MONASH

13/93

MONASH UNIVERSITY

ANCORA IMPARO

AUSTRALIA

HYPOTHESIS TESTING IN THE PRESENCE OF NUISANCE PARAMETERS

Maxwell L. King

Working Paper No. 13/93

October 1993

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0381 1

HYPOTHESIS TESTING IN THE PRESENCE OF NUISANCE PARAMETERS

Maxwell L. King

Working Paper No. 13/93

October 1993

DEPARTMENT OF ECONOMETRICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

HYPOTHESIS TESTING IN THE PRESENCE OF NUISANCE PARAMETERS

Maxwell L. King*
Department of Econometrics
Monash University
Clayton, Victoria 3168
Australia

Abbreviated running title: Testing in the Presence of Nuisance Parameters.

Keywords and phrases: Bayesian methods; first-order autocorrelation; linear regression; marginal likelihood; Monte Carlo simulation; p-values.

AMS classification numbers: 62, 90.

Abstract

How to deal with nuisance parameters is an important problem in econometrics because of the non-experimental nature of economic data. This paper suggests a new approach to dealing with such parameters in the context of hypothesis testing. It involves calculating p-values conditional on values for key nuisance parameters and then taking a weighted average of these values with the weights reflecting the likelihood or posterior probabilities of these values being true. Two specific applications are discussed. These are testing linear regression coefficients in the presence of first-order autoregressive (AR(1)) disturbances and testing for AR(1) disturbances in the dynamic linear regression model. For the former testing problem, a Monte Carlo experiment demonstrates that the new procedure provides more accurate inferences than accepted conventional procedures.

Address for correspondence: as above.

* This research was supported by an ARC grant. I am grateful to Tan Swee Liang, Fong Lai and Alan Morgan for research assistance and to seminar participants at the Universities of Adelaide, Canterbury and Melbourne and Monash University for constructive suggestions.

1. Introduction

The lack of experimental data distinguishes econometrics from its parent discipline of statistics. Non-experimental data adds unwelcome complications to statistical modelling of economic phenomenon. For example, typically a range of extra factors which may not directly influence the phenomenon under study, have to be taken into account by econometric models. When modelling economic time-series, we are faced with the difficult task of correctly modelling the dynamics of the process under study. A consequence of both these complications is that, generally, econometric models contain large numbers of parameters, many of which are not of direct interest. Also, because of the non-experimental nature of economics, there is much less certainty about model specification. It is not possible, for example, to conduct experiments that might verify the existence and form of causal relationships. This increases the need for reliable diagnostic specification tests, particularly in the presence of nuisance parameters.

Typically, statistical theory is good at providing tests for simple models with few or no parametric complications. When nuisance parameters are involved, statistical theory is generally less helpful in suggesting reliable diagnostic tests. The standard approach is to replace nuisance parameters by consistent estimates and then rely on asymptotic theory. For the rather small sample sizes econometricians often have to work with, this approach does not inspire confidence particularly as a multitude of Monte Carlo studies have provided cause for concern.[1]

One-sided testing of a regression coefficient provides a simple example of the damaging effect a nuisance parameter can have. In the classical linear regression model with independent and identically distributed normal disturbances, the familiar one-sided t-test is an exact uniformly most powerful invariant (UMPI) test. When, however, the disturbances follow a stationary Gaussian first-order autoregressive (AR(1)) process, a number of authors have found the standard asymptotic tests to be unreliable. For example, Nakamura and Nakamura (1978) reported actual sizes of up to 0.35 and 0.28 for sample sizes of 30 and 50 respectively for the asymptotic test based on the Cochrane-Orcutt iterative estimator for AR(1) errors and a five per cent nominal level of significance. King and Giles (1984) found that the form of the regressors has considerable effect on the true size of the asymptotic t-tests based on Durbin's (1960) estimator or the maximum likelihood estimator (MLE). For one regressor set with 60 observations and for a five per cent nominal significance level, they reported actual sizes ranging up to 0.26 and 0.21 for tests based on Durbin's estimator and the MLE, respectively. Other authors who have reported similar findings include Park and Michell (1980), and Griffiths and Beesley (1984).

While the non-experimental nature of econometric data is a negative complication, recent advances in computer hardware and software have meant that vast computations can now be undertaken quickly and cheaply. The future would appear to be such that we can begin to ask: what kind of testing procedure would we wish to use if computation time were not a constraint?

In this paper, a new approach to testing in the presence of nuisance parameters is proposed. It involves first identifying key nuisance parameters - those which if we knew their value we could apply a standard exact test that is possibly optimal. Then a p-value is calculated as an average of p-values from the standard test, conditional on the key nuisance parameters' values.

The plan of the paper is as follows. The theory of the new test procedure is outlined in section 2. Details of its application to (i) testing regression coefficients in the presence of AR(1) disturbances and (ii) testing for autocorrelation in the dynamic linear regression model are given as examples. Section 3 reports a Monte Carlo experiment that investigates the small-sample properties of the proposed test for example (i) and compares these properties with those of existing tests. Some concluding remarks are made in the final section.

2. Theory

Let y be an observable n×1 vector which is known to have probability density

$$f(y;\ \omega,\phi,\theta)$$

where ω, ϕ and θ are q×1, r×1 and s×1 vectors of unknown parameters. Suppose we wish to test

$$H_0 : \theta = \theta_0$$

against

$$H_a : \theta > \theta_0$$

where θ_0 is a known r×1 vector.[2] Then ω and ϕ are vectors of nuisance parameters. They are written as two vectors with ω being the vector of key nuisance parameters in the sense that if we knew their value, we would be able to apply an exact, possibly optimal, test. The aim is to choose ω such that q is as small as possible. Hence, conditional on $\omega = \omega_1$, an exact p-value can be calculated. The p-value can be interpreted as the probability of Evidence as Extreme or More Extreme as that in the

Data (EEMED) assuming H_0 is true and $\omega = \omega_1$, i.e., we can calculate

$$\Pr(\text{EEMED} \mid \omega = \omega_1, y, H_0) \ . \tag{1}$$

In theory, we can calculate this p-value for any value of ω_1.

We could examine these p-values and if they are all below α, the desired level of significance, we could reject H_0 at the α significance level. What should we do if some are above α and some below α?

There is information in the data. The data can tell us that some values of ω are more likely than others. Also, as the sample size increases, the data provides more accurate information concerning likely ω values. An obvious approach is to take a Bayesian view. Using non-informative priors, the marginal posterior distribution for ω provides a distribution of possible ω values based on the information in the data. An alternative is to take a likelihood approach. Given the nature of the key nuisance parameters, the concentrated or profile likelihood for ω seems worthy of consideration. Kalbfleisch and Sprott (1970) criticised its use on the grounds that it assumes the parameter values concentrated out are known to be equal to their maximum likelihood estimates. No account is taken of the uncertainty due to these parameter values being estimates. Kalbfleisch and Sprott advocated the use of the marginal likelihood function.[3] Under H_0, its construction involves finding a transformation of y for which the marginal distribution of a subset of the transformed variables is independent of ϕ. The marginal likelihood indicates the relative likelihood of different values of ω occurring. It is also typically proportional to the marginal posterior distribution for non-informative priors.

In the case of the general linear model

$$y = X(\mu)\beta + u , \qquad u \sim N(0,\sigma^2\Omega) \tag{2}$$

where y and u are n×1, $X(\mu)$ is an n×k matrix dependent on a parameter vector μ, β is a k×1 parameter vector and Ω is an n×n matrix, Bellhouse (1978) has shown that the marginal likelihood for μ and Ω is proportional to

$$L_m(\mu,\Omega|y) = \frac{\left|X'(\mu)X(\mu)\right|^{1/2}\left\{y'y-y'X(\mu)\left(X'(\mu)X(\mu)\right)^{-1}X'(\mu)y\right\}}{\left|\Omega\right|^{1/2}\left|X'(\mu)\Omega^{-1}X(\mu)\right|^{1/2} s^{n-k}} \tag{3}$$

where

$$s^2 = \left\{y'\Omega^{-1}y - y'\Omega^{-1}X(\mu)\left(X'(\mu)\Omega^{-1}X(\mu)\right)^{-1}X'(\mu)\Omega^{-1}y\right\}$$

$$= \hat{u}'(\mu,\Omega)\Omega^{-1}\hat{u}(\mu,\Omega)$$

in which $\hat{u}(\mu,\Omega)$ is the generalized least squares (GLS) residual vector from (2) assuming covariance matrix Ω and given μ. Model (2) covers a range of linear models used in econometrics including the linear regression model, the dynamic regression model and the reduced form of a simultaneous equation model.

For our problem of testing H_0 against H_a, let $L_m(\omega|y)$ denote the marginal likelihood (up to a constant of proportionality). We can turn it into a density function by taking account of the constant of integration,

$$c = \int_R L_m(\omega|y)d\omega , \tag{4}$$

where R is the range of ω. Then define

$$f(\omega|y) = L_m(\omega|y)/c$$

so that

$$\Pr(\text{EEMED}|y \text{ and } H_0)$$

$$= \int_R \Pr(\text{EEMED}|\omega = \omega_1, y, H_0) f(\omega|y) d\omega \ . \qquad (5)$$

Provided $L_m(\omega|y)$ is known and (1) can be computed, (4) and (5) can be calculated using standard numerical integration algorithms such as those in IMSL. (5) can be used as a p-value for a test of H_0 and is essentially a weighted average of p-values conditional on different values of ω_1 with the weights determined by the likelihood of different ω_1 values as indicated by the data.

Although the general linear model (2) covers a wide range of models of practical interest, it should be noted that marginal likelihoods cannot be constructed in all situations. Alternatives that might be used to replace $L_m(\omega|y)$ in (4) and (5) in such cases include modified profile likelihoods (see Barndorff-Nielsen (1986)) and conditional profile likelihoods (see Cox and Reid (1987)). We shall now illustrate the new approach by considering its application to two important econometric testing problems.

2.1 Testing regression coefficients in the presence of AR(1) disturbances

Consider the linear regression model

$$y_t = x_t'\beta + u_t \ , \qquad t = 1,\ldots,n \ , \qquad (6)$$

where y_t is the dependent variable at time t, x_t is a k×1 vector of nonstochastic regressors at time t, β is a k×1 vector of coefficients and u_t is a disturbance term generated by the stationary AR(1) process

$$u_t = \rho u_{t-1} + e_t , \quad |\rho| < 1 , \qquad t = 1,\ldots,n , \tag{7}$$

where $e = (e_1,\ldots,e_n)' \sim N(0,\sigma^2 I_n)$. Suppose we wish to test

$$H_0^1 : \beta_k = \beta_k^0$$

against

$$H_a^1 : \beta_k > \beta_k^0 .$$

where β_k^0 is a known fixed value. If ρ is known to be ρ_1, say, then we can transform (6) to

$$\begin{aligned} (1 - \rho_1^2)^{1/2} y_1 &= (1 - \rho_1^2)^{1/2} x_1'\beta + e_1 , \\ y_t - \rho_1 y_{t-1} &= (x_t - \rho_1 x_{t-1})'\beta + e_t, \quad t = 2,\ldots, n , \end{aligned} \tag{8}$$

and then apply the standard one-sided t test of β_k. The resultant test is UMPI and exact p-values can be calculated from the Student's t distribution with n-k degrees of freedom.

Thus for this problem, ρ is a key nuisance parameter. Hence $\omega = \rho$, $\phi = (\beta_1,\ldots,\beta_{k-1},\sigma^2)'$ and $\theta = \beta_k$. Let X be the n×k matrix of regressors, and let X* be X with the last column (the regressor corresponding to β_k) deleted. Then from (3), the marginal distribution of ρ under H_0^1 is proportional to

$$L_m(\rho|y) = \left|\Sigma(\rho)\right|^{-1/2} \left|X^{*\prime}\Sigma^{-1}(\rho)X^*\right|^{-1/2} s^{-(n-k+1)} \tag{9}$$

where

$$\sigma^2\Sigma(\rho) = \sigma^2/(1-\rho^2)\begin{bmatrix} 1 & \rho & \rho^2 & . & . & . & \rho^{n-1} \\ \rho & 1 & \rho & & & & . \\ \rho^2 & \rho & 1 & & & & . \\ . & & & . & & & . \\ . & & & & . & & . \\ . & & & & & . & . \\ . & & & & & 1 & \rho \\ \rho^{n-1} & . & . & . & . & \rho & 1 \end{bmatrix}$$

is the covariance matrix of u, and

$$s^2 = \left\{y'\Sigma^{-1}(\rho)y-y'\Sigma^{-1}(\rho)X^*\left(X^{*\prime}\Sigma^{-1}(\rho)X^*\right)^{-1}X^{*\prime}\Sigma^{-1}(\rho)y\right\}.$$

The terms in the numerator of (3) have been omitted along with other constants that do not involve ρ. Note that the marginal likelihood of ρ under H_a^1 is proportional to (9) with X^* replaced by X and $s^{-(n-k+1)}$ replaced by $s^{-(n-k)}$.

In theory, the range of ρ values is the open interval (-1,1). For practical purposes such as calculating (4) and (5), it must be approximated by a closed interval. In the calculations reported in Section 3, [-0.999,0.999] was used so that the constant of integration is

$$c = \int_{-0.999}^{0.999} L_m(\rho|y)d\rho$$

and $f(\rho|y) = L_m(\rho|y)/c$. If $t^*(\rho_1,y)$ denotes the calculated value of the standard t-statistic for β_k from the transformed regression (8) and $\Phi(\cdot)$ denotes the cumulative distribution of the Student's t distribution with n-k degrees of freedom, then the proposed test is based on the p-value calculated as

$$p(y) = \int_{-0.999}^{0.999} \left(1 - \Phi(t^*(\rho_1, y))\right) L_m(\rho_1|y)/c \, d\rho_1 . \qquad (10)$$

At the α level of significance, H_0 is rejected if $p(y) < \alpha$.

2.2 Testing for autocorrelation in the dynamic linear regression model

Consider the dynamic linear regression model

$$y_t = \gamma y_{t-1} + x_t'\beta + u_t , \qquad t = 2,\ldots,n , \qquad (11)$$

where x_t and β are as defined in (6) and the disturbance term, u_t, follows the stationary AR(1) process (7). Suppose we wish to test

$$H_0^2 : \rho = 0$$

against

$$H_a^2 : \rho > 0 .$$

Clearly, if the value of γ was known to be γ_1 say, then we could write (11) as

$$y_t - \gamma_1 y_{t-1} = x_t'\beta + u_t , \qquad t = 2,\ldots, n , \qquad (12)$$

and our testing problem becomes one of testing H_0^2 against H_a^2 in the context of the static regression model (6) with $y_t - \gamma_1 y_{t-1}$ replacing y_t. The standard test in this situation would be to apply the Durbin-Watson test which is approximately locally best invariant. Obviously γ is a key nuisance parameter so that $\omega = \gamma$, $\phi = (\beta', \sigma^2)'$ and $\theta = \rho$.

It is assumed that data are available on y_t and x_t for n observations, $t = 1,2,\ldots,n$. In order to derive the marginal likelihood function of γ, it is necessary to make further assumptions about y_1 and

u_1 so that the distribution of y can be determined. We adopt the approach used by Tse (1981), Nankervis and Savin (1985) and Inder (1985, 1986). This involves the following two assumptions:

(i) The mean of y_t is stable at t = 1 so that $E(y_1) = E(y_0)$

(ii) The variance of y_t is the same for all t = 1,...,n.

Inder (1985) shows that this is observationally equivalent to assuming y_1 is generated by

$$y_1 = x_1'\beta / (1-\gamma) + d_1 u_1 \tag{13}$$

where $d_1^2 = (1+\gamma\rho) / \{(1-\gamma\rho)(1-\gamma^2)\}$. Thus assuming $|\gamma| < 1$, (11) and (13) can be written as

$$\Gamma y = X\beta + Du \tag{14}$$

where Γ is the n×n matrix

$$\Gamma = \begin{bmatrix} (1-\gamma) & 0 & 0 & . & . & . & 0 \\ -\gamma & 1 & 0 & & & & . \\ 0 & -\gamma & 1 & & & & . \\ . & & & \ddots & & & . \\ . & & & & \ddots & & . \\ . & & & & & 1 & 0 \\ 0 & & & & & -\gamma & 1 \end{bmatrix},$$

$y = (y_1,\ldots,y_n)'$, X is the n×k matrix whose t^{th} row is x_t' and D is the n×n diagonal matrix

$$D = \mathrm{diag}(d_1(1-\gamma),1,1,\ldots,1) .$$

Equation (14) implies that

$$y = \Gamma^{-1}X\beta + \Gamma^{-1}Du$$

which under H_0^2 is of the form of (2) with

$$X(\mu) = \Gamma^{-1}X \tag{15}$$

and

$$\Omega = \Gamma^{-1}D^2\Gamma^{-1\prime}. \tag{16}$$

The marginal likelihood of γ under H_0^2, denoted $L_m(\gamma|y)$, therefore can be computed using (3), (15) and (16) with $d_1 = (1-\gamma^2)^{-1/2}$.

Assuming γ_1 is known, the Durbin-Watson statistic from (12) is

$$d(\gamma_1) = \sum_{t=3}^{n}\left(z_t - z_{t-1}\right)^2 \Big/ \sum_{t=2}^{n} z_t^2$$

where $z = (z_2,\ldots,z_n)'$ is the ordinary least squares (OLS) residual vector from (12). Its p-value can be found by computing

$$p^*(\gamma_1,y) = \Pr\left[\sum_{i=1}^{n-k-1}(\nu_i - d(\gamma_1))\xi_i^2 < 0\right]$$

using standard algorithms (see for example, King (1987)), where $\nu_1,\ldots,\nu_{n-k-1}$ are the eigenvalues of $(I - \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}')A_1$ excluding k zeros, $\bar{X}$ is X with its first row omitted, A_1 is the $(n-1)\times(n-1)$ matrix

$$A_1 = \begin{bmatrix} 1 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & & & & \cdot \\ 0 & -1 & 2 & & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & & 2 & -1 \\ 0 & & & & & -1 & 1 \end{bmatrix}$$

and $\xi_i \sim IN(0,1)$, $i = 1,\dots,n-k-1$. Observe that $z = z^* - \gamma_1 z^\dagger$ where z^* and $z^\dagger$ are $(n-1)\times 1$ OLS residual vectors from the regressions of y_t on x_t and y_{t-1} on x_t, $t = 2,\dots,n$, respectively.

Typically, one might wish to assume γ is restricted to the open interval (0,1) which can be approximated by [0,0.999] say. Then the constant of integration of $L_m(\gamma|y)$ can be calculated numerically as

$$c = \int_0^{0.999} L_m(\gamma|y)d\gamma .$$

The proposed test of H_0^2 against H_a^2 is based on calculating

$$\bar{p}(y) = \int_0^{0.999} p^*(\gamma_1,y)L_m(\gamma_1|y) / c \; d\gamma_1$$

and interpreting $\bar{p}(y)$ as a p-value.

3. Monte Carlo Experiment

A Monte Carlo experiment was conducted in order to investigate the small-sample properties of the proposed procedure in the case of testing regression coefficients in the presence of AR(1) disturbances. The sizes and powers of five test procedures were calculated for a range of X matrices in the context of (6) and (7). The following test procedures were studied:

(i) *The OLS test.* The OLS-based t test; i.e., ignoring the presence of autocorrelation.

(ii) *The p(y) test.* The recommended procedure based on using p(y) from (10) as a p-value. The marginal likelihood used is that under the alternative hypothesis; i.e. (9) with X* replaced by

X and k-1 replaced by k.

(iii) *The Durbin procedure.* ρ in (7) is estimated using the modification of Durbin's (1960) estimator outlined by King and Giles (1984). Then a t test is applied to (6) after transformation to (8) using the estimated ρ value as ρ_1.

(iv) *The ML test.* As for the Durbin procedure except the MLE of ρ based on Beach and MacKinnon's (1978) algorithm is used as ρ_1 in (8).

(v) *The R test.* The t test based on Wooldridge's (1989) computationally simple heteroskedasticity and serial correlation robust standard errors.

The following X matrices, each with sample sizes of n = 20 and 60, were used in the study:

X1: The regressors are a constant dummy, the Australian quarterly consumers' price index commencing 1959(1) and the same index lagged one quarter (k = 3).

X2: X1 augmented by three quarterly seasonal dummy variables.

X3: The regressors are a constant dummy, and the quarterly seasonally adjusted Australian household disposable income and private consumption expenditure series commencing 1967(2) (k = 3).

X4: X3 augmented by the income and consumption series lagged one quarter.

X5: The regressors are the eigenvectors corresponding to the k = 3 smallest eigenvalues of the n×n A_1 matrix.

X6: The regressors are a_1, $(a_2+a_n)/\sqrt{2},\ldots,(a_k+a_{n-k+2})/\sqrt{2}$, where $a_1,\ldots,a_n$ are the eigenvectors corresponding to the eigenvalues of the n×n A_1 matrix ranked in ascending order and k = 3.

X7: X6 with k = 5.

These X matrices were used by King and Giles (1984) in their study because they reflect a variety of economic and statistical phenomena. If $\Sigma(\rho)$ is approximated by $\left[(1-\rho)^2 I_n + \rho A_1\right]^{-1}$ then the OLS and GLS estimators are identical for X5 and Watson (1955) has shown that within the class of orthogonal X matrices with an intercept, OLS has minimum efficiency relative to GLS for X6 and X7.

Our choice of the marginal likelihood under the alternative hypothesis for use in the p(y) procedure deserves comment. While it may be expected to result in less accurate sizes, it is more likely to result in near optimal power in the following sense. As the sample size increases, the marginal likelihood becomes more concentrated around the true value of ρ, thus increasing the weights given to p-values based on the true or near-true values of ρ. In other words the test becomes more like the UMPI test based on the true value of ρ. Some preliminary experiments confirmed that our choice of marginal likelihood works better than does the H_0^1 marginal likelihood (9).

The tests were conducted on the coefficient of the first non-constant regressor at nominal levels of 10 and 5 per cent. Two thousand replications were used in the study with the random disturbances $\{e_t\}$, generated as outlined by King and Giles. Probabilities of rejection were estimated at ρ = 0.0, 0.3, 0.6, 0.9. Where required, numerical integration was performed by the IMSL subroutine DCADRE.

A selection of estimated sizes are presented in Tables 1 and 2 while corresponding estimated powers are given in Tables 3 and 4.

The proposed p(y) procedure is by far the most accurate of the five tests in terms of having estimated sizes close to the nominal significance levels. If confidence intervals are constructed around the estimated sizes, the p(y) procedure is the only test whose 99 per cent confidence intervals include the nominal size more than half the time. In fact they bracket the nominal significance level in more than 80 per cent of cases. The next best is the test based on Durbin's procedure which is reasonably accurate for the larger sample size and is clearly superior to the maximum likelihood based test. Wooldridge's simple robust procedure is most disappointing. It is very hard to argue based on the results of this study that the R test is better than wrongly applying the OLS-based test. All tests have poor sizes for the artifically generated X5 matrix although the p(y) procedure is best at controlling the probability of a Type I error in this case.

Because the tests have different sizes, it is difficult to make valid comparisons of powers. In most cases, the ordering of powers reflects the ordering of the corresponding sizes. Only when one test has both lower size and higher power than another test can a definite conclusion be drawn. There is some evidence to suggest that when $\rho = 0$, the OLS test has a slight power advantage over the other tests while when $\rho = 0.9$, the p(y) procedure has a clear power advantage over the OLS and R tests and is also slightly more powerful than those based on the ML and modified Durbin procedures.

Overall, the results of the Monte Carlo experiment are very encouraging. They certainly suggest that at least for the cases considered in this study, the new procedure provides more accurate inferences than accepted conventional procedures.

4. Concluding Remarks

How to successfully deal with nuisance parameters is an important problem that econometricians are forced to face because of the non-experimental nature of their data. This paper suggests a new approach to dealing with such parameters in the context of hypothesis testing. It involves calculating p-values conditional on values for key nuisance parameters and then essentially taking a weighted average of these values with weights reflecting the marginal likelihood or posterior probabilities of these values being true. In the case of testing a linear regression coefficient in the presence of autocorrelation, our Monte Carlo experiment demonstrates that the new procedure provides more accurate inferences than accepted conventional procedures.

The idea of taking weighted averages of inferences conditional on key nuisance parameters has also been found to work well for estimation and forecasting in the context of the linear regression model with AR(1) disturbances. Based on Bayesian arguments, Kennedy and Simons (1991) proposed a weighted average of GLS estimates of β in (6) and (7) conditional on ρ with weights determined by the marginal posterior distribution of ρ. In a simulation study, they found that this estimator had a much smaller mean squared error than conventional empirical GLS estimators of β. Latif and King (1993) suggested the use of forecasts from (6) and (7) constructed as weighted averages of forecasts conditional on ρ with weights proportional to the marginal likelihood of ρ. They reported a Monte Carlo experiment that showed this procedure has a distinct edge over conventional procedures.

It does appear, therefore, that at least in the context of the linear regression model, weighted averages of inferences with weights determined by marginal likelihoods or posterior distributions have a lot to commend them.

REFERENCES

Barndorff-Nielsen, O. (1986). Inference on full or partial parameters based on the standardized log likelihood ratio. *Biometrika* 73, 307-322.

Beach, C.M. and J.G. MacKinnon (1978). A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica* 46, 51-58.

Bellhouse, D,.R. (1978). Marginal likelihood methods for distributed lag models. *Statist. Hefte* 19, 2-14.

Bewley, R. (1986). *Allocation Models: Specification, Estimation and Applications*, Ballinger, Cambridge, Massachusetts.

Chesher, A. and G. Austin (1991). The finite sample distributions of heteroskedasticity robust Wald statistics. *J. Econometrics* 47, 153-173.

Cooper, D.M. and R. Thompson (1977). A note on the estimation of parameters of the autoregressive-moving average process. *Biometrika* 64, 625-628.

Cox, D.R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Royal Statist. Soc.* B, 49, 1-39.

Durbin, J. (1960). Estimation of parameters in time series regression models. *J. Royal Statist. Soc. B* 22, 139-153.

Griffiths, W.E. and P.A.A. Beesley (1984). The small sample properties of some preliminary-test estimators in a linear model with auto-correlated errors. *J. Econometrics* 25, 49-62.

Inder, B.A. (1985). Testing for first order autoregressive disturbances in the dynamic linear regression model. Unpublished Ph.D. thesis, Monash University.

Inder, B.A. (1986). An approximation to the null distribution of the Durbin-Watson statistic in models containing lagged dependent variables. *Econometric Theory* 2, 413-428.

Kalbfleisch, J.D. and D.A. Sprott (1970). Application of likelihood methods to models involving large numbers of parameters. *J. Royal Statist. Soc. B* 32, 175-194.

Kennedy, P. and D. Simons (1991). Fighting the teflon factor: Comparing classical and Bayesian estimators for autocorrelated errors. *J. Econometrics* 48, 15-27.

King, M.L. (1987). Testing for autocorrelation in linear regression models: A survey. In: M.L. King and D.E.A. Giles, Eds., *Specification Analysis in the Linear Model*, Routledge and Kegan Paul, London, 19-73.

King, M.L. and D.E.A. Giles (1984). Autocorrelation pre-testing in the linear model: Estimation, testing and prediction. *J. Econometrics* 25, 35-48.

King, M.L. and M. McAleer (1987). Further results on testing AR(1) against MA(1) disturbances in the linear regression model. *Rev. Econ. Studies* 54, 649-663.

Latif, A. and M.L. King (1993). Linear regression forecasting in the presence of AR(1) disturbances. *J. Forecasting* forthcoming.

Levenbach, H. (1972). Estimation of autoregressive parameters from a marginal likelihood function. *Biometrika* 59, 61-71.

Levenbach, H. (1973). Estimating heteroscedasticity from a marginal likelihood function. *J. Amer. Statist. Assoc.* 68, 436-439.

Moulton, B.R. and W.C. Randolph (1989). Alternative tests of the error components model. *Econometrica* 57, 685-693.

Nakamura, A. and M. Nakamura (1978). On the impact of tests for serial correlation upon the test of significance for the regression coefficient. *J. Econometrics* 7, 199-210.

Nankervis, J.C. and N.E. Savin (1985). Testing the autoregressive parameter with the t statistic. *J. Econometrics* 27, 143-162.

Park, R.E and B.M. Mitchell (1980). Estimating the autocorrelated error model with trended data. *J. Econometrics* 13, 185-201.

Patterson, H.D. and R. Thompson (1975). Maximum likelihood estimation of components of variance. In: L.C.A. Corsten and T. Postelnicu, Eds., *Proceedings of the 8th International Biometric Conference*, Academy of the Socialist Republic of Rumania, Bucharest, 197-208.

Tse, Y.K (1981). Edgeworth Approximation to the Finite Sample Distribution of Econometric Estimators and Test Statistics. Unpublished Ph.D. thesis, University of London.

Tunnicliffe Wilson, G. (1989). On the use of marginal likelihood in time series model estimation. *J. Royal Statist. Soc. B* 51, 15-27.

Watson, G.S. (1955). Serial correlation in regression analysis I. *Biometrika* 42, 327-341.

Wooldridge, J.M. (1989). A computationally simple heteroskedasticity and serial correlation robust standard error for the linear regression model. *Econ. Letters* 31, 239-243.

Table 1: Estimated sizes of one-sided tests of H_0: $\beta_2 = 1$ when n = 20.

ρ	10% nominal level OLS	p(y)	Durbin	ML	R	5% nominal level OLS	p(y)	Durbin	ML	R
					X1					
0.0	.100	.100	.130	.127	.191	.052	.054	.070	.072	.126
0.3	.146	.106	.136	.140	.222	.088	.054	.082	.083	.162
0.6	.198	.114	.144	.132	.250	.130	.063	.082	.080	.200
0.9	.221	.104	.132	.094	.276	.166	.052	.074	.052	.220
					X2					
0.0	.094	.094	.124	.119	.203	.050	.048	.068	.066	.136
0.3	.146	.101	.132	.136	.234	.086	.053	.076	.078	.174
0.6	.197	.112	.139	.128	.277	.138	.059	.086	.082	.216
0.9	.226	.112	.134	.091	.290	.168	.057	.076	.060	.239
					X3					
0.0	.104	.116	.127	.117	.150	.052	.060	.076	.069	.092
0.3	.104	.114	.124	.116	.154	.052	.058	.065	.062	.097
0.6	.097	.108	.110	.088	.153	.046	.054	.055	.034	.099
0.9	.072	.089	.084	.051	.138	.031	.042	.044	.020	.080
					X4					
0.0	.108	.115	.129	.122	.154	.055	.060	.071	.062	.100
0.3	.098	.109	.115	.110	.156	.044	.058	.060	.056	.096
0.6	.084	.104	.100	.092	.144	.036	.052	.053	.038	.082
0.9	.051	.094	.091	.061	.115	.024	.048	.046	.022	.060
					X5					
0.0	.106	.065	.148	.130	.184	.049	.028	.083	.075	.119
0.3	.180	.064	.184	.164	.225	.116	.030	.118	.106	.163
0.6	.278	.100	.245	.216	.294	.220	.043	.182	.156	.239
0.9	.393	.225	.344	.324	.390	.356	.158	.309	.282	.354
					X6					
0.0	.102	.106	.110	.134	.157	.049	.054	.056	.082	.096
0.3	.123	.106	.116	.132	.180	.062	.050	.062	.077	.116
0.6	.196	.102	.135	.109	.248	.127	.052	.081	.062	.182
0.9	.316	.098	.184	.090	.338	.266	.048	.114	.056	.288
					X7					
0.0	.102	.101	.121	.162	.174	.050	.053	.065	.102	.120
0.3	.126	.104	.133	.169	.204	.064	.052	.074	.114	.136
0.6	.200	.105	.166	.164	.262	.134	.057	.113	.121	.198
0.9	.320	.103	.252	.190	.352	.268	.054	.186	.158	.307

Table 2: Estimated sizes of one-sided tests of H_0: $\beta_2 = 1$ when n = 60.

	10% nominal level					5% nominal level				
ρ	OLS	p(y)	Durbin	ML	R	OLS	p(y)	Durbin	ML	R
					X1					
0.0	.099	.102	.114	.108	.130	.050	.050	.056	.058	.084
0.3	.148	.102	.110	.106	.154	.086	.052	.056	.053	.111
0.6	.178	.098	.108	.074	.178	.121	.051	.055	.032	.124
0.9	.256	.100	.100	.024	.242	.198	.049	.050	.008	.175
					X2					
0.0	.098	.101	.108	.107	.134	.049	.052	.058	.058	.084
0.3	.147	.092	.105	.099	.163	.084	.052	.054	.053	.109
0.6	.176	.097	.103	.069	.183	.119	.052	.052	.035	.126
0.9	.258	.098	.100	.012	.244	.200	.048	.052	.010	.171
					X3					
0.0	.092	.091	.104	.101	.120	.051	.047	.060	.060	.076
0.3	.152	.088	.104	.099	.150	.094	.048	.063	.056	.102
0.6	.232	.088	.112	.082	.204	.175	.046	.066	.046	.146
0.9	.351	.095	.125	.040	.325	.320	.048	.072	.018	.285
					X4					
0.0	.094	.091	.094	.096	.112	.043	.044	.057	.046	.064
0.3	.097	.090	.095	.088	.122	.047	.040	.054	.038	.070
0.6	.123	.083	.094	.064	.141	.056	.041	.057	.026	.080
0.9	.178	.094	.115	.032	.205	.116	.044	.070	.014	.142
					X5					
0.0	.092	.084	.108	.101	.115	.042	.040	.054	.051	.065
0.3	.174	.082	.126	.111	.156	.108	.040	.066	.060	.098
0.6	.270	.074	.160	.115	.220	.210	.036	.098	.066	.156
0.9	.388	.138	.286	.184	.354	.362	.068	.241	.136	.321
					X6					
0.0	.078	.083	.090	.094	.092	.042	.042	.044	.050	.050
0.3	.102	.086	.088	.084	.116	.050	.040	.046	.040	.064
0.6	.188	.092	.095	.058	.184	.126	.046	.045	.024	.120
0.9	.340	.103	.108	.017	.332	.308	.056	.057	.004	.294
					X7					
0.0	.078	.086	.084	.098	.102	.040	.044	.046	.054	.054
0.3	.104	.086	.088	.091	.123	.052	.041	.044	.040	.068
0.6	.194	.094	.097	.062	.195	.130	.044	.050	.026	.134
0.9	.346	.108	.146	.022	.338	.314	.054	.085	.006	.305

Table 3: Estimated powers of one-sided tests of H_0: $\beta_2 = 1$ when n = 20.

	10% nominal level					5% nominal level				
ρ	OLS	p(y)	Durbin	ML	R	OLS	p(y)	Durbin	ML	R
					X1 β_2/σ = 100.0					
0.0	.638	.604	.686	.680	.774	.486	.457	.570	.548	.678
0.3	.627	.566	.624	.606	.712	.502	.418	.496	.482	.627
0.6	.580	.545	.578	.536	.644	.484	.403	.457	.408	.575
0.9	.540	.556	.574	.459	.594	.455	.417	.440	.316	.522
					X2 β_2/σ = 100.0					
0.0	.600	.564	.644	.630	.762	.434	.397	.507	489	.658
0.3	.578	.499	.566	.544	.697	.458	.345	.432	.416	.610
0.6	.539	.466	.503	.457	.624	.443	.338	.391	.334	.549
0.9	.506	.468	.490	.360	.578	.416	.326	.358	.246	.520
					X3 β_2/σ = 15.0					
0.0	.660	.659	.674	.658	.720	.513	.522	.544	.526	.601
0.3	.655	.681	.684	.669	.726	.506	.545	.546	.532	.608
0.6	.624	.753	.782	.682	.714	.467	.593	.579	.522	.596
0.9	.568	.823	.800	.686	.682	.402	.673	.656	.496	.563
					X4 β_2/σ = 15.0					
0.0	.608	.606	.623	.608	.683	.457	.468	.478	.456	.572
0.3	.620	.640	.640	.622	.704	.458	.493	.496	.476	.590
0.6	.584	.684	.676	.630	.698	.434	.534	.528	.473	.566
0.9	.539	.754	.741	.622	.670	.358	.593	.586	.461	.542
					X5 β_2/σ = 2.0					
0.0	.972	.876	.978	.976	.986	.926	.682	.947	.938	.968
0.3	.918	.673	.904	.892	.938	.860	.430	.837	.814	.890
0.6	.806	.522	.758	.732	.808	.750	.314	.692	.652	.758
0.9	.661	.484	.618	.590	.660	.630	.356	.566	.532	.626
					X6 β_2/σ = 1.6					
0.0	.727	.694	.716	.709	.796	.581	.556	.590	.584	.698
0.3	.708	.706	.720	.723	.772	.574	.570	.583	.587	.673
0.6	.628	.740	.699	.730	.678	.532	.608	.580	.589	.600
0.9	.544	.800	.652	.702	.567	.504	.662	.568	.532	.524
					X7 β_2/σ = 1.6					
0.0	.728	.690	.729	.715	.823	.580	.555	.590	.595	.738
0.3	.706	.691	.722	.714	.796	.570	.557	.584	.596	.714
0.6	.631	.686	.668	.714	.696	.538	.549	.560	.596	.628
0.9	.550	.680	.588	.668	.577	.509	.552	.518	.545	.540

Table 4: Estimated powers of one-sided tests of H_0: $\beta_2 = 1$ when n = 60.

	10% nominal level					5% nominal level				
ρ	OLS	p(y)	Durbin	ML	R	OLS	p(y)	Durbin	ML	R
					X1 β_2/σ = 50.0					
0.0	.886	.876	.889	.885	.918	.802	.788	.814	.810	.860
0.3	.828	.800	.816	.804	.840	.744	.683	.716	.688	.764
0.6	.726	.766	.776	.694	.719	.631	.638	.656	.536	.631
0.9	.554	.788	.782	.480	.531	.482	.659	.660	.276	.451
					X2 β_2/σ = 50.0					
0.0	.878	.864	.878	.874	.908	.787	.772	.794	.786	.852
0.3	.814	.778	.716	.775	.829	.733	.664	.692	.665	.756
0.6	.710	.729	.744	.654	.709	.614	.600	.618	.484	.617
0.9	.544	.747	.746	.416	.522	.469	.612	.615	.230	.440
					X3 β_2/σ = 7.5					
0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.3	1.000	1.000	1.000	1.000	1.000	.999	.998	1.000	1.000	1.000
0.6	.988	.994	.996	.992	.982	.974	.980	.988	.973	.960
0.9	.785	.982	.972	.910	.754	.751	.964	.951	.767	.716
					X4 β_2/σ = 7.5					
0.0	.982	.978	.978	.978	.985	.950	.946	.950	.947	.958
0.3	.970	.985	.985	.984	.974	.934	.950	.956	.946	.943
0.6	.925	.986	.986	.974	.934	.866	.958	.965	.928	.883
0.9	.740	.984	.977	.896	.765	.640	.958	.950	.754	.682
					X5 β_2/σ = 2.0					
0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.3	.998	.987	.997	.995	.998	.996	.959	.988	.983	.995
0.6	.947	.739	.892	.823	.930	.932	.534	.822	.730	.890
0.9	.666	.326	.550	.403	.630	.646	.191	.486	.324	.590
					X6 β_2/σ = 1.6					
0.0	.976	.972	.974	.970	.980	.946	.933	.942	.939	.953
0.3	.952	.978	.974	.979	.958	.910	.948	.946	.947	.914
0.6	.830	.992	.981	.988	.829	.768	.980	.956	.962	.759
0.9	.574	.998	.961	.972	.560	.529	.996	.936	.894	.507
					X7 β_2/σ = 1.6					
0.0	.976	.970	.972	.970	.982	.947	.932	.942	.934	.958
0.3	.950	.976	.970	.977	.958	.909	.946	.944	.946	.924
0.6	.833	.992	.966	.988	.834	.773	.980	.930	.964	.777
0.9	.584	.998	.884	.980	.573	.540	.996	.845	.918	.530

FOOTNOTES

1. See for example Bewley (1986, section 3.3), King (1987), King and McAleer (1987), Moulton and Randolph (1989), Chesher and Austin (1991).

2. If a and b are vectors of the same dimension then $a > b$ denotes $a_i \geq b_i$ for every i with at least one strict inequality.

3. Others to have proposed the use of the marginal likelihood function, particularly in the context of estimation, include Levenbach (1972, 1973), Patterson and Thompson (1975), Cooper and Thompson (1977) and Tunnicliffe Wilson (1989). The main theme of this literature is that the use of the marginal likelihood helps reduce bias in maximum likelihood estimation.