A JONCKHEERE TYPE TEST FOR THE PARALLELISM OF REGRESSION LINES

Eunsook Jee

ABSTRACT. In this paper, we propose a Jonckheere type test statistic for testing the parallelism of k regression lines against ordered alternatives. The order restriction problems could arise in various settings such as location, scale, and regression problems. But most of theory about the statistical inferences under order restrictions has been developed to deal with location parameters. The proposed test is an application of Jonckheere's procedure to regression problem. Asymptotic normality and asymptotic distribution-free properties of the test statistic are obtained under some regularity conditions.

1. INTRODUCTION

Consider the linear regression model

$$(1.1) \qquad Y_{ij} = \alpha_i + \beta_i x_{ij} + \epsilon_{ij}, \quad j = 1, \cdots, n; \ i = 1, \cdots, k$$

Proper assumptions are mentioned in [3]. We are interested in testing the parallelism of k regression lines against ordered alternatives. That is, we want to test

$$(1.2) \qquad H_0 : \beta_1 = \cdots = \beta_k = \beta \ \ (\text{unknown})$$

against the ordered alternatives

$$(1.3) \qquad H_1 : \beta_1 \leq \cdots \leq \beta_k$$

where at least one inequality is strict. Jonckheere [4] has suggested some non-parametric approaches to test the equality of location parameters against ordered alternatives and the author [2] also has considered a distribution-free test which applied the Jonckheere's to regression problem. Parametric or nonparametric tests for the parallelism of regression lines against the ordered alternatives (1.3) have been

considered by Adichie [1], Rao and Gore [7], among others. Rao and Gore [7] have developed a distribution-free test. They have assumed that the independent variables x_{ij} are equispaced. Song, Huh, and Kang [8] have considered a distribution-free test which utilizes the Kepner and Robinson's contrast idea [5]. The results of a Monte Carlo study show that the proposed test is robust, but it has somewhat low efficiencies.

2. THE PROPOSED TEST STATISTIC

Now we want to construct an asymptotically distribution-free rank test statistic based on residuals which is Jonckheere type statistic.

Let $\hat{\beta}$ be a consistent estimator of the common slope β under H_0. The proposed test statistic is then defined by

$$(2.1) \qquad J(\hat{\beta}) = \sum_{u<v}^{k} U(\hat{\beta}),$$

where Mann-Whitney type statistic

$$(2.2) \qquad U_{uv}(\hat{\beta}) = \sum_{s=1}^{n_u} \sum_{t=1}^{n_v} \Phi(Z_{vt}(\hat{\beta}) - Z_{us}(\hat{\beta})),$$

where

$$(2.3) \qquad Z_{ij}(\hat{\beta}) = (Y_{ij} - \hat{\beta}x_{ij})\text{sign}(x_{ij}), \quad j = 1, \cdots, n_i; \ i = 1, \cdots, k.$$

Under the ordered alternatives H_1 in (1.3) the values of $J(\hat{\beta})$ is expected to be large. The exact distribution of $J(\hat{\beta})$ may be too complicated to be useful. We thus want to use the approximate distribution of $J(\hat{\beta})$.

3. ASYMPTOTIC PROPERTIES

Assuming the regression model (1.1), the proposed statistic $J(\hat{\beta})$ in (2.1) is the sum of the form of (2.2) which is the Mann-Whitney statistic applied to the residuals from the uth and v^{th} regression lines. While, if $\hat{\beta}$ is replaced by β in (2.2), the statistic becomes

$$(3.1) \qquad U_{uv}(\beta) = \sum_{s=1}^{n_u} \sum_{t=1}^{n_v} \Phi(Z_{vt}(\beta) - Z_{us}(\beta)),$$

which is the Mann-Whitney statistic applied to independent observations.

We now want to prove the asymptotic equivalence of $U_{uv}(\hat{\beta})$ and $U_{uv}(\beta)$ for each (u,v) to show that $J(\hat{\beta})$ and $J(\beta)$ have the same limiting distributions. Now we

assume that

(3.2) $$\lim_{N\to\infty} \frac{n_i}{N} = \lambda_i, \quad 0 < \lambda_i < 1, \ i = 1, \cdots, k$$

with

$$N = \sum_{i=1}^{k} n_i.$$

The U statistics corresponding to $U_{uv}(\hat{\beta})$ and $U_{uv}(\beta)$ are respectively given by

(3.3) $$U_{uv}^*(\hat{\beta}) = \frac{1}{n_u n_v} U_{uv}(\hat{\beta})$$

and

(3.4) $$U_{uv}^*(\beta) = \frac{1}{n_u n_v} U_{uv}(\beta),$$

which are two-sample U statistics.

The following theorem gives some conditions under which the two-sample U statistics $U_{uv}^*(\hat{\beta})$ and $U_{uv}^*(\beta)$ are asymptotically equivalent.

Theorem 3.1. *Assume that the density f of the error terms and the design points x_{ij} satisfy the following conditions:*

D_1 : *f is bounded by M_1 and symmetric about zero.*

D_2 : *There exists an $M_2 > 0$ such that for each (u, v)*

$$\max_{s,t} \big| |x_{vt}| - |x_{us}| \big| \le M_2.$$

D_3 : *Let*

$$|x|_{k\cdot} = \frac{1}{n_i} \sum_{j}^{n_i} |x_{ij}|, \quad i = 1, \cdots, k.$$

Then, for each (u, v)

$$|x|_{v\cdot} - |x|_{u\cdot} \to 0 \quad as \quad N \to \infty.$$

Then, under H_0,

$$\sqrt{n}\big[U_{uv}^*(\hat{\beta}) - U_{uv}^*(\beta)\big] \to 0,$$

where $\hat{\beta}$ is a $\sqrt{n}$-consistent estimator of β defined in Section 1.

Proof. It is enough to show that the conditions, for the two-sample U statistic U_{uv}, are satisfied. Let

$$h_{s,t}(x_{us}; x_{vt}; \gamma) = \Phi(z_{vt}(\gamma) - z_{us}(\gamma)).$$

For convenience, we omit the subscripts u and v, if necessary. Then $h_{s,t}(\cdot)$ is the corresponding kernel of degree (1.1), and we have

$$
\begin{aligned}
\mu_{s,t}(\gamma) &= E_\beta\big[h_{s,t}(x_{us}; x_{vt}; \gamma)\big] \\
&= E_\beta\big[\Phi\big[(\epsilon_{vt} + \beta x_{vt} - \gamma x_{vt})\mathrm{sign}(x_{vt}) \\
&\qquad\qquad - (x_{us} + \beta x_{us} - \gamma x_{us})\mathrm{sign}(x_{us})\big]\big] \\
&= E_\beta\big[\Phi\big[W_{vtus} - (\gamma - \beta)(|x_{vt}| - |x_{us}|)\big]\big],
\end{aligned}
$$

(3.5)

where W_{vtus} is defined by

$$
W_{vtus} = \epsilon_{vt}\mathrm{sign}(x_{vt}) - \epsilon_{us}\mathrm{sign}(x_{us}).
$$

Note that, since ϵ_{ij}'s are symmetric about zero, W_{vtus} is identically distributed for every (s,t). Let G and g be the cdf and density of W_{vtus}, respectively. Then, $\mu_{s,t}(\gamma)$ in (3.5) becomes

$$
\begin{aligned}
\mu_{s,t}(\gamma) &= P_\beta\big[W \geq (\gamma - \beta)(|x_{vt}| - |x_{us}|)\big] \\
&= 1 - G((\gamma - \beta)(|x_{vt}| - |x_{us}|)).
\end{aligned}
$$

Thus, according to D_1 and D_2,

(3.6)
$$
\frac{\partial}{\partial \gamma}\mu_{s,t}(\gamma)\big|_{\gamma=\beta} = -g(0)(|x_{vt}| - |x_{us}|)
$$

exists for all (s,t), and is achieved uniformly in (s,t). Randles [6] investigated a number of settings in the following condition

(3.7)
$$
\binom{n}{r}^{-1} \sum_{a \in B_n} \left[\frac{\partial \mu_a(r)}{\partial r}\big|_{r=\theta}\right] \to 0
$$

was satisfied. Now we check the condition (3.7). Let

$$
\mu(r) = E_\beta\big[U^*_{uv}(r)\big].
$$

Then, from (3.6) we have

$$
\begin{aligned}
\frac{\partial \mu(r)}{\partial r}\big|_{r=\beta} &= \frac{1}{n_u n_v}\sum_{s=1}^{n_u}\sum_{t=1}^{n_v}\frac{\partial}{\partial r}\mu_{s,t}(r)\big|_{r=\beta} \\
&= \frac{1}{n_u n_v}\sum_{s=1}^{n_u}\sum_{t=1}^{n_v}\big[-g(0)(|x_{vt}| - |x_{us}|)\big] \\
&= -g(0)(|x|_{v\cdot} - |x|_{u\cdot}) \to 0
\end{aligned}
$$

as $N \to \infty$ by the condition D_3. Thus, we have

$$
\sqrt{N}\big[U^*_{uv}(\hat{\beta}) - U_{uv}(\beta) \to 0,
$$

which completes the proof. □

From Theorem 3.1 we have the following theorem which indicates the asymptotic equivalence of $J(\hat{\beta})$ and $J(\beta)$.

Theorem 3.2. *Assume that the conditions $D_1 \sim D_3$ are satisfied for every (u,v). Assume also that the sample sizes satisfy the condition (3.2). Then, under H_0,*

$$N^{-\frac{3}{2}}\left[J(\hat{\beta}) - J(\beta)\right] \to 0,$$

where $\hat{\beta}$ is a $\sqrt{N}$-consistent estimator of β.

Since the null distribution of $J(\beta)$ is same as that of the Jonckheere statistic, we have the following corollary.

Table 1. Empirical Powers for Tests on Parallelism ($k = 3$; $n_1 = n_2 = n_3 = 6$)

α	0.05	0.05	0.05	0.10	0.10	0.10
m	0	1	2	0	1	2
(a) Uniform Distribution						
$\overline{x_k^2}$	0.035	0.107	0.280	0.085	0.232	0.490
S	0.037	0.139	0.336	0.088	0.271	0.538
G	0.061	0.173	0.367	0.097	0.245	0.468
J	0.063	0.435	0.894	0.105	0.619	0.957
(b) Normal Distribution						
$\overline{x_k^2}$	0.035	0.124	0.320	0.083	0.259	0.494
S	0.046	0.138	0.360	0.098	0.282	0.538
G	0.063	0.176	0.383	0.091	0.264	0.489
J	0.063	0.467	0.896	0.112	0.647	0.964
(c) Double Exponential Distribution						
$\overline{x_k^2}$	0.039	0.163	0.390	0.090	0.301	0.589
S	0.051	0.178	0.444	0.098	0.329	0.622
G	0.064	0.206	0.427	0.100	0.278	0.532
J	0.052	0.581	0.926	0.097	0.728	0.769
(d) Cauchy Distribution						
$\overline{x_k^2}$	0.037	0.138	0.282	0.078	0.265	0.459
S	0.037	0.160	0.330	0.090	0.290	0.478
G	0.053	0.137	0.280	0.084	0.200	0.360
J	0.063	0.454	0.746	0.120	0.603	0.850

Corollary 3.3. *Under the conditions in Theorem 3.2, the limiting distribution of*

$$\left[J(\hat{\beta}) - E_0(J(\beta))\right] / \left[Var_0(J(\beta))\right]^{\frac{1}{2}}$$

is standard normal when H_0 is true, where

$$E_0(J(\beta)) = \frac{1}{4}\left[N^2 - \sum_{i=1}^{k} n_i^2\right]$$

and

$$Var_0(J(\beta)) = \frac{1}{72}\left[N^2(2N + 3) - \sum_{i=1}^{k} n_i^2(2n_i + 3)\right]$$

Table 2. Empirical Powers for Tests on Parallelism ($k = 5$; $n_1 = n_2 = n_3 = 10$)

α	0.05	0.05	0.05	0.10	0.10	0.10
m	0	1	2	0	1	2
(a) Uniform Distribution						
$\overline{x_k^2}$	0.034	0.296	0.674	0.079	0.448	0.820
S	0.043	0.347	0.760	0.104	0.509	0.886
G	0.050	0.319	0.707	0.105	0.464	0.831
J	0.040	0.786	1.000	0.078	0.891	1.000
(b) Normal Distribution						
$\overline{x_k^2}$	0.032	0.272	0.766	0.086	0.417	0.874
S	0.034	0.343	0.829	0.101	0.478	0.910
G	0.048	0.307	0.744	0.102	0.460	0.867
J	0.050	0.831	1.000	0.098	0.903	1.000
(c) Double Exponential Distribution						
$\overline{x_k^2}$	0.038	0.371	0.856	0.100	0.531	0.940
S	0.048	0.415	0.899	0.104	0.585	0.964
G	0.048	0.354	0.803	0.104	0.502	0.896
J	0.063	0.931	1.000	0.117	0.964	1.000
(d) Cauchy Distribution						
$\overline{x_k^2}$	0.042	0.253	0.632	0.095	0.403	0.788
S	0.051	0.302	0.703	0.102	0.443	0.828
G	0.061	0.197	0.468	0.126	0.309	0.624
J	0.049	0.753	0.985	0.108	0.851	0.993

4. SMALL SAMPLE MONTE CARLO STUDY

In this chapter a series of Monte Carlo simulations is conducted to compare the empirical powers and the empirical significance levels of our proposed test statistic with other nonparametric test statistics. We compare the efficiencies of our proposed statistic J with Adichie's nonparametric statistics $\overline{\chi_k^2}$, S and Rao-Gore nonparametric statistic G.

In our Monte Carlo study the powers and significance levels are compared for various underlying distributions such as the uniform, normal, double exponential, and Cauchy distributions.

To design the simulation experiment, two cases of number of k regression lines are considered. Those are $k = 3$ and $k = 5$. The numbers of observations for each line are $n_i = 6$ and $n_i = 10$.

The design points x_{ij} are fixed with $(2, 4, 6, 8, 10)$ for $n_i = 6$ and with $(1, 2, \cdots, 10)$ for $n_i = 10$, and Y_{ij}'s are obtained from the model (1.1).

For the Adichie's test and our proposed test the intercepts $\alpha_i = 0$, $i = 1, \cdots, k$, for convenience. For the choice of β_i's under the ordered alternatives (1.3), we considered the equally-spaced slopes given by

$$\beta_i = \beta_0 + (i - 1)m\Delta, \quad i = 1, \cdots, k,$$

where Δ is the standard deviation of the least squares estimator of β for the combined sample. The initial value β_0 was set to be 1. The values of m indicate the significance of ordering of the slope parameters. We have chosen the values of m as $m = 0, 1$ or 2.

The error terms e_{ij} were generated from the uniform, normal, double exponential, and Cauchy distributions.

The total number of replications in each case was 1000. Two levels of significance, $\alpha = 0.05$ and $\alpha = 0.10$ were applied. To compute the empirical powers of the tests, we count the number of times that the null hypothesis H_0 is rejected for each test. Then the empirical power is the number of times of rejecting H_0 divided by 1000. The simulated proportions of rejecting H_0 in favor of the ordered alternatives H_1 among 1000 replications are presented in Table 1 through Table 2.

5. Summary and Conclusions

In this paper we proposed a Jonckheere type test statistic which is based on the ranks of residuals. We showed that the proposed test statistic is not distribution-free but asymptotically distribution-free under some regularity conditions. The Asymptotic null distribution of the proposed test statistic is the same as that of the Jonckheere test statistic in location problem. According to the simulation results it is more efficient than any other tests considered in this paper. The procedure for testing the parallelism of k regression lines without the assumption of equal intercepts is an interesting subject for a further study.

References

1. J.N. Adichie: Testing Parallelism of Regression Lines against Ordered Alternatives. *Communications in Statistics, Theory and Mathematics* **A5** (1976), no. 11, 985-997.

2. E.S. Jee: A Study on a Nonparametric Test for the Parallelism of k Regression Lines Against Ordered Alternatives. *J. Appl. Math. & Computing* **8** (2001), 669-682.

3. _____: A Study on a Nonparametric Test with Regression Depth, *Journal of Institute of Statistics Korea University* **20** (2005), 65-76.

4. A.R. Jonckheere: A Distribution-free k-sample Test Against Ordered Alternatives. *Biometrika* **41** (1954), 133-145.

5. J.L. Kepner & D.H. Robinson: A Distribution-free Rank Test for Ordered Alternatives in Randomized Complete Block Designs. *Journal of American Statistical Association* **79** (1984), 212-217.

6. R.H. Randles: On Tests Applied to Residuals. *Journal of American Statistical Association* **79** (1984), 349-354.

7. K.S.M. Rao & A.P. Gore: Testing Concurrence and Parallelism of Several Sample Regressions Against Ordered Alternatives. *U. Stat. Ser. Statist.* **15** (1984), 43-50.

8. M.S. Song, M.Y. Huh & H.J. Kang: On a Distribution-free Test for Parallelism of Regression Lines Against Ordered Alternatives. *Journal of Korean Society for Quality Control* **15** (1987), no. 2, 50-54.

Department of Mathematics, Kwangwoon University, Seoul 139-701, Republic of Korea
Email address: esjee@kw.ac.kr