



*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

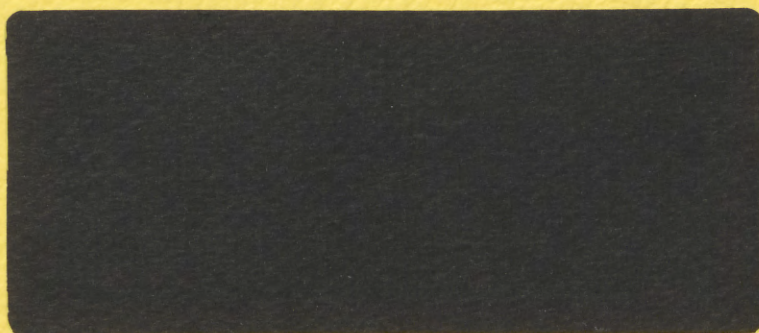


QUEEN'S

#710

ISSN 0316-5078

INSTITUTE FOR ECONOMIC RESEARCH



QUEEN'S UNIVERSITY.



GIANNINI FOUNDATION OF  
AGRICULTURAL ECONOMICS  
LIBRARY

AUG 29 1988

GIANNINI FOUNDATION OF  
AGRICULTURAL ECONOMICS  
LIBRARY

AUG 29 1988

Kingston, Ontario, Canada K7L 3N6

THE THEORY AND PRACTICE OF  
POINT-OPTIMAL TESTING\*

Naorayex K. Dastoor  
University of Alberta

and

Gordon Fisher  
Queen's University

Discussion Paper #710

\*Forthcoming in *Econometric Reviews* as a Comment on  
Maxwell L. King (1988), "Towards a Theory of Point-Optimal  
Testing".

# ABSTRACT

It is argued that the general formulation of point-optimal tests, as summarized in King (1988), (i) makes use of a benchmark which is excessively stringent, (ii) fails to integrate the method with traditional methods of testing, (iii) involves using data to form a specific null hypothesis, (iv) offers no guidance as to proper procedures in the event the null is rejected, (v) eschews the use of Bayesian methods while demanding the use of prior information, and (vi) yields tests with unknown power in the presence of specification error. For these and other reasons, point-optimal testing cannot yet be wholeheartedly recommended as part of "the econometrician's repertoire".

Key Words: Autocorrelation, Cox Principle, Hypothesis Testing, Linear Regression, Point-Optimal Testing, Power, Specification Error



## THE THEORY AND PRACTICE OF POINT-OPTIMAL TESTING

### Preamble

We would like to begin by congratulating Max King for providing a thought-provoking summary of the work on point-optimal testing and various issues concerning its application in econometrics. This type of testing has not yet been widely applied in econometrics and his paper will encourage an appreciation of its limitations and advantages.

King's general attitude to econometric analysis seems consistent with the idea that there should be a change in the perspective from which we view formal systems of inference, by virtue of the availability of modern computing power [Durbin, 1987]. Yet such a change, it has been argued, should also be augmented by the adoption of an ecumenical philosophy as the basis of statistical inference. This would bring together Bayesian and non-Bayesian analyses as complementary means of extracting relevant information from data [Durbin (1987), section 4.3]. King sees point-optimal tests as permitting the extraction of a larger amount of relevant information than would be extracted by using traditional methods on their own. However, he is less ecumenical than he might have been in seeing, indeed seeking, links to Bayesian analysis - more on this later.

A good survey of the literature in a field should have at least two properties: firstly, it should put the development of thought into historical perspective, by highlighting the important contributions, and secondly, it should synthesize the range of literature surveyed. On the matter of historical perspective and important contributions, King has done an excellent job in drawing to the attention of econometricians various papers from the statistical literature. However, he has been less than generous in highlighting the work of Kadiyala (1970) in econometrics. It should be made clear that Kadiyala

was the first to introduce a point-optimal test into econometrics and his 1970 paper still remains, after 18 years, the only paper on the subject to appear in *Econometrica*. Therefore, in terms of historical perspective and importance, Kadiyala's work should be regarded as a signal contribution.

On the matter of synthesis, King has been kind enough in section 3 to mention our paper, Dastoor and Fisher (1988). He notes (p. 16) that we "observed that the test can be interpreted as a special case of the [sic] Cox test of non-nested hypotheses". The quotation unfortunately leaves aside the second aim of Dastoor and Fisher (1988) which was (p. 98) "to synthesize a whole range of point-optimal tests". By drawing a distinction between a general and a point-optimal formulation of a test based on the Cox Principle (McAleer and Pesaran, 1987), we were able to synthesize (in four pages) ten of the point-optimal tests for regression disturbances and to bring these tests and their structure into perspective in relation to more traditional methods of testing. In section 3 of his paper, King provides a summary of the same tests and related work, but falls short of bringing the structure of these into relation with traditional tests.

#### Underlying Principles

In the paper, the theory of point-optimal testing is developed within a very general framework. At one extreme, the null and alternative hypotheses may be completely separate in distributions, in parameters and in parameter subspaces. At the other extreme the hypotheses may have a common distribution, corresponding parameters and nested parameter subspaces. In all cases, the unifying principle is the likelihood ratio statistic of the two models, notwithstanding the fact that this statistic, in other forms of testing, would be suitable for the second type of problem but not the first (see pp. 5-7). The likelihood ratio criterion may be regarded as a special case of the Cox Principle. This states that the null may be tested against the

alternative by considering the difference between some criterion function of the performance of the one relative to the other (such as the likelihood ratio statistic) and its expectation under the null. In certain problems, e.g. normal regressions, this is equivalent to a test based upon the difference between a function of the maximum-likelihood estimator (under the alternative) of the alternative parameters and a consistent estimator of its pseudo-true value under the null (Gourieroux et al., 1983). The same principle also goes under the name of 'encompassing' (Mizon and Richard, 1986). By using an appropriate criterion function in the common-distribution-nested-parameters case, the Cox Principle reduces to the traditional likelihood ratio statistic as a special case. With a point-optimal formulation, when some parameters are specified a priori, the Cox Principle could also lead to point-optimal tests as special cases. Thus a wider class of traditional likelihood ratio and point-optimal tests could be 'encompassed' using the Cox Principle. Moreover, acknowledging the Cox Principle would not lead the author (p. 45) to make the misleading and misdirected statement that "the [sic] Cox and related tests have been found to perform poorly in small samples". For example, Dastoor and Fisher (1988) show that a particular class of point-optimal tests can be obtained as a particular type of Cox test, provided the two are formulated under the same conditions. Hence, there is no such thing as the Cox test; rather, there are many Cox tests based upon alternative formulations of the Cox Principle.

#### Theory and Practical Implications

It is worth observing that the benchmark used by King for designing point-optimal tests is extremely stringent. This benchmark is obtained by evaluating the power of different statistics at different values in the alternative parameter space (i.e. under different hypotheses) whereas the usual power curve considers the same statistic under different hypotheses.

Since power is defined with respect to a particular hypothesis, a test can be powerful for one hypothesis but not another. Normally, then, testing involves a trade-off in power across the alternative parameter space. With the King benchmark, no such trade-off is possible.

In section 2, the null hypothesis is determined by the particular value of  $\omega_1$ . Hence if  $\omega_1$  is varied, so is the null. It could be argued that a value for  $\omega_1$  is chosen by treating it as a parameter to be estimated, equation (2) being the criterion of estimation. Does this not imply that the data are being used to formulate the null hypothesis? Of course, the natural retort to this is that  $\omega_1$  is chosen merely as a representative point of a given composite null. Nevertheless the conundrum remains:  $\omega_1$  is selected by using the data thereby defining a likelihood ratio to test the null which is determined by  $\omega_1$ ! In a similar way, the reader is informed (p. 29) that the choice of  $\theta_1$  "should depend on  $n$  and perhaps  $k$ ". While  $n$  and  $k$  may alter the testing problem by affecting the distribution of the test statistic, what is suggested by the statement is that  $n$  and  $k$  should affect the hypothesis to be tested. Why should this be so? Both the determination of  $\omega_1$  and the choice of  $\theta_1$  seem to be at odds with the accepted principles of statistical inference.

Generally it is argued that point-optimal tests have "reasonable power" for values other than those specified in the alternative parameter space. Unfortunately no formal metric is introduced on the space and so "reasonable" is ill-defined. Evidently what is meant by "reasonable" is subjective. Again, it is stated (pp. 3 and 6) that if the likelihood ratio leads to a MP test at  $\theta = \theta_1$ , then it will be most powerful in the neighbourhood of  $\theta = \theta_1$ . There does not seem to be a justification for this statement (c.f. the construction of the power envelope as a benchmark).

The paper is almost entirely about designing tests, yet a



theory of testing ought to offer guidance on practical procedures. How ought point-optimal tests to be used and what should be done in the event that the null is rejected by a point-optimal test? In the case of serial correlation in a normal linear regression, for example, suppose we begin by being surprised that  $H_0: \rho=0$  is not rejected by the standard DW test. We could take the view that the DW test may lack power and apply a point-optimal test. Let this reject  $H_0$  using the chosen value  $\rho=0.6$ , say. What this could tell us is that  $\rho=0.6$  is a better rule to work with for this set of data than  $\rho=0$ . How do we now proceed? And on what logical ground? Is  $\rho$  now to be estimated or do we stick with  $\rho=0.6$ ? Or do we test again with  $\rho=0.5$  and, having rejected  $H_0$  a second time, find the value of  $\rho$  which maximizes the posterior distribution with rectangular prior over  $0.5 \leq \rho < 1$ ? What is the next logical step after rejecting  $H_0$  and what is its justification?

Notice carefully that when we applied the DW statistic to the original problem, we were firmly in the realm of significance testing. But when we move to point-optimal testing in the same example, we actually practise discrimination; that is, we choose which of two hypotheses is the better for this set of data. In particular, if we test for positive serial correlation when in fact it is negative, point-optimal testing should not reject  $H_0$  because the DGP lies in a direction 'opposite' to the specified alternative.

Going further, if we have such strong views about  $\rho=0.6$  to the extent that this is selected to represent the alternative (we rejected the evidence of the DW test in the first place), why is  $\rho=0.6$  not the null and  $\rho=0$  the alternative? In significance testing,  $\rho=0$  is a point in the maintained parameter space  $0 \leq \rho < 1$ , and using it to represent the null is natural in the sense that it is a restriction on the maintained, every other point being in the alternative parameter space. But there is no such restriction with discrimination. Hence which point

is the null and which the alternative is open to question. In particular it would be possible not to reject  $H_0: \rho=0$  against  $H_\alpha: \rho=0.6$  and also not to reject  $H_0: \rho=0.6$  against  $H_\alpha: \rho=0$ , the test statistic of one being the reciprocal of the test statistic of the other.

What is brought out by this example is not so much the need to clarify practical procedures as the need to clarify the logic of such procedures. The reason there is confusion is that, in point-optimal testing, the test-statistic is not defined for the original hypotheses that are the subject of testing but for a new pair of hypotheses that are intended to represent the original ones. Then the test results are interpreted as if they apply to the original hypotheses although, in logic, they apply to the representative hypotheses which have an independent meaning; and such independent meaning is different from the meaning of the original hypotheses under test.

#### Prior Information

It is stated (pp. 3-4) that when a point-optimal test is used it is necessary to nominate part of the parameter space (under the alternative or under each hypothesis) for the test. This needs very precise information, far more precise than economic theory is normally able to provide. Sometimes economic theory is able to predict signs of coefficients — but often signs turn out to be ambiguous. Generally, point-optimal testing requires far more information than an investigator is able to provide. Is it not more reasonable to suppose that an economist is able to choose a region in the alternative parameter space? Incidentally, King argues (p. 2) that Bayesian analysis implies the adoption of prior distributions that may not be generally acceptable and thereby sets such analysis aside in favour of considering point-optimal testing. But point-optimal testing may be interpreted in the same way, a chosen point having a degenerate prior at that point. Why is the author so confident about the selection of points while

doubting an investigator's ability to choose a region with an uninformative prior? Again, King (p. 11) writes: "It is not clear how one can find  $\lambda$  and conduct the test when  $\lambda$  has some different form." Surely economic theory is more likely to provide information about  $\lambda$  than about  $\omega$ ? e.g. a uniform distribution over a feasible interval of parameter values. It is suggested (p. 28) that a test may be chosen by maximizing "some weighted average of powers". In this case the weighting could be the investigator's prior probabilities about plausible values of the parameter.

Generally the author eschews the use of Bayesian analysis while advocating methods that require the use of exact prior information. He even concludes (p. 26): "While prior information about the sign(s) of the nuisance parameter(s) is not necessary, it does seem that prior information about sign(s) of the parameter(s) being tested is essential for the method to work successfully." Why does the author believe that he can be convincing in choosing points, while he cannot be as convincing in choosing a range and a uniform (or other) prior over it?

#### Specification Error

Many of the claims for point-optimal testing rest on the specific alternative considered, that is, a known alternative without any specification error whatsoever. Many of the properties of traditional tests rest on power properties, not only for the stated alternative, but also for other alternatives arising from specification error that are locally equivalent, or otherwise 'similar'. In this way, even when the assumptions of the test do not hold, the procedure yields substantially the 'right' answer, namely, that the null should or should not be rejected, even though the cause may not be the specific alternative of the test. Without knowing the behaviour of a test in the presence of specification error of different types, it is difficult to pass judgment on a test, let alone on a whole test procedure. This point is very important in economics where

generally we do not have the advantage of controlled experiments and hence, in a fundamental sense, we do not know the truth. Moreover, the concept of a UMP test is only useful when the family of the DGP is considered. Since the DGP is not known, the UMP argument is conditional on the maintained model being the same as the DGP. What happens when the maintained model is not the DGP? In the paper, all comparisons with traditional tests take the maintained model and the DGP to be the same. More comparisons with traditional tests in the presence of specification error are necessary before a reasonable judgment can be made about point-optimal tests in general.

The linear regressions examined in Section 3 assume normality. This assumption and application of the principle of invariance lead to the elimination of nuisance parameters and thereby (using Imhof's procedure) to exact tests involving at most two parameters. What about other models? When do point-optimal tests lose their exactness? Is the method at all useful then? Is the method made viable by invariance, only to become less than practicable when nuisance parameters cannot be eliminated? Also, as the number of parameters to be chosen increases, the problem becomes unmanageable; the author is aware of this shortcoming because he states (p. 43) that "the approach is not recommended if one wishes to test for general  $AR(j)$  in (10) against general  $MA(j)$  disturbances".

It is argued, in particular, that the author's  $X$  matrices are representative of real economic data. It is misleading to advocate widespread use of point-optimal testing when extensions to non-linear models or models which violate the author's basic assumptions have not been considered.

#### Some Specific Comments

There is a certain amount of confusion in the paper over comparisons with more traditional tests, since the argument is couched in terms that make the two seem directly comparable. Thus the author tells us (p. 23) that "POI tests have more

desirable small-sample power properties than their competitors including LBI tests". This is obvious insofar as LBI tests are portmanteau tests (in a particular direction) whereas POI tests concentrate attention on a smaller part of the parameter space (in the same direction).

The author notes (p. 30) that Nyblom found that "Pitman efficiencies are rather insensitive to the choice of  $\alpha$  and  $p_1$  and suggests the use of  $\alpha = 0.05$  and  $p_1 = 0.8$ ". This is a good idea, but how does this test compare with a LBI test? Note that Pitman ARE is an asymptotic concept. Have other asymptotic concepts been considered?

It is proposed to solve equation (17) by iteration. However, this is not really necessary. It is surely easier to work out  $p \equiv \Pr[\sum (\lambda_i - s_c) \xi_i^2 < 0]$ ,  $s_c$  being the calculated value of  $s$ , and then to reject the null if  $p < \alpha$ . Moreover, there is no need to approximate  $s^*$  when  $p$  can be calculated. Is this not the natural path to take with the available computer technology, rather than calculating approximate critical values?

On pp. 19-20, the test takes a special form and is UMPI because the error dispersion has  $k$  eigenvectors which span the column space of  $X$ . It is perhaps worth noting that this condition is sufficient for OLS and GLS estimators to coincide. Thus the error dispersion is only needed to calculate standard errors.

On p. 15, equation (16) is a restricted form of equation (11). Therefore, the statement on p. 27 that "this represents a wider class of tests than those ... given by (11)" is rather puzzling.

The author states (p. 28) that "locally best tests optimize power where high power is needed least". There is a confusion created here between the consequences for estimation and the needs of testing. Being close to the null implies less severe consequences for estimation, but is desirable in testing in order to distinguish, clearly and precisely, the null from other



points in the alternative parameter space. It is precisely this idea which provides an important justification for constructing tests that are LBI.

#### Concluding Remarks

The theory of point-optimal testing can be viewed as a transformation of a general testing problem to a particular one in which the Neyman-Pearson Lemma is directly applicable. This transformation requires the specification of representative points in parameter spaces. If an investigator is able to do this, then clearly the theory of point-optimal testing is useful. Unfortunately, in most cases, economic theory will be unable to provide specific values for parameters. Nonetheless, King's survey should be welcomed for his argument (p. 44) that a point-optimal test forces an investigator "to think about where he wants his test to have best possible power and to justify his choice". However, the requirement that "the researcher should consider checking the small-sample properties of his chosen test", while desirable, is plainly unmanageable in most instances. Finally, in certain cases, the theory of point-optimal testing has a role to play in econometrics but, given the above comments, further research is required before it can be wholeheartedly recommended as part and parcel of the econometrician's repertoire.

Naorayex K. Dastoor  
University of Alberta

Gordon Fisher  
Queen's University

#### ADDITIONAL REFERENCES

Gourieroux, C., A. Monfort and A. Trognon (1983), "Testing nested or non-nested hypotheses", Journal of Econometrics, 21, 83-115.

Durbin, J. (1987), "Statistics and statistical science", Journal of the Royal Statistical Society, Series A, 150, p. 177-191.

McAleer, M. and M.H. Pesaran (1986), "Statistical inference in non-nested econometric models", Applied Mathematics and Computation, 20, 271-311.

Mizon, G.E. and J.-F. Richard (1986), "The encompassing principle and its application to testing non-nested hypotheses", Econometrica, 54, 657-678.

