ISSN 0729-0683

ISBN 0 86746 679 0

MONASH UNIVERSITY

TOWARDS A THEORY OF POINT OPTIMAL TESTING

Maxwell L. King

Working Paper No. 4/88

January 1988

DEPARTMENT OF ECONOMETRICS

ISSN 0729-0683

ISBN 0 86746 679 0

TOWARDS A THEORY OF POINT OPTIMAL TESTING

Maxwell L. King

Working Paper No. 4/88

January, 1988

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS AND POLITICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

TOWARDS A THEORY OF POINT OPTIMAL TESTING

Maxwell L. King

Department of Econometrics,

Monash University,

Clayton 3168,

Australia.

Key Words and Phrases: autocorrelation; hypothesis testing; linear regression model; most powerful test; Neyman-Pearson theory; normality; power envelope.

ABSTRACT

This paper puts the case for the inclusion of point optimal tests in the econometrician's repertoire. They do not suit every testing situation but the current evidence, which is reviewed here, indicates that they can have extremely useful small-sample power properties. As well as being most powerful at a nominated point in the alternative hypothesis parameter space, they may also have optimum power at a number of other points and indeed be uniformly most powerful when such a test exists. Point optimal tests can also be used to trace out the maximum attainable power envelope for a given testing problem, thus providing a benchmark against which test procedures can be evaluated. In some cases, point optimal tests can be constructed from tests of a simple null hypothesis against a simple alternative. For a wide range of models of interest to econometricians, this paper shows how one can check whether a point optimal test can be constructed in this way. When it cannot, one may wish to consider approximately point optimal tests. As an illustration, the approach is applied to the non-nested problem of testing for AR(1) disturbances against MA(1) disturbances in the linear regression model.

1. INTRODUCTION

Given that the scientific method essentially involves the formulation of an hypothesis followed by attempts to refute it using observed data, one might argue that statistical hypothesis testing is the cutting-edge that allows the discipline of economics to claim it is a science. Typically, when faced with the problem of testing an economic hypothesis, the econometrician has a fixed amount of data at his disposal. Furthermore, the data are generally gathered by observing the economic process rather than by conducting controlled experiments. Hence it is essential that test procedures using such data should be as powerful as possible.

In recent years there has been a rapid expansion in the availability of computer time. Advances at all levels have reduced the costs of computing to such an extent that highly computational procedures are becoming more and more feasible. Indeed, vast amounts of computing can now be done for the sums of money being spent on experiments in the physical sciences. It appears that we are now reaching the stage where we should be asking: what kind of test procedure would we wish to use if computation time were not a constraint?

One possibility is to use Bayesian posterior odds; the computer revolution has certainly helped remove the computational barrier that had previously hindered their use. Yet there still seems to be a reluctance by many in the profession to use Bayesian methods. Reasons for this reluctance have recently been discussed by Efron (1986). He notes that an important requirement of any statistical theory used in scientific work is that it results in widespread agreement that the data have been interpreted fairly. Unfortunately there is the perception that the outcome of a Bayesian analysis can be sensitive to the choice of prior distribution. One researcher's prior may not be that of his readers. Possible solutions to this problem might be to investigate the sensitivity of posterior odds to the prior distribution or the adoption of public priors which are generally accepted by the profession (see for example Doan, Litterman and Sims, (1984)). But until this issue is resolved satisfactorily, econometricians will continue to

turn to non-Bayesian testing procedures[1] so we shall address our question to these procedures.

Obviously, we would like to use a uniformly most powerful (UMP) test when such a test exists, which does seem to be rarely. This implies that our preferred test procedure should result in the UMP test when it exists. There may be much less agreement about what the preferred test should be when no UMP test exists, given that no single test can dominate in terms of power. Clearly, we should rule out any test whose power curve can be dominated by the power curve of another test. If we focus on one point in the alternative hypothesis parameter space, then for a given significance level, all tests have a single power value at this point. In theory, for a given class of tests, the maximum (or supremum) of these power values exists and a test whose power attains this maximum is a most powerful (MP) test in the neighbourhood of the predetermined point. We shall call such a test a point optimal test.

An insightful way of looking at the problem of choosing a test procedure is to regard each test, at a given significance level, as a power curve over the parameter space. A choice of test is in fact a choice of power curve. If a UMP test exists, the choice is straightforward because then one power curve dominates all others. When no UMP test exists, the choice is very difficult because at different parts of the parameter space, different tests will have highest power. Then any test we choose will favour particular parts of the parameter space where the test's relative power performance is best. Furthermore, typically we won't know over which parts of the parameter space our test performs best. One can view the class of point optimal tests as a collection of power curves which have highest possible power at predetermined points in the parameter space. As we shall see, they may also have optimum power at a number of other points and good relative power over large parts of the parameter space. When we choose to use a point optimal test, in effect we are nominating which part of the parameter space we want our test to have good relative power.

[1] Box (1980) advocates the use of Bayesian procedures for estimation and sampling theory procedures for the diagnostic checking or "criticism" of models.

Point optimal tests are useful in a number of ways. As tests they are most attractive for problems in which the parameter space can be restricted in size by theoretical considerations. For example, economic theory is usually good at providing information about the signs of parameters. Because of their power properties, point optimal tests are particularly attractive when testing one economic theory against another, perhaps a new theory against an existing theory. Based on the literature discussing the existing theory, one could nominate idealized values of the alternative model's parameters. A point optimal test would ensure optimal power at this nominated point and, depending on the structure of the problem, could give good power over the entire parameter space. If the critical regions of a point optimal test are invariant to the choice of point then the test is UMP. For some problems, the point optimal test may be approximately UMP or UMP over certain subspaces of the alternative parameter space. In such cases, the choice of point is not critical. For other problems, the test's power may be quite different from the maximum attainable power as one moves away from the predetermined point. In such circumstances, the choice of point assumes much greater importance.

Point optimal tests can also be used to trace out the maximum attainable power envelope for a given testing problem. The power envelope provides an obvious benchmark against which test procedures can be evaluated. It is very reassuring if one can show that the power of the test of interest is always close to the power envelope, say not less than one or five per cent below it. On the other hand, a large difference between the test's power curve and the power envelope suggests that a more powerful test may exist.

It is not suggested that point optimal tests suit every testing situation in econometrics. For example, they are not suited to problems in which the alternative parameter space cannot be restricted by theoretical considerations such as knowledge of signs of parameters. Also, little is known about their performance in problems involving a moderate or large number of parameters.

The aim of this paper is to review and extend understanding of point optimal testing. The following section begins by describing the construction of a

point optimal test from tests of a simple null hypothesis against a simple alternative. It also explains how one can discover whether a point optimal test can be constructed in this way for a wide range of econometric models. This section closes by briefly considering Lehmann and Stein's (1948) and Lehmann's (1959) general approach to constructing point optimal tests as well as approximate point optimal tests for situations in which it is extremely difficult to find true point optimal tests. Currently, our knowledge of the small-sample power properties of point optimal tests is confined to problems involving the linear regression model. These studies are surveyed in section 3. In particular, attention is focused on the use of the principle of invariance to reduce the parameter space and methods of choosing the point at which power is optimized. The theory discussed in sections 2 and 3 is applied in sections 4 and 5 to the problem of testing for first-order autoregressive (AR(1)) disturbances against first-order moving average (MA(1)) disturbances in the linear regression model. Section 4 discusses the testing problem and the construction of point optimal tests while section 5 reports an empirical power comparison involving two point optimal tests and two tests suggested by King (1983a) for this problem. Some concluding remarks may be found in the final section.

2. THEORY

Let x be an observable $n \times 1$ vector and suppose we wish to test,

$$H_0 : x \text{ has density } f(x, \omega),$$

where ω is a $j \times 1$ vector of parameters restricted to the set Ω, against

$$H_a : x \text{ has density } g(x, \phi),$$

where ϕ is an $i \times 1$ vector of parameters restricted to the set Φ. This is a very general form of testing problem and includes both nested and non-nested problems as special cases. It is assumed that any knowledge about the possible range of parameter values has been used to keep the parameter sets, Ω and Φ, as small as possible.

For the simpler problem of testing

$$H_0^1 : x \text{ has density } f(x,\omega_1)$$

against

$$H_a^1 : x \text{ has density } g(x,\phi_1),$$

where $\omega_1 \in \Omega$ and $\phi_1 \in \Phi$ are fixed and known, we have simple null and alternative hypotheses. Therefore, the Neyman-Pearson lemma (see e.g. Lehmann (1959, p.65)) implies that rejecting H_0^1 for large values of

$$r = g(x,\phi_1)/f(x,\omega_1)$$

is a MP test. If r is used as a test statistic for the wider problem of testing the simple null hypothesis, H_0^1, against the composite alternative, H_a, then this test is, by construction, MP in the neighbourhood of $\phi = \phi_1$. However, its use for the more general problem of testing the composite hypothesis, H_0, against the composite alternative, H_a, does not necessarily result in a test which is MP in the neighbourhood of $\phi = \phi_1$. To see this observe that the critical value for the former test is found by solving

$$Pr[r > r' \mid x \text{ has density } f(x,\omega_1)] = \alpha$$

for r', where α is the desired level of significance. Because H_0 is composite, the distribution of r under H_0 and hence the probability of a Type I error for the latter test may be a function of ω. The standard approach in this case (see for example Lehmann and Stein (1948)) is to control the maximum probability of a Type I error by ones choice of critical value. Hence, for the latter test, the critical value is found by solving

$$\sup_{\omega\in\Omega} Pr[r > r^* \mid x \text{ has density } f(x,\omega)] = \alpha \qquad (1)$$

for r^*. In general $r' > r^*$. Note that if Ω is closed, then

$$Pr[r > r^* \mid x \text{ has density } f(x,\omega)] = \alpha \qquad (2)$$

will be true for at least one $\omega \in \Omega$. If ω_1 can be chosen to be such a value, then $r' = r^*$ and the critical regions of the two tests correspond. The test of H_0 against H_a is then MP in the neighbourhood of $\phi = \phi_1$. If this were not the case there would exist a test which was more powerful at $\phi = \phi_1$ and such a test would contradict the Neyman-Pearson lemma because, as a test of H_0^1 against H_a^1, it would be more powerful than the test based on r.

If such an ω_1 exists, then we have a straightforward method of constructing the point optimal test. In order to check for its existence, we need a method of calculating the left hand side of (2).

Suppose y is an observable $n \times 1$ random vector such that under H_0,

$$y \sim N(a(\omega), A(\omega)), \qquad \omega \in \Omega,$$

while under H_a,

$$y \sim N(b(\phi), B(\phi)), \qquad \phi \in \Phi.$$

This formulation includes a wide variety of models of interest in econometrics such as Box-Jenkins time-series models, linear and non-linear regression models, linear dynamic models after repeated substitution for lagged dependent variables with y_0 assumed nonstochastic, and simultaneous and other systems-of-equation models where y is the vector of stacked endogenous variable vectors. It also allows any parameterization of the covariance matrix of the disturbances in such models, and therefore permits a very wide range of nested and non-nested testing problems. For ease of exposition, we shall assume both $A(\omega)$ and $B(\phi)$ are nonsingular matrices, at least for ω and ϕ values of interest.

We wish to evaluate probabilities of the form

$$Pr[r > r^* \mid y \sim N(a(\omega), A(\omega))], \tag{3}$$

where

$$r = \frac{|B(\phi_1)|^{-1/2} exp\{-\frac{1}{2}(y - b(\phi_1))' B^{-1}(\phi_1)(y - b(\phi_1))\}}{|A(\omega_1)|^{-1/2} exp\{-\frac{1}{2}(y - a(\omega_1))' A^{-1}(\omega_1)(y - a(\omega_1))\}}.$$

If $A(\omega_1) \neq B(\phi_1)$, an equivalent critical region is to reject for small values of

$$\begin{aligned}(y-b(\phi_1))'B^{-1}(\phi_1)(y-b(\phi_1)) - (y-a(\omega_1))'A^{-1}(\omega_1)(y-a(\omega_1)) \\ = (y-d)'(B^{-1}(\phi_1)-A^{-1}(\omega_1))(y-d) - d'(B^{-1}(\phi_1)-A^{-1}(\omega_1))d \\ +b'(\phi_1)B^{-1}(\phi_1)b(\phi_1) - a'(\omega_1)A^{-1}(\omega_1)a(\omega_1)\end{aligned}$$

in which

$$d' = (b'(\phi_1)B^{-1}(\phi_1) - a'(\omega_1)A^{-1}(\omega_1))(B^{-1}(\phi_1)-A^{-1}(\omega_1))^{-1} \tag{4}$$

or to reject for small values of

$$s(\phi_1,\omega_1) = (y-d)'(B^{-1}(\phi_1)-A^{-1}(\omega_1))(y-d).$$

In other words, probabilities of the form

$$Pr[s(\phi_1,\omega_1) < s^* \mid y \sim N(a(\omega), A(\omega))]$$

are equivalent to those of the form of (3).

We can write

$$s(\phi_1,\omega_1) = z'Dz$$

where

$$z = A^{-1/2}(\omega)(y-d)$$

and

$$D = (A^{1/2}(\omega))'(B^{-1}(\phi_1)-A^{-1}(\omega_1))A^{1/2}(\omega). \tag{5}$$

Note that when $y \sim N(a(\omega), A(\omega))$, $z \sim N(A^{-1/2}(\omega)(a(\omega)-d), I_n)$. Hence

$$Pr[s(\phi_1,\omega_1) < s^* \mid y \sim N(a(\omega), A(\omega))] = Pr[\sum_{i=1}^{n} \lambda_i \xi_i^2 < s^*], \tag{6}$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of D and $\xi_1^2, ..., \xi_n^2$ are independent non-central chi-squared variates with one degree of freedom and non-centrality parameters

$$\delta_i^2 = [P'A^{-1/2}(\omega)(a(\omega)-d)]_i^2, \qquad i = 1, \cdots, n,$$

in which P is the orthogonal matrix of eigenvectors of D. This probability can be evaluated using Imhof's (1961) algorithm for computing the distribution function of quadratic forms in normal variables. This may be achieved using Koerts and Abrahamse's (1969) FQUAD subroutine or the version of Imhof's algorithm coded in Algol by Davies (1980). An alternative approach to calculating (6) that should cut computational time, especially for large n, has been proposed by Farebrother (1985).

If $A(\omega_1) = B(\phi_1)$, critical regions of the form $r > r^*$ are equivalent to those which reject for small values of

$$2[a(\omega_1) - b(\phi_1)]'A^{-1}(\omega_1)y - b'(\phi_1)A^{-1}(\omega_1)b(\phi_1) + a'(\omega_1)A^{-1}(\omega_1)a(\omega_1). \quad (7)$$

The probability of rejection can be calculated by noting that (7) has a normal distribution whenever y is multivariate normal.

Thus, for any testing problem that fits into this framework one can explore whether an ω_1 value exists such that (2) holds for $\omega = \omega_1$ subject to the constraint that (1) also holds. This involves the following iterative procedure: fix a value for ω_1, find r^* by solving (2) and then check to see if (1) holds. If it does, we have found ω_1 and r^*. If not, choose a new value of ω_1 by moving it in the direction of the ω values which cause (1) to be violated and repeat the process.

If we knew in advance that the testing problem was one for which an ω_1 value exists, then we could proceed as follows. When the ω_1 value is known, we can apply the test by first calculating $s(\phi_1, \omega_1)$ for our given sample and finding

$$p = Pr[s(\phi_1, \omega_1) < s_c(\phi_1, \omega_1) \mid y \sim N(a(\omega_1), A(\omega_1))],$$

where $s_c(\phi_1, \omega_1)$ denotes the calculated value of $s(\phi_1, \omega_1)$. If p is less than our desired level of significance, then we reject H_0, otherwise H_0 is not rejected.

Consider the case where ω_1 is unknown. At the p significance level it will be such that

$$Pr[s(\phi_1, \omega_1) < s_c(\phi_1, \omega_1) \mid y \sim N(a(\omega), A(\omega))]$$

is maximized at $\omega = \omega_1$. This implies that we can find both ω_1 and p by maximizing

$$Pr[s(\phi_1, \omega_1) < s_c(\phi_1, \omega_1) \mid y \sim N(a(\omega_1), A(\omega_1))] \tag{8}$$

with respect to ω_1. This can be done using any standard non-linear optimization computer package. For any given value of ω_1, the objective function is evaluated as follows:

(a) compute the calculated value of $s(\phi_1, \omega_1)$, namely $s_c(\phi_1, \omega_1)$,

(b) find d and D for $\omega = \omega_1$ using equations (4) and (5) with $\omega = \omega_1$,

(c) compute (8) either using the algorithm outlined by Farebrother or by Koerts and Abrahamse's or Davies' algorithms after standard computer packages have been used to find the eigenvalues and eigenvectors of D and the noncentrality parameters $\delta_1^2, ..., \delta_n^2$ have been calculated.

Note that the iterative maximization process can be stopped and H_0 accepted whenever (c) yields a probability greater than the desired significance level.

An obvious question is: under what conditions does such an ω_1 value exist? One set of sufficient conditions for the existence of an ω_1 value can be obtained as follows. Given fixed values of α and ϕ_1 and assuming Ω is closed, for any $\omega \in \Omega$ let $f_{\alpha,\phi_1}(\omega_1)$ denote the value of ω for which (2) holds subject to (1). If necessary we assume $f_{\alpha,\phi_1}(\omega_1)$ has been made into a single valued function by a suitable selection function in cases in which (2) has multiple solutions. Then f_{α,ϕ_1} is a mapping of Ω onto itself. If Ω is a convex, compact subspace of a Banach space and f_{α,ϕ_1} is continuous then by Schauder's Fixed Point Theorem, an appropriate ω_1 value will exist. While the first condition is easily verified, this is not always the case for the second condition. If f_{α,ϕ_1} maps every $\omega \in \Omega$ to a single point, then obviously this is the desired ω_1 value. Also if one can show that the left hand side of (2) has for any ω_1, ϕ_1 and r^* only one local maximum then f_{α,ϕ_1} will be continuous.

It is also worth noting that the problem we are attempting to solve is one of finding the optimal test of a composite null hypothesis against a simple alternative. This problem was first addressed by Lehmann and Stein (1948) who were able to find such tests for a range of straightforward testing problems. They gave some clues as to how such tests might be found but did not give a general method of test construction. It turns out that all the tests they discussed could have been constructed using the method outlined above. Their main aim was to show that best similar tests are not always most powerful within the class of all tests including nonsimilar tests. Consequently, they sometimes only showed the existence of an ω_1 value without considering how it might be found. The above method fills this gap.

Unfortunately, not all problems will be such that our desired ω_1 value exists. The Lehmann-Stein approach, which is further developed in Lehmann (1959, pp.90-94), does not rely on such a value existing. Their approach involves finding a probability distribution, λ, defined over the null hypothesis parameter space, Ω, and then replacing H_0 with the simple hypothesis, H_λ, that the density function of x is given by

$$h_\lambda(x) = \int_\Omega f(x,\omega)d\lambda(\omega).$$

If the most powerful test of H_λ against H_a^1 of size α is also of size less than or equal to α with respect to H_0 then the resultant test is a point optimal test of H_0 against H_a. We have considered how one can look for and construct such a test when λ is the degenerate distribution at $\omega = \omega_1$. It is not clear how one can find λ and conduct the test when λ has some other form.

In such cases, in order to have an operational test, one may wish to consider tests which are approximately point optimal. For example, a test based on r with critical value r^* determined by (1) could be considered to be approximately point optimal if $Pr(r > r^* \mid H_0^1)$ is close to α, say within five per cent of α. This is because if we modify[2] the H_0 distributions such that this probability is

[2] Durbin and Watson (1950) used a similar approach in the construction of their test for autocorrelation. They found a test which has optimal properties for a modified version of their original testing problem.

equal to α, then for this new problem our test is point optimal. Hence one may wish to search for an ω_1 value for which

$$\alpha - Pr[r > r^* \mid x \text{ has density } f(x,\omega_1)] \tag{9}$$

is minimized. Furthermore, because (9) depends on ϕ_1, one might also consider varying ϕ_1 in order to minimize this measure of approximation further.

Suppose after attempting to construct a point optimal test, one is forced to consider an approximate point optimal test, and finds that, even after minimization, (9) is closer to α than to zero. Although the resultant test might be a long way from being an approximate point optimal test, it can still have power properties that make it worthy of use. A successful application of this approach (King (1987c)) is discussed in the next section. This is a promising aspect of point optimal testing theory that may reward further research.

The parameter vectors ω and ϕ may include common parameters which, for our testing problem, are nuisance parameters. If the critical region $r > r'$ does not depend on the choice of values for these parameters in ϕ_1 and ω_1 then the final test is similar with respect to these nuisance parameters. If a search of the parameter space, Ω, is required to find an appropriate ω_1 value, either for a point optimal test or an approximate point optimal test, then it helps if Ω is as small as possible. One should check whether economic theory allows one to impose restrictions on the ranges of the various parameters both under H_0 and H_a. One may also wish to use standard similarity arguments such as conditioning on sufficient statistics of the nuisance parameters to help reduce the dimension of the problem. See Hillier (1987) for an excellent survey article on the construction of similar tests in econometrics.

In most of the examples of point optimal tests reviewed in the following section, invariance arguments have been used to eliminate the nuisance parameter problem with considerable success. The reasoning behind this approach is that if an hypothesis testing problem is invariant to a class of transformations on the observed sample, it is desirable that the test procedure also have this property. For example, the presence of serial correlation in the disturbances of

the standard linear regression model is independent of the scale of y and hence we should only consider tests which have this invariance property.

A test can be viewed as a partition of the sample space into two regions, a rejection region and a non-rejection region. If our observed sample (y in the linear regression example) falls in the rejection region, H_0 is rejected. Otherwise H_0 is not rejected. An invariant test is one for which each pair of points in the sample space that can be related by a transformation (i.e., one can be obtained as a transformation of the other) either both fall into the rejection region or both fall into the non-rejection region. We do not want to be able to move from one region to the other by transforming y. Hence the problem of deciding how to partition the sample space simplifies to one of deciding which sets of points related by transformations should be in the rejection region and which should not. Fortunately, instead of considering sets of points, we can use a convenient summary statistic called a maximal invariant. A maximal invariant is a function of the observations which is invariant to a class of transformations yet takes on different values for any two samples that are not related by a member of the class. Thus, instead of dealing with the observed sample we can base our inference on the maximal invariant in the sense that we can treat it as our observed sample. Hopefully its distributions under H_0 and H_a will involve fewer parameters. A drawback with this approach is that it is often difficult to find the distribution of the maximal invariant under H_0 and H_a. It is often necessary to write these distributions as integrals over the group of transformations.

The invariance approach is illustrated in the next section in which we review the literature on point optimal tests in the context of the linear regression model. For further discussion of the theory of invariance see Lehmann (1959, chapter 6) and Cox and Hinkley (1974, pp.157-171).

3. POINT OPTIMAL TESTS OF THE LINEAR REGRESSION MODEL

3.1 Tests of the disturbance covariance matrix

3.1.1 Theory

We begin by considering tests of the disturbance covariance matrix of the linear regression model

$$y = X\beta + u, \tag{10}$$

where y is $n \times 1$, X is an $n \times k$ matrix that is assumed independent of u and of rank $k < n$, β is a $k \times 1$ vector of parameters and u is the $n \times 1$ disturbance vector. The result which provides the key to much of what follows involves testing

$$H_0 : u \sim N(0, \sigma^2 I_n)$$

against

$$H_1 : u \sim N(0, \sigma^2 \Upsilon),$$

where Υ is a known $n \times n$ positive definite matrix. Observe that this testing problem is invariant with respect to transformations of the form

$$y \rightarrow \eta_0 y + X\eta, \tag{11}$$

where η_0 is a positive scalar and η is a $k \times 1$ vector. In other words, changing the scale of y and adding a known linear combination of regressors to the rescaled y does not change the truth of either H_0 or H_1.

Let $m = n - k$, $M = I_n - X(X'X)^{-1}X'$, $z = My$ be the ordinary least squares (OLS) residual vector from (10) and P be an $m \times n$ matrix such that $PP' = I_m$ and $P'P = M$. Note that the $m \times 1$ vector Pz is a LUS residual vector[3] given that under H_0, $Pz \sim N(0, \sigma^2 I_m)$. The vector

$$v = Pz/(z'P'Pz)^{1/2}$$

[3] For an introduction to LUS residuals see Theil (1971, chapter 5) or King (1987a, section 5).

is a maximal invariant under the group of transformations given by (11) for our problem. The density of v under H_1 can be shown to be

$$f(v)dv = \frac{1}{2}\Gamma(m/2)\pi^{-m/2}|P\Upsilon P'|^{-1/2}(v'(P\Upsilon P')^{-1}v)^{-m/2}dv, \tag{12}$$

where dv denotes the uniform measure on the surface of the unit m-sphere. If $\Upsilon = I_n$, as it does under H_0, (12) reduces to

$$f(v)dv = \frac{1}{2}\Gamma(m/2)\pi^{-m/2}dv \tag{13}$$

which is the uniform density on the surface of the unit m-sphere. Observe that invariance has removed the nuisance parameters β and σ^2 as both (12) and (13) do not involve any unknown parameters. Our problem has now become one of testing a simple null hypothesis against a simple alternative with v representing the observations. The Neyman-Pearson lemma implies that a MP test within the class of invariant tests can be based on rejection or critical regions of the form

$$s = v'(P\Upsilon P')^{-1}v < s^*, \tag{14}$$

where s^* is a suitably chosen critical value. King (1980, Lemma 2) shows that s can also be written as

$$s = \hat{u}'\Upsilon^{-1}\hat{u}/z'z, \tag{15}$$

where $\hat{u}$ is the generalized least squares (GLS) residual vector assuming covariance matrix Υ.

This test was first constructed by Kadiyala (1970) using a different approach. He noted that if u was observable, a MP test can be constructed using a result[4] due to Lehmann and Stein (1948). He then proposed that his analysis start with the observable random vector of OLS residuals, z, rather than y, and showed that a MP test with respect to this new problem is to reject H_0 for small values of s. That the two approaches yield the same test is not surprising given that z is a maximal invariant with respect to transformations of the form

$$y \rightarrow y + X\eta, \tag{16}$$

[4] See section 3.2.

where η is a $k \times 1$ vector. Kadiyala also showed the test to be unbiased and noted that it could be derived as a likelihood ratio test. More recently, Dastoor and Fisher (1988) observed that the test can be interpreted as a special case of the Cox (1961, 1962) test of non-nested hypotheses.

The usefulness of the above result becomes evident when one wishes to test H_0 against

$$H_a : u \sim N(0, \sigma^2 \Upsilon(\theta)),$$

where θ is a $q \times 1$ parameter vector and $\Upsilon(\theta)$ is a symmetric matrix that is positive definite for all $\theta \in \Theta$, where Θ denotes the range of θ vectors of interest. For any choice of $\theta = \theta_1 \in \Theta$ and $\theta_1 \neq 0$, rejecting H_0 for small values of (15), where $\Upsilon = \Upsilon(\theta_1)$, is a MP invariant (MPI) test at $\theta = \theta_1$.

We shall call such a test a point optimal invariant (POI) test. It is often convenient to compute (15) using the Cholesky decomposition of Υ, i.e. L such that

$$L'L = \Upsilon^{-1},$$

to transform (10) because the numerator of (15) is the sum of squared OLS residuals from

$$Ly = LX\beta + Lu.$$

Appropriate critical values can be computed by noting that (14) can be written as

$$s = \xi'(P\Upsilon P')^{-1}\xi/\xi'\xi = \sum_{i=1}^{m} \lambda_i \xi_i^2 / \sum_{i=1}^{m} \xi_i^2,$$

where $\xi = \{P'z/\sigma\} \sim N(0, I_m)$ under H_0 and $\lambda_1, ..., \lambda_m$ are the reciprocals of the non-zero eigenvalues of ΥM or, equivalently, the non-zero eigenvalues of

$$\Delta = \Upsilon^{-1} - \Upsilon^{-1}X(X'\Upsilon^{-1}X)^{-1}X'\Upsilon^{-1}.$$

Thus, like for the Durbin-Watson (DW) statistic, α-level critical values of s can be found by solving

$$Pr[\sum_{i=1}^{m}(\lambda_i - s^*)\xi_i^2 < 0] = \alpha \qquad (17)$$

for s^*. This can be done iteratively using either Koerts and Abrahamse's (1969) FQUAD subroutine, Farebrother's (1980) PAN procedure or Davies' (1980) algorithm to evaluate the left hand side of (17).

Alternative methods of computing $Pr[s < s^*]$ which do not require finding eigenvalues have been suggested by Palm and Sneek (1984) and Shively, Ansley and Kohn (1987). Palm and Sneek's approach involves using Householder transformations to tridiagonalize Δ. Shively *et al.* explain how a modification of the Kalman filter can be used to calculate the probability in $O(n)$ arithmetic operations in contrast to the $O(n^3)$ operations needed to compute eigenvalues. As a result, one might expect considerable computational savings over methods based on eigenvalues, especially for larger sized samples.

Observe that because $M\Delta = \Delta = \Delta M$, (15) can also be written as

$$\begin{aligned} s &= u'\Delta u / u'Mu \\ &= u'M\Delta Mu / u'Mu \\ &= z'\Delta z / z'z, \end{aligned} \tag{18}$$

which is in a similar form to the DW statistic. Thus the following extension to Durbin and Watson's (1950) lemma can be used to compute bounds for the critical values that are independent of X or particular columns of X but which depend on $\Upsilon(\theta_1)$, n and k.

Lemma: (King, 1981a). Consider the class of regression models (10) whose design matrices can be partitioned as $X = [X_1 \vdots X_2]$, such that the columns of X_1 span the same j-dimensional space, where X_1 is $n \times j$ and X_2 is $n \times (k-j)$. If

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_m$$

are the reciprocals of the non-zero eigenvalues of ΥM (or, equivalently, the non-zero eigenvalues of Δ) and if

$$\nu_1 \leq \nu_2 \leq \ldots \leq \nu_{n-j}$$

are the reciprocals of the non-zero eigenvalues of ΥM_1, where

$$M_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$$

then

$$\nu_i \leq \lambda_i \leq \nu_{i+k-j}.$$

Corollary:

$$s_L \leq s \leq s_U,$$

where

$$s_L = \sum_{i=1}^{m} \nu_i \xi_i^2 / \sum_{i=1}^{m} \xi_i^2, \qquad s_U = \sum_{i=1}^{m} \nu_{i+k-j} \xi_i^2 / \sum_{i=1}^{m} \xi_i^2$$

and under $H_0 : u \sim N(0, \sigma^2 I_n)$, $\xi \sim N(0, I_m)$.

Most commonly, X_1 would be the $n \times 1$ vector of ones representing the constant regressor.

The fact that s can be written as (18) means that many of the methods of approximating the critical value of the DW test (see King (1987a, pp.25-27)) can be used to approximate s^*. Evans and King (1985b) empirically compared the accuracy of three of these methods, namely the normal, two-moment beta and four-moment beta approximations, for a variety of POI tests against autocorrelation and heteroscedasticity. Overall, these three approximations were found to provide reasonably accurate critical values, with the four-moment beta and normal approximations being the most and least accurate, respectively. Also the approximations seemed to be more accurate for autocorrelation tests as opposed to tests against heteroscedasticity. Skewness of the test statistic was found to be a determining factor with respect to accuracy. It was suggested that if the statistic's coefficient of skewness exceeds 0.14 in absolute value, one should think twice about using the two-moment beta and normal approximations.

It is worth noting that when $q = 1$ (i.e., θ is a scalar), the limit of the critical regions of a series of POI tests with $\theta_1 > 0$ approaching zero is that of the locally best invariant (LBI) test which is also known as the locally most powerful invariant test. It is the test with the steepest sloping power curve at $H_0 : \theta = 0$ within the class of invariant tests of the same significance level. It can also be viewed as the test which has optimal power within the neighbourhood of H_0. King and Hillier (1985) show that the critical region of this test is of the form

$$s_0 = z'A_0z/z'z < c,$$

where

$$A_0 = \partial\Upsilon^{-1}(\theta)/\partial\theta|_{\theta=0} = -\partial\Upsilon(\theta)/\partial\theta|_{\theta=0},$$

z is the OLS residual vector from (10) and c is a suitably chosen critical value. They also demonstrated the equivalence of this test to a one-sided version of the Lagrange multiplier test.

In some situations, the POI critical region is invariant to the choice of θ_1 vector. When this is the case, the resultant test is uniformly most powerful invariant (UMPI). For example, consider

$$\Upsilon(\theta) = (f_1(\theta)I_n + f_2(\theta)A)^{-1},$$

where $f_1(\theta)$ and $f_2(\theta)$ are known non-zero scalar-valued functions of θ such that $f_2(0) = 0$ and A is a known $n \times n$ symmetric matrix such that $\Upsilon(\theta)$ is positive definite. If the regressors of (10) are such that they can all be written as linear combinations of k of the eigenvectors of A (i.e., the space spanned by the columns of X is also spanned by k of the eigenvectors of A) then

$$(P\Upsilon(\theta)P')^{-1} = P\Upsilon^{-1}(\theta)P'$$

and from (14) we have

$$\begin{aligned} s &= z'\Upsilon^{-1}(\theta_1)z/z'z \\ &= f_1(\theta_1) + f_2(\theta_1)z'Az/z'z. \end{aligned}$$

Therefore, for any θ_1 such that $f_2(\theta_1) > 0$ $(f_2(\theta_1) < 0)$, rejecting H_0 for small values of s is equivalent to rejecting H_0 for small (large) values of $z'Az/z'z$ and is thus a UMPI test. This is an example of a point optimal test that is UMP in certain circumstances.

3.1.2 The literature

An early suggested application of a POI test is provided by Spjotvoll (1967). He considered testing $\delta \leq \delta_0$ against $\delta > \delta_0$, where δ is the ratio of variances in the one-way classification of the analysis of variance with variance components. For this problem, the POI test which optimizes power at $\delta = \delta_1$, also maximizes the minimum power over $\delta \geq \delta_1$.

Examples of point optimal tests were used by Davies (1969) to illustrate his new class of optimal tests which he called beta-optimal tests. He constructed the point optimal test of $H_0 : \sigma = 0$ against $\sigma > 0$ when $y_i \sim IN(0, 1+\sigma^2 x_i^2)$, where x_i^2 are known constants, and compared its small-sample power with that of the locally best test and Wald's test. Davies' point optimal test, which optimizes power at a power value of 0.8, seems to have the best overall power. He also discussed the POI test of $H_0 : \sigma = 0$ against $\sigma > 0$ when $y_i \sim IN(x_i\beta, 1+\sigma^2 x_i^2)$, where β is an unknown scalar parameter.

A number of authors have proposed tests for AR(1) disturbances in the linear regression model. Berenblut and Webb (1973) considered the problem of testing $H_0 : \rho = 0$ against $\rho > 0$ in the nonstationary AR(1) process

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad u_1 = \varepsilon_1, \qquad \varepsilon_t \sim IN(0, \sigma^2), \tag{19}$$

and proposed a test that optimized power at $\rho = 1$. A different approach was taken by Fraser, Guttman and Styan (1976) who focussed on the stationary AR(1) process

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim IN(0, \sigma^2). \tag{20}$$

They constructed their Likelihood Ratio Observable (LRO) test by factoring the distribution of $w = u/\sigma$ into a distribution of the observable part of w and

a conditional distribution of the unobservable part. Application of Neyman-Pearson theory to the observable distribution results in a test which is a function of ρ . In order to make their test operational, they suggested the point optimal solution of setting $\rho = 0.5$ in the test statistic and called this test the LRO ($\rho_1 = 0.5$) test. The theory of test construction outlined in section 3.1.1 was applied by King (1985a) to the problem of testing H_0 against the alternative that the disturbances of (10) are generated by the stationary AR(1) process (20). The resultant test is identical to the operational version of Fraser *et al.*'s LRO test.

It is worth noting that for certain X matrices, POI tests are approximately UMPI against positive AR(1) disturbances. Details for the nonstationary and stationary cases are given by Berenblut and Webb (1973) and King (1985a), respectively. These results indicate that, at least for certain X matrices, the tests are insensitive to the choice of ρ value at which power is optimized. Also for stationary AR(1) disturbances, Breusch and Pagan (1984) have shown that all tests based on s as well as the DW and Berenblut-Webb tests are asymptotically equivalent. Essentially, the latter two tests can be regarded as s tests which optimize power at $\rho = 0$ and $\rho = 1$, respectively.

Empirical power comparisons (see for example King (1985a)) suggest that for the majority of economic regressor sets, the difference in power between the DW, Berenblut-Webb and POI tests is small. However, in some cases there can be a clear power advantage to be gained over the DW test from using a POI test which optimizes power at a middle value of ρ, say $\rho = 0.5$. This is especially true for Watson's (1955) X matrix for which the OLS estimator has minimum efficiency relative to the best linear unbiased estimator. Furthermore, some of the regressors in Watson's X matrix are rather similar to time series comprised purely of a business cycle and a seasonal component. Judge *et al.* (1985, p.330) note the wisdom of using a POI test which optimizes power at 0.5 in preference to the DW test. In most applications, the choice may not matter. On the other hand, cases in which there is a large power difference are likely to coincide with situations in which the OLS estimator is relatively inefficient and a powerful test is needed most.

Analogous results for the simple AR(j) case may be found in Webb (1973), King (1984) and Evans and King (1985c).

A related problem is that of testing the null hypothesis that the disturbances of (10) are generated by the Gaussian random walk process,

$$u_t = u_{t-1} + \varepsilon_t, \qquad t = 1, ..., n,$$

where $\varepsilon_t \sim IN(0, \sigma^2)$, against the alternative of AR(1) disturbances. LBI and POI tests have been proposed in the stationary AR(1) case by Sargan and Bhargava (1983) and in the nonstationary case by King (1981b). Again, for certain X matrices, these tests are approximately UMPI and empirical power comparisons suggest that POI tests have better small-sample power properties than LBI tests. Variations and generalizations of these tests have been investigated by Bhargava, Franzini and Narendranathan (1982), Bhargava (1986) and Dufour and King (1986).

It would seem that the choice of θ_1 value for use in (15) can be more critical for testing against processes other than simple AR(j) processes. Small-sample powers of POI tests have been investigated by:

(i) King (1983b, 1985b) in the context of testing $H_0 : \gamma = 0$ against $H_a^+ : \gamma > 0$ (or $H_a^- : \gamma < 0$) in regression disturbances generated by the MA(1) process

$$u_t = \varepsilon_t + \gamma \varepsilon_{t-1}, \tag{21}$$

where $\varepsilon^* = (\varepsilon_0, ..., \varepsilon_n)' \sim N(0, \sigma^2 I_{n+1})$;

(ii) Evans and King (1985a) when testing $H_0 : b = 0$ against $H_a : b > 0$ assuming heteroscedastic disturbances with variances

$$var(u_t) = (a + bz_t)^d, \tag{22}$$

where z_t is a known exogenous variable, $d \neq 0$ is a known scalar while a and b are unknown scalars;

(iii) King (1986) when testing for autocorrelation in regression disturbances which are the sum of independent white noise and AR(1) components and therefore a special case of an ARMA(1,1) process ($q = 2$);

(iv) King and Skeels (1984) in the context of testing for joint AR(1) and heteroscedastic regression disturbances ($q = 2$).

These studies all found that certain POI tests have more desirable small-sample power properties than their competitors including LBI tests. As might be expected, in each case the relative performance of the LBI test was best for parameter values close to H_0 while that of the POI test was best in a region of the parameter space that includes the point at which power is optimized. A choice of middle value for this point often resulted in a POI test with relatively good power over the entire parameter space.

A further class of POI tests for testing against (22) when z_t is unknown but the ranking of σ_t^2 is known are discussed by Evans and King (1988). These tests are constructed by reordering the observations so that σ_t^2 is an increasing function of t and then approximating the unknown z_t with t. The tests, therefore, are point optimal in a parameter space which includes $z_t, t = 1, \ldots, n$, as well as b and d. An advantage is that bounds, independent of the regressors and z_t, can be tabulated for the critical values of these tests. Evans and King report evidence that suggests the small-sample power of these tests is superior to that of other tests which only assume knowledge of the ranking of σ_t^2, although it does seem that the *relative* performance of LBI tests is better when z_t is unknown. It also appears that the less one knows about z_t in (22), the more advisable it is to optimize power at a smaller b value.

There is a growing literature which considers the application of POI tests to problems which involve testing for random coefficients in the linear regression model. As we shall see in section 3.3, this literature has made a number of advances in determining how the point at which power is optimized should be chosen.

With respect to the time series-model comprising the sum of trend, seasonal and irregular components, Franzini and Harvey (1983) constructed point optimal tests for deterministic trend and seasonal components. Their testing problem can be formulated as one of testing $H_0 : \theta = 0$ against $H_a : \theta > 0$ in the context of (10) with $u \sim N(0, \sigma^2 \Upsilon(\theta))$, where θ is a 3×1 vector of non-negative

variances. They also investigated testing for a partially deterministic model and recommended the application of POI tests to the first differenced model.

A special case of Franzini and Harvey's time-series model, namely

$$y_t = \mu_t + \varepsilon_t, \tag{23}$$

$$\mu_t = \mu_{t-1} + \beta + \delta_t, \qquad t = 1, ..., n,$$

where $\varepsilon_t \sim N(0, \sigma^2)$ and $\delta_t \sim N(0, \tau^2)$, $t = 1, ..., n$, are mutually independent, was considered by Nyblom (1986). He investigated testing $\rho = \tau^2/\sigma^2 = 0$ against $\rho > 0$ and compared the small-sample power of various forms of LaMotte and McWhorter's (1978) test with the LBI and two POI tests, one of which is the Franzini-Harvey POI test. The two POI tests were found to have superior power over a wide range of the more interesting alternatives.

The Hildreth-Houck (1968) random coefficient model assumes regression coefficients in (10) of the form

$$\beta_{it} = \beta_i + \varepsilon_{it}, \qquad t = 1, ..., n,$$

where $\varepsilon_{it} \sim IN(0, \sigma_i^2)$. When the first regressor is the intercept term, testing $H_0 : \sigma_i^2 = 0$, $i = 2, ..., k$, is equivalent to testing for a specific form of heteroscedasticity which, when $k = 2$, is a special case of (22). Milan (1984) compared the small-sample performances of some POI tests for this problem with a range of alternative tests for heteroscedasticity. The POI test based on a θ_1 value which assumes $\sigma_1^2 = ... = \sigma_k^2$ and a coefficient of variation of the implied disturbances of 0.5 was found to be best in terms of power.

Shively (1986a) turned his attention to testing for a random walk coefficient in the model

$$y_t = x_t'\beta + z_t\alpha_t + \varepsilon_t, \tag{24}$$

$$\alpha_t = \alpha_{t-1} + \delta_t, \qquad t = 1, ..., n, \tag{25}$$

where $\varepsilon_t \sim N(0, \sigma^2)$ and $\delta_t \sim N(0, \lambda\sigma^2)$, $t = 1, ..., n$, are mutually independent, x_t is a $(k-1) \times 1$ vector of fixed regressors and z_t is a fixed regressor. He compared the small-sample power of two versions of the POI test of $\lambda = 0$

against $\lambda > 0$ with the power envelope. Both tests were found to have powers very close to the power envelope. Consequently, as Shively observes, they can be regarded as approximately UMPI tests. He was also able to confirm Nyblom's (1986) finding concerning the superiority of POI tests over the LBI and LaMotte and McWhorter's tests in the context of (23).

A related problem is that of testing for a stochastic coefficient based on Rosenberg's (1973) "return to normalcy" model. Here, (25) is replaced with

$$(\alpha_t - \mu) = \phi(\alpha_{t-1} - \mu) + \delta_t, \qquad t = 1, ..., n,$$

where again $\varepsilon_t \sim N(0, \sigma^2)$ and $\delta_t \sim N(0, \lambda\sigma^2)$, $t = 1, ..., n$, are mutually independent. Shively's (1986b, 1987) simulation results indicate that if $\theta_1 = (\phi_1, \lambda_1)'$ is chosen carefully, the resultant POI test outperforms Watson and Engle's (1985) test of $\lambda = 0$ against $\lambda > 0$ as well as King's (1987b) LBI test. Shively also computed the small-sample powers of a POI test of $\lambda = 0$ when α_t in (24) is generated by the ARIMA(1,1,0) process,

$$\alpha_t = \alpha_{t-1} + \phi(\alpha_{t-1} - \alpha_{t-2}) + \delta_t, \qquad t = 1, ..., n,$$

with $\varepsilon_t \sim N(0, \sigma^2)$ and $\delta_t \sim N(0, \lambda\sigma^2)$, $t = 1, ..., n$, mutually independent. Again he was able to conclude that if θ_1 is chosen carefully, the resultant POI test has high power across most ARIMA(1,1,0) alternatives.

In the context of testing $H_0 : \rho_4 = 0$ against $H_a : \rho_4 > 0$ when the disturbances of (10) are generated by the special AR(5) process

$$u_t = \rho_1 u_{t-1} + \rho_4 u_{t-4} - \rho_1\rho_4 u_{t-5} + e_t, \tag{26}$$

King (1987c) found that the method outlined in section 2 nearly always fails to allow the construction of point optimal tests. This testing problem is invariant to transformations of the form of (11) so that v is a maximal invariant with density given by (12) where Υ is the covariance matrix of the AR(5) process (26). Hence, after reduction through invariance, the testing problem becomes one of testing ρ_4 in the presence of a nuisance parameter, ρ_1. In the notation of section 2, $\omega = \rho_1$ and so is a scalar while $\phi = (\rho_1, \rho_4)'$ is a 2×1 vector.

The likelihood ratio of simple hypotheses, r, therefore is a function of three parameters; ρ_{10} which is the $\omega = \rho_1$ value under H_0^1 and ρ_{11} and ρ_{41} which are the ρ_1 and ρ_4 values under H_a^1. King turned to the class of "approximate" point optimal tests discussed in section 2 and investigated their small-sample power properties for this testing problem. He found that with ρ_{10} chosen to minimize (9) and using sensible rules for choosing ρ_{11} and ρ_{41} (see section 3.3), "approximate" POI tests have good small-sample power. This is despite (9) generally being closer to α, the significance level, than to zero which makes it difficult to claim that the tests are approximately optimal.

Clearly this approach shows considerable promise and could readily be applied to a number of other testing problems. Some of the more obvious examples include testing regression disturbances in the presence of autocorrelation, heteroscedasticity, a random regression coefficient or an error component for higher-order effects or combinations of these effects. While prior knowledge of the sign(s) of the nuisance parameter(s) is not necessary, it does seem that prior information about the sign(s) of the parameter(s) being tested is essential for the method to work successfully.

3.2 Other tests

Lehmann and Stein (1948) provide a number of examples of point optimal tests which did not require the use of invariance or conditioning on sufficient statistics to construct. Their aim was to show the existence of non-similar tests which can be more powerful than best similar tests. Hence they did not generally advocate the use of point optimal tests nor did they compute small-sample powers. However, they did note that the use of such tests might be desirable in situations in which no similar test exists or when the class of alternatives is sufficiently restricted.

In the context of x being an $n \times 1$ random sample from the $N(\xi, \sigma^2)$ distribution, Lehmann and Stein constructed tests of

$$H_0 : \sigma = \sigma_0 \qquad \text{against} \qquad H_1 : \xi = \xi_1, \sigma = \sigma_1,$$

and

$$H_0 : \xi = 0 \qquad \text{against} \qquad H_1 : \xi = \xi_1, \sigma = \sigma_1.$$

They also discussed point optimal tests for the Behrens-Fisher problem and for testing

$$H_0 : \sigma_1 = ... = \sigma_m$$

when x_i, $i = 1, ..., m$, are $n_i \times 1$ random samples from the $N(\xi_i, \sigma_i^2)$ distribution. Another problem of some interest is that of testing

$$H_0 : x \sim N(0, \sigma^2 A) \qquad \text{against} \qquad H_1 : x \sim N(0, \sigma_1^2 B),$$

where x is $n \times 1$, A and B are known $n \times n$ positive definite matrices and $\sigma^2 > 0$ is an unknown scalar while $\sigma_1^2 > 0$ is a known scalar. In this case, they showed that rejecting H_0 for small values of

$$\frac{x' A^{-1} x}{x' B^{-1} x}$$

yields a MP test. Observe that because the critical region is invariant to the σ_1^2 value, the point optimal test of

$$H_0 : x \sim N(0, \sigma^2 A)$$

against

$$H_1 : x \sim N(0, \tau^2 B),$$

where both σ^2 and τ^2 are unknown, gives rise to a UMP test. An implication of this result is that the POI tests discussed in section 3.1 are optimal within the class of tests invariant to transformations of the form of (16). This represents a wider class of tests than those invariant to transformations given by (11).

Another application of point optimal testing not concerned with the covariance matrix of the linear regression model involves testing whether a subset of regression coefficients are jointly zero assuming knowledge of the coefficients' signs. King and Smith (1986) showed that a one-sided t-test applied to a weighted sum of the associated regressors is UMPI along the ray whose direction is defined by the weights. Hence, we have an example of a point optimal test whose power is

optimal over a subspace of the alternative hypothesis parameter space. Various forms of the test were used by King and Smith to trace out the power envelope for a range of testing situations thus providing a benchmark against which the performance of point optimal tests (with given choices of weights) and other tests can be evaluated. King and Smith's comparison indicates that point optimal tests can have power within five per cent of the power envelope over a wide range of the parameter space, especially when the associated regressors are correlated. In contrast, the power of the F test was almost never found to be within five per cent of the power envelope. Hillier (1986) gives an alternative derivation of the test and shows that within the class of similar tests, it is UMP along the ray whose direction is defined by the weights.

3.3 The choice of point at which power is optimized

When discussing the problem of testing a simple null against a composite alternative for which no UMP test exists, Cox and Hinkley (1974, p.102) considered three possible approaches. These are (a) the point optimal solution which involves picking "somewhat arbitrarily a 'typical' point", (b) removing this arbitrariness by choosing the point to be close to the null hypothesis which leads to the locally best (or locally most powerful) test and (c) choosing the test which maximizes some weighted average of powers. Thus one non-arbitrary choice of point is to optimize power at the null hypothesis resulting in the locally best test. In some cases, this approach works well, while in others, such as testing for Gaussian random walk disturbances against non-stationary AR(1) disturbances in (10), it leads to a test with disappointing power away from the null hypothesis (see King (1981b)). Unfortunately, locally best tests optimize power where high power is needed least. Current empirical evidence plainly suggests that, by shifting the point at which power is optimized a small distance away from the origin, the small-sample power performance of the test is improved.

If there is a natural outer boundary to the parameter space under the alternative hypothesis, then an alternative approach is to choose a point on this boundary. Power is now optimized at a point at which high power is required while some of the arbitrariness of choosing a point is removed. Berenblut and

Webb (1973) favoured this approach for testing against (19) in (10) as did Sargan and Bhargava (1983). Again, current empirical evidence suggests that the performance of the test is improved by choosing a point slightly closer to the null hypothesis. In addition, for reasonably sized samples, power may be one at the outer boundary for a range of other point choices, so that nothing is gained by optimizing power at this point.

The most popular approach has been to recommend a choice based on an empirical comparison of the small-sample powers of two or more choices of "representative" or "mid-range" points with the powers of competing tests such as the locally best test. A good example of this approach is King's (1983b) study of POI tests against MA(1) disturbances (21) in the linear regression model (10). After comparing the power functions of the LBI and POI tests which maximize power at $\gamma = 0.25, 0.5, 0.75$ and 1.0, over a range of X matrices, King recommended $\gamma = 0.5$ as a good choice of point at which to optimize power. However, he did note that a case could be made for $\gamma = 0.75$ in very small samples ($n \leq 15$) and $\gamma = 0.25$ in large samples ($n \geq 60$). Others who have followed this approach include Franzini and Harvey (1983), King (1981b, 1984, 1985a, 1986), Evans and King (1985a, 1985c, 1988), King and Skeels (1984) and King and Smith (1986). We shall call this the empirical approach.

Obviously, there is a degree of arbitrariness in the empirical approach, given that choices need to be made concerning which points to include and under what circumstances (sample sizes, data sets, etc) powers are compared. In the context of testing the covariance matrix of the linear regression model (10), it does seem that a good choice of θ_1, for use in (15), should depend upon n and perhaps also k. This point is taken up by King (1985b) who re-examined his previous (1983b) study of POI tests against MA(1) disturbances. Using arguments similar to those Davies (1969) used to introduce beta-optimal tests, King concluded that θ_1 should be chosen such that the optimized power is between the size of the test and one. He showed how, in theory for a given X matrix and the scalar θ case, one can find a value of θ_1 such that the power of s at θ_1, is some predetermined value, say p_1. Assuming the power function of (15) is always a monotonically

non-decreasing function of θ, which is probably not unreasonable in many cases, the resultant test is also a beta-optimal invariant test.

Unfortunately, such a θ_1 value will vary with X, and the choice of p_1 and significance level while also requiring a large amount of computation to find. Recognizing this, King gives formulae for approximate θ_1 values for testing against $\gamma > 0$ and $\gamma < 0$, respectively, in (21). These formulae were computed based on actual θ_1 values for a representative class of X matrices, $p_1 = 0.65$ and the five per cent significance level. The p_1 value of 0.65 was chosen after initial experiments with $p_1 = 0.5$ suggested that a higher value would be more appropriate in view of the greater importance of detecting larger $|\gamma|$ values. Note that this choice of p_1 value is midway between 0.5 and the value of 0.8 suggested by Davies (1969) when introducing beta-optimal tests.

In the context of testing for a random walk coefficient in (24) and (25), Shively (1986a) compares POI tests based on $p_1 = 0.5$ and $p_1 = 0.8$. He finds his POI tests are reasonably insensitive to the choice of θ_1 value. The POI test which optimizes power at 0.5 is favoured because of the overall closeness of its power curve to the power envelope.

Nyblom (1986) suggests that Pitman asymptotic relative efficiency offers a good method of determining θ_1 as a function of sample size. This approach requires a sequence of θ_1 values for which $\theta_1 \to 0$ as $n \to \infty$, such that the limiting power stays below one for a fixed significance level. For testing $\rho = 0$ against $\rho > 0$ in (23), the special case of Franzini and Harvey's time-series model, Nyblom finds

$$\rho_1 = \delta/n^2 + O(n^{-2}),$$

where δ is chosen to give a limiting power of p_1 at significance level α, to be such a sequence. He notes that the Pitman efficiencies are rather insensitive to the choice of α and p_1 and suggests the use of $\alpha = 0.05$ and $p_1 = 0.8$. As Nyblom observes, this approach can be generalized to the θ vector, although possibly not uniquely.

When testing zero restrictions on nonnegative regression coefficients, the POI test is UMPI along a ray from the origin (see King and Smith (1986) and Hillier (1986)). Thus the choice of point in this case is effectively a choice of

direction from the origin in the parameter space. While King and Smith consider an empirical approach to this problem, Hillier suggests a minimax solution. It involves choosing that direction which maximizes the minimum power on a relevant subset of the surface of a unit sphere in a transformed parameter space.

The "approximately" POI tests of $H_0 : \rho_4 = 0$ against $H_a : \rho_4 > 0$ when the disturbances of (10) are generated by the special AR(5) process (26), investigated by King (1987c), involve the choice of three parameter values, namely ρ_{10}, ρ_{11} and ρ_{41}. The value of ρ_{10} is always chosen to minimize (9) while the ρ_{11} and ρ_{41} values determine the point at which the test "attempts" to optimize power. King found that the following rules for choosing the latter values result in a test with good small-sample power properties. Because we are interested in testing for non-zero ρ_4 values, for any $\rho_4 > 0$, say ρ_4^*, we would obviously like the minimum power along $\rho_4 = \rho_4^*$ in the $(\rho_1, \rho_4)'$ parameter space to be as large as possible. Hence, for a given ρ_4^*, an obvious choice of ρ_{11} value is that which, with $\rho_{41} = \rho_4^*$, maximizes the minimum power along $\rho_4 = \rho_4^*$. Thus, for any ρ_{41} value, we have rules for determining ρ_{11} and ρ_{10}. The ρ_{41} value is then chosen so that the maximized minimum power has a value of 0.5. This represents a new approach in the treatment of nuisance parameters in hypothesis testing. Those that cannot be eliminated through invariance are used to advantage in the final choice of test statistic.

Unfortunately, rules which involve optimizing or "approximately" optimizing power at a predetermined level generally require a high degree of computation to apply. They frequently need to be applied individually to each model and for each level of significance. On the other hand, knowledge of the test's power function is a valuable by-product of these calculations. Improved computer algorithms (see for example, Shively (1987) and Shively *et al.* (1987)) and advances in computer hardware may do much to reduce this computational burden.

4. TESTING AR(1) AGAINST MA(1) DISTURBANCES

The majority of the successful applications of point optimal tests reviewed in the previous section involve testing a null hypothesis which, after reduction

through invariance, is a simple hypothesis nested in a composite alternative. The discussion in section 2 applies to the construction of a point optimal test for a general testing problem which involves a composite null hypothesis that may or may not be nested. In the remainder of this paper, we illustrate this approach by applying it to the problem of testing AR(1) against MA(1) disturbances in the linear regression model (10). After reduction through invariance, this involves a non-nested composite null hypothesis and a composite alternative, each of which is characterized by a single parameter. This problem has previously been investigated by King (1983a) and King and McAleer (1987). King proposed a test similar to that discussed in section 2 but with ϕ_1 and ω_1 set to arbitrary values so that the resultant test is not necessarily MPI in the neighbourhood of $\phi = \phi_1$. King and McAleer compared the small-sample sizes and powers of this test with those of the Cox test, some linearized Cox tests and the Lagrange multiplier test of AR(1) disturbances against ARMA(1,1) disturbances. This comparison showed King's test to have the best power properties and the Cox and linearized Cox tests to be somewhat unreliable in small samples.

AR(1) disturbances of the form (20) imply $u \sim N(0, \sigma^2\Sigma(\rho))$, where

$$\Sigma(\rho) = 1/(1-\rho^2)\begin{pmatrix} 1 & \rho & \rho^2 & \cdot & \cdot & \cdot & \rho^{n-1} \\ \rho & 1 & \rho & & & & \cdot \\ \rho^2 & \rho & 1 & & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & & \cdot & \cdot \\ \rho^{n-1} & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix},$$

while for MA(1) disturbances generated by (21), $u \sim N(0, \sigma^2\Lambda(\gamma))$ in which $\Lambda(\gamma)$ is the tridiagonal matrix whose main diagonal elements are $1+\gamma^2$ and whose off diagonal elements are γ. Following King (1983a) and King and McAleer (1987), we shall be concerned with testing

$$H_0 : y \sim N(X\beta, \sigma^2\Sigma(\rho)), \qquad 0 \leq \rho < 1$$

against

$$H_a : y \sim N(X\beta, \sigma^2\Lambda(\gamma)), \qquad 0 < \gamma \leq 1.$$

This problem is invariant to transformations of the form of (11). Using the approach discussed in section 3.1.1, one can show that for the simpler problem of testing

$$H_0' : y \sim N(X\beta, \sigma^2\Sigma(\rho_1))$$

against

$$H_a' : y \sim N(X\beta, \sigma^2\Lambda(\gamma_1)),$$

where $0 \leq \rho_1 < 1$ and $0 < \gamma_1 \leq 1$ are fixed and known, MPI critical regions are of the form

$$r(\gamma_1, \rho_1) = \tilde{u}'\Lambda^{-1}(\gamma_1)\tilde{u}/\hat{u}'\Sigma^{-1}(\rho_1)\hat{u} < c_\alpha, \tag{27}$$

in which $\tilde{u}$ and $\hat{u}$ are the GLS residual vectors from (10) assuming covariance matrices $\Lambda(\gamma_1)$ and $\Sigma(\rho_1)$, respectively, and c_α is an appropriate critical value.

King suggested the use of $r(\gamma_1, \rho_1)$ as a statistic to test H_0 against H_a. Note that the resultant test is not necessarily a POI test of H_0. This is because c_α may not be the appropriate critical value for testing H_0. The distribution of $r(\gamma_1, \rho_1)$, and hence its probability of a Type I error, is a function of ρ under H_0. While c_α is determined by solving

$$Pr[r(\gamma_1, \rho_1) < c_\alpha | u \sim N(0, \Sigma(\rho_1))] = \alpha, \tag{28}$$

the appropriate α level critical value for testing H_0 is obtained by solving the conditional probability inequality

$$Pr[r(\gamma_1, \rho_1) < c_\alpha' \mid u \sim N(0, \Sigma(\rho)), 0 \leq \rho < 1] \leq \alpha \tag{29}$$

for the largest possible value of c_α' . Therefore $c_\alpha' \leq c_\alpha$. As discussed in section 2, if we can choose ρ_1 such that $c_\alpha' = c_\alpha$ then the test based on (27) will be POI.

For any given value of ρ

$$Pr[r(\gamma_1, \rho_1) < c_\alpha'] = Pr[u'\Gamma(\gamma_1, \rho_1, c_\alpha')u < 0]$$

$$= Pr[\sum_{t=1}^{n} \varsigma_t \xi_t^2 < 0], \tag{30}$$

where

$$\Gamma(\gamma_1,\rho_1,c'_\alpha) = \Lambda^{-1}(\gamma_1) - \Lambda^{-1}(\gamma_1)X(X'\Lambda^{-1}(\gamma_1)X)^{-1}X'\Lambda^{-1}(\gamma_1) - c'_\alpha\{\Sigma^{-1}(\rho_1) - \Sigma^{-1}(\rho_1)X(X'\Sigma^{-1}(\rho_1)X)^{-1}X'\Sigma^{-1}(\rho_1)\},$$

$\varsigma_1,...,\varsigma_n$ are the eigenvalues of $\Gamma(\gamma_1,\rho_1,c'_\alpha)\Sigma(\rho)$ and $\xi = (\xi_1,...,\xi_n)' \sim N(0,I_n)$. Thus (30) is of the form of the left hand side of (17) and can be evaluated using any of the algorithms discussed in section 3.1.1.

For any given γ_1 value, our aim is to find ρ_1 and c_α values such that (28) and (29) with $c'_\alpha = c_\alpha$, hold simultaneously. When this is the case, we have a test of H_0 against H_a which optimizes power at $\gamma = \gamma_1$ and is thus POI. Of course, there is no guarantee that such values do exist. If they do not, our method of constructing a POI test breaks down, although we are still left with the option of constructing an approximately POI test. There is also the question of how γ_1 should be chosen. Because of the exploratory nature of this study, the empirical approach is used to recommend an appropriate γ_1 value for practical use. An obvious alternative is to choose the γ_1 value that gives a predetermined level of power at $\gamma = \gamma_1$, of say 0.5 or 0.65.

After some experimentation, it became obvious that for any fixed γ_1 and ρ_1 values, (30) is a well behaved function of ρ, typically increasing and then decreasing as ρ increases from zero. The ρ value at which (30) is maximized holds the key. If this can be made to coincide with the choice of ρ_1, then (28) and (29) with $c'_\alpha = c_\alpha$ will hold simultaneously.

One method of iterating to the required ρ_1 and c'_α values is as follows:

(i) Pick a ρ_1 value and solve (28) for c_α.

(ii) For $c'_\alpha = c_\alpha$, evaluate (30) at ρ values around ρ_1. If (30) is a maximum at $\rho = \rho_1$, we have the required solution.

(iii) Otherwise move ρ_1 towards the ρ value which maximizes (30) and solve (28) for c_α. Beginning at (ii), repeat the process.

Because one may not always wish to tackle the amount of computation involved in solving for c_α and ρ_1, it is convenient to have available, solved values for an X matrix which is representative in some sense. Such values might be

used to apply an approximate test. At the very least, they should provide good starting values for the iterative process of solving for c_α and ρ_1. Our choice of representative X matrices are those comprised of the eigenvectors associated with the k smallest eigenvalues of the $n \times n$ matrix

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

Dubbelman (1972) and Dubbelman, Louter and Abrahamse (1978) present empirical evidence which suggests that in economic time-series regression analysis, it is often likely that the space spanned by such eigenvectors is a good approximation to the column space of the X matrix. Similar views have been expressed by Theil and Nagar (1961), Hannan and Terrell (1968), Abrahamse and Koerts (1971) and Abrahamse and Louter (1971). The approximation is likely to be particularly good if the regression has an intercept and if the regressors are few in number and smoothly evolving.

Values of c'_α and ρ_1 at the five per cent significance level for selected $n \times k$ matrices comprised of the eigenvectors associated with the k smallest eigenvalues of A_1 and for $\gamma_1 = 0.5$ and $\gamma_1 = 0.75$ are presented in Tables I and II, respectively. They were computed using a modified version of Koerts and Abrahamse's (1969) FQUAD subroutine.

In the remainder of this paper we shall denote by $r_\rho(\gamma_1)$ the test of H_0 against H_a that is MPI in the neighbourhood of $\gamma = \gamma_1$.

TABLE I

Calculated values of ρ_1 and c_α for the $r_\rho(0.5)$

test at the five per cent significance level

	$k=2$		$k=3$		$k=4$		$k=5$		$k=6$	
n	ρ_1	c_α	ρ_1	c_α	ρ_1	c_α	ρ_1	c_α	ρ_1	c_α
15	.425	.838	.485	.798	.593	.721	.820	.572	.999	.470
16	.423	.844	.475	.809	.568	.743	.755	.651	.940	.503
17	.420	.850	.468	.818	.547	.762	.695	.657	.900	.528
18	.418	.856	.460	.827	.530	.777	.655	.688	.860	.554
19	.416	.861	.455	.834	.517	.790	.620	.715	.815	.585
20	.416	.865	.451	.841	.506	.802	.595	.736	.760	.622
22	.414	.873	.444	.852	.489	.820	.557	.770	.671	.687
24	.412	.880	.438	.863	.476	.836	.530	.795	.616	.731
26	.411	.887	.434	.871	.466	.848	.512	.814	.579	.763
28	.410	.893	.431	.878	.459	.858	.498	.829	.552	.788
30	.410	.898	.428	.885	.453	.867	.486	.842	.531	.808
33	.409	.905	.425	.894	.446	.878	.474	.858	.510	.830
36	.409	.911	.423	.901	.441	.888	.464	.870	.494	.847
40	.408	.918	.421	.909	.436	.898	.455	.884	.479	.865
45	.409	.926	.419	.918	.432	.909	.447	.897	.466	.883
50	.409	.933	.418	.926	.428	.918	.442	.908	.457	.896
55	.409	.938	.417	.932	.426	.925	.438	.916	.450	.906
60	.410	.943	.416	.938	.425	.931	.434	.924	.446	.915
65	.410	.948	.416	.943	.424	.937	.432	.930	.442	.922
70	.410	.951	.416	.947	.422	.942	.430	.936	.439	.929
80	.411	.958	.416	.954	.421	.950	.428	.945	.434	.939
90	.411	.964	.416	.960	.420	.959	.426	.952	.432	.947
100	.412	.969	.416	.965	.420	.962	.424	.958	.430	.954

TABLE II

Calculated values of ρ_1 and c_α for the $r_\rho(0.75)$

test at the five per cent significance level

	$k=2$		$k=3$		$k=4$		$k=5$		$k=6$	
n	ρ_1	c_α	ρ_1	c_α	ρ_1	c_α	ρ_1	c_α	ρ_1	c_α
15	.540	.738	.641	.674	.851	.545	.976	.470	.999	.458
16	.537	.751	.628	.693	.815	.575	.965	.483	.999	.463
17	.535	.762	.617	.710	.780	.604	.960	.495	.999	.469
18	.534	.773	.607	.725	.749	.632	.953	.507	.999	.476
19	.534	.783	.600	.739	.722	.657	.930	.527	.999	.483
20	.533	.793	.593	.752	.700	.680	.910	.547	.999	.489
22	.532	.811	.584	.775	.670	.715	.840	.603	.982	.512
24	.533	.827	.577	.796	.650	.744	.780	.654	.953	.542
26	.533	.842	.572	.814	.632	.770	.736	.696	.910	.580
28	.534	.855	.568	.830	.620	.792	.705	.730	.855	.626
30	.535	.868	.566	.845	.612	.810	.683	.757	.805	.670
33	.536	.885	.563	.865	.601	.835	.658	.792	.747	.726
36	.538	.901	.561	.882	.594	.857	.640	.821	.710	.767
40	.539	.919	.560	.903	.587	.881	.625	.851	.677	.809
45	.542	.939	.559	.925	.582	.906	.611	.882	.651	.850
50	.544	.957	.559	.944	.578	.928	.602	.908	.634	.881
55	.546	.973	.559	.961	.576	.947	.596	.629	.622	.907
60	.547	.987	.559	.976	.574	.963	.592	.948	.614	.929
65	.549	.999	.560	.989	.573	.978	.589	.964	.608	.947
70	.551	1.010	.560	1.002	.572	.991	.586	.979	.602	.964
80	.554	1.030	.562	1.022	.572	1.013	.583	1.003	.597	.991
90	.556	1.047	.563	1.040	.571	1.032	.581	1.023	.592	1.013
100	.558	1.061	.564	1.055	.571	1.048	.580	1.040	.589	1.031

5. AN EMPIRICAL POWER COMPARISON

In order to assess the small-sample performance of the $r_\rho(\gamma_1)$ test as well as to help with the choice of γ_1 value, the tests' sizes and powers were computed at the five per cent significance level for $\gamma_1 = 0.5$ and 0.75 and compared with those of King's (1983a) $g(0.75, 0.75)$ and $g(1/3, 0.3)$ tests. For computational reasons, King originally assumed an approximate AR(1) process under H_0 but following King and McAleer, we shall use the $g(\gamma_1, \rho_1)$ tests which assume true AR(1) disturbances under H_0. In their empirical power comparison, King and McAleer found that the $g(0.75, 0.75)$ test generally dominates the Cox and linearized Cox tests and, of the tests they considered, was the best overall. The $g(1/3, 0.3)$ test was included in the current comparison because King (1983a, p.49) recommended its use when "one strongly suspects a value of γ between 0.3 and 0.5 under H_a." Both tests can be regarded as approximately POI.

The sizes and powers were computed at $\rho = 0.0, ..., 0.9$ under H_0 and $\gamma = 0.1, ..., 0.9$ under H_a using the methodology outlined by King but with true rather than approximate AR(1) disturbances. The following design matrices chosen by King were used in the comparison.

X1: $(n \times 3;\ n = 15, 30, 60)$. The first n observations of Durbin and Watson's (1951, p.159) consumption of spirits example.

X2: $(n \times 3;\ n = 15, 30, 60)$. A constant dummy, the quarterly Australian Consumers' Price Index commencing 1959(1) and the same index lagged one quarter as regressors.

X3: $(n \times 5;\ n = 15, 30, 60)$. X2 with the Australian Consumers' Price Index lagged two and three quarters as additional regressors.

X4: $(n \times 3;\ n = 15, 30)$. A constant and logarithms of Chow's (1957, Table 1) automobile stock per capita and personal money stock per capita variables for the United States 1921-1950.

These design matrices were also used to assess the usefulness of the approximate $r_\rho(0.5)$ and $r_\rho(0.75)$ tests which result from using the ρ_1 values and nominal five per cent significance points from Tables I and II, respectively. In each case, the maximum probability of rejecting H_0 when H_0 is true was found,

this being the true size of the test. These results are presented in Table III and are discussed first.

TABLE III

True sizes of approximate $r_\rho(0.5)$ and $r_\rho(0.75)$ tests applied using ρ_1 and five per cent critical values from Tables I and II.

	approx. $r_\rho(0.5)$ test			approx. $r_\rho(0.75)$ test		
$n=$	15	30	60	15	30	60
Design matrix						
$X1(n \times 3)$	.025	.039	.047	.028	.040	.048
$X2(n \times 3)$	.028	.039	.045	.034	.046	.057
$X3(n \times 5)$	.000	.020	.035	.008	.030	.052
$X4(n \times 3)$	.026	.039	–	.029	.039	–

With one exception, the true sizes move towards the nominal size of 0.05 as the sample size increases or the number of regressors decreases, *ceteris paribus*. Again with one exception, the $r_\rho(0.75)$ sizes are closer to 0.05 than the corresponding $r_\rho(0.5)$ sizes. Also the true sizes are almost always below the nominal size. At least for the particular X matrices of this study, it seems that the approximate $r_\rho(0.75)$ test has reasonable sizes for samples of 60 observations or more as does the approximate $r_\rho(0.5)$ test when $k = 3$.

Selected sizes and powers for design matrices X2 and X3 calculated using exact five per cent critical values may be found in Tables IV and V.

Under H_0 the tests' size functions are well behaved, first increasing as ρ moves away from zero, and then, after reaching 0.05, decreasing as ρ continues to increase. The point of inflection, which is of course ρ_1, always occurs at a larger ρ value for the $r_\rho(0.75)$ test than for $r_\rho(0.5)$. For both tests, the point of inflection tends to decrease as the sample size increases, more so for $r_\rho(0.5)$ than for $r_\rho(0.75)$. For $\rho > 0.5$, the power functions of all tests decrease towards zero as n increases. The desirability of this property is discussed by King (1983a).

TABLE IV

Calculated sizes and powers for X2 at the five per cent significance level.

	$H_0 : u_t = \rho u_{t-1} + \varepsilon_t, \rho \geq 0$					$H_a : u_t = \varepsilon_t + \gamma \varepsilon_{t-1}, \gamma > 0$				
$\rho =$	0.9	0.7	0.5	0.3	0.1	$\gamma = 0.1$	0.3	0.5	0.7	0.9
Tests				$n = 15$						
$r_\rho(0.5)$	.036	.045	.050	.047	.036	.038	.071	.141	.252	.344
$r_\rho(0.75)$	.042	.049	.049	.042	.030	.031	.062	.133	.265	.401
$g(0.75, 0.75)$	.035	.044	.050	.048	.038	.040	.071	.137	.255	.371
$g(1/3, 0.3)$	.039	.047	.050	.044	.031	.032	.066	.137	.245	.327
				$n = 30$						
$r_\rho(0.5)$	.012	.030	.048	.047	.029	.032	.094	.265	.541	.706
$r_\rho(0.75)$	.023	.043	.050	.039	.023	.025	.073	.232	.588	.869
$g(0.75, 0.75)$	.009	.024	.044	.050	.038	.041	.094	.239	.531	.762
$g(1/3, 0.3)$	.008	.022	.043	.050	.036	.039	.102	.251	.447	.556
				$n = 60$						
$r_\rho(0.5)$	.002	.017	.046	.045	.021	.024	.126	.482	.864	.957
$r_\rho(0.75)$	.009	.038	.049	.033	.014	.016	.084	.403	.910	.998
$g(0.75, 0.75)$	.001	.013	.041	.049	.032	.035	.124	.429	.865	.977
$g(1/3, 0.3)$	.000	.006	.031	.050	.032	.037	.149	.431	.703	.801

TABLE V

Calculated sizes and powers for X3 at the five per cent significance level.

	$H_0 : u_t = \rho u_{t-1} + \varepsilon_t, \rho \geq 0$					$H_a : u_t = \varepsilon_t + \gamma \varepsilon_{t-1}, \gamma > 0$				
$\rho =$	0.9	0.7	0.5	0.3	0.1	$\gamma = 0.1$	0.3	0.5	0.7	0.9
Tests				$n = 15$						
$r_\rho(0.5)$	.044	.048	.050	.046	.036	.038	.064	.112	.179	.228
$r_\rho(0.75)$	.047	.050	.048	.040	.029	.030	.056	.107	.186	.252
$g(0.75,0.75)$	.045	.049	.049	.043	.033	.034	.060	.110	.185	.247
$g(1/3,0.3)$	.049	.050	.045	.036	.023	.024	.050	.102	.176	.231
				$n = 30$						
$r_\rho(0.5)$	.015	.033	.048	.046	.030	.033	.090	.242	.492	.652
$r_\rho(0.75)$	.026	.044	.050	.039	.024	.025	.071	.214	.535	.815
$g(0.75,0.75)$	.011	.027	.045	.049	.038	.041	.090	.221	.484	.709
$g(1/3,0.3)$	.011	.026	.045	.049	.035	.038	.097	.233	.419	.528
				$n = 60$						
$r_\rho(0.5)$	.002	.017	.046	.045	.021	.024	.123	.460	.841	.945
$r_\rho(0.75)$	.010	.038	.049	.034	.015	.016	.083	.385	.891	.996
$g(0.75,0.75)$	.001	.012	.040	.049	.033	.037	.122	.408	.839	.968
$g(1/3,0.3)$	.000	.006	.031	.050	.033	.038	.145	.411	.674	.773

Under H_a, the power functions of all tests increase as γ increases. All tests have power less than 0.05 at $\gamma = 0.1$ and occasionally also at $\gamma = 0.2$. This is not a particularly troublesome feature because AR(1) and MA(1) processes are good approximations to each other for ρ and γ close to zero. Also the Monte Carlo study by Griffiths and Beesley (1984) suggests that, wrongly assuming an AR(1) process when the disturbances of (10) follow an MA(1) process with small $|\gamma|$, improves the efficiency of the β estimates. Except for $\gamma = 0.1$ and, in the case of $r_\rho(0.75)$, $\gamma = 0.2$, the power functions of all tests increase as n increases or k decreases, *ceteris paribus*.

The power functions of the two $r_\rho(\gamma_1)$ tests always cross somewhere between $\gamma = 0.6$ and $\gamma = 0.7$ with the $r_\rho(0.5)$ test being more powerful for small γ values and $r_\rho(0.75)$ being more powerful for $\gamma \geq 0.7$. The results suggest that $r_\rho(0.75)$ is the better overall test when $n = 15$, 30 because the gains in power generally far outweigh the losses. For example, the maximum power differences in favour of the $r_\rho(0.75)$ test are 0.073 and 0.158 for the 15×3 and 30×3 X1 matrices, respectively, while the corresponding differences in favour of $r_\rho(0.5)$ are 0.009 and 0.035. The same reasoning leads one to conclude that $r_\rho(0.5)$ is the better overall test when $n = 60$. This test has a maximum power advantage for the 60×3 X1 matrix of 0.084 while that in favour of the $r_\rho(0.75)$ test is 0.051.

Similarly, the $g(0.75, 0.75)$ test has better overall power than the $g(1/3, 0.3)$ test. Typically, the latter test has a slight power advantage for $0.2 \leq \gamma \leq 0.5$ although this is not the case for the X2 and X3 matrices with $n = 15$. This power advantage is clearly negated by the large power loss which occurs for $\gamma \geq 0.7$ and is generally greater than 0.1 for $n = 30$ and 60.

The $r_\rho(0.75)$ test is typically more powerful than the $g(0.75, 0.75)$ test for $\gamma \geq 0.6$ with the reverse being true for smaller γ values. At least for $n = 15$ and 30, one can argue that the $r_\rho(0.75)$ test has better overall power because the gains more than compensate for the losses, although for the 15×5 X3 matrix and, to a lesser extent, the 15×3 X2 matrix, there is very little difference between the two power curves. It is not obvious which is the better test when $n = 60$. The $g(0.75, 0.75)$ test might be marginally preferred for the 60×3 X1

and X2 matrices but not for the 60×5 X3 matrix. In one sense, this is reassuring because $g(0.75, 0.75)$ can be regarded as an approximate point optimal test.

The $r_\rho(0.5)$ test is almost always more powerful than the $g(0.75, 0.75)$ test for $0.4 \leq \gamma \leq 0.6$ and frequently also for $\gamma = 0.3$ and 0.7. Except when $n = 60$, one may wish to favour the $g(0.75, 0.75)$ test because of its greater power for larger γ values, especially $\gamma = 0.9$.

Thus, assuming all values of γ are equally likely, the best of the four tests would appear to be the $r_\rho(0.75)$ test when $n = 15$ or 30 and $r_\rho(0.5)$ when $n = 60$. If, however, one strongly suspected that under H_a γ might lie in a particular range, then it would make sense to apply the $r_\rho(\gamma_1)$ test with γ_1 chosen to be the mid-point or perhaps slightly beyond the mid-point.

In theory, the approach outlined in sections 2 and 4 could be used to construct tests of higher-order autoregressive disturbances in (10) against higher-order moving average disturbances. However, it does seem that prior information about the signs of the parameters[5] of the moving average processes being tested is essential for the resultant test to have useful power properties. For example, one might expect the approach to provide a good test of the null hypothesis that the disturbances of (10) are generated by the special AR(5) process (26) where $0 \leq \rho_1 < 1$ and $0 \leq \rho_4 < 1$ against the analogous MA(5) process with similar restrictions on its parameters. On the other hand, the approach is not recommended if one wishes to test for general AR(j) disturbances in (10) against general MA(j) disturbances.

6. CONCLUDING REMARKS

The computer revolution has meant that we can now begin to ask, what kind of test procedure would we like to apply, rather than, what procedure can

[5] Prior information about the signs of the low-order autocorrelation coefficients of the processes being tested might also satisfy this requirement.

we apply? When, as is usually the case, no UMP test exists, there is no simple answer to the former question. In this paper we have put the case for the inclusion of point optimal and approximately point optimal tests in the econometrician's repertoire. They seem best suited to testing problems which involve a small number of parameters and prior knowledge of the signs of the parameters under the alternative hypothesis. In econometric applications, knowledge of the parameter's signs often can be deduced from the underlying economic theory. Also, the number of parameters in a testing problem, particularly if it involves a linear model, can sometimes be reduced by the use of invariance or by conditioning on sufficient statistics. The available evidence concerning this category of testing problem indicates that point optimal and approximately point optimal tests can have extremely useful small-sample power properties. A point optimal test is UMP when such a test exists. They can also result in tests which are UMP in a subspace of the alternative hypothesis parameter space. However, it is plain that they do not suit all testing problems. Also, little is known about their performance in problems involving a moderate or large number of parameters.

A possible objection to their use is that they require the researcher to nominate the point at which power is to be maximized . As noted in the introduction, any choice of test when a UMP test does not exist will frequently favour certain parts of the parameter space over other parts of the space. Worse still, we are often ignorant of which parts of the space our chosen test favours. By using a point optimal test, the researcher is forced to think about where he wants his test to have best possible power and to justify his choice. If good power is required over a wide range of the parameter space, then the researcher should consider checking the small-sample properties of his chosen test, whether it be a point optimal or some other test. As indicated in section 3.3, a good deal of recent research has focussed on making less arbitrary choices of point. Hopefully, we can look forward to further developements in this area in the future.

As well as providing possible tests, point optimal tests have another use in that they allow one to trace out the maximum attainable power for each point in the alternative hypothesis parameter space. By comparing two such

power envelopes, one can uncover the potential loss of power that results from restricting attention to similar tests or tests with a particular invariance property. For example, the familiar F test of parameter restrictions in the linear regression model is UMP within the class of invariant tests where invariance is with respect to four different transformations (see Seber (1980, p.35)), some of which may seem unnatural. The power envelope calculations reported by King and Smith (1986) give some indication of the potential improvement in power that might result from a partial relaxation of the invariance requirement. A further use might be to assess the potential of a new or currently favoured test by comparing its small-sample power with that of the power envelope.

This paper reports some progress in the search for a complete theory of point optimal testing. We find that in some cases point optimal tests can be constructed from tests of a simple null hypothesis against a simple alternative. For nested or non-nested testing problems involving a wide range of models of interest to the econometrician, we show how one can check whether a point optimal test can be constructed in this way. When it can, exact critical values can be computed. When it cannot, one may wish to consider approximately point optimal tests.

As an illustration, the approach was applied to the problem of testing for AR(1) disturbances against MA(1) disturbances in the linear regression model with encouraging results. This is a non-nested testing problem for which the Cox and related tests have been found to perform poorly in small samples.

ACKNOWLEDGEMENTS

The research reported in this paper was supported by the Australian Research Grants Scheme and the ESRC under grant HR8323. An earlier draft was written while I was visiting the Department of Economics at the University of Southampton. I am most grateful to Nicole Avery, Beryl Leavesley and Phillip Wright for research assistance, to Tom Hettmansperger, Grant Hillier, Esfandiar

Maasoumi, Jukka Nyblom, Dale Poirier and four referees for helpful suggestions and to Gene Savin for coining the phrase "point optimal test".

REFERENCES

Abrahamse, A.P.J. & Koerts, J., (1971). New estimators of disturbances in regression analysis. Journal of the American Statistical Association, 66, 71-74.

Abrahamse, A.P.J. & Louter, A.S., (1971). On a new test for autocorrelation in least squares regression. Biometrika, 58, 53-60.

Berenblut, I.I. & Webb, G.I., (1973). A new test for autocorrelated errors in the linear regression model. Journal of the Royal Statistical Society B, 35, 33-50.

Bhargava, A., (1986). On the theory of testing unit roots in observed time series. Review of Economic Studies, 53, 369-384.

Bhargava, A., Franzini, L. & Narendranathan, W., (1982). Serial correlation and the fixed effects model. Review of Economic Studies, 49, 533-549.

Box, G.E.P., (1980). Sampling and Bayes' inference in scientific modelling and robustness. Journal of the Royal Statistical Society A, 143, 383-430.

Breusch, T.S. & Pagan, A.R., (1984). Model evaluation by differencing transformations. Paper presented at a conference in honour of Denis Sargan.

Chow, G.C., (1957). Demand for Automobiles in the United States. Amsterdam: North Holland.

Cox, D.R., (1961). Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1, 105-123.

Cox, D.R., (1962). Further results on tests of separate families of hypotheses. Journal of the Royal Statistical Society B, 24, 406-424.

Cox, D.R. & Hinkley, D.V., (1974). Theoretical Statistics. London: Chapman and Hall.

Dastoor, N.K., & Fisher, G., (1988). On point-optimal Cox tests. Econometric Theory, 4, forthcoming.

Davies, R.B., (1969). Beta-optimal tests and an application to the summary evaluation of experiments. Journal of the Royal Statistical Society B, 31, 524-538.

Davies, R.B., (1980). Algorithm AS155: The distribution of a linear combination of χ^2 random variables. Applied Statistics, 29, 323-333.

Doan, T., Litterman, R. & Sims, C., (1984). Forecasting and conditional projection using realistic prior distributions. Econometric Reviews, 3, 1-100.

Dubbelman, C., (1972). A priori fixed covariance matrices of disturbance estimators. European Economic Review, 3, 413-436.

Dubbelman, C., Louter, A.S. & Abrahamse, A.P.J., (1978). On typical characteristics of economic time series and the relative qualities of five autocorrelation tests. Journal of Econometrics, 8, 295-306.

Dufour, J.M. & King, M.L., (1986). Optimal invariant tests for the autocorrelation coefficient in linear regressions with autocorrelated errors. Mimeo, Université de Montréal.

Durbin, J. & Watson, G.S., (1950). Testing for serial correlation in least squares regression I. Biometrika, 37, 409-428.

Durbin, J. & Watson, G.S., (1951). Testing for serial correlation in least squares regression II. Biometrika, 38, 159-178.

Efron, B., (1986). Why isn't everyone a Bayesian? The American Statistician, 40, 1-5.

Evans, M.A. & King, M.L., (1985a). A point optimal test for heteroscedastic disturbances. Journal of Econometrics, 27, 163-178.

Evans, M.A. & King, M.L., (1985b). Critical value approximations for tests of linear regression disturbances. Australian Journal of Statistics, 27, 68-83.

Evans, M.A. & King, M.L., (1985c). Higher order generalisations of first order autoregressive tests. Communications in Statistics: Theory and Methods, 14, 2907-2918.

Evans, M.A. & King, M.L., (1988). A further class of tests for heteroscedasticity. Journal of Econometrics, 37, forthcoming.

Farebrother, R.W., (1980). Algorithm AS153: Pan's procedure for the tail probabilities of the Durbin-Watson statistic. Applied Statistics, 29, 224-227 and 30, 189.

Farebrother, R.W., (1985). Eigenvalue-free methods for computing the distribution of a quadratic form in normal variables. Statistische Hefte, 26, 287-302.

Franzini, L. & Harvey, A.C., (1983). Testing for deterministic trend and seasonal components in time series models. Biometrika, 70, 673-682.

Fraser, D.A.S., Guttman, I. & Styan, G.P.H., (1976). Serial correlation and distributions on the sphere. Communications in Statistics, A5, 97-118.

Griffiths, W.E. & Beesley, P.A.A., (1984). The small sample properties of some preliminary test estimators in a linear model with autocorrelated errors. Journal of Econometrics, 25, 49-61.

Hannan, E.J. & Terrell, R.D., (1968). Testing for serial correlation after least squares regression. Econometrica, 36, 133-150.

Hildreth, C. & Houck, J.P., (1968). Some estimators for a linear model with random coefficients. Journal of the American Statistical Association, 63, 584-595.

Hillier, G.H., (1986). Joint tests for zero restrictions on non-negative regression coefficients. Biometrika, 73, 657-669.

Hillier, G.H., (1987). Classes of similar regions and their power properties for some econometric testing problems. Econometric Theory, 3, 1-44.

Imhof, P.J., (1961). Computing the distribution of quadratic forms in normal variables. Biometrika, 48, 419-426.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H. & Lee, T.C., (1985). The Theory and Practice of Econometrics. 2^{nd} ed. New York: John Wiley.

Kadiyala, K.R., (1970). Testing for the independence of regression disturbances, Econometrica, 38, 97-117.

King, M.L., (1980). Robust tests for spherical symmetry and their application to least squares regression. Annals of Statistics, 8, 1265-1271.

King, M.L., (1981a). The Durbin-Watson test for serial correlation: Bounds for regressions with trend and/or seasonal dummy variables. Econometrica, 49, 1571-1581.

King, M.L., (1981b). Testing for random walk disturbances in the linear regression model. Mimeo, Monash University.

King, M.L., (1983a). Testing for autoregressive against moving average errors in the linear regression model. Journal of Econometrics, 21, 35-51.

King, M.L., (1983b). Testing for moving average regression disturbances. Australian Journal of Statistics, 25, 23-34.

King, M.L., (1984). A new test for fourth-order autoregressive disturbances. Journal of Econometrics, 24, 269-277.

King, M.L., (1985a). A point optimal test for autoregressive disturbances. Journal of Econometrics, 27, 21-37.

King, M.L., (1985b). A point optimal test for moving average regression disturbances. Econometric Theory, 1, 211-222.

King, M.L., (1986). Efficient estimation and testing of regressions with a serially correlated error component. Journal of Quantitative Economics, 2, 231-247

King, M.L., (1987a). 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. London: Rutledge and Kegan Paul, 19-73.

King, M.L., (1987b). An alternative test for regression coefficient stability. Review of Economics and Statistics, 69, 379-381.

King,M.L., (1987c). Testing for fourth-order autocorrelation in regression disturbances when first-order autocorrelation is present. Mimeo, Monash University.

King, M.L. & Hillier, G.H., (1985). Locally best invariant tests of the error covariance matrix of the linear regression model. Journal of the Royal Statistical Society B, 47, 98-102.

King, M.L. & McAleer, M., (1987). Further results on testing AR(1) against MA(1) disturbances in the linear regression model. Review of Economic Studies, 54, 649-663.

King, M.L. & Skeels, C.L., (1984). Joint testing for serial correlation and heteroscedasticity in the linear regression model. Paper presented at the Australasian Meeting of the Econometric Society, Sydney.

King, M.L. & Smith, M.D., (1986). Joint one-sided tests of linear regression coefficients. Journal of Econometrics, 32, 367-383.

Koerts, J. & Abrahamse, A.P.J., (1969). On the Theory and Application of the General Linear Model. Rotterdam: Rotterdam University Press.

LaMotte, L.R. & McWhorter, A., (1978). An exact test for the presence of random walk coefficients in a linear regression model. Journal of the American Statistical Association, 73, 816-820.

Lehmann, E.L., (1959). Testing Statistical Hypotheses. New York: John Wiley.

Lehmann, E.L. & Stein, C., (1948). Most powerful tests of composite hypotheses I. Normal distributions. Annals of Mathematical Statistics, 19, 495-516.

Milan, L., (1984). Testing for the Hildreth-Houck random coefficient model. Unpublished M.Ec. thesis, Monash University.

Nyblom, J., (1986). Testing for deterministic linear trend in time series. Journal of the American Statistical Association, 81, 545-549.

Palm, F.C. & Sneek, J.M., (1984). Significance tests and spurious correlation in regression models with autocorrelated errors. Statistische Hefte, 25, 87-105.

Rosenberg, B., (1973). The analysis of a cross-section of time series by stochastically convergent parameter regression. Annals of Economic and Social Measurement, 2, 399-428.

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

Seber, G.A.F., (1980). The Linear Hypothesis: A General Theory. 2nd ed. London: Griffin.

Shively, T.S., (1986a). An exact test for a stochastic coefficient in a time series regression model. Australian Graduate School of Management, Working Paper 86-008.

Shively, T.S., (1986b). Issues in the analysis of stochastic coefficient regression models. Unpublished Ph.D. thesis, University of Chicago.

Shively, T.S., (1987). An analysis of tests for regression coefficient stability. Mimeo, University of Texas at Austin.

Shively, T.S., Ansley, C.F. & Kohn, R., (1987). Fast evaluation of the distribution of the Durbin-Watson and other invariant test statistics in regression. Mimeo, University of New South Wales.

Spjotvoll, E., (1967). Optimum invariant tests in unbalanced variance components models. Annals of Mathematical Statistics, 38, 422-428.

Theil, H., (1971). Principles of Econometrics. New York: Wiley.

Theil, H. & Nagar, A.L., (1961). Testing the independence of regression disturbances. Journal of the American Statistical Association, 63, 242-251.

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

Watson, M.W. & Engle, R.F., (1985). Testing for regression coefficint stability with a stationary AR(1) alternative. Review of Economics and Statistics, 67, 341-346.

Webb, G.I., (1973). Autocorrelations and the general linear model. Unpublished Ph.D. thesis, The City University, London.