Confidence Interval For Sum Of Variance Components In A Simple Linear Regression Model With Unbalanced Nested Error Structure

Dong Joon Park[1)]

Abstract

Those who are interested in making inferences concerning linear combination of variance components in a simple linear regression model with unbalanced nested error structure can use the confidence intervals proposed in this paper. Two approximate confidence intervals for the sum of two variance components in the model are proposed. Simulation study is peformed to compare the methods.

Keywords : variance components, mixed model, inference

1. A Simple Regression Model With An Unbalanced Nested Error Structure

A simple linear regression model with an unbalanced nested error structure is written as

$$Y_{ij} = \mu + \beta X_{ij} + A_i + E_{ij} \qquad (1.1)$$
$$i = 1, \ldots, I; \quad j = 1, \ldots, J_i$$

where Y_{ij} is the j th observation in the i th primary level, μ and β are unknown constants, X_{ij} is a fixed predictor variable, and A_i and E_{ij} are jointly independent normal random variables with zero means and variances σ_A^2 and σ_E^2, respectively, $I > 2$, $J_i \geq 1$, and $J_i > 1$ for at least one value of i. A_i is an error term associated with the first-stage sampling unit(primary level) and E_{ij} is an error term associated with the second-stage sampling unit. One possible partitioning of model (2.1) was shown in Park et al.(2002). Model (2.1) is written in matrix notation,

$$\boldsymbol{y} = \boldsymbol{X}\underline{\boldsymbol{\alpha}} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e} \qquad (1.2)$$

where $\boldsymbol{y}$ is a $J_. \times 1$ vector of observations, $\boldsymbol{X}$ is a $J_. \times 2$ matrix of known values with a column of 1's in the first column and a column of X_{ij}'s in the second column, $\underline{\boldsymbol{\alpha}}$ is a 2×1 vector of parameters with μ and β as elements, $\boldsymbol{Z}$ is a $J_. \times I$ design matrix with 0's and 1's, i.e. $\boldsymbol{Z} = \oplus_{i=1}^{I} \mathbf{1}_{J_{i\times 1}}$, $\boldsymbol{u}$ is an $I \times 1$ vector of random effects, $\boldsymbol{e}$ is a $J_. \times 1$ vector of random error terms, $J_. = \Sigma_{i=1}^{I} J_i$, and $\boldsymbol{D}_{J_.}$ is a $J_. \times J_.$ identity

1) Associate Professor, Division of Mathematical Sciences, Pukyong National University Nam-Gu Daeyeon 3-Dong, Pusan, Korea, 608-737, Email: djpark@pknu.ac.kr

matrix.

2. Confidence Intervals For Sum of Two Variance Components

We propose two approximate confidence intervals for a linear combination of variance components, $\gamma = \sigma_A^2 + \sigma_E^2$. Olsen et al.(1976) used spectral decomposition method to obtain following statistics. They proposed a statistic $SSM = \boldsymbol{U}' \boldsymbol{U}$ which is asymptotically chi-squared distributed. In particular, $\boldsymbol{U}' \boldsymbol{U} / (\sigma_A^2 + \sigma_E^2/\lambda_H) \rightarrow \chi^2_{(I-1)}$ as $\sigma_E^2 \rightarrow 0$ where $\boldsymbol{U} = \boldsymbol{C}^+ \boldsymbol{Z}' (\boldsymbol{D}_{J.} - \boldsymbol{X} (\boldsymbol{X}' \boldsymbol{X})^{-1} \boldsymbol{X}') \boldsymbol{y}$, $\boldsymbol{C}^+$is the Moore-Penrose inverse of $\boldsymbol{C}$, $\boldsymbol{C} = \boldsymbol{Z}' (\boldsymbol{I} - \boldsymbol{X} (\boldsymbol{X}' \boldsymbol{X})^{-1} \boldsymbol{X}') \boldsymbol{Z}$, λ_H is the harmonic mean of positive eigenvalues λ_i of $\boldsymbol{C}$. Thus, $E(SSM) = (I - 1) (\sigma_A^2 + \sigma_E^2/\lambda_H)$. It was also shown that $SSM / (\sigma_A^2 + \sigma_E^2 / \lambda_H)$ and R_T / σ_E^2 are independent. The error sum of squares R_T associated with within regression coefficient estimator is written as $R_T = \boldsymbol{y}' \boldsymbol{T} \boldsymbol{y}$ where $\boldsymbol{T} = \boldsymbol{D}_{J.} - \boldsymbol{X}^* (\boldsymbol{X}^{*'} \boldsymbol{X}^*)^- \boldsymbol{X}^{*'}$, $(\boldsymbol{X}^{*'} \boldsymbol{X}^*)^-$ is a generalized inverse of $\boldsymbol{X}^{*'} \boldsymbol{X}^*$, and $\boldsymbol{X}^* = [\boldsymbol{X} \ \boldsymbol{Z}]$. In addition, R_T / σ_E^2 is shown to be a chi-squared random variable with $J_. - I - 1$ degrees of freedom. The expected mean squares are thus summarized using the distributional properties of error sums of squares

$$E(S_M^2) \doteq \sigma_A^2 + \frac{1}{\lambda_H} \sigma_E^2 = \theta_M \quad \text{and} \tag{2.1a}$$

$$E(S_T^2) = \sigma_E^2 = \theta_T \tag{2.1b}$$

where $S_M^2 = SSM/(I - 1)$ and $S_T^2 = R_T / (J_. - I - 1)$. Thus an approximate $1 - 2\alpha$ two-sided confidence interval for γ using S_M^2 and S_T^2 with Ting et al.(1990) is

$$[S_M^2 + (1-\frac{1}{\lambda_H})S_T^2 - (L_1^2 S_M^4 + (1-\frac{1}{\lambda_H})^2 L_2^2 S_T^4 + (1-\frac{1}{\lambda_H})^2 L_{12} S_M^2 S_T^2)^{\frac{1}{2}};$$

$$S_M^2 + (1-\frac{1}{\lambda_H})S_T^2 + (H_1^2 S_M^4 + (1-\frac{1}{\lambda_H})^2 H_2^2 S_T^4 + (1-\frac{1}{\lambda_H})^2 H_{12} S_M^2 S_T^2)^{\frac{1}{2}}] \tag{2.2}$$

where $F_1 = F_{(\alpha : I-1, J.-I-1)}$, $F_2 = F_{(1-\alpha : I-1, J.-I-1)}$, $L_1 = 1 - 1 / F_{(1-\alpha : I-1, \infty)}$, $L_2 = 1 / F_{(\alpha : J.-I-1, \infty)} - 1$, $L_{12} = [(F_2 - 1)^2 - L_1^2 F_2^2 - L_2^2] / F_2$, $H_1 = 1 / F_{(\alpha : I-1, \infty)} - 1$, $H_2 = 1 - 1 / F_{(1-\alpha : J.-I-1, \infty)}$, $H_{12} = [(1 - F_1)^2 - H_1^2 F_1^2 - H_2^2] / F_1$ and $F_{(\delta : n_1, n_2)}$ is the F -percentile with degrees of freedom of n_1 and n_2 degrees of freedom with δ area to the left. This method is referred to as TING method.

As an alternative generalized p-values method proposed by Khuri et al.(1998) can be

applied to (2.1) to construct a confidence interval on γ. The estimates of σ_E^2 are obtained by $(J_. - I - 1)\, s_T^2 / U^*$ where s_T^2 is an observed value of S_T^2 and U^* has a chi-squared distribution with $(J_. - I - 1)$ degrees of freedom. The estimates of σ_{AE}^2 are obtained by $(I - 1)\, s_M^2 / V^*$ where $\sigma_{AE}^2 = \sigma_A^2 + \sigma_E^2 / \lambda_H$, s_M^2 is an observed value of S_M^2, and V^* has a chi-squared distribution with $(I - 1)$ degrees of freedom. Thus, a generalized pivotal quantity γ can be represented as $\gamma = (I-1)\, s_M^2 / V^* + (1 - 1/\lambda_H)\,(J_. - I - 1)\, s_T^2 / U^*$ Accordingly, an approximate $1-2\alpha$ two-sided confidence interval for γ is

$$[F_\alpha \ ; \ F_{1-\alpha}] \tag{2.3}$$

where F_α is the α th percentile of the distribution constructed by the generalized pivotal quantities. Interval (2.3) is referred to as GEN method.

3. Simulation Study

Four unbalanced patterns were selected for simulation study and are shown in Table 1.

TABLE 1. Unbalanced Patterns Used in Simulation

Pattern	I	J_i
1	3	3 5 10
2	5	1 3 5 7 10
3	7	1 2 4 6 8 10
4	10	1 1 1 5 5 5 5 10 10 10

The values with * in Table 2 represent simulated confidence coefficients less than 0.8866. TING method is too conservative when $\rho < 0.5$ for pattern 1 where $\rho = \sigma_A^2 / (\sigma_A^2 + \sigma_E^2)$ and it is too liberal when $\rho < 0.4$ for pattern 4. Except these values of ρ TING method generates the simulated confidence coefficients close to 0.9. However, GEN method generally maintains the stated confidence coefficients through four patterns although its average interval lengths are slightly wider than TING method as Table 3 shows.

References

[1] Khuri, A. I., Mathew, T., and Sinha, B. (1998), *Statistical Tests for Mixed Linear Models*, John Wiley & Sons, New York.

[2] Olsen, A., Seely, J., and Birkes, D.(1976), Invariant Quadratic Unbiased Estimation for Two Variance Components, *The Annals of Statistics*, vol. 4, no. 5, 878-890.

[3] Park, D. J., Park, S.Y., and Han, M. H.(2002), Comparison of Confidence Intervals on

Variance component In a Simple Linear Regression Model with Unbalanced Nested Error Structure, *The Korean Communications in Statistics*, vol. 9, no. 2, 459-471.

[4] Ting, N., Burdick, R.K., Graybill, F. A., Jeyaratnam, S., and Lu, T.-F. C.(1990), Confidence Intervals on Linear Combinations of Variance Components, *Journal of Statistical Computation*, vol. 35, 135-143.

TABLE 2. 90% Range of Simulated Confidence Coefficients

Pattern	1		2		3		4	
ρ	TING	GEN	TING	GEN	TING	GEN	TING	GEN
0.001	0.9530	0.9105	0.8930	0.9055	0.8875	0.9110	**0.8470***	0.8995
0.1	0.9510	0.9110	0.8930	0.9130	0.9030	0.9155	**0.8635***	0.8960
0.2	0.9455	0.9070	0.9150	0.9100	0.9045	0.8960	**0.8815***	0.8970
0.3	0.9335	0.9140	0.9090	0.8960	0.9035	0.8955	**0.8780***	0.8990
0.4	0.9340	0.9120	0.8985	0.9000	0.8910	0.8945	0.8880	0.9045
0.5	0.9080	0.8995	0.8975	0.9000	0.8980	0.9080	0.9010	0.8890
0.6	0.9170	0.8960	0.8985	0.9125	0.9060	0.8975	0.9060	0.8940
0.7	0.9145	0.8935	0.8990	0.9085	0.9055	0.9005	0.8985	0.8970
0.8	0.9045	0.8935	0.8955	0.8995	0.8995	0.9090	0.8960	0.9045
0.9	0.8980	0.8970	0.8960	0.8915	0.9095	0.8915	0.9115	**0.8860***
0.999	0.9120	0.8970	0.9045	0.8960	0.9000	0.9065	0.9190	0.9000
MAX	0.9530	0.9140	0.9150	0.9130	0.9095	0.9155	0.9190	0.9045
MIN	0.8980	0.8935	0.8930	0.8915	0.8875	0.8915	**0.8470***	**0.8860***

TABLE 3. 90% Range of Average Interval Lengths

Pattern	1		2		3		4	
ρ	TING	GEN	TING	GEN	TING	GEN	TING	GEN
0.001	4.8931	8.1532	2.4860	2.7934	1.7815	2.0020	1.0820	1.1891
0.1	6.3293	9.0126	2.7064	2.9286	1.9360	2.1774	1.1783	1.2550
0.2	7.9243	10.0143	2.9111	3.2356	2.1202	2.3474	1.2595	1.3549
0.3	9.2495	11.2606	3.0704	3.3884	2.3489	2.5690	1.3694	1.4430
0.4	10.3255	12.0546	3.4287	3.6621	2.6040	2.7367	1.4733	1.5422
0.5	11.8025	13.5760	3.7395	3.9111	2.7158	2.9801	1.6043	1.6601
0.6	13.7805	15.4741	4.0792	4.0829	2.9748	3.0298	1.6865	1.7602
0.7	15.0355	15.8402	4.3017	4.4095	3.1640	3.3119	1.8065	1.8265
0.8	16.0754	17.1787	4.5853	4.6070	3.4785	3.5441	1.9643	1.9696
0.9	17.5622	18.2204	4.9296	4.9230	3.6558	3.7980	2.0522	2.0432
0.999	18.7191	19.2593	5.2110	5.1309	3.8416	3.8855	2.1746	2.1635
MAX	18.7191	19.2593	5.2110	5.1309	3.8416	3.8855	2.1746	2.1635
MIN	4.8931	8.1532	2.4860	2.7934	1.7815	2.0020	1.0820	1.1891