

## Three Score and 15 Years (1948-2023) of Rao's Score Test: A Brief History

Anil K. Bera<sup>1</sup> and Yannis Biliass<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign, U.S.A.

<sup>2</sup>Athens University of Economics and Business, Greece

Received: 15 June 2024; Revised: 03 September 2024; Accepted: 16 September 2024

आचख्युः कवयः केचित्संप्रत्याचक्षते परे ।  
आख्यास्यन्ति तथैवान्ये इतिहासमिमं भुवि ॥  
*ācakhyuḥ kavayaḥ kecitsampratyācakṣate pare |*  
*ākhyāsyanti tathavānye itihāsamimam bhuvi ||*

Mahabharata (~500BC) by Krishna-Dwaipayana Vyasa, Verse 1.1.24

“Some bards have already published this history, some are now teaching it,  
and others, in like manner, will hereafter promulgate it upon the earth.”

### Abstract

Rao (1948) introduced the score test statistic as an alternative to the likelihood ratio and Wald test statistics. In spite of the optimality properties of the score statistic shown in Rao and Poti (1946), the Rao score (RS) test remained unnoticed for almost 20 years. Today, the RS test is part of the “Holy Trinity” of hypothesis testing and has found its place in Statistics and Econometrics textbooks and related software. Reviewing the history of the RS test we note that remarkable test statistics proposed in the literature earlier or around the time of Rao (1948) mostly from intuition, such as Pearson (1900) goodness-fit-test, Moran (1948) **I** test for spatial dependence and Durbin and Watson (1950) test for serial correlation, can be given RS test statistic interpretation. At the same time, recent developments in the robust hypothesis testing under certain forms of misspecification, make the RS test an active area of research in Statistics and Econometrics. From our brief account of the history of the RS test we conclude that its impact in science goes far beyond its calendar starting point with promising future research activities for many years to come.

*Key words:* Applications to Econometrics and Statistics; Hypothesis testing; Rao's score; Robust tests; Sequential testing.

## 1. Prologue

C. R. Rao's work was always inspired by some practical problems. In 1946, he was deputed from the Indian Statistical Institute (ISI), Calcutta, to work on an anthropometric project in the Museum of Anthropology and Ethnology at the Cambridge University, U.K. While at Cambridge, Rao took the opportunity to contact R. A. Fisher, then the Belfour Professor of Genetics, and registered for a Ph.D. degree in Statistics under Fisher's guidance. As recollected in Rao (2001), Fisher agreed under the condition that Rao spends time in the Genetics Laboratory where Fisher was breeding mice to map their chromosomes. Rao started by mating mice of different genotypes to collect the necessary data and additionally, he was trying to develop appropriate statistical methodology to analyze the experimental data. The problem was estimation of linkage *parameters* (recombination probabilities in the various segments of the chromosomes) using data sets from different experiments, designed in such a way that each data set had information on the *same parameters*. It was thus necessary to test whether the parameters in different experiments are the same or not.

Rao wrote and published two papers based on this work. The *first* paper, Rao (1948), deals with the *general problem of testing* simple and composite hypotheses concerning a vector parameter. The test was based on the *scores*, derivatives of the log-likelihood function with respect to the individual parameters. The paper was published in the *Proceedings of the Cambridge Philosophical Society*, where he termed the test principle as a *score test*. In this paper, we will refer to it as the Rao score (RS) test. The *other* paper, Rao (1950a), contains the detail steps for *analyzing the data* involving the segregation of several factors in mating of different genotypes. And it used the RS test for the meta-analysis of testing the equality of parameters coming from different experimental data sets. That paper was published in Fisher's new journal *Heredity*. For more, see Rao (2001).

The rest of the paper is organized as follows. In Section 2, we start with the first principle of testing, namely the Neyman-Pearson Lemma and derive the simplest version of RS test and then discuss it in its full generality. There, we also provide RS test interpretation to some of the classic tests in Econometrics and Statistics, such as the quintessential Pearson (1900) goodness-fit-test, which was suggested mostly by pure intuition, but its theoretical foundation can be buttressed by RS test principle. In Sections 3 and 4, we list a (somewhat incomplete) catalogue of RS tests in Econometrics and Statistics. In Section 5, we outline some of the possible ways an assumed probability model can be misspecified, and discuss how the various RS tests can be robustified to make them valid under misspecification. We close the paper in Section 6 (Epilogue) with some concluding remarks. At the outset let us mention that while compiling the 75 years (from 1948 to 2023) history of the RS test, we have included here some of our own past historical accounts and cited accordingly. Our aim is to have a comprehensive review as far as possible at one place, like a one-stop-shopping for the RS test history.

## 2. Score as an optimal test function: Rao and Poti (1946)

We start by introducing some notation and concepts. Suppose we have  $n$  independent observations  $y_1, y_2, \dots, y_n$  on a random variable  $Y$  with density function  $f(y; \theta)$ , where  $\theta$  is a  $p \times 1$  parameter vector with  $\theta \in \Theta \subseteq \mathbb{R}^p$ . It is assumed that  $f(y; \theta)$  satisfies the regularity conditions stated in Rao (1973, p.364) and Serfling (1980, p.144). The likelihood function is

given by

$$\mathbf{L}(\theta, \mathbf{y}) \equiv \mathbf{L}(\theta) = \prod_{i=1}^n f(y_i; \theta), \tag{1}$$

where  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  denotes the sample. Suppose we want to test  $H_0 : \theta = \theta_0$  against  $H_1 : \theta \neq \theta_0$  based on the sample  $\mathbf{y}$ .

The foundation of the theory of hypothesis testing was laid by Neyman and Pearson (1933) fundamental lemma. This lemma provides a way to find the most powerful (MP) and uniformly most powerful (UMP) tests. According to the Neyman-Pearson (N-P) Lemma, the MP critical region for testing  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta = \theta_1$  having size  $\alpha$ , is given by

$$\omega(\mathbf{y}) = \{\mathbf{y} \mid \mathbf{L}(\theta_1) > \kappa \mathbf{L}(\theta_0)\}, \tag{2}$$

where  $\kappa$  is such that  $\Pr[\omega(\mathbf{y}) \mid H_0] = \alpha$ .

If an MP test maximizes powers uniformly in  $\theta_1 \in \Theta_1 \subseteq \Theta$ , the test is called UMP test. Unfortunately, an UMP test rarely exists, and when it does not, there is no single critical region best for all alternatives. We, therefore, try to find a critical region that is good for alternatives *close* to the null hypothesis, called *local* alternatives, hoping that the region will also be good for alternatives away from the null. Lehmann (1999, p.529) advocated for such critical region when the sample size  $n$  is large, stating, “if the true value is at some distance from  $\theta_0$ , a large sample will typically reveal this so strikingly that a formal test may be deemed unnecessary.”

For a critical region  $\omega(\mathbf{y})$ , let us define the power function as

$$\gamma(\theta) = \Pr[\omega(\mathbf{y}) \mid \theta] = \int_{\omega(\mathbf{y})} \mathbf{L}(\theta) d\mathbf{y}. \tag{3}$$

Assuming a scalar  $\theta$  and that  $\gamma(\theta)$  admits Taylor series expansion, we have

$$\gamma(\theta) = \gamma(\theta_0) + (\theta - \theta_0)\gamma'(\theta_0) + \frac{(\theta - \theta_0)^2}{2}\gamma''(\theta^*), \tag{4}$$

where  $\theta^*$  is a value in between  $\theta$  and  $\theta_0$ . If we consider local alternatives of the form  $\theta = \theta_0 + n^{-\frac{1}{2}}\delta, 0 < \delta < \infty$ , the third term will be of the order  $\mathbf{O}(n^{-1})$ . Therefore, from (4), to obtain the highest power, we need to maximize

$$\gamma'(\theta_0) = \left. \frac{\partial}{\partial \theta} \gamma(\theta) \right|_{\theta=\theta_0} = \frac{\partial}{\partial \theta} \int_{\omega(\mathbf{y})} \mathbf{L}(\theta) d\mathbf{y} = \int_{\omega(\mathbf{y})} \frac{\partial}{\partial \theta} \mathbf{L}(\theta) d\mathbf{y}, \tag{5}$$

for  $\theta > \theta_0$ , assuming the regularity conditions that allow differentiation under the sign of intergration.

Using the generalized N-P Lemma given in Neyman and Pearson (1936), it is easy to see that the locally most powerful (LMP) critical region for  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta > \theta_0$ , is given by

$$\omega(\mathbf{y}) = \left\{ \mathbf{y} \mid \frac{\partial}{\partial \theta} \mathbf{L}(\theta_0) > \kappa \mathbf{L}(\theta_0) \right\}$$

or

$$\omega(\mathbf{y}) = \left\{ \mathbf{y} \mid \frac{\partial}{\partial \theta} \ln(\mathbf{L}(\theta_0)) = \frac{\partial}{\partial \theta} l(\theta_0) > \kappa \right\}, \tag{6}$$

where  $l(\theta)$  denotes the log-likelihood function and  $\kappa$  is a constant such that the size of the test is  $\alpha$ . The quantity  $S(\theta) = \partial l(\theta)/\partial \theta$  is called the Fisher-Rao *score function*. The above result in (6) was first discussed in Rao and Poti (1946), who stated that a LMP critical region for  $H_0 : \theta = \theta_0$  is given by

$$\omega(\mathbf{y}) = \{ \mathbf{y} \mid \kappa_1 S(\theta_0) > \kappa_2 \}, \tag{7}$$

where  $\kappa_2$  is so determined that the size of the test is equal to a preassigned value  $\alpha$  with  $\kappa_1$  as  $+1$  or  $-1$ , respectively, for alternatives  $\theta > \theta_0$  and  $\theta < \theta_0$ .

Let us define the Fisher information as

$$\mathcal{I}(\theta) = -E \left[ \frac{\partial^2 l(\theta)}{\partial \theta^2} \right] = Var[S(\theta)]. \tag{8}$$

The result that under  $H_0$ ,  $S(\theta_0)$  is asymptotically distributed as normal with mean zero and variance  $\mathcal{I}(\theta_0)$ , led Rao and Poti (1946) to suggest a test based on  $S(\theta_0)/\sqrt{\mathcal{I}(\theta_0)}$  as standard normal [or  $S^2(\theta_0)/\mathcal{I}(\theta_0)$  as  $\chi_1^2$ ], for large  $n$ .

**2.1. From Rao and Poti (1946) to Rao (1948): Test for the multiparameter case**

Rao and Poti (1946) can be viewed as a precursor to Rao (1948). Generalization of the LMP test in (7) to the multiparameter case ( $p \geq 2$ ) is not trivial. There will be scores for each individual paramter, and the problem is to combine them in an “optimal” way. Let  $H_0 : \theta = \theta_0$ , where now  $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$  and  $\theta_0 = (\theta_{10}, \theta_{20}, \dots, \theta_{p0})'$ . Consider a *scalar* linear combination

$$\sum_{j=1}^p \delta_j \frac{\partial l(\theta)}{\partial \theta_j} = \delta' S(\theta), \tag{9}$$

where  $\delta = (\delta_1, \delta_2, \dots, \delta_p)'$  is a fixed vector and test the hypothesis  $H_{0\delta} : \delta' \theta = \delta' \theta_0$  against  $H_{1\delta} : \delta' \theta \neq \delta' \theta_0$ ,  $\delta \in \mathbb{R}^p$ .

We rewrite the Fisher information in (8) as

$$\mathcal{I}(\theta) = -E \left[ \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} \right]. \tag{10}$$

Asymptotically, under  $H_0$ ,  $\delta' S(\theta_0)$  is distributed as normal, with mean zero and variance  $\delta' \mathcal{I}(\theta_0) \delta$ . Thus if  $\delta$ 's were known, a test could be based on

$$\frac{[\delta' S(\theta_0)]^2}{\delta' \mathcal{I}(\theta_0) \delta}, \tag{11}$$

which under  $H_0$  will be distributed as  $\chi_1^2$  as in Rao and Poti (1946). Note that our  $H_0 : \theta = \theta_0$  for  $p \geq 2$  can be expressed as  $H_0 \equiv \bigcap_{\delta \in \mathbb{R}^p} H_{0\delta}$ , i.e., the multiparameter testing problem can be

decomposed into a series of *single-parameter* problems. To obtain a linear function like (9), Rao (1948) maximized (11) with respect to  $\delta$ . Using the Cauchy-Schwarz inequality

$$\frac{(u'v)^2}{u'Au} \leq v'A^{-1}v, \tag{12}$$

where  $u$  and  $v$  are column vectors and  $A$  is a non-singular matrix, we have

$$\sup_{\delta \in \mathbb{R}^p} \frac{[\delta'S(\theta_0)]^2}{\delta'\mathcal{I}(\theta_0)\delta} = S(\theta_0)'\mathcal{I}(\theta_0)^{-1}S(\theta_0). \tag{13}$$

In (13), the supremum reaches at  $\delta = \mathcal{I}(\theta_0)^{-1}S(\theta_0)$  and this provides an optimal linear combination of scores.

Roy (1953) used Rao's maximization technique (13) to develop his *union-intersection* (UI) method of testing. Let  $H_0 \equiv \bigcap_{j \in J} H_{0j}$ , where  $J$  is an index set. Roy's UI method gives the rejection region for  $H_0$  as the *union* of rejection regions for all  $H_{0j}$ ,  $j \in J$ . Consider testing  $H_{0\delta} : \delta'\theta = \delta'\theta_0$  against  $H_{1\delta} : \delta'\theta \neq \delta'\theta_0$ ,  $\delta \in \mathbb{R}^p$ . Let  $H_0 = \bigcap_{\delta \in \mathbb{R}^p} H_{0\delta}$  and  $H_1 \equiv \bigcap_{\delta \in \mathbb{R}^p} H_{1\delta}$ . If  $T_\delta$  is the likelihood ratio (LR) statistic for testing  $H_{0\delta}$  against  $H_{1\delta}$ , then

$$T = \sup_{\delta \in \mathbb{R}^p} T_\delta \tag{14}$$

is Roy's LR statistic for testing  $H_0$  against  $H_1$ . This is the same principle that was used by Rao (1948) to convert a "multivariate" problem into a series of "univariate" ones, as we have seen in equation (13).

When the null hypothesis is composite, like  $H_0 : h(\theta) = c$ , where  $h(\theta)$  is an  $r \times 1$  vector function of  $\theta$  with  $r \leq p$ , the general form of the RS test statistic is

$$RS = S(\tilde{\theta})'\mathcal{I}(\tilde{\theta})^{-1}S(\tilde{\theta}), \tag{15}$$

where  $\tilde{\theta}$  is the restricted maximum likelihood estimator (MLE) of  $\theta$ , i.e.,  $h(\tilde{\theta}) = c$ . Asymptotically, under  $H_0$ , the RS test statistic is distributed as  $\chi_r^2$ . Therefore, we observe *two* optimality principles behind the RS test; first, in terms of LMP test as given in (6), and second, in deriving the "optimal" direction for the multiparameter case, as in (13).

Rao (1948) suggested the score test as an alternative to the Wald (1943) statistic, which for testing  $H_0 : h(\theta) = c$  is given by

$$W = [h(\hat{\theta}) - c]' [H(\hat{\theta})'\mathcal{I}(\hat{\theta})^{-1}H(\hat{\theta})]^{-1} [h(\hat{\theta}) - c], \tag{16}$$

where  $\hat{\theta}$  is the unrestricted MLE of  $\theta$ , and  $H(\theta) = \partial h(\theta)/\partial \theta$  is a  $r \times p$  matrix with full column rank  $r$ . Rao (1948, p.53) stated that his test (15), "besides being simpler than Wald's has some theoretical advantages." For more on this see Bera (2000) and Bera and Biliias (2001).

Neyman and Pearson (1928) suggested their LR test as

$$LR = 2 \left[ \ln \frac{\mathbf{L}(\hat{\theta})}{\mathbf{L}(\tilde{\theta})} \right] = 2 [l(\hat{\theta}) - l(\tilde{\theta})]. \tag{17}$$

Their suggestion did not come from any search procedure satisfying an optimality criterion. It was purely based on intuitive grounds; as Neyman (1980, p.6) stated, “The intuitive background of the likelihood ratio test was simply as follows: if among the contemplated admissible hypotheses there are some that ascribe to the facts observed probabilities much larger than that ascribed by the hypothesis tested, then it appears ‘reasonable’ to reject the null hypothesis.”

The three statistics RS, W, and LR, given respectively in (15), (16), and (17) are referred to as the “Holy Trinity.” These tests can be viewed as three different *distance measures* between  $H_0$  and  $H_1$ . When  $H_0$  is true, we should expect the restricted and unrestricted MLEs of  $\theta$ , namely  $\tilde{\theta}$  and  $\hat{\theta}$  to be close, and likewise the log-likelihood functions  $l(\tilde{\theta})$  and  $l(\hat{\theta})$ , respectively. The LR statistic in (17) measures the distance through the log-likelihood function and is based on the difference  $l(\hat{\theta}) - l(\tilde{\theta})$ . To see the intuition behind the RS test, note that  $S(\hat{\theta}) = 0$  by construction, and thus we should expect  $S(\tilde{\theta})$  to be close to zero if  $H_0$  is true. Therefore, the basis of the RS test is  $S(\tilde{\theta}) - S(\hat{\theta}) = S(\tilde{\theta})$ , distance between  $\tilde{\theta}$  and  $\hat{\theta}$  measured through the function  $S(\theta)$ . Finally to test  $H_0 : h(\theta) = c$ , W considers the distance directly in terms of  $h(\theta)$ , and is based on  $[h(\hat{\theta}) - c] - [h(\tilde{\theta}) - c] = h(\hat{\theta}) - c$ , where  $h(\tilde{\theta}) = c$  by construction, as we see in expression (16). It is interesting to note the similarity between the Wald and the RS tests based on  $h(\hat{\theta})$  and  $S(\tilde{\theta})$ , respectively. Therefore the RS test statistic is closer to W than LR. Therefore, it makes sense that Rao (1948, p.53) mentioned his test as an alternative to W.

The interrelationships among these three tests can be brought home to the students of Statistics through the following amusing story [see Bera and Premaratne (2001, p.58)]: Once around 1946 Ronald Fisher invited Jerzy Neyman, Abraham Wald, and C.R. Rao to his Cambridge University lodge for afternoon tea. During their conversation, Fisher mentioned the problem of deciding whether his dog, who had been going to an “obedience school” for some time, was disciplined enough. Neyman quickly came up with an idea: leave the dog free for some time and then put him on leash. If there is not much difference in his behavior, the dog can be thought of as having completed the course successfully. Wald, who lost his family in the concentration camps, was adverse to any kind of restrictions and simply suggested leaving the dog free and seeing whether it behaved properly. Rao, who had observed the nuisances of stray dogs in Calcutta streets, did not like the idea of letting the dog roam freely, and suggested keeping the dog on a leash at all times and observing how hard it pulls on the leash. If it pulled too much, it needed more training. That night when Rao was back in his Cambridge dormitory after tending Fisher’s mice at the genetics laboratory, he suddenly realized the connection of Neyman and Wald’s recommendations to the Neyman–Pearson LR and Wald tests, respectively. He got an idea and the rest, as they say, is history.

At this stage, it will be instructive to provide a geometric illustration highlighting the fundamental connections and contrasts among the three tests [see Bera (1983, pp.56-60)]. For simplicity, let us consider the case of scalar  $\theta$ , i.e.,  $p = 1$ , and that the null hypothesis is  $H_0 : \theta = \theta_0$ . In Figure 1, we plot the score function  $S(\theta) = dl(\theta)/d\theta$  against  $\theta$ , the solid curved line. The unrestricted MLE  $\hat{\theta}$  is obtained by setting  $S(\hat{\theta}) = 0$ , i.e., at the point D.

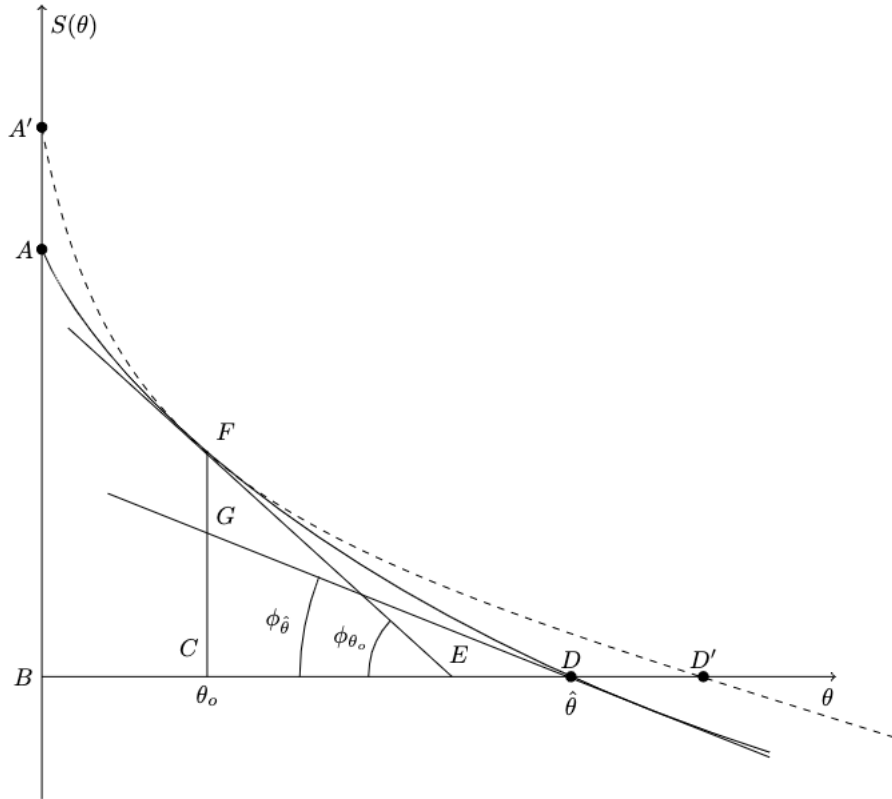


Figure 1: Geometry of LR, W and RS statistics

From the Figure 1, it is easily seen that

$$\begin{aligned} l(\hat{\theta}) - l(\theta_0) &= \int_{\theta_0}^{\hat{\theta}} S(\theta) d(\theta) = \\ &= \text{Area under the curve } S(\theta) \text{ from } \theta_0(\text{point C}) \text{ to } \hat{\theta}(\text{point D}). \end{aligned} \quad (18)$$

Therefore,

$$LR = 2[l(\hat{\theta}) - l(\theta_0)] = 2 \cdot \text{Area}(CDF). \quad (19)$$

For our particular case,  $h(\theta) = \theta - \theta_0$ ,  $H(\theta) = 1$  and  $c = 0$ . Thus, W in (16) can be expressed as

$$W = (\hat{\theta} - \theta_0)^2 \mathcal{I}(\hat{\theta}) = CD^2 \cdot \mathcal{I}(\hat{\theta}). \quad (20)$$

$\mathcal{I}(\hat{\theta})$  can be obtained from  $-d^2l(\theta)/d\theta^2 = -dS(\theta)/d\theta$ , evaluated at  $\theta = \hat{\theta}$ , i.e., from  $\tan \phi_{\hat{\theta}} = CG/CD$ . Therefore,

$$W = CD^2 \cdot \frac{CG}{CD} = CD \cdot CG = 2 \cdot \text{Area}(\triangle CDG). \quad (21)$$

On the other hand, the RS test will be based on  $S(\theta)$  at  $\theta_0$ , i.e., on the distance CF. The variance of  $S(\theta_0)$  can be estimated by  $-dS(\theta_0)/d\theta = \tan \phi_{\theta_0} = CF/CE$ . Hence,

$$RS = CF^2 \cdot \frac{CE}{CF} = CF \cdot CE = 2 \cdot \text{Area}(\triangle CEF). \quad (22)$$

From above, we can note the following features of the three tests. *First*, since the tests are based on three different areas in general, they will yield conflicting inference if the same critical value is used [see Berndt and Savin (1977)]. *Second*, the RS tests depends only on  $S(\theta)$  and the slope of  $S(\theta)$  at  $\theta_0$ . We can draw many curved lines through F with the same slope at F, and the dotted line  $A'FD'$  is an example. This implies that there may be other  $S(\theta)$  functions, i.e., other likelihood function representing different alternative hypothesis, with the same slope at  $\theta_0$ , giving rise to the same RS test statistic. In the literature, this property is known as invariance property of the RS test principle [see Godfrey (1988, p.70)]. *Finally*, for both the RS and W tests, the variances can be calculated in a number of ways which are asymptotically equivalent. This can lead to different versions of the test statistics. It is not clear which versions will give better results in finite samples.

**Example 1:** Let us start with a simple example where  $y_i \sim IIDN(\theta, 1)$ ,  $i = 1, 2, \dots, n$ , and we test  $H_0 : \theta = \theta_0 = 0$  against  $\theta > 0$ . Here the log-likelihood and score functions are respectively

$$l(\theta) = \text{Constant} - \frac{1}{2} \sum_{i=1}^n (y_i - \theta)^2, \quad (23)$$

$$\text{and } S(\theta) = \sum_{i=1}^n (y_i - \theta) = n(\bar{y} - \theta),$$

where  $\bar{y} = \sum_i y_i/n$ . Note that here  $S(\theta)$  is *linear* in  $\theta$ , and thus from Figure 1, all the three tests LR, W, and RS will be identical. Given that  $S(\theta_0) = n\bar{y}$  with  $Var[S(\theta_0)] = n$ , we will reject  $H_0$ , if  $\sqrt{n}\bar{y} > \mathbf{z}_\alpha$ , where  $\mathbf{z}_\alpha$  is the upper  $\alpha$  percent cut-off point of standard normal distribution. For fixed  $n$ , the power of this test goes to 1 as  $\theta \rightarrow \infty$ . Hence the score test  $\sqrt{n}\bar{y} > \mathbf{z}_\alpha$  is not only LMP, but also UMP for all  $\theta > 0$ .

**Example 2:** [Ferguson (1967, p.235)] Consider testing for the median of a Cauchy distribution with density

$$f(y; \theta) = \frac{1}{\pi} \cdot \frac{1}{1 + (y - \theta)^2}, \quad -\infty < y < \infty. \quad (24)$$

Since here  $\mathcal{I}(\theta) = n/2$ , the RS test will reject  $H_0 : \theta = \theta_0$  against  $H_1 : \theta > \theta_0$ , if

$$\frac{S(\theta_0)}{\sqrt{\mathcal{I}(\theta_0)}} = \sqrt{\frac{2}{n}} \sum_{i=1}^n \frac{2(y_i - \theta_0)}{1 + (y_i - \theta_0)^2} > \mathbf{z}_\alpha. \quad (25)$$

As  $\theta \rightarrow \infty$  with  $n$  remaining fixed,  $\min(y_i - \theta_0) \xrightarrow{p} \infty$ , and  $S(\theta_0)/\sqrt{\mathcal{I}(\theta_0)} \xrightarrow{p} 0$ . Thus the power of the test tends to zero as  $\theta \rightarrow \infty$ . Therefore what works for *local* alternatives may not work for *not-so-local* alternatives. This is in contrast to Example 1 where the LMP test is also the UMP.

In the example below we illustrate one of the most famous tests in the Statistics literature that was suggested long before 1948 and the theoretical foundation of which can be buttressed by the RS test principle.

**Example 3:** [Pearson (1900) Goodness-of-fit test]. Consider a multinomial distribution with  $p$  classes and let the probability of an observation belonging to the  $j$ -th class be  $\theta_j (\geq 0)$ ,



$j = 1, 2, \dots, p$ , so that  $\sum_{i=1}^p \theta_j = 1$ . Denote the observed frequency of the  $j$ -th class by  $n_j$  with  $\sum_{j=1}^p n_j = n$ . We are interested in testing  $\theta_j = \theta_{j0}$ ,  $j = 1, 2, \dots, p$ , where  $\theta_{j0}$  are known constants. Pearson (1900) suggested the statistic

$$P = \sum_{j=1}^p \frac{(n_j - n\theta_{j0})^2}{n\theta_{j0}} = \sum \frac{(O - E)^2}{E}, \quad (26)$$

where  $O$  and  $E$  denote respectively, the observed and expected frequencies. Given the profound importance of  $P$  in almost all branches of science, we demonstrate the theoretical underpinnings of  $P$  based on the RS test principle. The log-likelihood function, score and information matrix are respectively, given by [see Bera and Biliias (2001, p.17)]

$$l(\theta) = Constant + \sum_{j=1}^p n_j \ln(\theta_j) \quad (27)$$

$$S(\theta)_{[(p-1) \times 1]} = \begin{bmatrix} \frac{n_1}{\theta_1} - \frac{n_p}{\theta_p} \\ \frac{n_2}{\theta_2} - \frac{n_p}{\theta_p} \\ \dots \\ \frac{n_{p-1}}{\theta_{p-1}} - \frac{n_p}{\theta_p} \end{bmatrix} \quad (28)$$

and

$$\mathcal{I}(\theta)_{[(p-1) \times (p-1)]} = n \left[ \text{diag} \left( \frac{1}{\theta_1}, \frac{1}{\theta_2}, \dots, \frac{1}{\theta_{p-1}} \right) + \frac{1}{\theta_p} \mathbf{1}\mathbf{1}' \right] \quad (29)$$

where  $\mathbf{1} = (1, 1, \dots, 1)'$  is a  $(p - 1) \times 1$  vector of ones. We end up with effectively  $(p - 1)$  parameters since  $\sum_{j=1}^p \theta_j = \sum_{j=1}^p \theta_{j0} = 1$ . Using the above expressions, it is easy to see that

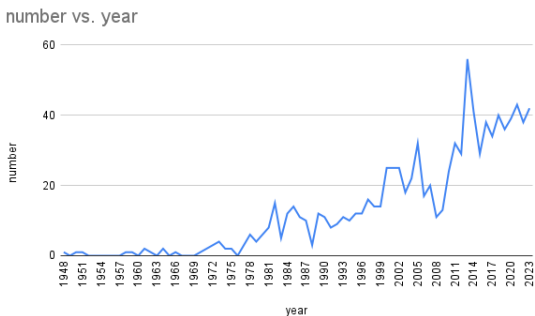
$$S(\theta_0)' \mathcal{I}(\theta_0)^{-1} S(\theta_0) = P, \quad (30)$$

where  $\theta_0 = (\theta_{10}, \theta_{20}, \dots, \theta_{p0})'$  [see Rao (1973, p.442) and (Cox and Hinkley, 1974, p.316)]. The coincidence that  $P$  is same as the RS test, is an amazing result. Pearson (1900) suggested his test mostly based on intuitive grounds almost 50 years before Rao (1948).

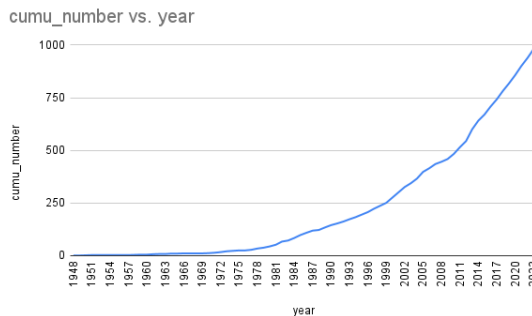
### 3. Some applications of the RS test in econometrics

RS test was well ahead of its time. It went unnoticed for very many years. It is fair to say that econometricians can claim major credit in recognizing its importance and applying the RS test in several useful contexts and coming up with closed form, neat test statistics. Rao himself acknowledged this fact by writing [see Rao (2005, p.15)] "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." More recently, statisticians are catching up with innovative applications. To obtain a quantitative perception of the influence of Rao (1948), we plot the yearly citations for the last 75 years in Figure 2. The corresponding cumulative citations are depicted in Figure 3. *First* thing to note is that the total number of citations in the last 75 years is only 980, apparently a very low number for such a seminal paper. Of course, we need to take into consideration of the fact that there are many papers, especially in the Statistics literature, that use the score test without making any reference to Rao (1948). *Second*, there are only a handful citations

during the first thirty years, i.e., until around 1978. That was the time econometricians recognized the usefulness of the Rao test principle, and used it in developing several model specification tests. There was another surge in its use after another 30 years, i.e., around 2008, in both the Statistics and Econometrics literature. *Finally*, from both Figures 2 and 3, it is clear that overall, the number of citations is still going up at an increasing rate, indicating continuing influence of Rao (1948), as far as the citation numbers go.



**Figure 2: Yearly number of citations of Rao (1948): 1948-2023**



**Figure 3: Cumulative number of citations**

Byron (1968) was probably the first to apply the RS test in Econometrics. He used Silvey (1959) Lagrange multiplier (LM) version along with the LR statistic for testing homogeneity and symmetry restrictions in the demand system. In the Econometrics literature, the RS test is known as the LM test - the terminology came from Silvey (1959). Note that the restricted MLE  $\tilde{\theta}$  under the restriction  $H_0 : h(\theta) = c$  can be obtained from the first order condition of the Lagrangian function

$$\mathcal{L} = l(\theta) - \lambda'[h(\theta) - c], \tag{31}$$

where  $\lambda$  is an  $r \times 1$  vector of Lagrange multipliers. The first order conditions are

$$S(\tilde{\theta}) - H(\tilde{\theta})\tilde{\lambda} = 0 \tag{32}$$

$$h(\tilde{\theta}) = c, \tag{33}$$

where  $H(\theta) = dh(\theta)/d\theta$ . Therefore, from (32) we have  $S(\tilde{\theta}) = H(\tilde{\theta})\tilde{\lambda}$ . Given that  $H(\theta)$  has full rank,  $S(\tilde{\theta}) = 0$  is equivalent to  $\tilde{\lambda} = 0$ . These multipliers can be interpreted as the implicit cost (shadow prices) of imposing the restrictions  $h(\theta) = c$ . It can be shown that

$$\tilde{\lambda} = \frac{dl(\tilde{\theta})}{dc}, \tag{34}$$

i.e., the multipliers give the rate of change of the maximum attainable value of the log-likelihood function with respect to the change in constraints. If  $H_0 : h(\theta) = c$  is true and  $l(\tilde{\theta})$  gives the optimal value,  $\tilde{\lambda}$  should be close to zero. Given this “economic” interpretation in terms of Lagrange multipliers, it is not surprising that econometricians prefer the term LM rather than RS. In terms of Lagrange multipliers, (15) can be expressed as

$$RS = LM = \tilde{\lambda}'H(\tilde{\theta})'\mathcal{I}(\tilde{\theta})^{-1}H(\tilde{\theta})\tilde{\lambda}. \tag{35}$$

After Byron (1968), it took another decade for econometricians to realize the potential of the RS test. The earlier notable contributions include Savin (1976), Berndt and Savin (1977), Breusch (1978, 1979) and Godfrey (1978a,b,c). Possibly Breusch and Pagan (1980) had been the most influential. They collected relevant research reported in the Statistics literature, presented the RS test in a general framework in the context of evaluating various econometric models, and discussed many applications. In a full length research monograph, Godfrey (1988) provided a comprehensive account of most of the available RS tests in Econometrics. Bera and Ullah (1991) and Bera and Biliias (2001) demonstrated that many of the commonly used specification tests could be given a score-test interpretation. For the last two-score years the RS tests had been the most common items in econometricians' kit for testing tools. It is not hard to understand the popularity of the score test principle in economics. In most cases, the algebraic forms of W and LR tests can hardly be simplified beyond their original formulae (16) and (17). On the other hand, in the majority of the cases the RS test statistics can explicitly be reduced to neat and elegant explicit formulae enabling its easy incorporation into computer software.

We will not make any attempt to provide a comprehensive list of applications of the RS test in Econometrics, for there are far too many. For instance, consider the workhorse of basic econometric modeling, the linear regression model:

$$y_i = x_i' \beta + \epsilon_i, \tag{36}$$

where  $y_i$  is the  $i$ -th observation on the dependent variable,  $x_i$  is the  $i$ -th observation on  $k$  exogenous variables and  $\epsilon_i \sim IIDN(0, \sigma^2)$ ,  $i = 1, 2, \dots, n$ . The ordinary least squares (OLS) estimation and the related hypotheses tests are based on the four basic assumptions: correct linear functional form; the assumptions of disturbance normality; homoskedasticity; and serial independence. Just to name some of the uses of the RS test principle, test for normality was derived by Bera and Jarque (1981) and Jarque and Bera (1987); Breusch and Pagan (1979) proposed a test for homoskedasticity; and Godfrey (1978a,b) developed tests for serial independence which are very close to the earlier Durbin and Watson (1950) test.

To see the attractiveness of the RS test, let us briefly consider the popular Jarque and Bera (JB) test for normality. Bera and Jarque (1981) started with the Pearson (1895) family of distributions for the disturbance term  $\epsilon_i$  in (36). That means if the pdf of  $\epsilon_i$  is  $f(\epsilon_i)$ , we can write

$$\frac{d \log f(\epsilon_i)}{d \epsilon_i} = \frac{c_1 - \epsilon_i}{\sigma^2 - c_1 \epsilon_i + c_2 \epsilon_i^2}, \quad i = 1, 2, \dots, n, \tag{37}$$

where  $c_1$  and  $c_2$  are constants. The null hypothesis of normality can be stated as  $H_0 : c_1 = c_2 = 0$  in (37). Given the complexity of ML estimation of  $\sigma^2$ ,  $c_1$ , and  $c_2$  in the Pearson family of distributions, W and LR tests are ruled out from a practical point of view. However, the score functions corresponding to  $c_1$  and  $c_2$  in (37), evaluated under the normality assumption, are given respectively by

$$S(\tilde{c}_1) = \frac{n \sqrt{b_1}}{3} \tag{38}$$

and

$$S(\tilde{c}_2) = \frac{n}{4}(b_2 - 3), \tag{39}$$

where  $\sqrt{b_1} = m_3/m_2^{3/2}$  and  $b_2 = m_4/m_2^2$  with  $m_j = \frac{1}{n} \sum_{i=1}^n \tilde{\epsilon}_i^j$ ,  $\tilde{\epsilon}_i = y_i - x_i' \tilde{\beta}$  as OLS residuals,  $j = 2, 3, 4$ . For large  $n$ , under normality

$$E \left[ \sqrt{b_1} \right] = 0, \quad Var \left[ \sqrt{nb_1} \right] = 6, \quad (40)$$

$$E [b_2] = 3, \quad Var \left[ \sqrt{nb_2} \right] = 24, \quad (41)$$

and they are asymptotically normally distributed. Thus, a simple test statistic for normality is given by

$$JB = n \left[ \frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{24} \right], \quad (42)$$

which is asymptotically distributed as  $\chi_2^2$ . It turns out that this test was mentioned by Bowman and Shenton (1975) but was hardly used in practice due to its lack of theoretical underpinnings. The RS test principle uncovered the theoretical justification of (42), ensuing the asymptotic optimality of the test. As it is obvious, JB is based on the two moments, third and fourth. One could have started with these two moments directly without going through the full derivations. From that point of view this RS test has a *moment* test interpretation.

It is quite common to express specification tests in Econometrics as *moment* tests. In a way “any” moment test can be obtained as a RS test under a suitably defined density function. To see this, let us write the  $r$  moment restrictions as

$$E_f[m(y; \theta)] = 0, \quad (43)$$

where  $E_f[\cdot]$  means that (43) is true only when  $f(y; \theta)$  is the correct pdf. A test for the hypothesis  $H_0 : E_f[m(y; \theta)] = 0$  can be based on the estimate of the sample counterpart of  $E_f[m(y; \theta)]$ , namely,

$$\frac{1}{n} \sum_{i=1}^n m(y_i; \theta). \quad (44)$$

Now consider an auxiliary density function

$$f^*(y; \theta, \gamma) = f(y; \theta) \exp[\gamma' m(y; \theta) - \phi(\theta, \gamma)], \quad (45)$$

where  $\phi(\theta, \gamma) = \ln \int \exp[\gamma' m(y; \theta)] f(y; \theta) dy$ , with  $\gamma$  as  $(r \times 1)$  parameter vector.

Note that if  $f(y; \theta)$  is the correct pdf, then  $\gamma = 0$  in (45). The log-likelihood function under the alternative hypothesis is

$$l^*(\theta, \gamma) = \sum_{i=1}^n \ln f^*(y_i; \theta, \gamma). \quad (46)$$

Therefore, the score function for testing  $\gamma = 0$  in (45) is given by

$$\left. \frac{\partial l^*(\theta, \gamma)}{\partial \gamma} \right|_{\gamma=0} = \sum_{i=1}^n m(y_i, \theta), \quad (47)$$

and it provides the identical moment test as in (44). This interpretation of the moment test as a score test was first noted by White (1994). It is easy to see that there are many choices of auxiliary pdf  $f^*(y; \theta, \gamma)$  and the score test will be invariant with respect to these choices, as depicted in Figure 1. The LR and W tests, however, will be sensitive to the forms of  $f^*(y; \theta, \gamma)$ . This ends our coverage of the use of the RS test in Econometrics.

#### 4. Some applications of the RS test in statistics

Rao (1950b) proposed a sequential test of null hypotheses based on the score statistic and his work on locally most powerful (LMP) tests in the case of one-sided alternative hypotheses. His proposal was a reaction to Wald (1945) sequential probability ratio test (SPRT) which was based on the idea of likelihood ratio test for the fixed sample case.

Wald's SPRT statistic was devised to discriminate between different alternative hypotheses for the value of the unknown parameter  $\theta$ . On the other hand, Rao (1950b) seeks to test a null hypothesis  $H_0: \theta = \theta_0$  against a one-sided alternative hypothesis  $H_1: \theta > \theta_0$  with a test statistic that depends only on the null value.

For the fixed sample case, with sample of size  $N$ , the LMP test suggested by Rao and Poti (1946) is defined by [also see equation (6)]

$$P'_N(\theta_0) \geq \mu P_N(\theta_0), \quad (48)$$

where  $P'_N(\theta_0)$  is the first derivative of  $P_N(\theta)$  at  $\theta = \theta_0$ , with  $\mu$  chosen so as to maintain Type-I error at a predetermined level. Motivated by this result, Rao (1950b) proposes a sequential test of the form

$$P'_n(\theta_0) \geq A(N) P_n(\theta_0), \quad (49)$$

with  $n \leq N$ ,  $A(N)$  a properly determined constant depending on the overall level of significance, and  $N$  being the upper limit to the number of observations. According to this sequential testing scheme, the sampling stops with rejection of the null hypothesis, at the smallest value of  $n$  for which the inequality (49) holds true. If by the  $N$ th sampled unit (49) is not realized, the null is not rejected.

Berk (1953) proved that the sequential score tests against a one-sided alternative, where the stopping rule is the first time a certain random walk exceeds a bounded interval, are LMP tests *asymptotically*.

Sequential testing procedures that perform interim analyses during the evolution of the experiment, with the goal of obtaining the result earlier than the termination time suggested by the fixed sample analysis due to time or monetary cost considerations or ethical reasons, are easier validated with the use of score-based test statistics rather than the analogues of LR statistics. We may refer to chapters 9-11 of Sen (1981) for the role of score processes in sequential nonparametrics, where it is mentioned (p.339) "it is comparatively simpler to verify these regularity conditions [i.e., for the score] than those for the likelihood function."

Lombardi (1951), in a thesis on how to select a panel of judges for taste testing and quality evaluation using scientifically sound methods, appears to be one of the first applications of the sequential testing using Rao (1950b) methodology and its comparison with Wald's approach.

To diagnose the potential ability of candidate judges and to decide on the selection of a taste panel, each candidate judge is required to perform a prespecified number of sample comparisons. The number of sample comparisons that should be performed by each candidate judge before reaching to a decision on who to include in the taste panel is always a concern.

Bradley (1953), in conjunction with Lombardi, adapted Rao's method to binomial distribution, and is an early effort to communicate these statistical procedures for the selection of a taste panel to food technologists.

What makes the Rao procedure relatively more appropriate than that of Wald is that a limit to the testing of any one potential judge may be set. In this application,  $N$  denotes the maximum number of tests to be given to any judge. As it is noted by (Bradley, 1953, p.28): "The theory of the procedure needs further investigation since its properties are not well known. However, when it is applied to sequences of triangle tests, apparently satisfactory results are obtained."

In another context, time-series researchers use sequential analysis to determine and test for structural breaks. In a recent application, Bucci (2024) proposes a sequentially computed score statistic to test for the number of regimes in multivariate nonlinear models.

#### 4.1. The role of the score statistic in survival analysis

Another context where the RS test statistic has found fruitful applications for inference is the analysis of survival data. The semiparametric proportional hazards model proposed by Cox (1972) is a standard tool of analysis for time-to-an-event data met in medical, engineering and economic applications. The parameter estimation using the partial likelihood of Cox (1975) initiated an intense research activity for the validation of inference. The majority of the test statistics are special cases of weighted score statistics for different weighting functions and different type of covariates.

The partial likelihood score statistic has a natural martingale characterization. By rewriting the model within the counting process framework, Andersen and Gill (1982) were able to obtain a general asymptotic theory of the score statistic and the associated estimator. In a research related to sequentially computed score test statistic for repeated significance tests, Tsiatis (1981) established the joint asymptotic normality of efficient scores test for the proportional hazards model calculated over time. In a fundamental breakthrough, Sellke and Siegmund (1983) showed that the score process (over time) of the partial likelihood is approximated by a suitable martingale and thus behaves asymptotically like Brownian motion.

Bilias (2000) offered an application of a repeated significance test in a retrospective analysis of the Pennsylvania 'Reemployment Bonus' controlled experiments conducted by the US Department of Labor. Their main purpose was to determine whether the offer of a bonus amount to the unemployment insurance (UI) claimants, provided that they find a job with some required permanence within a given period of time, can act as an incentive for more intensive job-seeking with subsequent reduction of the unemployment spells. The response of primary interest is the *length of insured unemployment spell* and it is assumed that it follows a proportional hazard regression model. The statistic for measuring the effect of the various bonus packages on the duration of insured unemployment relative to the existing scheme is the partial likelihood score statistic. In carrying out sequential analysis, the score statistic is evaluated repeatedly, at different points in chronological time, each time with the available data. The retrospective sequential analysis concluded that the experiment could be concluded earlier than the fixed sample analysis with gains in time and monetary savings.

## 5. Robust RS tests under distributional and parametric misspecifications

As we have narrated in the previous sections the success of the RS test had been phenomenal. However the main problem in these specification tests is that they are developed under the assumption that the underlying probability model is *correctly specified*. When the assumed model is misspecified, it is well known that the RS test loses its local optimal properties.

While discussing the problem in statistical hypothesis testing, Haavelmo (1944, pp.65-66) stated, “Whatever be the principles by which we choose a “best” critical region of size  $\alpha$ , the essential thing is that a test is always developed with respect to a *given fixed* set of possible *alternatives*  $\Omega^0$ .” Haavelmo called  $\Omega^0$ , the *a priori admissible hypotheses* and according to him, a test is not robust if we shift our attention to another admissible set  $\Omega'$  (that may be obtained by extending  $\Omega^0$  to include new/different alternatives), for which the proposed test has poor size and power properties.

Very often it is difficult to interpret the results of a test applied to a *misspecified* model. For instance, while testing the significance of some of the regression coefficients in the *linear* regression models, the results are not easily interpretable when a *nonlinear* model is the appropriate one [see, White (1980), Bera and Byron (1983) and Byron and Bera (1983)]. In the Statistics and Econometrics literature, most emphasis has been put on the minimization of type-I and type-II error probabilities. There are, however, only a few works that seriously consider the consequences and suggest remedies of misspecifying the a priori admissible hypothesis – which can be called the type-III error.

Note that the model under our a priori admissible hypothesis could be misspecified in a variety of ways. Here we consider only two kinds: distributional and parametric. In the former case, the assumed probability density function differs from the true data generating process (DGP). Kent (1982) and White (1982) analyzed this case and suggested a modified version of the RS test that involves adjustment of the variance of the score function. In the parametric misspecification case, the dimension of the assumed parameter space does not match with the true one. Bera and Yoon (1993) developed a modified RS test that is valid under the *local* parametric misspecification.

### 5.1. Robust RS test under distributional misspecification

Let the true DGP be described by the unknown density  $g(y)$  and  $f(y; \theta)$  be our assumed distribution. The RS test statistic given in (15) is not valid when  $g(y)$  and  $f(y; \theta)$  differ. This is because some of the standard results breakdown under distributional misspecification. For instance, consider the information matrix (IM) equality:

$$E_f \left[ \frac{\partial \ln f(y; \theta)}{\partial \theta} \cdot \frac{\partial \ln f(y; \theta)}{\partial \theta'} \right] = E_f \left[ -\frac{\partial^2 \ln f(y; \theta)}{\partial \theta \partial \theta'} \right], \quad (50)$$

where  $E_f[\cdot]$  denotes expectation under  $f(y; \theta)$ . Let us now define

$$J(\theta_g) = nE_g \left[ \frac{\partial \ln f(y; \theta)}{\partial \theta} \cdot \frac{\partial \ln f(y; \theta)}{\partial \theta'} \right] \quad (51)$$

$$K(\theta_g) = nE_g \left[ -\frac{\partial^2 \ln f(y; \theta)}{\partial \theta \partial \theta'} \right], \quad (52)$$

where  $\theta_g$  minimizes the Kullback-Leibler information criterion [see White (1982)]

$$\mathcal{I}_{KL} = E_g \left[ \ln \frac{g(y)}{f(y; \theta)} \right]. \tag{53}$$

One can easily see that  $J(\theta_g) \neq K(\theta_g)$ , in general.

**Example 4:** Suppose we take  $f(y; \theta) \equiv N(\mu, \sigma^2)$ , and let the DGP  $g(y)$  satisfy  $E_g[y] = \mu$ ,  $E_g[y - \mu]^2 = \sigma^2$ ,  $E_g[y - \mu]^3 = \mu_3$  and  $E_g[y - \mu]^4 = \mu_4$ . Then it is easy to show that

$$J(\theta_g) = \begin{bmatrix} \frac{1}{\sigma^2} & \frac{\mu_3}{2\sigma^6} \\ \frac{\mu_3}{2\sigma^6} & \frac{\mu_4}{4\sigma^2} - \frac{1}{4\sigma^4} \end{bmatrix} \tag{54}$$

and

$$K(\theta_g) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{bmatrix}. \tag{55}$$

Hence,  $J(\theta_g) = K(\theta_g)$  if and only if  $\mu_3 = 0$  and  $\mu_4 = 3\sigma^4$ . We can clearly see the connection of these conditions and the JB test for normality given in (42).

Due to this divergence between  $J$  and  $K$ , and noting that we defined the information matrix  $\mathcal{I}(\theta)$  in (10) by taking expectation under  $f(y; \theta)$  instead of under the DGP  $g(y)$ , the standard RS test in (15) is not valid. Let us define an estimator of  $\theta$  by maximizing a likelihood function based on the misspecified density  $f(y; \theta)$  in place of the unknown DGP  $g(y)$ . Such an estimator is called quasi-MLE (QMLE). An early reference to QMLE can be found in Koopmans *et al.* (1950, p.135) [for more on this see Bera *et al.* (2020)]. We will denote QMLE of  $\theta$  (under  $H_0$ ) by  $\tilde{\theta}$ . Kent (1982) and White (1982) suggested the following robust form of the RS test statistic for testing the  $H_0 : h(\theta) = c$ :

$$RS^*(D) = S(\tilde{\theta})' K(\tilde{\theta})^{-1} H(\tilde{\theta}) [H(\tilde{\theta})' B(\tilde{\theta}) H(\tilde{\theta})] H(\tilde{\theta})' K(\tilde{\theta})^{-1} S(\tilde{\theta}), \tag{56}$$

where  $H(\theta) = \partial h(\theta) / \partial \theta$ ,  $B(\theta) = K(\theta)^{-1} J(\theta) K(\theta)^{-1}$  and the notation  $RS^*(D)$  is used to signify *robust* RS test statistic under *distributional* misspecification.

Under  $H_0 : h(\theta) = c$ ,  $RS^*(D)$  is asymptotically distributed as  $\chi_r^2$  even under distributional misspecification, that is, when the assumed density  $f(y; \theta)$  does not coincide with the true DGP  $g(y)$ . This approach of finding the asymptotically correct formula for variance has its origin in Koopmans *et al.* (1950, pp.148-150); for more on this see Bera *et al.* (2021). Expression (56) can be simplified if the parameter vector  $\theta$  ( $p \times 1$ ) can be partitioned as  $\theta = (\gamma', \psi')'$  where  $\gamma$  and  $\psi$  have dimensions  $m$  and  $r$ , respectively,  $m + r = p$ , and we test  $H_0 : \psi = \psi_*$  (say). Let us also partition the score function  $S(\theta)$  and  $J(\theta)$  [similarly  $K(\theta)$ ] as

$$S(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial l(\theta)}{\partial \gamma} \\ \frac{\partial l(\theta)}{\partial \psi} \end{bmatrix} = \begin{bmatrix} S_\gamma(\theta) \\ S_\psi(\theta) \end{bmatrix} \quad (\text{say}) \tag{57}$$

and

$$J(\theta) = \begin{bmatrix} J_\gamma & J_{\gamma\psi} \\ J_{\psi\gamma} & J_\psi \end{bmatrix}. \tag{58}$$



While testing  $H_0 : \psi = \psi_*$ , under this setup  $h(\theta) = \psi - \psi_*$  and  $H(\theta) = [0_{r \times (p-r)}, I_{(r \times r)}]$ , and we can express  $RS^*(D)$  in (56) as [see also Bera *et al.* (2020)]

$$RS_{\psi}^*(D) = S'_{\psi}(\tilde{\theta}) \left[ K_{\psi}(\tilde{\theta}) + J_{\psi\gamma}(\tilde{\theta}) J_{\gamma}^{-1}(\tilde{\theta}) K_{\gamma}(\tilde{\theta}) J_{\gamma}^{-1}(\tilde{\theta}) J_{\gamma\psi}(\tilde{\theta}) - J_{\psi\gamma}(\tilde{\theta}) J_{\gamma}^{-1}(\tilde{\theta}) K_{\gamma\psi}(\tilde{\theta}) - K_{\psi\gamma}(\tilde{\theta}) J_{\gamma}^{-1}(\tilde{\theta}) J_{\gamma\psi}(\tilde{\theta}) \right]^{-1} S_{\psi}(\tilde{\theta}), \quad (59)$$

where  $\tilde{\theta} = (\tilde{\gamma}', \psi_*')'$ , the restricted MLE under  $H_0 : \psi = \psi_*$ .

**Example 5:** In (37) the parameters  $c_1$  and  $c_2$  of the Pearson family of distributions can be treated, respectively, as the “skewness” and “kurtosis” parameters. Suppose we test the symmetry ignoring the (excess) kurtosis. Then we can start with the system (37) with  $c_2 = 0$ , that is,

$$\frac{d \log f(\epsilon_i)}{d\epsilon_i} = \frac{c_1 - \epsilon_i}{\sigma^2 - c_1 \epsilon_i}. \quad (60)$$

After some derivation, it can be shown that the standard RS test for  $c_1 = 0$  is given by

$$RS_{c_1} = n \frac{(\sqrt{b_1})^2}{6}, \quad (61)$$

which is essentially the first part of JB in (42). If  $f(\epsilon_i)$  in (60) is not the true DGP,  $RS_{c_1}$  will not be valid; in particular, the asymptotic variance formula used in (61),  $Var(\sqrt{nb_1}) = 6$  is incorrect [see also equation (40)]. For instance in the presence of excess kurtosis, there will be proportionately more outliers, resulting in higher variance, and thus “6” will underestimate the true variance of  $\sqrt{nb_1}$ . After incorporating the variance correction as in  $RS_{\psi}^*(D)$  in (59) the robust RS test statistic can be written as [for further details see Premaratne and Bera (2017)]:

$$RS_{c_1}^*(D) = n \frac{(\sqrt{b_1})^2}{[9 + m_6 m_2^{-3} - 6 m_4 m_2^{-2}]}, \quad (62)$$

where  $m_j = \frac{1}{n} \sum_{i=1}^n \tilde{\epsilon}_i^j$ ,  $j = 2, 4, 6$ . From (62) we can write the population counterpart of the  $Var(\sqrt{nb_1})$  as

$$Var(\sqrt{nb_1}) = 9 + \mu_6 \mu_2^{-3} - 6 \mu_4 \mu_2^{-2}, \quad (63)$$

where  $\mu_j$  denotes the  $j$ -th population moment of  $\epsilon$ . Therefore, the construction of the robust RS test statistic  $RS_{c_1}^*(D)$  indicates that the true variance of  $\sqrt{nb_1}$  that is valid under excess kurtosis is given by (63). If we impose normality,  $\mu_6 = 15\sigma^6$  and  $\mu_4 = 3\sigma^4$ , then with  $\mu_2 = \sigma^2$ , (63) reduces to  $Var(\sqrt{nb_1}) = 9 + 15 - 6 \times 3 = 6$ , as in (61).

**Example 6:** Consider the test for homoskedasticity under the regression framework of (36), where we now explicitly specify the heteroskedastic structure as  $Var(\epsilon_i) = \sigma_i^2 = \sigma^2 + \delta' z_i$ , where  $\delta$  is a  $r \times 1$  vector and  $z_i$ 's are fixed exogenous variables,  $i = 1, 2, \dots, n$ . Assuming normality of  $\epsilon_i$  the RS statistic for testing homoskedasticity hypothesis  $H_0 : \delta = 0$  is given by [see Breusch and Pagan (1979)]

$$RS_{\delta} = \frac{\nu' Z (Z' Z)^{-1} Z' \nu}{2\tilde{\sigma}^4}, \quad (64)$$

where  $\nu_i = \tilde{\epsilon}_i^2 - \tilde{\sigma}^2$ ,  $\nu = (\nu_1, \nu_2, \dots, \nu_n)'$  and  $Z = (z_1, z_2, \dots, z_n)'$ . The factor “ $2\tilde{\sigma}^4$ ” is the consequence of the normality assumption, and therefore, the test in (64) will not be valid

even asymptotically if  $\epsilon_i$ 's are not distributed as normal. Using (61), the robust form of RS test statistics can be derived as

$$RS_{\delta}^*(D) = \frac{\nu'Z(Z'Z)^{-1}Z'\nu}{\frac{\nu'\nu}{n}}. \tag{65}$$

This is the same modification suggested by Koenker (1981). Note that the modification amounts to replacing  $Var(\epsilon_i^2) = \mu_4 - \mu_2^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4$  (derived under normality) by a robust estimate, namely, by,  $\frac{1}{n} \sum_{i=1}^n (\tilde{\epsilon}_i^2 - \tilde{\sigma}^2)^2 = (\nu'\nu)/n$ . For other applications of  $RS^*(D)$  see for instance, Lucas (1998) and Premaratne and Bera (2017).

In a similar fashion the Wald statistic in (16) can be robustified as [see Kent (1982), White (1982), and Pace and Salvan (1997)]:

$$W^* = [h(\hat{\theta}) - c]'[H(\hat{\theta})B(\hat{\theta})H(\hat{\theta})]^{-1}[h(\hat{\theta}) - c], \tag{66}$$

and asymptotically it has  $\chi_r^2$  distribution under the null hypothesis  $H_0 : h(\theta) = c$ . Thus, robust  $RS^*$  and  $W^*$  are obtained by robustifying the variance expressions, respectively, of  $S(\tilde{\theta})$  and  $h(\hat{\theta})$ . However, similar robustification of LR statistic in (17) is not possible. Kent (1982) showed that under distributional misspecification LR statistic is asymptotically distributed as a weighted sum of  $r$  independent  $\chi_1^2$  variables, and thus no obvious “variance” adjustment is possible.

### 5.2. Robust RS tests under parametric misspecification

Consider a general statistical model represented by the log-likelihood function  $l(\gamma, \psi, \phi)$  where  $\gamma, \psi$ , and  $\phi$  are parameter vectors with dimensions  $(m \times 1)$ ,  $(r \times 1)$  and  $(q \times 1)$ , respectively. Thus our  $(p \times 1)$  parameter vector is  $\theta = (\gamma', \psi', \phi')'$  and  $p = m + r + q$ . Suppose an investigator sets  $\phi = 0$  and tests  $H_0 : \psi = 0$  using the log-likelihood function  $l_1(\gamma, \psi) = l(\gamma, \psi, 0)$ . We will denote the RS statistic for testing  $H_0$  in  $l_1(\gamma, \psi)$  by  $RS_{\psi}$ . Let us also denote  $\tilde{\theta} = (\tilde{\gamma}', 0, 0)'$ , where  $\tilde{\gamma}$  is MLE of  $\gamma$  when  $\psi = 0$  and  $\phi = 0$ . The score vector and the information matrix are defined, respectively, as

$$S(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial l(\theta)}{\partial \gamma} \\ \frac{\partial l(\theta)}{\partial \psi} \\ \frac{\partial l(\theta)}{\partial \phi} \end{bmatrix} = \begin{bmatrix} S_{\gamma}(\theta) \\ S_{\psi}(\theta) \\ S_{\phi}(\theta) \end{bmatrix} \quad (\text{say}) \tag{67}$$

$$\mathcal{I}(\theta) = E_{\theta} \left[ -\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} \right] = \begin{bmatrix} \mathcal{I}_{\gamma} & \mathcal{I}_{\gamma\psi} & \mathcal{I}_{\gamma\phi} \\ \mathcal{I}_{\psi\gamma} & \mathcal{I}_{\psi} & \mathcal{I}_{\psi\phi} \\ \mathcal{I}_{\phi\gamma} & \mathcal{I}_{\phi\psi} & \mathcal{I}_{\phi} \end{bmatrix}. \tag{68}$$

If  $l_1(\gamma, \psi)$  were correctly specified, then the RS test statistic of (15), in the current context can be written as

$$RS_{\psi} = S_{\psi}(\tilde{\theta})'\mathcal{I}_{\psi\cdot\gamma}^{-1}(\tilde{\theta})S_{\psi}(\tilde{\theta}), \tag{69}$$

where  $\mathcal{I}_{\psi\cdot\gamma} = \mathcal{I}_{\psi} - \mathcal{I}_{\psi\gamma}\mathcal{I}_{\gamma}^{-1}\mathcal{I}_{\gamma\psi}$  and it will be asymptotically distributed as *central*  $\chi_r^2$ . Under this set-up, asymptotically  $RS_{\psi}$  will have the correct size and will be locally optimal.

Let us now consider the case of *parametric misspecification*. Suppose the true log-likelihood function is  $l_2 = (\gamma, \phi) = l(\gamma, 0, \phi)$ , so that the alternative  $l_1(\gamma, \psi)$  becomes misspecified. Using the sequence of *local DGP*  $\phi = \delta/\sqrt{n}$ , Davidson and MacKinnon (1987) and Saikkonen (1989) showed that under  $l_2(\gamma, \phi)$  with  $\phi = \delta/\sqrt{n}$ ,  $RS_\psi$  in (69), under  $H_0 : \psi = 0$  will be distributed as *non-central*  $\chi_r^2$  with noncentrality parameter,

$$\lambda(\delta) = \delta' \mathcal{I}_{\phi\psi\cdot\gamma} \mathcal{I}_{\psi\cdot\gamma}^{-1} \mathcal{I}_{\psi\phi\cdot\gamma} \delta, \tag{70}$$

with  $\mathcal{I}'_{\phi\psi\cdot\gamma} = \mathcal{I}_{\psi\phi\cdot\gamma} = \mathcal{I}_{\psi\phi} - \mathcal{I}_{\psi\gamma} \mathcal{I}_{\gamma}^{-1} \mathcal{I}_{\gamma\phi}$ . Owing to the presence of this non-centrality parameter,  $RS_\psi$  will reject the null hypothesis  $H_0 : \psi = 0$  more often than allowed by the preassigned size of the test, even when  $\psi = 0$ . Therefore, under parametric misspecification,  $RS_\psi$  will have an excessive size. For the expression of  $\lambda(\delta)$  in (70), we note that the crucial quantity is  $\mathcal{I}_{\psi\phi\cdot\gamma}$ , which can be interpreted as the conditional covariance between the scores  $S_\psi$  and  $S_\phi$  given  $S_\gamma$ . If  $\mathcal{I}_{\psi\phi\cdot\gamma} = 0$ , then the local presence of the misspecified parameter  $\phi = \delta/\sqrt{n}$  will have no effect on the performance of  $RS_\psi$ .

Using the expression in (70), Bera and Yoon (1993) suggested a modification to  $RS_\psi$  so that the resulting test is robust to the presence of  $\phi$ . The modified statistic is given by

$$\begin{aligned} RS_\psi^*(P) = & [S_\psi(\tilde{\theta}) - \mathcal{I}_{\psi\phi\cdot\gamma}(\tilde{\theta}) \mathcal{I}_{\phi\cdot\gamma}^{-1}(\tilde{\theta}) S_\phi(\tilde{\theta})]' \\ & [\mathcal{I}_{\psi\cdot\gamma}(\tilde{\theta}) - \mathcal{I}_{\psi\phi\cdot\gamma}(\tilde{\theta}) \mathcal{I}_{\phi\cdot\gamma}^{-1}(\tilde{\theta}) \mathcal{I}_{\phi\psi\cdot\gamma}(\tilde{\theta})]^{-1} \\ & [S_\psi(\tilde{\theta}) - \mathcal{I}_{\psi\phi\cdot\gamma}(\tilde{\theta}) \mathcal{I}_{\phi\cdot\gamma}^{-1}(\tilde{\theta}) S_\phi(\tilde{\theta})]. \end{aligned} \tag{71}$$

Here the notation  $RS^*(P)$  is used to signify *robust* RS test statistic under *parametric misspecification*. Under  $H_0 : \psi = 0$ ,  $RS_\psi^*(P)$  is asymptotically distributed as *central*  $\chi_r^2$ , i.e.,  $RS_\psi^*(P)$  has the *same* asymptotic distribution as of  $RS_\psi$  in (69) based on the correct specification. Thus,  $RS_\psi^*(P)$  provides an asymptotically correct-size test under the locally misspecified alternative  $l_2(\gamma, \phi)$ .

$RS_\psi^*(P)$  essentially adjusts the asymptotic mean and variance of standard (unadjusted)  $RS_\psi$ . Another way to look at  $RS_\psi^*(P)$  is to view the quantity,  $\mathcal{I}_{\psi\phi\cdot\gamma}(\tilde{\theta}) \mathcal{I}_{\phi\cdot\gamma}^{-1}(\tilde{\theta}) S_\phi(\tilde{\theta})$  as the prediction of  $S_\psi(\tilde{\theta})$  by  $S_\phi(\tilde{\theta})$ . Here  $S_\phi(\tilde{\theta})$  is the score function of the parameter vector  $\phi$  whose effect we want to take into account in constructing the robust version of the test. Therefore, the net score  $S_\psi^*(\tilde{\theta}) = S_\psi(\tilde{\theta}) - \mathcal{I}_{\psi\phi\cdot\gamma}(\tilde{\theta}) \mathcal{I}_{\phi\cdot\gamma}^{-1}(\tilde{\theta}) S_\phi(\tilde{\theta})$  is the part of  $S_\psi(\tilde{\theta})$  that remains after eliminating the effect of  $S_\phi(\tilde{\theta})$ . In summary,  $S_\psi^*(\tilde{\theta}) \perp S_\phi(\tilde{\theta})$ , though  $S_\phi(\tilde{\theta})$  has “peer” effect on  $S_\psi(\tilde{\theta})$ . Three more things regarding  $RS_\psi^*(P)$  are worth noting. *First*,  $RS_\psi^*(P)$  requires estimation only under the joint null, namely for the constrained model in which both  $\psi = 0$  and  $\phi = 0$ . Given the full specification of the model  $l(\gamma, \psi, \phi)$ , it is of course possible to derive a RS test for  $H_0 : \psi = 0$  in the presence of  $\phi$ . However, that requires the MLE of  $\phi$ , which could be difficult to obtain in some cases. *Second*, when  $\mathcal{I}_{\psi\phi\cdot\gamma} = 0$ ,  $RS_\psi^*(P) = RS_\psi$ . This is a simple condition to check in practice. As mentioned earlier, if this condition is true,  $RS_\psi$  is an asymptotically valid test in the local presence of  $\phi$ . *Finally*, Bera and Yoon (1993) showed that for local misspecification  $RS_\psi^*(P)$  is asymptotically equivalent to Neyman (1959)  $C(\alpha)$  test, and therefore, shares its optimality properties.

**Example 7:** To illustrate the usefulness of the robust score statistic  $RS_\psi^*(P)$ , we now consider the tests developed in Anselin *et al.* (1996) for the mixed regressive - spatial au-

toregressive (SAR) model with a SAR disturbance

$$\begin{aligned}y &= \phi W y + X \gamma + u \\u &= \psi W u + \epsilon \\ \epsilon &\sim N(0, I \sigma^2).\end{aligned}\tag{72}$$

In this model,  $y$  is an  $(n \times 1)$  vector of observations on a dependent variable recorded at each of  $n$  locations,  $X$  is an  $(n \times m)$  matrix of exogenous variables, and  $\gamma$  is a  $(m \times 1)$  vector of parameters,  $\phi$  and  $\psi$  are scalar spatial parameters and  $W$  is an observable spatial weight matrix with positive elements, associated with the spatially lagged dependent variable and SAR disturbance  $u$ . This spatial weight matrix represents “degree of potential interactions” among neighboring locations and are scaled so that the sum of the each row elements of  $W$  is equal to one.

The conventional RS statistic for testing  $H_0 : \psi = 0$  is given by

$$RS_\psi = \frac{[\tilde{u}' W \tilde{u} / \tilde{\sigma}^2]^2}{T},\tag{73}$$

where  $\tilde{u} = y - X \tilde{\gamma}$  are the OLS residuals,  $\tilde{\sigma}^2 = \tilde{u}' \tilde{u} / n$  and  $T = \text{tr}[(W' + W)W]$ . One very interesting observation here is that  $RS_\psi$  is essentially same as the widely used Moran (1948) **I** test. Let us now consider testing  $H_0$  under the local presence of  $\phi$ . First, the crucial quantity to consider is  $\mathcal{I}_{\psi\phi\gamma}$  which is equal to  $T$  and that can never be zero. Therefore robustification of  $RS_\psi$  is needed. Anselin *et al.* (1996) derived the robust test as

$$RS_\psi^*(P) = \frac{[(\tilde{u}' W \tilde{u}) / \tilde{\sigma}^2 - T(\mathcal{I}_{\phi\gamma})^{-1}(\tilde{u}' W y) / \tilde{\sigma}^2]^2}{T[1 - T(\mathcal{I}_{\phi\gamma})^{-1}]},\tag{74}$$

where

$$\mathcal{I}_{\phi\gamma} = \frac{[(W X \tilde{\gamma})' M (W X \tilde{\gamma}) + T \tilde{\sigma}^2]}{\tilde{\sigma}^2},\tag{75}$$

with  $M = I - X(X'X)^{-1}X'$ . A comparison of (73) and (74) clearly reveals that  $RS_\psi^*(P)$  modifies the standard  $RS_\psi$  by correcting the asymptotic mean and variance of the score function  $S_\psi$ .

In a similar way we can find  $RS_\phi$  and  $RS_\phi^*(P)$  which are given, respectively, by

$$RS_\phi = \frac{[(\tilde{u}' W y) / \tilde{\sigma}^2]^2}{\mathcal{I}_{\phi\gamma}}\tag{76}$$

and

$$RS_\phi^*(P) = \frac{[(\tilde{u}' W y) / \tilde{\sigma}^2 - (\tilde{u}' W \tilde{u}) / \tilde{\sigma}^2]^2}{\mathcal{I}_{\phi\gamma} - T},\tag{77}$$

where  $\mathcal{I}_{\phi\gamma} = \mathcal{I}_\phi - \mathcal{I}_{\phi\gamma} \mathcal{I}_\gamma^{-1} \mathcal{I}_{\gamma\phi}$  using the submatrices of the partitioned form of  $\mathcal{I}(\theta)$  given in (68). Anselin (1988) derived a joint RS test for  $H_0 : \psi = \phi = 0$  under the framework of (72) and that takes the following form

$$RS_{\psi\phi} = \frac{[(\tilde{u}' W \tilde{u}) / \tilde{\sigma}^2]^2}{T} + \frac{[(\tilde{u}' W y) / \tilde{\sigma}^2 - (\tilde{u}' W \tilde{u}) / \tilde{\sigma}^2]^2}{\mathcal{I}_{\phi\gamma} - T}.\tag{78}$$

This statistic is asymptotically distributed  $\chi_2^2$ . It is easy to verify that [see Bera *et al.* (2020, Corollary 1)]

$$RS_{\psi\phi} = RS_{\psi} + RS_{\phi}^*(P) = RS_{\phi} + RS_{\psi}^*(P). \quad (79)$$

In other words, the directional RS test for  $\psi$  and  $\phi$  can be decomposed into sum of the unadjusted one-directional test for one type of alternative and the adjusted form for the other alternative. Equalities in (79) can facilitate computations of the adjusted (robust) RS tests after having the unadjusted versions which are easy to obtain and are reported in most of the spatial software.

Anselin and Florax (1995) and Anselin *et al.* (1996) provided simulation results on the finite sample performance of the unadjusted and adjusted RS tests and some related tests. The adjusted tests  $RS_{\psi}^*(P)$  and  $RS_{\phi}^*(P)$  performed remarkably well. Those had very reasonable empirical sizes, remaining within the confidence intervals in all cases. In terms of power they performed exactly the way they were supposed to.

### 5.3. Robust RS tests under *both* the distributional and parametric misspecifications

Now we combine the results of Sections 5.1 and 5.2 and develop robust tests  $RS_{\psi}^*(DP)$  which provides a two-way protection against both types of misspecifications, distributional ( $D$ ) and parametric ( $P$ ). As we have noted in (59),  $RS_{\psi}^*(D)$  involves both the  $J(\theta)$  and  $K(\theta)$  matrices in the variance expression of  $S_{\psi}(\tilde{\theta})$ . While to account of the parametric misspecification, as we did in (71),  $\mathcal{I}_{\psi\phi\cdot\gamma}(\tilde{\theta})\mathcal{I}_{\phi\cdot\gamma}^{-1}(\tilde{\theta})S_{\phi}(\tilde{\theta})$  must be subtracted from  $S_{\psi}(\tilde{\theta})$  to center its mean to zero. The expression for  $RS_{\psi}^*(DP)$  is given by [for details see Bera *et al.* (2020)]:

$$\begin{aligned} RS_{\psi}^*(DP) = & \left[ S_{\psi}(\tilde{\theta}) - J_{\psi\phi\cdot\gamma}(\tilde{\theta})J_{\phi\cdot\gamma}^{-1}(\tilde{\theta})S_{\phi}(\tilde{\theta}) \right]' \\ & \left[ B_{\psi\cdot\gamma}(\tilde{\theta}) + J_{\psi\phi\cdot\gamma}(\tilde{\theta})J_{\phi\cdot\gamma}^{-1}(\tilde{\theta})B_{\phi\cdot\gamma}(\tilde{\theta})J_{\phi\cdot\gamma}^{-1}(\tilde{\theta})J_{\phi\psi\cdot\gamma}(\tilde{\theta}) \right. \\ & \left. - J_{\psi\phi\cdot\gamma}(\tilde{\theta})J_{\phi\cdot\gamma}^{-1}(\tilde{\theta})B_{\phi\psi\cdot\gamma}(\tilde{\theta})B_{\psi\phi\cdot\gamma}(\tilde{\theta})J_{\phi\cdot\gamma}^{-1}(\tilde{\theta})J_{\phi\psi\cdot\gamma}(\tilde{\theta}) \right]^{-1} \\ & \left[ S_{\psi}(\tilde{\theta}) - J_{\psi\phi\cdot\gamma}(\tilde{\theta})J_{\phi\cdot\gamma}^{-1}(\tilde{\theta})S_{\phi}(\tilde{\theta}) \right], \end{aligned} \quad (80)$$

where

$$B_{\psi\cdot\gamma} = K_{\psi} + J_{\psi\gamma}J_{\gamma}^{-1}K_{\gamma}J_{\gamma}^{-1}J_{\gamma\psi} - J_{\psi\gamma}J_{\gamma}^{-1}K_{\gamma\psi} - K_{\psi\gamma}J_{\gamma}^{-1}J_{\gamma\psi}, \quad (81)$$

similarly  $B_{\phi\cdot\gamma}$  and

$$B_{\psi\phi\cdot\gamma} = K_{\psi\phi} - J_{\psi\gamma}J_{\gamma}^{-1}K_{\gamma\phi} - K_{\psi\gamma}J_{\gamma}^{-1}J_{\gamma\phi} + J_{\psi\gamma}J_{\gamma}^{-1}K_{\gamma}J_{\gamma}^{-1}J_{\gamma\phi}, \quad (82)$$

and similarly  $B_{\phi\psi\cdot\gamma}$ . Expressions for the general forms of  $J(\theta)$  and  $K(\theta)$  are given in (51) and (52) and here we are using their partitioned forms for  $\theta = (\gamma', \psi', \phi)'$ . Under  $H_0 : \psi = 0$ , the  $RS_{\psi}^*(DP)$  test statistic will be asymptotically distributed as  $\chi_r^2$  in the presence of both distributional and parametric misspecifications. Although  $RS_{\psi}^*(DP)$  has rather a lengthy expression as in (80), it is actually easy to compute requiring only  $\tilde{\theta} = (\tilde{\gamma}', 0', 0)'$ . It can be

easily seen that under no distributional misspecification, i.e., when  $f(y; \theta) \equiv g(y)$ , resulting in  $K(\tilde{\theta}) = J(\tilde{\theta})$ ,

$$RS_{\psi}^*(DP) = RS_{\psi}^*(P), \quad (83)$$

and similarly under no parametric misspecification, i.e., when  $\delta = 0$  in  $\phi = \delta/\sqrt{n}$ ,

$$RS_{\psi}^*(DP) = RS_{\psi}^*(D). \quad (84)$$

Finally, trivially when  $K = J$  and  $\delta = 0$ ,

$$RS_{\psi}^*(DP) = RS_{\psi} \quad (85)$$

as given in (69).

**Example 8:** Let us briefly go back to Example 7 and now introduce distributional misspecification along with the presence of parametric misspecification. This case has been rigorously considered by Fang *et al.* (2014) and they demonstrated both analytically and through extensive simulations that  $RS_{\psi}^*(P)$  and  $RS_{\phi}^*(P)$  as given, respectively in (74) and (77) are valid under non-normality. Therefore,  $RS_{\psi}^*(DP) = RS_{\psi}^*(P)$  and  $RS_{\phi}^*(DP) = RS_{\phi}^*(P)$ . This is a somewhat unusual situation. For this model as given in (72), information matrix equality does not hold, i.e.,  $J(\theta_g) \neq K(\theta_g)$  [see equations (51)-(53)]. However, still  $J^{-1}KJ^{-1} = J^{-1}$ . This is a serendipitous situation, since no additional adjustment is needed for the distributional misspecification. The intuition behind this serendipity is that the hypotheses  $\psi = 0$  and  $\phi = 0$  relate to the conditional *mean* (first moment) of  $y$  in (72) (conditional on the neighborhood as captured by the  $W$  matrix). However, in general, only tests for variance (second moment) and higher moments get affected by non-normality. A similar case appeared in Bera *et al.* (2020) where they considered testing for random effects and serial correlation within an error component model. Extensive simulation results are also given in Koley and Bera (2022, 2024) demonstrating the robustness of the RS tests under non-normality in finite sample in *spatial regression* model set up.

## 6. Epilogue

We started this survey paper by stating that C.R. Rao's work was always inspired by some practical problems. In his 2003 *Econometric Theory (ET)* Interview [see Bera (2003, p.349)], on the RS test, Rao had the following to say, "The test evolved in a natural way while I was analyzing some genetic data. As I recall, the problem was the estimation of a linkage parameter using data sets from different experiments designed in such a way that each data set had information on the same linkage parameter. It was, however, necessary to test whether such an assumption could be made because of unforeseen factors affecting the experiments. This required a test for consistency of estimates derived from different experimental data sets." Thus we had a new statistical test principle, after LR and W, motivated by a practical problem in *genetics*. However, as we have narrated here, the resulting RS test principle has a far reaching influence even beyond the Statistics, and in particular, it has become one of the most useful model misspecification testing tools in the Econometrics literature. In the history of any scientific field, once in a while there comes a moment for a major breakthrough. Appearance of Rao (1948) was such a historical moment. In fact, after that we have not witnessed any new test principle, beyond the trinity, LR, W and RS.

To keep our exposition simple and to be close to the spirit of Rao (1948), we have stuck to the *likelihood framework*. However, it is easy to extend the RS test and its various ramifications to the generalized method of moments (GMM) and estimating functions (EF) frameworks [for more on these, see for instance Basawa (1991) and Bera *et al.* (2010)]. We have also largely confined ourselves to the asymptotic properties and distributions of the tests. However there is a huge literature on the investigation of the finite sample performance of the RS, particularly, in relation to that of LR and W and finding bootstrap critical values; for example see, Mukerjee (1990, 1993) and Horowitz (1997).

To conclude, we can only speculate what is stored in the future. Given the current vastness of the field we have lost “sharp moments of birth”, like that of Rao (1948). However, considering that 75 years have already been passed, it might be a time for a brand new equally good test principle.

### Acknowledgements

We are profoundly thankful to an anonymous referee for her/his careful reading and offering many pertinent comments which led to improvement of the paper. An earlier version of the paper was presented at the Invited Memorial Session for Professor C. R. Rao, Joint Statistical Meetings (JSM), Portland, Oregon, August 3-8, 2024. We are thankful to the participants of that conference for their comments, especially to the organizer of the Session, Professor Ronald L. Wasserstein, Executive Director of the American Statistical Association (ASA), for giving us the opportunity to present our paper. This paper was also presented at the Department of Statistics, University of Illinois at Urbana-Champaign (UIUC). We would like to thank the attendees of the UIUC seminar for their constructive feedback that further helped in producing an improved version. We are grateful to our research assistant (RA) Anirudh Adhikary. Without his assistance this paper wouldn't have taken off. We are also grateful to our other RAs, Rong Yuwen, Tiancheng Guo, Scarlett He, Yice Zhang and Wenqi Zeng for their diligent work on compiling the citation numbers and a very careful analysis. We thank Professor Osman Doğan for reading an earlier version of the paper with great care and thoroughness, and providing many pertinent comments with detail suggestions for improvements. We wish we could incorporate all his suggestions. Thanks are also due to Dr. Malabika Koley for comments that improved the exposition of the paper. At the publication stage, the Chair Editor, Professor V. K. Gupta paid painstaking attention to the finest details. Indeed, his suggestions, we believe, greatly improved the final presentation. We, however, retain the responsibility of any remaining errors. Part of this work was completed when the second author visited UIUC.

### References

- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, **10**, 1100–1120.
- Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, **20**, 1–17.
- Anselin, L., Bera, A. K., Florax, R., and Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, **26**, 77–104.

- Anselin, L. and Florax, R. (1995). Small sample properties of tests for spatial dependence in regression models: Some further results. In Anselin, L. and Florax, R., editors, *New Directions in Spatial Econometrics*, pages 21–74. Springer-Verlag, Berlin.
- Basawa, I. V. (1991). Generalized score tests for composite hypotheses. In Godambe, V. P., editor, *Estimating Functions*, pages 121–131. Clarendon Press, Oxford.
- Bera, A. K. (1983). Aspects of econometric modeling. Unpublished Ph.D. Dissertation, The Australian National University, Canberra.
- Bera, A. K. (2000). Hypothesis testing in the 20th century with a special reference to testing with misspecified models. In Rao, C. and Szekely, G., editors, *Statistics for the 21st Century*, pages 33–92. Marcel Dekker, New York.
- Bera, A. K. (2003). The ET interview: Professor C.R. Rao: Interviewed by Anil K. Bera. *Econometric Theory*, **19**, 331–400.
- Bera, A. K. and Biliyas, Y. (2001). Rao's score, Neyman's  $C(\alpha)$  and Silvey's LM tests: An essay on historical developments and some new results. *Journal of Statistical Planning and Inference*, **97**, 9–44.
- Bera, A. K., Biliyas, Y., Yoon, M. J., Doğan, O., and Taşpınar, S. (2020). Adjustments of Rao's score test for distributional and local parametric misspecifications. *Journal of Econometric Methods*, **9**. doi:10.1515/jem-2017-0022.
- Bera, A. K. and Byron, R. P. (1983). A note on the effects of linear approximation on hypothesis testing. *Economics Letters*, **12**, 251–254.
- Bera, A. K., Doğan, O., and Taşpınar, S. (2021). Asymptotic variance of test statistics in the ML and QML frameworks. *Journal of Statistical Theory and Practice*, **15**. doi:10.1007/s42519-020-00137-0.
- Bera, A. K. and Jarque, C. M. (1981). An efficient large-sample test for normality of observations and regression residuals. Working Paper in Economics and Econometrics, No. 40, The Australian National University, Canberra.
- Bera, A. K., Montes-Rojas, G., and Sosa-Escudero, W. (2010). General specification testing with locally misspecified models. *Econometric Theory*, **26**, 1838–1845.
- Bera, A. K. and Premaratne, G. (2001). General hypothesis testing. In Baltagi, B. H., editor, *A Companion to Theoretical Econometrics*, pages 38–61. Blackwell Publishing Ltd.
- Bera, A. K. and Ullah, A. (1991). Rao's score test in econometrics. *Journal of Quantitative Economics*, **7**, 189–220.
- Bera, A. K. and Yoon, M. J. (1993). Specification testing with locally misspecified alternatives. *Econometric Theory*, **9**, 649–658.
- Berk, R. H. (1953). Locally most powerful sequential tests. *The Annals of Statistics*, **3**, 373–381.
- Berndt, E. R. and Savin, N. E. (1977). Conflict among criteria for testing hypotheses in the multivariate linear regression model. *Econometrica*, **45**, 1263–1277.
- Biliyas, Y. (2000). Sequential testing of duration data: The case of the pennsylvania 'reemployment bonus' experiment. *Journal of Applied Econometrics*, **15**, 575–594.
- Bowman, K. O. and Shenton, L. R. (1975). Omnibus test contours for departures from normality based on  $\sqrt{b_1}$  and  $b_2$ . *Biometrika*, **62**, 243–250.
- Bradley, R. A. (1953). Some statistical methods in taste testing and quality evaluation. *Biometrics*, **9**, 22–38.



- Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, **17**, 334–355.
- Breusch, T. S. (1979). Conflict among criteria for testing hypotheses: Extensions and comments. *Econometrica*, **47**, 203–207.
- Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, **47**, 1287–1294.
- Breusch, T. S. and Pagan, A. R. (1980). The Lagrange Multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, **47**.
- Bucci, A. (2024). A sequential test procedure for the choice of the number of regimes in multivariate nonlinear models. arXiv:2406.02152v1.
- 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.
- Byron, R. P. and Bera, A. K. (1983). Least squares approximation to unknown regression functions: A comment. *International Economic Review*, **24**, 255–260.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62**, 269–276.
- Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall/CRC.
- Davidson, R. and MacKinnon, J. G. (1987). Implicit alternatives and the local power of test statistics. *Econometrica*, **55**, 1305–1329.
- Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, **37**, 409–428.
- Fang, Y., Park, S. Y., and Zhang, J. (2014). A simple spatial dependence test robust to local and distributional misspecifications. *Economics Letters*, **124**, 203–206.
- Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. New York and London: Academic Press, Inc.
- Godfrey, L. G. (1978a). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, **46**, 1293–1301.
- Godfrey, L. G. (1978b). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1310.
- Godfrey, L. G. (1978c). Testing for multiplicative heteroskedasticity. *Journal of Econometrics*, **8**, 227–236.
- Godfrey, L. G. (1988). *Misspecification Tests in Econometrics*. Cambridge University Press, Cambridge.
- Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, **12**, iii–115.
- Horowitz, J. L. (1997). Bootstrap methods in econometrics: theory and numerical performance. In Kreps, D. M. and Wallis, K. F., editors, *Advances in Economics and Econometrics: Theory and Applications*, volume 3, pages 188–222. Cambridge University Press.
- Jarque, C. M. and Bera, A. K. (1987). Test for normality of observations and regression residuals. *International Statistical Review*, **55**, 163–172.
- Kent, J. T. (1982). Robust properties of likelihood ratio test. *Biometrika*, **69**, 19–27.

- Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*, **17**, 107–112.
- Koley, M. and Bera, A. K. (2022). Testing for spatial dependence in a spatial autoregressive (SAR) model in the presence of endogenous regressors. *Journal of Spatial Econometrics*, **3**. doi:10.1007/s43071-022-00026-7.
- Koley, M. and Bera, A. K. (2024). To use, or not to use the spatial Durbin model?—that is the question. *Spatial Economic Analysis*, **19**, 30–56.
- Koopmans, T. C., Rubin, H., and Leipnik, R. B. (1950). Measuring the equation systems of dynamic economics. In Koopmans, T. C., editor, *Statistical Inference in Dynamic Economic Models*, Cowles Commission for Research in Economics, Monograph No. 10, pages 53–237. John Wiley and Sons, Inc.
- Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer-Verlag, New York, Inc.
- Lombardi, G. J. (1951). The sequential selection of judges for organoleptic testing. Unpublished thesis, Virginia Polytechnic Institute.
- Lucas, A. (1998). Inference on cointegrating ranks using LR and LM tests based on pseudo-likelihoods. *Econometric Reviews*, **17**, 185–214.
- Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B (Methodological)*, **10**, 243–251.
- Mukerjee, R. (1990). Comparison of tests in the multiparameter case i: Second order power. *Journal of Multivariate Analysis*, **33**, 17–30.
- Mukerjee, R. (1993). Rao's score test: Recent asymptotic results. In G. S. Maddala, C. R. R. and Vinod, H., editors, *Handbook of Statistics 11*, pages 363–379. North-Holland Science Publishers, Amsterdam.
- Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypothesis. In Grenander, U., editor, *Probability and Statistics*, pages 213–234. John Wiley, New York.
- Neyman, J. (1980). Some memorable incidents in probabilistic/statistical studies. In Chakravarti, I. M., editor, *Asymptotic Theory of Statistical Tests and Estimation*, pages 1–32. Academic Press, New York.
- Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20**, 175–240.
- Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A*, **231**, 694–706.
- Neyman, J. and Pearson, E. S. (1936). Contributions to the theory of testing statistical hypotheses. *Statistical Research Memoirs*, **1**, 1–37.
- Pace, L. and Salvani, A. (1997). *Principles of Statistical Inference: From a Neo-Fisherian Perspective*. World Scientific, River Edge, N.J.
- Pearson, K. (1895). Contribution to the mathematical theory of evolution—ii. skewed variation in homogeneous material. *Philosophical Transactions of the Royal Society A*, **186**, 343–414.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **50**, 157–175.

- Premaratne, G. and Bera, A. K. (2017). Adjusting the tests for skewness and kurtosis for distributional misspecifications. *Communications in Statistics - Simulation and Computation*, **46**, 3599–3613.
- Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, **44**, 50–57.
- Rao, C. R. (1950a). Methods of scoring linkage data giving the simultaneous segregation of three factors. *Heredity*, **4**, 37–59.
- Rao, C. R. (1950b). Sequential tests of null hypotheses. *Sankhyā*, **10**, 361–370.
- Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. John Wiley & Sons, Inc.
- Rao, C. R. (2001). Two score and 10 years of score tests. *Journal of Statistical Planning and Inference*, **97**, 3–7.
- Rao, C. R. (2005). Score test: Historical review and recent developments. In Balakrishnan, N., Nagaraja, H. N., and Kannan, N., editors, *Advances in Ranking and Selection, Multiple Comparisons, and Reliability: Methodology and Applications*, pages 3–20. Birkhäuser Boston, Boston, MA.
- Rao, C. R. and Poti, S. J. (1946). On locally most powerful tests when alternative are one sided. *Sankhyā*, **7**, 439.
- Roy, S. N. (1953). On a Heuristic Method of Test Construction and its use in Multivariate Analysis. *The Annals of Mathematical Statistics*, **24**, 220 – 238.
- Saikkonen, P. (1989). Asymptotic relative efficiency of the classical test statistics under misspecification. *Journal of Econometrics*, **42**, 351–369.
- Savin, N. E. (1976). Conflict among testing procedures in a linear regression model with autoregressive disturbances. *Econometrica*, **44**, 1303–1315.
- Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika*, **70**, 315–326.
- Sen, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*. Wiley, New York.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- Silvey, S. D. (1959). The Lagrangian Multiplier test. *The Annals of Mathematical Statistics*, **30**, 389–407.
- Tsiatis, A. A. (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*, **68**, 311–315.
- Wald, A. (1943). Tests of statistical hypothesis concerning several parameters when the number of observation is large. *Transactions of the American Mathematical Society*, **54**, 426–482.
- Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, **16**, 117–186.
- White, H. (1980). Using least squares to approximate unknown regression functions. *International Economic Review*, **21**, 149–170.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge.