PERMUTATION TEST FOR A MULTIPLE LINEAR REGRESSION MODEL

Siriwan Tantawanich
Department of Mathematics and Statistics,
Huachiew Chalermphakiet University, Samutphrakhran, Bangkok 10540
EMAIL: s_tantawanit@yahoo.com

Pachitjanut Siripanich
School of Applied Statistics,
The National Institute of Development Administration (NIDA), Bangkapi, Bangkok 10240
EMAIL: oui52@as.nida.ac.th

ABSTRACT

To test the partial regression coefficients in a multiple linear regression model, a nonparametric permutation test based on the partial F statistic is proposed and, when sample size is large, it is significantly as good as the partial F test even in the case of normal distribution. This test can be applied to any distribution of errors. Simulation results show that, for non-normal distributions and in term of power, the proposed test performs quite well compared with the partial F test and the other permutation tests.

Keyword: Nonparametric permutation test, partial F statistic, exchangeable

INTRODUCTION

In a multiple linear regression model, the parametric partial F test is a common procedure for testing the partial regression coefficients. The partial F test is believed to be efficient in terms of power of the test. However, this is true only when the regression errors are assumed to be normally distributed. If this is not the case, then the test may not even be consistent. Recently, the nonparametric permutation test, introduced by Sir Fisher [7], is applied to test the partial regression coefficients. Because this test is based on permutation theory, it does not require the normality assumption. Two procedures are proposed by Manly [11] and Ter Braak [20]. These procedures are revised by Kennedy and Cade [10] and Anderson and Robison [2]. Both tests, however, determine the distribution of the test statistic, under the null hypothesis, by using permutation theory. The simulation study by Anderson and Legendre [1] suggested that these tests are not difference in power, but all have higher power than the partial F test when errors are non-normal.

In this paper, an alternative permutation test is proposed so that its power is higher than that of some existing tests under any distributions of i.i.d. errors.

TESTS FOR PARTIAL REGRESSION COEFFICIENTS

Review of Literatures

Consider a multiple linear regression model,

\[ y = W\gamma + e = X\beta + Z\theta + e, \]

where \( y \) is an \( n \times 1 \) vector of responses, \( e \) is an \( n \times 1 \) vector of i.i.d. errors, \( W = (X, Z) \) is an \( n \times (p + q) \) matrix of predictors, \( X \) is an \( n \times p \) and \( Z \) is an \( n \times q \), \( \gamma = (\beta, \theta)' \) is a \( (p + q) \times 1 \) vector of partial regression coefficients such that \( \beta \) is a \( p \times 1 \) and \( \theta \) is a \( q \times 1 \). The hypotheses of interest is

\[ H_0 : \theta = 0 \quad \text{against} \quad H_1 : \theta \neq 0 \]

where \( \theta \) is a \( q \times 1 \) vector of zeroes. Let \( H = I - W(W'W)^{-1}W' \) and \( M = I - X(X'X)^{-1}X' \); then the OLS estimator of \( \theta \) and \( \beta \) are [19, P. 66]

\[ \hat{\theta} = (Z'MZ)^{-1}Z'My \quad \text{and} \quad \hat{\beta} = (X'X)^{-1}X'(y - Z\hat{\theta}) \]

The parametric partial F statistic [17, P. 181] for testing \( H_0 : \theta = 0 \) is of the form:
\[
F = \frac{y' \left[ W(W'W)^{-1} W' - X(X'X)^{-1} X' \right] y / q}{y' \left[ I - W(W'W)^{-1} W' \right] y / (n - p - q)}, \quad (4)
\]

\[
\hat{\theta} = \frac{\hat{\theta}(ZMZ) \hat{\theta}}{y'Hy / (n - p - q)}. \quad (5)
\]

Under normality assumption, the partial F statistic in (4) is distributed as \( F_{\alpha, (q,n - p - q)} \) when \( H_0 : \theta = 0 \) is true and the null hypothesis is rejected if \( F \geq F_{\alpha, (q,n - p - q)} \).

In recent years, a nonparametric permutation test was applied for testing \( H_0 : \theta = 0 \) where a distribution of the partial F statistic in (4) is obtained by using permutation theory. Hence the normality assumption need not be assumed.

Two well-known permutation tests are proposed by Manly [11, p.156-162] and Ter Braak [20, p. 84].

Under the null hypothesis (2), the i.i.d. random vector \( y \) is exchangeable and hence all possible permutations of \( y \) have the same distribution [18, p.15]. A permutation vector of \( y \), denoted by \( y^* \), is a vector where its elements are permutation of elements of \( y \). Substituting a vector \( y \) in (4) by \( y^* \), Manly’s permutation statistic, denoted by \( F_M \), is

\[
F_M = \frac{y^* \left[ W(W'W)^{-1} W' - X(X'X)^{-1} X' \right] y^* / q}{y^* \left[ I_n - W(W'W)^{-1} W' \right] y^* / (n - p - q)}. \quad (6)
\]

Since there are \( n! \) possible values of \( y^* \), so an empirical permutation distribution for testing the hypothesis (2) can be obtained from \( n! \) \( F_M \)'s. Note that each value of \( F_M \) is equal likely and so, at a significance level \( \alpha \), \( H_0 : \theta = 0 \) is rejected if

\[
P(F_M \geq F) = \frac{\text{Number of } F_M \text{'s } \geq F}{n!} \leq \alpha .
\]

Ter Braak [20] proposed a permutation on the vector of residuals, \( r = y - (X\hat{\beta} + Z\theta) \). Let \( r^* \) denote a random permutation of the vector \( r \). An empirical permutation distribution for testing the hypothesis (2) can be obtained from \( n! \) values of Ter Braak’s permutation statistic \( F_T \) which is defined as follow.

\[
F_T = \frac{y^* \left[ W(W'W)^{-1} W' - X(X'X)^{-1} X' \right] y^* / q}{y^* \left[ I_n - W(W'W)^{-1} W' \right] y^* / (n - p - q)}, \quad (7)
\]

where \( y^* = X\hat{\beta} + Z\theta + r^* \). Similar to Manly’s procedure, \( H_0 : \theta = 0 \) is rejected if

\[
P(F_T \geq F) = \frac{\text{Number of } F_T \text{'s } \geq F}{n!} \leq \alpha .
\]

It is a huge and messy task to calculate all \( n! \) permuted statistics when \( n \) is greater than 15. One way to cope this problem is selecting \( m \) out of \( n! \) permuted statistics with or without replacement at random. The consequence test is called “sampling permutation test” and was first introduced by Vadeviloo [21]. Ter Braak [20] suggested that at significance level \( \alpha \), \( H_0 : \theta = 0 \) is rejected when \( k/(m + 1) \leq \alpha \), where \( k \) is the total number out of \( m \) permuted statistics for which its value is greater than \( F \) statistic. The proposed test is a sampling permutation test developed based on the ideas of Vadeviloo and Ter Braak.

**The Proposed Test**

For the OLS estimator of \( \theta \) that is shown in (3),

\[
\hat{\theta} - \theta = (Z'MZ)^{-1} Z'M y - \theta = (Z'MZ)^{-1} Z'M \left[ y - (X\hat{\beta} + Z\theta) \right] = (Z'MZ)^{-1} Z'M e, \quad (8)
\]

since \( MX = \left[ I - X(X'X)^{-1} X' \right] X = 0 \).

Regard the vector of residuals, \( r = y - (X\hat{\beta} + Z\theta) \), as an estimator of the vector of errors \( e \), equation (8) suggests a permuted estimator for \( \theta \) to be

\[
\hat{\theta}^* = \hat{\theta} + (Z'MZ)^{-1} Z'M r^*, \quad (9)
\]

where \( r^* \) is a vector of random permutation of \( r \).
Multiplying the partial F statistic in (5) by \( \frac{q}{n-p-q} \) yields the same result as omitting the degrees of freedom, such a result is

\[
\tilde{F} = \frac{\hat{\theta} (Z'MZ) \hat{\theta}}{y'Hy}.
\]

(10)

Substituting \( \hat{\theta} \) in (10) by \( \hat{\theta}^* \) in (9) the statistic \( \tilde{F} \) becomes a permuted statistic \( F_{S} \) and is defined as follow.

\[
F_{S} = \frac{\hat{\theta}^* (Z'MZ) \hat{\theta}^*}{y'Hy}
\]

(11)

For a proposed sampling permutation test of hypotheses in (2), an (estimated) empirical distribution of \( F \) is obtained from \( m \) permutation statistics \( F_{S} \) (selected at random) and \( H_0: \theta = 0 \) is rejected if

\[
\text{(number of } F_{S} \geq \tilde{F})/(m+1) \leq \alpha.
\]

**Simulation Study**

Consider a model

\[ y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \theta_1 z_{i1} + \theta_2 z_{i2} + e_i, \]

where \( i = 1, 2, \ldots, n \). For each \( i, x_{i1}, x_{i2}, z_{i1} \) and \( z_{i2} \) are generated from a uniform distribution with mean 1 and variances 2, 3, 5 and 6 respectively, \( e_i \) is generated from either \( N(0, 1) \), \( U(0, 1) \) or Laplace \( (0, 1) \) and hence \( y_i \) can be calculated for each combination of \( (\beta_1, \beta_2) \) and \( (\theta_1, \theta_2) \). In this study \( (\beta_1, \beta_2) = (1, 1), (2, 2); (\theta_1, \theta_2) = (0, 0), (0.1, 0.1), (0.1, 0.3), (0.1, 0.7), (0.1, 0.12); \) and \( n = 8 \) (small), 12 (moderate), 16, 20 (large) are given. As recommended by Manly (1997, p. 32-36), a permuted sampling number \( m \) is chosen to be 999 where the significance level \( \alpha \) is 0.05. In addition, a number of simulation runs for each combination is 5,000 runs.

The empirical type I error is estimated from the rejection rates out of 5,000 when the \( H_0 \) is true, i.e., \( (\theta_1, \theta_2) = (0, 0) \) and the empirical power is estimated from the rejection rates out of 5,000 when the null hypothesis is false, i.e., \( (\theta_1, \theta_2) = (0.1, 0.1), (0.1, 0.3), (0.1, 0.7), (0.1, 0.12). \)

Table 1 exhibits empirical type I errors (the rejection rate out of 5000 simulations) of all 4 tests, namely the proposed test, Ter Braak test, Manly test and the partial F test. Further, the values in Table 1 are used for testing whether the type I error is statistically different from \( \alpha = 0.05 \) or not.

For normal errors, all tests have type I errors which are not significantly different from 0.05 in all sets of simulations. Unsurprisingly, the partial F test has type I errors which are closer to 0.05 than the others. For non-normal errors (uniform and Laplace), the partial F test has type I errors which are significantly lower than 0.05 when sample size are small to moderate (\( n = 8, 12 \)), whereas the proposed test and the other permutation tests have type I errors which are not significantly different from 0.05 in all sets of simulations.

For (estimated) empirical powers of all tests, results are shown in Figures 1-3 and can be summarized as follow:

For normal errors, the partial F test has highest power as expected, follows descending by the proposed test, Ter Braak test and Manly test in all set of simulations. When the sample size is large (\( n = 16 \) or 20), there are no difference in power between the partial F test and the proposed test. Moreover all tests show increasing in power as \( (\theta_1, \theta_2) \) increases. On the contrary, there are decreasing in the power as \( (\beta_1, \beta_2) \) increases. See Figure 1.

When Uniform and Laplace distribution are implemented, the proposed test has highest power whereas the partial F test has lowest power in all sets of simulations. The differences of powers among the proposed test and the others can be detected for small sample size (\( n = 8 \)). When the sample size increases, the power curves of the three permutation tests (proposed test, Manly test and Ter Braak test) are not much difference. In this situation, proposed test still has highest power. See Figure 2 and Figure 3.

**CONCLUSION**

Conclusion are the following.

1. For normal errors case, it is no doubt about the efficiency of the partial F test. It has highest power and the type I error are closest to the significant level \( \alpha \) than the other tests in all situations. The proposed test performs as good as the partial F test in this case for large sample size (\( n = 16, 20 \)). Among three permutation tests, the proposed test has highest power and its type I error are closer to \( \alpha \) than that of Ter Braak test and Manly test, in all situations.

2. For non-normal errors case, the proposed test has highest power and its type I error are closest to the significant level \( \alpha \).
in all situations. On the contrary, the partial F test has lowest power, especially, for small sample size.

Among these 4 tests, the proposed test seems to be the best choice. However, there is not much difference among the powers of proposed test, Ter Braak test, and Manly test as sample size increases.

In this study, the sampling permutation test was applied to the proposed test and the estimated significant level was determined from $m$ out of $(n! - 1)$ values of the permutation statistic. It is interesting to find a sufficient value of $m$.

<table>
<thead>
<tr>
<th>Tests</th>
<th>n</th>
<th>$e_i \sim \text{Normal}(0,1)$</th>
<th>$e_i \sim \text{Uniform}(0,1)$</th>
<th>$e_i \sim \text{Laplace}(0,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1,1)</td>
<td>(1,1)</td>
<td>(1,1)</td>
</tr>
<tr>
<td>Proposed Test</td>
<td>8</td>
<td>0.0488</td>
<td>0.0495</td>
<td>0.0490</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.0491</td>
<td>0.0497</td>
<td>0.0495</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.0499</td>
<td>0.0501</td>
<td>0.0501</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.0502</td>
<td>0.0500</td>
<td>0.0500</td>
</tr>
<tr>
<td>Ter Braak Test</td>
<td>8</td>
<td>0.0481</td>
<td>0.0470</td>
<td>0.0500</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.0487</td>
<td>0.0495</td>
<td>0.0493</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.0492</td>
<td>0.0501</td>
<td>0.0495</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.0501</td>
<td>0.0500</td>
<td>0.0497</td>
</tr>
<tr>
<td>Manly Test</td>
<td>8</td>
<td>0.0476</td>
<td>0.0485</td>
<td>0.0489</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.0488</td>
<td>0.0488</td>
<td>0.0492</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.0490</td>
<td>0.0493</td>
<td>0.0492</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.0503</td>
<td>0.0494</td>
<td>0.0502</td>
</tr>
<tr>
<td>Partial F Test</td>
<td>8</td>
<td>0.0496</td>
<td>0.0433</td>
<td>0.0438</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.0500</td>
<td>0.0437</td>
<td>0.0464</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.0499</td>
<td>0.0437</td>
<td>0.0464</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.0500</td>
<td>0.0488</td>
<td>0.0488</td>
</tr>
</tbody>
</table>

Note: The bold face implies that the type I error is statistically different from $\alpha = 0.05$.

REFERENCES

Figure 1 Power comparison of the four tests at $\alpha = 0.05$ for different sample size (n) and different values of $(\beta_1, \beta_2)$ when the errors are i.i.d. normal.
Figure 2  Power comparison of the four tests at $\alpha = 0.05$ for different sample size (n) and different values of $(\beta_1, \beta_2)$ when the errors are i.i.d. uniform.
Figure 3  Power comparison of the four tests at $\alpha = 0.05$ for different sample size ($n$) and different values of $(\beta_1, \beta_2)$ when the errors are i.i.d. Laplace.