한국통계학회논문집 제2권 2호
1995년 12월 pp. 296-302

# Constant Error Variance Assumption in Random Effects Linear Model[1)]

Chul Hwan Ahn[2)]

## Abstract

When heteroscedasticity occurs in random effects linear model, the error variance may depend on the values of one or more of the explanatory variables or on other relevant quantities such as time or spatial ordering. In this paper we derive a score test as a diagnostic tool for detecting non-constant error variance in random effects linear model based on the model expansion on error variance. This score test is compared to loglikelihood ratio test.

## 1. Introduction

Diagnostics in general are used to decide if assumptions made in fitting a model are appropriate. Box(1980) suggests that these methods should be done conditionally given the fitted model, so any principles to be developed for them are likely to be the same for both a Bayesian and a frequentist. How do we check the appropriateness of the assumption given the fitted model? Several diagnostic tools have been developed for this purpose. In linear regression Cook and Weisberg(1983) proposed a diagnostic test for examining the assumption of nonconstant variance. Their idea is to convert the constant variance assumption to a testable parametric hypothesis. In one way random effects model Dempster and Ryan(1985) proposed weighted normal plots as graphical checks on the normality of random effects. Their weighted normal plots involve a modification that gives the i-th observation a sample dependent weight. They showed that weighted normal plots are more sensitive than unweighted plots to several departures from the assumed distribution on the random effects. In mixed model analysis of variance Cook, Beckman and Nachtsheim(1987) applied the idea of local influence (Cook, 1986) to assessing the effects of perturbation from the usual assumption of constant error variance and from the assumption that each realization of a given random factor has been drawn from the same normal population. In the area of generalized linear model, Smyth(1989) considered the case for which the dispersion submodel is a gamma generalized linear model with log-link. The score test statistic in his paper is essentially of

1) Department of Applied Statistics, Sejong University, Seoul 133-747, KOREA.
2) This paper was supported in part by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1993.

the same form as the one in this study - half the sum of squares due to linear regression from the constructed model.

In this study we consider methods for examining the assumption of homoscedasticity in a special case of the random effects linear model, given by

$$Y_{ij} = X_{ij}^T \underline{\beta} + A_i + \varepsilon_{ij} \tag{1}$$

$A_i$ is NID$(0, \sigma_A^2)$, $\varepsilon_{ij}$ is NID$(0, \sigma^2)$, and A's and $\varepsilon$'s are independen

We now question the assumption on constant error variance. When heteroscedasticity occurs the error variance may depend on the values of one or more of the explanatory variables or on other relevant quantities such as time or spatial ordering. There will be several methods to seek diagnostics concerning heteroscedasticity but a standard method of deriving diagnostics will be through the technique of model expansion, in which a model like (1) is embedded in a large class and a relatively simple procedure like a score test (Chen, 1983) is then used to turn a test into a diagnostic. For model (1), useful expansion can be obtained by setting

$$\text{var}(\varepsilon_{ij}) = \sigma^2 \exp(\underline{\lambda}^T \underline{Z}_i) \tag{2}$$

where $\underline{\lambda}$ is an unknown parameter vector and $\underline{Z}_i$ is the value of a vector of covariates for group i, which can be related to the $X_{ij}, E(Y_{ij})$, or other relevant items. This type of model expansion was considered for a diagnostics of homoscedasticity in linear regression by Cook and Weisberg(1983), and for a diagnostics of non-constant variance on random effects in mixed linear model by Ahn(1990). The form (2) implies that the variance depends on $\underline{Z}_i$ and $\underline{\lambda}$ but only through $\underline{\lambda}^T \underline{Z}_i$, and $\text{var}(\varepsilon_{ij}) = \sigma^2$ when $\underline{\lambda} = \underline{0}$. Since the variance is a monotonic function of $\underline{\lambda}^T \underline{Z}_i$, we can think of this method as specifying a direction $\underline{\lambda}^T \underline{Z}_i$ such that the variance increases in that direction. In the following section, the score test for $\underline{\lambda} = \underline{0}$ (that is, for homoscedasticity) in the expanded model (1) is derived. Section 3 reports a simulation study to investigate the $\chi^2$ approximation to the null distribution of the score statistic. It is also compared to loglikelihood ratio test.

## 2. Score test

The model (1) can be written in matrix form as follows:

$$\underline{Y} = X \underline{\beta} + R \underline{A} + \underline{\varepsilon} \tag{3}$$

where $\underline{Y}$ is an Nx1 response vector with N = $n_1 + n_2 + \cdots + n_t$, and $n_i$ equal to the size of group i, $X$ is an Nxp fixed explanatory variables of matrix, $\underline{\beta}$ is a px1 unknown parameter vector, $R=(\underline{R}_1\ \underline{R}_2 \cdots \underline{R}_t)$ where $\underline{R}_i$ is an Nx1 indicator vector, with $n_i$'s of 1 for group i and 0 elsewhere, $\underline{A}=(A_1\ A_2 \cdots A_t)$where $A_i$ denotes random effect for group i with $\underline{A} \sim N(0,\sigma_A^2 I)$, and $\underline{\varepsilon}$ is an Nx1 error vector with $\underline{\varepsilon} \sim N(0,\sigma^2 D)$ where $D$ is a diagonal matrix with i-th diagonal element equal to $d_i = \exp(\underline{\lambda}^T \underline{Z}_i)$ with $\underline{Z}_i$, a qx1 fixed and known vector and $\underline{\lambda}$, a qx1 unknown parameter vector. And, we let $\xi = \sigma_A^2/\sigma^2$.

Considering the balanced case with $n_1, n_2, \cdots, n_t = n$, we get the following result.

Theorem 1 Let $\underline{U}$ be a tx1 vector with elements

$$U_i = \frac{1}{\hat{\sigma}^2}\left(n - \frac{n\hat{\xi}}{1+n\hat{\xi}} - \frac{n\hat{\xi}}{(1+n\hat{\xi})^2}\right)^{-1/2} \sum_{j=1}^{n}\left(e_{ij} - \frac{n\hat{\xi}}{1+n}\bar{e}_{i.}\right)^2 \quad (4)$$

where $\bar{e}_{i.} = \sum_{j=1}^{n}(y_{ij} - x_{ij}^T\hat{\underline{\beta}})/n$ is the average residual in group i and $\hat{\underline{\beta}}$, $\hat{\sigma}^2$ and $\hat{\xi}$ are maximum likelihood estimates of $\underline{\beta}$, $\sigma^2$, and $\xi$, respectively under the model (1) with $\underline{\lambda} = \underline{0}$. Let $Z$ denote a txq matrix with its i-th row $\underline{Z}_i^T$, a lxq vector of covariates for group i. Let $\overline{Z} = Z - \underline{1}\,\underline{1}^T Z / t$ with $\underline{1}$, a tx1 vector of ones. $\overline{Z}$ is a txq matrix obtained from $Z$ by subtracting column averages. Then, the score test statistic S for NH: $\lambda = 0$ vs AH: $\underline{\lambda} \neq \underline{0}$ with $\underline{\theta} = (\underline{\beta}^T, \sigma^2, \xi)^T$as a nuisance parameter vector is

$$S = \frac{1}{2}\underline{U}^T \overline{Z}(\overline{Z}^T\overline{Z})^{-1}\overline{Z}^T\underline{U} \quad . \quad (5)$$

Computationally, $S$ is one half of the regression sum of squares from the constructed model $\underline{U} = \lambda_o \underline{1} + Z\underline{\lambda} + \underline{\varepsilon}_u$. Asymptotically as $t \to \infty$, under NH, $S \sim \chi^2(q)$.

proof: The covariance matrix of $\underline{Y}$ can be written as follows.

$$Cov(\underline{Y}) = \sum_{i=1}^{t}\mathrm{var}(\varepsilon_{ij})\,G_i + \sigma_A^2 R R^T = \sigma^2\left(\sum_{i=1}^{t}\exp(\underline{\lambda}^T\underline{Z}_i)\,G_i + \xi R R^T\right) \quad (6)$$

where $G_i$ is an NxN diagonal matrix with one in the i-th group and zero elsewhere.

By letting $Q_\lambda = \sum_{i=1}^{t}\exp(\underline{\lambda}^T\underline{Z}_i)\,G_i + \xi R R^T$, we can write $Cov(\underline{Y}) = \sigma^2 Q_\lambda$.

Note that when $\underline{\lambda} = \underline{0}$, $Q_o = I + \xi R R^T$, and var(y|x) $= \sigma^2 Q_o$. In the sequel, we write $Q$ for $Q_o$. $Q_\lambda$ is a function of $\underline{\lambda}$ and $\xi$, and $Q$ is a fundtion of $\xi$ only. Let $\underline{\theta} = (\underline{\beta}^T, \sigma^2, \xi)^T$ denote a parameter vector and $L(\underline{\theta},\underline{\lambda})$ the log-likelihood under the expanded model. Then

$$L(\underline{\theta},\underline{\lambda}) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log\sigma^2 - \frac{1}{2}\log|Q_\lambda| - \frac{1}{2\sigma^2}(\underline{Y}-X\underline{\beta})^T Q_\lambda^{-1}(\underline{Y}-X\underline{\beta}) \quad (7)$$

The information matrix $I$ of ( $\underline{\theta},\underline{\lambda}$ ) can then be partitioned as

$$I = \begin{Bmatrix} i_{\theta\theta} & i_{\theta\lambda} \\ i_{\lambda\theta} & i_{\lambda\lambda} \end{Bmatrix}, \quad \text{where} \quad i_{\theta\lambda} = -E\frac{\partial^2 L(\theta,\lambda)}{\partial\theta\ \partial\lambda^T}. \tag{8}$$

The total score statistic for $\underline{\lambda}$ evaluated under the null hypothesis $\underline{\lambda} = \underline{0}$ is $\underline{V} = \frac{\partial L}{\partial\lambda}$ evaluated at $\underline{\lambda}=\underline{0}$, $\underline{\theta} = \hat{\underline{\theta}}_o$ where $\hat{\underline{\theta}}_o = ( \hat{\underline{\beta}}^T, \widehat{\sigma^2}, \hat{\xi} )^T$ given $\lambda$=0, that is, $\hat{\underline{\theta}}_o$ is the m.l.e of $\underline{\theta}$ under the null hypothesis. And its asymptotic covariance matrix is estimated by $C=i_{\lambda\lambda}( \hat{\underline{\theta}}_o,\underline{\lambda}_o ) - i_{\lambda\theta} ( \hat{\underline{\theta}}_o,\underline{\lambda}_o )i_{\theta\theta}( \hat{\underline{\theta}}_o,\underline{\lambda}_o )^{-1}i_{\theta\lambda}( \hat{\underline{\theta}}_o,\underline{\lambda}_o )$ or simply $\widehat{i_{\lambda\lambda}} - \widehat{i_{\lambda\theta}}\ \widehat{i_{\theta\theta}^{-1}}\ \widehat{i_{\theta\lambda}}$. Then the score test statistic S for testing $\underline{\lambda} = \underline{0}$ can be expressed in terms of the quadratic form as $S = \underline{V}^T C^{-1}\underline{V}$. It is well known from Cox and Hinkley (1974) that the limiting distribution of S is chi-squared with degrees of freedom equal to dim($\lambda$) and is central under $H_0$. We now evaluate $\underline{V}$ and $C$. First,

$$\frac{\partial L}{\partial\lambda} = -\frac{1}{2}\frac{\partial}{\partial\lambda}\log|Q_\lambda| - \frac{1}{2\sigma^2}(\underline{Y}-X\underline{\beta})^T\frac{\partial}{\partial\lambda}Q_\lambda^{-1}(\underline{Y}-X\underline{\beta}) \tag{9}$$

The following two results for matrix derivatives from Rogers(1980) will be used to evaluate the first and second partial derivatives of L.

$$1.\ \frac{\partial}{\partial\lambda}\log|Q_\lambda| = TR( Q_\lambda^{-1}\frac{\partial Q_\lambda}{\partial\lambda}) \qquad 2.\ \frac{\partial}{\partial\lambda}Q_\lambda^{-1} = -Q_\lambda^{-1}\frac{\partial Q_\lambda}{\partial\lambda}Q_\lambda^{-1}$$

Let $\frac{\partial L}{\partial\lambda_k}$ be the k-th element of $\frac{\partial L}{\partial\lambda}$. Then,

$$\left.\frac{\partial L}{\partial\lambda_k}\right|_{\lambda=o} = -\frac{1}{2}TR(Q^{-1}A_k) + \frac{1}{2\sigma^2}(\underline{Y}-X\underline{\beta})^T Q^{-1}A_k Q^{-1}(\underline{Y}-X\underline{\beta}) \tag{10}$$

where $A_k=\sum_{i=1}^{t} z_i^k G_i$. The $z_i^k$ is the k th value in the vector $\underline{Z}_i$. Let $\phi = \frac{n}{1+n\xi}$ and $\gamma = n - \xi\phi - \frac{\xi\phi^2}{n}$. Then $\underline{V}$ is simplified as $\underline{V} = \frac{1}{2}\gamma^{\frac{1}{2}}\underline{Z}^T\underline{U}$ where $\underline{U}$ is a t x 1 vector with its i th elemen

$$u_i = \left.\frac{\sum_{i=1}^{t}(e_{ij} - \xi\phi\overline{e}_{i.})^2}{\gamma^{\frac{1}{2}}\sigma^2}\right|_{\theta_o} \tag{11}$$

We turn to evaluation of $C=i_{\lambda\lambda}( \hat{\underline{\theta}}_o,\underline{\lambda}_o ) - i_{\lambda\theta} ( \hat{\underline{\theta}}_o,\underline{\lambda}_o )i_{\theta\theta}( \hat{\underline{\theta}}_o,\underline{\lambda}_o )^{-1}i_{\theta\lambda}( \hat{\underline{\theta}}_o,\underline{\lambda}_o )$.

The following identities can be shown under the null hypothesis:

a) $i_{\lambda\lambda}(\underline{\widehat{\theta_o}}, \underline{\lambda_o}) = \frac{1}{2}\gamma\, Z^T Z$

b) $i_{\lambda\theta}(\underline{\widehat{\theta_o}}, \underline{\lambda_o}) = \left(0 \quad \frac{1}{2\sigma^2}\sum_{i=1}^{t}(n - \xi\phi)\underline{Z}_i \quad \frac{1}{2}\sum_{i=1}^{t}\phi^2\underline{Z}_i\right)$

c)

$$i_{\theta\theta}(\underline{\widehat{\theta_o}}, \underline{\lambda_o}) = \begin{bmatrix} \frac{x^T Q^{-1} x}{\sigma^4} & 0 & 0 \\ 0 & \frac{N}{2\sigma^4} & \frac{t}{2\sigma^2}\phi \\ 0 & \frac{t}{2\sigma^2}\phi & \frac{t}{2}\phi^2 \end{bmatrix}$$

Now, $\widehat{i_{\lambda\theta}}\ \widehat{i_{\theta\theta}^{-1}}\ \widehat{i_{\theta\lambda}} = \gamma\, E^T E$, where $E = \frac{1}{t}\underline{1}\ \underline{1}^T Z$. Therefore,

$$C = \gamma(Z^T Z - E^T E) = \gamma(Z^T Z - \frac{1}{t}Z^T \underline{1}\ \underline{1}^T Z) = \gamma\, \overline{Z}^T \overline{Z}, \quad \text{where } \overline{Z} = Z - \frac{1}{t}\underline{1}\ \underline{1}^T Z$$

Combining these we obtain in the balanced case the score test statistic S ;

$$S = \frac{1}{2}\underline{U}^T \overline{Z}\, (\overline{Z}^T \overline{Z})^{-1}\, \overline{Z}^T \underline{U} \tag{12}$$

Computationally, S is one half of the regression sum of squares from the constructed model, $\underline{U} = \lambda_o \underline{1} + Z\underline{\lambda} + \underline{\varepsilon}_u$ where $\lambda_o$ denotes a parameter for intercept, $\underline{1}$, a column vector of ones, and $\underline{\varepsilon}_u$, an error vector.

## 3. Simulation results

A simulation study was conducted in the balanced case to investigate the chi-squared approximation to the null distribution of the score tests. It was expected that as the number of group increases the levels of the test will be close to those of appropriate Chi-square distribution. It was also compared to the loglikelihood ratio test. The score test statistic S is obtained from the Taylor expansion of the loglikelihood ratio statistic (we'll call it W) up to the second order. It is well known that W and S are asymptotically equivalent (Cox and Hinkley, 1974). An advantage of S over W is that the maximum likelihood estimate of $\lambda$ need not be calculated. The responses $y_{ij}$ are generated from $y_{ij} = x_{ij}^T\beta + A_i + \varepsilon_{ij}$. An 80x3 matrix with all entries generated from standard normal distribution is prepared for the matrix X and a 3x1 vector $(5\ 10\ 15\ )^T$ for the parameter $\beta$. The two factors were varied in the simulation;

the number of of groups (t) : 5, 10, 20, and the size of dimension of $\beta$ (p) : 1, 3. And, The six combinations of (t,p) are considered - (5,1), (10,1), (20,1), (5,3), (10,3), and (20,3). Each pair of (t,p) specifies the matrix X. For example, when t=10 and p=3, we use the 40x3 submatix in the top left-hand corner of the 80x3 matrix described above since X is N by p where N is t times n. The number of observations per group, n was fixed as 4.

The covariates $z_{ij}$ are set equal to $x_{ij}$ and so, q=p. The $A_i$'s and $\varepsilon_{ij}$'s are generated as independent, standard normal deviates. Normal deviates are produced from the uniform stream. For each of the 499 replications, S and W are computed. Table 1 gives the 0.90, 0.95, 0.975 and 0.99 points of the sample distributions of S and W. The nominal values from the appropriate Chi-square distributions are also given.

From Table 1, we can see that as the number of groups, t increases, the percntage points of both S and W become closer to the corresponding Chi-square nominal values. The values in parenthesis are those for W. As t increases, S approaches to Chi-square nominal values from below and W from above. The Chi-squared approximation to S seems appropriate for large t. However, W is not close to Chi-square nominal values for t=20 and p=3.

Table 1. Simulated percentage points from small-sample null distribution of S and W

| P | Level | t=5 | t=10 | t=20 | $\chi^2$ |
|---|---|---|---|---|---|
| 1 | 0.90 | 2.42 (3.51) | 2.42 (2.98) | 2.30 (2.54) | 2.71 |
| | 0.95 | 3.12 (5.46) | 4.14 (4.51) | 3.94 (3.98) | 3.84 |
| | 0.975 | 4.06 (6.30) | 4.96 (5.75) | 4.78 (5.26) | 5.02 |
| | 0.99 | 5.34 (8.32) | 5.63 (6.55) | 6.51 (6.11) | 6.63 |
| 3 | 0.90 | 5.14 (10.6) | 5.48 (7.88) | 6.31 (7.31) | 6.25 |
| | 0.95 | 6.50 (12.9) | 7.59 (9.88) | 8.12 (9.20) | 7.81 |
| | 0.975 | 9.65 (15.6) | 9.19 (11.7) | 9.43 (11.5) | 9.35 |
| | 0.99 | 11.2 (19.2) | 12.6 (16.9) | 11.1 (12.8) | 11.3 |

## 4. Conclusion

It was seen in Section 3 that as the number of groups increases the score statistic approaches to Chi-square nominal values. It can be concluded that the Chi-square approximation to the score statistic is appropriate for large t. In the problem of diagnostics for nonconstant variances in random effect linear model the score test developed in this study may be very useful. There are a number of open research questions concerning diagnostics in random effects linear model. It may include the score test for checking the various assumption, for example, checking normality (Hinkley, 1985) on the variance components in mixed linear model, and the improvement of the Chi-squared approximation to the null distribution of score statistic. It is always recommended to use both the test and the graphical method in diagnostics. Graphical equivalents to score test may be found similar to the one proposed by Cook and Weisberg (1983).

## Acknowledgment

The author is grateful to a professor Sanford Weisberg at the University of Minnesota for his extensive comments that improved this paper.

## References

[1] Ahn, Chul H. (1990). Diagnostics for Heteroscedasticity in Mixed Linear Models, *Journal of the Korean Statistical Society*, Vol 19, No 2, 171-175.

[2] Box, G.E.P. (1980). Sampling and Bayes inference in scientific modelling and robustness(with discussion), Journal of Royal Statistical Society, Series A, 143, 1-35.

[3] Chen, C.(1983). Score tests for regression models, *Journal of the American Statistical Association*, Vol. 78, 158-161.

[4] Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman & Hall, London.

[5] Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in Regression, *Biometrika*, Vol. 70, 1-10.

[6] Cook, R. D. (1986). Assessment of Local Influence(with discussion), *Journal of Royal Statistical Society*, Series B, 48, 133-169.

[7] Cook, R. D., Beckman R. and Nachtsheim, C. (1987). Diagnostics for Mixed-Model Analysis of Variance, *Technometrics*, Vol. 29, 413-426.

[8] Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman & Hall, London.

[9] Dempster, A. P. and Ryan, L. M. (1985). Weighted normal plot, *Journal of the American Statistical Association*, Vol. 80, 845-850.

[10] Graham, A. (1981). *Kronecker Products and Matrix Calculus with Application*, Ellis Horwood Ltd., England.

[11] Hinkley, D. V. (1985). Transformation diagnostics for linear model, *Biometrika*, Vol. 72, 487-496.

[12] Hocking, R. R. (1984). Diagnostic methods in variance component estimation, *Proceedings of International Biometrics Conference*, Tokyo, Japan.

[13] Rogers, G. S. (1980). *Matrix Derivatives*, Marcel Dekker, New York.

[14] Smyth, Gordon (1989). Generalized linear models with varying dispersion, *Journal of Royal Statistical Society*, Series B, 51, 47-60.