**Advances and Applications in Statistics**
© 2013 Pushpa Publishing House, Allahabad, India
Published Online: November 2013
Available online at http://pphmj.com/journals/adas.htm
Volume 37, Number 2, 2013, Pages 187-193

# NOTE ON ESTIMATING PARAMETERS IN HETEROGENEOUS RANDOM EFFECTS MODELS FOR TWO SAMPLE PROBLEM

**Kengo Ueda and Hiroto Hyakutake**

Graduate School of Mathematics
Kyushu University
Motooka, Fukuoka 819-0395, Japan

Faculty of Mathematics
Kyushu University
Motooka, Fukuoka 819-0395, Japan
e-mail: hyakutak@math.kyushu-u.ac.jp

## Abstract

In heterogeneous random effects models, a weighted estimator of parameter in the model is recommended as a good estimator in mean square error. However, a confidence region based on the weighted estimator is conservative. Confidence regions in two sample problem are given by an adjusted weighted estimator and others, approximately. The accuracy of approximation is examined by simulation.

## 1. Introduction

Linear random effects models are often applied to the longitudinal data, see Laird and Ware [5], Diggle et al. [2] and so on. Statistical inferences for the linear random effects models are usually considered under equal error

---

Received: September 18, 2013; Accepted: October 9, 2013

2010 Mathematics Subject Classification: 62F25, 62J05.

Keywords and phrases: confidence region, heteroscedastic error, random effects, weighted estimator.

variances. Rukhin [6] gave a weighted estimator of a parameter in the linear random effects model under the heteroscedastic error and considered the mean square errors for some weights. Let $\mathbf{y}_{ij}$ be the $p_{ij}$-dimensional observation and consider the model

$$\mathbf{y}_{ij} = C_{ij}(\boldsymbol{\theta}_i + A_{ij}\mathbf{b}_{ij}) + \boldsymbol{\varepsilon}_{ij} \qquad (i = 1, 2;\ j = 1, 2, \ldots, n_i), \tag{1}$$

where $C_{ij} : p_{ij} \times q$ and $A_{ij} : q \times r$ are known with $\operatorname{rank} C_{ij} = p(< p_{ij})$ and $\operatorname{rank} A_{ij} = r(\leq q)$, respectively, $\boldsymbol{\theta}_i$ is unknown parameter, $\mathbf{b}_{ij}$ is random effect and $\boldsymbol{\varepsilon}_{ij}$ is the error. $\mathbf{b}_{ij}$ and $\boldsymbol{\varepsilon}_{ij}$ are independent. Let the distributions of $\mathbf{b}_{ij}$ and $\boldsymbol{\varepsilon}_{ij}$ be the $r$-variate normal $N_r(\mathbf{0}, \Psi_i)$ and $p_{ij}$-variate normal $N_{p_{ij}}(\mathbf{0}, \sigma_{ij}^2 I_{p_{ij}})$, respectively. Then $\mathbf{y}_{ij}$ is distributed as $N_{p_{ij}}(C_{ij}\boldsymbol{\theta}_i, \sigma_{ij}^2 I_{p_{ij}} + C_{ij}A_{ij}\Psi_i A'_{ij}C'_{ij})$. By Rukhin [6],

$$\mathbf{x}_{ij} = (C'_{ij}C_{ij})^{-1}C'_{ij}\mathbf{y}_{ij} \quad \text{and} \quad v_{ij} = (\mathbf{y}_{ij} - C_{ij}\mathbf{x}_{ij})'(\mathbf{y}_{ij} - C_{ij}\mathbf{x}_{ij})$$

are sufficient statistics in the model (1), then $\mathbf{x}_{ij}$ is distributed as $N_q(\boldsymbol{\theta}_i, \Sigma_{ij})$ and $v_{ij}/\sigma_{ij}^2$ has the chi-square distribution with $\nu_{ij} = p_{ij} - q$ degrees of freedom (df), $\chi^2_{\nu_{ij}}$, where $\Sigma_{ij} = \sigma_{ij}^2(C'_{ij}C_{ij})^{-1} + A_{ij}\Psi_i A'_{ij}$. Note that $(C'_{ij}V(\mathbf{y}_{ij})^{-1}C_{ij})^{-1}C'_{ij}V(\mathbf{y}_{ij})^{-1}\mathbf{y}_{ij}$ is equal to $(C'_{ij}C_{ij})^{-1}C'_{ij}\mathbf{y}_{ij}$, where $V(\mathbf{y}_{ij}) = \sigma_{ij}^2 I_{p_{ij}} + C_{ij}A_{ij}\Psi_i A'_{ij}C'_{ij}$ for a fixed $(i, j)$, see, e.g., Laird et al. [4] or Davidian and Giltinan [1].

We discuss about the difference of the parameters. This is considered as the multivariate Behrens-Fisher problem, in which approximated methods are given by Yao [9], Johansen [3] and Yanagihara and Yuan [8]. In Section 2, we give a modified Rukhin's estimator and approximated confidence regions of $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$. The accuracy of approximation is examined by simulation in Section 3.

## 2. Confidence Region

Rukhin [6] considered the weighted means $\tilde{\mathbf{x}}_i$ of the form

$$\tilde{\mathbf{x}}_i = \left(\sum_{j=1}^{n_i} W_{ij}\right)^{-1} \sum_{j=1}^{n_i} W_{ij}\mathbf{x}_{ij}, \tag{2}$$

where $W_{ij}$ are non-negative definite matrix weights such that $\sum_j W_{ij}$ is a non-singular matrix. When $\sigma_{ij}^2$ and $\Psi_i$ are known, $\tilde{\mathbf{x}}_i$ with $W_{ij} = V(\mathbf{x}_{ij})^{-1}$ is the best linear unbiased estimator of $\boldsymbol{\theta}_i$. If $W_{ij} = (1/n_i)I_{p_{ij}}$ in (2), then $\tilde{\mathbf{x}}_i = \bar{\mathbf{x}}_i$. Rukhin [6] gave a method for estimating $V(\mathbf{x}_{ij})$ as follows:

Let $\omega_{ij} = \left(\sum_{\ell=1}^{n_i} V(\mathbf{x}_{i\ell})^{-1}\right)^{-1} V(\mathbf{x}_{ij})^{-1}$ and $\tilde{\sigma}_{ij}^2 = v_{ij}/\nu_{ij}$. Define

$$\max[V_{ij}, \tilde{\sigma}_{ij}^2(C'_{ij}C_{ij})^{-1}] = \begin{cases} V_{ij} & (V_{ij} > \tilde{\sigma}_{ij}^2(C'_{ij}C_{ij})^{-1}), \\ \tilde{\sigma}_{ij}^2(C'_{ij}C_{ij})^{-1} & (\text{otherwise}), \end{cases}$$

where $B_1 > B_2$ means that $B_1 - B_2$ is positive definite. Then an estimator $V_{ij} = \hat{V}(\mathbf{x}_{ij})$ of $V(\mathbf{x}_{ij})$ is computed by the iteration scheme

$$V_{ij}^{(r+1)} = \left(I_{p_{ij}} - \frac{1}{2}\omega_{ij}\right)^{-1}(\mathbf{x}_{ij} - \tilde{\mathbf{x}}_i)(\mathbf{x}_{ij} - \tilde{\mathbf{x}}_i)'\left(I_{p_{ij}} - \frac{1}{2}\omega_{ij}\right)^{-1}$$

$$+ \frac{1}{4}\left(I_{p_{ij}} - \frac{1}{2}\omega_{ij}\right)^{-1}\omega_{ij}\max[V_{ij}^{(r)}, \tilde{\sigma}_{ij}^2(C'_{ij}C_{ij})^{-1}]\omega'_{ij}\left(I_{p_{ij}} - \frac{1}{2}\omega'_{ij}\right)^{-1},$$

$r = 0, 1, \ldots, V_{ij}^{(0)} = 0$.

Let the estimator of $V(\tilde{\mathbf{x}}_i)$ be $\hat{V}(\tilde{\mathbf{x}}_i) = \sum_j \omega_{ij}\hat{V}(\mathbf{x}_{ij})\omega'_{ij}$. When $W_{ij} = \hat{V}(\mathbf{x}_{ij})^{-1}$ in (2), that is $\omega_{ij} = \left(\sum_{\ell=1}^{n_i} \hat{V}(\mathbf{x}_{i\ell})^{-1}\right)^{-1} \hat{V}(\mathbf{x}_{ij})^{-1}$, then the estimator of $V(\tilde{\mathbf{x}}_i)$ is $\left(\sum_{\ell=1}^{n_i} \hat{V}(\mathbf{x}_{i\ell})^{-1}\right)^{-1}$. When $W_{ij} = (1/n_i)I_q$, $\tilde{\mathbf{x}}_i$ in (2) is the sample mean $\bar{\mathbf{x}}_i$ and $\hat{V}(\bar{\mathbf{x}}_i) = \{1/n_i(n_i - 1)\}\sum_j (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$.

Based on the above, Rukhin [6] gave an approximate $1 - \alpha$ confidence region for $\boldsymbol{\theta}_i$ of the form

$$(\tilde{\mathbf{x}}_i - \boldsymbol{\theta}_i)' \hat{V}(\tilde{\mathbf{x}}_i)^{-1} (\tilde{\mathbf{x}}_i - \boldsymbol{\theta}_i) \leq qF_{q,\, n_i - q}(\alpha), \tag{3}$$

where $F_{q,\, n_i - q}(\alpha)$ is the upper $\alpha$ point of the $F$-distribution with $(q,\ n_i - q)$ df. By simulation in Rukhin [6], the coverage probabilities for (3) with $\bar{\mathbf{x}}_i$ are less than $1 - \alpha$ in many cases. But (3) with $\tilde{\mathbf{x}}_i$ using $W_{ij} = \hat{V}(\mathbf{x}_{ij})^{-1}$ is conservative. We modify the estimator $\hat{V}(\mathbf{x}_{ij})$ and extend to two sample problem. In the above estimating method, the estimator of error variance $\tilde{\sigma}_{ij}^2$ is changed into $\hat{\sigma}_{ij}^2 = v_{ij}/p_{ij}$. The estimator proposed by Rukhin [6] is denoted by $\tilde{\mathbf{x}}_i^R$, which is based on $\tilde{\sigma}_{ij}^2$. Let $\tilde{\mathbf{x}}_i^m$ be the modified estimator of $\boldsymbol{\theta}_i$, which is based on $\hat{\sigma}_{ij}^2$.

Next we give approximated confidence regions for $\boldsymbol{\delta} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$. By using Yao [9], see, e.g., Siotani et al. [7], the confidence region with confidence coefficient $100(1 - \alpha)\%$ can be approximated by

$$(\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2 - \boldsymbol{\delta})' U^{-1} (\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2 - \boldsymbol{\delta}) \leq \frac{qf_Y}{f_Y - q + 1} F_{q,\, f_Y}(\alpha), \tag{4}$$

where $U = \hat{V}(\tilde{\mathbf{x}}_1) + \hat{V}(\tilde{\mathbf{x}}_2)$,

$$\frac{1}{f_Y} = \sum_{i=1}^{2} \frac{1}{n_i - 1} \left( \frac{(\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2)' U^{-1} \hat{V}(\tilde{\mathbf{x}}_i) U^{-1} (\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2)}{T^{*2}} \right)$$

and $T^{*2} = (\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2)' U^{-1} (\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2)$.

By using Johansen [3], the approximated $1-\alpha$ confidence region is given by

$$(\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2 - \boldsymbol{\delta})' U^{-1} (\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2 - \boldsymbol{\delta}) \le \phi_J F_{q, f_J}(\alpha), \tag{5}$$

where

$$\phi_J = q + \frac{(q-1)(\hat{\psi}_1 + \hat{\psi}_2)}{(q+2)(n_1+n_2-2)}, \quad f_J = \frac{2q(q+2)(n_1+n_2-2)}{3(\hat{\psi}_1 + \hat{\psi}_2)},$$

$$\hat{\psi}_1 = \frac{n_1+n_2-2}{(n_1+n_2)^2}\left(\frac{n_2^2}{n_1-1}\{\mathrm{tr}(\hat{V}(\tilde{\mathbf{x}}_1)\overline{V}^{-1})\}^2 + \frac{n_1^2}{n_2-1}\{\mathrm{tr}(\hat{V}(\tilde{\mathbf{x}}_2)\overline{V}^{-1})\}^2\right),$$

$$\hat{\psi}_2 = \frac{n_1+n_2-2}{(n_1+n_2)^2}\left(\frac{n_2^2}{n_1-1}\mathrm{tr}(\hat{V}(\tilde{\mathbf{x}}_1)\overline{V}^{-1}\hat{V}(\tilde{\mathbf{x}}_1)\overline{V}^{-1}) \right.$$

$$\left. + \frac{n_1^2}{n_2-1}\mathrm{tr}(\hat{V}(\tilde{\mathbf{x}}_2)\overline{V}^{-1}\hat{V}(\tilde{\mathbf{x}}_2)\overline{V}^{-1})\right),$$

and $\overline{V} = (1/(n_1+n_2))\{n_2\hat{V}(\tilde{\mathbf{x}}_1) + n_1\hat{V}(\tilde{\mathbf{x}}_2)\}$.

Three approximate solutions to the multivariate Behrens-Fisher problem are given by Yanagihara and Yuan [8], in which the *F* approximation is recommended in simulation. Then the approximate confidence region of $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$ by the *F* approximation in Yanagihara and Yuan [8] is given by

$$(\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2 - \boldsymbol{\delta})' U^{-1} (\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2 - \boldsymbol{\delta}) \le \frac{q(n_1+n_2-2)}{n_1+n_2-\hat{\eta}_1} F_{q, f_{YY}}(\alpha), \tag{6}$$

where

$$f_{YY} = \frac{(n_1+n_2-\hat{\eta}_1)^2}{(n_1+n_2-2)\hat{\eta}_2 - \hat{\eta}_1}, \quad \hat{\eta}_1 = \frac{q\hat{\psi}_1 + (q-2)\hat{\psi}_2}{q(q+2)}, \quad \hat{\eta}_2 = \frac{\hat{\psi}_1 + 2\hat{\psi}_2}{q(q+2)}.$$

## 3. Simulation

In Section 2, we gave three approximated confidence regions of $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$. Here we examine the accuracy of approximation by using three types of estimators $\tilde{\mathbf{x}}_i^R$, $\tilde{\mathbf{x}}_i^m$, $\bar{\mathbf{x}}_i$. In simulation study, we choose $\alpha = 0.05$, $q = 2$, $r = 2$, $n_1 = 7$, $n_2 = 8$,

$$\boldsymbol{\theta}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{\theta}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Psi_1 = \begin{pmatrix} 0.50 & 0.05 \\ 0.05 & 0.15 \end{pmatrix}, \quad \Psi_2 = \begin{pmatrix} 0.75 & 0.15 \\ 0.15 & 0.03 \end{pmatrix},$$

and the error variances $\sigma_{ij}^2$ were taken to be different multiple $\kappa$ of $\chi^2$ distribution with 2 df, where $\kappa = 0.1(0.2)1.9$. Let $A_{ij} = I_2$ and $C'_{ij} = (C', \ldots, C') : q \times p_{ij}$, where $p_{ij} = jr \ (j = 1, \ldots, n_i)$ and

$$C' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2.0 & 2.5 & 3.0 & 5.0 & 7.5 \end{pmatrix}.$$

For these values, 10,000 confidence regions for each of (4), (5) and (6) are constructed. The proportion, that regions include the true values $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$, is calculated. The results are in Table 1, in which the boldface is the closest to 0.95 in each case.

**Table 1.** Coverage probability

| | (4) | | | (5) | | | (6) | | |
|---|---|---|---|---|---|---|---|---|---|
| $\kappa$ | $\tilde{\mathbf{x}}_i^R$ | $\tilde{\mathbf{x}}_i^m$ | $\bar{\mathbf{x}}_i$ | $\tilde{\mathbf{x}}_i^R$ | $\tilde{\mathbf{x}}_i^m$ | $\bar{\mathbf{x}}_i$ | $\tilde{\mathbf{x}}_i^R$ | $\tilde{\mathbf{x}}_i^m$ | $\bar{\mathbf{x}}_i$ |
| 0.1 | 0.9587 | 0.9539 | 0.9590 | 0.9546 | **0.9495** | 0.9558 | 0.9576 | 0.9534 | 0.9580 |
| 0.3 | 0.9647 | 0.9610 | 0.9629 | 0.9628 | **0.9584** | 0.9597 | 0.9658 | 0.9612 | 0.9623 |
| 0.5 | 0.9643 | 0.9587 | 0.9551 | 0.9624 | 0.9564 | **0.9525** | 0.9646 | 0.9586 | 0.9552 |
| 0.7 | 0.9692 | 0.9651 | 0.9599 | 0.9681 | 0.9634 | **0.9581** | 0.9701 | 0.9663 | 0.9610 |
| 0.9 | 0.9689 | 0.9621 | 0.9618 | 0.9664 | **0.9589** | 0.9592 | 0.9690 | 0.9622 | 0.9610 |
| 1.1 | 0.9743 | 0.9687 | 0.9616 | 0.9728 | 0.9677 | **0.9606** | 0.9737 | 0.9698 | 0.9630 |
| 1.3 | 0.9708 | 0.9659 | 0.9605 | 0.9698 | 0.9635 | **0.9595** | 0.9717 | 0.9660 | 0.9616 |
| 1.5 | 0.9710 | 0.9664 | 0.9608 | 0.9701 | 0.9656 | **0.9581** | 0.9719 | 0.9677 | 0.9614 |
| 1.7 | 0.9724 | 0.9672 | 0.9592 | 0.9712 | 0.9653 | **0.9565** | 0.9738 | 0.9673 | 0.9598 |
| 1.9 | 0.9747 | 0.9698 | 0.9623 | 0.9731 | 0.9670 | **0.9607** | 0.9748 | 0.9695 | 0.9633 |

In Table 1, almost all values are greater than 0.95, say the confidence regions are conservative. In two sample problem, Johansen's [3] approximation by using estimator $\tilde{\mathbf{x}}_i^m$ (for small error variances) and $\bar{\mathbf{x}}_i$ (for large error variances) would be better. Rukhin's estimator $\tilde{\mathbf{x}}_i^R$ may be better in mean square error, but is not adequate for interval estimation, see Rukhin [6].

## Acknowledgement

This work was supported by JSPS KAKENHI Grant number 2540206.

## References

[1] M. Davidian and D. M. Giltinan, Nonlinear Models for Repeated Measurement Data, Chapman & Hall/CRC, 1995.

[2] P. J. Diggle, P. Heagerty, K. Y. Liang and S. L. Zeger, Analysis of Longitudinal Data, 2nd ed., Oxford University Press, 2002.

[3] S. Johansen, The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression, Biometrika 67 (1980), 85-92.

[4] N. M. Laird, N. Lange and D. Stream, Maximum likelihood computations with repeated measures, J. Amer. Statist. Assoc. 82 (1987), 97-105.

[5] N. M. Laird and J. H. Ware, Random effects models for longitudinal data, Biometrics 38 (1982), 963-974.

[6] A. L. Rukhin, Estimating common parameters in heterogeneous random effects model, J. Statist. Plann. Inference 141 (2011), 3181-3192.

[7] M. Siotani, T. Hayakawa and Y. Fujikoshi, Modern Multivariate Statistical Analysis: A Graduate Course and Handbook, American Sciences Press, 1985.

[8] H. Yanagihara and K. Yuan, Three approximate solutions to the multivariate Behrens-Fisher problem, Comm. Statist. Simulation Comput. 34 (2005), 975-988.

[9] Y. Yao, Approximate degrees of freedom solution to the multivariate Behrens-Fisher problem, Biometrika 52 (1965), 139-147.