ISSN 0729-0683

ISBN 0 86746 751 7

ANCORA IMPARO

MONASH UNIVERSITY

THE POWER OF STUDENT'S t TEST:

CAN A NON-SIMILAR TEST DO BETTER?

Maxwell L. King

Working Paper No. 13/88

September 1988

DEPARTMENT OF ECONOMETRICS

ISSN 0729-0683

ISBN 0 86746 751 7

THE POWER OF STUDENT'S t TEST:
CAN A NON-SIMILAR TEST DO BETTER?

Maxwell L. King

Working Paper No. 13/88

September 1988

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS AND POLITICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

THE POWER OF STUDENT'S t TEST: CAN A NON-SIMILAR TEST DO BETTER?

Maxwell L. King
Department of Econometrics, Monash University

Summary

Lehmann and Stein (1948) proved the existence of non-similar tests which can be more powerful than best similar tests. They used Student's problem of testing for a non-zero mean given a random sample from the normal distribution with unknown variance as an example. This raises the question: should we use a non-similar test instead of Student's t test? Questions like this can be answered by comparing the power of the test with the power envelope. This paper discusses the difficulties involved in computing power envelopes. It reports an empirical comparison of the power of the t test and the power envelope and finds that the two are almost identical especially for sample sizes greater than 20. These findings suggest that, as well as being uniformly most powerful (UMP) within the class of similar tests, Student's t test is approximately UMP within the class of all tests. For practical purposes it might also be regarded as UMP when moderate or large sample sizes are involved.

KEY WORDS: Most powerful tests; Power envelope; Similar tests; Uniformly most powerful tests.

1. INTRODUCTION

When testing a composite null hypothesis against a composite alternative, statisticians generally prefer to work with a similar test. Such tests have the same probability of a Type I error for all distributions in the null hypothesis. If a test which is Uniformly Most Powerful (UMP) within the class of similar tests can be found, then this is nearly always regarded as the ideal test. Of course, the class of similar tests is only a subset of the class of all tests. It may be that by restricting the choice of test to those that are similar, some quite powerful tests are being ruled out unnecessarily. In other words, there may be a cost, in terms of power, involved in restricting one's choice of test to a similar test. Also, given that the aim in any testing problem is to minimize the probability of making an error, it may be advantageous to use a test whose probability of a Type I error varies with the null distribution.

In a seminal paper, Lehmann and Stein (1948) proved the existence of non-similar tests which can be more powerful than best similar tests. This raises the question of whether UMP similar (UMPS) tests should continue to be thought of as ideal tests . One of Lehmann and Stein's examples is Student's problem of testing for a non-zero mean given a random sample from the normal distribution with unknown variance. For the one-sided testing problem, Student's t test is UMPS. Lehmann and Stein showed that, provided the level of significance is less than 0.5, Student's t test is not most powerful within the class of all tests. Should we not look for a better test within the class of non-similar tests?

This paper provides an answer to this question by empirically comparing the power of the t test with the power envelope for this problem. The power envelope traces out the maximum (or supremum) of the powers of all tests over the alternative hypothesis parameter space. A small difference between the power curve of the t test and the power envelope would indicate that there is little to be gained by looking for another test. A large difference, on the other hand, would suggest that it may be worthwhile to look for another test. The next section considers the problem of evaluating the power envelope while the results of the comparison are discussed in section 3. They suggest that the t test is approximately UMP within the class of all tests so there is little point in looking for a better test.

2. THEORY

In order to compute the power envelope for testing a composite null hypothesis against a composite alternative, one needs to construct the most powerful (MP) test of the composite null against a simple alternative and compute its power. The simple alternative hypothesis represents the point at which the power envelope is being evaluated. An equivalent approach is to choose a series of points at which the power envelope is to be calculated. For each point, the point optimal test, which optimizes power at that point (see King (1987)), is constructed and its optimal power at the point evaluated.

The existence and form of MP tests of composite null hypotheses against simple alternatives are discussed by Lehmann and Stein (1948) and Lehmann (1959, pp.90-97) but few details are given on how such tests can be constructed. In a survey on point optimal testing, King (1987) explains how in some situations a most powerful test of a composite null hypothesis can be constructed from likelihood ratio tests of simple null and simple alternative hypotheses.

Suppose we wish to test

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

where ω is a vector of parameters restricted to the set Ω, against the simple alternative

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

Note that for the simpler problem of testing

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

against H_a, where $\omega_1 \varepsilon \Omega$ is fixed and known, the Neyman-Pearson lemma implies that rejecting H_0^1 for large values of

$$r(\omega_1) = g(x)/f(x,\omega_1) \tag{1}$$

gives a MP test. The critical value for this test is found by solving

$$\Pr[r(\omega_1) > r' | x \text{ has density } f(x,\omega_1)] = \alpha \tag{2}$$

for r' where α is the desired level of significance.

For testing the composite null, H_0, against H_a, (1) can be used as a test statistic but its critical value must now be found by solving

$$\sup_{\omega\varepsilon\Omega} \Pr[r(\omega_1) > r^* | x \text{ has density } f(x,\omega)] = \alpha \tag{3}$$

for r^*. If an ω_1 value exists such that (2) and (3) hold for $r' = r^*$, then the resultant test is the required MP test of H_0 against H_a. This can be verified by noting that if a more powerful test does exist then as a test of H_0^1 against H_a, it would contradict the Neyman-Pearson lemma. King (1987) discusses how Imhof's (1961) algorithm for computing the distribution function of quadratic forms in normal variables can be used to compute the probabilities required to solve (2) and (3) for a range of statistical models with Gaussian error terms. These include Box-Jenkins time-series models, linear and non-linear regression models, linear dynamic regression models and simultaneous equation models. At present, for most testing problems, the existence of an ω_1 value such that (2) and (3) hold for $r' = r^*$ is uncertain and can only be confirmed by numerical methods. In some cases, however (see for example Lehmann and Stein (1948)), the existence of appropriate ω_1 values can be proved analytically. The problem of interest in this paper, that of testing for a non-zero mean given a random sample from the normal distribution with unknown variance, is one such special case.

Let $x_1, \ldots, x_n$ be independent random variables distributed $N(\mu, \sigma^2)$ where μ and σ^2 are unknown. Let

$$\bar{x} = \sum_{i=1}^{n} x_i/n$$

and

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1) .$$

Consider testing $H_0 : \mu = 0$ against $H_1 : \mu > 0$. The UMPS test for this problem is to reject H_0 for large values of

$$\sqrt{n}\ \bar{x}/s .$$

This statistic has a (central) Student's t distribution with n-1 degrees of freedom under H_0 and a noncentral Student's t distribution with noncentrality parameter $\delta = \sqrt{n}\mu/\sigma$ and n-1 degrees of freedom under H_1.

Suppose we wish to evaluate the power envelope at $\mu = \mu_1$ and $\sigma^2 = \sigma_1^2$. Then

$$g(x) = (2\pi\sigma_1^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_1^2}\Sigma(x_i-\mu_1)^2\right\}$$

and

$$f(x,\sigma^2) = (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\Sigma x_i^2\right\} .$$

For $\sigma^2 = \sigma_0^2$, (1) has the form

$$(\sigma_0^2/\sigma_1^2)^{n/2}\exp\left\{ \frac{1}{2\sigma_0^2}\Sigma x_i^2 - \frac{1}{2\sigma_1^2}\Sigma(x_i-\mu_1)^2\right\} . \qquad (4)$$

Rejecting H_0 for large values of (4) is equivalent to rejecting when

$$\left[\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right] \Sigma x_i^2 - 2\frac{\mu_1}{\sigma_1^2}\Sigma x_i \le c . \qquad (5)$$

In order to find the MP test at $\mu = \mu_1$ and $\sigma^2 = \sigma_1^2$, we need to find an $\omega_1 = \sigma_0^2$ value such that (2) and (3) hold for $r' = r^*$. This requires being able to evaluate the probability of (5) being true when the x_i are independent $N(0,\sigma^2)$ random variables.

Lehmann and Stein (1948, p.507) note that if we set $y_i = x_i/\sigma$, i=1, ..., n, so that y_i are independent N(0,1) random variables, then (5) can be rewritten as

$$\Sigma(y_i - a)^2 \le (n - k)a^2$$

where

$$a = \frac{\mu\sigma_0^2}{\sigma(\sigma_0^2 - \sigma_1^2)} , \qquad k = \frac{-c\sigma_1^2(\sigma_0^2 - \sigma_1^2)}{\mu_1^2\sigma_0^2} .$$

Thus, the required probabilities can be calculated by noting that $\Sigma(y_i-a)^2$ has a noncentral chi-squared distribution with n degrees of freedom and noncentrality parameter $\delta = \sqrt{n}a$. Lehmann and Stein prove the existence of an appropriate $\omega_1 = \sigma_0^2$ value such that $\sigma_0^2 > \sigma_1^2$.

The following numerical procedure was used to find $\omega_1 = \sigma_0^2$ such that (2) and (3) hold for $r' = r^*$:

(i) Choose a σ_0^2 value such that $\sigma_0^2 > \sigma_1^2$.

(ii) For $\sigma = \sigma_0$, solve

$$\Pr[\Sigma(y_i - a)^2 \le (n - k)a^2] = \alpha \qquad (6)$$

for c where α is the desired significance level.

(iii) Check whether

$$\sup_{\sigma^2} \Pr[\Sigma(y_i - a)^2 \le (n - k)a^2] = \alpha \qquad (7)$$

holds. If it does the $\omega_1 = \sigma_0^2$ value has been found. If not, change σ_0^2 by moving it closer to the σ^2 values which violate (7) and repeat (ii) and (iii).

The noncentral chi-squared distribution function was calculated using the IMSL subroutine MDCHN.

Once σ_0^2 has been determined, the power of the resultant test when x_i are independent $N(\mu_1,\sigma_1^2)$ random variables can be computed by noting that then

$$\Sigma(y_i - a)^2$$

has a noncentral chi-squared distribution with n degrees of freedom and noncentrality parameter $\sqrt{n}(\mu_1/\sigma_1 - a)$.

3. POWER CALCULATIONS

Using the method described above, the power envelope (PE) was computed and compared with the power curve of the one-sided Student's t test at three levels of significance, namely $\alpha = 0.01, 0.05, 0.1$. This was done at $\sigma_1^2 = 1.0$ and $\mu_1 = 0.1, 0.2, 0.3, \ldots, 1.0$ for $n = 10, 20$; $\mu_1 = 0.1, 0.2, 0.3, \ldots, 0.8$ for $n = 30$; $\mu_1 = 0.05, 0.1, 0.15, \ldots, 0.5$ for $n = 50, 70$; and $\mu_1 = 0.05, 0.1, 0.15, \ldots, 0.4$ for $n = 100$. Only one value of σ_1^2 was used because it can be shown that both the power of the t test and the power envelope depend only on μ/σ under H_1. The power of the t test was evaluated using Bulgren and Amos' (1968, pp.1016-1018) algorithm for computing values of the doubly noncentral t distribution function with the second noncentrality parameter set to zero. The program used, reproduced Bulgren and Amos' table of doubly non-central t distribution function values exactly to six decimal places.

The results of the power comparison for n = 10 and 20 are given in Table 1. They show that there is very little difference between the power envelope and the power of the t-test. The difference is greatest for small n and low significance levels. It quickly disappears as n increases and for $n \geq 30$, the maximum difference is less than 0.005 for all three significance levels. The fact that the difference decreases as the significance level increases is consistent with Lehmann and Stein's finding that a more powerful test exists only if the significance level is less than 0.5.

These results suggest that, at the significance levels typically used in practice, Student's t test is approximately UMP within the class of all tests. For practical purposes it might also be regarded as UMP when moderate or large sample sizes are involved.

It may be wrong to assume that the above conclusions also apply to other cases involving UMPS or UMP invariant tests. An example of a wide gap between a power envelope and a UMP invariant test is given by King and Smith (1986). They considered testing linear restrictions on coefficients in the linear regression model. Here the standard F-test is UMP within the class of tests invariant to a set of four transformations on the regressand (see for example Seber (1980, p.35)). King and Smith used the point optimal approach to trace out the power envelope for the class of tests invariant to two of the four transformations. For a range of regression models they found that the power of the F-test was almost never above 95 per cent of the power envelope and reported power differences as high as 0.4.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Australian Research Council. The author is grateful to Yeoh Teow Eng and Simone Grose for research assistance.

REFERENCES

Bulgren, W.G. & Amos, D.E. (1968). A note on representations of the doubly non-central t distribution. *J. Amer. Statist. Assoc.* 63, 1013-1019.

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

King, M.L. (1987). Towards a theory of point optimal testing. *Econometric Reviews* 6, forthcoming.

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

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. *Ann. Math. Statist.* 19, 495-516.

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

Table 1: Comparison of the power of Student's t test with the power envelope for n = 10, 20 and α = 0.01, 0.05, 0.1.

μ/σ =	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
n = 10										
α = 0.01										
PE	.020	.038	.066	.108	.167	.243	.335	.437	.543	.645
t	.020	.037	.065	.107	.165	.240	.331	.431	.537	.639
α = 0.05										
PE	.088	.145	.223	.318	.429	.545	.657	.756	.838	.899
t	.088	.145	.222	.317	.427	.543	.655	.754	.836	.898
α = 0.1										
PE	.164	.249	.354	.471	.590	.702	.797	.871	.924	.958
t	.164	.249	.353	.470	.589	.701	.796	.870	.923	.958
n = 20										
α = 0.01										
PE	.028	.068	.141	.254	.403	.567	.720	.840	.921	.966
t	.028	.068	.140	.254	.402	.566	.719	.839	.920	.965
α = 0.05										
PE	.113	.217	.363	.532	.696	.827	.915	.964	.987	.996
t	.112	.217	.363	.532	.695	.827	.915	.964	.987	.996
α = 0.1										
PE	.199	.342	.512	.680	.818	.910	.962	.987	.996	.999
t	.199	.342	.512	.680	.817	.910	.962	.987	.996	.999

Recent Working Papers

(Available on request)

7/87 A.R. Tremayne, "A Synthesis of Econometrics and Time Series Methods to Model and Forecast UK Demands for Imports".

8/87 Merran A. Evans and Maxwell L. King, "A Further Class of Tests for Heteroscedasticity".

9/87 Asraul Hoque, "Size of Holding and efficiency in Bangladesh: An Application of a Random Coefficient Model".

10/87 Brett A. Inder, "Bias in the Ordinary Least Squares Estimator in the Dynamic Linear Regression Model with Autocorrelated Disturbances".

11/87 Russel J. Cooper and Keith R. McLaren, "Approaches to the Solution of Intertemporal Consumer demand Models".

12/87 Esme Preston, "On Choosing a Technical Wordprocessor".

1988

1/88 John Preston and Esme Preston, "Two Papers on Simple Word Processors for the PC".

2/88 Esme Preston, "Multilingual and Mathematical Text Processing".

3/88 Esme Preston and John Preston, "The Accountant's PC: Getting Started".

4/88 Maxwell L. King, "Towards a Theory of Point Optimal Testing".

5/88 Ralph D. Snyder, "Statistical Foundations of Exponential Smoothing".

6/88 Grant H. Hillier, "On the Interpretation of Exact Results for Structural Equation Estimators".

7/88 Kuldeep Kumar, "Some Recent Developments in Non-Linear Time Series Modelling.

8/88 Maxwell L. King, "Testing for Fourth-Order Autocorrelation in Regression Disturbances When First-Order Autocorrelation is Present".

9/88 Ralph D. Snyder, "Kalman Filtering with Partially Diffuse Initial Conditions".

10/88 Asraul Hoque, "Indirect Rational Expectations and Estimation in a Single Equation Macro Model".

11/88 Asraul Hoque, "Efficiency of OLS Relative to C-O for Trended x and Positive Autocorrelation Coefficient".