CIRJE-F-559

On Testing Linear Hypothesis in a Nested Error Regression Model

Tatsuya Kubokawa
Nyambaa Erdembat
University of Tokyo

April 2008

CIRJE Discussion Papers can be downloaded without charge from:

http://www.e.u-tokyo.ac.jp/cirje/research/03research02dp.html

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

On Testing Linear Hypothesis in a Nested Error Regression Model

Tatsuya Kubokawa and Nyambaa Erdembat
University of Tokyo

April 6, 2008

Abstract

Consider the problem of testing the linear hypothesis on regression coefficients in the nested error regression model. The standard F-test statistic based on the ordinary least squares (OLS) estimator has the serious shortcoming that its type I error rates (sizes) are much larger than nominal significance levels, because the covariance matrix of data is not the identity but has the intraclass correlation structure. One of methods for fixing the problem is to consider an F-test statistic based on the generalized least squares (GLS) estimator, and the resulting GLS F-test performs well in controlling the sizes. However, numerical investigations show that the sizes remain still slightly larger than nominal levels. In this paper, we derive two test procedures: One is an exact test based on the within analysis of variance, and the other is a testing procedure based on the asymptotic correction of the GLS method. It is numerically shown that both procedures are superior to the GLS F-test in controlling the sizes and that the latter test is more powerful than the exact test.

Key word and Phrases: asymptotic correction, generalized least squares method, intraclass correlation, F-test statistic, linear mixed model, nested error regression model.

1 Introduction

Consider the nested error regression model which have been used in the small area estimation. Let k be the number of small areas and let n_i be size of a sample from the i-th small area. Then the model is described as

$$y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \ldots, k, \ j = 1, \ldots, n_i, \tag{1.1}$$

where y_{ij} is a scalar observation and $\boldsymbol{x}_{ij}$ is a p-dimensional vector of the corresponding covariates, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$ is an unknown vector of the regression coefficients, v_i is a random effect representing an area effect and e_{ij} is an error term. It is assumed that these random variables are mutually independent and

$$v_i \sim \mathcal{N}(0, \sigma_v^2), \quad e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$$

where σ_v^2 and σ_e^2 are unknown and called 'between' and 'within' components of variance, respectively. Let $N = \sum_{i=1}^k n_i$ and assume that $N > k + p$ and $k \geq p$. Letting $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{i,n_i})'$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{i,n_i})'$, $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{i,n_i})'$ and $\boldsymbol{j}_{n_i} = (1, \ldots, 1)' \in \boldsymbol{R}^{n_i}$, we can rewrite (1.1) as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{j}_{n_i}v_i + \boldsymbol{e}_i, \quad i = 1, \ldots, k, \tag{1.2}$$

and the covariance matrix of $\boldsymbol{y}_i$ is

$$\mathbf{Cov}\,(\boldsymbol{y}_i) = \sigma_e^2\boldsymbol{V}_i(\psi) = \sigma_e^2\left\{\boldsymbol{I}_{n_i} + \psi\boldsymbol{J}_{n_i}\right\},$$

where $\boldsymbol{I}_{n_i}$ is the $n_i \times n_i$ identity matrix, $\boldsymbol{J}_{n_i} = \boldsymbol{j}_{n_i}\boldsymbol{j}_{n_i}'$, and $\psi = \sigma_v^2/\sigma_e^2$. It is noted that the covariance of $\boldsymbol{y}_i$ has the intra-class correlation structure, namely $y_{i1}, \ldots, y_{i,n_i}$ are not mutually independent when $\psi \neq 0$. Letting $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_k')'$, $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_k')'$, $\boldsymbol{v} = (v_1, \ldots, v_k)'$ and $\boldsymbol{e} = (\boldsymbol{e}_1', \ldots, \boldsymbol{e}_k')'$, we express (1.1) in the matrix notation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{v} + \boldsymbol{e},$$

where $\boldsymbol{Z} =$ block diag$(\boldsymbol{j}_{n_1}, \ldots, \boldsymbol{j}_{n_k})$, a block diagonal matrix. It is assumed that $\boldsymbol{X}$ is of full rank. The covariance matrix of $\boldsymbol{y}$ is given by

$$\mathbf{Cov}\,(\boldsymbol{y}) = \sigma_e^2\boldsymbol{V}(\psi) = \sigma_e^2 \cdot \text{block diag}(\boldsymbol{V}_1(\psi), \ldots, \boldsymbol{V}_k(\psi)).$$

In this paper, we consider the problem of testing the following linear hypothesis on the regression coefficients $\boldsymbol{\beta}$:

$$H_0 \;:\; \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{b}$$

against $H_1 : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{b}$, where $\boldsymbol{C}$ is a $q \times p$ $(q \leq p)$ known matrix with rank q and a $p \times 1$ known vector $\boldsymbol{b}$. The F-test statistic based on the ordinary least squares (OLS) estimator $\widehat{\boldsymbol{\beta}}_0 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is given by

$$F_{OLS} = \frac{(\boldsymbol{C}\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{b})'(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{b})/q}{(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0)'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0)/(N-p)}, \tag{1.3}$$

where $\boldsymbol{X}_c = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'$ and $(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1} = (\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}')^{-1}$. The OLS F test F_{OLS} is a standard procedure when $\psi = 0$, while it has the serious drawback of having inflated type I error (size) when ψ is away from zero. To fix this problem, two procedures have been proposed in the literature: One is from Wu, Holt and Holmes (1988) and the other from Rao, Sutradhar and Yue (1993). The expectations of the numerator and denominator of F_{OLS} are $E[(\boldsymbol{C}\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{b})'(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{b})] = \sigma_e^2\text{tr}\,\boldsymbol{P}_c\boldsymbol{V}(\psi)$ and $E[(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0)'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0)] = \sigma_e^2((1+\psi)N - \text{tr}\,\boldsymbol{P}\boldsymbol{V}(\psi))$, respectively, for

$$\boldsymbol{P}_c = \boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c' = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}')^{-1}\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}',$$

and $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Wu *et al.* (1988) proposed the method of exchanging the degrees of freedom q and $N - p$ in F_{OLS} with the above expectations, namely,

$$F_{WHH}(\psi) = \frac{(\boldsymbol{C}\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{b})'(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{b})/\text{tr}\,\boldsymbol{P}_c\boldsymbol{V}(\psi)}{(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0)'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0)/\{(1+\psi)N - \text{tr}\,\boldsymbol{P}\boldsymbol{V}(\psi)\}}. \tag{1.4}$$

On the other hand, Rao *et al.* (1993) proposed an F-type test statistic based on the generalized least squares (GLS) estimator $\widetilde{\boldsymbol{\beta}}(\psi) = (\boldsymbol{X}'\boldsymbol{V}(\psi)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}(\psi)^{-1}\boldsymbol{y}$, and their test is given by

$$F_{RSY}(\psi) = \frac{(\boldsymbol{C}\widetilde{\boldsymbol{\beta}}(\psi) - \boldsymbol{b})'\{\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{V}(\psi)^{-1}\boldsymbol{X})^{-1}\boldsymbol{C}'\}^{-1}(\boldsymbol{C}\widetilde{\boldsymbol{\beta}}(\psi) - \boldsymbol{b})/q}{(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\psi))'\boldsymbol{V}(\psi)^{-1}(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\psi))/(N-p)}. \tag{1.5}$$

Since ψ is unknown, an estimator $\widehat{\psi}$ is substituted to get the test statistics $F_{WHH}(\widehat{\psi})$ and $F_{RSY}(\widehat{\psi})$. Rao and Wang (1995) proved that the power function of $F_{RSY}(\widehat{\psi})$ increases in ψ. It has been numerically shown that these test procedures $F_{WHH}(\widehat{\psi})$ and $F_{RSY}(\widehat{\psi})$ have sizes much smaller than F_{OLS}, epecially F_{RSY} performs well in cotrolling the sizes. As shown in Section 3, however, the sizes of F_{RSY} remain still slightly larger than significance levels.

In this paper, we try to derive test statistics with further improvements in controlling the sizes. To this end, we consider the two testing procedures: One is an exact test, denoted by F_{EXT}, which is constructed based on the 'within' analysis of variance, and the other is the test, denoted by F_{ACG}, which is constructed based on the asymptotic correction of the GLS test statistic. These test statistics are derived in Section 2. In Section 3, we investigate the size performances of all the test statistics treated in this paper through simulation studies, and show that the tests F_{EXT} and F_{ACG} have sizes close to nominal levels. Also the powers of the tests F_{RSY}, F_{EXT} and F_{ACG} are compared, and it is shown that F_{RSY} and F_{ACG} are more powerful than F_{EXT}. Combining the numerical properties of the sizes and the powers suggests the use of the asymptotic corrected GLS test statistic F_{ACG}. The proof of the main result is given in Section 4.

2 Main results

2.1 An exact test statistic

We first derive an exact test for the linear hypothesis $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{b}$. The exact test can be obtained from the 'within' part of analysis of variance in the likelihood function. Letting $f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma_v^2, \sigma_e^2)$ be a marginal density of $\boldsymbol{y}$, we can decompose the likelihood as

$$-2\log f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma_v^2, \sigma_e^2) = g_1(\boldsymbol{y}|\sigma_e^2, \boldsymbol{\beta}) + g_2(\overline{\boldsymbol{y}}|\sigma_e^2, \boldsymbol{\beta}, \psi), \tag{2.1}$$

where for $\gamma_i = \gamma_i(\psi) = 1/(1 + n_i\psi)$ and $\overline{\boldsymbol{y}} = (\overline{y}_1, \ldots, \overline{y}_k)'$,

$$g_1(\boldsymbol{y}|\sigma_e^2, \boldsymbol{\beta}) = (N-k)\log(2\pi\sigma_e^2) + \frac{1}{\sigma_e^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{(y_{ij} - \overline{y}_i) - (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)'\boldsymbol{\beta}\}^2, g_2(\overline{\boldsymbol{y}}|\sigma_e^2, \boldsymbol{\beta}, \psi) = \quad k\log(2\pi\sigma_e^2) -$$

which correspond to the 'within' and 'between' parts of analysis of variance, respectively.

Since $g_1(\boldsymbol{y}|\sigma_e^2, \boldsymbol{\beta})$ does not depend on ψ, we can construct an exact test. Let

$$\widehat{\boldsymbol{\beta}}_1 = \boldsymbol{B}^+ \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)(y_{ij} - \overline{y}_i),$$

$$S_1 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left\{(y_{ij} - \overline{y}_i) - (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}_1\right\}^2,$$

where $\boldsymbol{B} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)'$, and $\boldsymbol{B}^+$ denotes the Moore-Penrose generalized inverse matrix of $\boldsymbol{B}$. Assume that the rank of $\boldsymbol{B}$ is $p - \lambda$, and an unbiased estimator of σ_e^2 is given by

$$\hat{\sigma}_e^{2U} = S_1/(N - k - p + \lambda).$$

The statistic for testing $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{b}$ is

$$F_{EXT} = (\boldsymbol{C}\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{b})'(\boldsymbol{C}\boldsymbol{B}^+\boldsymbol{C}')^-(\boldsymbol{C}\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{b})/\mathrm{rank}\,(\boldsymbol{C}\boldsymbol{B}^+\boldsymbol{C}')/\hat{\sigma}_e^{2U}, \tag{2.2}$$

which, under H_0, has an F-distribution with $(\mathrm{rank}\,(\boldsymbol{C}\boldsymbol{B}^+\boldsymbol{C}'), N - k - p + \lambda)$ degrees of freedom. This means that F_{EXT} is an exact test.

2.2 Asymptotically corrected GLS test procedure

Although F_{EXT} is an exact test, the power gets worse when $\boldsymbol{C}\boldsymbol{B}^+\boldsymbol{C}'$ is not of full rank. It is also noted that the estimator $\widehat{\boldsymbol{\beta}}_1$ used in F_{EXT} is based on only the 'within' analysis of variance, which means that $\widehat{\boldsymbol{\beta}}_1$ is less efficient than the GLS estimator $\widetilde{\boldsymbol{\beta}}(\psi)$ given by

$$\widetilde{\boldsymbol{\beta}}(\psi) = (\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}\left\{\boldsymbol{B}\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{A}(\psi)\widetilde{\boldsymbol{\beta}}_2(\psi)\right\},$$

where $\boldsymbol{A}(\psi) = \sum_{i=1}^{k} n_i\gamma_i\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'$ and $\widetilde{\boldsymbol{\beta}}_2(\psi) = \boldsymbol{A}(\psi)^{-1}\sum_{i=1}^{k} n_i\gamma_i\overline{\boldsymbol{x}}_i\overline{y}_i$. It is assumed that $\boldsymbol{A}(\psi)$ is of full rank. Since $\widetilde{\boldsymbol{\beta}}(\psi)$ is more efficient, it is reasonable to consider a test statistic constructed based on the GLS $\widetilde{\boldsymbol{\beta}}(\psi)$.

To derive a test statistic based on the GLS estimator, we need to estimate ψ since it is unknown. For this purpose, we use the Henderson method III to get estimators of the variance component σ_v^2 and the ratio $\psi = \sigma_v^2/\sigma_e^2$. For $\widehat{\boldsymbol{\beta}}_0 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, let $S = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0)'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0)$ and $N_* = N - \mathrm{tr}\,\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\sum_{i=1}^{k} n_i^2\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'\}$. Then, an unbiased estimator of σ_v^2 is given by

$$\hat{\sigma}_v^{2U} = \left\{S - (N - p)\hat{\sigma}_e^{2U}\right\}/N_*,$$

and ψ is estimated by

$$\widehat{\psi}^U = \frac{\hat{\sigma}_v^{2U}}{\hat{\sigma}_e^{2U}} = \frac{1}{N_*}\left\{\frac{S}{\hat{\sigma}_e^{2U}} - (N - p)\right\}. \tag{2.3}$$

Since it takes negative values with a positive probability, it is reasonable to truncate $\widehat{\psi}^U$ at zero as

$$\widehat{\psi}_0 = \max\left\{\widehat{\psi}^U, 0\right\}. \tag{2.4}$$

It can be shown that for $a \geq 1/2$, $\widehat{\psi}_0 = \widehat{\psi}^U + o_p(k^{-a})$ (see Kubokawa and Srivastava (2007)). Then we can use the GLS estimator $\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0)$ substituting the truncated estimator $\widehat{\psi}_0$. Based on $\widehat{\psi}_0$, Rao *et* (1993) proposed the GLS F-test $F_{RSY}(\widehat{\psi}_0)$ and numerically compared its size and power with Wu *et al.*'s test $F_{WHH}(\widehat{\psi}_0)$.

We now consider the test statistic given by

$$F_G(\widehat{\psi}_0) = \frac{(\boldsymbol{C}\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0) - \boldsymbol{b})'\{\boldsymbol{C}[\boldsymbol{B} + \boldsymbol{A}(\widehat{\psi}_0)]^{-1}\boldsymbol{C}'\}^{-1}(\boldsymbol{C}\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0) - \boldsymbol{b})/q}{\hat{\sigma}_e^{2U}},$$

where the denominator uses the unbiased estimator $\hat{\sigma}_e^{2U}$ constructed from the 'within' analysis of variance. Along the arguments as in Wu, *et al.* (1988), we try to modify the numerator of $F_G(\widehat{\psi}_0)$. Let

$$G(\widehat{\psi}_0) = (\boldsymbol{C}\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0) - \boldsymbol{b})'\{\boldsymbol{C}[\boldsymbol{B} + \boldsymbol{A}(\widehat{\psi}_0)]^{-1}\boldsymbol{C}'\}^{-1}(\boldsymbol{C}\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0) - \boldsymbol{b})/\sigma_e^2.$$

In the context of small area estimation, it is common to consider the asymptotic approximation when the number of small areas, k, is large. We thus derive an asymptotic approximation to $E[G(\widehat{\psi}_0)]$ as $k \to \infty$. Let $\boldsymbol{D}(\psi) = \boldsymbol{C}'\{\boldsymbol{C}[\boldsymbol{B} + \boldsymbol{A}(\psi)]^{-1}\boldsymbol{C}'\}^{-1}\boldsymbol{C}$, $\boldsymbol{D}^{(1)}(\psi) = (\mathrm{d}/\mathrm{d}\psi)\boldsymbol{D}(\psi)$ and $\boldsymbol{D}^{(2)}(\psi) = (\mathrm{d}^2/\mathrm{d}\psi^2)\boldsymbol{D}(\psi)$. Also let $\boldsymbol{A}(\psi) = \sum_i n_i\gamma_i\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'$, $\boldsymbol{A}^{(1)}(\psi) = -\sum_i n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'$ and $\boldsymbol{A}^{(2)}(\psi) = 2\sum_i n_i^3\gamma_i^3\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'$. Define $h(\psi)$ by

$$\begin{aligned} h(\psi) =& \frac{2k}{N(N-k)}\sum_i \gamma_i^{-1}\mathrm{tr}\,(\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}\boldsymbol{D}^{(1)}(\psi) \\ &+ \frac{2k}{N^2}\Big\{\sum_i \gamma_i^{-2} + \frac{1}{N-k}(\sum_i \gamma_i^{-1})^2\Big\} \\ &\quad \times \Big\{\frac{1}{2}\mathrm{tr}\,(\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}\boldsymbol{D}^{(2)} + \frac{1}{2}\mathrm{tr}\,\boldsymbol{D}_0(\psi)\boldsymbol{A}^{(2)}(\psi) \\ &\qquad - \mathrm{tr}\,\boldsymbol{D}_0(\psi)\boldsymbol{A}^{(1)}(\psi)(\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}\boldsymbol{A}^{(1)}(\psi)\Big\} \end{aligned} \tag{2.5}$$

where $\boldsymbol{D}_0(\psi) = (\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}\boldsymbol{D}(\psi)(\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}$.

Theorem 2.1 *Assume that $\boldsymbol{X}$ and $\boldsymbol{A}(\psi)$ are of full rank and that $\boldsymbol{B} = O(k)$ and $\boldsymbol{A}(\psi) = O(k)$ as $k \to \infty$. Then, the expectation of $G(\widehat{\psi}_0)$ is approximated as $E[G(\widehat{\psi}_0)] = q + h(\psi)/k + o(k^{-1})$.*

From this theorem, we obtain the asymptotically corrected test statistic given by

$$F_{ACG}(\widehat{\psi}_0, \hat{\sigma}_e^2) = F_G(\widehat{\psi}_0, \hat{\sigma}_e^2)/(q + h(\widehat{\psi}_0)/k). \tag{2.6}$$

When $\boldsymbol{C} = \boldsymbol{I}$, we have that $q = p$, $\boldsymbol{D} = \boldsymbol{B} + \boldsymbol{A}(\psi)$, $\boldsymbol{D}^{(1)}(\psi) = \boldsymbol{A}^{(1)}(\psi)$ and $\boldsymbol{D}^{(2)}(\psi) = \boldsymbol{A}^{(2)}(\psi)$, so that $h(\psi)$ is expressed as

$$\begin{aligned} h(\psi) =& \frac{2k}{N(N-k)}\sum_i \gamma_i^{-1}\mathrm{tr}\,(\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}\boldsymbol{A}^{(1)}(\psi) \\ &+ \frac{2k}{N^2}\Big\{\sum_i \gamma_i^{-2} + \frac{1}{N-k}(\sum_i \gamma_i^{-1})^2\Big\} \\ &\quad \times \Big\{\mathrm{tr}\,(\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}\boldsymbol{A}^{(2)} - \mathrm{tr}\,[(\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}\boldsymbol{A}^{(1)}(\psi)]^2\Big\}. \end{aligned}$$

For the general matrix $\boldsymbol{C}$, $\boldsymbol{D}^{(2)}(\psi)$ gives a complicated form, but we can compute it as follows: Let $\boldsymbol{E}(\psi) = (\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}$ and $\boldsymbol{F}(\psi) = \{\boldsymbol{C}(\boldsymbol{B} + \boldsymbol{A}(\psi))^{-1}\boldsymbol{C}'\}^{-1}$. Then, $\boldsymbol{D}^{(1)}(\psi) = \boldsymbol{C}'\boldsymbol{F}^{(1)}(\psi)\boldsymbol{C}$, which can be obtained from $\boldsymbol{F}^{(1)}(\psi) = -\boldsymbol{F}(\psi)\boldsymbol{C}\boldsymbol{E}^{(1)}(\psi)\boldsymbol{C}'\boldsymbol{F}(\psi)$ and $\boldsymbol{E}^{(1)}(\psi) = -\boldsymbol{E}(\psi)\boldsymbol{A}^{(1)}(\psi)\boldsymbol{E}(\psi)$. Also, we have $\boldsymbol{D}^{(2)}(\psi) = \boldsymbol{C}'\boldsymbol{F}^{(2)}(\psi)\boldsymbol{C}$, which can be obtained from

$$\boldsymbol{F}^{(2)}(\psi) = -\boldsymbol{F}^{(1)}(\psi)\boldsymbol{C}\boldsymbol{E}^{(1)}(\psi)\boldsymbol{C}'\boldsymbol{F}(\psi) - \boldsymbol{F}(\psi)\boldsymbol{C}\boldsymbol{E}^{(2)}(\psi)\boldsymbol{C}'\boldsymbol{F}(\psi) - \boldsymbol{F}(\psi)\boldsymbol{C}\boldsymbol{E}^{(1)}(\psi)\boldsymbol{C}'\boldsymbol{F}^{(1)}(\psi)$$

and

$$\boldsymbol{E}^{(2)}(\psi) = -\boldsymbol{E}^{(1)}(\psi)\boldsymbol{A}^{(1)}(\psi)\boldsymbol{E}(\psi) - \boldsymbol{E}(\psi)\boldsymbol{A}^{(2)}(\psi)\boldsymbol{E}(\psi) - \boldsymbol{E}(\psi)\boldsymbol{A}^{(1)}(\psi)\boldsymbol{E}^{(1)}(\psi).$$

These expressions can be used in the numerical investigations studied in the next section.

3 Numerical Studies

In this section, we shall investigate the performances of the sizes of the test statistics proposed in the previous section through simulation experiments.

Consider k small areas, and the sample sizes n_i's for small areas are generated as $n_i = 1 + Bin(10, 1/2)$ for $i = 1, \ldots, k$, where $Bin(10, 1/2)$ is a random variable distributed as a binomial distribution with mean 5 and success probability $1/2$. The observations from each small area are generated as for $i = 1, \ldots, k$,

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{j}_{n_i}v_i + \boldsymbol{e}_i, \tag{3.1}$$

where $\boldsymbol{y}_i$ is an n_i-dimensional vector, $\boldsymbol{X}_i$ is a $n_i \times p$ matrix of the regressor variables and the other notations are defined around (1.2). Here, v_i and $\boldsymbol{e}_i$ are random observations from $\mathcal{N}(0, \psi)$ and $\mathcal{N}_{n_i}(\boldsymbol{0}, \boldsymbol{I}_{n_i})$, respectively. The regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$ are set up as $\beta_i = 5(-1)^i(U_i + 1)$ for $i = 0, \ldots, p-1$ where U_i is a random number from a uniform distribution on the interval $(0, 1)$. For the regressor variables, it is supposed that the model has an intercept term, namely, $\boldsymbol{X}_i' = (\boldsymbol{j}_{n_i}, \boldsymbol{X}_i^{*\prime})$ for $n_i \times (p-1)$ matrix $\boldsymbol{X}_i^*$, where $\boldsymbol{X}_i^*$ is generated as

$$\boldsymbol{X}_i^* = \boldsymbol{j}_{n_i}\boldsymbol{u}_i' + \boldsymbol{Z}_i,$$

where $\boldsymbol{Z}_i$ is an $n_i \times (p-1)$ random matrix having $\mathcal{N}_{n_i,p-1}(\boldsymbol{0}, (10\boldsymbol{I}_{n_i}) \otimes \boldsymbol{I}_{p-1})$, and $\boldsymbol{u}_i$ is a $(p-1)$-dimensional vector having $\mathcal{N}_{p-1}(\boldsymbol{0}, 10\boldsymbol{\Sigma}_u)$ for $\boldsymbol{\Sigma}_u = (1-\rho_u)\boldsymbol{I}_{p-1} + \rho_u\boldsymbol{j}_{p-1}\boldsymbol{j}_{p-1}'$ and $\rho_u = 0.6$.

In the simulation experiments, we handle the three cases: (A) $k = 10$, $p = 3$, $q = 2$ and $H_0 : \beta_1 = \beta_2 = 0$, (B) $k = 10$, $p = 3$, $q = 2$ and $H_0 : \beta_0 = \beta_1 = 0$, and (C) $k = 20$, $p = 5$, $q = 3$ and $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$, where ψ takes the values $\psi = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. A set of observations of the regressor variables $\boldsymbol{X}$ is generated, and 10,000 observations of the response variable $\boldsymbol{y}$ are generated from the model (3.1). Sizes of test statistics can be approximated based on these simulation experiments. The test statistics we want to investigate are the OLS F-test statistic F_{OLS}, Wu *et al.*'s test statistic $F_{WHH}(\widehat{\psi}_0)$, Rao *et al.*'s test statistic $F_{RSY}(\widehat{\psi}_0)$, the exact test statistic F_{EXT} and the asymptotically

Table 1: Size Estimates (%) of Tests F_{OLS}, F_{WHH}, F_{RSY}, F_{EXT} and F_{ACG}

	Nominal level **5%**					Nominal level **1%**				
ψ	F_{OLS}	F_{WHH}	F_{RSY}	F_{EXT}	F_{ACG}	F_{OLS}	F_{WHH}	F_{RSY}	F_{EXT}	F_{ACG}
(A) $k = 10$, $p = 3$, $q = 2$ and $H_0 : \beta_1 = \beta_2 = 0$										
0.0	5.2	4.6	5.1	4.7	2.2	1.0	0.9	0.9	0.8	0.3
0.2	12.0	8.3	7.0	4.5	4.1	4.0	2.7	1.8	0.9	0.8
0.4	19.8	11.1	8.0	4.7	5.1	8.3	4.3	2.2	1.0	1.2
0.6	24.1	11.3	7.5	5.0	4.9	11.2	4.6	1.9	1.0	1.0
0.8	27.9	11.6	7.1	4.8	4.6	13.3	4.8	2.0	1.0	1.1
1.0	29.5	12.2	6.9	4.8	4.7	15.0	5.0	1.6	1.0	0.9
(B) $k = 10$, $p = 3$, $q = 2$ and $H_0 : \beta_0 = \beta_1 = 0$										
0.0	5.6	4.4	4.7	5.0	1.9	0.9	0.8	0.7	0.8	0.2
0.2	15.9	9.0	7.7	4.9	4.5	5.7	3.2	2.2	1.0	0.9
0.4	24.6	11.0	8.3	4.8	5.2	11.3	4.6	2.5	1.0	1.3
0.6	30.4	11.8	8.4	4.7	5.4	15.3	5.3	2.5	1.0	1.4
0.8	34.3	12.4	8.0	5.1	5.4	19.2	5.9	2.6	0.9	1.7
1.0	37.5	13.0	8.1	4.9	5.1	21.7	6.7	2.4	1.0	1.3
(C) $k = 20$, $p = 5$, $q = 3$ and $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$										
0.0	5.2	4.7	4.9	4.7	2.9	0.9	0.9	0.8	0.9	0.5
0.2	12.0	7.3	6.1	5.1	4.7	3.7	2.2	1.4	1.0	1.1
0.4	19.3	9.3	6.5	4.8	5.0	7.7	3.0	1.2	1.0	0.8
0.6	23.3	9.0	5.7	5.0	4.6	9.9	3.1	1.5	1.0	1.0
0.8	26.4	8.9	5.7	5.0	4.6	12.4	3.1	1.4	1.2	1.0
1.0	29.3	9.6	5.6	5.0	4.5	14.9	3.6	1.1	1.0	0.8

corrected test statistic $F_{ACG}(\widehat{\psi}_0, \hat{\sigma}_e^{2U})$, which are given in (1.3), (1.4), (1.5), (2.2) and (2.6). The sizes of these test statistics for the nominal significance levels $\alpha = 5\%, 1\%$ are reported in Table 1, where the notations $F_{WHH}(\widehat{\psi}_0)$, $F_{RSY}(\widehat{\psi}_0)$ and $F_{ACG}(\widehat{\psi}_0, \hat{\sigma}_e^{2U})$ are abbreviated as F_{WHH}, F_{RSY} and F_{ACG}. In the cases of (A) and (B), $\boldsymbol{X}_i^*$'s are two dimensional and the sample correlation coefficient for the simulated data is 0.350. For the case (C), the correlation matrix of the simulated sample $\boldsymbol{X}_i^*$'s is

$$\text{Sample Correlation Matrix of } \boldsymbol{X}_i^*\text{'s} = \begin{pmatrix} 1.000 & 0.347 & 0.316 & 0.264 \\ 0.347 & 1.000 & 0.357 & 0.386 \\ 0.316 & 0.357 & 1.000 & 0.345 \\ 0.264 & 0.386 & 0.345 & 1.000 \end{pmatrix}$$

As seen from Table 1, the sizes of the Rao *et al.*'s test F_{RSY} are much better than those of F_{OLS} and F_{WHH}, but they are slightly larger than the nominal levels. The asymptocally corrected test F_{ACG} has sizes close to the nominal levels and improves on F_{RSY} in controlling the sizes for $\psi \geq 0.2$. When $\psi = 0$, F_{ACG} is very conservative, and we need a further modification, which will be done in a future. Since F_{EXT} is an exact

Table 2: Power Estimates (%) of Tests F_{RSY}, F_{EXT} and F_{ACG}
(The size of F_{RSY} is adjusted so that its size is about 5%.)

ψ	F_{RSY}	F_{EXT}	F_{ACG}	F_{RSY}	F_{EXT}	F_{ACG}
	$\beta_1 = 0.00$, $\beta_2 = 0.00$			$\beta_1 = -0.05$, $\beta_2 = 0.04$		
0.0	5.0	4.7	2.2	46.7	26.1	33.5
0.2	5.0	4.5	4.1	34.5	26.3	31.5
0.4	5.0	4.8	5.1	28.5	25.6	29.4
0.6	5.0	5.0	5.0	27.7	25.7	27.5
0.8	5.0	4.8	4.6	29.3	25.8	27.8
1.0	5.0	4.8	4.8	27.5	26.4	26.7
	$\beta_1 = -0.10$, $\beta_2 = 0.08$			$\beta_1 = -0.15$, $\beta_2 = 0.12$		
0.0	97.2	79.6	94.3	100.0	99.0	100.0
0.2	90.9	79.9	89.7	100.0	99.1	99.9
0.4	86.2	80.2	86.8	99.6	99.1	99.6
0.6	84.2	79.4	84.2	99.6	99.1	99.6
0.8	83.5	79.4	82.6	99.4	99.1	99.4
1.0	82.3	79.1	81.9	99.4	99.0	99.3

test, its sizes satisfy the nominal level.

The powers of the tests F_{RSY}, F_{EXT} and F_{ACG} are examined in the case (A) and reported in Table 2 for $(\beta_1, \beta_2) = (-0.05, 0.04)$, $(-0.10, 0.08)$ and $(-0.15, 0.12)$, where the test F_{RSY} is adjusted so that its size satisfies the nominal level $\alpha = 5\%$, but F_{ACG} is not adjusted. The column for $\beta_1 = \beta_2 = 0$ means the sizes of the tests, and in most cases the three tests satisfy the nominal level. It is seen that F_{RSY} and F_{ACG} are more powerful than the exact test F_{EXT} and that F_{ACG} performs as well as F_{RSY}. From the numerical results reported in Tables 1 and 2, the asymptotically corrected test F_{ACG} is recommendable.

4 Proof of Theorem 2.1

For notational convenience, let $\boldsymbol{D}(\psi) = \boldsymbol{C}'\{\boldsymbol{C}[\boldsymbol{B} + \boldsymbol{A}(\psi)]^{-1}\boldsymbol{C}'\}^{-1}\boldsymbol{C}$. Under the null hypothesis $H_0 : \boldsymbol{b} = \boldsymbol{C\beta}$, the expectation $E[G(\widehat{\psi})]$ is written as

$$\begin{aligned} E_\beta[G(\widehat{\psi}_0)] =& E_\beta[(\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0) - \boldsymbol{\beta})'\boldsymbol{D}(\widehat{\psi}_0)(\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0) - \boldsymbol{\beta})]/\sigma_e^2 \\ =& E_0[\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0)'\boldsymbol{D}(\widehat{\psi}_0)\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0)]/\sigma_e^2, \end{aligned}$$

so that we can put $\boldsymbol{\beta} = \boldsymbol{0}$ without any loss of generality, and we shall omit 0 in the expectation notation $E_0[\cdot]$.

Let $\widetilde{\boldsymbol{\beta}}^{(i)}(\psi) = (\mathrm{d}^i/\mathrm{d}\psi^i)\widetilde{\boldsymbol{\beta}}(\psi)$ and $\boldsymbol{D}^{(i)}(\psi) = (\mathrm{d}^i/\mathrm{d}\psi^i)\boldsymbol{D}(\psi)$ for $i = 1, 2$. Using the Taylor expansion of $\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0)$ and $\boldsymbol{D}(\widehat{\psi}_0)$ around $\widehat{\psi}_0 = \psi$, we can approximate them as

$\widetilde{\boldsymbol{\beta}}(\widehat{\psi}_0) = \widetilde{\boldsymbol{\beta}}(\psi) + \widetilde{\boldsymbol{\beta}}^{(1)}(\psi)(\widehat{\psi}_0 - \psi) + 2^{-1}\widetilde{\boldsymbol{\beta}}^{(2)}(\psi)(\widehat{\psi}_0 - \psi)^2 + o_p(k^{-1})$ and $\boldsymbol{D}(\widehat{\psi}_0) = \boldsymbol{D}(\psi) + \boldsymbol{D}^{(1)}(\psi)(\widehat{\psi}_0 - \psi) + 2^{-1}\boldsymbol{D}^{(2)}(\psi)(\widehat{\psi}_0 - \psi)^2 + o_p(k^{-1})$. It can be shown that $\widehat{\psi}_0 = \widehat{\psi}^U + o_p(k^{-1})$ for $\widehat{\psi}_0$ and $\widehat{\psi}^U$ given in (2.4) and (2.3), respectively. Evaluating $G(\widehat{\psi}_0)$ up to $O(k^{-1})$, we observe that

$$
\begin{aligned}
G(\widehat{\psi}_0) = & \frac{1}{\sigma_e^2}\Big\{\widetilde{\boldsymbol{\beta}}(\psi)'\boldsymbol{D}(\psi)\widetilde{\boldsymbol{\beta}}(\psi) + 2\widetilde{\boldsymbol{\beta}}(\psi)'\boldsymbol{D}(\psi)\widetilde{\boldsymbol{\beta}}^{(1)}(\psi)(\widehat{\psi}^U - \psi) \\
& + \widetilde{\boldsymbol{\beta}}(\psi)'\boldsymbol{D}(\psi)\widetilde{\boldsymbol{\beta}}^{(2)}(\psi)(\widehat{\psi}^U - \psi)^2 + \widetilde{\boldsymbol{\beta}}^{(1)}(\psi)'\boldsymbol{D}(\psi)\widetilde{\boldsymbol{\beta}}^{(1)}(\psi)(\widehat{\psi}^U - \psi)^2 \\
& + \widetilde{\boldsymbol{\beta}}(\psi)'\boldsymbol{D}^{(1)}(\psi)\widetilde{\boldsymbol{\beta}}(\psi)(\widehat{\psi}^U - \psi) + 2\widetilde{\boldsymbol{\beta}}(\psi)'\boldsymbol{D}^{(1)}(\psi)\widetilde{\boldsymbol{\beta}}^{(1)}(\psi)(\widehat{\psi}^U - \psi)^2 \\
& + \frac{1}{2}\widetilde{\boldsymbol{\beta}}(\psi)'\boldsymbol{D}^{(2)}(\psi)\widetilde{\boldsymbol{\beta}}(\psi)(\widehat{\psi}^U - \psi)^2\Big\} + o_p(k^{-1}).
\end{aligned}
$$

Hereafter, let $\widehat{\psi} = \widehat{\psi}^U$ for simplicity. Also we use the notations $\widetilde{\boldsymbol{\beta}}$, $\widetilde{\boldsymbol{\beta}}^{(1)}$, $\boldsymbol{D}$, $\boldsymbol{D}^{(1)}$ and $\boldsymbol{A}$ instead of $\widetilde{\boldsymbol{\beta}}(\psi)$, $\widetilde{\boldsymbol{\beta}}^{(1)}(\psi)$, $\boldsymbol{D}(\psi)$, $\boldsymbol{D}^{(1)}(\psi)$ and $\boldsymbol{A}(\psi)$, respectively. Since $E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}\widetilde{\boldsymbol{\beta}}/\sigma_e^2] = q$, the expectation $E[G(\widehat{\psi})]$ can be approximated as $E[G(\widehat{\psi})] = q + I_1 + 2^{-1}I_2 + 2I_3 + 2I_4 + I_5 + I_6 + o(k^{-1})$, where $I_1 = E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}^{(1)}\widetilde{\boldsymbol{\beta}}(\widehat{\psi} - \psi)]/\sigma_e^2$, $I_2 = E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}^{(2)}\widetilde{\boldsymbol{\beta}}(\widehat{\psi} - \psi)^2]/\sigma_e^2$, $I_3 = E\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}\widetilde{\boldsymbol{\beta}}^{(1)}(\widehat{\psi} - \psi)]/\sigma_e^2$, $I_4 = E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}^{(1)}\widetilde{\boldsymbol{\beta}}^{(1)}(\widehat{\psi} - \psi)^2]/\sigma_e^2$, $I_5 = E[\widetilde{\boldsymbol{\beta}}^{(1)\prime}\boldsymbol{D}\widetilde{\boldsymbol{\beta}}^{(1)}(\widehat{\psi} - \psi)^2]/\sigma_e^2$ and $I_6 = E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}\widetilde{\boldsymbol{\beta}}^{(2)}(\widehat{\psi} - \psi)^2]/\sigma_e^2$. Hence, we need to evaluate the terms I_1 to I_6. For the purpose, we note the following distributional properties for $\boldsymbol{\beta} = \boldsymbol{0}$:

(1) From (2.1), it is seen that $\widehat{\boldsymbol{\beta}}_1$, S_1 and $(\overline{y}_1, \ldots, \overline{y}_k)$ are mutually independently distributed as $S_1/\sigma_e^2 \sim \chi^2_{N-k-p+\lambda}$ and $\overline{y}_i \sim \mathcal{N}(\overline{\boldsymbol{x}}_i'\boldsymbol{\beta}, \sigma_e^2/(n_i\gamma_i))$ for $i = 1, \ldots, k$. For a distribution of $\widehat{\boldsymbol{\beta}}_1$, let $\boldsymbol{H}$ and $\boldsymbol{L}$ be a $p \times p$ orthogonal matrix and a diagonal matrix with positive diagonal elements such that

$$
\boldsymbol{B} = \boldsymbol{H}'\begin{pmatrix} \boldsymbol{L} & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}, = \boldsymbol{H}_1'\boldsymbol{L}\boldsymbol{H}_1
$$

where $\boldsymbol{H}' = (\boldsymbol{H}_1', \boldsymbol{H}_2')$. Then, $\boldsymbol{B}\widehat{\boldsymbol{\beta}}_1 = \boldsymbol{H}_1'\boldsymbol{L}\boldsymbol{z}$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \sigma_e^2\boldsymbol{L}^{-1})$.

(2) The GLS estimator $\widetilde{\boldsymbol{\beta}}$ is expressed as $\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\psi) = (\boldsymbol{B}+\boldsymbol{A})^{-1}(\boldsymbol{H}_1'\boldsymbol{L}\boldsymbol{z} + \sum_i n_i\gamma_i\overline{\boldsymbol{x}}_i\overline{y}_i)$ and it has $\mathcal{N}_p(\boldsymbol{\beta}, \sigma_e^2[\boldsymbol{B} + \boldsymbol{A}]^{-1})$.

(3) From the result of Kackar and Harville (1984), it follows that $\widehat{\psi}$ is independent of $\widetilde{\boldsymbol{\beta}}$.

(4) **(Stein identity)** Let $\boldsymbol{Y}$ be a p-dimensional random vector having $\mathcal{N}_p(\boldsymbol{\eta}, \boldsymbol{\Sigma})$. Then, for p-dimensional absolutely continuous function $\boldsymbol{g}(\boldsymbol{Y})$, Stein (1973, 81) derived the following identity:

$$
E[\boldsymbol{Y}'\boldsymbol{g}(\boldsymbol{Y})] = E[(\partial/\partial\boldsymbol{Y})'\boldsymbol{\Sigma}\boldsymbol{g}(\boldsymbol{Y})],
$$

where $(\partial/\partial\boldsymbol{Y})' = \partial/\partial\boldsymbol{Y}' = (\partial/\partial Y_1, \ldots, \partial/\partial Y_p)$ for $\boldsymbol{Y} = (Y_1, \ldots, Y_p)'$.

(5) The following equation is also useful: for scalar function $f(\boldsymbol{Y})$,

$$
(\partial/\partial\boldsymbol{Y})'[\boldsymbol{g}(\boldsymbol{Y})f(\boldsymbol{Y})] = [(\partial/\partial\boldsymbol{Y})'\boldsymbol{g}(\boldsymbol{Y})]f(\boldsymbol{Y}) + \boldsymbol{g}(\boldsymbol{Y})'[\partial f(\boldsymbol{Y})/\partial\boldsymbol{Y}].
$$

To evaluate $I_1 = E[\widetilde{\boldsymbol{\beta}}' \boldsymbol{D}^{(1)} \widetilde{\boldsymbol{\beta}} (\widehat{\psi} - \psi)]/\sigma_e^2$, note that $\widetilde{\boldsymbol{\beta}}$ is independent of $\widehat{\psi}$ and that $E[\widetilde{\boldsymbol{\beta}}' \boldsymbol{D}^{(1)} \widetilde{\boldsymbol{\beta}}] = E[\mathrm{tr}\,\{\boldsymbol{D}^{(1)} \widetilde{\boldsymbol{\beta}} \widetilde{\boldsymbol{\beta}}'\}] = \sigma_e^2 \mathrm{tr}\,\{\boldsymbol{D}^{(1)} [\boldsymbol{B} + \boldsymbol{A}(\psi)]^{-1}\}$. Thus, it is observed that

$$
\begin{aligned}
I_1 =& E[\widetilde{\boldsymbol{\beta}}' \boldsymbol{D}^{(1)} \widetilde{\boldsymbol{\beta}}] Bias(\widehat{\psi})/\sigma_e^2 \\
=& \mathrm{tr}\,\{\boldsymbol{D}^{(1)} [\boldsymbol{B} + \boldsymbol{A}(\psi)]^{-1}\} Bias(\widehat{\psi}).
\end{aligned}
$$

Similarly,

$$
I_2 = E[\widetilde{\boldsymbol{\beta}}' \boldsymbol{D}^{(2)} \widetilde{\boldsymbol{\beta}} (\widehat{\psi} - \psi)^2]/\sigma_e^2 = \mathrm{tr}\,\{\boldsymbol{D}^{(2)} [\boldsymbol{B} + \boldsymbol{A}(\psi)]^{-1}\} E[(\widehat{\psi} - \psi)^2].
$$

To evaluate $I_3 = E\widetilde{\boldsymbol{\beta}}' \boldsymbol{D} \widetilde{\boldsymbol{\beta}}^{(1)} (\widehat{\psi} - \psi)]/\sigma_e^2$, note that $\widehat{\psi}$ is not independent of $\widetilde{\boldsymbol{\beta}}^{(1)}$, which is written as

$$
\widetilde{\boldsymbol{\beta}}^{(1)} = -(\boldsymbol{B} + \boldsymbol{A})^{-1} \boldsymbol{A}^{(1)} \widetilde{\boldsymbol{\beta}} - (\boldsymbol{B} + \boldsymbol{A})^{-1} \sum_i n_i^2 \gamma_i^2 \overline{\boldsymbol{x}}_i \overline{y}_i, \tag{4.1}
$$

where $\boldsymbol{A}^{(1)} = (\mathrm{d}/\mathrm{d}\psi) \boldsymbol{A}(\psi)$. Then,

$$
\begin{aligned}
I_3 =& - E[\widetilde{\boldsymbol{\beta}}' \boldsymbol{D} (\boldsymbol{B} + \boldsymbol{A})^{-1} \boldsymbol{A}^{(1)} \widetilde{\boldsymbol{\beta}} (\widehat{\psi} - \psi)]/\sigma_e^2 \\
& - E[\widetilde{\boldsymbol{\beta}}' \boldsymbol{D} (\boldsymbol{B} + \boldsymbol{A})^{-1} \sum_i n_i^2 \gamma_i^2 \overline{\boldsymbol{x}}_i \overline{y}_i (\widehat{\psi} - \psi)]/\sigma_e^2 \\
=& I_{31} + I_{32}. \quad \text{(say)}
\end{aligned}
$$

Let $\boldsymbol{D}_0 = (\boldsymbol{B} + \boldsymbol{A})^{-1} \boldsymbol{D} (\boldsymbol{B} + \boldsymbol{A})^{-1}$. The independence of $\widehat{\psi}$ and $\widetilde{\boldsymbol{\beta}}$ implies that $I_{31} = -\mathrm{tr}\, \boldsymbol{D}_0 \boldsymbol{A}^{(1)} Bias(\widehat{\psi})$. The Stein identity is used to get that

$$
\begin{aligned}
I_{32} =& - \sum_i E[n_i^2 \gamma_i^2 \overline{\boldsymbol{x}}_i' \overline{y}_i \boldsymbol{D}_0 (\boldsymbol{H}_1' \boldsymbol{L} \boldsymbol{z} + \sum_j n_j \gamma_j \overline{\boldsymbol{x}}_j \overline{y}_j) (\widehat{\psi} - \psi)]/\sigma_e^2 \\
=& - \sum_i E[n_i \gamma_i \overline{\boldsymbol{x}}_i' \boldsymbol{D}_0 n_i \gamma_i \overline{\boldsymbol{x}}_i (\widehat{\psi} - \psi)] - \sum_i E[n_i \gamma_i \overline{\boldsymbol{x}}_i' \boldsymbol{D}_0 (\boldsymbol{H}_1' \boldsymbol{L} \boldsymbol{z} + \sum_j n_j \gamma_j \overline{\boldsymbol{x}}_j \overline{y}_j) \frac{\partial \widehat{\psi}}{\partial \overline{y}_i}] \\
=& I_{321} + I_{322}. \quad \text{(say)}
\end{aligned}
$$

Since $\boldsymbol{A}^{(1)} = -\sum_i n_i^2 \gamma_i^2 \overline{\boldsymbol{x}}_i \overline{\boldsymbol{x}}_i'$, it is easy to see that $I_{321} = \mathrm{tr}\, \boldsymbol{D}_0 \boldsymbol{A}^{(1)} Bias(\widehat{\psi})$. The Stein identity is used again to rewrite I_{322} as

$$
\begin{aligned}
I_{322} =& - \sum_i n_i \gamma_i E[(\boldsymbol{z}' \boldsymbol{L} \boldsymbol{H}_1 + \sum_j n_j \gamma_j \overline{\boldsymbol{x}}_j' \overline{y}_j) \boldsymbol{D}_0 \overline{\boldsymbol{x}}_i \frac{\partial \widehat{\psi}}{\partial \overline{y}_i}] \\
=& - \sigma_e^2 \sum_i n_i \gamma_i E[\mathrm{tr}\, \boldsymbol{H}_1 \boldsymbol{D}_0 \overline{\boldsymbol{x}}_i \frac{\partial^2 \widehat{\psi}}{\partial \boldsymbol{z}' \partial \overline{y}_i}] - \sigma_e^2 \sum_i n_i \gamma_i \sum_j E[\overline{\boldsymbol{x}}_j' \boldsymbol{D}_0 \overline{\boldsymbol{x}}_i \frac{\partial^2 \widehat{\psi}}{\partial \overline{y}_i \partial \overline{y}_j}].
\end{aligned}
$$

From Lemma 4.1 given below, it follows that

$$\begin{aligned}I_{322} =&E[\frac{2\sigma_e^2}{N_*\hat{\sigma}_e^2}]\sum_i n_i\gamma_i \mathrm{tr}\, \boldsymbol{H}_1\boldsymbol{D}_0\overline{\boldsymbol{x}}_i n_i \overline{\boldsymbol{x}}_i'(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{H}_1'\boldsymbol{L}\\
&- E[\frac{2\sigma_e^2}{N_*\hat{\sigma}_e^2}]\sum_i n_i\gamma_i\overline{\boldsymbol{x}}_i'\boldsymbol{D}_0\overline{\boldsymbol{x}}_i n_i\\
&+ E[\frac{2\sigma_e^2}{N_*\hat{\sigma}_e^2}]\sum_i n_i\gamma_i\sum_j E[\overline{\boldsymbol{x}}_j'\boldsymbol{D}_0\overline{\boldsymbol{x}}_i n_i n_j \overline{\boldsymbol{x}}_i'(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\overline{\boldsymbol{x}}_j\\
=&E[\frac{2\sigma_e^2}{N_*\hat{\sigma}_e^2}]\mathrm{tr}\,\boldsymbol{D}_0\sum_i n_i^2\gamma_i\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'\Big\{(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{B}-(\boldsymbol{I}-(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0)\Big\},\end{aligned}$$

which is equal to zero. Hence,

$$I_3 = 0.$$

For $I_4 = E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}^{(1)}\widetilde{\boldsymbol{\beta}}^{(1)}(\widehat{\psi}-\psi)^2]/\sigma_e^2$, from (4.1), it is written as

$$\begin{aligned}I_4 =&- E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}\widetilde{\boldsymbol{\beta}}(\widehat{\psi}-\psi)^2]/\sigma_e^2\\
&- E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}\sum_i n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i\overline{y}_i(\widehat{\psi}-\psi)^2]/\sigma_e^2\\
=&I_{41}+I_{42}.\end{aligned}$$

Let $\boldsymbol{D}_1 = (\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{D}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}$. The independence of $\widehat{\psi}$ and $\widetilde{\boldsymbol{\beta}}$ implies that $I_{41} = -\mathrm{tr}\,\boldsymbol{D}_1\boldsymbol{A}^{(1)}E[(\widehat{\psi}-\psi)^2]$. Using the Stein identity, we have

$$\begin{aligned}I_{42} =&-\sum_i E[n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i'\overline{y}_i\boldsymbol{D}_1(\boldsymbol{H}_1'\boldsymbol{L}\boldsymbol{z}+\sum_j n_j\gamma_j\overline{\boldsymbol{x}}_j\overline{y}_j)(\widehat{\psi}-\psi)^2]/\sigma_e^2\\
=&-\sum_i E[n_i\gamma_i\overline{\boldsymbol{x}}_i'\boldsymbol{D}_1 n_i\gamma_i\overline{\boldsymbol{x}}_i(\widehat{\psi}-\psi)^2]\\
&-2\sum_i E[n_i\gamma_i\overline{\boldsymbol{x}}_i'\boldsymbol{D}_1(\boldsymbol{H}_1'\boldsymbol{L}\boldsymbol{z}+\sum_j n_j\gamma_j\overline{\boldsymbol{x}}_j\overline{y}_j)\frac{\partial\widehat{\psi}}{\partial\overline{y}_i}(\widehat{\psi}-\psi)]\\
=&I_{421}+I_{422}.\end{aligned}$$

It can be seen that $I_{421} = \mathrm{tr}\,\boldsymbol{D}_1\boldsymbol{A}^{(1)}E[(\widehat{\psi}-\psi)^2]$. The Stein identity is used again to get that

$$\begin{aligned}I_{422} =&-2\sum_i n_i\gamma_i E[(\boldsymbol{z}'\boldsymbol{L}\boldsymbol{H}_1+\sum_j n_j\gamma_j\overline{\boldsymbol{x}}_j'\overline{y}_j)\boldsymbol{D}_1\overline{\boldsymbol{x}}_i\frac{\partial\widehat{\psi}}{\partial\overline{y}_i}(\widehat{\psi}-\psi)]\\
=&-2\sigma_e^2\sum_i n_i\gamma_i E[\mathrm{tr}\,\boldsymbol{H}_1\boldsymbol{D}_1\overline{\boldsymbol{x}}_i\{\frac{\partial^2\widehat{\psi}}{\partial\boldsymbol{z}'\partial\overline{y}_i}(\widehat{\psi}-\psi)+\frac{\partial\widehat{\psi}}{\partial\overline{y}_i}\frac{\partial\widehat{\psi}}{\partial\boldsymbol{z}'}\}]\\
&-2\sigma_e^2\sum_i n_i\gamma_i\sum_j E[\overline{\boldsymbol{x}}_j'\boldsymbol{D}_1\overline{\boldsymbol{x}}_i\{\frac{\partial^2\widehat{\psi}}{\partial\overline{y}_i\partial\overline{y}_j}(\widehat{\psi}-\psi)+\frac{\partial\widehat{\psi}}{\partial\overline{y}_i}\frac{\partial\widehat{\psi}}{\partial\overline{y}_j}\}].\end{aligned}$$

Hence from Lemma 4.2,

$$I_4 = o(k^{-1}). \tag{4.2}$$

For $I_5 = E[\widetilde{\boldsymbol{\beta}}^{(1)\prime}\boldsymbol{D}\widetilde{\boldsymbol{\beta}}^{(1)}(\widehat{\psi}-\psi)^2]/\sigma_e^2$, from (4.1), I_5 is rewritten as

$$\begin{aligned}
I_5 =& - E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{A}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{D}\widetilde{\boldsymbol{\beta}}^{(1)}(\widehat{\psi}-\psi)^2]/\sigma_e^2 \\
&+ E[\sum_i n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i'\overline{y}_i\boldsymbol{D}_0\boldsymbol{A}^{(1)}\widetilde{\boldsymbol{\beta}}(\widehat{\psi}-\psi)^2]/\sigma_e^2 \\
&+ E[\sum_i n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i'\overline{y}_i\boldsymbol{D}_0\sum_j n_j^2\gamma_j^2\overline{\boldsymbol{x}}_j\overline{y}_j(\widehat{\psi}-\psi)^2]/\sigma_e^2 \\
&= I_{51}+I_{52}+I_{53}.
\end{aligned}$$

From the same arguments as in the evaluation of I_4, it follows that $I_{51}=o(k^{-1})$. For I_{52}, let $\boldsymbol{D}_5=\boldsymbol{D}_0\boldsymbol{A}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}=(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{D}(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}$. Similar to I_{42},

$$\begin{aligned}
I_{52} =& \sum_i E[n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i'\overline{y}_i\boldsymbol{D}_5(\boldsymbol{H}_1'\boldsymbol{L}\boldsymbol{z}+\sum_j n_j\gamma_j\overline{\boldsymbol{x}}_j\overline{y}_j)(\widehat{\psi}-\psi)^2]/\sigma_e^2 \\
=& \sum_i E[n_i\gamma_i\overline{\boldsymbol{x}}_i'\boldsymbol{D}_5 n_i\gamma_i\overline{\boldsymbol{x}}_i(\widehat{\psi}-\psi)^2] \\
&+ 2\sum_i E[n_i\gamma_i\overline{\boldsymbol{x}}_i'\boldsymbol{D}_5(\boldsymbol{H}_1'\boldsymbol{L}\boldsymbol{z}+\sum_j n_j\gamma_j\overline{\boldsymbol{x}}_j\overline{y}_j)\frac{\partial\widehat{\psi}}{\partial\overline{y}_i}(\widehat{\psi}-\psi)] \\
=& - \operatorname{tr}\boldsymbol{D}_5\boldsymbol{A}^{(1)}E[(\widehat{\psi}-\psi)^2] \\
&+ 2\sigma_e^2\sum_i n_i\gamma_i E[\operatorname{tr}\boldsymbol{H}_1\boldsymbol{D}_5\overline{\boldsymbol{x}}_i\{\frac{\partial^2\widehat{\psi}}{\partial\boldsymbol{z}'\partial\overline{y}_i}(\widehat{\psi}-\psi)+\frac{\partial\widehat{\psi}}{\partial\overline{y}_i}\frac{\partial\widehat{\psi}}{\partial\boldsymbol{z}'}\}] \\
&+ 2\sigma_e^2\sum_i n_i\gamma_i\sum_j E[\overline{\boldsymbol{x}}_j'\boldsymbol{D}_5\overline{\boldsymbol{x}}_i\{\frac{\partial^2\widehat{\psi}}{\partial\overline{y}_i\partial\overline{y}_j}(\widehat{\psi}-\psi)+\frac{\partial\widehat{\psi}}{\partial\overline{y}_i}\frac{\partial\widehat{\psi}}{\partial\overline{y}_j}\}],
\end{aligned}$$

which implies that $I_{52}=-\operatorname{tr}\boldsymbol{D}_5\boldsymbol{A}^{(1)}E[(\widehat{\psi}-\psi)^2]+o(k^{-1})=-\operatorname{tr}\boldsymbol{D}_0\boldsymbol{A}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}E[(\widehat{\psi}-\psi)^2]+o(k^{-1})$. Finally, I_{53} is evaluated as

$$\begin{aligned}
I_{53} =& \sum_i n_i\gamma_i\overline{\boldsymbol{x}}_i'\boldsymbol{D}_0 n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i E[(\widehat{\psi}-\psi)^2] + 2\sum_i n_i\gamma_i E[\overline{\boldsymbol{x}}_i'\boldsymbol{D}_0\sum_j n_j^2\gamma_j^2\overline{\boldsymbol{x}}_j\overline{y}_j\frac{\partial\widehat{\psi}}{\partial\overline{y}_i}(\widehat{\psi}-\psi)] \\
=& \operatorname{tr}\boldsymbol{D}_0\sum_i n_i^3\gamma_i^3\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'E[(\widehat{\psi}-\psi)^2] \\
&+ 2\sigma_e^2\sum_i\sum_j n_i\gamma_i n_j\gamma_j E[\overline{\boldsymbol{x}}_i'\boldsymbol{D}_0\overline{\boldsymbol{x}}_j\{\frac{\partial^2\widehat{\psi}}{\partial\overline{y}_i\overline{y}_j}(\widehat{\psi}-\psi)+\frac{\partial\widehat{\psi}}{\partial\overline{y}_i}\frac{\partial\widehat{\psi}}{\partial\overline{y}_j}] \\
=& \frac{1}{2}\operatorname{tr}\boldsymbol{D}_0\boldsymbol{A}^{(2)}E[(\widehat{\psi}-\psi)^2]+o(k^{-1}).
\end{aligned}$$

Hence,

$$I_5 = -\operatorname{tr}\boldsymbol{D}_0\boldsymbol{A}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}E[(\widehat{\psi}-\psi)^2]+\frac{1}{2}\operatorname{tr}\boldsymbol{D}_0\boldsymbol{A}^{(2)}E[(\widehat{\psi}-\psi)^2]+o(k^{-1}).$$

For $I_6 = E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}\widetilde{\boldsymbol{\beta}}^{(2)}(\widehat{\psi}-\psi)^2]/\sigma_e^2$, we need to derive $\widetilde{\boldsymbol{\beta}}^{(2)}$, which is written as

$$
\begin{aligned}
\widetilde{\boldsymbol{\beta}}^{(2)} =&(\boldsymbol{B}+\boldsymbol{A})^{-1}\{\boldsymbol{A}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}-\boldsymbol{A}^{(2)}\}\widetilde{\boldsymbol{\beta}}-(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}\widetilde{\boldsymbol{\beta}}^{(1)}\\
&+(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}\sum_i n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i\overline{y}_i+2(\boldsymbol{B}+\boldsymbol{A})^{-1}\sum_i n_i^3\gamma_i^3\overline{\boldsymbol{x}}_i\overline{y}_i.
\end{aligned}
$$

Since $\sum_i n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i\overline{y}_i = -(\boldsymbol{B}+\boldsymbol{A})\widetilde{\boldsymbol{\beta}}^{(1)}-\boldsymbol{A}^{(1)}\widetilde{\boldsymbol{\beta}}$, $\widetilde{\boldsymbol{\beta}}^{(2)}$ is rewritten as

$$
\widetilde{\boldsymbol{\beta}}^{(2)} = -(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(2)}\widetilde{\boldsymbol{\beta}}-2(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}\widetilde{\boldsymbol{\beta}}^{(1)}+2(\boldsymbol{B}+\boldsymbol{A})^{-1}\sum_i n_i^3\gamma_i^3\overline{\boldsymbol{x}}_i\overline{y}_i,
$$

which is used to express I_6 as

$$
\begin{aligned}
I_6 =&-E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(2)}\widetilde{\boldsymbol{\beta}}(\widehat{\psi}-\psi)^2]/\sigma_e^2\\
&-2E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}\widetilde{\boldsymbol{\beta}}^{(1)}(\widehat{\psi}-\psi)^2]/\sigma_e^2\\
&+2E[\widetilde{\boldsymbol{\beta}}'\boldsymbol{D}(\boldsymbol{B}+\boldsymbol{A})^{-1}\sum_i n_i^3\gamma_i^3\overline{\boldsymbol{x}}_i\overline{y}_i(\widehat{\psi}-\psi)^2]/\sigma_e^2\\
=&I_{61}+I_{62}+I_{63}.
\end{aligned}
$$

It is easy to see that $I_{61} = -\mathrm{tr}\,\boldsymbol{D}_0\boldsymbol{A}^{(2)}E[(\widehat{\psi}-\psi)^2]$. Also from (4.2), $I_{62} = o(k^{-1})$. Similar to I_{52},

$$
\begin{aligned}
I_{63} =&2\mathrm{tr}\,\boldsymbol{D}_0\sum_i n_i^3\gamma_i^3\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'E[(\widehat{\psi}-\psi)^2]+o(k^{-1})\\
=&\mathrm{tr}\,\boldsymbol{D}_0\boldsymbol{A}^{(2)}E[(\widehat{\psi}-\psi)^2]+o(k^{-1}),
\end{aligned}
$$

which implies that

$$
I_6 = o(k^{-1}).
$$

Combining the evaluations I_1-I_6 and recalling $\widehat{\psi}=\widehat{\psi}^U$, we obtain the approximation as

$$
\begin{aligned}
E[G(\widehat{\psi}_0)] =&q+\mathrm{tr}\,(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{D}^{(1)}Bias(\widehat{\psi}^U)\\
&+E[(\widehat{\psi}^U-\psi)^2]\Big\{\frac{1}{2}\mathrm{tr}\,(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{D}^{(2)}+\frac{1}{2}\mathrm{tr}\,\boldsymbol{D}_0\boldsymbol{A}^{(2)}\\
&\qquad-\mathrm{tr}\,\boldsymbol{D}_0\boldsymbol{A}^{(1)}(\boldsymbol{B}+\boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}\Big\}+o_p(k^{-1})
\end{aligned}
$$

Finally, we shall evaluate the bias $Bias(\widehat{\psi})$ and $E[(\widehat{\psi}-\psi)^2]$ for $\widehat{\psi}=\widehat{\psi}^U$, which is written by $\widehat{\psi}=\hat{\sigma}_v^{2U}/\hat{\sigma}_e^{2U}$ for unbiased estimators $\hat{\sigma}_v^{2U}$ and $\hat{\sigma}_e^{2U}$. By the Taylor approximation, it is observed that $(\widehat{\psi}-\psi)/\psi = T_v-T_e+T_e^2-T_vT_e+o_p(k^{-1})$, where $T_v=(\hat{\sigma}_v^{2U}-\sigma_v^2)/\sigma_v^2$

and $T_e = (\hat{\sigma}_e^{2U} - \sigma_e^2)/\sigma_e^2$. From the unbiasedness of $\hat{\sigma}_v^{2U}$ and $\hat{\sigma}_e^{2U}$, it is seen that $E[T_v] = 0$ and $E[T_e] = 0$. Also,

$$\begin{aligned}E[\widehat{\psi} - \psi] =& \psi E[T_e^2 - T_v T_e] + o(k^{-1}),\\ E[(\widehat{\psi} - \psi)^2] =& \psi^2 E\left[T_v^2 - 2T_v T_e + T_e^2\right] + o(k^{-1}).\end{aligned}$$

From Battese and Fuller (1981) and Prasad and Rao (1891), it follows that

$$\begin{aligned}E[T_e^2] =& \frac{1}{\sigma_e^4} Var(\hat{\sigma}_e^{2U}) = \frac{2}{N-k-p+\lambda},\\ E[T_v^2] =& \frac{1}{\sigma_v^4} Var(\hat{\sigma}_v^{2U}) = \frac{2}{N_*^2}\left[\frac{(k-\lambda)(N-p)}{(N-k-p+\lambda)\psi^2} + 2\frac{N_*}{\psi} + N_{**}\right],\\ E[T_v T_e] =& \frac{1}{\sigma_e^2\sigma_v^2} Cov(\hat{\sigma}_e^{2U}, \hat{\sigma}_v^{2U}) = -\frac{2(k-\lambda)}{N_*(N-k-p+\lambda)\psi},\end{aligned}$$

where $N_{**} = \mathrm{tr}\,[(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Z}\boldsymbol{Z}']^2$. Noting that $\psi^2 N_{**}/N_* = \psi^2 \sum_{i=1}^k n_i^2/N + o(k^{-1}) = (\sum_{i=1}^k \gamma_i^{-2} - k - 2N\psi)/N + o(k^{-1})$, we get the approximations as $E[T_e^2] = 2/(N-k) + o(k^{-1})$, $E[T_v T_e] = -2k/\{N(N-k)\psi\} + o(k^{-1})$ and

$$E[T_v^2] = \frac{2}{N^2\psi^2}\left[\frac{k^2}{N-k} + \sum_{i=1}^k \gamma_i^{-2}\right] + o(k^{-1}).$$

Thus, we obtain that

$$\begin{aligned}E[\widehat{\psi} - \psi] =& \frac{2}{N(N-k)}\sum_i \gamma_i^{-1} + o(k^{-1}),\\ E[(\widehat{\psi} - \psi)^2] =& \frac{2}{N^2}\Big\{\sum_i \gamma_i^{-2} + \frac{1}{N-k}(\sum_i \gamma_i^{-1})^2\Big\} + o(k^{-1}),\end{aligned}$$

which are used to get the approximation as

$$\begin{aligned}G(\widehat{\psi}_0) =& q + \frac{2}{N(N-k)}\sum_i \gamma_i^{-1}\mathrm{tr}\,(\boldsymbol{B} + \boldsymbol{A})^{-1}\boldsymbol{D}^{(1)}\\ &+ \frac{2}{N^2}\Big\{\sum_i \gamma_i^{-2} + \frac{1}{N-k}(\sum_i \gamma_i^{-1})^2\Big\}\\ &\times \Big\{\frac{1}{2}\mathrm{tr}\,(\boldsymbol{B} + \boldsymbol{A})^{-1}\boldsymbol{D}^{(2)} + \frac{1}{2}\mathrm{tr}\,\boldsymbol{D}_0\boldsymbol{A}^{(2)} - \mathrm{tr}\,\boldsymbol{D}_0\boldsymbol{A}^{(1)}(\boldsymbol{B} + \boldsymbol{A})^{-1}\boldsymbol{A}^{(1)}\Big\}\\ &+ o_p(k^{-1}).\end{aligned}$$

This is the expression given in Theorem 2.1.

To compete the proof, we show the following lemmas which have been used in the above proof.

Lemma 4.1 *For $\widehat{\psi}=\widehat{\psi}^U$ given in (2.3), the second partial derivatives of $\widehat{\psi}$ with respect to $\boldsymbol{z}$, $\overline{y}_i$ are given by*

$$
\begin{aligned}
\frac{\partial^2\widehat{\psi}}{\partial\boldsymbol{z}\partial\boldsymbol{z}'} =& \frac{2}{N_*\hat{\sigma}_e^2}\boldsymbol{L}\boldsymbol{H}_1(\boldsymbol{B}^+ - (\boldsymbol{B}+\boldsymbol{A}_0)^{-1})\boldsymbol{H}_1'\boldsymbol{L},\\
\frac{\partial^2\widehat{\psi}}{\partial\boldsymbol{z}\partial\overline{y}_i} =& -\frac{2}{N_*\hat{\sigma}_e^2}\boldsymbol{L}\boldsymbol{H}_1(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}n_i\overline{\boldsymbol{x}}_i,\\
\frac{\partial^2\widehat{\psi}}{\partial\overline{y}_i^2} =& \frac{2}{N_*\hat{\sigma}_e^2}\{n_i - n_i^2\overline{\boldsymbol{x}}_i'(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\overline{\boldsymbol{x}}_i\},\\
\frac{\partial^2\widehat{\psi}}{\partial\overline{y}_i\partial\overline{y}_j} =& -\frac{2}{N_*\hat{\sigma}_e^2}n_in_j\overline{\boldsymbol{x}}_i'(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\overline{\boldsymbol{x}}_j,\ (i\neq j).
\end{aligned}
$$

Proof. Note that $\widehat{\psi}=N_*^{-1}\{S/\hat{\sigma}_e^2-(N-p)\}$, and S is expressed as

$$
S = S_1 + S_2 + (\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{A}_0-\boldsymbol{A}_0(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0)(\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2),
$$

where $\boldsymbol{A}_0=\boldsymbol{A}(0)$, $\widehat{\boldsymbol{\beta}}_2=\boldsymbol{A}_0^{-1}\sum_{i=1}^k n_i\overline{\boldsymbol{x}}_i\overline{y}_i$ and $S_2=\sum_{i=1}^k n_i(\overline{y}_i-\overline{\boldsymbol{x}}_i'\widehat{\boldsymbol{\beta}}_2)^2$. Let $S_{(1)}=\sum_{i=1}^k\sum_{j=1}^{n_i}\{(y_{ij}-\overline{y}_i)-(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}_0\}^2$ and $S_{(2)}=\sum_{i=1}^k n_i(\overline{y}_i-\overline{\boldsymbol{x}}_i'\widehat{\boldsymbol{\beta}}_0)^2$. Then, S is written as $S=S_{(1)}+S_{(2)}$. Since $\widehat{\boldsymbol{\beta}}_0=\widehat{\boldsymbol{\beta}}_1-(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0(\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2)$, $S_{(1)}$ is written as

$$
S_{(1)} = S_1 + (\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2)'\boldsymbol{A}_0(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0(\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2).
$$

On the other hand, since $\widehat{\boldsymbol{\beta}}_0=\widehat{\boldsymbol{\beta}}_2+(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{B}(\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2)$, $S_{(2)}$ is written as

$$
S_{(2)} = S_2 + (\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2)'\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{B}(\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2),
$$

so that

$$
S = S_1 + S_2 + (\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2)'\boldsymbol{R}(\widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2),
$$

where

$$
\boldsymbol{R}=\boldsymbol{A}_0(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0+\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{B}.
$$

Note that

$$
\boldsymbol{R}=\boldsymbol{B}(\boldsymbol{B}^+-(\boldsymbol{B}+\boldsymbol{A}_0)^{-1})\boldsymbol{B}=\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0=\boldsymbol{A}_0(\boldsymbol{A}_0^{-1}-(\boldsymbol{B}+\boldsymbol{A}_0)^{-1})\boldsymbol{A}_0.
$$

Using these equations, we can rewrite S as

$$
\begin{aligned}
S =& S_1+S_2+\widehat{\boldsymbol{\beta}}_1'(\boldsymbol{B}-\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{B})\widehat{\boldsymbol{\beta}}_1\\
&-2\widehat{\boldsymbol{\beta}}_1'\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0\widehat{\boldsymbol{\beta}}_2+\widehat{\boldsymbol{\beta}}_2'(\boldsymbol{A}_0-\boldsymbol{A}_0(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0)\widehat{\boldsymbol{\beta}}_2\\
=& S_1+S_2+\boldsymbol{z}'\boldsymbol{L}\boldsymbol{H}_1(\boldsymbol{B}^+-(\boldsymbol{B}+\boldsymbol{A}_0)^{-1})\boldsymbol{H}_1'\boldsymbol{L}\boldsymbol{z}\\
&-2\boldsymbol{z}'\boldsymbol{L}\boldsymbol{H}_1(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\sum_i n_i\overline{\boldsymbol{x}}_i\overline{y}_i+\sum_i n_i\overline{\boldsymbol{x}}_i'\overline{y}_i(\boldsymbol{A}_0^{-1}-(\boldsymbol{B}+\boldsymbol{A}_0)^{-1})\sum_i n_i\overline{\boldsymbol{x}}_i\overline{y}_i.
\end{aligned}
$$

Note that S_1, $\widehat{\boldsymbol{\beta}}_1$, $\overline{y}_1, \ldots, \overline{y}_k$ are mutually independent. The first partial derivatives of $\widehat{\psi}$ with respect to $\boldsymbol{z}$, $\overline{y}_i$ are given by

$$\begin{aligned}\frac{\partial \widehat{\psi}}{\partial \boldsymbol{z}} =& \frac{1}{N_* \hat{\sigma}_e^2} \frac{\partial S}{\partial \boldsymbol{z}} \\ =& \frac{2}{N_* \hat{\sigma}_e^2} \Big\{ \boldsymbol{L}\boldsymbol{H}_1(\boldsymbol{B}^+ - (\boldsymbol{B}+\boldsymbol{A}_0)^{-1})\boldsymbol{H}_1\boldsymbol{L}\boldsymbol{z} - \boldsymbol{L}\boldsymbol{H}_1(\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\sum_i n_i \overline{\boldsymbol{x}}_i \overline{y}_i \Big\}, \\ \frac{\partial \widehat{\psi}}{\partial \overline{y}_i} =& \frac{2}{N_* \hat{\sigma}_e^2} \{ n_i(\overline{y}_i - \overline{\boldsymbol{x}}_i' \widehat{\boldsymbol{\beta}}_2) - n_i \overline{\boldsymbol{x}}_i' (\boldsymbol{I} - (\boldsymbol{B}+\boldsymbol{A}_0)^{-1}\boldsymbol{A}_0)(\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2) \},\end{aligned}$$

so that we can derive the second partial derivatives given in Lemma 4.1. ■

The following lemmas can be shown directly from Lemma 4.1.

Lemma 4.2 *Assume that* $\boldsymbol{B} = O(k)$ *and* $\boldsymbol{A}(\psi) = O(k)$ *as* $k \to \infty$. *Then, as* $k \to \infty$,

$$\frac{\partial^2 \widehat{\psi}}{\partial \boldsymbol{z} \partial \boldsymbol{z}'} = O_p(1), \quad \frac{\partial^2 \widehat{\psi}}{\partial \boldsymbol{z} \partial \overline{y}_i} = O_p(k^{-1}), \quad \frac{\partial^2 \widehat{\psi}}{\partial \overline{y}_i^2} = O_p(k^{-1}), \quad \frac{\partial^2 \widehat{\psi}}{\partial \overline{y}_i \partial \overline{y}_j} = O_p(k^{-2}), \ (i \neq j).$$

Lemma 4.3 *Assume that* $\boldsymbol{B} = O(k)$ *and* $\boldsymbol{A}(\psi) = O(k)$ *as* $k \to \infty$. *Then, as* $k \to \infty$,

$$\begin{aligned} &E[\frac{\partial \widehat{\psi}}{\partial \boldsymbol{z}} \frac{\partial \widehat{\psi}}{\partial \boldsymbol{z}'}] = O(k^{-1}), \quad E[\frac{\partial \widehat{\psi}}{\partial \boldsymbol{z}} \frac{\partial \widehat{\psi}}{\partial \overline{y}_i}] = O(k^{-2}), \\ &E[(\frac{\partial \widehat{\psi}}{\partial \overline{y}_i})^2] = O(k^{-2}), \quad E[\frac{\partial \widehat{\psi}}{\partial \overline{y}_i} \frac{\partial \widehat{\psi}}{\partial \overline{y}_j}] = O(k^{-3}), \ (i \neq j). \end{aligned}$$

5 Concluding Remarks

In the problem of testing the linear hypothesis on regression coefficients in the nested error regression model, the standard F-test based on the OLS method is known to have the serious shortcoming of having inflated type I error rates (sizes) due to the intraclass correlation structure. To fix this problem, in this paper, we have obtained the exact test F_{EXT} and the asymptotically corrected GLS test F_{ACG}. Through some simulation studies, we have shown that the two tests F_{EXT} and F_{ACG} have sizes close to nominal levels and that the size of F_{ACG} is slightly better than Rao *et al.*'s GLS test F_{RSY}. Also it is shown that F_{ACG} is more powerful than F_{EXT}. Thus we can recommend the test F_{ACG} from the numerical results.

Acknowledgments. The research of the first author was supported in part by a grant from the Ministry of Education, Japan, No. 16500172 and in part by a grant from the 21st Century COE Program at Faculty of Economics, University of Tokyo.

References

[1] Fuller, W.A., and Battese, G.E. (1973). Transformations for estimation of linear models with nested-error structures. *Journal of the American Statistical Association*, **68**, 626-632.

[2] Kackar, R.N., and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, **79**, 853-862.

[3] Kubokawa, T. and Srivastava, M.S. (2007). *Akaike information criterion for selecting variables in a nested error regression model.* Discussion Paper Series, CIRJE-F-525, Faculty of Economics, University of Tokyo.

[4] Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163-171.

[5] Rao, J.N.K, Sutradhar, B.C., and Yue, K. (1993). Generalized least squares F test in regression analysis with two-stage cluster samples. *Journal of the American Statistical Association*, **88**, 1388-1391.

[6] Rao, J.N.K, and Wang, S.-G. (1995). On the power of F tests under regression models with nested error structure. *Journal of Multivariate Analysis*, **53**, 237-246

[7] Wu, C.F.J, Holt, D., and Holmes, D.J. (1988). The effect of two-stage sampling on the F statistic. *Journal of the American Statistical Association*, **83**, 150-159.