UIT Report Vol. 12, No. 1, pp. 151~154, 1981　　연구논문집 제 12 권 제 1 호 pp. 151~154, 1981.

On the nonparametric test for parallelism and confidence interval of slope estimators

Dong Hee Kim
Dept. of Applied Mathematics
(Received October 6, 1980)

〈Abstract〉

For testing $\beta_i=\beta$, $i=1,\cdots,k$, in the regression model $Y_{ij}=\alpha_i+\beta_i x_{ij}+e_{ij}$, $j=1,\cdots n_i$, a robust test using weighted median and confidence interval of β_i are proposed. Its asymptotic distribution is proved to be chi-square under the null hypothesis and noncentral chi-square under an appropriate sequence of alternatives.

평행성에 대한 비모수 검정과 slope estimator의 신뢰구간에 관하여

김　동　희
응용 수학과
(1980. 10. 6 접수)

〈요 약〉

회귀모델 $Y_{ij}=\alpha_i+\beta_i x_{ij}+e_{ij}$, $j=1,\cdots,n_i$에서 $\beta_i=\beta$, $i=1,\cdots,k$를 검정하는데 weighted median을 이용한 robust한 검정을 고찰하고 β_i의 신뢰구간을 알아보았다. 귀무가설 하에서의 극한 분포는 χ^2 분포를 하고 대립가설하에서는 noncentral χ^2 분포를 한다는 것을 증명하였다.

I. Introduction

Consider the regression model

(1.1) $Y_{ij}=\alpha_i+\beta_i x_{ij}+e_{ij}$, $j=1,\cdots,n_i$, $i=1,\cdots,k$, where e_{ij}'s are independent and identically distributed (iid) random variables, x_{ij}'s are known constants, α_i's are nuisance parameters, and β_i's are the regression parameters.

Our problem is to test the null hypothesis

(1,2) H_0;　$\beta_1=\beta_2=\cdots\beta_k=\beta$ (unknown)

against the set of alternatives that $\beta_1,\cdots,\beta_k$ are not all equal and to get the confidence interval of β_i in (1.1). Comparisons of several regression coefficients are considered by Hollander (1970), Potthoff(1974), etc.

For the regression model

(1.3) $Y_i=\alpha+\beta x_i+e_i$, $i=1\cdots,n$, where e_i's are iid random variables, Sen (1968) proposed an estimator of β based on Kendall's tau.

Based upon Sen's idea, Song(1978) developed a test for the parallelism of k regression lines. Scholz(1978) investigated weighted median estimator for β in (1.1) and showed that any two weighted median estimators which achieved optimun efficiency were asymptotically equivalent.

In this paper, using Scholz's estimator and Song's technique, a nonparametric test for H_0 using weighted median and a confidence interval of β_i in (1.1) are proposed.

II. Definitions and assumptions

For each $i=1,\cdots,k$, let Y_{ij}, $i=1,\cdots,n_i$, be indendent random variables with distributions

$$P(Y_{ij}\leq y)=F_{ij}(y)=F(y-\alpha_i-\beta_i x_{ij})$$

where $F(y)$ is a continuous cumulative distribution function (cdf), x_{ij}'s are known constants, and α_i's and β_i's are unknown parameters. Without loss of generality, we may assume that

$$x_{i1}\leq x_{i2},\cdots,\leq x_{in_i},\quad i=1,\cdots,k,$$

where all the equality signs are not strict.

Definition 2.1 For $i=1,\cdots,k$, we define

$$\bar{x}_i=(1/n_i)\sum_{j=1}^{n_i}x_{ij}\ ;\ C_{ni}^2=\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$$

Definition 2.2 We define $w=\{w_{ist}\ ;\ s\leq t,\ i=1,\cdots,k\}$ as a set of weights with the following properties;

$w_{ist}\geq 0$,

$w_{ist}=0$ whenever $x_{it}=x_{is}$, and

$$\sum_{1\leq s<t\leq n_i}' w_{ist}=1.$$

Definition 2.3 We extend the definition of w_{ist} to all $1\leq s,\ t\leq n_i$ as follows;

$w_{ist}=-w_{its}$ for $s\geq t$.

Definition 2.4 We now define for $i=1,\cdots,k$,

$$W_{ni}=\sum_{s=1}^{n_i}(\sum_{t=1}^{n_i}w_{ist})^2$$

$$\rho_{ni}=\sum_{1\leq s<t\leq n_i}w_{ist}(x_{it}-x_{is})/(W_{ni}^{\frac{1}{2}}C_{ni})$$

$$T_n=\sum_{i=1}^{k}\rho_{ni}C_{ni}^2,\quad \gamma_{ni}=\rho_{ni}^2C_{ni}^2/T_n^2$$

Assumption 1. As $n\to\infty$

$$n_i/(\sum_{i=1}^{k}n_i)\longrightarrow c_i\ ;\ 0<c_0\leq c_1\leq,\cdots,c_k\leq 1-c_0,$$

$$\gamma_{ni}\longrightarrow\gamma_i\ ;\ 0<\gamma_0\leq\gamma_1\leq,\cdots,\gamma_k\leq 1-\gamma_0.$$

where $c_0,\ \gamma_0\leq 1/k$.

Assumption 2.

(i) ρ_{ni} stays bounded away from zero as $n\to\infty$.

(ii) $C_{ni}^2\to\infty$ as $n\to\infty$ for $i=1,\cdots,k$.

Assumption 3. $F(x)$ is an absolutely continuous cdf having a continuous density function satisfying

$$B(F)=\int_{-\infty}^{\infty}f^2(x)dx<\infty.$$

Assumption 4.

$$n_i\sum_{s=1}^{n_i}\sum_{t=1}^{n_i}w_{ist}^2/W_{ni}=0(1)\ \text{as}\ n\to\infty$$

$$\sup_{1\leq s<t\leq n_i}(\Sigma|w_{ist}|)^2=o(1)$$

Definition 2.5 We now define $c(u)$ to be 1, 0, or -1 according as u is $>$, $=$, or <0. For any real b,

$$Z_{ij}(b)=Y_{ij}-bx_{ij}\ \text{for}\ j=1,\cdots,n_i,\ i=1,\cdots,k.$$

Definition 2.6 To estimate β_i, we construct the statistic

$$U_{ni}(b)=\sum_{1\leq s<t\leq n_i}w_{ist}c(Z_{it}(b)-Z_{is}(b))\ \text{for}\ i=1,\cdots,k.$$

Since $x_{is}\leq x_{it}$ for all $s<t$, $Z_{it}(b)-Z_{is}(b)$ is nonincreasing in b for all $1\leq s<t\leq n_i$, and $U_{ni}(b)$ is also nonicreasing in b. $Z_{i1}(\beta_i),\cdots,Z_{in_i}(\beta_i)$ are n_i *iid* random variables having *cdf* $F(y-\alpha_i)$, *for* $i=1,\cdots,k$, independent of $X=(x_{i1},\cdots,x_{in_i})$.

Consequently $U_{ni}(\beta_i)$ is a strictly distribution-free statistic having a distribution symmetric about zero by the definition 2.6. Thus we may estimate β_i by choosing b which makes $U_{ni}(b)$ as close to zero as possible.

Definition 2.7 We define the estimators

$$\beta_{n,i}^{(1)}=\sup\{b\ ;\ U_{ni}(b)>0\},$$

$$\beta_{n,i}^{(2)}=\inf\{b\ ;\ U_{ni}(b)<0\}.$$

$$\beta_{n,i}^{*}=(1/2)(\beta_{n,i}^{(1)}+\beta_{n,i}^{(2)}).$$

Definition 2.8 We also define

$$G_i(t)=\sum_{s<t}w_{ist}I_{(s_{ist}\leq t)}\ \text{with}$$

$$I_{(s_{ist}\leq t)}=1\ \text{if}\ s_{ist}\leq t$$

$$=0\ \text{otherwise for}\ i=1,\cdots,k.$$

Then $G_i(t)$ is the distribution function of the probability distribution of weights w_{ist} over the s_{ist} where

$$s_{ist}=(Y_{it}-Y_{is})/(x_{it}-x_{is})$$

$$=\beta_i+(e_{it}-e_{is})/(x_{it}-x_{is}).$$

Definition 2.9 We define

$$\tilde{\beta}_1=\inf\{t\ :\ G_i(t)>0.5\},$$

$$\tilde{\beta}_2=\sup\ \{t\ ;\ G_i(t)<0.5\}$$

$$\tilde{\beta}_{n,i}=(1/2)(\tilde{\beta}_1+\tilde{\beta}_2).$$

Theorem 2.1 $\beta_{n,i}^{*}$ in definition 2.7 is equivalent to $\tilde{\beta}_{n,i}$ in definition 2.9.

Proof) This theorem is easily proved by definition 2.7 and definition 2.9

Definition 2.10 The proposed sample estimators of β is defined by

$$\beta_n^* = \sum_{j=1}^{k} \gamma_{ni} \beta^*_{n,i}$$

where γ_{ni}'s are defined in definition 2.4 and $\beta_{n,i}^*$'s are defined in definition 2.7.

Definition 2.10 means that β_n^* is a weighted average of the individual weighted median slope estimators for k lines.

Definition 2.11 To construct a test statistic, we define $V_{ni}^2=(1/3)W_{ni}$, where V_{ni}^2 is the variance of $U_{ni}(\beta_i)$ whose proof is in Kim(1979).

Definition 2.12 Now our proposed test-statistic is

$$\hat{U}_n = \sum_{j=1}^{k} (U_{ni}(\beta_n^*)/V_{ni})^2$$

Theorem 2.2 Under H_0 in (1.2) and conditions in this section, $\hat{U}_n$ in definition 2.12 has asymptotically a chi-square distribution with k-1 degrees of freedom(df).

In view of Theorem 2.2, an asymptotic level ε test rejects H_0 in (1.2) if $\hat{U}_n$ is greater than the upper 100ε% point of the chi-square distribution with k-1 df.

Theorem 2.3 $\hat{U}_{ni}(\beta_i)$ has asymptotically a normal distribution with mean zero and variance $V_{ni}^2=(1/3)W_{ni}$.

Proof) By Hoeffeing(1948), $U_{ni}(b)$ is U-statistic and normal, where its mean is zero and variance is $V_{ni}^2=(1/3)W_{ni}$.

III. Asymptotic distribution of $\hat{U}_n$

In order to scudy of the asymptotic power properties of the proposed $\hat{U}_n$-test, we consider the following sequence of alternatives.

(3.1) H_n ; $\beta_i=\beta+(\theta_i/T_n)$, $i=1,\cdots,k$;

$$\sum_{j=1}^{k} \gamma_i \theta_i = 0.$$

Lemma 3.1 (Theorem 7.1 of Sen(1968)) For $i=1,\cdots,k$, under $H;\beta_i=0$ and assumption 2

$U_{ni}(b/(\rho_{ni}C_{ni}))/V_{ni}$

has asymptotically a normal distribution with mean $-2bB(F)W_{ni}^2/V_{ni}$ and unit variance.

If we note that $W^{\frac{1}{2}}_{ni}/V_{ni} \to \sqrt{3}$ as $n\to\infty$, the following lemma can be easily proved using lemma 3.1.

Lemma 3.2 For any real $(a.b)$,

$[U_{ni}(\beta_i-a/(\rho_{ni}C_{ni}))-U_{ni}(\beta_i-b/(\rho_{ni}C_{ni}))]/V_{ni}$

converges (in probability) to $2\sqrt{3}(a-b)B(F)$.

Lemma 3.3 (Theorem 1 of Scholz (1978)) For $i=1,\cdots,k$

$\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_i)$

has asymptotically a normal distribution with mean zero and variance $1/(12B^2(F))$.

Lemma 3.4 (Lemma 3.4 of Song(1978))

(i) For $i=1,\cdots,k$

$|\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_i)|$ is bounded(in probability).

(ii) Under$\{H_n\}$ in (3.1)

$|T_n(\beta_n^*-\beta)|$ is bounded (in probability).

Now we have the main theorem.

Theorem 3.1 Under $\{H_n\}$ in (3.1) and the conditions in section 2, $\hat{U}_n$ has asymptotically a noncentral chi-square distribution with k-1 df and the nocentrality parameter

$$\Delta_u = 12B^2(F)\sum_{j=1}^{k}\gamma_i\theta_i^2$$

Proof) Since $\beta_{n,i}^*$ is a weighted median, it follows that $U_{ni}(\beta^*_{n,i})=0$ for $i=1,\cdots,k$. Thus

$$\hat{U}_n = \sum_{j=1}^{k}(U_{ni}(\beta_n^*)/V_{ni}-U_{ni}(\beta^*_{n,i})/V_{ni})^2 \quad (3.2)$$

By lemma 3.2 and lemma 3.4, the right side of (3.2) is equivalent (in probability) to

$$12B^2(F)\sum_{i=1}^{k}(\rho_{ni}C_{ni}(\beta^*_{n,i}-\beta_n^*))^2 \quad (3.3)$$

Now according to lemma 3.3, $\rho_{ni}C_{ni}(\beta^*_{n,i}-\beta_i)$, for $i=1,\cdots,k$, are independent and each has asymptotically a normal distribution with mean zero and common variance $1/(12B^2(F))$. Note that

$$\rho_{ni}C_{ni}(\beta^*_{n,i}-\beta_i)=\rho_{ni}C_{ni}(\beta^*_{n,i}-\beta)-\gamma^{\frac{1}{2}}_{ni}\theta_i \quad (3.4)$$

We also notice that

$$\sum_{j=1}^{k}(\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta))^2=\sum_{i=1}^{k}(\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_n^*))^2$$
$$+T_n^2(\beta_n^*-\beta)^2-2T_n^2(\beta_n^*-\beta)\sum_{i=1}\gamma_{ni}(\beta_n^*-\beta_{n,i}^*)$$

By (3.1), Since $T_n(\beta_n^*-\beta)=\sum_{i=1}^{k}\gamma_{ni}^{\frac{1}{2}}\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_i)+\sum_{i=1}^{k}\gamma_{ni}\theta_i$, $T_n(\beta_n^*-\beta)$ converges to $\sum_{i=1}^{k}\gamma^{\frac{1}{2}}_{ni}\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_i)$.

Then by lemma 3.3, $T_n(\beta_n^*-\beta)$ has asymptotically a normal distribution with mean zero and variance $1/(12B^2(F))$. From (3.4) and (3.5), we have

$$\begin{aligned}\sum_{i=1}^{k}(\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_n^*))^2 &= \sum_{i=1}^{k}(\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_i))^2 \\ &+\sum_{i=1}^{k}\gamma_{ni}\theta_i-T_{ni}(\beta_n^*-\beta)^2 \\ &+2\sum_{i=1}^{k}\gamma^{\frac{1}{2}}_{ni}\theta_i\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_i).\end{aligned} \tag{3.6}$$

By lemma 3.3, the fourth term in (3.6) is zero.

Thus

$$\sum_{i=1}^{k}(\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_n^*))^2-\sum_{i=1}^{k}\gamma_{ni}\theta_i^2 = \sum_{i=1}^{k}(\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_i))^2-T_n^2(\beta_n^*-\beta)^2. \tag{3.7}$$

Since $\sum_{i=1}^{k}(\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_i))^2$ is a chi-square distribution with k *df* from lemma 3.3, and since $T_n^2(\beta_n^*-\beta)^2$ is also a chi-square distribution with 1 *df*, $\sum_{i=1}^{k}(\rho_{ni}C_{ni}(\beta_{n,i}^*-\beta_n^*))^2$ is a noncentral chi-square distribution with k-1 *df* and noncentrality parameter $\sum_{i=1}^{k}\gamma_{ni}\theta_i^2$.

Hence, from (3.2) and (3.3), the theorem follows.

Ⅲ. Confidence interval of β_i

As $V_{ni}^2=\text{var } U_{ni}(\beta_i)=(1/3)W_{ni}(1+0(1))$ and $E(U_{ni}(\beta_i))=0$,

$(U_{ni}(\beta_i)-0)/V_{ni}\to N(0,1)$ as $n\to\infty$.

Define $U_{ni}^{-1}(g)=\inf\{b\,;\,U_{ni}(b)>g\}$ for $g\in(0,1)$ and let $S_1=U_{ni}^{-1}(Z_{\alpha/2}V_{ni})$, and $S_2=U_{ni}^{-1}(Z_{1-\alpha/2}V_{ni})$.

Then for large n we have

$$1-\alpha\cong P(Z_{\alpha/2}\le U_{ni}(\beta_i)/V_{ni}<Z_{1-\alpha/2}) =P(S_1\le\beta_i\le S_2).$$

i.e., we may consider $[S_1, S_2]$ as a large sample confidence interval for β_i with approximate confidence level $1-\alpha$. S_1 and S_2 are particular ordered slopes s_{ist}.

Theorem 4.1 (Theorem 2 of Scholz (1978)) For weights we have

$\rho_{ni}C_{ni}(S_2-S_1)\longrightarrow Z_{1-\alpha/2}/\sqrt{3}\,F(x)$.

References

1. Gibbons, J.D. "Nonparametric statistical inference." International Student Education, Mcgrow-hill Kogakusha, Ltd, 205—218.
2. Hodges, J.L and Lehmann, E.L. "The efficiency of some nonparametric compatitors of the t-test." Ann. Math. Statist. 7, 324—335 (1956).
3. Hoeffding, W.C. "A class of statistics with asymptotically normal distribution." Ann. Math. Statist. 19, 293—325. (1948).
4. Hollander, M. "A distribution-free test for parallelism." I. Amer. Statist. Assoc. 65, 387—394, (1970).
5. Kim, D.H. "a nonparametric test for the parallelism of regression lines using weighted median." (1979).
6. Potthoff, R.F. "A nonparametric test of whether two simple regression lines are parallel." Ann. Statist. 2, 295—310. (1974).
7. Scholz, F.W. "Weighted median regression estimators." Ann. Statist. 6, 603—609 (1978).
8. Sen, P.K. "Estimates of the regression coefficient based on Kendall's tau." J. Amer. Statist. Assc. 63, 1379—1389 (1968).
9. Song, M.S. "A nonparametric test for the parallelism of regression lines based on Kendall's tau." J. Korean-Statist. Society 7. 17—26. (1978).