CLASSICAL AND BAYESIAN HYPOTHESIS TESTING: A COMPROMISE

Richard E. Quandt

Econometric Research Program
Research Memorandum No. 279

April 1981

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, New Jersey

Classical and Bayesian Hypothesis Testing: A Compromise*

by

Richard E. Quandt

1. Introduction

Classicists and Bayesians are sharply divided on the question of hypothesis testing. Consider a random variable y with pdf $f(y|\theta)$ and the testing of a sharp null hypothesis H_0: $\theta = 0$ against a nonsharp, composite alternative H_1: $\theta \neq 0$. A traditional classical test procedure may be based on the sampling distribution of the sufficient statistic $\hat{\theta}$, $g(\hat{\theta}|\theta)$. A critical value c is chosen and the test procedure is to

$$\begin{array}{lll} \text{Reject } H_0 & \text{if } |\hat{\theta}| > c \\ \text{Accept } H_0 & \text{otherwise} \end{array} \qquad (1\text{-}1)$$

In acting according to the rule (1-1) two types of error may be committed: the hypothesis may be rejected even though it is true (Type I) or it may be accepted even though it is false (Type II). Characteristically, c is chosen so as to make the probability of Type I error, α , of fixed size:

$$\alpha(c) = 1 - \int_{-c}^{c} g(\hat{\theta}|0)\,d\hat{\theta} \qquad (1\text{-}2)$$

---

* In am indebted to David Brownstone, Gregory Chow, Cheng Hsiao and Kenneth Small for helpful comments and to the National Science Foundation for financial support.

Bayesian analysis requires a prior distribution over θ which may be stated in the present problem as

$$p(\theta) = \begin{cases} \lambda & \text{if } \theta = 0 \\ (1-\lambda)h(\theta) & \text{otherwise} \end{cases} \tag{1-3}$$

where λ is simply the prior probability that $\theta = 0$. Employing Bayes' Theorem one may then compute the posterior pdf and more specifically the posterior odds ratio

$$R = \frac{P(H_1|\hat{\theta})}{P(H_0|\hat{\theta})} = \frac{(1-\lambda)\int_{\theta\neq 0} g(\hat{\theta}|\theta)h(\theta)d\theta}{\lambda g(\hat{\theta}|0)} \tag{1-4}$$

H_0 is then rejected if $R > 1$ and accepted otherwise. If the losses ℓ_1 and ℓ_0 incurred from taking the wrong action (acting as if H_1 were false when in fact it is true, and conversely) are unequal, the decision is based on the ratio

$$R' = \frac{\ell_1 P(H_1|\hat{\theta})}{\ell_0 P(H_0|\hat{\theta})} \tag{1-5}$$

and H_0 is rejected if $R' > 1$.[1]

The Classical approach neglects prior beliefs about the parameter in question and the arbitrary choice of c (or $\alpha(c)$) in (1-2) has the unfortunate consequence that the probability of Type I error remains constant, although the probability of Type II error diminishes with sample size. This implies a peculiar utility function over the probabilities of the two types of errors and bases the decision between the hypotheses purely on the size of the estimated $\hat{\theta}$. By contrast, in the Bayesian approach the size of the sample also matters

---

1. A lucid and much more detailed exposition of these ideas is in Leamer (1978), Ch. 4. See also Zellner (1979).

and in large samples we require a larger departure of $\hat{\theta}$ from H_0 in order to constitute adequate evidence against it.[2] On the other hand, one may reasonably ask over what domain the econometrician's utility function can most reasonably be defined. In the Bayesian approach we are required to state the losses (disutilities) associated with making the wrong decision. Since this has to be done individually for every single hypothesis one may ever test, it may strain one's ability to produce the appropriate loss estimates. It may place less of a burden on the individual to require him to define his utility function over the domain of Type I and Type II probabilities. It must be emphasized that, except in the case of testing a point hypothesis against a point alternative, this represents a departure from the framework of a Bayesian for whom loss depends only on the true state of affairs and on the action taken.

In the remainder of this paper we outline a procedure by which such utility functions can be defined in practice. This will be done in a framework in which prior beliefs about parameter values may be, but need not be, utilized. In Section 2 we introduce the simple example of testing a hypothesis concerning the mean of a normal distribution and discuss the measures over which the utility function will be defined. In Section 3 we introduce the utility function and solve the appropriate constrained utility maximization problem. Section 4 compares the classical and Bayesian approaches with the present one.

2. An Example and Domain of the Utility Function

We assume that the random variable y has pdf $N(\mu,1)$ and that we wish to test H_0: $\mu = 0$ against H_1: $\mu \neq 0$. A sample of n observations is taken and μ is estimated by $\hat{\mu} = \sum_{i=1}^{n} y_i/n$. We write the significance level,

---

2. As Leamer (1978) argues in discussing this Lindley paradox (P. 105), the Bayesians are clearly right in this.

the probability of Type I error, $\alpha(n)$, as a function of the sample size. Since $\hat{\mu}$ is distributed under H_0 as $N(0,1/n)$, this determines boundaries of the critical region $(-\infty,-c(n))$ and $(c(n),\infty)$ for the standardized variable $\hat{\mu}/(1/n^{1/2})$ by

$$Pr\{-c(n) < \frac{\hat{\mu}}{1/n^{1/2}} < c(n)\} = \int_{-c(n)}^{c(n)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1 - \alpha(n) \qquad (2\text{-}1)$$

The hypothesis is accepted if the observed $\hat{\mu}$ is in the interval $(-c(n)/n^{1/2}, c(n)/n^{1/2})$ and if the mean is $\mu \neq 0$, the probability of Type II error is

$$\beta(\mu,n) = \int_{-c(n)-\mu n^{1/2}}^{c(n)-\mu n^{1/2}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \Phi(c-\mu n^{1/2}) - \Phi(-c-\mu n^{1/2}) \qquad (2\text{-}2)$$

where $\Phi(\ )$ is the cumulative standard normal integral. The power of the test is defined as

$$\psi(\mu,n) = 1 - \beta(\mu,n) \qquad (2\text{-}3)$$

It is obvious that $\lim_{n\to\infty}\beta(\mu,n) = 0$ for $\mu \neq 0$ and $\lim_{n\to\infty}\beta(0,n) = 1 - \alpha(n)$.

Small α and β are desirable but even if α is fixed, β depends on the unknown value of μ . It seems reasonable to aim for a low value of expected β and we define[3]

$$T(n) = \int_{-\infty}^{\infty} \beta(\mu,n)p(\mu)d\mu \qquad (2\text{-}4)$$

---

3. $T(n)$ exists even if $p(\mu)$ is the improper prior $p(\mu) = 1$ since $\beta(\mu,n)$ goes to zero faster as $\mu \to \pm\infty$ than any polynomial.

where $p(\ )$ is a prior density of μ . Then low values of $\alpha(n)$ and $T(n)$ are desirable. Of course, arbitrarily small values of $\alpha(n)$ and $T(n)$ are not simultaneously attainable since $\alpha(n)$ and $T(n)$ are inversely related. Let $\phi(x) = \exp\{-x^2/2\}/\sqrt{2\pi}$. Then

$$\frac{d\beta(\mu,n)}{d\alpha(n)} = \phi(c(n)-\mu n^{1/2})\frac{dc(n)}{d\alpha(n)} + \phi(-c(n)-\mu n^{1/2})\frac{dc(n)}{d\alpha(n)} \tag{2-5}$$

and

$$\frac{dc(n)}{d\alpha(n)} = -\frac{1}{\phi(c(n)) + \phi(-c(n))} = -\frac{1}{2\phi(c(n))} \tag{2-6}$$

Hence $dc(n)/d\alpha(n) < 0$, $d\beta(\mu,n)/d\alpha(n) < 0$ and $dT(n)/d\alpha(n) < 0$.

The locus of $\alpha(n)$, $T(n)$ points generated by (2-1), (2-2) and (2-4) is the feasibility locus, i.e. the set of pairs that are attainable. A different locus is generated for each possible value of n . It is also straightforward to verify that the feasible locus gives $T(n)$ as a convex function of $\alpha(n)$. It is sufficient to establish that $d^2\beta(\mu,n)/d\alpha(n)^2 > 0$. For simplicity we write $c(n)$ as c and $\alpha(n)$ as α .

$$\frac{d^2\beta(\mu,n)}{d\alpha^2} = \left[\phi(c-\mu n^{1/2}) + \phi(-c-\mu n^{1/2})\right]\frac{d^2c}{d\alpha^2} + \\ + \left[\phi'(c-\mu n^{1/2}) - \phi'(-c-\mu n^{1/2})\right]\left(\frac{dc}{d\alpha}\right)^2 \tag{2-7}$$

Substituting for $dc/d\alpha$ from (2-6) and for $d^2c/d\alpha^2$ its value $-\phi'(c)/\phi(c)^3$, we obtain

$$\frac{d^2\beta(\mu,n)}{d\alpha^2} = \frac{1}{4\phi(c)^2}\left[\phi'(c-\mu n^{1/2}) - \phi'(-c-\mu n^{1/2}) - (\phi(c-\mu n^{1/2}) + \phi(-c-\mu n^{1/2}))\frac{\phi'(c)}{\phi(c)}\right] \tag{2-8}$$

This is positive if and only if

$$\mu n^{1/2}(\exp\{-\frac{(c-\mu n^{1/2})^2}{2}\} - \exp\{-\frac{(-c-\mu n^{1/2})^2}{2}\}) > 0$$

This is clearly the case, for when $\mu > 0$, $(c-\mu n^{1/2})^2 < (-c-\mu n^{1/2})^2$ and when $\mu < 0$, $(c-\mu n^{1/2})^2 > (-c-\mu n^{1/2})^2$.

3. The Utility Function and Constrained Optimization

The utility function of the investigator is assumed to have α and T as arguments:

$$U = U(\alpha,T) \tag{3-1}$$

where we omit the dependence of α and T on n for sake of notational simplicity. One could, in principle, stop at this point by simply requiring the investigator to specify his own utility function and maximize it subject to the constraint given by the feasibility locus. But it is interesting to examine the consequences of assuming a general class of utility functions, particularly because a specific member of that class has been recommended before.[4] For this purpose we select the CES function

4. Leamer (1978) argues, following Savage, that indifference curves between α and β are straight lines.

4. Results and Comparisons

In the present section we present the appropriate significance levels (α's) for two procedures: (1) the Bayesian posterior-odds computation on the assumption that $h(\theta)$ in (1-3) is normal with mean zero and variance ω^2, and (2) the method outlined in Section 3.

Assuming that the prior probability λ in (1-3) is 1/2, it is easy to show that the posterior-odds ratio in favor of H_1 is (Leamer (1978))

$$R = \left(1 + \frac{n}{\sigma^2/\omega^2}\right)^{-1/2} \exp\left\{z^2\left(1 + \frac{\sigma^2/\omega^2}{n}\right)^{-1}/2\right\} \qquad (4\text{-}1)$$

where σ^2, the variance of y, has been previously assumed to be unity, and where $z^2 = \bar{y}^2 n/\sigma^2$. It follows that H_0 is rejected whenever

$$|\bar{y}| > \left\{\frac{1}{n}\left(1 + \frac{1/\omega^2}{n}\right)^{-1/2} \ell n\left(1 + \frac{n}{1/\omega^2}\right)\right\}^{1/2}$$

from which one easily obtains the implied significance level of the test.

The optimal significance levels can also be computed for the method of Section 3 under the assumption of a normal prior density for μ (with mean 0 and variance ω^2) as well as ignoring the prior (i.e. assuming that $p(\mu) = 1$ for all μ). This computation was performed for several values of the CES parameter ρ. The appropriate value of δ in each case was determined by setting the level α for $n = 10$ (α_0) equal to the implied α- value from the corresponding Bayesian case. The Bayesian and utility maximizing significance levels are contained in Table 1. The same values of α_0 were employed when the prior is ignored as when it is not. The significance levels of this computation are in Table 2.

Table 1. Significance Levels

	Bayesian α	Utility Maximizing α for $\rho = 1.0$	$\rho = 1.5$	$\rho = 1.9$
$\omega^2 = 0.1$				
n = 10	.239	.239	.239	.239
20	.200	.220	.230	.232
30	.174	.202	.224	.226
40	.156	.188	.216	.220
50	.143	.176	.210	.214
60	.132	.166	.206	.210
70	.123	.156	.200	.206
$\omega^2 = 1.0$				
n = 10	.105	.105	.105	.105
20	.074	.080	.088	.090
30	.060	.066	.078	.080
40	.048	.058	.070	.074
50	.045	.052	.066	.070
60	.041	.046	.062	.066
70	.038	.044	.058	.064
$\omega^2 = 10.0$				
n = 10	.031	.031	.031	.031
20	.021	.022	.024	.024
30	.017	.018	.020	.020
40	.014	.014	.018	.019
50	.013	.012	.016	.018
60	.011	.011	.015	.016
70	.010	.010	.014	.015

Table 2

Utility Maximizing α for $\rho(\mu) = 1$, $\rho = 1.5$

	$\alpha_0 = .239$	$\alpha_0 = .105$	$\alpha_0 = .031$
n = 10	.239	.105	.031
20	.168	.074	.022
30	.130	.060	.018
40	.114	.052	.016
50	.102	.046	.014
60	.092	.042	.013
70	.084	.038	.012

First of all, we note that the variations in ρ (implying elasticities of substitution ranging from .50 to .34) have a fairly substantial effect on the significance levels. Where ρ is high, the optimal significance level declines with sample size much more slowly than when ρ is small. For large samples this causes a substantial difference in the Bayesian and the corresponding utility maximizing α , with the difference being proportionately greatest when the prior is not tight. When the prior is ignored (Table 2) the utility maximizing α's are generally closer to the Bayesian one's but for a tight prior $\omega^2 = .1$) the optimal α's decline with sample size faster than do the Bayesian α's . This is to be expected, for ignoring the prior tends to have the same effect as a diffuse prior.

The most important result is that in all cases the significance level declines with sample size, whether an explicit prior is or is not employed. The classical procedure of fixing α at, say, .05 and leaving it invariant cannot be justified, even if the prior is ignored.

5. Concluding Comments

The standard, classical procedure fixes the significance level α and "accepts" any value of n as determined by sample size. It is difficult to think of an optimizing framework in which such a procedure would be rational.

Alternatively one may employ the Bayesian posterior odds for selecting among hypotheses. This procedure leads to significance levels that decline with sample size. However, many investigators shy away from using posterior odds because of the need to specify prior distributions.

A third alternative is the utility maximizing framework which may be employed with or without explicitly specifying prior distributions. The cost of using this procedure is that the investigator must be able to specify his utility function, its general form and the values of its parameters and that he must abandon the standard Bayesian and decision-theoretic approach of letting loss be a function only of the true state and the action taken. Yet it does not appear to be too much to ask the investigator to specify a utility function. Although one may reasonably argue about the proper domain of the utility function or about its mathematical form, problems of choice are not soluble in a satisfactory manner without a suitable optimand. It seems natural to accept this in statistical decision making as it has been accepted in the theory of the consumer and other portions of microeconomic theory.

References

Leamer, E.E. (1978), Specification Searches, Wiley.

Zellner, A. (1979), "Posterior Odds Ratios for Regression Hypotheses: General Considerations and Some Specific Results," University of Chicago, mimeo.