that $\mathrm{d}\Omega^{-1}(\theta)/\mathrm{d}\theta|_{\theta=0} = -A$.

If there is no restriction on β, then the (unrestricted) generalized least-squares (GLS) estimator for β is

$$\hat{\beta}(\theta) = (X'\Omega^{-1}(\theta)X)^{-1}X'\Omega^{-1}(\theta)y. \tag{3}$$

If there is a restriction on β, say $R\beta = r$, where R is a $q \times k$ matrix of rank $q \geqslant 1$, then the restricted GLS estimator for β is

$$\tilde{\beta}(\theta) = \hat{\beta}(\theta) - (X'\Omega^{-1}(\theta)X)^{-1}R'(R(X'\Omega^{-1}(\theta)X)^{-1}R')^{-1}(R\hat{\beta}(\theta) - r). \tag{4}$$

If we assume that θ is known, then the usual F-statistic for testing the hypothesis $R\beta = r$ can be written as

$$F(\theta) = \frac{(R\hat{\beta} - r)'(R(X'\Omega^{-1}(\theta)X)^{-1}R')^{-1}(R\hat{\beta} - r)}{\hat{u}'(\theta)\Omega^{-1}(\theta)\hat{u}(\theta)}\frac{n-k}{q}, \tag{5}$$

or alternatively as

$$F(\theta) = \frac{\tilde{u}'(\theta)\Omega^{-1}(\theta)\tilde{u}(\theta) - \hat{u}'(\theta)\Omega^{-1}(\theta)\hat{u}(\theta)}{\hat{u}'(\theta)\Omega^{-1}(\theta)\hat{u}(\theta)}\frac{n-k}{q}, \tag{6}$$

where

$$\hat{u}(\theta) = y - X\hat{\beta}(\theta), \qquad \tilde{u}(\theta) = y - X\tilde{\beta}(\theta). \tag{7}$$

Note that the equality of (5) and (6) holds whether or not the restriction $R\beta = r$ is satisfied. Of course, under the null hypothesis H_0: $R\beta = r$, $F(\theta)$ is distributed as $F(q, n-k)$.

Suppose we believe that $\theta = 0$, which may or may not be the case. Then we would use the OLS estimator $\hat{\beta}(0)$ or the restricted OLS estimator $\tilde{\beta}(0)$. We now define the symmetric idempotent $n \times n$ matrices

$$M = I_n - X(X'X)^{-1}X' \tag{8}$$

and

$$B = X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}X', \tag{9}$$

satisfying

$$MB = 0, \quad \text{rank}(M) = n - k, \quad \text{rank}(B) = q. \tag{10}$$

We then have, with $u \sim \mathrm{N}(0, \sigma^2 \Omega(\theta))$, $\hat{u} = \hat{u}(0) = Mu$, and, if the restriction $R\beta = r$ is satisfied, $\tilde{u} = \tilde{u}(0) = (M + B)u$.

3. Sensitivity of the *F*-test

We want to find out how sensitive the F-statistic is with respect to small changes in θ when θ is close to 0. As in Banerjee and Magnus (1999), we do *not* ask the question whether θ is 0 or not, using for example a Durbin–Watson test. Instead, we think of θ as a nuisance parameter whose estimate may or may not be 'significantly' different from 0. But even when θ is 'far' from 0, this does not imply that $F(\theta)$ is 'far' from $F(0)$. And this is what interests us in this paper: Is it legitimate to use $F(0)$ — based on OLS residuals — instead of $F(\theta)$?

Thus motivated we define the sensitivity of the F-statistic $F(\theta)$ as

$$\varphi = \left.\frac{\mathrm{d}F(\theta)}{\mathrm{d}\theta}\right|_{\theta=0}, \tag{11}$$

where $F(\theta)$ is given in (5) or (6). Large values of φ indicate that $F(\theta)$ is sensitive to small changes in θ when θ is close to 0 and hence that setting $\theta = 0$ is not justified. The statistic φ depends only on y and X (and, of course, on R and r) and can therefore be observed. The *distribution* of φ does, however, depend on θ (and, if the restriction $R\beta = r$ is not satisfied, on σ^2 as well).

Theorem 1. We have

$$\varphi = 2\left(F(0) + \frac{n-k}{q}\right)(\hat{\theta} - \tilde{\theta}), \tag{12}$$

where

$$\hat{\theta} = \frac{1}{2}\frac{\hat{u}'A\hat{u}}{\hat{u}'\hat{u}}, \quad \tilde{\theta} = \frac{1}{2}\frac{\tilde{u}'A\tilde{u}}{\tilde{u}'\tilde{u}}, \tag{13}$$

$$F(0) = \frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\hat{u}'\hat{u}}\frac{n-k}{q}, \tag{14}$$

where $\hat{u}$ and $\tilde{u}$ denote the unrestricted and restricted OLS residuals, and A is defined in (2).

We note that Theorem 1 is valid whether or not the restriction $R\beta = r$ is satisfied, and also whether or not the distribution of y is evaluated at $\theta = 0$. We see from Theorem 1 that φ is a function of quadratic forms in normal variables, but, since these quadratic forms are not independent, it does not appear feasible to obtain the density of φ in closed form. We shall obtain certain limiting results (Sections 4 and 5) and also the first two moments of φ exactly.

The notation $\hat{\theta}$ and $\tilde{\theta}$ in (12) and (13) suggests that these statistics can be interpreted as estimators of θ. This suggestion is based on the following argument. We expand $\Omega(\theta)$ as

$$\Omega(\theta) = I_n + \theta A + \tfrac{1}{2}\theta^2 H + \mathcal{O}(\theta^3), \tag{15}$$

where A is defined in (2), and H denotes the second derivative at 0. Then,

$$\Omega^{-1}(\theta) = I_n - \theta A + \tfrac{1}{2}\theta^2(2A^2 - H) + \mathcal{O}(\theta^3). \tag{16}$$

If the y-process is covariance stationary, we may assume that the diagonal elements of Ω are all ones. Then, $\operatorname{tr}A = \operatorname{tr}H = 0$ and

$$\operatorname{tr}\left(\frac{\mathrm{d}\Omega^{-1}(\theta)}{\mathrm{d}\theta} \cdot \Omega(\theta)\right) = \theta \operatorname{tr} A^2 + \mathcal{O}(\theta^2). \tag{17}$$

We next expand $\hat{u}(\theta)$ as

$$\hat{u}(\theta) = \hat{u}(0) + \theta X(X'X)^{-1}X'A\hat{u}(0) + \mathcal{O}(\theta^2), \tag{18}$$

so that, writing $\hat{u}$ instead of $\hat{u}(0)$,

$$\hat{u}'(\theta)\frac{\mathrm{d}\Omega^{-1}(\theta)}{\mathrm{d}\theta}\hat{u}(\theta) = -\hat{u}'A\hat{u} + \theta(2\hat{u}'AMA\hat{u} - \hat{u}'H\hat{u}) + \mathcal{O}(\theta^2). \tag{19}$$

The maximum-likelihood estimator for θ is obtained by equating (17) and (19), see Magnus (1978). This gives

$$\hat{\theta}_{\mathrm{ML}} \approx \frac{\hat{u}'A\hat{u}}{2\hat{u}'AMA\hat{u} - \hat{u}'H\hat{u} - \operatorname{tr}A^2} = \frac{1}{2}\frac{\hat{u}'A\hat{u}}{\hat{u}'\hat{u}}(1 + n^{-1/2}\delta), \tag{20}$$

where δ will be bounded in probability if $(1/n)\operatorname{tr}A^2 \to 2$. This will usually be the case, certainly for low-order ARMA processes; see Section 5 for the particular case of AR(1). In essence, therefore, all properties of the distribution of φ are determined by the behaviour of $n(\hat{\theta} - \tilde{\theta})$, the difference between the unrestricted and the restricted 'estimator' of θ.

4. Finite and asymptotic behaviour of φ

Using Pitman's Lemma (see Lemma 7 in Appendix B), we obtain the first two exact moments of φ.

Theorem 2. Assume that the distribution of y is evaluated at $\theta = 0$ *and that the restriction* $R\beta = r$ *is satisfied. Then,*

$$\mathrm{E}\phi = -\frac{\operatorname{tr} AB}{q} - \frac{1}{n-k-2}\left(2\frac{\operatorname{tr} AB}{q} - \operatorname{tr} AM\right)$$

and

$$\mathrm{E}\phi^2 = \frac{n-k}{n-k-2}\left(\frac{n-k}{n-k-4}\frac{2\operatorname{tr}(AB)^2 + (\operatorname{tr} AB)^2}{q^2} + \frac{4}{q^2}\operatorname{tr} ABAM\right) + \frac{n-k}{(n-k-2)(n-k-4)} \left(\frac{q+2}{q}\cdot\frac{2\operatorname{tr}(AM)^2 + (\operatorname{tr} AM)^2}{n-k+2} - 2(q+2)(\operatorname{tr} AB)(\operatorname{tr} AM)\right).$$

While the exact moments of φ are easy to calculate in specific situations, we gain further insight in the behaviour of φ by studying its large sample approximation. This approximation is given in Theorem 3.

Theorem 3. Assume that the distribution of y is evaluated at $\theta = 0$ *and that the restriction* $R\beta = r$ *is satisfied. Assume further that* (i) $\Omega(\theta)$ *is normalized such that* $\operatorname{tr} \Omega(\theta) = n$ *for all* $\theta \in \Theta$, *and* (ii) *that the eigenvalues of A are bounded. Then, letting* $z = \tilde{u}' A \tilde{u} - \hat{u}' A \hat{u}$, *we obtain for large n,*

$$\varphi = -\frac{z}{q} + \mathcal{O}_p(1/\sqrt{n}). \tag{21}$$

Condition (i) is quite harmless since we can always redefine σ^2 to force the condition to hold. Note that (i) implies that $\operatorname{tr} A = 0$. Condition (ii) will also be satisfied in standard applications, but not in unit-root processes (see Section 5). If μ_n denotes the largest eigenvalue (in absolute value) of A, then condition (ii) guarantees the existence of a positive number μ such that $\mu_n \leqslant \mu$ for all n. Then,

$$|\operatorname{tr} AB| \leqslant q\mu, \quad \operatorname{tr}(AB)^2 \leqslant q\mu^2, \quad \operatorname{tr}(AM)^2 \leqslant (n-k)\mu^2, \quad \operatorname{tr} ABAM \leqslant q\mu^2, \tag{22}$$

and, using condition (i), $|\operatorname{tr} AM| = |\operatorname{tr} A(I-M)| \leqslant k\mu$. The leading term in the expansion (21) is a simple sum of quadratic forms all whose moments can be easily calculated.

In the special case $q=1$ (one restriction) we can say more.

Theorem 4. ($q=1$). *Let $q=1$ and let the assumptions of Theorem 3 hold. Then, for large n,*

$$\varphi = -z + \mathcal{O}_p(1/\sqrt{n}), \tag{23}$$

where z satisfies $z = cv$,

$$c^2 = (b'Ab)^2 + 4b'AMAb, \quad b = (w'(X'X)^{-1}w)^{-1/2}X(X'X)^{-1}w, \tag{24}$$

and v follows a product normal distribution pn(r) with parameter $r = b'Ab/c$.

Note: The product normal distribution is defined and briefly discussed in Appendix A.

Theorem 4 is important because the distribution of φ at $\theta = 0$ is intractable, but the distribution of z is known. To assess the sensitivity of the F-test (in fact, t-test since $q=1$) we consider the equation

$$\Pr(|\varphi| > \varphi^*) = \alpha. \tag{25}$$

According to Theorem 4 this is approximately equal to $\Pr(|v| > \varphi^*/c) = \alpha$. We thus obtain an *asymptotic* sensitivity statistic v whose distribution is simple and depends only on one parameter r. We stated — after defining φ in (11) — that large values of φ indicate that $F(\theta)$ is sensitive to small changes in θ when θ is close to 0. We did not discuss what we mean by 'large'. We can now discuss this matter in the context of Theorem 4.

For a given data set we know c and r. Hence, given α, ϕ^* can be obtained from published tables of the product normal distribution; see Appendix A. If $|\varphi| > \varphi^*$, we say that the t-test is sensitive to covariance misspecification; if $|\varphi| \leqslant \varphi^*$ we say it is insensitive or robust. There is, of course, some arbitrariness in the choice of α. The most common choice would be $\alpha = 0.05$ or 0.01, in which case we would (too) frequently conclude that the t-test is robust. In our view the most sensible choice is $\alpha = 0.50$, in which case φ^*/c is the median of $|v|$. As shown in Fig. 1, the median of $|v|$ does not depend much on r. In fact,

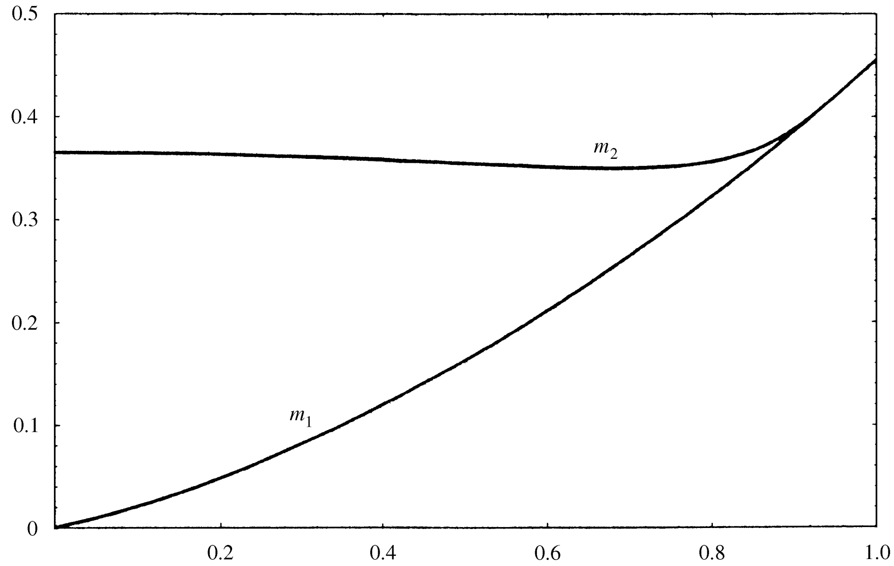

Fig. 1. Median m_1 of v and m_2 of $|v|$ as a function of r.

$0.35 \leqslant \text{median}(|v|) \leqslant 0.45$. Hence, at $\alpha = 0.50$ we obtain the following 'rule of thumb' based on the above asymptotic sensitivity argument.

Rule of thumb: The t-statistic is sensitive (at the 50% level) to covariance misspecification if and only if $|\varphi|/c > 0.40$.

So, in practice, we compute φ from (12) and c from (24) and check whether $|\varphi| > 0.40c$. If we know the type of covariance misspecification which could occur, we use the A-matrix corresponding to this type of misspecification. In most situations we would not know this. Then we use the Toeplitz matrix T defined in Section 5 as our A-matrix. This is the appropriate matrix for AR(1), MA(1) and ARMA(1, 1) misspecification and appears to work well in other situations too; see Banerjee and Magnus (1999). There is evidence (not reported here) that the probability that $|\varphi| > 0.40c$ is extremely close to 0.50. In other words, $0.40c$ is an excellent approximation to the exact (finite sample) median of $|\phi|$.

There is one further interesting consequence of Theorem 4. If $r > 0$ (and this is usually the case), then, using the results of Appendix A,

$$\Pr(v > 0) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{r}{\sqrt{1 - r^2}}\right) > \frac{1}{2} \tag{26}$$

and hence, since $\varphi \approx -cv$ for large n, we have

$$\Pr(\varphi > 0) \approx \Pr(v < 0) < \tfrac{1}{2}. \tag{27}$$

The typical situation is therefore $\varphi < 0$. If this is the case then, using the expansion $F(\theta) \approx F(0) + \theta\varphi$ for $\theta > 0$, we see that $F(\theta) < F(0)$ and hence that the decision to accept the restricted model is a robust one, that is, if we accept (fail to reject) the restriction at $\theta = 0$, there is a high probability that we shall continue to accept the restriction when $\theta \neq 0$. On the other hand, if $\varphi > 0$, then the decision to *reject* the restricted model is robust.

5. AR(1) misspecification: Behaviour near the unit root

When the disturbances are white noise (at $\theta = 0$), we have seen that φ follows a somewhat intractable distribution, which can however be approximated (when $q = 1$) by $-z$ whose distribution is known. We now ask how φ behaves when the disturbances follow a more general stationary process, in particular an AR(1) process with parameter θ. The covariance matrix of the disturbances is proportional to

$$\Omega(\theta) = \begin{bmatrix} 1 & \theta & \dots & \theta^{n-1} \\ \theta & 1 & \dots & \theta^{n-2} \\ \vdots & \vdots & & \vdots \\ \theta^{n-1} & \theta^{n-2} & \dots & 1 \end{bmatrix}. \tag{28}$$

We define the $n \times n$ symmetric Toeplitz matrix T as $T_{ij} = 1$ if $|i-j| = 1$, $T_{ij} = 0$ otherwise, and we notice that, with Ω given in (28), $A = T$. In this case the 'estimators' $\hat{\theta}$ and $\tilde{\theta}$, defined in Theorem 1, take the form

$$\hat{\theta} = \frac{\sum_2^n \hat{u}_t \hat{u}_{t-1}}{\sum_1^n \hat{u}_t^2}, \quad \tilde{\theta} = \frac{\sum_2^n \tilde{u}_t \tilde{u}_{t-1}}{\sum_1^n \tilde{u}_t^2}. \tag{29}$$

We are primarily interested in how φ behaves near the unit root, because the conventional statistics fail at the unit root and also because identification problems can occur (Dufour, 1990). Our analysis leads quite naturally to a distinction between three cases (i denotes the $n \times 1$ vector of ones):

case 1: $Mi = 0$, $Bi \neq 0$ — the regression has an intercept and at least one constraint involves the intercept,

case 2: $Mi = 0$, $Bi = 0$ — the regression has an intercept and none of the constraints involve the intercept, *case 3*: $Mi \neq 0$ — the regression has no intercept.

Suppose $Mi = 0$. Then the regression has an intercept (or at least i lies in the column space of X). Without loss of generality, we may assume that $Xe = i$ where $e' = (1, 0, \ldots, 0)$. Then we find that $Bi = 0 \leftrightarrow i'Bi = 0 \leftrightarrow e'R'(R(X'X)^{-1}R')^{-1}Re = 0 \leftrightarrow Re = 0$. Hence, $Bi = 0$ if and only if the first component of each constraint (row of R) is zero, that is, if and only if none of the constraints involve the intercept.

We now obtain Theorem 5, which gives the distribution of $F(0)$ near the unit root.[1]

Theorem 5. Let $u \sim \mathrm{N}(0, \sigma^2\Omega(\theta))$ *where* $\Omega(\theta)$ *is given in (28). If the restriction* $R\beta = r$ *is satisfied, we can write* $F(0) = (u'Bu/q)/(u'Mu/(n-k))$, *where* M *and* B *are defined in (8) and (9). Then, as* $\theta \to 1$,

$$\Pr(F(0) > \delta) \to \begin{cases} \Pr(F^*(0) > \delta) & \text{if } Mi = 0,\ Bi = 0, \\ 1 & \text{if } Mi = 0,\ Bi \neq 0, \\ 1 & \text{if } Mi \neq 0,\ \bar{F}(0) > \delta, \\ 0 & \text{if } Mi \neq 0,\ \bar{F}(0) \leqslant \delta, \end{cases}$$

where

$$F^*(0) = \frac{\eta'\bar{P}'B\bar{P}\eta}{\eta'\bar{P}'M\bar{P}\eta} \cdot \frac{n-k}{q}, \quad \bar{F}(0) = \frac{i'Bi}{i'Mi} \cdot \frac{n-k}{q},$$

i *is an* $n \times 1$ *vector of ones*, $\bar{P} = JP$, J *is the* $n \times (n-1)$ *matrix such that* $J' = (0 : I_{n-1})$, P *is the lower triangular* $(n-1) \times (n-1)$ *matrix with ones on and below the diagonal and zeroes elsewhere, and* $\eta \sim \mathrm{N}(0, I_{n-1})$.

Theorem 5 provides the size of the F-test when θ is near the unit root. In case 1 the size approaches 1, but in case 3 (no intercept), the size approaches either 0 or 1, while in case 2 it approaches a number between 0 and 1. Theorem 5 also solves a little puzzle raised by Dufour (1990, p. 488). In the case $Mi = 0$, $Bi \neq 0$ (testing for the intercept), Dufour encounters an identification problem at $\theta = 1$. This problem does not occur in Theorem 5, because (in contrast to Dufour) we are working with ratios where the scaling parameter $1/(1 - \theta^2)$ cancels, so that the limit exists.

We next prove Theorem 6, which provides the distribution of φ near the unit root.

[1] A special case of Theorem 5 was considered by Krämer (1989).

Theorem 6. Let $u \sim \mathrm{N}(0, \sigma^2 \Omega(\theta)$ *where* $\Omega(\theta)$ *is given in (28). If the restriction* $R\beta = r$ *is satisfied, we can write*

$$\varphi = 2\left(F(0) + \frac{n-k}{q}\right)(\hat{\theta} - \tilde{\theta}),$$

where

$$\hat{\theta} = \frac{1}{2}\frac{u'MTMu}{u'Mu}, \quad \tilde{\theta} = \frac{1}{2}\frac{u'(M+B)T(M+B)u}{u'(M+B)u},$$

$F(0)$ *is defined in Theorem 5, and* T *is the Toeplitz matrix defined above. Then, as* $\theta \to 1$,

$$\Pr(\varphi > \delta) \to \begin{cases} \Pr(\varphi^* > \delta) & \text{if } Mi = 0,\ Bi = 0, \\ \Pr(\eta' Q \eta > 0) & \text{if } Mi = 0,\ Bi \neq 0, \\ 1 & \text{if } Mi \neq 0,\ \bar{\varphi} > \delta, \\ 0 & \text{if } Mi \neq 0,\ \bar{\varphi} \leqslant \delta, \end{cases}$$

where

$$\varphi^* = 2\left(F^*(0) + \frac{n-k}{q}\right)(\hat{\theta}^* - \tilde{\theta}^*),$$

$$\bar{\varphi} = 2\left(\bar{F}(0) + \frac{n-k}{q}\right)\left(\frac{1}{2}\frac{i'MTMi}{i'Mi} - \frac{1}{2}\frac{i'(M+B)T(M+B)i}{i'(M+B)i}\right),$$

$$\hat{\theta}^* = \frac{1}{2}\frac{\eta'\bar{P}'MTM\bar{P}\eta}{\eta'\bar{P}'M\bar{P}\eta}, \quad \tilde{\theta}^* = \frac{1}{2}\frac{\eta'\bar{P}'(M+B)T(M+B)\bar{P}\eta}{\eta'\bar{P}'(M+B)\bar{P}\eta},$$

$$Q = \bar{P}'M\left(T - \frac{i'BTBi}{i'Bi} I_n\right)M\bar{P}$$

and $F^*(0), \bar{F}(0), i, \bar{P}$, *and* η *are defined in Theorem 5.*

Theorem 6 allows us to calculate the sensitivity of φ when θ approaches one, and also suggests that we should distinguish between cases 1, 2, and 3.

6. Conclusions

In this paper we have studied the sensitivity of the F-statistic (t-statistic) for testing linear restrictions on the coefficients of a linear regression model to covariance misspecification. We have tried to answer two questions. First, is the F-statistic (t-statistic) based on OLS residuals robust? The answer, in general, is no and one should therefore be cautious in using the F-statistic if our 'rule of thumb' indicates sensitivity. Simulation experiments (not reported here) for the AR(1) case with $q = 1$, distinguishing between the three cases discussed in Section 5, lead to the conclusion that the median of z is a good approximation to the median of φ when the distribution of y is evaluated at $\theta = 0$.

Our second question was: is accepting the null hypothesis using the OLS-based F-statistic a robust decision? The answer to this question is, in general, yes. In the case of AR(1) disturbances, Rothenberg (1988) finds that null rejection probabilities are considerably larger than their nominal levels, that is, we reject too often. Our simulation results support Rothenberg's findings. Roughly speaking, we find that the F-statistic is sensitive to covariance misspecification but that nevertheless accepting the null hypothesis is a robust procedure, that is, if the null hypothesis is accepted using the usual F-statistic, it will also be accepted if the disturbances are not white noise.

While the simulations are for the case of AR(1) only, we expect — based on our experience from Banerjee and Magnus (1999) — that the sensitivity ϕ based on AR(1) misspecification (that is, using the Toeplitz matrix T) will perform well under more general stationary covariance misspecifications.

The proposed method generalizes easily to more than one nuisance parameter. If $\theta = (\theta_1, \dots, \theta_m)'$ is a vector of nuisance parameters, we approximate $F(\theta)$ as

$$F(\theta) \approx F(0) + \sum_{i=1}^{m} \theta_i \varphi_i, \quad \varphi_i = \left. \frac{\partial F(\theta)}{\partial \theta_i} \right|_{\theta = 0}$$

and, as discussed in Banerjee and Magnus (1999), we have $F(\theta) \approx F(0)$ if and only if $\varphi_i = 0$ for all i. This allows us to measure the impact of each θ_i on the F-statistic individually.

Appendix A. Some properties of the product of two standardnormal variables

Let x and y be two normally distributed random variables with

$$\mathrm{E}x = \mathrm{E}y = 0, \quad \mathrm{var}(x) = \mathrm{var}(y) = 1$$

and correlation coefficient r. We shall say that the random variable $v = xy$ follows the *product normal* distribution with parameter r, and we write $v \sim pn(r)$. The product normal distribution was first studied by Craig (1936), who established that

$$\mathrm{E}v = r, \quad \mathrm{var}(v) = 1 + r^2.$$

It is easy to see that, when $r = 1, v$ follows a $\chi^2(1)$ distribution and that, when $r = -1, -v$ follows a $\chi^2(1)$ distribution. It is sufficient to consider the case $0 \leqslant r \leqslant 1$, since the density $h(v;r)$ of v possesses the symmetry property $h(-v;-r) = h(v;r)$. The density $h(v;r)$ has a singularity at $v = 0$. Various aspects of the product normal distribution were studied by Aroian et al. (1978), Meeker et al. (1981), and Springer (1983). We can write v alternatively as $v = \lambda u_1^2 - \mu u_2^2$, where u_1 and u_2 are *independent* N(0, 1) variables, $\lambda = (1+r)/2$ and $\mu = (1-r)/2$. Using this representation and the properties of the Cauchy distribution, we can show that

$$\Pr(v > 0) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{r}{\sqrt{1-r^2}}\right),$$

a fact which does not seem to have been noticed by previous authors.

The medians m_1 and m_2 of v and $|v|$, respectively, are graphed in Fig. 1 (Section 4). These are given implicitly by

$$\Pr(v > m_1) = \tfrac{1}{2}, \quad \Pr(|v| > m_2) = \tfrac{1}{2}.$$

Both m_1 and m_2 converge to the same number as $r \to 1$, namely the median 0.4549 of a $\chi^2(1)$ variable. We note that m_2 is not very sensitive to changes in r. In particular, $0.35 \leqslant m_2(r) \leqslant 0.45$.

Appendix B. Proof of theorems

Proof of Theorem 1. Using standard tools of differential calculus (Magnus and Neudecker, 1999), and letting $G = R(X'X)^{-1}R'$, we obtain, evaluated at $\theta = 0$,

$$\mathrm{d}\hat{\beta}(\theta) = -(X'X)^{-1}X'A\hat{u}, \quad \mathrm{d}\tilde{\beta}(\theta) = -((X'X)^{-1} - (X'X)^{-1}R'G^{-1}R(X'X)^{-1})X'A\tilde{u},$$

$$\mathrm{d}\hat{u}(\theta) = X(X'X)^{-1}X'A\hat{u}, \quad \mathrm{d}\tilde{u}(\theta) = X((X'X)^{-1} - (X'X)^{-1}R'G^{-1}R(X'X)^{-1})X'A\tilde{u},$$

$$\hat{u}'\mathrm{d}\hat{u}(\theta) = 0, \qquad \tilde{u}'\mathrm{d}\tilde{u}(\theta) = 0,$$

$$\mathrm{d}\hat{u}'(\theta)\Omega^{-1}(\theta)\hat{u}(\theta) = -\hat{u}'A\hat{u}, \qquad \mathrm{d}\tilde{u}'(\theta)\Omega^{-1}(\theta)\tilde{u}(\theta) = -\tilde{u}'A\tilde{u},$$

where $A = \mathrm{d}\Omega(\theta)$ by (2) and hence $\mathrm{d}\Omega^{-1}(\theta) = -A$, since all differentials are evaluated at $\theta = 0$. (Notice that the above results are true whether or not the constraint $R\beta = r$ is satisfied.) Hence,

$$\begin{aligned}
\mathrm{d}F(\theta) &= \mathrm{d}\frac{\tilde{u}'(\theta)\Omega^{-1}(\theta)\tilde{u}(\theta) - \hat{u}'(\theta)\Omega^{-1}(\theta)\hat{u}(\theta)}{\hat{u}'(\theta)\Omega^{-1}(\theta)\hat{u}(\theta)}\cdot\frac{n-k}{q} \\
&= \frac{\mathrm{d}(\tilde{u}'(\theta)\Omega^{-1}(\theta)\tilde{u}(\theta) - \hat{u}'(\theta)\Omega^{-1}(\theta)\hat{u}(\theta))}{\hat{u}'\hat{u}}\cdot\frac{n-k}{q} \\
&\quad - \frac{\mathrm{d}(\hat{u}'(\theta)\Omega^{-1}(\theta)\hat{u}(\theta))}{\hat{u}'\hat{u}}F(0) \\
&= \frac{\hat{u}'A\hat{u} - \tilde{u}'A\tilde{u}}{\hat{u}'\hat{u}}\cdot\frac{n-k}{q} + \frac{\hat{u}'A\hat{u}}{\hat{u}'\hat{u}}F(0) \\
&= 2\hat{\theta}\frac{n-k}{q} - 2\tilde{\theta}\left(F(0) + \frac{n-k}{q}\right) + 2\hat{\theta}F(0) \\
&= 2\left(F(0) + \frac{n-k}{q}\right)(\hat{\theta} - \tilde{\theta}). \qquad \square
\end{aligned}$$

In order to prove Theorem 2 we need Pitman's Lemma.

Lemma 7 (Pitman, 1937; Laha, 1954). *Let $x_1, x_2, \ldots, x_n$ be identically and independently distributed random variables with a finite second moment. Then $\sum_i a_i x_i / \sum_i x_i$ and $\sum_i x_i$ are independent if and only if each x_i follows a gamma distribution.*

Proof of Theorem 2. At $\theta = 0$ and assuming $R\beta = r$ we can write

$$\varphi = \frac{n-k}{q}\cdot\frac{u'MAMu\cdot u'Bu - u'Mu\cdot u'(BAB + BAM + MAB)u}{(u'Mu)^2},$$

where $u \sim \mathrm{N}(0, I_n)$. Let $M = SS'$, $S'S = I_{n-k}$, $B = TT'$, $T'T = I_q$, so that $S'T = 0$. Define $x = S'u$ and $y = T'u$, so that x and y are independent. Then,

$$\varphi = \frac{n-k}{q}\cdot\frac{x'S'ASx(y'y) - (x'x)y'T'ATy - 2(x'x)x'S'ATy}{(x'x)^2}$$

and hence

$$\mathrm{E}(\phi\,|\,x) = \frac{1}{w}\left(R_1 - \frac{1}{q}\operatorname{tr} AB\right),$$

$$\mathrm{E}(\phi^2\,|\,x) = \frac{1}{q^2w^2}(q(q+2)R_1^2 + 2\operatorname{tr}(AB)^2 + (\operatorname{tr} AB)^2 + 4(n-k)wR_2 - 2(q+2)R_1 \operatorname{tr} AB),$$

where

$$R_1 = \frac{x'S'ASx}{x'x}, \quad R_2 = \frac{x'S'ABASx}{x'x}, \quad w = \frac{x'x}{n-k}.$$

Now, since R_1 and w are independent (Lemma 7) and, similarly, R_2 and w are independent, and using

$$\mathrm{E}\left(\frac{1}{w}\right) = \frac{n-k}{n-k-2}, \quad \mathrm{E}\left(\frac{1}{w^2}\right) = \frac{(n-k)^2}{(n-k-2)(n-k-4)},$$

the results follow. □

Proof of Theorem 3. Using the same notation as in the proof of Theorem 2, we obtain

$$q\varphi + z = (1/w)R_1(y'y) + (1-1/w)y'T'ATy + 2(1-1/w)x's'ATy$$

and hence

$$\mathrm{E}(q\varphi + z)\,|\,x = qR_1/w + (\operatorname{tr} AB)(1-1/w),$$

$$\mathrm{E}(q\varphi + z)^2\,|\,x = q(q+2)R_1^2/w^2 + (2\operatorname{tr}(AB)^2 + (\operatorname{tr} AB)^2)(1-1/w)^2 + 2(q+2)(\operatorname{tr} AB)R_1(1/w - 1/w^2) + 4(n-k)R_2(w - 2 + 1/w).$$

Taking expectations and using the five inequalities in (22) and below, we see that

$$\mathrm{E}(q\varphi + z) = \mathcal{O}(1/n), \quad \operatorname{var}(q\varphi + z) = \mathcal{O}(1/n),$$

and the result follows. □

Proof of Theorem 4. Since φ does not depend on σ^2 we can write $u \sim \mathrm{N}(0, I_n)$ and, noting that $B = bb'$, we find

$$z = \tilde{u}'A\tilde{u} - \hat{u}'A\hat{u} = u'((M + bb')A(M + bb') - MAM)u$$

$$= (b'u)(2b'AM + (b'Ab)b')u = cxy \sim c \cdot pn(r),$$

where

$$x = b'u \sim \mathrm{N}(0, 1),$$

$$y = (1/c)(2b'AM + (b'Ab)b')u \sim \mathrm{N}(0, 1),$$

$$r = \mathrm{E}\, xy = (b'Ab)/c,$$

and $pn(r)$ denotes the product normal distribution discussed in Appendix A. □

In order to prove Theorems 5 and 6 we need the following result, which is a special case of a theorem in Banerjee and Magnus (1999). It is related to earlier results by Sargan and Bhargava (1983) and Krämer (1985).

Lemma 8. Assume that the random variables $u = (u_1, \ldots, u_n)'$ are generated by a stationary AR(1) process, $u_t = \theta u_{t-1} + \varepsilon_t$, where the ε_t are i.i.d. $\mathrm{N}(0, \sigma^2)$. *Let A and B be symmetric positive semidefinite $n \times n$ matrices, let S be a symmetric $n \times n$ matrix, and assume that $B \neq 0$. Then, as $\theta \to 1$,*

$$\frac{u'Au}{u'Bu} \xrightarrow{p} \begin{cases} \dfrac{\eta'\bar{P}'A\bar{P}\eta}{\eta'\bar{P}'B\bar{P}\eta} & \text{if } Bi = 0,\ Ai = 0, \\ +\ \infty & \text{if } Bi = 0,\ Ai \neq 0, \\ \dfrac{i'Ai}{i'Bi} & \text{if } Bi \neq 0 \end{cases}$$

and

$$\frac{u'BSBu}{u'Bu} \xrightarrow{p} \begin{cases} \dfrac{\eta'\bar{P}'BSB\bar{P}\eta}{\eta'\bar{P}'B\bar{P}\eta} & \text{if } Bi = 0, \\ \dfrac{i'BSBi}{i'Bi} & \text{if } Bi \neq 0 \end{cases},$$

where $\bar{P} = JP$, J is the $n \times (n-1)$ matrix such that $J' = (0 : I_{n-1})$, P is the lower triangular $(n-1) \times (n-1)$ matrix with ones on and below the diagonal and zeroes elsewhere, i is an $n \times 1$ vector of ones, $\eta \sim$ $\mathrm{N}(0, I_{n-1})$, *and $\xrightarrow{p}$ indicates convergence in probability.*

Proof of Lemma 8. Both parts of the lemma follow directly from Theorem A1 of Banerjee and Magnus (1999). □

Proof of Theorem 5. We have, using the first part of Lemma 8,

$$F(0) \xrightarrow{p} \begin{cases} F^*(0) & \text{if } Mi = 0,\ Bi = 0, \\ +\ \infty & \text{if } Mi = 0,\ Bi \neq 0, \\ \bar{F}(0) & \text{if } Mi \neq 0, \end{cases}$$

and the result follows. □

Proof of Theorem 6. We have, now using the second part of Lemma 8, as $\theta \to 1$,

$$\hat{\theta} \xrightarrow{p} \begin{cases} \hat{\theta}^* & \text{if } Mi = 0, \\ \dfrac{1}{2}\dfrac{i'MTMi}{i'Mi} & \text{if } Mi \neq 0 \end{cases}$$

and

$$\tilde{\theta} \xrightarrow{p} \begin{cases} \tilde{\theta}^* & \text{if } (M+B)i = 0, \\ \dfrac{1}{2}\dfrac{i'(M+B)T(M+B)i}{i'(M+B)i}, & \text{if } (M+B)i \neq 0. \end{cases}$$

Hence,

$$\hat{\theta} - \tilde{\theta} \xrightarrow{p} \begin{cases} \hat{\theta}^* - \tilde{\theta}^* & \text{if } Mi = 0,\ Bi = 0, \\ \hat{\theta}^* - \dfrac{1}{2}\dfrac{i'BTBi}{i'Bi} & \text{if } Mi = 0,\ Bi \neq 0, \\ \dfrac{1}{2}\dfrac{i'MTMi}{i'Mi} - \dfrac{1}{2}\dfrac{i'(M+B)T(M+B)i}{i'(M+B)i} & \text{if } Mi \neq 0. \end{cases}$$

The result follows easily if $Mi = 0, Bi = 0$ or if $Mi \neq 0$. If $Mi = 0, Bi \neq 0$, let E denote the event $\hat{\theta} > \tilde{\theta}$ and let $\bar{E}$ denote its complement. Then, using the fact that $F(0) \xrightarrow{p} +\ \infty$ (see the proof of Theorem 5) and hence that

$$\Pr(\varphi > \delta \mid E) \to 1 \quad \text{and} \quad \Pr(\varphi > \delta \mid \bar{E}) \to 0,$$

we obtain

$$\begin{aligned} \Pr(\varphi > \delta) &= \Pr(\varphi > \delta, E) + \Pr(\varphi > \delta, \bar{E}) = \Pr(\varphi > \delta \mid E)\Pr(E) \\ &\quad + \Pr(\varphi > \delta \mid \bar{E})\Pr(\bar{E}) \\ &\to \Pr(\hat{\theta}^* > \tfrac{1}{2} i'BTBi/i'Bi) = \Pr(\eta' Q \eta > 0). \end{aligned}$$ □

References

Aroian, L.A., Taneja, V.A., Cornwell, L.W., 1978. Mathematical forms of the distribution of the product of two normal variables. Communications in Statistics, Theory and Methods A7, 165–172.

Banerjee, A.N., Magnus, J.R., 1999. The sensitivity of OLS when the variance matrix is (partially) unknown. Journal of Econometrics, to appear.

Craig, C.C., 1936. On the frequency function XY. The Annals of Mathematical Statistics 7, 1–15.

Dufour, J.M., 1990. Exact tests and confidence sets in linear regressions with autocorrelated errors. Econometrica 58, 475–494.

Durbin, J., Watson, G., 1950. Testing for serial correlation in least squares regression – I. Biometrika 37, 409–428.

Durbin, J., Watson, G., 1951. Testing for serial correlation in least squares regression – II. Biometrika 38, 159–178.

Hillier, G.H., King, M.L., 1987. Linear regression with correlated errors: Bounds on coefficient estimates and t-values. In: King, M.L., Giles, D.E.A. (Eds.). Specification Analysis in the Linear Model, Routledge and Kegan Paul, London, pp. 74–80.

King, M.L., Giles, D.E.A., 1984. Autocorrelation pre-testing in the linear model. Journal of Econometrics 25, 35–48.

Kiviet, J.F., 1980. Effects of ARMA errors on tests for regression coefficients: Comments of Vinod's article: improved and additional results. Journal of the American Statistical Association 75, 353–358.

Krämer, W., 1985. The power of the Durbin–Watson test for regressions without an intercept. Journal of Econometrics 28, 363–370.

Krämer, W., 1989. On the robustness of the F-test to autocorrelation among disturbances. Economics Letters 30, 37–40.

Laha, R.G., 1954. On a characterization of the gamma distribution. The Annals of Mathematical Statistics 25, 784–787.

Magnus, J.R., 1978. Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. Journal of Econometrics 7, 281–312.

Magnus, J.R., 1999. The traditional pretest estimator. Theory of Probability and Its Applications, to appear.

Magnus, J.R., Neudecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics. Revised Edition. Wiley, Chichester/New York.

Meeker, W.Q., Cornwell, L.W., Aroian, L.A., 1981. The product of two normally distributed random variables. In: Selected Tables in Mathematical Statistics, vol. VII. American Mathematical Society, Providence, RI.

Nakamura, A., Nakamura, M., 1978. On the impact of the tests for serial correlation upon the test of significance for the regression coefficient. Journal of Econometrics 7, 199–210.

Park, R.E., Mitchell, B.M., 1980. Estimating the autocorrelated error model with trended data. Journal of Econometrics 13, 185–201.

Pitman, E.J.G., 1937. The "closest" estimates of statistical parameters. Proceedings of the Cambridge Philosophical Society 33, 212–222.

Rothenberg, T.J., 1984a. Hypothesis testing in linear models when the error covariance matrix is nonscalar. Econometrica 52, 827–842.

Rothenberg, T.J., 1984b. Approximate normality of generalized least squares estimates. Econometrica 52, 811–825.

Rothenberg, T.J., 1988. Approximate power functions for some robust tests of regression coefficients. Econometrica 56, 997–1019.

Sargan, J.D., Bhargava, A., 1983. Testing residuals from least squares regression for being generated by the Gaussian random walk. Econometrica 51, 153–174.

Springer, M.D., 1983. Review of Meeker, Cornwell and Aroian (1981). Technometrics 25, 211.

Vinod, H.D., 1976. Effects of ARMA errors on the significance tests for regression coefficients. Journal of the American Statistical Assocation 71, 929–933.

Vinod, H.D., Ullah, A., 1981. Recent Advances in Regression Methods. Marcel Dekker, New York.