1

Score Test: Historical Review and Recent Developments

C.R. Rao

Department of Statistics, Pennsylvania State University, State College, Pennsylvania, U.S.A.

Abstract: The three asymptotic tests, Neyman and Pearson Likelihood Ratio (LR), Wald's statistic (W) and Rao's score (RS)are referred to in statistical literature on testing of hypotheses as the *Holy Trinity*. All these tests are equivalent to the first-order of asymptotics, but differ to some extent in the second-order properties. Some of the merits and defects of these tests are presented.

Some applications of the score test, recent developments on refining the score test and problems for further investigation are presented.

Keywords and phrases: Composite hypothesis, Lagrangian multiplier (LM) test, Likelihood ratio (LR), Neyman's C(α), Neyman-Rao test, Rao's score (RS), Wald's statistic (W)

1.1 Introduction

The Score test was introduced in Rao (1948) as an alternative to the likelihood ratio test of Neyman and Pearson (1928) and Wald (1943) test. A few years later Aitchison and Silvey (1958) and Silvey (1959) gave an interpretation of the score statistic in terms of a Lagrangian Multiplier used in optimizing a function subject to restrictions, and called it the Lagrangian Multiplier (LM) test.

The score (RS) test went unnoticed for a number of years after it was introduced. The first application of the score test, apart from the examples given in Rao (1948, 1950, 1961) appeared in econometric literature [Byron (1968)]. During the late 1970s and 1980s, the RS test was applied to a variety of problems in econometrics. Reference may be made to survey papers by Breusch and Pagan (1980), Engle (1984), Kramer and Sonnberger (1986), and Godfrey (1988). Most of the recent textbooks on econometrics also discuss the RS test. Some of them are by White (1984, pp. 72–74), Amemiya (1985, pp. 141–146), Judge et al. (1985, pp. 182–187), Kmenta (1986, pp. 493–495), Spanos (1986, pp. 326–336), Maddala (1988, pp. 137–139), Green (1990, pp. 357–359), and Harvey (1990, pp. 169–177).

The distributional aspects of the RS statistic are covered in books by Rao (1973, pp. 418–419), Serfling (1980, pp. 156–160), Godfrey (1988, pp. 13–15), Lehmann (1999, pp. 451, 529, 532, 534, 539, 541, 570), and Bickel and Doksum (2001, pp. 335–336, 399–402).

The study of the power properties of the RS test started with a paper by Chandra and Joshi (1983) and continued by Chandra and Mukherjee (1984, 1985), Chandra and Samanta (1988), Ghosh (1991) and others. Reference may be made to Peers (1971) for a comment on a conjecture I made about the local properties of the LR test, which motivated the work of others on power properties.

In this chapter, a brief review is given of the RS statistic and its merits and demerits in terms of power properties compared to LR and W are discussed. Some of the recent developments and refinements and modifications of the RS statistic are presented and some problems for future research are indicated.

1.2 Asymptotic Tests of a Simple Hypothesis

1.2.1 Notation

Let $X = (x_1, \ldots, x_n)$ be an iid sample of size n from the density function $p(x, \theta)$ where θ is a p-vector parameter, and denote the joint density by $P(X, \theta) = p(x_1, \theta) \ldots p(x_n, \theta)$ and the log likelihood by $L(\theta|X) = \log P(X, \theta)$. The score vector of p components, as defined by Fisher, is

$$\begin{aligned} s(\theta) &= \frac{1}{P}\frac{\partial P}{\partial \theta} = (s_1(\theta), \ldots, s_p(\theta))', \qquad (1.1) \\ s_i(\theta) &= \frac{1}{P}\frac{\partial P}{\partial \theta_i}, \ i = 1, \ldots, p. \end{aligned}$$

The Fisher information matrix of order $p \times p$ is defined by

$$ni(\theta) = I(\theta) = E[s(\theta)s'(\theta)] = (i_{rs}(\theta)) \tag{1.2}$$

where $i_{rs}(\theta) = E\left[s_r(\theta)s_s(\theta)\right]$. The maximum likelihood estimate of θ is obtained as a solution of the p equations

$$s_i(\theta) = 0, \ i = 1, \ldots, p \tag{1.3}$$

which we represent by $\hat{\theta}$. Under suitable regularity conditions [Lehmann (1999, pp. 499–501)], using the multivariate cental limit theorem

$$n^{-1/2}s(\theta_0) \sim N_p\left(0, i(\theta_0)\right) \tag{1.4}$$

where θ_0 is the true value, and

$$n^{1/2}\left(\hat{\theta}-\theta_0\right) \sim N_p\left(0, [i(\theta_0)]^{-1}\right) \tag{1.5}$$

where $N_p(0, A)$ is a p variate normal distribution with mean zero and covariance matrix A.

1.2.2 Three possible tests of a simple hypothesis: The Holy Trinity

Let $H_0 : \theta = \theta_0$ (a specified p-vector) be the null hypothesis to be tested. Three tests which are in current use are as follows.

1. Likelihood ratio test [Neyman and Pearson (1928)]

$$LR = 2\left[L(\hat{\theta}|X) - L(\theta_0|X)\right] \tag{1.6}$$

   where $L(\theta|X) = \log P(X, \theta)$.

2. Wald test [Wald (1943)]

$$W = (\hat{\theta}-\theta_0)' I(\hat{\theta})(\hat{\theta}-\theta_0). \tag{1.7}$$

3. Rao Score test [Rao (1948)]

$$RS = [s(\theta_0)]' \, [I(\theta_0)]^{-1} \, [s(\theta_0)] \, . \tag{1.8}$$

All the three statistics known as the *Holy Trinity* have an asymptotic chi-square distribution on p degrees of freedom.

1.2.3 Motivation for the score test of a simple hypothesis

Consider the case of a single parameter θ and $H_0 : \theta = \theta_0$. If $w \subset R^n$ is the critical region of size α in the sample space, then the power of the test is

$$\pi(\theta) = \int_w P(X, \theta) dv \text{ with } \pi(\theta_o) = \int_w P(X, \theta_0) dv = \alpha.$$

To find a locally most powerful one-sided test ($\theta > \theta_0$) we maximize

$$\pi'(\theta_0) = \int_w P'(X, \theta_0) dv$$

subject to $\pi(\theta_0) = \alpha$. Using the Neyman–Pearson Lemma, the optimal region is defined by

$$\frac{P'(x, \theta_0)}{P(x, \theta_0)} \geq \lambda \text{ or } s(\theta_0) \geq \lambda$$

where λ is chosen such that the size of the region is α, as shown in Rao and Poti (1946). The test can be written in the form

$$\frac{s(\theta_0)}{\sqrt{I(\theta_0)}} \geq \lambda \text{ or } \frac{[s(\theta_0)]^2}{I(\theta_0)} \geq \lambda^2. \tag{1.9}$$

In the multiparameter case, the slope of the power function in the direction $a = (a_1, \ldots, a_p)'$, at θ_0 is

$$a_1 s_1(\theta_0) + \cdots + a_p s_p(\theta_0) = a' s(\theta_0) \tag{1.10}$$

and the statistic (1.9) takes the form

$$\frac{[a' s(\theta_0)]^2}{a' I(\theta_0) a}. \tag{1.11}$$

Maximizing with respect to a yields the statistic

$$[s(\theta_0)]'[I(\theta_0)]^{-1}[s(\theta_0)] \tag{1.12}$$

which is the same as (1.8).

1.2.4 Test of a composite hypothesis

Under the same setup as in Section 1.2.1, let the hypothesis to be tested be $H_0 : h(\theta) = c$, where h is an $r \times 1$ vector function of the p-vector θ with $p \geq r$ and c is a given r-vector of constants. The corresponding *Holy Trinity* is as follows:

1. Likelihood ratio test [Neyman and Pearson (1928)]

$$LR = 2\left[l(\hat{\theta}|X) - l(\tilde{\theta}|X)\right] \tag{1.13}$$

   where $\tilde{\theta}$ is the ml of θ under the restriction $h(\theta) = c$.

2. Wald test [Wald (1943)]

$$W = \left[h(\hat{\theta}) - c\right]' \left[A(\hat{\theta})\right]^{-1} \left[h(\hat{\theta}) - c\right] \tag{1.14}$$

   where

$$A(\theta) = [H(\theta)][I(\theta)]^{-1}[H(\theta)]',$$
$$H(\theta) = (\partial h_i(\theta)/\partial\theta_j), h(\theta) = (h_1(\theta), \ldots, h_r(\theta))',$$

   and $I(\theta)$ is as defined in (1.2).

3. Rao Score test [Rao (1948)]

$$RS = [s(\tilde{\theta})]'[I(\tilde{\theta})]^{-1}[s(\tilde{\theta})]. \tag{1.15}$$

All the three statistics have an asymptotic chi-square distribution on r degrees of freedom.

An alternative way of expressing the RS statistic is as follows. Note that $\tilde{\theta}$, the restricted ml of θ, is a solution of the equation

$$s(\theta) + [H(\theta)]'\lambda = 0,\ h(\theta) = c$$

where λ is an r-vector of the Lagrangian Multiplier so that $[s(\tilde{\theta})]' = -\lambda' H(\tilde{\theta})$. Substituting in (1.15) we have

$$\text{RS} = \lambda' H(\tilde{\theta})[I(\tilde{\theta})]^{-1}[H(\tilde{\theta})]'\lambda = \lambda'[A(\tilde{\theta})]\lambda \tag{1.16}$$

where $A(\theta)$ is as defined in (1.14). Silvey (1959) expressed the RS statistic (1.15) in the form (1.16) and called it the Lagrangian Multiplier (LM) test. (In econometric literature, the RS test is generally referred to as the LM test.)

1.2.5 Special form of composite hypothesis

In many problems, the p-vector parameter θ consists of two parts, θ_1 an r vector and θ_2 a $(p-r)$ vector and the null hypothesis is of the form $H_0 : \theta_1 = \theta_{10}$ (a specified vector) and θ_2 (known as a nuisance parameter) is arbitrary. This becomes a special case of the composite hypothesis considered in Subsection 1.2.4 if we take $h(\theta) = \theta_1$. Denote the unrestricted ml of (θ_1, θ_2) by $(\hat{\theta}_1, \hat{\theta}_2)$ and its asymptotic covariance matrix by

$$\begin{aligned} cov(\hat{\theta}, \hat{\theta}) &= [I(\theta)]^{-1} \\ &= \begin{pmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{pmatrix}^{-1} = \begin{pmatrix} A & B \\ B' & C \end{pmatrix} \end{aligned}$$

where the partitions of the information matrix, I_{11}, I_{12}, and I_{22} are matrices of orders $r \times r$, $r \times (p-r)$ and $(p-r) \times (p-r)$, respectively, The Wald statistic can be written as

$$\begin{aligned} W &= (\hat{\theta}_1 - \theta_{10})'\hat{A}^{-1}(\hat{\theta}_1 - \theta_{10}),\ \hat{A} = A(\hat{\theta}) \\ &= (\hat{\theta}_1 - \theta_{10})' I_{1.2}(\hat{\theta})(\hat{\theta}_1 - \hat{\theta}_{10}) \end{aligned} \tag{1.17}$$

where

$$I_{1.2} = I_{11} - I_{12} I_{22}^{-1} I_{21}$$

the Schur complement of I_{22}.

To compute LR and RS statistics, we need to find the restricted ml estimates of θ_1, θ_2 under the restriction $\theta_1 = \theta_{10}$. Using the Lagrangian multiplier we have to maximize

$$L(\theta|x) - \lambda(\theta_1 - \theta_{10})$$

with respect to θ. The estimating equations are

$$s_1(\tilde{\theta}) = \lambda,\ s_2(\tilde{\theta}) = 0,\ \tilde{\theta}_1 = \theta_{10}.$$

The Rao score statistic is

$$\begin{aligned} \mathrm{RS} &= [s_1(\tilde{\theta})', 0']'[I(\tilde{\theta})]^{-1}[s_1(\tilde{\theta})', 0'] \\ &= [s_1(\tilde{\theta})]'[I_{1.2}(\tilde{\theta})]^{-1}[s_1(\tilde{\theta})] \\ &= \lambda'[I_{1.2}(\tilde{\theta})]^{-1}\lambda. \end{aligned} \tag{1.18}$$

The LR statistic is

$$\mathrm{LR} = 2\left[L(\hat{\theta}) - L(\tilde{\theta})\right] \tag{1.19}$$

All the three statistics have asymptotically chi-square distribution on r d.f.

1.3 Neyman's $C(\alpha)$ Test and Neyman-Rao Test

Neyman (1959, 1979) considered the problem of testing the hypothesis $H_0 : \theta_1 = \theta_{10}$ (given) and $\theta_2, \ldots, \theta_p$ are arbitrary (nuisance) parameters. Hall and Mathiason (1990) considered the more general problem of testing the composite hypothesis

$$H_0 : \theta_1 = \theta_{10}, \ldots, \theta_q = \theta_{q0} \text{ and } \theta_{q+1}, \ldots, \theta_p$$

are arbitrary by generalizing Neyman's results using the type of the argument used in Rao (1948) as in Section 1.2.3. Consider the slope of the power curve in the direction $(a_1, \ldots, a_q, 0, \ldots, 0)$

$$a_1 s_1 + \cdots + a_q s_q$$

where s_i is the derivative of the log likelihood with respect to θ_i, and define the Neyman statistic N as

$$N = \max_a \frac{(a_1 s_1 + \cdots + a_q s_q)^2}{V(a_1 s_1 + \cdots + a_q s_q)} \tag{1.20}$$

subject to

$$cov(s_i, a_1 s_1 + \cdots + a_q s_q) = 0,\ i = q+1, \ldots, p. \tag{1.21}$$

Using notation

$$\begin{aligned} S_1 &= (s_1, \ldots, s_q)', \; S_2 = (s_{q+1}, \ldots, s_p)', \\ a &= (a_1, \ldots, a_q)', \\ E(S_1 S_1') &= I_{11}, \; E(S_1 S_2') = I_{12}, E(S_2 S_2') = I_{22}, \end{aligned}$$

the problem (1.20), (1.21) can be written as

$$N(\Theta_{10}, \Theta_2) = \max_a \frac{(a' s_1)^2}{a' I_{11} a} \tag{1.22}$$

subject to $I_{21} a = 0$, where $\Theta_{10} = (\theta_{10}, \ldots, \theta_{q0})', \Theta_2 = (\theta_{q+1}, \ldots, \theta_p)'$. Using standard algebra, the optimum N is obtained as

$$N(\Theta_{10}, \Theta_2) = (S_1 - I_{12} I_{22}^{-1} S_2)' (I_{1.2})^{-1} (S_1 - I_{12} I_{22}^{-1} S_2) \tag{1.23}$$

where $I_{1.2} = I_{11} - I_{12} I_{22}^{-1} I_{21}$.

Neyman chose $\sqrt{n}$ as the consistent estimate of Θ_2 to obtain his statistic

$$N = N\left(\Theta_{10}, \tilde{\Theta}_2\right). \tag{1.24}$$

This form of the N statistic, obtained as a generalization of Neyman's single parameter test, is called the Neyman–Rao test by Hall and Mathiason (1990). The asymptotic distribution of N as in (1.24) is chi-square on q degrees of freedom. If Θ_2 is estimated by the constrained ml method, the test reduces to the RS test (1.19).

1.4 Some Examples of the RS Test

Godfrey (1988) gives a comprehensive account of the applications of the RS test in econometrics. A few examples mentioned in the paper by Bera and Ullah (1991) are as follows.

Chi-square goodness-of-fit: Given a parametric specification of the cell probabilities in a multinomial distribution, Pearson developed the chi-square goodness-of-fit test based on observed frequencies. This test can be seen to be the RS test of a composite hypothesis [Rao (1948)].

Linear model: The analysis of the linear model $y_i = x_i'\beta + \epsilon_i$, $i = 1, \ldots, n$, is based on four basic assumptions: correct linear functional form, normality of the distribution of the error term, homoscedasticity and serial independence. The RS test for normality has been derived by Bera and Jarque (1981), for homoscedasticity by Breusch and Pagan (1979). for serial independence by Breusch (1978) and Godfrey (1978a,b) and for linearity by Byron (1968).

For further examples and interpretation of several well-known tests in terms of the score functions, reference may be made to Bera and Ullah (1991) and the papers in the special issue on Rao's score test, Vol. **97**. pp. 1–200 of *Journal of Statistical Planning and Inference* (2001).

1.5 Some Advantages of the RS Test

1. In general, it is simple to compute the RS statistic as it depends only on estimates of parameters under H_0.

2. The test is invariant under transformation of the parameters, unlike the Wald test (see Section 1.6 for examples). Transformation of parameters may simplify the estimation of parameters without effecting the value of the statistic.

3. The RS test has the same local efficiency as the Wald and LR tests.

4. The distribution of RS is not affected by parameters being on the boundary of the parameter space under H_0. In such a case the LR test, and in some cases the W test, is not applicable.

5. There are situations where nuisance parameters are not identifiable under H_0 leading to singular information matrix. In such cases the RS test can be suitably modified as illustrated in Davies (1977, 1987).

1.6 Some Anomalies

1.6.1 Behavior of the power function

The LR, W and RS tests are consistent in the sense that for a fixed alternative to the null hypothesis the power tends to unity as the sample size $n \to \infty$. However, for a fixed sample size, the power function may not be monotonically increasing with increase in the distance (defined in some sense) of the alternative hypothesis from the null.

Example 1.6.1 Let $x_1, \ldots, x_n$ be an iid sample from the Cauchy distribution with density $\pi^{-1}[1 + (x - \theta)^2]^{-1}$.

The RS test for $H_0 : \theta = \theta_0$ against the alternatives $\theta > \theta_0$ rejects when

$$2\sqrt{\frac{2}{n}}\sum_{i=1}^{n}\frac{(x_i-\theta_0)}{1+(x_i-\theta_0)^2} \geq u_\alpha. \tag{1.25}$$

As the alternative $\theta \to \infty$, $\min(x_i - \theta_0) \to \infty$ in probability, so that for fixed n, the left-hand side of (1.25) tends to zero. Since $u_\alpha > 0$ (for $\alpha < 1/2$), the power of the test as $\theta \to \infty$ for fixed n tends to zero. [See Lehmann (1999, p. 532) for further details].

Example 1.6.2 Let $x_1, \ldots, x_n$ be independent binary response variables such that

$$P(x_i = 1) = \left[1 + \exp\left(-\sum_{i=1}^{q}\beta_i z_{ij}\right)\right]^{-1}, \; i = 1, \ldots, n, \tag{1.26}$$

where $z_{i1} = 1$ and $z_{i1}, \ldots, z_{iq}$ are observations on q covariables. To test the hypothesis $H_0 : \beta_q = 0$ against the alternative $H : \beta_q \neq 0$, the Wald statistic is

$$W = \hat{\beta}_q^2/\hat{i}_{qq} \tag{1.27}$$

where $\hat{\beta}_q$ is the ml estimate of β_q and $\hat{i}_{qq}$ is the estimated variance of $\hat{\beta}_q$. Hauck and Donner (1977) show that for fixed n, $W \to 0$ as $\beta_q \to \infty$ for fixed $\beta_1, \ldots, \beta_{q-1}$, so that the power of the test decreases as β_q increases. For further examples of such anamolies associated with Wald's statistic, reference may be made to Vaeth (1985) and Le Cam (1990).

The above examples do not contradict the claims made about RS and W about the local power of the tests. Nonetheless, they suggest a caution in the use of these tests [see Mantel (1987)]. It would be of interest to construct an example of the type of anomaly noted above for the RS and W tests in the case of the LR test.

1.6.2 Examples of non-invariance of the Wald test

The Wald test is not invariant for transformations of the parameter while the LR and RS statistics are. Different choices of parameters using the Wald statistic may lead to different inferences.

Example 1.6.3 Consider the likelihood $P(X, \theta)$ based on observed data X and a single unknown parameter θ. Let $\hat{\theta}$ be the ml estimate of θ and $I(\hat{\theta})$, the estimated information.

The Wald statistic for testing the hypothesis $H_0 : \theta = 0$ is

$$\hat{\theta}\sqrt{I(\hat{\theta})} \tag{1.28}$$

which is asymptotically distributed as $N(0,1)$. An equivalent hypothesis is $H_0 : \theta^3 = 0$ and the Wald test based on the parameter θ^3 (using the δ-method to compute the variance of θ^3, Rao (1973, p. 388) is

$$\left(\hat{\theta}^3/3\hat{\theta}^2\right)\sqrt{I(\hat{\theta})} = \frac{\hat{\theta}}{3}\sqrt{I(\hat{\theta})} \tag{1.29}$$

which is asymptotically normal as $N(0,1)$. The p-values based on (1.28) and (1.29) can be quite different.

Example 1.6.4 [Gregory and Veal (1985)]. Consider the linear model

$$y = \beta x + \gamma z + u, \ u \sim N(0, \sigma^2) \tag{1.30}$$

and tests based on n observations. Let $\hat{\beta}$ and $\hat{\gamma}$ be the maximum likelihood estimates (MLEs) of β and γ with the estimated variance-covariance matrix

$$\hat{\sigma}^2 \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$$

where $\hat{\sigma}$ is the least squares estimate of σ. To test the hypothesis $H_0 : \beta\gamma = 1$, the Wald statistic is

$$\frac{(\hat{\beta}\hat{\gamma} - 1)^2}{\hat{\sigma}^2(\hat{\gamma}^2 w_{11} + 2\hat{\beta}\hat{\gamma}w_{12} + \hat{\beta}^2 w_{22})} \tag{1.31}$$

which is asymptotically chi-square on 1 d.f., while the test for the equivalent hypothesis $\beta = \gamma^{-1}$ is

$$\frac{(\hat{\beta} - \hat{\gamma}^{-1})^2}{\hat{\sigma}^2(w_{11} + 2\hat{\gamma}^{-2}w_{12} + \hat{\gamma}^{-4}w_{22})} = \frac{(\hat{\beta}\hat{\gamma} - 1)^2}{\sigma^2(\hat{\gamma}^2 w_{11} + 2w_{12} + \hat{\gamma}^{-2}w_{22})} \tag{1.32}$$

which is different from (1.31) and is also asymptotically chi-square on 1 d.f.

For another example of non-invariance of Wald's test, reference may be made to Fears, Benichow and Gail (1996) and Pawitan (2000).

1.6.3 Weak dependence of the RS statistic on alternatives to the null hypothesis

In general, when the null hypothesis is rejected, one looks for alternative stochastic models for the observed data. The score test depends on the slope of the likelihood function at the null hypothesis. There may be different likelihoods all giving the same score statistic. If the score test is significant, there is no way of knowing what the alternative is.

Test for normality: Suppose we start with the Pearson family or Gram-Charlier type of distributions and construct a test for normality. The same RS statistic is obtained for both alternatives [Bera and Bilias (2001)].

Test for homoscedasticity: The RS statistic for testing homoscedasticity is the same for alternatives such as multiplicative and additive homoscedasticity [Breusch and Pagan (1979) and Godfrey and Wickens (1981)].

Testing for serial independence: The RS statistic for testing serial independence is the same whether we consider as alternatives the pth order autoregressive or pth order moving average model [Breusch (1978), Godfrey (1978a)].

Such difficulties may exist with other test criteria and it would be of interest to construct some examples.

1.7 Power Comparisons

The following is a summary of numerous papers devoted to power comparisons of LR, W and RS tests.

Taniguchi (1988; 1991): The first-order local powers are the same for all the tests. The second-order local powers are different but no one dominates the other.

Taniguchi (2001): In terms of Bahadur efficiency, they are the same up to the second order.

Bing Li (2001): They are all sensitive to changes in the values of the nuisance parameters.

Chandra and Joshi (1983): Rao's test is more powerful to the order $(1/n)$ than LR and W, when one modifies the critical regions to have the same size up to order $(1/n)$.

Ghosh and Mukherjee (2001): RS is more (or equally) efficient than LR and W under the criteria of maximinity and average local power. See also Mukherjee (1990, 1993) for results on asymptotic efficiency of Rao's Score.

Further investigation of power properties of LR, W and RS tests would be of interest.

1.8 Some Recent Developments

In this Section, we consider some modifications and refinements made on the RS statistic and indicate the need for further research in some cases.

1. In testing a composite hypothesis the estimated score vector $s(\tilde{\theta})$, where $\tilde{\theta}$ is the restricted ml estimate of θ under the hypothesis, is used in computing the RS statistic. It was argued that $s(\tilde{\theta})$ is close to zero if the hypothesis is true. But $E[s(\tilde{\theta})]$ may not be zero unless the null hypothesis is a simple one. In such a case Conniffe (1990) suggested the use of the quadratic form

$$\left[s(\tilde{\theta}) - Es(\tilde{\theta})\right]' J(\tilde{\theta}) \left[s(\tilde{\theta}) - Es(\tilde{\theta})\right] \tag{1.33}$$

   where J is the inverse of the covariance matrix of $s(\tilde{\theta}) - E[s(\tilde{\theta})]$. The computation of (1.33) and its improvement over the RS statistic needs further study.

2. White (1982) developed score type of statistics based on estimating equations and the quasi-likelihood functions. This introduces some robustification in inference procedures. See also Godfrey and Orme (2001).

3. Several authors tried to adjust the RS statistic similar to a Bartlett (1937) type of adjustment to the LR statistic. Harris (1985) suggested an adjustment based on Edgeworth-type expansion. Dean and Lawless (1989) suggested a different type of adjustment in certain models. Ghosh and Mukherjee (2001) developed a method of adjustment when the RS statistic is based on quasilikelihood. This is an area where further research is needed. Reference may also be made to a recent contribution by Tu, Chen and Shi (2004) on Bartlett type correction to the Score test in the Cox regression model.

4. The RS statistic (1.8) for testing a simple hypothesis $H_0 : \theta = \theta_0$ is

$$[s(\theta_0)]' \left[I(\theta_0)\right]^{-1} [s(\theta_0)]$$

   which involves the computation of the information matrix

$$I(\theta_0) = E\left[s(\theta_0)s(\theta_0)'\right].$$

   Instead of $I(\theta_0)$, one could use the $p \times p$ matrix of second derivatives of the log likelihood with a minus sign

$$A(\theta) = -\left(\frac{\partial^2 L}{\partial\theta_r \partial\theta_s}\right) \tag{1.34}$$

leading to the statistic

$$[s(\theta_0)]' (A(\theta_0))^{-1} [s(\theta_0)]. \tag{1.35}$$

Terril (2001) suggests further simplification by using what he calls the gradient statistic

$$F^2 = [s(\theta_0)]'(\hat{\theta} - \theta_0) \tag{1.36}$$

where $\hat{\theta}$ is the ml estimate of θ. The suggestion by Terril is attractive as it is simple to compute. It would be of interest to investigate the performance of the statistic (1.36).

5. In considering the score statistic, Rao (1948) used the ml estimates of parameters. A similar theory can be developed using BAN estimators.

6. Rao (1951) suggested the use of score tests in sequential analysis for testing a simple null versus a simple alternative hypothesis. Bradley (1953) considered a nice application of Rao's sequential test in clinical trials. An application in quality control is given by Box and Ramirez (1992). For some comments on sequential score test and possible applications reference may be made to Sen (1997).

Acknowledgments: I am gratified to see the large number of papers contributed by econometricians on the application of the score statistic to problems in econometrics and the extensions and improvements they have made. My special thanks are due to Professor A. K. Bera whose papers provided me all the information I needed in preparing this chapter.

References

1. Aitchison, J. and Silvey, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints, *Annals of Mathematical Statistics*, **29**, 813–828.

2. Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press, Boston, Massachusetts.

3. Bartlett, M. S. (1937). Properties of sufficiency and statistical tests, *Proceedings of the Royal Society, Series A*, **160**, 268–282.

4. Bera, A. K. and Bilias, Y. (2001). Rao's score, Neyman's $C(\alpha)$) and Silvey's LM tests: An essay on historical developments and new results, *Journal of Statistical Planning and Inference*, **97**, 9–44.

5. Bera, A. K. and Jarque, C. M. (1981). An efficient large sample test for normality of observations and regression residuals, *Working Papers in Economics and Econometrics*, **40**, The Australian National University, Canberra, Australia.

6. Bera, A. K. and Ullah, A. (1991). Rao's score test in Econometrics, *Journal of Quantitative Economics*, **7**, 189–220.

7. Bickel, P. J. and Doksum, K. A. (2001). *Mathematical Statistics*, Vol. 1, Second Edition, Prentice-Hall, Inglewood Cliffs, New Jersey.

8. Box, G. E. P. and Raminez, J. (1992). Cumulative score charts, *Quality and Reliability Engineering*, **8**, 17–27.

9. Bradley, R. A. (1953). Some statistical methods in taste testing and quality evaluation, *Biometrics*, **9**, 22–38.

10. Breusch, T.S. (1978). Testing for autocorrelation in dynamic linear models, *Australian Economic Papers*, **17**, 334–355.

11. Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation, *Econometrics*, **47**, 1287–1294.

12. Breusch, T. S. and Pagan, A. R. (1980). The Lagrange Multiplier test and its application to model specification in econometrics, *Review of Economic Studies*, **47**, 239–253.

13. Byron, R. P. (1968). Methods for estimating demand equations using prior information, A series of experiments with Australian data, *Australian Economic Papers*, **7**, 227–248.

14. Chandra, T. K. and Joshi, S. N. (1983). Comparison of the likelihood ratio, Wald's and Rao's tests, *Sankhyā, Series A*, **45**, 226–246.

15. Chandra, T. K. and Mukherjee, R. (1984). On the optimality of Rao's statistic, *Communications in Statistics—Theory and Methods*, **13**, 1507–1515.

16. Chandra, T. K. and Mukherjee, R. (1985). Comparison of the likelihood ratio, Wald's and Rao's tests, *Sankhyā, Series A*, **47**, 271–284.

17. Chandra, T. K. and Samanta, T. (1988). On second order local comparisons between perturbed maximum likelihood estimators and Rao's statistic as test statistics, *Journal of Multivariate Analysis*, **25**, 201–222.

18. Conniffe, D. (1990). Testing hypotheses with estimated scores, *Biometrika*, **77**, 97–106.

19. Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **64**, 247-254.

20. Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **74**, 33–43.

21. Dean, C. and Lawless, J. F. (1989). Tests for detecting over dispersion in Poisson regression models, *Journal of the American Statistical Association*, **84**, 467–472.

22. Engle, R. F. (1984). Wald, likelihood ratio and Lagrangian Multiplier test in econometrics, In *Handbook of Econometrics*, Vol. 2 (Eds., Z. Griliches and M. Intriligator), North-Holland Science Publishers, Amsterdam.

23. Fears, T. R., Benichow, J. and Gail, M. H. (1996). A remainder of the fallibility of the Wald statistic, *The American Statistician*, **50**, 226–227.

24. Ghosh, J. K. (1991). Higher order asymptotics for the likelihood ratio, Rao's and Wald's test, *Statistics & Probability Letters*, **12**, 505–509.

25. Ghosh, J. K. and Mukherjee, R. (2001). Test statistic arising from quasi-likelihood: Bartlett adjustment and higher-order power, *Journal of Statistical Planning and Inference*, **97**, 45–55.

26. Godfrey, L. G. (1978a). Testing against general autoregressive and moving average models when the regression include lagged dependent variables, *Econometrica*, **46**, 227–236.

27. Godfrey, L. G. (1978b). Testing for higher order serial correlation in regression equations when the regression includes lagged dependent variables, *Econometrica*, **46**, 1303–1310.

28. Godfrey, L. G. (1988). *Misspecification Tests in Econometrics*, Cambridge University Press, London, England.

29. Godfrey, L. G. and Orme, C. D. (2001). On improving the robustness and reliability of Rao's score test, *Journal of Statistical Planning and Inference*, **97**, 153–176.

30. Godfrey, L. G. and Wickens, M. R. (1981). Testing linear and log-linear regression for functional form, *Review of Economic Studies*, **48**, 487–496.

31. Green, W. H. (1990). *Econometric Analysis*, Macmillan Publishers Limited, Hampshire, England.

32. Gregory, A. W. and Veal, M. R. (1985). Formulating Wald tests of non-linear restrictions, *Econometrica*, **53**, 1465–1468.

33. Hall, W. J. and Mathiason, D. J. (1990). On large sample estimation and testing in parameter models, *International Statistical Review*, **58**, 77–97.

34. Harris, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistics, *Biometrika*, **72**, 653–659.

35. Harvey, A. (1990). *The Econometric Analysis of Time Series*, Philip Allan, Oxford, England.

36. Hauck, W. W. and Donner, A. (1977). Wald test as applied to hypotheses in logit analysis, *Journal of the American Statistical Association*, **72**, 851–853.

37. *Journal of Statistical Planning and Inference*, 2001, Vol. **97**, No. 1, Special Issue on Rao's Score Test (12 papers, pp. 1–200).

38. Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T. C. (1985). *The Theory and Practice of Econometrics*, Second edition, John Wiley & Sons, New York.

39. Kmenta, J. (1986). *Elements of Econometrics*, Second edition, Macmillan Publishers Limited, Hampshire, England.

40. Kramer, W. and Sonnberger, H. (1986). *The Linear Regression Model Test*, Physica-Verlag, Heidelberg, Germany.

41. Le Cam, L. (1990). On the standard asymptotic confidence ellipsoids of Wald, *International Statistical Review*, **58**, 129–152.

42. Lehmann, E. (1999). *Theory of Asymptotic Inference*, Springer-Verlag, New York.

43. Li, Bing (2001). Sensitivity of Rao's score test, the Wald test and the likelihood ratio test to nuisance parameters, *Journal of Statistical Planning and Inference*, **97**, 57–66.

44. Maddala, G. S. (1988). *Introduction to Econometrics*, Macmillan Publishers Limited, Hampshire, England.

45. Mantel, N. (1987). Understanding Wald's test for exponential families, *The American Statistician*, **41**, 147–149.

46. Mukherjee, R. (1990). Comparison of tests in the multiparameter case I: Second order power, *Journal of Multivariate Analysis*, **33**, 17–30.

47. Mukherjee, R. (1993). Rao's score test: Recent asymptotic results, In *Handbook of Statistics 11* (Eds., G. S. Maddala, C. R. Rao and H. Vinod), pp. 363–379, North- Holland Science Publishers, Amsterdam.

48. Neyman, J. (1959). Optimal asymptotic test of composite statistical hypothesis, In *Probability and Statistics* (Ed., U. Grenander), John Wiley & Sons, New York.

49. Neyman, J, (1979). $C(\alpha)$ tests and their uses, *Sankhyā, Series A*, **41**, 1–21.

50. Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria, *Biometrika*, **20A**, 175–240, 263–294.

51. Pawitan, Y. (2000). A remainder of the fallibility of the Wald statistic: Likelihood explanation, *The American Statistician*, **54**, 54–56.

52. Peers, H. W. (1971). Likelihood ratio and associated test criteria, *Biometrika*, **58**, 577–587.

53. Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation, *Proceedings of the Cambridge Philosophical Society*, **44**, 50–57.

54. Rao, C. R. (1950). Methods of scoring linkage data giving the simultaneous segregation of three factors, *Heredity*, **4**, 37–59.

55. Rao, C .R. (1951). Sequential tests of null hypotheses, *Sankhyā*, **10**, 361–370.

56. Rao, C. R. (1961). A study of large sample test criteria through properties of efficient estimates. *Sankhyā, Series A*, **23**, 25–40.

57. Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, Second edition, John Wiley & Sons, New York.

58. Rao, C. R. and Poti, S. J. (1946). On locally most powerful tests when alternatives are one-sided, *Sankhyā*, **7**, 439–440.

59. Sen, P. K. (1997). Introduction to Rao (1948). In *Breakthroughs in Statistics* (Eds., S. Kotz and N. L. Johnson), Vol. III, Springer-Verlag, New York.

60. Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.

61. Silvey, S. D. (1959). The Lagrangian multiplier test, *Annals of Mathematical Statistics*, **30**, 389–407.

62. Spanos, A. (1986). *Statistical Foundations of Econometric Modeling*, Cambridge University Press, London, England.

63. Taniguchi, M. (2001). On large deviation asymptotics of some tests in time series, *Journal of Statistical Planning and Inference*, **97**, 191–200.

64. Taniguchi, M. (1988). Asymptotic expansions of the distribution of some statistics for Gaussian ARMA process, *Journal of Multivariate Analysis*, **27**, 494–511.

65. Taniguchi, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*, Lecture Notes in Statistics, Vol. 68, Springer-Verlag, New York.

66. Terril, G. R. (2001). The gradient statistic, *Personal communication.*

67. Tu, D., Chen, J. and Shi, P. (2004). A Bartlett type correction for Rao Score test in Cox regression model, *Technical Report.*

68. Vaeth, M. (1985). On the use of Wald's test in exponential families, *International Statistical Review*, **53**, 199–214.

69. Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Transactions of the Americal Mathematical Soceity*, **54**, 426–482.

70. White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica*, **50**, 1–25.

71. White, H. (1984). *Asymptotic Theory for Econometricians*, Academic Press, New York.