This is an open access article distributed under the Creative Commons Attribution License

Hindawi Publishing Corporation
Advances in Decision Sciences
Volume 2010, Article ID 292013, 8 pages
doi:10.1155/2010/292013

Review Article

Generalised Score and Wald Tests

Paul Rippon and J. C. W. Rayner

School of Mathematical and Physical Sciences, University of Newcastle, Newcastle, NSW 2308, Australia

Correspondence should be addressed to J. C. W. Rayner, john.rayner@newcastle.edu.au

Received 17 September 2009; Accepted 16 December 2009

Academic Editor: Chin Lai

Copyright © 2010 P. Rippon and J. C. W. Rayner. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The generalised score and Wald tests are described and related to their nongeneralised versions. Two interesting applications are discussed. In the first a new test for the Behrens-Fisher problem is derived. The second is testing homogeneity of variances from multiple univariate normal populations.

1. Introduction

This paper is intended to be a tutorial for those wishing to inform themselves about the generalised score and Wald Tests. It extends the content of [1] and has similar objectives; that is, it focuses on the use of these tests rather than their properties. It is intended to be very accessible. Readers need only some prior knowledge of partitioned matrices, score and Wald tests, see, for example, [1] and [2, Chapter 3].

The score test is particularly valuable when maximum likelihood (ML) estimation under the full model is not preferred, but ML estimation under the null model is. The converse holds for the Wald test. Thus when ML estimation under one of the null and full models is not preferred, the likelihood ratio test is problematic, but one of the score and Wald tests is not. Here by not preferred we mean that, for example, estimates may be calculated by some iterative scheme with dubious convergence. Other possibilities are that estimates may have a particularly convoluted expression or the finite sample properties (such as large bias) may be inappropriate for the problem of interest.

When ML estimation under both the null and full models is not preferred, we need another way forward. This is provided by the generalised score and Wald Tests. These tests are especially valuable when the model may be misspecified, but that will not be the focus here.

In Section 2 the generalised score and Wald Tests are described. In Section 3 this material is applied to deriving a new test for the Behrens-Fisher problem, while Section 4 looks at testing equality of variances from multiple independent normal samples.

2. M-Estimators and Generalised Score Tests

The class of M-estimators includes both ML and method of moments estimators. An M-estimator $\tilde{\gamma}$ satisfies

$$\sum_{j=1}^{n} \Psi(X_j, \tilde{\gamma}) = 0_p, \tag{2.1}$$

in which $X_1, \ldots, X_n$ are independent but not necessarily identically distributed, Ψ is a known $p \times 1$ function not depending on j or n, γ is a p-dimensional parameter, and in general 0_m denotes an $m \times 1$ vector of zeros. The estimating function Ψ must be sufficiently 'smooth.' In particular, its derivatives up to second order, and their expectations, must exist. Hence the matrices A and B defined subsequently are assumed to exist. Also, the expectation of the second-order derivatives must be bounded in probability. More technical details on M-estimators may be found in [3, Chapter 5].

In our setting we assume that $\gamma = (\theta^T, \beta^T)^T$ and that we wish to test H_0: $\theta = 0_k$ against the alternative K: $\theta \neq 0_k$ with θ being the $k \times 1$ vector of primary interest, with β a $q \times 1$ vector of nuisance parameters, and with $p = k + q$. The generalised score test is based on the partial M-estimator that satisfies

$$\sum_{j=1}^{n} \Psi_\beta(X_j, \tilde{\gamma}_0) = 0_q, \tag{2.2}$$

where Ψ is partitioned similarly to γ, so that $\Psi^T = (\Psi_\theta^T, \Psi_\beta^T)$, and where $\tilde{\gamma}_0 = (0_k^T, \tilde{\beta}_0^T)^T$ in which $\tilde{\beta}_0$ is the M-estimator of β under the null hypothesis. Define

$$\begin{aligned} U_\theta(\gamma) &= \sum_{j=1}^{n} \Psi_\theta(X_j, \gamma), \\ A(\gamma) &= -E_0\left[\frac{\partial \Psi(X, \gamma)}{\partial \gamma}\right] = \begin{pmatrix} A_{\theta\theta} & A_{\theta\beta} \\ A_{\beta\theta} & A_{\beta\beta} \end{pmatrix}, \\ B(\gamma) &= E_0\left[\Psi\Psi^T\right] = \begin{pmatrix} B_{\theta\theta} & B_{\theta\beta} \\ B_{\beta\theta} & B_{\beta\beta} \end{pmatrix}, \end{aligned} \tag{2.3}$$

in which E_0 denotes expectation under the null hypothesis. Here $A(\gamma)$ and $B(\gamma)$ are $p \times p$ and $A_{\theta\theta}$ and $B_{\theta\theta}$ are $k \times k$. We note that $A(\gamma)$ is not necessarily symmetric while $B(\gamma)$ is. This means that the form of the generalised tests given by, for example, [4], needs to be slightly modified. The generalised score test statistic is given by

$$S_G = U_\theta^T \left(\left(A^T\right)^{-1}\right)_{\theta\theta} \left(A^{-1} B \left(A^T\right)^{-1}\right)_{\theta\theta}^{-1} \left(A^{-1}\right)_{\theta\theta} U_\theta \tag{2.4}$$

in which, as can readily be shown, $(A^{-1})_{\theta\theta} = (A_{\theta\theta} - A_{\theta\beta}A_{\beta\beta}^{-1}A_{\beta\theta})^{-1}$ and the arguments in U_θ, A, and B are suppressed; here all are $\tilde{\gamma}_0$. Similarly the generalised Wald test statistic is given by

$$W_G = \tilde{\theta}^T \left(A^{-1} B \left(A^T \right)^{-1} \right)_{\theta\theta}^{-1} \tilde{\theta} \tag{2.5}$$

in which all arguments are $\tilde{\gamma}$. In the exposition in [4] parameters are estimated by ML but the data do not come from the parametric model: this is ML under misspecification. In [5], Kent's definitions are given but in place of ML estimators any M-estimators are permitted. It is also noted in [4] that A and B can in practice be replaced by any consistent estimates.

An alternative form of S_G that is more convenient for calculation is given in [2], where it is applied to the construction of generalized smooth tests of goodness of fit. This form gives

$$S_G = U_\theta^T(\tilde{\gamma}_0)\Sigma_{GS}^{-1}(\tilde{\gamma}_0)U_\theta(\tilde{\gamma}_0) \tag{2.6}$$

in which

$$\Sigma_{GS}(\gamma) = B_{\theta\theta} - A_{\theta\beta}A_{\beta\beta}^{-1}B_{\beta\theta} - B_{\theta\beta}\left(A_{\beta\beta}^{-1}\right)^T A_{\theta\beta}^T + A_{\theta\beta}A_{\beta\beta}^{-1}B_{\beta\beta}\left(A_{\beta\beta}^{-1}\right)^T A_{\theta\beta}^T. \tag{2.7}$$

The equivalence of the two forms requires routine but tedious matrix algebra and is omitted here. The asymptotic distribution of both S_G and W_G under H_0 is χ_k^2.

If $\Psi(X, \gamma)$ is the derivative of logarithm of the likelihood, which is the usual score function, then $A = B$ is the usual (symmetric) information matrix, and $A^{-1}B(A^T)^{-1} = A^{-1}$. If ML estimation is used, then $W_G = \widehat{\theta}^T\{A_{\theta\theta} - A_{\theta\beta}A_{\beta\beta}^{-1}A_{\beta\theta}\}\widehat{\theta}$, the usual Wald test statistic, and $S_G = U_\theta^T\{A_{\theta\theta} - A_{\theta\beta}A_{\beta\beta}^{-1}A_{\beta\theta}\}^{-1}U_\theta$, the usual score test statistic. Both are given in this form in [1]. For more information see [5, 6].

In [5, page 328] replacing the inverse of the asymptotic covariance matrix $\Sigma_{GS}(\tilde{\gamma}_0)$ in S_G by a generalised inverse of a consistent estimate of $\Sigma_{GS}(\tilde{\gamma}_0)$ is recommended. Although it may sound trivial, when calculating any of the ordinary or generalised score or Wald test statistics, we are finding $(X - E[X])^T\Sigma^{-1}(X - E[X])$ where X is at least asymptotically multivariate normal and Σ is at least asymptotically the full rank covariance matrix of X. Very occasionally it may be more convenient to find the exact covariance matrix rather than one that is asymptotically equivalent. If so the exact covariance matrix can be used in the above expressions; similarly when appropriate a generalised inverse of the exact or an asymptotically equivalent covariance matrix can be used.

3. The Behrens-Fisher Problem

In the Behrens-Fisher problem, $Y_1, \ldots, Y_m$ is a random sample from an $N(\mu_Y, \sigma_Y^2)$ population, and $Z_1, \ldots, Z_n$ is an independent random sample from an $N(\mu_Z, \sigma_Z^2)$ population. It is desired to test H: $\mu_Y = \mu_Z$ against K: $\mu_Y \neq \mu_Z$, with the standard deviations σ_Y and σ_Z being nuisance parameters. In [2, Example 3.3.2] the likelihood ratio, score, and Wald tests are derived. The score test requires the solution of an inconvenient cubic equation; so this is one situation in

which the Wald statistic looks distinctly more appealing than both the likelihood ratio and score test statistics.

When the estimating function $\sum_{j=1}^{n}\Psi(X_j,\gamma)$ is the usual score function, the generalised score test is the usual score test. To conform to our notation put $(Y_1,\ldots,Y_m,Z_1,\ldots,Z_n)=X^T$, $\mu_Y-\mu_Z=2\theta$, $\mu_Y+\mu_Z=2\beta_1$, $\sigma_Y^2=\beta_2$ and $\sigma_Z^2=\beta_2$. We test $H:\theta=0$ against $K:\theta\neq 0$, with nuisance parameters β_1, β_2 and β_3. The logarithm of the likelihood is

$$\text{constant}-\left(\frac{m}{2}\right)\log\beta_2-\left(\frac{n}{2}\right)\log\beta_3-(2\beta_2)^{-1}\sum_i\left(y_i-\beta_1-\theta\right)^2-(2\beta_3)^{-1}\sum_j\left(z_j-\beta_1+\theta\right)^2, \tag{3.1}$$

and therefore the score function has the following components:

$$\begin{aligned}
S_\theta(\gamma)&=\sum_i\frac{(y_i-\beta_1-\theta)}{\beta_2}-\sum_j\frac{(z_j-\beta_1+\theta)}{\beta_3},\\
S_{\beta_1}(\gamma)&=\sum_i\frac{(y_i-\beta_1-\theta)}{\beta_2}+\sum_j\frac{(z_j-\beta_1+\theta)}{\beta_3},\\
S_{\beta_2}(\gamma)&=-\frac{m}{(2\beta_2)}+\sum_i\frac{(y_i-\beta_1-\theta)^2}{(2\beta_2^2)},\\
S_{\beta_3}(\gamma)&=-\frac{n}{(2\beta_3)}+\sum_j\frac{(z_j-\beta_1+\theta)^2}{(2\beta_3^2)}.
\end{aligned} \tag{3.2}$$

These are the partial derivatives of the logarithm of the likelihood. Under the null hypothesis the estimating equations are $S_{\beta_1}(\gamma_0)=S_{\beta_2}(\gamma_0)=S_{\beta_3}(\gamma_0)=0$. This leads to the inconvenient cubic equation mentioned previously. If we proceed with this model, the cubic must be solved to find $\tilde{\beta}_{10}$, and hence $\tilde{\beta}_{20}=\sum_i\left(Y_i-\tilde{\beta}_{10}\right)^2/m$ and $\tilde{\beta}_{30}=\sum_j\left(Z_j-\tilde{\beta}_{10}\right)^2/n$. We also find

$$\begin{aligned}
S_\theta(\tilde{\gamma}_0)&=\frac{2\left(\overline{Y}-\overline{Z}\right)}{\tilde{\beta}_{20}/m+\tilde{\beta}_{30}/n},\\
A&=\begin{pmatrix}
\frac{m}{\beta_2}+\frac{n}{\beta_3} & \frac{m}{\beta_2}-\frac{n}{\beta_3} & 0 & 0\\
\frac{m}{\beta_2}-\frac{n}{\beta_3} & \frac{m}{\beta_2}+\frac{n}{\beta_3} & 0 & 0\\
0 & 0 & \frac{m}{(2\beta_2^2)} & 0\\
0 & 0 & 0 & \frac{n}{(2\beta_3^2)}
\end{pmatrix}=B,
\end{aligned} \tag{3.3}$$

whence

$$\Sigma_{GS} = \frac{4}{\{\beta^2/m + \beta_3/n\}}, \tag{3.4}$$

and the generalised score test statistic is $S_G = (\overline{Y} - \overline{Z})^2/(\tilde{\beta}_{20}/m + \tilde{\beta}_{30}/n)$. This is just the ordinary score test statistic.

While solving the cubic is not a great difficulty, if we modify $S_{\beta_1}(\gamma)$ so that it becomes

$$S_{\beta_1}(\gamma) = \sum_i (y_i - \beta_1 - \theta) + \sum_j (z_j - \beta_1 + \theta), \tag{3.5}$$

a possibly less efficient but certainly more convenient estimator of the common mean under the null hypothesis may be found. This estimator is the solution to $S_{\beta_1}(\gamma_0) = 0$, namely, $\tilde{\beta}^*_{10} = (m\overline{Y} + n\overline{Z})/(m+n)$. If we also modify $S_\theta(\gamma)$ so that

$$S_\theta(\gamma) = \sum_i (y_i - \beta_1 - \theta) - \sum_j (z_j - \beta_1 + \theta), \tag{3.6}$$

while leaving the other two equations unchanged, the generalised score test is based on

$$S_\theta(\tilde{\gamma}_0) = \frac{2mn}{m+n}\left(\overline{Y} - \overline{Z}\right). \tag{3.7}$$

The estimators of β_2 and β_3 are slightly different from those found previously, being $\tilde{\beta}^*_{20} = \sum_i (Y_i - \tilde{\beta}^*_{10})^2/m$ and $\tilde{\beta}^*_{30} = \sum_j (Z_j - \tilde{\beta}^*_{10})^2/n$. Modifying the previous derivation gives

$$A = \begin{pmatrix} m+n & m-n & 0 & 0 \\ m-n & m+n & 0 & 0 \\ 0 & 0 & \dfrac{m}{(2\beta_2^2)} & 0 \\ 0 & 0 & 0 & \dfrac{n}{(2\beta_3^2)} \end{pmatrix},$$

$$B = \begin{pmatrix} m\beta_2 + n\beta_3 & m\beta_2 - n\beta_3 & 0 & 0 \\ m\beta_2 - n\beta_3 & m\beta_2 + n\beta_3 & 0 & 0 \\ 0 & 0 & \dfrac{m}{(2\beta_2^2)} & 0 \\ 0 & 0 & 0 & \dfrac{n}{(2\beta_3^2)} \end{pmatrix}, \tag{3.8}$$

whence

$$\Sigma_{GS} = \frac{4mn\left(m\beta_3 + n\beta_2\right)}{\left(m + n\right)^2}, \tag{3.9}$$

and the generalised score test statistic is

$$S_G = \frac{\left(\overline{Y} - \overline{Z}\right)^2}{\left(\widetilde{\beta}^*_{20}/m + \widetilde{\beta}^*_{30}/n\right)}. \tag{3.10}$$

It may be shown that the Wald test statistic is a one-one function of this S_G, so that these two tests are equivalent. However, if using the asymptotic χ^2_1 critical values, the generalised score test has actual test sizes much closer to the nominal sizes than the Wald test. When using simulated critical values that are virtually exact, the generalised score test power is within 1% of the entrenched test due to Welch [7]. So on this criterion the Welch and generalised score tests are virtually indistinguishable.

The Welch test is very similar to the Wald test. Using Satterthwaite's approximation to the null distribution of the Welch test gives excellent agreement between the nominal and actual test sizes. However Satterthwaite's approximation does not work nearly as well for S_G. Hence, in terms of agreement between nominal and actual test sizes using approximations and asymptotic critical values, the Welch test is to be preferred. Support for these assertions and more numerical details are available in [8].

4. Testing Equality of Variances

Suppose that we have m independent random samples, with the jth, $j = 1, \ldots, m$, being of size n_j and from a normal $N(\mu_j, \sigma_j^2)$ population. The total sample size is $n = n_1 + \cdots + n_m$. We seek to test equality of variances: H: $\sigma_1^2 = \cdots = \sigma_m^2 = \sigma^2$ say against the alternative K: not H. Popular tests include the likelihood ratio test, frequently referred to as Bartlett's test, and Levene's test. The former is known to be nonrobust, while the latter is more robust in that its actual levels are closer to the nominal levels. Levene's test is less powerful than Bartlett's when the data are consistent with normality.

We now construct a Wald test of H against K. We could use the generalised Wald test construction with $\Psi(X, \gamma)$ being the derivative of logarithm of the likelihood, but we leave that as an exercise for the interested reader. We could also calculate one of the forms of the asymptotic covariance matrix, but this is a case where it is simpler to calculate the exact covariance matrix. Moreover the exact covariance matrix involves an inconvenient inverse; so we instead use the Moore-Penrose inverse. This is defined in the appendix, along with some relevant useful results. This approach leads to a simpler test statistic.

Throughout this example, since we are calculating the Wald test statistic, all estimation is ML. As a consequence estimators are denoted by hats (^) instead of tildes (~). We also use unbiased versions of the sample variances (with divisors $n - 1$ instead of n). These are asymptotically equivalent to the usual ML estimators, and the corresponding test statistic is asymptotically equivalent to the usual Wald test statistic.

Before proceeding with the construction, note that if S^2 is the unbiased sample variance from a random sample of size n from a $N(\mu, \sigma^2)$ distribution, then $(n-1)\ S^2/\sigma^2$ has the χ^2_{n-1} distribution. As is well known, $\text{var}\{(n-1)\ S^2/\sigma^2\} = 2(n-1)$, so that

$$\text{var}\left(S^2\right) = \frac{2\sigma^4}{(n-1)} \quad \text{and} \quad E\left[S^4\right] = \frac{(n+1)\sigma^4}{(n-1)}. \tag{4.1}$$

From the Rao-Blackwell theorem $(n-1)S^4/(n+1)$ is an optimal estimator of σ^4, being the unique estimator with minimum variance in the class of unbiased estimators of σ^4. This optimality is conferred upon $2S^4/(n+1)$ when estimating $\text{var}(S^2)$. Writing S_j^2 for the unbiased estimator of the jth population variance σ_j^2, $j = 1, \ldots, m$, the optimal estimator of $\text{var}(S_j^2) = 2\sigma_j^4/(n_j - 1)$ is $d_j = 2S_j^4/(n_j+1)$ for $j = 1, \ldots, m$.

Should the null hypothesis be true, an unbiased estimator of the common population variance σ^2 is the pooled sample variance $S^2 = \sum_j w_j S_j^2$ where $w_j = (n_j - 1)/(n-m)$ for $j = 1, \ldots, m$. Note that since $\sum_j (n_j - 1) = n - m$, $\sum_j w_j = 1$. Now define

$$\sigma^2 = \sum_j w_j \sigma_j^2, \qquad \phi = \left(\sigma_j^2 \sqrt{w_j}\right), \qquad u = \left(\sqrt{w_j}\right), \qquad C = I_m - uu^T. \tag{4.2}$$

Then $\theta = C\phi = ((\sigma_j^2 - \sigma^2)\sqrt{w_j})$. This is zero if and only if $\sigma_j^2 = \sigma^2$ for all j. Hence testing equality of variances is equivalent to testing $H\colon \theta = 0_m$ against $K\colon \theta \neq 0_m$. An unbiased estimator of θ is $\widehat{\theta} = ((S_j^2 - S^2)\sqrt{w_j})$ and since C is symmetric, $\widehat{\theta} = C\widehat{\phi}$ has covariance matrix estimated by $\widehat{\text{cov}}(\widehat{\theta}) = \text{CDC}$ where now $D = \text{diag}(d_j w_j)$. Now CDC is not of full rank, and in order to use results on quadratic forms of multivariate normal random variables generalised or pseudoinverses are required. Here we use M^+, the Moore-Penrose inverse of the matrix M. See the appendix.

Because C is idempotent, the Moore-Penrose inverse of CDC is given by

$$(\text{CDC})^+ = \text{C}^+\text{D}^+\text{C}^+ = \text{CD}^{-1}\text{C}. \tag{4.3}$$

A Wald test statistic for testing $H\colon \theta = 0_m$ against $K\colon \theta \neq 0_m$ is

$$\begin{aligned} \widehat{\theta}^T \left(\widehat{\text{cov}}(\widehat{\theta})\right)^+ \widehat{\theta} &= \left(\text{C}\widehat{\phi}\right)^T (\text{CDC})^+ \left(\text{C}\widehat{\phi}\right) = \widehat{\phi}^T \text{CCD}^{-1}\text{CC}\widehat{\phi} \\ &= \widehat{\phi}^T \text{CD}^{-1}\text{C}\widehat{\phi} = \widehat{\theta}^T \text{D}^{-1}\widehat{\theta} = \sum_{j=1}^{m} \frac{\left(S_j^2 - S^2\right)^2}{d_j} = T_{\text{MP}} \text{ say.} \end{aligned} \tag{4.4}$$

Since $\text{rank}(\text{CDC}) = m - 1$, T_{MP} should be compared with the χ^2_{m-1} distribution to assess significance. Should the test indicate significance at an appropriate level, rough pairwise comparisons can be made as in the comparison of means in the analysis of variance. To see how to do this first note that, as above, $(n-1)S^2/\sigma^2$ has the χ^2_{n-1} distribution which, for large n, is approximately $N((n-1), 2(n-1))$. Hence S^2 is approximately $N(\sigma^2, 2\sigma^4/(n-1))$ and under the null hypothesis of equality of variances for any $i \neq j$ the difference $S_i^2 - S_j^2$ is

approximately $N(0, 2\sigma_i^4/(n_i-1) + 2\sigma_j^4/(n_j-1))$, and $\operatorname{var}(S_i^2 - S_j^2)$ can be estimated by $d_i + d_j$. A least significant difference may be constructed in the usual way.

Appendix

The Moore-Penrose Inverse

One of several pseudo-inverses or generalised inverses is the Moore-Penrose inverse: see, for example, [9, section 8.11]. The unique Moore-Penrose inverse B^+ of a real symmetric matrix B satisfies

$$\begin{aligned} \mathrm{B}^+\mathrm{B}\mathrm{B}^+ &= \mathrm{B}^+, \\ \mathrm{B}\mathrm{B}^+\mathrm{B} &= \mathrm{B}, \\ \left(\mathrm{B}^+\mathrm{B}\right)^T &= \mathrm{B}^+\mathrm{B}, \\ \left(\mathrm{B}\mathrm{B}^+\right)^T &= \mathrm{B}\mathrm{B}^+. \end{aligned} \tag{A.1}$$

It is routine to show the following.

(i) If $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$, then $\Lambda^+ = \operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_r^{-1}, 0, \ldots, 0)$.

(ii) If H is orthogonal, then $\mathrm{H}^+ = \mathrm{H}^T$.

(iii) If A is idempotent, then $\mathrm{A}^+ = \mathrm{A}$.

(iv) If the subsequent matrix products are defined, then $(\mathrm{BC})^+ = \mathrm{C}^+\mathrm{B}^+$ and $(\mathrm{ABC})^+ = \mathrm{C}^+\mathrm{B}^+\mathrm{A}^+$.

It is well known that if X is $N_p(0, \Sigma)$ with rank $(\Sigma) = r < p$, then $X^T\Sigma^+X$ has the χ_r^2 distribution where Σ^+ is a pseudoinverse of Σ. For the scenario here it is reasonable to test H against K using the test statistic $\widehat{\theta}^T(\widehat{\operatorname{cov}}(\widehat{\theta}))^+\widehat{\theta}$.

References

[1] J. C. W. Rayner, "The asymptotically optimal tests," *Journal of the Royal Statistical Society Series D*, vol. 46, no. 3, pp. 337–346, 1997.

[2] J. C. W. Rayner, O. Thas, and D. J. Best, *Smooth Tests of Goodness of Fit: Using R*, John Wiley & Sons, Singapore, 2nd edition, 2009.

[3] A. W. van der Vaart, *Asymptotic Statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, UK, 1998.

[4] J. T. Kent, "Robust properties of likelihood ratio tests," *Biometrika*, vol. 69, no. 1, pp. 19–27, 1982.

[5] D. Boos, "On generalised score tests," *The American Statistician*, vol. 47, pp. 327–333, 1992.

[6] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 50, no. 1, pp. 1–25, 1982.

[7] B. L. Welch, "The significance of the difference between two means when the population variances are unequal," *Biometrika*, vol. 29, pp. 350–362, 1937.

[8] P. Rippon, J. Rayner, and O. Thas, "A competitor for the test Welch test in the Behrens-Fisher problem," Unpublished Report, 2008.

[9] J. L. Goldberg, *Matrix Theory with Applications*, McGraw-Hill, New York, NY, USA, 1991.