# TESTING NONNESTED HYPOTHESES*

by

Richard E. Quandt

ECONOMETRIC RESEARCH PROGRAM
Research Memorandum No. 140

May 1972

*I am indebted to Gregory C. Chow, David R. Cox, Stephen M. Goldfeld and Alan Stuart for helpful comments and suggestions. Of course, I am alone responsible for errors. Financial support from the National Science Foundation is gratefully acknowledged.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

# TESTING NONNESTED HYPOTHESES

by

Richard E. Quandt

## 1. Theoretical Considerations

The standard problem of testing hypotheses arises as follows: Let $y$ be a random variable with probability density function (pdf) $f(y,\theta)$, where $\theta$ is a p-dimensional vector of parameters. Letting $\Omega$ represent p-dimensional Cartesian space, $\theta\varepsilon\Omega$ in general. The null hypothesis $H_o$ is expressed by the restriction $\theta \varepsilon \Omega_o$, where $\Omega_o \cap \Omega = \Omega_o$. It is this circumstance which makes likelihood ratio tests possible, for denoting the loglikelihood function by $L(\theta)$, it guarantees that

$$\ell = \sup_{\theta\varepsilon\Omega_o} L(\theta) - \sup_{\theta\varepsilon\Omega} L(\theta) \leq 0 \qquad (1\text{-}1)$$

This condition is obviously also necessary for the asymptotic theorem, which holds under certain broad regularity conditions, that $-2\ell$ has $\chi^2$ distribution with degrees of freedom equal to the number of restrictions imposed by the hypothesis on $\Omega$.[1]

The standard theory is not applicable if the null hypothesis and the alternative hypothesis are disparate or nonnested in the following sense. Let there be two hypotheses denoted by $H_1$ and $H_2$. According to $H_1$, the random variable $y$ has pdf $f_1(y, \theta_1)$ with $\theta_1 \varepsilon \Omega_1$ and according to $H_2$, $y$ has pdf $f_2(y, \theta_2)$ with $\theta_2 \varepsilon \Omega_2$. Finally, assume that $\Omega_1 \cap \Omega_2 \neq \Omega_1$

---

[1] See [10], pp. 230-1.

and $\Omega_1 \cap \Omega_2 \neq \Omega_2$ . Cases of this type may be more common in econometrics than is usually recognized. A brief list of examples is as follows: (1) Test of the hypothesis that personal incomes have the lognormal distribution as against the alternative that they have Pareto distribution; (2) Test of the hypothesis that the error term in a Cobb-Douglas type production function enters the equation multiplicatively as $.e^u$ against the alternative that it enters additively as $+u$; (3) Test of the hypothesis that the regression equation has one form against the alternative that it has some other form.

Although these and other questions of this type have been investigated frequently by various more or less ad hoc methods,[2] only few studies have explored systematically the statistical aspects of choosing between competing models and of testing hypotheses of this type. In a regression context and under the classical assumptions Theil has shown that a choice between models based on maximizing the multiple $r^2$, adjusted for degrees of freedom, will on the average produce the correct statistical decision.[3] More general are the works of Cox and Atkinson.[4] Cox's main procedure is based on the generalized loglikelihood ratio $L_1(\theta_1) - L_2(\theta_2)$ where $L_1$ and $L_2$ are the likelihood functions appropriate for $H_1$ and $H_2$ respectively. In a recent paper Pesaran specializes this approach to the choice between competing regression models, both under the classical assumptions and also in the presence of first order serial correlation of the error term.[5]

---

[2]See [2], [4], [15].

[3]See [16], pp. 211-214 for Theil's argument and Pesaran's work [12] for some criticisms of it.

[4]See [1], [5], [6].

[5]See [12].

Cox introduces and Atkinson further elaborates another procedure by which two competing pdf's are combined and their respective parameters are embedded in a larger space with the aid of a choice parameter $\lambda$ . The pdf generating the sample is taken as

$$h(y,\theta_1,\theta_2,\lambda) = kf_1(y,\theta_1)^{\lambda} f_2(y,\theta_2)^{1-\lambda} \qquad (1\text{-}2)$$

and it is suggested that this combined pdf, representing a compound statistical model, be used to make inferences about $\lambda$ .[6]

The difficulty with (1-2) is that in order for $h(y,\theta_1, \theta_2,\lambda)$ to be a density function, the factor of proportionality must be

$$k = \frac{1}{\int_{-\infty}^{\infty} f_1(y,\theta_1)^{\lambda} f_2(y,\theta_2)^{1-\lambda} dy} \qquad (1\text{-}3)$$

Hence the loglikelihood function is

$$L = \sum_{i=1}^{n} \log h(y_i,\theta_1,\theta_2,\lambda) = \lambda \sum_{i=1}^{n} \log f_1(y_i,\theta_1) + (1-\lambda) \sum_{i=1}^{n} \log f_2(y_i,\theta_2)$$
$$- \sum_{i=1}^{n} \log \int_{-\infty}^{\infty} f_1(y,\theta_1)^{\lambda} f_2(y,\theta_2)^{1-\lambda} dy \qquad (1\text{-}4)$$

which must be maximized with respect to $\theta_1,\theta_2$ and $\lambda$ . Since, in general, the integral in (1-4) will be computable only by numerical quadrature, the maximization of (1-4) is likely to be a difficult task.

---

[6] For further discussion of these and related issues see [7].

Even though the exponential combination of pdf's is convenient for many frequently occurring distributions, it is of some interest to explore the possibilities of hypothesis testing with the linear combination of pdf's given by

$$h(y,\theta_1,\theta_2,\lambda) = \lambda f_1(y,\theta_1) + (1-\lambda)\, f_2(y,\theta_2) \qquad (1\text{-}5)$$

There are several reasons for preferring (1-5) to (1-4). (1) The resulting pdf is a convex combination of pdf's and is capable of intuitively easier interpretation. (2) It is formally identical with the pdf of a random variable that is produced by a mixture of two distributions.[7] Under the interpretation of a mixture distribution, $\lambda$ may be interpreted (unlike in (1-4)) as the probability that nature has chosen pdf $f_1(y,\theta_1)$ for generating values of $y$. (3) No normalizing constant is required in this case and the maximization of the likelihood function becomes a relatively straightforward numerical optimization problem.

There are also some difficulties associated with using a composite pdf such as (1-5). (1) Departure from one hypothesis not in the direction of the other but in the opposite direction may yield values of $\lambda$ outside the (0,1) interval which may cause some values of the density $h(y,\theta_1,\theta_2,\lambda)$ to become negative. (2) The reasonable restriction that $\lambda$ not be outside the (0,1) interval means that we are attempting to estimate a boundary point of the range of $\lambda$'s ; hence there will not exist an interval containing the true value of $\lambda$ in which the regularity conditions guaranteeing the asymptotic normality of maximum likelihood estimates hold.[8]

---

[7]See [9] and [14].

[8]See [10], pp. 43-44. There is a singularity of a sort in the likelihood function at $\lambda = 0$ or 1 since when $\lambda = 0$, $h(y,\theta_1,\theta_2,\lambda)$ is not a function of $\theta_1$ and when $\lambda = 1$ it is not a function of $\theta_2$.

In spite of some possible advantages that (1-5) may have over (1-4), it is clear that any embedding of this type retains a degree of arbitrariness, since a compound statistical model can be constructed in numerous ways. It is recognized that serious problems can arise if one type of embedding results in rejecting $H_1$ in favor of $H_2$ and another type in the reverse.

The remainder of this paper reports some computational experience with the formulation resulting from (1-5). Section 2 contains the results of some illustrative sampling experiments, while Section 3 is devoted to applying the test procedures to two concrete economic examples.

## 2. Some Sampling Experiments

Experiments were performed in a regression context designed to test the hypothesis that a linear regression equation generated the data as against the alternative that a linear regression equation holds for the data deflated by some other variable.[9] Specifically, let the two hypotheses be given by

$$H_1: \quad y_i = a_1 + b_1 x_i + u_i \tag{2-1}$$

$$H_2: \quad \frac{y_i}{z_i} = a_2 + b_2 \frac{x_i}{z_i} + \frac{v_i}{z_i} \tag{2-2}$$

with $u_i$ and $v_i$ being distributed as $N(0,\sigma^2)$ in either case and with $x_i$ and $z_i$ being nonstochastic. Equ. (2-2) is equivalent to

---

[9]Obviously this problem could be approached more directly by employing the residual sums of squares of a composite equation incorporating both hypotheses and of the equation incorporating only one of them. See [1], p. 327.

$$y_i = a_2 z_i + b_2 x_i + v_i \tag{2-3}$$

and the density function (1-5) is then

$$h(y_i) = \frac{\lambda}{\sqrt{2\pi}\sigma_1} \exp\{-\frac{1}{2\sigma_1^2}(y_i - a_1 - b_1 x_i)^2\} + \frac{(1-\lambda)}{\sqrt{2\pi}\sigma_1} \exp\{-\frac{1}{2\sigma_2^2}(y_i - a_2 z_i - b_2 x_i)^2\} \tag{2-4}$$

Maximum likelihood estimates are obtained by maximizing

$$L = \sum_{i=1}^{n} \log h(y_i) \tag{2-5}$$

with respect to $a_1$, $b_1$, $\sigma_1$, $a_2$, $b_2$, $\sigma_2$ and $\lambda$ .

The embedding of the parameters $\theta_1$ and $\theta_2$ in a larger space with the aid of $\lambda$ has the consequence that we no longer test $H_1$ against $H_2$ but rather $H_1$ against the compound hypothesis $H_3$ (in which $H_1$ and $H_2$ are weighted by $\lambda$ and $(1-\lambda)$ respectively) and also $H_2$ against $H_3$ . Whatever the decision criterion, the following cases may arise: (1) $H_1$ is rejected and $H_2$ is not; (2) $H_2$ is rejected and $H_1$ is not; (3) both $H_1$ and $H_2$ are rejected, (4) neither $H_1$ nor $H_2$ is rejected. The first two cases are straightforward and answer the original question of choice between $H_1$ and $H_2$ . Case (3) is equivalent to accepting the compound hypothesis $H_3$ . Case (4) signals that there is inadequate information to discriminate between the hypotheses.

There are at least two standard procedures that could be employed for the test procedure. First, one could attempt to estimate the asymptotic covariance matrix of the maximum likelihood estimates by[10]

---

[10]See [10], pp. 52-55.

$$\left\| - \frac{\partial^2 L}{\partial \hat{\theta} \partial \hat{\theta}'} \right\|^{-1} \tag{2-6}$$

Denoting by $\hat{\lambda}$ the maximum likelihood estimate for $\lambda$ and by $\hat{\sigma}_{\hat{\lambda}}$ the squareroot of the asymptotic variance for $\hat{\lambda}$, $H_1$ would be rejected (at some appropriate significance level) if the interval $(\hat{\lambda} - k_1\hat{\sigma}_{\hat{\lambda}}, \hat{\lambda} + k_2\hat{\sigma}_{\hat{\lambda}})$ did not overlap 1.0 for some appropriate positive constants $k_1$ and $k_2$ ; $H_2$ will be rejected if the interval does not overlap 0.0; both hypotheses will be rejected if the interval overlaps neighter 1.0 nor 0.0 and finally neither hypothesis will be rejected if the interval overlaps both 1.0 and 0.0. If asymptotic normality could be assumed, the value of 1.96 would be appropriate for $k_1$ and $k_2$ for a .05 level of significance. We shall examine the extent to which the use of 1.96 for $k_1$ and $k_2$ produces the correct statistical decision.

Secondly, we may consider the loglikelihood ratios $\ell_1$ and $\ell_2$ defined in (1-1) where $\Omega_o$ is the space of parameters obtained by setting $\lambda$ alternately equal to 1.0 or 0.0 and $\Omega$ is the space of all parameters. A large value of either likelihood ratio leads to rejection of the corresponding hypothesis. Since the asymptotic theorem concerning $-2\ell$ alluded to in Section 1 cannot be expected to hold, we will attempt to ascertain critical values by the sampling experiments.

Sampling experiments were carried out for varying values of the number of observations $n$ . Data were generated from (2-1) with $a = b = 1.0$, $\sigma_u^2 = 20.0$. The set of x's used was identical in repeated samples and was chosen from the uniform distribution over the (0,20) interval. The set of z's used was also identical in repeated samples and was chosen to be distributed uniformly over the (0.8, 1.2) interval and independently of $x$ . Error terms were normally distributed with variance $\sigma_u^2$ . In the experiments $H_1$ (i.e., Equ. (2-1)) was true and $H_2$ was false.

The experiments were replicated as many times as was necessary to produce 30 successful replications. Failures occurred because at the point at which the maximization algorithms terminated the matrix of second partial derivatives was not always negative definite.[11]

Table 1 displays the failure rate and the results of testing $\lambda$ directly on the assumption of asymptotic normality.

TABLE 1. Results of Sampling Experiments when $H_1$ is True, Using Estimate of Asymptotic Variance of $\hat{\lambda}$.

| Sample Size | Failure Rate | Fraction of Cases in Which $\hat{\lambda} < .5$ | Fraction of Cases in Which $H_1$ is Accepted[1] | Fraction of Cases in Which $H_2$ is Accepted[2] |
|---|---|---|---|---|
| 30 | .30 | .400 | .300 | .200 |
| 60 | .09 | .500 | .267 | .267 |
| 90 | .26 | .067 | .567 | .000 |
| 180 | .29 | .000 | .633 | .000 |
| 360 | .21 | .000 | .833 | .000 |

The failures appeared to be caused by flatness of the likelihood surface, making it difficult to get a good estimate of $\partial^2 L/\partial\hat{\theta}\partial\hat{\theta}'$ by numerical differencing. The results employing an interval around $\hat{\lambda}$ exhibit the expected qualitative behavior but convergence is not quite as rapid as might be desired; for n = 60, for example, only 26.7 percent of the cases are decided correctly although for n = 90 and greater no wrong decisions are made and the fraction of undecided cases shrinks progressively as n increases. Less formally one might accept

[11]Maximization employed the two algorithms described in [8] and [13].

$H_1$ if $\hat{\lambda} > .5$ and $H_2$ otherwise. The fraction of cases in which $\hat{\lambda} \leq .5$ is also displayed and shows the correct qualitative behavior (but note the reversal between $n = 30$ and $n = 60$) but, again, satisfactory convergence cannot be said to have taken place for $n < 90$.

The conjecture that the sample values of $-2\ell$ will not have the $\chi^2$ distribution was tested by the Kolmogorov-Smirnov test for $\chi^2$ distributions with, alternately, 1, 2,...,10 degrees of freedom. If the asymptotic theorem concerning $-2\ell$ were true, one would expect $\chi^2(1)$ to provide a good fit. This null hypothesis is categorically rejected for every value of $n$. For example, if in the case of $n = 360$ we had employed the appropriate critical value at the .05 level, we would have rejected $H_1$ when it was in fact true in 80 percent of the cases.

The best-fitting $\chi^2$ distributions are $\chi^2(9)$ for $n = 30$ and, $\chi^2(8)$ for $n = 60$ and 90. For larger values of $n$ one must reject altogether the hypothesis of a $\chi^2$ distribution with degrees of freedom $\leq 10$. Since it appears extremely difficult to determine exactly the distribution of $-2\ell$ we obtain empirically determined critical values from the sampling experiments. These are displayed in Table 2.

TABLE 2. Empirically Determined Critical Values for $2\ell_1$ for the .05 level

| n | Critical Value |
|---|---|
| 30 | 17.88 |
| 60 | 14.11 |
| 90 | 19.32 |
| 180 | 13.85 |
| 360 | 14.90 |

We can now examine the behavior of the likelihood ratio based on $H_2$ when $H_1$ is true. Denote the loglikelihood ratios based on $H_1$ and $H_2$ by $\ell_1$ and $\ell_2$. For $n = 30$ and 60 the behavior of these two is very similar. For larger values of $n$ the values of $\ell_2$ become quite large compared to $\ell_1$. If the empirically determined critical values derived from the actual sample distribution of $\ell_1$ were used, the test of $H_2$ based on $\ell_2$ would have resulted correctly in rejection of $H_2$ in 63 percent of the cases for $n = 90$, and 100 percent of the cases for $n = 180$ and 360. These empirically determined critical values may therefore be useful in performing tests of hypotheses. Since the sample distributions of $-2\ell_1$ appear fairly stable for varying values of $n$, a better estimate may be given by the average of the values in Table 2 which is 16.01. A corresponding estimate obtained by pooling the samples for different values of $n$ is 18.08

## 3. Some Economic Examples.

Klein's Model II. Klein's Model II is a reduced form model which is given by[12]

$$\frac{y_t}{p_t} = \alpha_o + \alpha_1 \frac{y_{t-1}}{P_{t-1}} + \alpha_2 \frac{M_{t-1}}{p_{t-1}} + \alpha_3 \left(\frac{I_t + G_t - T_t}{p_t}\right) + v_t \qquad (3\text{-}1)$$

where $y_t$ is per capita disposable income, $M_t$ is total per capita deposits and currency outside banks, $I_t$ is gross per capita investment, $G_t$ is per capita government expenditures on goods and services, $p_t$ is the cost of living index; and $T_t$ is per capita $GNP_t$ minus $y_t$. The parameters $\alpha_o$, $\alpha_1$, $\alpha_2$, $\alpha_3$

---

[12] See [11], pp. 80-84.

are related to the original structural parameters but this relationship is of no interest here. We shall contrast the hypothesis expressed by (3-1), referred to as $H_2$, with the hypothesis $H_1$ that the equation is in terms of money terms rather than real terms,[13] i.e., that

$$y_t = \alpha_{o1} + \alpha_{11} y_{t-1} + \alpha_{21} M_{t-1} + \alpha_{31} (I_t + G_t - T_t) + u_t \qquad (3\text{-}2)$$

Equ. (3-1) is transformed to

$$y_t = \alpha_{o2} p_t + \alpha_{12} y_{t-1} + \alpha_{22} M_{t-1} + \alpha_{32} (I_t + G_t - T_t) + w_t \qquad (3\text{-}3)$$

where $u_t$ and $w_t$ are both assumed to be homoscedastic. It will be assumed here that both $u_t$ and $w_t$ have first-order autocorrelation and maximum likelihood estimates will be obtained on this assumption. The results are therefore not directly comparable with those of Klein. Rewrite (3-2) and (3-3) as

$$y_t = f_t + u_t \qquad (3\text{-}4)$$

$$y_t = g_t + w_t \qquad (3\text{-}5)$$

where $f_t$ and $g_t$ represent the systematic parts on the right hand side of (3-2) and (3-3). The pdf's for $y_t$ $(t=2,\ldots,n)$ are taken to be

$$h_1(y_t) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\{-\frac{1}{2\sigma_1^2}(y_t - \rho_1 y_{t-1} - (f_t - \rho_1 f_{t-1}))^2\} \qquad (3\text{-}6)$$

$$h_2(y_t) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\{-\frac{1}{2\sigma_2^2}(y_t - \rho_2 y_{t-1} - (g_t - \rho_2 g_{t-1}))^2\} \qquad (3\text{-}7)$$

---

[13] Klein was primarily interested in whether $\alpha_2$ is significantly different from zero.

The density functions for $y_1$ are taken to be

$$h_1(y_1) = \frac{(1-\rho_1^2)^{1/2}}{\sqrt{2\pi}\sigma_1} \exp\{-\frac{1}{2\sigma_1^2}(y_1-f_1)^2\} \tag{3-8}$$

$$h_2(y_1) = \frac{(1-\rho_1^2)^{1/2}}{\sqrt{2\pi}\sigma_2} \exp\{-\frac{1}{2\sigma_2^2}(y_1-g_1)^2\} \tag{3-9}$$

from which the loglikelihood function is immediately obtained. The results of maximizing it are displayed and contrasted with Klein's results in Table 3. The results

TABLE 3. Results for Klein's Model II

| Coefficient | Klein's Result | Klein's Standard Error | Max. Like. Result | Max. Like. Standard Error |
|---|---|---|---|---|
| $\alpha_{01}$ | -** | -** | 85.570 | 15.651 |
| $\alpha_{11}$ | -** | -** | .073 | .044 |
| $\alpha_{21}$ | -** | -** | -.170 | .042 |
| $\alpha_{31}$ | -** | -** | .701 | .008 |
| $\alpha_{02}$ | 186.53 | -* | 77.310 | 2.082 |
| $\alpha_{12}$ | .30 | .13 | .176 | .004 |
| $\alpha_{22}$ | .13 | .10 | -.258 | .007 |
| $\alpha_{32}$ | 2.36 | .35 | .695 | .018 |
| $\rho_1$ | -** | -** | .269 | .341 |
| $\rho_2$ | -** | -** | .678 | .037 |
| $\lambda$ | -** | -** | .765 | .116 |

*Not reported

**Not relevant

of maximum likelihood estimation for the composite pdf differs from those of Klein in that the coefficient of lagged income is negative. The $(\hat{\lambda} - 1.96\, \hat{\sigma}_\lambda, \hat{\lambda} + 1.96\, \hat{\sigma}_\lambda)$ interval does not contain either zero or one. This suggests that the compound hypothesis should be accepted. The likelihood ratio statistic for the monetary hypothesis, $-2\ell_m$, is 5.02 and for the real hypothesis, $2\ell_r$, is 0.02. In terms of the empirically established critical values neither pure hypothesis is rejected in favor of the compound one (although the real hypothesis seems better than the monetary one; a view that tends to be confirmed by the very small asymptotic variances associated with the estimates for the real part of the compound pdf). The first type of test suggests that the consumption function is a hybrid between a version expressed in real terms and one expressed in money terms; the second one suggests that the two pure hypotheses and the compound one are of comparable degree of acceptability. Both lend some support to the findings of Branson and Klevorick [3] that money illusion is present in the consumption function.

Chow's Model of Computer Growth. In [4] Chow fitted Gompertz and logistic curves to explain the quantity of computers used in the years 1954-1965. Approximating derivatives by finite differences, the Gompertz formulation yields

$$\log\frac{y_t}{y_{t-1}} = \alpha_{o1} + \alpha_{11}\log p_t + \alpha_{21}\log y_{t-1} \quad (3\text{-}10)$$

and the logistic yields

$$\log\frac{y_t}{y_{t-1}} = \alpha_{o2} + \alpha_{12}p_t^{-\beta} + \alpha_{22}y_{t-1} \quad (3\text{-}11)$$

where $y_t$ measures the quantity of computers and $p_t$ is a price index of computers deflated by the GNP deflator.

In attacking this problem we have assumed that the error term enters both equations additively and that, following Chow, serial correlation of the error term may be disregarded. The parameter $\lambda$ was associated with the Gompertz formulation. Maximum likelihood estimates as well as the squareroots of the estimated asymptotic variances are given in Table 4.

TABLE 4. Results for Chow's Model

| Coefficient | Chow's Result | Chow's Standard Error | Max. Like. Result | Max. Like. Standard Error |
|---|---|---|---|---|
| $\alpha_{01}$ | 2.950 | -* | 3.144 | .681 |
| $\alpha_{11}$ | -.364 | .173 | -.408 | .196 |
| $\alpha_{21}$ | -.253 | .074 | -.279 | .070 |
| $\alpha_{02}$ | -* | -* | .787 | .021 |
| $\alpha_{12}$ | .054 | .045 | .054 | .012 |
| $\alpha_{22}$ | $-1.012\text{x}10^{-5}$ | $.608\text{x}10^{-5}$ | $-1.022\text{x}10^{-5}$ | $2.172\text{x}10^{-6}$ |
| $\beta$ | 2.500 | -* | 2.445 | .188 |
| $\lambda$ | -** | -** | .540 | .204 |

*Not reported

**Not relevant

The coefficients and standard errors estimated from the compound model are quite close (wherever such comparison is relevant) to the corresponding quantities estimated by Chow, particularly for the Gompertz model. The correspondence is more pronounced than for the Klein model and is presumably due, at least in part, to the fact that in this case the basic statistical model has not been altered. The value of $\hat{\lambda}$ is greater than .5 and thus gives slight evidence in favor of Chow's conclusion that the Gompertz hypothesis is preferable. The interval $(\hat{\lambda} - 1.96\, \hat{\sigma}_{\hat{\lambda}}, \hat{\lambda} + 1.96\, \hat{\sigma}_{\hat{\lambda}})$ does not include either zero or unity and leads to an inconclusive result which is not surprising in the light of the few observations used. The value of the likelihood ratio test statistic $-2\ell$ is 1.94 for the Gompertz and 19.46 for the logistic hypothesis and on the basis of the empirically established critical values discussed in Section 2 the Gompertz hypothesis is accepted and the logistic hypothesis is rejected.

## 4. Summary

To test the nonnested hypotheses it is proposed to transform the problem so that it becomes one of testing either pure hypothesis against a compound hypothesis obtained by weighing the pdf's of the pure hypotheses by $\lambda$ and $1-\lambda$. Maximum likelihood estimates are obtained in sampling experiments and although the maximum likelihood estimates cannot be assumed to be asymptotically normal, the qualitative behavior of $\hat{\lambda}$ as well as of the likelihood ratio test statistic makes the test procedure workable. They have been applied to Klein's Model II and Chow's Model of Growth in computer demand and produce reasonable answers.

Several important questions remain for further research: (1) What is the asymptotic distribution of the maximum likelihood estimates and of the likelihood ratio statistic? (2) Are these distributions invariant with respect to sufficiently large classes of econometric models? (3) Are tests based on asymptotic distributions going to be sufficiently powerful to be useful in concrete examples? (4) What difference will result from different types of embeddings? Successful answers to even some of these may permit testing of models against specific alternatives more routinely than is now possible.

REFERENCES

[1] Atkinson, A.C., "A Method for Discriminating Between Models," Journal of the Royal Statistical Society, Series B, 32 (1970), 323-44.

[2] Bodkin, R.G., and L. R. Klein, "Nonlinear Estimation of Aggregate Production Functions," Review of Economics and Statistics, XLIX (1967), 28-44.

[3] Branson, W.H., and A.K. Klevorick, "Money Illusion and the Aggregate Consumption Function," American Economic Review, LIX (1969), 832-849.

[4] Chow, G.C., "Technological Change and Demand for Computers," American Economic Review, LVII (1967), 1117-1130.

[5] Cox, D.R., "Further Results on Tests of Separate Families of Hypotheses," Journal of the Royal Statistical Society, Series B, 24 (1962), 406-24.

[6] Cox, D.R., "Tests of Separate Families of Hypotheses," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California Press, 1961, pp. 105-23.

[7] Dhrymes, P.J., E. P. Howrey, S. H. Hymans, J. Kmenta, E.E. Leamer, R.E. Quandt, J.B. Ramsey, H.T. Shapiro, V. Zarnowitz, "Criteria for Evaluation of Econometric Models," Annals of Economic and Social Measurement, forthcoming.

[8] Goldfeld, S.M., R.E. Quandt, H.F. Trotter, "Maximization by Quadratic Hill Climbing," Econometrica, 34 (1961), 541-551.

[9] Hill, B.M., "Information for Estimating the Propositions in Mixtures of Exponential and Normal Distributions," Journal of the American Statistical Association, 58 (1963), 918-32.

[10] Kendall, M.G., and A. Stuart, The Advanced Theory of Statistics, Vol. 2, Hafner Publishing Co., 1961.

[11] Klein, L.R., Economic Fluctuations in the United States 1921-1941, Cowles Monograph No. 11, Wiley, 1950.

[12] Pesaran, M.H., "On the General Problem of Model Selection," mimeographed, Trinity College, Cambridge, Feb. 1972, pp. 1-30.

[13] Powell, M.J.D., "An Efficient Method for Findings the Minimum of a Function of Several Variables Without Calculating Derivatives," Computer Journal, 7 (1964), 155-162.

[14] Quandt, R.E., "A New Approach to Estimating Switching Regressions," Journal of the American Statistical Association, 67 (1972).

[15] Quandt, R.E., "On the Size Distribution of Firms," American Economic Review, LVI (1966), 416-32.

[16] Theil, H., Economic Forecasts and Policy, North Holland Publishing Co., 1958.