CIRJE-F-884

Bartlett Adjustments for Hypothesis Testing in Linear Models with General Error Covariance Matrices

Masahiro Kojima
Graduate School in Economics, University of Tokyo

Tatsuya Kubokawa
University of Tokyo

April 2013

Bartlett Adjustments for Hypothesis Testing in Linear Models with General Error Covariance Matrices

Masahiro Kojima* and Tatsuya Kubokawa†

University of Tokyo

April 2, 2013

Abstract

Consider the problem of testing a linear hypothesis of regression coefficients in a general linear regression model with an error term having a covariance matrix involving several nuisance parameters. Three typical test statistics of Wald, Score and Likelihood Ratio (LR) and their Bartlett adjustments have been derived in the literature when the unknown nuisance parameters are estimated by maximum likelihood (ML). On the other hand, statistical inference in linear mixed models has been studied actively and extensively in recent years with applications to small-area estimation. The marginal distribution of the linear mixed model is included in the framework of the general linear regression model, and the nuisance parameters correspond to the variance components and others in the linear mixed model. Although the restricted ML (REML), minimum norm quadratic unbiased estimator (MINQUE) and other specific estimators are available for estimating the variance components, the Bartlett adjustments given in the literature are not correct for those estimators other than ML.

In this paper, using the Taylor series expansion, we derive the Bartlett adjustments of the Wald, Score and modified LR tests for general consistent estimators of the unknown nuisance parameters. These analytical results may be harder to calculate for a model with a complicate structure of the covariance matrix. Thus, we propose the simple parametric bootstrap methods for estimating the Bartlett adjustments and show that they have the second order accuracy. Finally, it is shown that both Bartlett adjustments work well through simulation experiments in the nested error regression model.

Key words and phrases: Asymptotic power function, Bartlett adjustment, general consistent estimator, likelihood Ratio(LR) test, linear mixed model, linear regression model, nested error regression model(NERM), parametric Bootstrap, re-

*Graduate School in Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: kojimasahi@gmail.com

†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: tatsuya@e.u-tokyo.ac.jp

stricted maximum likelihood(REML) estimator, restricted general consistent estimator, score test, Wald test.

1 Introduction

The linear mixed models (LMM) and the empirical best linear unbiased predictor (EBLUP) have been actively studied and extensively recognized useful in many applications including small-area estimation. Especially, the problem of selecting the explanatory variables has received much attention in recent years, and the Akaike Information Criterion (AIC) and the conditional AIC have been studied by Vaida and Blanchard (2005), Srivastava and Kubokawa (2010), Kubokawa (2011b) and others. To investigate whether the selected variables are significant, we need to consider the hypothesis testing for regression parameters. The standard F test statistic based on the ordinary least squares statistics is known to have a serious drawback of having incorrect type I error (size). To fix this problem, Wu, Holt and Holmes (1988), Rao, Sutradhar and Yue (1993), Rao and Wang (1995) and Kubokawa and Erdembat (2010) proposed modified procedures, but they are not guaranteed to have second-order corrections analytically. Thus, we want to derive Bartlett corrections so that the adjusted test statistics have second-order corrections in type I errors (Bartlett (1937)).

This problem was resolved by Rothenberg (1984) in a general linear regression model with an error term having a covariance matrix involving several unknown nuisance parameters, since the marginal distribution of the linear mixed model is in the framework of the general linear regression model. In fact, Rothenberg (1984) derived Bartlett adjustments of the Wald, Score and Likelihood Ratio (LR) test statistics in the general linear regression model. However, their Bartlett adjustments are limited to the case that the nuisance parameters are estimated by Maximum Likelihood (ML). In the linear mixed model with variance components, one can use the Restricted ML (REML), Minimum Norm Quadratic Unbiased Estimator (MINQUE) and other specific estimators for the variance components. For example, simple estimators proposed by Prasad and Rao (1990) and Fay and Herriot (1979) are available for specific linear mixed models. However, the Bartlett adjustments given by Rothenberg (1984) are not correct for those estimators other than ML.

In this paper, we consider an extension of the results of Rothenberg (1984), which is based on ML, to the case of general consistent estimators. That is, we treat the classical Wald, Score and LR test statistics based on the general consistent estimators of the nuisance parameters, and we want to derive their Bartlett corrections. However, we are faced with the following difficulties:

(I) The null hypothesis is a linear constraint of the regression coefficients, and the score and LR test statistics use an estimator which restricts the general consistent estimators on the linear constraint. How can we construct such an estimator restricting the general estimator on the null hypothesis?

(II) When we substitute the general consistent estimators instead of ML into the LR test statistic, the corresponding Bartlett adjustment produces many additional terms, which implies that the Bartlett adjustment yields a large variance.

(III) The Bartlett adjustments are hard to compute in a model with a covariance matrix having a complicate structure, since they include various kinds of differentiations of the covariance matrix with respect to unknown parameters, and the second-order bias and variance of the general consistent estimators. In general, it is a painful task to derive the second-order bias of consistent estimators analytically.

For the query (I), we suggest the use of an equation induced from the Taylor series expansion of the likelihood function. For maximum likelihood estimators, a relation between the unrestricted ML and the ML restricted on the null hypothesis can be expressed as an explicit equation through the likelihood function. To construct a restricted estimator for the general consistent estimator, we use the same equation where the consistent estimator is substituted instead of ML.

For (II), we suggest the use of the modified LR which is defined by the average of the Wald and the score test statistics. The modified LR is asymptotically identical to the original LR when the nuisance parameters are estimated by ML.

For (III), we propose the parametric bootstrap method for estimating the Bartlett adjustments of the three test statistics and show that this approach guarantees the second-order correction. Using the parametric bootstrap, we do not have to derive differentiations of the covariance matrix and the second-order bias and variance of the general consistent estimators. Rayner (1990) proposed another types of the parametric bootstrap methods for estimating the Bartlett corrections for the three test statistics, where the estimators of the nuisance parameters are limited to ML. Although his approach works for LR test, we cannot obtain the second-order adjustments of the Wald and score test statistics via his parametric bootstrap method.

The paper is organized as follows: In Section 2, we propose three kinds of test statistics for linear hypothesis of regression parameters in a general linear regression model. These test statistics are the Wald, score and modified likelihood ratio tests based on the general consistent estimators. In Section 3, we give analytical expressions of the Bartlett adjustments for those three tests. The parametric bootstrap methods for estimating the Bartlett corrections are given in Section 4. Section 5 gives an application to the nested error regression model, which has been used in the context of the small-area estimation. In Section 6, we investigate numerical performances of the proposed three test statistics modified by the Bartlett adjustments by simulation. It is shown that the type I errors (size) for the test statistics with Bartlett adjustments are much improved. The paper is concluded in Section 7, and all the proofs are given in Appendix.

2 Test Statistics in a General Linear Regression Model

Consider the general linear regression model given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\boldsymbol{y}$ is an $N \times 1$ vector of observations, $\boldsymbol{X}$ is an $N \times p$ ($N \geq p$) known matrix of explanatory variables with rank p, $\boldsymbol{\beta}$ is a $p \times 1$ unknown vector of regression coefficients, $\boldsymbol{\epsilon}$ is an $N \times 1$ vector of random errors having $\mathcal{N}_N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ for a q-dimensional vector $\boldsymbol{\theta}$ of unknown parameters.

Let $\boldsymbol{R}$ be an $r \times p$ ($p \geq r$) constant matrix with rank r and let $\boldsymbol{r}$ be an r-dimensional constant vector. Then we consider the problem of testing the null hypothesis H_0 against the alternative hypothesis H_1, namely

$$H_0 \ : \ \boldsymbol{R\beta} = \boldsymbol{r} \quad vs \quad H_1 \ : \ \boldsymbol{R\beta} \neq \boldsymbol{r}. \tag{2.2}$$

Given $\boldsymbol{\theta}$, the generalized least squares estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{y}.$$

The nuisance parameter $\boldsymbol{\theta}$ can be estimated by ML, REML and moment methods. In a variance components model, $\boldsymbol{\theta}$ corresponds to a vector of variance components, which can be estimated by various methods including MINQUE and Henderson's methods. In specific linear mixed models, various specific estimators like the Fay-Herriot estimator are also available. In this paper, we consider a general consistent estimator, denoted by $\widehat{\boldsymbol{\theta}}$, for $\boldsymbol{\theta}$. Then $\boldsymbol{\beta}$ is estimated by $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$.

For the testing problem (2.2), we look at the three classical test statistics based on the consistent estimator $\widehat{\boldsymbol{\theta}}$.

[1] **Wald test.** The Wald test statistic is given by $W = V(\widehat{\boldsymbol{\theta}})$, where

$$V(\boldsymbol{\theta}) = (\boldsymbol{R}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{r})'(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{r}).$$

[2] **Score test.** The score test statistic is described by $S = V(\widetilde{\boldsymbol{\theta}})$, where $\widetilde{\boldsymbol{\theta}}$ is a restricted estimator of $\boldsymbol{\theta}$ induced from $\widehat{\boldsymbol{\theta}}$ under the constraint $\boldsymbol{R\beta} = \boldsymbol{r}$. The problem is how we should construct the consistent and restricted estimator from $\widehat{\boldsymbol{\theta}}$. An important point is that we should take the restricted estimator $\widetilde{\boldsymbol{\theta}}$ so that $\widetilde{\boldsymbol{\theta}}$ is independent of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ under the constraint $\boldsymbol{R\beta} = \boldsymbol{r}$. Otherwise, it may be very difficult to evaluate the moments for deriving the Bartlett corrections. We here suggest the use of the restricted estimator given by

$$\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}) \begin{pmatrix} V_{(1)}(\widehat{\boldsymbol{\theta}}) \\ \vdots \\ V_{(q)}(\widehat{\boldsymbol{\theta}}) \end{pmatrix}, \tag{2.3}$$

where $V_{(i)}(\boldsymbol{\theta}) = (\partial/\partial\theta_i)V(\boldsymbol{\theta})$ and $\boldsymbol{\Lambda}(\boldsymbol{\theta}) = 2^{-1}E[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'] + o(N^{-1})$. This equation is motivated from the relation between the unrestricted maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{M}}$ and the restricted ML $\widetilde{\boldsymbol{\theta}}_{\mathrm{M}}$ under the constraint $\boldsymbol{R\beta} = \boldsymbol{r}$. That is, the restricted ML is expressed based on the unrestricted ML $\widehat{\boldsymbol{\theta}}_{\mathrm{M}}$ as

$$\widetilde{\boldsymbol{\theta}}_{\mathrm{M}} = \widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}) \begin{pmatrix} V_{(1)}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}) \\ \vdots \\ V_{(q)}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}) \end{pmatrix},$$

where the derivation of this equation is given in Appendix. Along this line, we define the restricted estimator induced from $\widehat{\boldsymbol{\theta}}$ by (2.3). It is noted that $\widetilde{\boldsymbol{\theta}}$ is independent of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ if $\widehat{\boldsymbol{\theta}}$ is independent of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$.

[3] Modified likelihood ratio test. The likelihood ratio (LR) test statistic is defined by

$$LR(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}) = -2\{L(\widetilde{\boldsymbol{\beta}}(\widetilde{\boldsymbol{\theta}}),\widetilde{\boldsymbol{\theta}}) - L(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}),\widehat{\boldsymbol{\theta}})\},$$

where $L(\boldsymbol{\beta},\boldsymbol{\theta})$ is the log likelihood function, and $\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta})$ is the restricted general least squares estimator satisfying $\boldsymbol{R}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{r}$. In general, the Bartlett correction of $LR(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$ produces many additional terms which can vanish in the case of ML $\widehat{\boldsymbol{\theta}}_{\mathrm{M}}$ and $\widetilde{\boldsymbol{\theta}}_{\mathrm{M}}$. This implies that the Bartlett adjustment of $LR(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$ yields a larger variance. Thus, we suggest the use of the modified LR test statistic

$$mLR = mLR(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}) = (W+S)/2 = (V(\widehat{\boldsymbol{\theta}}) + V(\widetilde{\boldsymbol{\theta}}))/2, \tag{2.4}$$

which is the average of the Wald and score test statistics. It is noted that mLR is asymptotically identical to LR for ML, namely

$$LR(\widehat{\boldsymbol{\theta}}_{\mathrm{M}},\widetilde{\boldsymbol{\theta}}_{\mathrm{M}}) = (V(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}) + V(\widetilde{\boldsymbol{\theta}}_{\mathrm{M}}))/2 + o_p(N^{-1}) = mLR(\widehat{\boldsymbol{\theta}}_{\mathrm{M}},\widetilde{\boldsymbol{\theta}}_{\mathrm{M}}) + o_p(N^{-1}).$$

Under the null hypothesis, all the three test statistics are asymptotically distributed as the chi-square distribution χ^2_r with r degrees of freedom, since both estimators $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ are consistent. However, as numerically shown in Section 6, all the three testing procedures have the incorrect type I errors. Thus, we need to derive the Bartlett corrections for the three test statistics in the next section.

3 Bartlett Adjustments via Analytical Approach

We now derive the Bartlett corrections for the test statistics W, S and mLR, so that the type I errors of the corresponding tests with the Bartlett adjustments are identical to the nominal significance level up to the second-order $O(N^{-1})$. To this end, we use the following notations

$$\begin{aligned}
\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)_{(i)} =& \frac{\partial\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)}{\partial\theta_i}, \quad (\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1})_{(ij)} = \frac{\partial\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)}{\partial\theta_i\partial\theta_j},\\
\mathbf{col}(a_i) =& \begin{pmatrix} a_1 \\ \vdots \\ a_q \end{pmatrix}, \quad \mathbf{mat}(a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & & \vdots \\ a_{q1} & \cdots & a_{qq} \end{pmatrix},\\
\boldsymbol{\nabla} =& (\partial/\partial y_1,\ldots,\partial/\partial y_N)'.
\end{aligned}$$

Without confusion, we use the simple notations $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$ by dropping $(\boldsymbol{\theta})$ in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$.

The aim of this paper is to extend the results of Rothenberg (1984), which is limited to ML for $\boldsymbol{\theta}$, to the case of the general consistent estimators $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$. We assume the following conditions for $\boldsymbol{\theta}$.

Assumption 1

[1] $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ satisfies that $\widehat{\boldsymbol{\theta}}(-\boldsymbol{y}) = \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ and $\widehat{\boldsymbol{\theta}}(\boldsymbol{y} + \boldsymbol{X\alpha}) = \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ for any p-dimensional vector $\boldsymbol{\alpha}$.

[2] $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is expanded as

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{\dagger} + \widehat{\boldsymbol{\theta}}^{\dagger\dagger} + o_p(N^{-1}),$$

where $\widehat{\boldsymbol{\theta}}^{\dagger} = (\hat{\theta}_1^{\dagger}, \ldots, \hat{\theta}_q^{\dagger})' = O_p(N^{-1/2})$, $\widehat{\boldsymbol{\theta}}^{\dagger\dagger} = (\hat{\theta}_1^{\dagger\dagger}, \ldots, \hat{\theta}_q^{\dagger\dagger})' = O_p(N^{-1})$ and every element of $\boldsymbol{\Sigma}\boldsymbol{\nabla}\hat{\theta}_i^{\dagger}$ is of $O_p(N^{-1})$.

We begin by looking at the Wald test statistic W. Since $W = V(\widehat{\boldsymbol{\theta}})$ is based on the unrestricted consistent estimator $\widehat{\boldsymbol{\theta}}$, the second-order correction terms are easier to calculate than the other tests S and mLR. It is noted that W can be decomposed as

$$W = V(\widehat{\boldsymbol{\theta}}) = (\boldsymbol{x} + \boldsymbol{s})' (\boldsymbol{I} - \boldsymbol{S})^{-1} (\boldsymbol{x} + \boldsymbol{s}), \tag{3.1}$$

where $\boldsymbol{x}$, $\boldsymbol{s}$ and $\boldsymbol{S}$ are denoted by

$$\begin{aligned}
\boldsymbol{x} =& \boldsymbol{H}\big\{\boldsymbol{R}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{r}\big\},\\
\boldsymbol{s} =& \boldsymbol{H}\boldsymbol{R}\big\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})\big\},\\
\boldsymbol{S} =& -\boldsymbol{H}\boldsymbol{R}\big\{(\boldsymbol{X}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{X})^{-1} - (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\big\}\boldsymbol{R}'\boldsymbol{H},
\end{aligned}$$

for $\boldsymbol{H} = (\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1/2}$. It is noted that $\boldsymbol{s}$ and $\boldsymbol{S}$ are location-invariant statistic since $\widehat{\boldsymbol{\theta}}$ is a location-invariant from Assumption 1 [1]. Then form Kackar and Harville (1984), it is seen that $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and that $\boldsymbol{x}$ is independent of $(\widehat{\boldsymbol{\theta}}, \boldsymbol{s}, \boldsymbol{S})$. In fact, the independence follows from Basu's theorem and Assumption 1 [1], since $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ is a complete sufficient statistic for $\boldsymbol{\beta}$, and $(\widehat{\boldsymbol{\theta}}, \boldsymbol{s}, \boldsymbol{S})$ is an ancillary statistic for $\boldsymbol{\beta}$. The independence is useful for evaluating moments in order to derive the Bartlett correction.

The score test is defined by $S = V(\widetilde{\boldsymbol{\theta}})$ for a restricted estimator $\widetilde{\boldsymbol{\theta}}$. In general, a restricted estimator cannot be guaranteed to be independent of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$, which implies that the moments for deriving the Bartlett correction are difficult to evaluate. Thus, we use the restricted estimator given in (2.3), namely,

$$\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}})\mathbf{col}(V_{(i)}(\widehat{\boldsymbol{\theta}})).$$

Since $\widetilde{\boldsymbol{\theta}}$ is a function of $\widehat{\boldsymbol{\theta}}$, $\widetilde{\boldsymbol{\theta}}$ is still independent of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$. Since $\boldsymbol{\Lambda}(\boldsymbol{\theta}) = O(N^{-1})$, it is seen that $\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}} = O_p(N^{-1})$. Using the Taylor series expansion and the relation (2.3), we can approximate the score test statistic as

$$\begin{aligned}
S = V(\widetilde{\boldsymbol{\theta}}) &= W + \mathbf{col}(V_{(i)}(\widehat{\boldsymbol{\theta}}))'(\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}) + o_p(N^{-1})\\
&= W - \mathbf{col}(V_{(i)}(\widehat{\boldsymbol{\theta}}))'\boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}})\mathbf{col}(V_{(i)}(\widehat{\boldsymbol{\theta}})) + o_p(N^{-1}),
\end{aligned}$$

since $W = V(\widehat{\boldsymbol{\theta}})$. Similarly, the modified LR test statistic defined in (2.4) is expressed as

$$mLR = (W+S)/2 = W - \frac{1}{2}\mathbf{col}(V_{(i)}(\widehat{\boldsymbol{\theta}}))\boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}})\mathbf{col}(V_{(i)}(\widehat{\boldsymbol{\theta}})) + o_p(N^{-1}).$$

The cumulative distribution functions of the test statistics $P(W \le x)$, $P(LR \le x)$, $P(S \le x)$ are approximated as $F_r(x) + O(N^{-1})$ under the null hypothesis for a distribution function $F_r(x)$ of the χ^2_r distribution. This means that the significance levels of the three test statistics are not identical to the nominal significance $\chi^2_{r,\alpha}$ in the sense of second-order. Thus, we need to derive their Bartlett corrections. Define $a(\boldsymbol{\theta})$, $b(\boldsymbol{\theta})$ and $c(\boldsymbol{\theta})$ by

$$\begin{aligned} a(\boldsymbol{\theta}) =& E_{H_0}\left[\boldsymbol{s}'\boldsymbol{s}\right], \\ b(\boldsymbol{\theta}) =& \frac{1}{2}E_{H_0}\Big[\mathrm{tr}\ \left(\boldsymbol{S}^2\right) + \frac{1}{2}\left(\mathrm{tr}\ (\boldsymbol{S})\right)^2\Big], \\ c(\boldsymbol{\theta}) =& E_{H_0}\left[\mathrm{tr}\ (\boldsymbol{S}) + \mathrm{tr}\ \left(\boldsymbol{S}^2\right)\right], \end{aligned} \tag{3.2}$$

where $E_{H_0}[\cdot]$ denotes the expectation under the null hypothesis. Based on these functions, we obtain the test statistics with the Bartlett adjustments, given by

$$\begin{aligned} W^{\mathrm{BC}} &= W\left(1 - \frac{1}{r}\left(a(\widehat{\boldsymbol{\theta}}) - b(\widehat{\boldsymbol{\theta}}) + c(\widehat{\boldsymbol{\theta}})\right) - \frac{x}{r(r+2)}b(\widehat{\boldsymbol{\theta}})\right), \\ mLR^{\mathrm{BC}} &= mLR\left(1 - \frac{1}{r}\left(-b(\widehat{\boldsymbol{\theta}}) + c(\widehat{\boldsymbol{\theta}})\right)\right), \\ S^{\mathrm{BC}} &= S\left(1 - \frac{1}{r}\left(-a(\widehat{\boldsymbol{\theta}}) - b(\widehat{\boldsymbol{\theta}}) + c(\widehat{\boldsymbol{\theta}})\right) + \frac{x}{r(r+2)}b(\widehat{\boldsymbol{\theta}})\right). \end{aligned} \tag{3.3}$$

To establish the second-order accuracy of the Bartlett adjustments, we assume the following conditions:

Assumption 2

[**1**] The elements of $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{G}(\boldsymbol{\theta})$, $\boldsymbol{R}(\boldsymbol{\theta})$, are uniformly bounded, and p, r, M are bounded.

[**2**] $\boldsymbol{X}'\boldsymbol{X}/N$, $\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{X}/N$, $\boldsymbol{X}'\boldsymbol{Z}^2\boldsymbol{X}/N$, $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}/N$, $\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1})_{(i)}\boldsymbol{X}/N$, $\boldsymbol{X}'\left(\boldsymbol{\Sigma}^{-1}\right)_{(ij)}\boldsymbol{X}/N$ and $\boldsymbol{X}'\left(\boldsymbol{\Sigma}^{-1}\right)_{(i)}\boldsymbol{\Sigma}(\boldsymbol{\theta})\left(\boldsymbol{\Sigma}^{-1}\right)_{(j)}\boldsymbol{X}/N$ converge to finite matrices as $N \to \infty$.

[**3**] $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon} = O_p(N^{1/2})$ and $\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1})_{(i)}\boldsymbol{\epsilon} = O_p(N^{1/2})$ for $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$.

Define $\boldsymbol{A}_{(i)}$, $\boldsymbol{B}_{(i)}$ and $\boldsymbol{B}_{(ij)}$ by

$$\begin{aligned} \boldsymbol{A}_{(i)} =& \boldsymbol{H}\boldsymbol{R}((\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1})_{(i)}, \\ \boldsymbol{B}_{(i)} =& -\boldsymbol{H}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}_{(i)}\boldsymbol{R}'\boldsymbol{H}, \\ \boldsymbol{B}_{(ij)} =& -\boldsymbol{H}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}_{(ij)}\boldsymbol{R}'\boldsymbol{H}. \end{aligned} \tag{3.4}$$

Then we get the following theorem which will be proved in Appendix.

Theorem 3.1 *Let $\widehat{\boldsymbol{\theta}}$ be a general consistent estimator satisfying Assumption 1. Let $\widetilde{\boldsymbol{\theta}}$ be the restricted estimator defined in* (2.3). *Under Assumption 2, the cumulative distribution functions of the Wald, score and modified likelihood ratio test statistics with the Bartlett adjustments $P\left(W^{BC} \leq x\right)$, $P\left(mLR^{BC} \leq x\right)$ and $P\left(S^{BC} \leq x\right)$ are approximated as $F_r(x) + o(N^{-1})$ under the null hypothesis. Also, the functions $a(\boldsymbol{\theta})$, $b(\boldsymbol{\theta})$ and $c(\boldsymbol{\theta})$ are approximated as*

$$
\begin{aligned}
a(\boldsymbol{\theta}) =& \sum_{i,j}^{q} \mathrm{tr}\,(\boldsymbol{\Sigma} \boldsymbol{A}'_{(i)} \boldsymbol{A}_{(j)}) E_{H_0}[\hat{\theta}_i^{\dagger} \hat{\theta}_j^{\dagger}] + o(N^{-1}), \\
b(\boldsymbol{\theta}) =& \frac{1}{2} \sum_{i,j}^{q} \Big\{ \mathrm{tr}\,(\boldsymbol{B}_{(i)} \boldsymbol{B}_{(j)}) + \frac{1}{2} \mathrm{tr}\,(\boldsymbol{B}_{(i)}) \mathrm{tr}\,(\boldsymbol{B}_{(j)}) \Big\} E_{H_0}[\hat{\theta}_i^{\dagger} \hat{\theta}_j^{\dagger}] + o(N^{-1}), \qquad (3.5) \\
c(\boldsymbol{\theta}) =& \sum_{i}^{q} \mathrm{tr}\,(\boldsymbol{B}_{(i)}) E_{H_0}[\hat{\theta}_i^{\dagger} + \hat{\theta}_i^{\dagger\dagger}] + \sum_{i,j}^{q} \Big\{ \frac{1}{2} \mathrm{tr}\,(\boldsymbol{B}_{(ij)}) + \mathrm{tr}\,(\boldsymbol{B}_{(i)} \boldsymbol{B}_{(j)}) \Big\} E_{H_0}[\hat{\theta}_i^{\dagger} \hat{\theta}_j^{\dagger}] + o(N^{-1}),
\end{aligned}
$$

which are of order $O(N^{-1})$.

The estimators $a(\widehat{\boldsymbol{\theta}})$, $b(\widehat{\boldsymbol{\theta}})$ and $c(\widehat{\boldsymbol{\theta}})$ are provided by substituting the estimator $\widehat{\boldsymbol{\theta}}$ into the r.h.s. of (3.5). For this purpose, we need to derive the second-order bias $E_{H_0}[\hat{\theta}_i^{\dagger} + \hat{\theta}_i^{\dagger\dagger}]$, the limiting values of covariance $Cov_{H_0}(\hat{\theta}_i, \hat{\theta}_j)$ or $E_{H_0}[\hat{\theta}_i^{\dagger} \hat{\theta}_j^{\dagger}]$, and $\boldsymbol{A}_{(i)}$, $\boldsymbol{B}_{(i)}$ and $\boldsymbol{B}_{(ij)}$.

We next provide second-order approximations of the power functions of the Bartlett-adjustment test statistics W^{BC}, mLR^{BC} and S^{BC} at the point $\boldsymbol{\delta} = \boldsymbol{R\beta} - \boldsymbol{r}$. Let $\boldsymbol{A}(\boldsymbol{\theta}) = E_{\boldsymbol{\delta}}\left[\boldsymbol{s}\boldsymbol{s}'\right]$, $\boldsymbol{B}(\boldsymbol{\theta}) = 2^{-1} E_{\boldsymbol{\delta}}\left[\boldsymbol{S}^2 + 2^{-1} \boldsymbol{S} \mathrm{tr}\,(\boldsymbol{S})\right]$, $\boldsymbol{C}(\boldsymbol{\theta}) = E_{\boldsymbol{\delta}}\left[\boldsymbol{S} + \boldsymbol{S}^2\right]$ and $\boldsymbol{D}(\boldsymbol{\theta}) = E_{\boldsymbol{\delta}}\left[\boldsymbol{S\Delta S}\right]$ for $\boldsymbol{\Delta} = \boldsymbol{\delta\delta}'$, where $E_{\boldsymbol{\delta}}[\cdot]$ denotes the expectation under the alternative hypothesis at $\boldsymbol{\delta} = \boldsymbol{R\beta} - \boldsymbol{r}$. We use the notations given by

$$
\begin{aligned}
\lambda_1(\widehat{\boldsymbol{\theta}}) =& \mathrm{tr}\, \Big(\boldsymbol{A}(\widehat{\boldsymbol{\theta}}) \boldsymbol{\Delta} \Big), \quad \lambda_3(\widehat{\boldsymbol{\theta}}) = \mathrm{tr}\, \Big(\boldsymbol{D}(\widehat{\boldsymbol{\theta}}) \boldsymbol{\Delta} \Big), \\
\lambda_2(\widehat{\boldsymbol{\theta}}) =& r^{-1} \Big(-b(\widehat{\boldsymbol{\theta}}) + c(\widehat{\boldsymbol{\theta}}) \Big) \mathrm{tr}\, (\boldsymbol{\Delta}) + \mathrm{tr}\, \Big(\Big(\boldsymbol{B}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{C}(\widehat{\boldsymbol{\theta}}) \Big) \boldsymbol{\Delta} \Big), \\
\lambda'_2(\widehat{\boldsymbol{\theta}}) =& r^{-1} a(\widehat{\boldsymbol{\theta}}) \mathrm{tr}\, (\boldsymbol{\Delta}) - \mathrm{tr}\, \Big(\boldsymbol{A}(\widehat{\boldsymbol{\theta}}) \boldsymbol{\Delta} \Big), \qquad (3.6) \\
\lambda'_3(\widehat{\boldsymbol{\theta}}) =& r^{-1} b(\widehat{\boldsymbol{\theta}}) \mathrm{tr}\, (\boldsymbol{\Delta}) - \mathrm{tr}\, \Big(\boldsymbol{C}(\widehat{\boldsymbol{\theta}}) \boldsymbol{\Delta} \Big), \\
\lambda_4(\widehat{\boldsymbol{\theta}}) =& (r(r+2))^{-1} b(\widehat{\boldsymbol{\theta}}) \left(\mathrm{tr}\, (\boldsymbol{\Delta})\right)^2 - \mathrm{tr}\, \Big(\boldsymbol{D}(\widehat{\boldsymbol{\theta}}) \boldsymbol{\Delta} \Big).
\end{aligned}
$$

Then we obtain the following theorem which will be proved in Appendix.

Theorem 3.2 *Under the same assumptions as in Theorem 3.1, the power functions of the test statistics mLR^{BC}, W^{BC} and S^{BC} at $\boldsymbol{\delta} = \boldsymbol{R\beta} - \boldsymbol{r}$ are approximated as*

$$
\begin{aligned}
P(mLR^{BC} > x) &= 1 - G_r(x) - \lambda_1(\widehat{\boldsymbol{\theta}}) g_{r+2}(x) - \lambda_2(\widehat{\boldsymbol{\theta}}) g_{r+4}(x) - \lambda_3(\widehat{\boldsymbol{\theta}}) g_{r+6} + o(N^{-1}), \\
P(W^{BC} > x) &= P(mLR^{BC} > x) - \lambda'_2(\widehat{\boldsymbol{\theta}}) g_{r+4}(x) - \lambda'_3(\widehat{\boldsymbol{\theta}}) g_{r+6}(x) - \lambda_4(\widehat{\boldsymbol{\theta}}) g_{r+8}(x) + o(N^{-1}), \\
P(S^{BC} > x) &= P(mLR^{BC} > x) + \lambda'_2(\widehat{\boldsymbol{\theta}}) g_{r+4}(x) + \lambda'_3(\widehat{\boldsymbol{\theta}}) g_{r+6}(x) + \lambda_4(\widehat{\boldsymbol{\theta}}) g_{r+8}(x) + o(N^{-1}),
\end{aligned}
$$

where $G_r(x)$ and $g_r(x)$ are the cumulative distribution and the probability density functions, respectively, of a noncentral chi-squared random variable $\chi_r^2(\boldsymbol{\delta}'\boldsymbol{\delta})$ with r degrees of freedom and the noncentrality $\boldsymbol{\delta}'\boldsymbol{\delta}$.

4 Bartlett Adjustments via Parametric Bootstrap

As stated below Theorem 3.1, we need to derive the second-order bias and the covariance matrix of $\widehat{\boldsymbol{\theta}}$. It is not easy to derive these moments for ML and REML of $\boldsymbol{\theta}$, and it is hard to calculate $a(\boldsymbol{\theta})$, $b(\boldsymbol{\theta})$ and $c(\boldsymbol{\theta})$ for complicated models. A simple and useful method for estimating the Bartlett adjustments is the parametric bootstrap.

Let $\widehat{\boldsymbol{\theta}}$ be a general consistent estimator of $\boldsymbol{\theta}$ based on $\boldsymbol{y}$, and let $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$ be, respectively, the generalized least squares estimator of $\boldsymbol{\beta}$ based on $\boldsymbol{y}$ and its restricted estimator under the null hypothesis, where

$$\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{r}). \tag{4.1}$$

We first generate the parametric bootstrap sample under the null hypothesis. An $N \times 1$ random vector $\boldsymbol{y}^*$ given $\boldsymbol{y}$ has the general linear regression model

$$\boldsymbol{y}^* = \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) + \boldsymbol{\epsilon}^*, \tag{4.2}$$

where $\boldsymbol{X}$ is the same matrix as given in (2.1), and given $\boldsymbol{y}$, $\boldsymbol{\epsilon}^*$ is conditionally distributed as $\boldsymbol{\epsilon}^*|\boldsymbol{y} \sim \mathcal{N}_N(\boldsymbol{0}, \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}))$. Let $\widehat{\boldsymbol{\theta}}^*$ be a general consistent estimator of $\widehat{\boldsymbol{\theta}}$, where the calculation of $\widehat{\boldsymbol{\theta}}^*$ is the same as that of $\widehat{\boldsymbol{\theta}}$ except that $\widehat{\boldsymbol{\theta}}^*$ is calculated based on $\boldsymbol{y}^*$ instead of $\boldsymbol{y}$.

Define $\boldsymbol{s}^*$ and $\boldsymbol{S}^*$ by

$$\begin{aligned} \boldsymbol{s}^* =& \widehat{\boldsymbol{H}}\boldsymbol{R}\big\{\widehat{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\theta}}^*) - \widehat{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\theta}})\big\}, \\ \boldsymbol{S}^* =& -\widehat{\boldsymbol{H}}\boldsymbol{R}\big\{(\boldsymbol{X}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}^*)^{-1}\boldsymbol{X})^{-1} - (\boldsymbol{X}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{X})^{-1}\big\}\boldsymbol{R}'\widehat{\boldsymbol{H}}, \end{aligned} \tag{4.3}$$

for $\widehat{\boldsymbol{\beta}}^*(\boldsymbol{\theta}) = (\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{y}^*$ and $\widehat{\boldsymbol{H}} = (\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1/2}$. Based on these variables, we can estimate $a(\boldsymbol{\theta})$, $b(\boldsymbol{\theta})$ and $c(\boldsymbol{\theta})$ given in (3.4) with their parametric bootstrap estimators given by

$$\begin{aligned} a^* =& E_*[(\boldsymbol{s}^*)'\boldsymbol{s}^*|\boldsymbol{y}], \\ b^* =& \frac{1}{2}E_*\big[\mathrm{tr}\,\{(\boldsymbol{S}^*)^2\} + \frac{1}{2}(\mathrm{tr}\,\boldsymbol{S}^*)^2|\boldsymbol{y}\big], \\ c^* =& E_*\big[\mathrm{tr}\,(\boldsymbol{S}^*) + \mathrm{tr}\,\{(\boldsymbol{S}^*)^2\}|\boldsymbol{y}\big], \end{aligned} \tag{4.4}$$

where $E_*[\cdot|\boldsymbol{y}]$ denotes the expectation with respect to $\boldsymbol{y}^*$ given $\boldsymbol{y}$. Then, the Wald, the modified likelihood ratio and the score test statistics with the Bartlett adjustments via the parametric bootstrap are described as

$$
\begin{aligned}
W_*^{\mathrm{BC}} &= W\left(1 - \frac{1}{r}\left(a^* - b^* + c^*\right) - \frac{x}{r(r+2)}b^*\right), \\
mLR_*^{\mathrm{BC}} &= mLR(1 - \frac{1}{r}(-b^* + c^*)), \\
S_*^{\mathrm{BC}} &= S(1 - \frac{1}{r}\left(-a^* - b^* + c^*\right) + \frac{x}{r(r+2)}b^*).
\end{aligned}
\tag{4.5}
$$

It is noted that $a^* = a(\widehat{\boldsymbol{\theta}})$, $b^* = b(\widehat{\boldsymbol{\theta}})$ and $c^* = c(\widehat{\boldsymbol{\theta}})$, which are approximated as $a^* = a(\boldsymbol{\theta}) + o(N^{-1})$, $b^* = b(\boldsymbol{\theta}) + o(N^{-1})$ and $c^* = c(\boldsymbol{\theta}) + o(N^{-1})$, since $a(\boldsymbol{\theta}) = O(N^{-1})$, $b(\boldsymbol{\theta}) = O(N^{-1})$ and $c(\boldsymbol{\theta}) = O(N^{-1})$. Thus, we obtain the following theorem.

Theorem 4.1 *Let $\widehat{\boldsymbol{\theta}}$ be a general consistent estimator satisfying Assumption 1. Let $\widetilde{\boldsymbol{\theta}}$ be the restricted estimator defined in* (2.3). *Under Assumption 2, the cumulative distribution functions of the Wald, score and modified likelihood ratio tests statistics with the Bartlett adjustments via the parametric bootstrap $P\left(W_*^{BC} \leq x\right)$, $P\left(mLR_*^{BC} \leq x\right)$ and $P\left(S_*^{BC} \leq x\right)$ are approximated as $F_r(x) + o(N^{-1})$ under the null hypothesis.*

Rayner (1990) proposed another type of the Bartlett adjustments by the parametric bootstrap. Although his approach established that the type I error of the adjusted LR test is identical to the nominal significance level in the second-order, the type I errors of the adjusted Wald and score tests remain second-order terms, namely, their cumulative distributions are approximated as $F_r(x) + O(N^{-1})$. Theorem 4.1 shows that our Bartlett adjustments based on the parametric bootstrap provides the type I errors identical to the nominal significance level in the second-order for all three test statistics.

Corresponding to Theorem 3.2, we can construct the second-order approximations of the power functions based on the parametric bootstrap. To generate the parametric bootstrap sample under the alternative hypothesis at the point $\boldsymbol{\delta} = \boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r}$. An $N \times 1$ random vector $\boldsymbol{y}^*$ given $\boldsymbol{y}$ has the general linear regression model

$$
\boldsymbol{y}^{**} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\delta}}(\widehat{\boldsymbol{\theta}}) + \boldsymbol{\epsilon}^{**}, \tag{4.6}
$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\delta}}(\widehat{\boldsymbol{\theta}})$ is the estimator of $\boldsymbol{\beta}$ satisfying $\boldsymbol{\delta} = \boldsymbol{R}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\delta}}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{r}$, and given $\boldsymbol{y}$, $\boldsymbol{\epsilon}^{**}$ is conditionally distributed as $\boldsymbol{\epsilon}^{**}|\boldsymbol{y} \sim \mathcal{N}_N(\boldsymbol{0}, \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}))$. Let $\widehat{\boldsymbol{\theta}}^{**}$ be a general consistent estimator of $\widehat{\boldsymbol{\theta}}$, where the calculation of $\widehat{\boldsymbol{\theta}}^{**}$ is the same as that of $\widehat{\boldsymbol{\theta}}$ except that $\widehat{\boldsymbol{\theta}}^{**}$ is calculated based on $\boldsymbol{y}^{**}$ instead of $\boldsymbol{y}$.

Let $\boldsymbol{s}^{**}$ and $\boldsymbol{S}^{**}$ be the same as those of $\boldsymbol{s}^*$ and $\boldsymbol{S}^*$ except that the superscript $*$ in $\boldsymbol{s}^*$ and $\boldsymbol{S}^*$ is replaced with $**$. The variables a^{**}, b^{**} and c^{**} are similarly defined. Let $\boldsymbol{A}^{**} = E_{**}\left[\boldsymbol{s}^{**}(\boldsymbol{s}^{**})'|\boldsymbol{y}\right]$, $\boldsymbol{B}^{**} = 2^{-1}E_{**}\left[(\boldsymbol{S}^{**})^2 + 2^{-1}\boldsymbol{S}^{**}\mathrm{tr}\left(\boldsymbol{S}^{**}\right)|\boldsymbol{y}\right]$, $\boldsymbol{C}^{**} = E_{**}\left[\boldsymbol{S}^{**} + (\boldsymbol{S}^{**})^2|\boldsymbol{y}\right]$ and $\boldsymbol{D}^{**} = E_{**}\left[\boldsymbol{S}^{**}\boldsymbol{\Delta}\boldsymbol{S}^{**}|\boldsymbol{y}\right]$ for $\boldsymbol{\Delta} = \boldsymbol{\delta}\boldsymbol{\delta}'$. Corresponding to (3.6), we use the notations given by

$$\begin{aligned}
\lambda_1^{**} =& \mathrm{tr}\left(\boldsymbol{A}^{**}\boldsymbol{\Delta}\right), \quad \lambda_3^{**} = \mathrm{tr}\left(\boldsymbol{D}^{**}\boldsymbol{\Delta}\right), \\
\lambda_2^{**} =& r^{-1}\left(-b^{**}+c^{**}\right)\mathrm{tr}\left(\boldsymbol{\Delta}\right) + \mathrm{tr}\left(\left(\boldsymbol{B}^{**}-\boldsymbol{C}^{**}\right)\boldsymbol{\Delta}\right), \\
\lambda_2^{**\prime} =& r^{-1}a^{**}\mathrm{tr}\left(\boldsymbol{\Delta}\right) - \mathrm{tr}\left(\boldsymbol{A}^{**}\boldsymbol{\Delta}\right), \\
\lambda_3^{**\prime} =& r^{-1}b^{**}\mathrm{tr}\left(\boldsymbol{\Delta}\right) - \mathrm{tr}\left(\boldsymbol{C}^{**}\boldsymbol{\Delta}\right), \\
\lambda_4^{**} =& (r(r+2))^{-1}b^{**}\left(\mathrm{tr}\left(\boldsymbol{\Delta}\right)\right)^2 - \mathrm{tr}\left(\boldsymbol{D}^{**}\boldsymbol{\Delta}\right).
\end{aligned}$$

Theorem 4.2 *Under the same assumptions as in Theorem 4.1, the power functions of the test statistics mLR_*^{BC}, W_*^{BC} and S_*^{BC} at $\boldsymbol{\delta} = \boldsymbol{R\beta} - \boldsymbol{r}$ are approximated as*

$$\begin{aligned}
P(mLR_*^{BC} > x) &= G_r(x) - \lambda_1^{**}g_{r+2}(x) - \lambda_2^{**}f_{r+4}(x) - \lambda_3^{**}f_{r+6} + o_p(N^{-1}), \\
P(W_*^{BC} > x) &= P(mLR_*^{BC} > x) - \lambda_2^{**\prime}f_{r+4}(x) - \lambda_3^{**\prime}g_{r+6}(x) - \lambda_4^{**}g_{r+8}(x) + o_p(N^{-1}), \\
P(S_*^{BC} > x) &= P(mLR_*^{BC} > x) + \lambda_2^{**\prime}f_{r+4}(x) + \lambda_3^{**\prime}g_{r+6}(x) + \lambda_4^{**}g_{r+8}(x) + o_p(N^{-1}).
\end{aligned}$$

The leading terms in the r.h.s. of the approximations given in Theorem 4.2 give us the second-order approximations of the power functions of the adjusted test statistics.

5 An Application to a Linear Mixed Model

An important example of the general linear regression model (2.1) is a linear mixed model given by

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} = \boldsymbol{Zv} + \boldsymbol{u}, \tag{5.1}$$

where $\boldsymbol{y}$ is an $N \times 1$ vector of observation, $\boldsymbol{X}$ is an $N \times p$ design matrix that is fixed, $\boldsymbol{Z}$ is an $N \times M$ design matrix of random effect, $\boldsymbol{\beta}$ is a $p \times 1$ unknown vector of the regression coefficients, which are called fixed effects, $\boldsymbol{v}$ is an $M \times 1$ vector of the random effects distributed as $\mathcal{N}_M(\boldsymbol{0}, \boldsymbol{G}(\boldsymbol{\theta}))$, $\boldsymbol{u}$ is an $N \times 1$ vector of errors distributed as $\mathcal{N}_N(\boldsymbol{0}, \boldsymbol{R}(\boldsymbol{\theta}))$, $\boldsymbol{\theta}$ is a q-dimensional vector of unknown nuisance parameters, and $\boldsymbol{v}$ and $\boldsymbol{u}$ are mutually independent, $\boldsymbol{G}(\boldsymbol{\theta})$ and $\boldsymbol{R}(\boldsymbol{\theta})$ are the positive definite matrices. Then, we express a marginal distribution of $\boldsymbol{y}$ as $\mathcal{N}_N(\boldsymbol{X\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, which is in the framework of (2.1), where

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{R}(\boldsymbol{\theta}) + \boldsymbol{ZG}(\boldsymbol{\theta})\boldsymbol{Z}'.$$

In this section, we describe the test statistics with the Bartlett adjustments in a specific linear mixed model. The model we treat here is the nested error regression model which is often used for two stage sampling and the small area estimation. The model is given by

$$y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \varepsilon_{ij}, \quad \varepsilon_{ij} = v_i + u_{ij}, \quad i = 1, \ldots, k, \ j = 1, \ldots, n_i, \tag{5.2}$$

where $v_i \sim \mathcal{N}(0, \sigma_v^2)$, $u_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, and v_i and u_{ij} are mutually independent. Let $\boldsymbol{X}_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})'$, $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2', \ldots, \boldsymbol{X}_k')'$ and $\boldsymbol{Z} = \text{block diag}(\boldsymbol{J}_{n_1}, \boldsymbol{J}_{n_2}, \ldots, \boldsymbol{J}_{n_k})$, where $\boldsymbol{J}_{n_i} = \boldsymbol{j}_{n_i}\boldsymbol{j}_{n_i}'$ for $\boldsymbol{j}_{n_i} = (1, 1, \ldots, 1)' \in \boldsymbol{R}^{n_i}$. Also, let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$, $\boldsymbol{y} =$

$(\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_k')'$, $\boldsymbol{u}_i = (u_{i1}, u_{i2}, \ldots, u_{in_i})'$, $\boldsymbol{u} = (\boldsymbol{u}_1', \boldsymbol{u}_2', \ldots, \boldsymbol{u}_k')$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_k)$. Then, the model (5.2) is rewritten as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} = \boldsymbol{Z}\boldsymbol{v} + \boldsymbol{u},$$

which shows that the model (5.2) is a special case of (5.1). In this model, $\mathbf{Cov}\,(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = (\text{block diag}(\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \boldsymbol{\Sigma}_2(\boldsymbol{\theta}), \ldots, \boldsymbol{\Sigma}_k(\boldsymbol{\theta})))$ for $\boldsymbol{\theta} = (\theta_1, \theta_2)' = (\sigma_\varepsilon^2, \sigma_v^2)'$, where $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \theta_1 \boldsymbol{I}_{n_i} + \theta_2 \boldsymbol{J}_{n_i}$. It is easy to show that $\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) = \theta_1^{-1}(\boldsymbol{I}_{n_i} - \theta_2/(\theta_1 + n_i\theta_2)\boldsymbol{J}_{n_i})$.

As an estimator of $\boldsymbol{\theta}$, we here deal with the Prasad-Rao estimator. Define the two statistics S_1 and S_2 by

$$\begin{aligned} S_1 &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}_{\boldsymbol{E}})'\boldsymbol{E}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}_{\boldsymbol{E}}) = \boldsymbol{y}'\boldsymbol{M}_{\boldsymbol{E}}\boldsymbol{y}, \\ S_2 &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) = \boldsymbol{y}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}, \end{aligned}$$

where $\boldsymbol{E} = \text{block diag}\,(\boldsymbol{E}_1, \boldsymbol{E}_2, \ldots, \boldsymbol{E}_k)$, $\boldsymbol{E}_i = \boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}$, $\boldsymbol{M}_{\boldsymbol{E}} = \boldsymbol{E} - \boldsymbol{P}_{\boldsymbol{E}}$, $\boldsymbol{P}_{\boldsymbol{E}} = \boldsymbol{E}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{E}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{E}$, $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $\boldsymbol{b}_{\boldsymbol{E}} = (\boldsymbol{X}'\boldsymbol{E}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{E}\boldsymbol{y}$. Then the unbiased estimator $\widehat{\boldsymbol{\theta}}_U$ given by Prasad and Rao (1990) is given by

$$\widehat{\boldsymbol{\theta}}_U = \begin{pmatrix} \hat{\theta}_1^U \\ \hat{\theta}_2^U \end{pmatrix} = \begin{pmatrix} S_1/\left(N - k - p\right) \\ \left(S_2 - (N-p)\,\hat{\theta}_1^U\right)/N_1 \end{pmatrix},$$

where $N_1 = N - \text{tr}\,(\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Z})$ and $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$. The stochastic expansion of $\widehat{\boldsymbol{\theta}}^U$ in Assumption 1 can be written as

$$\widehat{\boldsymbol{\theta}}_U - \widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_U^{\dagger} + \widehat{\boldsymbol{\theta}}_U^{\dagger\dagger} + o_p(N^{-1}),$$

where

$$\widehat{\boldsymbol{\theta}}_U^{\dagger} = \begin{pmatrix} (N-k)^{-1}\text{tr}\,(\boldsymbol{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' - \boldsymbol{\Sigma}(\boldsymbol{\theta}))) \\ N^{-1}\text{tr}\,(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' - \boldsymbol{\Sigma}(\boldsymbol{\theta})) - (N-k)^{-1}\text{tr}\,(\boldsymbol{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' - \boldsymbol{\Sigma}(\boldsymbol{\theta}))) \end{pmatrix},$$

$$\widehat{\boldsymbol{\theta}}_U^{\dagger\dagger} = \begin{pmatrix} -(N-k)^{-1}\text{tr}\,(\boldsymbol{P}_{\boldsymbol{E}}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' - \boldsymbol{\Sigma}(\boldsymbol{\theta}))) \\ -N^{-1}\text{tr}\,(\boldsymbol{P}_{\boldsymbol{X}}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' - \boldsymbol{\Sigma}(\boldsymbol{\theta}))) + (N-k)^{-1}\text{tr}\,(\boldsymbol{P}_{\boldsymbol{E}}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' - \boldsymbol{\Sigma}(\boldsymbol{\theta}))). \end{pmatrix}$$

It can be easily seen that $E[\widehat{\boldsymbol{\theta}}_U^{\dagger}] = E[\widehat{\boldsymbol{\theta}}_U^{\dagger\dagger}] = \mathbf{0}$. From (5.4)-(5.6) of Prasad and Rao (1990), $E[\widehat{\boldsymbol{\theta}}_U^{\dagger}(\widehat{\boldsymbol{\theta}}_U^{\dagger})']$ can be approximated as $E[\widehat{\boldsymbol{\theta}}_U^{\dagger}(\widehat{\boldsymbol{\theta}}_U^{\dagger})'] = 2\boldsymbol{\Lambda}(\boldsymbol{\theta}) + O(N^{-2})$, where

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}) = \frac{\theta_1^2}{N-k}\begin{pmatrix} 1 & -k/N \\ -k/N & \{k^2 + (N-k)\sum_{i=1}^k (1 + n_i\theta_2/\theta_1)^2\}/N^2 \end{pmatrix}. \tag{5.3}$$

The Wald, score and modified LR test statistics given in Section 2 are expressed as $W = V(\widehat{\boldsymbol{\theta}}^U)$, $S = V(\widetilde{\boldsymbol{\theta}}^U)$ and $mLR = (W+S)/2$ through $V(\cdot)$, which is written as

$$V(\boldsymbol{\theta}) = \frac{1}{\theta_1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_\alpha(\boldsymbol{\theta}) - \boldsymbol{r}\right)'(\boldsymbol{X}_{\boldsymbol{R}}'\boldsymbol{X}_{\boldsymbol{R}})^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_\alpha(\boldsymbol{\theta}) - \boldsymbol{r}\right),$$

where $\boldsymbol{X_R} = \boldsymbol{X}_\alpha \left(\boldsymbol{X}'_\alpha \boldsymbol{X}_\alpha\right)^{-1} \boldsymbol{R}'$ and $\widehat{\boldsymbol{\beta}}_\alpha(\boldsymbol{\theta}) = (\boldsymbol{X}'_\alpha \boldsymbol{X}_\alpha)^{-1}\boldsymbol{X}'_\alpha \boldsymbol{y}_\alpha$ for $\boldsymbol{X}_{\alpha i} = \boldsymbol{X}_i - \alpha_i {n_i}^{-1}\boldsymbol{J}_{n_i}\boldsymbol{X}_i$, $\alpha_i(\boldsymbol{\theta}) = 1-(1+n_i\theta_2/\theta_1)^{-1/2}$ and $\boldsymbol{y}_{\alpha i} = \boldsymbol{y}_i - \alpha_i {n_i}^{-1}\boldsymbol{J}_{n_i}\boldsymbol{y}_i$. The restricted estimator induced from $\widehat{\boldsymbol{\theta}}^U$ is provided by

$$\widetilde{\boldsymbol{\theta}}^U = \widehat{\boldsymbol{\theta}}^U - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}^U)\mathbf{col}(V_{(i)}(\widehat{\boldsymbol{\theta}}^U)),$$

where $\boldsymbol{\Lambda}(\cdot)$ is given in (5.3). It is noted that $V_{(i)}(\boldsymbol{\theta})$ is expressed as

$$V_{(i)}(\boldsymbol{\theta}) = -(\boldsymbol{y}_\alpha - \boldsymbol{X}_\alpha\boldsymbol{\beta})'(2\boldsymbol{M}_\alpha + \boldsymbol{P_R})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{P_R}(\boldsymbol{y}_\alpha - \boldsymbol{X}_\alpha\boldsymbol{\beta}),$$

where $\boldsymbol{M}_\alpha = \boldsymbol{I}_N - \boldsymbol{X}_\alpha \left(\boldsymbol{X}'_\alpha \boldsymbol{X}_\alpha\right)^{-1} \boldsymbol{X}'_\alpha$, $\boldsymbol{P_R} = \boldsymbol{X_R}\left(\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{X_R}\right)^{-1}\boldsymbol{X}'_{\boldsymbol{R}}$ and $\boldsymbol{\Sigma}_i^{-1/2} = \theta_1^{-1/2}(\boldsymbol{I}_{n_i} - n_i^{-1}\alpha_i\boldsymbol{J}_{n_i})$.

The test statistics with the analytical Bartlett adjustments are given in (3.3), where the differentiations (3.4) are written as

$$\begin{aligned}
\boldsymbol{A}_{(i)} &= -(\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{X_R})^{-1/2}\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{M}_\alpha\boldsymbol{\Sigma}^{-1/2},\\
\boldsymbol{B}_{(i)} &= -(\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{X_R})^{-1/2}\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X_R}(\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{X_R})^{-1/2},\\
\boldsymbol{B}_{(ij)} &= (\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{X_R})^{-1/2}\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{M}_\alpha\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_{(j)}\\
&\qquad +\boldsymbol{\Sigma}_{(j)}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{M}_\alpha\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_{(i)})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X_R}(\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{X_R})^{-1/2},
\end{aligned}$$

where $\boldsymbol{\Sigma}_{(1)i} = \boldsymbol{I}_{n_i}$, $\boldsymbol{\Sigma}_{(2)i} = \boldsymbol{J}_{n_i}$.

Also, the test statistics with the Bartlett adjustments via the parametric bootstrap are given in (4.5), where the restricted estimator $\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta})$ in (4.1) is expressed as

$$\widetilde{\boldsymbol{\beta}}_\alpha(\boldsymbol{\theta}) = \widehat{\boldsymbol{\beta}}_\alpha(\boldsymbol{\theta}) - (\boldsymbol{X}'_\alpha\boldsymbol{X}_\alpha)^{-1}\boldsymbol{R}'(\boldsymbol{X}'_{\boldsymbol{R}}\boldsymbol{X_R})^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_\alpha(\boldsymbol{\theta}) - \boldsymbol{r}),$$

6 Simulation Study

In this section, we investigate the performances of the type I errors and the powers for the three classical tests and the adjusted tests with the Bartlett corrections through a Monte Carlo simulation.

In the simulation experiment, we use the nested error regression model (5.2), described again as

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

where u_{ij} is generated from $\mathcal{N}(0, 1)$. For v_i, we use the setup proposed in Datta, Rao and Smith (2005), namely, v_i is generated from three different distributions: $\mathcal{N}(0, 1)$, the double exponential distribution $DExp(0, 1/\sqrt{2})$ and the shifted exponential distribution $SExp(-1, 1)$. Using these three different distributions, we can examine the robustness on the significance level of the proposed tests. Let $p = 3$, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ for $\beta_0 = 1$. For n_i, we handle two patters: pattern A is $(n_1, n_2, n_3, n_4) = (4, 4, 5, 6)$ with $N = 19$ and pattern B is $(n_1, n_2, \ldots, n_9) = (2, 2, 4, 4, 4, 5, 5, 5, 10)$ with $N = 41$. For $\boldsymbol{x}_{ij}$, we consider the case that $(\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i})' = (\boldsymbol{j}_{n_i}, \boldsymbol{X}_i^{*\prime})$ for $n_i \times (p-1)$ matrix $\boldsymbol{X}_i^*$, where $\boldsymbol{X}_i^*$ is generated as

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

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

The null hypothesis we deal with in this simulation experiment is

$$H_0 \ : \ \beta_1 = \beta_2 = 0,$$

and the nominal significance level is $\alpha = 0.05$. We investigate the performance of the type I errors (size) and powers of the crude tests given by W, mLR and S, the adjusted tests given by W^{BC}, mLR^{BC} and S^{BC} with the analytical Bartlett corrections, and the adjusted tests given by W_*^{BC}, mLR_*^{BC} and S_*^{BC} with the parametric bootstrap Bartlett corrections.

When the Prasad-Rao estimator is used for $\boldsymbol{\theta}$, Table 1 reports the average of the sizes of the test statistics based on 10,000 replications were the size of the bootstrap sample is 1,000. For both patterns A and B, the type I errors of the W and S tests are not good, while the adjusted tests with the Bartlett corrections W^{BC}, S^{BC}, W_*^{BC} and S_*^{BC} give significant improvements in terms of their sizes. For the modified LR test, the sizes of the adjusted tests are better than those of mLR. When the ML and REML are used for $\boldsymbol{\theta}$, we have similar observations as shown in Table 2, where the Monte Carlo simulation was conducted with 1,000 times with a bootstrap sample of size 100, since it takes long time in the simulation experiments for ML and REML estimators.

The powers of the test statistics based on the Prasad-Rao estimator of $\boldsymbol{\theta}$ are reported in Table 3 where the powers at two points $\beta_1 = \beta_2 = 0.1$ and $\beta_1 = \beta_2 = 0.2$ are computed based on 10,000 replications with bootstrap samples of size 1,000. From the table, it is revealed that the powers of the tests with the analytical Bartlett adjustments are close to the powers of the tests with the parametric bootstrap adjustments. It is interesting to point out that the adjusted score tests S^{BC} and S_*^{BC} are more powerful than the adjusted LR tests mLR^{BC} and mLR_*^{BC}, which are a bit more powerful than the adjusted Wald tests W^{BC} and W_*^{BC}.

7 Concluding Remarks

In this paper, we have derived the Bartlett corrections of the Wald, score and modified likelihood ratio (LR) tests using two kinds of techniques, namely the analytical method via the Taylor series expansion and the numerical method via the parametric bootstrap. We have also shown that the three test statistics with these Bartlett adjustments have the second-order corrections in their type I errors. Although the analytical Bartlett corrections were provided by Rothenberg (1984) in the case that the nuisance parameters are estimated by the maximum likelihood (ML), we have extended his results to the case that the nuisance parameters are estimated by general consistent estimators. The nuisance parameters correspond to variance components in specific linear mixed models, and various estimators of the variance components including REML, MINQUE and other specific unbiased estimators are available. For these estimators other than ML, the results in this

Table 1: Comparison of the type I error (size, %) of the original W, mLR and S tests and the adjusted tests with the analytical and parametric bootstrap Bartlett corrections where $\boldsymbol{\theta}$ is estimated by the Prasad-Rao estimator and the nominal level is 5 %

Prasad-Rao estimator									
	W	W^{BC}	W^{BC}_{*}	mLR	mLR^{BC}	mLR^{BC}_{*}	S	S^{BC}	S^{BC}_{*}
	$v_i \sim \mathcal{N}(0,1)$								
pattern A	10.7	5.86	5.85	6.42	5.44	5.33	1.72	5.60	5.47
pattern B	6.95	4.99	4.79	5.39	4.96	4.91	3.64	4.99	5.07
	$v_i \sim DExp(0, 1/\sqrt{2})$								
pattern A	10.2	5.47	5.52	6.70	5.77	5.60	1.87	5.65	5.57
pattern B	6.95	5.06	4.86	5.35	5.04	5.15	3.67	5.03	5.29
	$v_i \sim SExp(-1, 1)$								
pattern A	10.2	5.36	5.42	6.21	5.42	5.33	1.74	5.30	5.24
pattern B	6.62	4.86	4.76	5.24	4.83	4.98	3.68	4.89	5.52

Table 2: Comparison of the type I error (size, %) of the original W, mLR and S tests and the adjusted tests with the analytical and parametric bootstrap Bartlett corrections where $\boldsymbol{\theta}$ is estimated by the ML and REML estimators and the nominal level is 5 %

ML and REML estimator									
	W	W^{BC}	W^{BC}_{*}	mLR	mLR^{BC}	mLR^{BC}_{*}	S	S^{BC}	S^{BC}_{*}
	$v_i \sim \mathcal{N}(0,1)$								
pattern B ML	9.5	5.1	5.5	6.6	5.2	5.3	3.3	4.7	4.2
pattern B REML	7.7	4.6	4.5	5.0	5.0	4.7	1.7	4.7	5.1
	$v_i \sim DExp(0, 1/\sqrt{2})$								
pattern B ML	9.5	5.9	5.8	6.9	5.8	5.6	3.9	5.1	5.0
pattern B REML	7.8	5.1	4.7	5.1	4.9	5.1	2.1	4.9	5.6
	$v_i \sim SExp(-1, 1)$								
pattern B ML	10.4	5.6	6.1	7.9	5.8	6.5	3.6	5.8	5.7
pattern B REML	8.1	5.3	5.0	5.3	5.3	5.6	2.7	6.1	6.3

paper can provide the Bartlett corrections of the W, S and mLR tests. In fact, we have treated the simple unbiased estimators given by Prasad and Rao (1990) in the nested error regression model, which has been used in the two-stage sampling and the small-area estimation, and through simulation experiments we have shown that the test statistics with the Bartlett adjustments via the analytical and parametric bootstrap methods provide significant improvements in the type I errors.

The above description implies that the testing procedures proposed in this paper possess a couple of merits from a practical aspect. One is that we can use any computationally simpler estimators among consistent estimators of the nuisance parameters. In general,

Table 3: Comparison of the powers (%) of the adjusted W, mLR and S tests with the analytical and parametric bootstrap Bartlett corrections where $\boldsymbol{\theta}$ is estimated by the Prasad-Rao estimator and the nominal level is 5 %

Prasad-Rao estimator						
	W^{BC}	W_*^{BC}	mLR^{BC}	mLR_*^{BC}	S^{BC}	S_*^{BC}
$\beta_1 = \beta_2 = 0.1$			$v_i \sim \mathcal{N}(0,1)$			
pattern A	56.89	57.05	64.80	61.28	64.80	66.36
pattern B	58.36	58.39	62.03	62.87	65.82	67.14
$\beta_1 = \beta_2 = 0.2$						
pattern A	98.70	98.75	98.97	98.98	99.13	99.24
pattern B	98.98	99.00	99.21	99.20	99.24	99.32
$\beta_1 = \beta_2 = 0.1$			$v_i \sim DExp(0, 1/\sqrt{2})$			
pattern A	57.56	58.00	61.68	62.67	66.00	67.97
pattern B	59.25	59.61	63.24	64.31	67.34	69.49
$\beta_1 = \beta_2 = 0.2$						
pattern A	98.56	98.59	98.89	99.01	99.11	99.26
pattern B	98.85	98.91	99.11	99.20	99.33	99.44
$\beta_1 = \beta_2 = 0.1$			$v_i \sim SExp(-1, 1)$			
pattern A	58.13	58.42	62.29	63.70	66.60	68.91
pattern B	59.25	59.52	62.67	63.69	66.65	69.01
$\beta_1 = \beta_2 = 0.2$						
pattern A	98.79	98.84	99.03	99.12	99.21	99.33
pattern B	98.93	98.97	99.18	99.23	99.32	99.41

ML and REML require numerical iterations to get the solutions, and the solutions are sometimes instable as well as it takes time to get them. Since the Prasad-Rao estimator is given as an explicit form, it is easier to compute the test statistics. Another merit is that we can compute the Bartlett correction numerically with the parametric bootstrap, namely, we do not have to derive the second-order bias, variance and covariance of the general consistent estimators. This is really useful, since we can obtain the bootstrap Bartlett adjustments with the second-order corrections for any consistent estimator in any complicated model (2.1).

A Appendix

A.1 Derivation of the restricted estimator (2.3) and the consistency

We begin by deriving the relation between the ML $\widehat{\boldsymbol{\theta}}_{\mathrm{M}}$ and the restricted ML $\widetilde{\boldsymbol{\theta}}_{\mathrm{M}}$ under the null hypothesis. Note that $V(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}) = -2(L(\widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}), \widehat{\boldsymbol{\theta}}_{\mathrm{M}}) - L(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}), \widehat{\boldsymbol{\theta}}_{\mathrm{M}}))$. From the definition of ML, it follows that $\mathbf{col}(L_{(i)}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}), \widehat{\boldsymbol{\theta}}_{\mathrm{M}})) = \mathbf{0}$, which implies that the partial derivative of $V(\widehat{\boldsymbol{\theta}}_{\mathrm{M}})$ with respect to $\widehat{\boldsymbol{\theta}}_{\mathrm{M}}$ is expressed as

$$\mathbf{col}(V_{(i)}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}})) = -2\mathbf{col}(L_{(i)}(\widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}), \widehat{\boldsymbol{\theta}}_{\mathrm{M}})).$$

From the definition of the restricted ML, it follows that $\mathbf{col}(L_{(i)}(\widetilde{\boldsymbol{\beta}}(\widetilde{\boldsymbol{\theta}}_{\mathrm{M}}), \widetilde{\boldsymbol{\theta}}_{\mathrm{M}})) = \mathbf{0}$, so that the Taylor series expansion of $\mathbf{col}(L_{(i)}(\widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}), \widehat{\boldsymbol{\theta}}_{\mathrm{M}}))$ around $\widetilde{\boldsymbol{\theta}}_{\mathrm{M}}$ gives

$$\begin{aligned}-2\mathbf{col}(L_{(i)}(\widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}), \widehat{\boldsymbol{\theta}}_{\mathrm{M}})) =& \mathbf{mat}(-2L_{(ij)}(\widetilde{\boldsymbol{\beta}}(\widetilde{\boldsymbol{\theta}}_{\mathrm{M}}), \widetilde{\boldsymbol{\theta}}_{\mathrm{M}}))(\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \widetilde{\boldsymbol{\theta}}_{\mathrm{M}}) + O_p(N^{-1/2}) \\ =& \boldsymbol{\Omega}(\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \widetilde{\boldsymbol{\theta}}_{\mathrm{M}}) + O_p(N^{-1/2}),\end{aligned}$$

under the null hypothesis, where $\boldsymbol{\Omega}(\boldsymbol{\theta}) = \mathbf{mat}(-2E[L_{(ij)}(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta})])$. Since $\boldsymbol{\Omega}(\boldsymbol{\theta}) = O(N)$,

$$\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \widetilde{\boldsymbol{\theta}}_{\mathrm{M}} = \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\mathbf{col}(\boldsymbol{V}_{(i)}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}})) + O_p(N^{-3/2}). \tag{A.1}$$

It is here noted that $E[\widehat{\boldsymbol{\theta}}^{\dagger}_{\mathrm{M}}(\widehat{\boldsymbol{\theta}}^{\dagger}_{\mathrm{M}})'] = 2\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}) + o(N^{-1})$, where $\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{\dagger}_{\mathrm{M}} + o_p(N^{-1/2})$. In fact, the Taylor series expansion of $\mathbf{col}(L_{(i)}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}), \widehat{\boldsymbol{\theta}}_{\mathrm{M}}))$ around $\boldsymbol{\theta}$ gives

$$\begin{aligned}0 =& -2\mathbf{col}(L_{(i)}(\widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}}), \widehat{\boldsymbol{\theta}}_{\mathrm{M}})) \\ =& -2\mathbf{col}(L_{(i)}(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta})) + \mathbf{mat}(-2L_{(ij)}(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}))(\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \boldsymbol{\theta}) + o_p(N^{1/2}),\end{aligned}$$

which yields that $\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \boldsymbol{\theta} = \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\mathbf{col}(NL_{(i)}(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta})) + o_p(N^{-1/2})$. From Kubokawa (2011a), it follows that $E[\widehat{\boldsymbol{\theta}}^{\dagger}_{\mathrm{M}}(\widehat{\boldsymbol{\theta}}^{\dagger}_{\mathrm{M}})'] = E[(\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \boldsymbol{\theta})'] + o_p(N^{-1}) = 2\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}) + o(N^{-1})$.

Substituting $\widehat{\boldsymbol{\theta}}_{\mathrm{M}}$ into $\boldsymbol{\Omega}(\boldsymbol{\theta})$ in (A.1), we get the restricted ML given by

$$\widetilde{\boldsymbol{\theta}}_{\mathrm{M}} = \widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \boldsymbol{\Omega}^{-1}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}})\mathbf{col}(\boldsymbol{V}_{(i)}(\widehat{\boldsymbol{\theta}}_{\mathrm{M}})), \tag{A.2}$$

where $\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}) = 2^{-1}E[(\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}_{\mathrm{M}} - \boldsymbol{\theta})'] + o(N^{-1}) = 2^{-1}E[\widehat{\boldsymbol{\theta}}^{\dagger}_{\mathrm{M}}(\widehat{\boldsymbol{\theta}}^{\dagger}_{\mathrm{M}})'] + o(N^{-1})$.

The general restricted estimator $\widetilde{\boldsymbol{\theta}}$ can be provided by replacing $\widehat{\boldsymbol{\theta}}_{\mathrm{M}}$ with $\widehat{\boldsymbol{\theta}}$ in (A.2), namely,

$$\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}})\mathbf{col}(\boldsymbol{V}_{(i)}(\widehat{\boldsymbol{\theta}})),$$

where $\boldsymbol{\Lambda}(\boldsymbol{\theta}) = 2^{-1}E[\widehat{\boldsymbol{\theta}}^{\dagger}(\widehat{\boldsymbol{\theta}}^{\dagger})']$. The consistency of $\widetilde{\boldsymbol{\theta}}$ under the null hypothesis follows from the fact that $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}$ and $\boldsymbol{\Lambda}(\boldsymbol{\theta}) = O(N^{-1})$.

A.2 Proofs of Theorems 3.1 and 3.2

We begin by preparing the following two lemmas which will be useful for proving these theorems.

Lemma A.1 *Let $\boldsymbol{X}$ be an r-dimensional random vector having $\mathcal{N}_r(\boldsymbol{\mu}, \boldsymbol{I})$. For an $r \times 1$ constant vector $\boldsymbol{a}$ and an $r \times r$ constant matrix $\boldsymbol{A}$, one gets the the following expression:*

$$E[e^{t\boldsymbol{X}'\boldsymbol{X}}g(\boldsymbol{a}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{a}', \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})] = \gamma^{\frac{r}{2}}e^{t\gamma\boldsymbol{\mu}'\boldsymbol{\mu}}E[g(\boldsymbol{a}'\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}'\boldsymbol{a}', \tilde{\boldsymbol{X}}'\boldsymbol{A}\tilde{\boldsymbol{X}})],$$

where $\tilde{\boldsymbol{X}} \sim \mathcal{N}_r(\gamma\boldsymbol{\mu}, \gamma\boldsymbol{I})$ for $\gamma = 1/(1-2t)$.

Proof. Combining $e^{t\boldsymbol{X}'\boldsymbol{X}}$ and the density of the multivariate normal distribution, we see that

$$
\begin{aligned}
&E[e^{t\boldsymbol{X}'\boldsymbol{X}}g(\boldsymbol{a}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{a}', \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})] \\
=&\int e^{t\boldsymbol{x}'\boldsymbol{x}}g(\boldsymbol{a}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{a}', \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})\frac{1}{(2\pi)^{r/2}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu})'(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right)d\boldsymbol{x} \\
=&\int g(\boldsymbol{a}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{a}', \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})\frac{1}{(2\pi)^{r/2}}\exp\left(t\gamma\boldsymbol{\mu}'\boldsymbol{\mu}-\frac{\left(\boldsymbol{x}-\frac{1}{2\gamma}\boldsymbol{\delta}\right)'\left(\boldsymbol{x}-\frac{1}{2\gamma}\boldsymbol{\delta}\right)}{2\gamma}\right)d\boldsymbol{x} \\
=&\gamma^{\frac{r}{2}}e^{t\gamma\boldsymbol{\mu}'\boldsymbol{\mu}}\int g(\boldsymbol{a}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{a}', \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})\left(\frac{1}{2\pi\gamma}\right)^{r/2}\exp\left(-\frac{(\boldsymbol{x}-\frac{1}{2\gamma}\boldsymbol{\delta})'(\boldsymbol{x}-\frac{1}{2\gamma}\boldsymbol{\delta})}{2\gamma}\right)d\boldsymbol{x} \\
=&\gamma^{\frac{r}{2}}e^{t\gamma\boldsymbol{\mu}'\boldsymbol{\mu}}E[g(\boldsymbol{a}'\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}'\boldsymbol{a}, \tilde{\boldsymbol{X}}'\boldsymbol{A}\tilde{\boldsymbol{X}})],
\end{aligned}
$$

which establishes Lemma A.1. □

Lemma A.2 *Let* $\boldsymbol{X}$ *be a random vector having* $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Then, the following identities are useful for evaluating moments.*

(1) $E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] = \mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$

(2) $E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{a}'\boldsymbol{X}] = \mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A})\boldsymbol{a}'\boldsymbol{\mu} + \boldsymbol{\mu}'(\boldsymbol{\Sigma}\boldsymbol{A} + \boldsymbol{A}\boldsymbol{\Sigma})\boldsymbol{a} + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{a}'\boldsymbol{\mu}$

(3) $E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] = (\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu})^2 + 2\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A})^2 + 4\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu}$

Proof. To evaluate moments, we use the following Stein identity given by Stein (1981) for $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$E[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{g}(\boldsymbol{y})] = E[\boldsymbol{\nabla}'\{\boldsymbol{\Sigma}\boldsymbol{g}(\boldsymbol{X})\}], \tag{A.3}$$

where $\boldsymbol{g}(\boldsymbol{X}) = (g_1(\boldsymbol{X}), \ldots, g_N(\boldsymbol{X}))'$ is an absolutely continuous function and $\boldsymbol{\nabla} = \partial/\partial\boldsymbol{X}$. For (1), the Stein identity is used to evaluate $E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}]$ as $E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] = E[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{A}\boldsymbol{X}] + \boldsymbol{\mu}'\boldsymbol{A}E[\boldsymbol{X}] = E[\boldsymbol{\nabla}'\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{X}] + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} = \mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$. For (2), the same argument is applied to get

$$
\begin{aligned}
E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{a}'\boldsymbol{X}] =&E[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{A}\boldsymbol{X}\boldsymbol{a}'\boldsymbol{X}] + E[\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{a}'\boldsymbol{X}] \\
=&E[\boldsymbol{\nabla}'\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{X}\boldsymbol{a}'\boldsymbol{X}] + E[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{a}'\boldsymbol{\mu}] + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}E[\boldsymbol{a}'\boldsymbol{X}] \\
=&E[\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A})\boldsymbol{a}'\boldsymbol{X} + \boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{a}] + E[\boldsymbol{\nabla}'\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{a}'\boldsymbol{X}] + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{a}'\boldsymbol{\mu} \\
=&\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A})\boldsymbol{a}'\boldsymbol{\mu} + \boldsymbol{\mu}'(\boldsymbol{\Sigma}\boldsymbol{A} + \boldsymbol{A}\boldsymbol{\Sigma})\boldsymbol{a} + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{a}'\boldsymbol{\mu}.
\end{aligned}
$$

For (3), it is similarly shown that

$$
\begin{aligned}
E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] =&E[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] + E[\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] \\
=&E[\boldsymbol{\nabla}'\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] + E[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] \\
=&\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A})E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] + 2E[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{X}] \\
&+ E[\boldsymbol{\nabla}'\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}(\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}) \\
=&(\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu})^2 + 2\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A})^2 + 4\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu},
\end{aligned}
$$

which proves Lemma A.2. $\square$

[Expansion of the Wald test] We now derive the asymptotic expansion of the Wald test statistic, which is expressed as $W = (\boldsymbol{x}+\boldsymbol{s})'(\boldsymbol{I}-\boldsymbol{S})^{-1}(\boldsymbol{x}+\boldsymbol{s})$ from (3.1), where $\boldsymbol{x} = \boldsymbol{H}\{\boldsymbol{R}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{r}\}$, $\boldsymbol{s} = \boldsymbol{H}\boldsymbol{R}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})\}$ and $\boldsymbol{S} = -\boldsymbol{H}\boldsymbol{R}\{(\boldsymbol{X}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{X})^{-1} - (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\}\boldsymbol{R}'\boldsymbol{H}$ for $\boldsymbol{H} = (\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1/2}$. Let $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$. The Taylor series expansions of $\boldsymbol{S}$ and $\boldsymbol{s}$ around $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ give

$$\boldsymbol{S} = -\boldsymbol{H}\boldsymbol{R}\Big(\sum_i (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}_{(i)}(\hat{\theta}_i^{\dagger} + \hat{\theta}_i^{\dagger\dagger}) + \frac{1}{2}\sum_{i,j}((\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}_{(ij)}\hat{\theta}_i^{\dagger}\hat{\theta}_j^{\dagger})\Big)\boldsymbol{R}'\boldsymbol{H} + o_p(N^{-1}),$$

and

$$\begin{aligned}\boldsymbol{s} =& \boldsymbol{H}\boldsymbol{R}\Big(\sum_i (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1})_{(i)}\boldsymbol{\epsilon}(\hat{\theta}_i^{\dagger} + \hat{\theta}_i^{\dagger\dagger}) \\ &+ \frac{1}{2}\sum_{i,j}((\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1})_{(ij)}\boldsymbol{\epsilon}\hat{\theta}_i^{\dagger}\hat{\theta}_j^{\dagger})\Big) + o_p(N^{-1}),\end{aligned}$$

both of which are of $O_p(N^{-1/2})$ from Assumption 1. Since $(\boldsymbol{I}-\boldsymbol{S})^{-1} = \boldsymbol{I} + \boldsymbol{S} + \boldsymbol{S}^2 + \boldsymbol{R}_N$ where every elements of $\boldsymbol{R}_N$ is of order $o_p(N^{-1})$, the Wald test statistic is expanded as

$$\begin{aligned}W =& (\boldsymbol{x}+\boldsymbol{s})'(\boldsymbol{I}+\boldsymbol{S}+\boldsymbol{S}^2)(\boldsymbol{x}+\boldsymbol{s}) + o_p(N^{-1}) \\ =& \boldsymbol{x}'\boldsymbol{x} + (2\boldsymbol{s}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{S}\boldsymbol{x}) + (\boldsymbol{s}'\boldsymbol{s} + 2\boldsymbol{s}'\boldsymbol{S}\boldsymbol{x} + \boldsymbol{x}\boldsymbol{S}^2\boldsymbol{x}) + o_p(N^{-1}).\end{aligned}$$

It is noted that $2\boldsymbol{s}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{S}\boldsymbol{x} = O_p(N^{-1/2})$ and $\boldsymbol{s}'\boldsymbol{s} + 2\boldsymbol{s}'\boldsymbol{S}\boldsymbol{x} + \boldsymbol{x}\boldsymbol{S}^2\boldsymbol{x} = O_p(N^{-1})$. Using these facts and Lemma A.1, we can evaluate the moment generating function of W as

$$\begin{aligned}E[e^{tW}] =& E[e^{t\boldsymbol{x}'\boldsymbol{x}}e^{t(2\boldsymbol{s}'\boldsymbol{x}+\boldsymbol{x}'\boldsymbol{S}\boldsymbol{x})}e^{t(\boldsymbol{s}'\boldsymbol{s}+2\boldsymbol{s}'\boldsymbol{S}\boldsymbol{x}+\boldsymbol{x}\boldsymbol{S}^2\boldsymbol{x})} + o_p(N^{-1})] \\ =& E[e^{t\boldsymbol{x}'\boldsymbol{x}}(1 + t(2\boldsymbol{s}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{S}\boldsymbol{x}) + t(\boldsymbol{s}'\boldsymbol{s} + 2\boldsymbol{s}'\boldsymbol{S}\boldsymbol{x} + \boldsymbol{x}\boldsymbol{S}^2\boldsymbol{x}) \\ &+ \frac{1}{2}t^2(2\boldsymbol{s}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{S}\boldsymbol{x})^2)] + o(N^{-1}) \\ =& t\gamma^{\frac{r}{2}}e^{\frac{\gamma-1}{2}\boldsymbol{\delta}'\boldsymbol{\delta}}E[(1 + t(\tilde{\boldsymbol{x}}'\boldsymbol{S}\tilde{\boldsymbol{x}} + \boldsymbol{s}'\boldsymbol{s} + 2\boldsymbol{s}'\boldsymbol{S}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{x}}\boldsymbol{S}^2\tilde{\boldsymbol{x}}) \\ &+ \frac{1}{2}t^2(2\boldsymbol{s}'\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{x}}'\boldsymbol{S}\tilde{\boldsymbol{x}})^2)] + o(N^{-1}),\end{aligned}$$

where $\tilde{\boldsymbol{x}} \sim \mathcal{N}_N(\gamma\boldsymbol{\delta}, \gamma\boldsymbol{I})$. Since $E[\boldsymbol{\epsilon}] = 0$, it is noted that $E[\boldsymbol{s}] = \boldsymbol{0}$. Lemma A.2 can be used to simplify $E[e^{tW}]$ as

$$\begin{aligned}E[e^{tW}] =& \gamma^{\frac{r}{2}}e^{\frac{\gamma-1}{2}\boldsymbol{\delta}'\boldsymbol{\delta}}E\Big[1 + \frac{\gamma-1}{2}(\boldsymbol{s}'\boldsymbol{s} + \mathrm{tr}\,(\boldsymbol{S}) + \mathrm{tr}\,(\boldsymbol{S}^2)) + \frac{\gamma(\gamma-1)}{2}(\mathrm{tr}\,(\boldsymbol{S}\boldsymbol{\Delta}) + \mathrm{tr}\,(\boldsymbol{S}^2\boldsymbol{\Delta})) \\ &+ \frac{(\gamma-1)^2}{2}\mathrm{tr}\,(\boldsymbol{s}\boldsymbol{s}'\boldsymbol{\Delta}) + \frac{(\gamma-1)^2}{4}(\mathrm{tr}\,(\boldsymbol{S}^2) + \frac{1}{2}(\mathrm{tr}\,(\boldsymbol{S}))^2) \\ &+ \frac{\gamma(\gamma-1)^2}{2}(\mathrm{tr}\,(\boldsymbol{S}^2\boldsymbol{\Delta}) + \frac{1}{2}\mathrm{tr}\,(\boldsymbol{S}\boldsymbol{\Delta})\mathrm{tr}\,(\boldsymbol{S})) + \frac{\gamma^2(\gamma-1)^2}{8}(\mathrm{tr}\,(\boldsymbol{S}\boldsymbol{\Delta}))^2\Big] + o(N^{-1}),\end{aligned}$$

where $\boldsymbol{\Delta} = \boldsymbol{\delta}\boldsymbol{\delta}'$.

We need to investigate the order of the moments $E[\boldsymbol{s}'\boldsymbol{s}]$, $E[\mathrm{tr}\,(\boldsymbol{S})]$, $E[(\mathrm{tr}\,(\boldsymbol{S}))^2]$ and $E[\mathrm{tr}\,(\boldsymbol{S}^2)]$. For the purpose, we use the notations given by

$$\begin{aligned}\boldsymbol{A}_{(i)} =& \boldsymbol{H}\boldsymbol{R}((\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1})_{(i)},\\ \boldsymbol{B}_{(i)} =& -\boldsymbol{H}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}_{(i)}\boldsymbol{R}'\boldsymbol{H},\\ \boldsymbol{B}_{(ij)} =& -\boldsymbol{H}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}_{(ij)}\boldsymbol{R}'\boldsymbol{H}.\end{aligned}$$

Under Assumption 1 [2], it is seen that

$$\begin{aligned}E[\boldsymbol{s}'\boldsymbol{s}] =& \sum_{i,j}^{q} E[\boldsymbol{u}'\boldsymbol{A}'_{(i)}\boldsymbol{A}_{(j)}\boldsymbol{u}\hat{\theta}^{\dagger}_i\hat{\theta}^{\dagger}_j] + o(N^{-1})\\ =& \sum_{i,j}^{q} E[\nabla'(\boldsymbol{\Sigma}\boldsymbol{A}'_{(i)}\boldsymbol{A}_{(j)}\boldsymbol{u})\hat{\theta}^{\dagger}_i\hat{\theta}^{\dagger}_j] + o(N^{-1}) \quad (\because \text{Stein identity})\\ =& \sum_{i,j}^{q} \mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}'_{(i)}\boldsymbol{A}_{(j)})E[\hat{\theta}^{\dagger}_i\hat{\theta}^{\dagger}_j] + 2\sum_{i,j}^{q} E[\boldsymbol{u}'A'_jA_i\Sigma(\boldsymbol{\theta})(\nabla\hat{\theta}^{\dagger}_i)\hat{\theta}^{\dagger}_j] + o(N^{-1})\\ =& \sum_{i,j}^{q} \mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}'_{(i)}\boldsymbol{A}_{(j)})E[\hat{\theta}^{\dagger}_i\hat{\theta}^{\dagger}_j] + o(N^{-1}),\end{aligned} \tag{A.4}$$

which is of order $O(N^{-1})$. The same argument is used to show that

$$E[\mathrm{tr}\,(\boldsymbol{S})] = \sum_{i}^{q} \mathrm{tr}\,(\boldsymbol{B}_{(i)})E[\hat{\theta}^{\dagger}_i + \hat{\theta}^{\dagger\dagger}_i] + \frac{1}{2}\sum_{i,j}^{q} \mathrm{tr}\,(\boldsymbol{B}_{(ij)})E[\hat{\theta}^{\dagger}_i\hat{\theta}^{\dagger}_j] + o(N^{-1}), \tag{A.5}$$

which is of order $O(N^{-1})$. Similarly,

$$E[(\mathrm{tr}\,(\boldsymbol{S}))^2] = \sum_{i,j}^{q} \mathrm{tr}\,(\boldsymbol{B}_{(i)})\mathrm{tr}\,(\boldsymbol{B}_{(j)})E[\hat{\theta}^{\dagger}_i\hat{\theta}^{\dagger}_j] + o(N^{-1}), \tag{A.6}$$

$$E[tr(\boldsymbol{S}^2)] = \sum_{i,j}^{q} \mathrm{tr}\,(\boldsymbol{B}_{(i)}\boldsymbol{B}_{(j)})E[\hat{\theta}^{\dagger}_i\hat{\theta}^{\dagger}_j] + o(N^{-1}) \tag{A.7}$$

both of which are of order $O(N^{-1})$.

Taking the above observations into account, we can evaluate the moment generating function of the Wald test statistic as

$$\begin{aligned}E[e^{tW}] =& \gamma^{\frac{r}{2}}e^{\frac{\gamma-1}{2}\boldsymbol{\delta}'\boldsymbol{\delta}}\Big(1 + \frac{\gamma-1}{2}(a+c) + \frac{(\gamma-1)^2}{2}(b + \mathrm{tr}\,(\boldsymbol{A}\boldsymbol{\Delta}))\\ &+ \frac{\gamma(\gamma-1)^2}{2}\mathrm{tr}\,(\boldsymbol{B}\boldsymbol{\Delta}) + \frac{\gamma(\gamma-1)^2}{2}\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{\Delta}) + \frac{\gamma^2(\gamma-1)^2}{2}\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{\Delta})\Big) + o(N^{-1})\end{aligned}$$

where $a = E[\boldsymbol{s}'\boldsymbol{s}]$, $b = 2^{-1}E\left[\mathrm{tr}\left(\boldsymbol{S}^2\right) + 2^{-1}\left(\mathrm{tr}\left(\boldsymbol{S}\right)\right)^2\right]$, $c = E\left[\mathrm{tr}\left(\boldsymbol{S}\right) + \mathrm{tr}\left(\boldsymbol{S}^2\right)\right]$, $\boldsymbol{A} = E[\boldsymbol{s}\boldsymbol{s}']$, $\boldsymbol{B} = E\left[\boldsymbol{S}^2 + 2^{-1}\boldsymbol{S}\mathrm{tr}\left(\boldsymbol{S}\right)\right]$, $\boldsymbol{C} = E\left[\boldsymbol{S} + \boldsymbol{S}^2\right]$ and $\boldsymbol{D} = 4^{-1}E\left[\boldsymbol{S}\boldsymbol{\Delta}\boldsymbol{S}\right]$. Let $G_r(x)$ be the distribution of the non-central chi-square distribution with r degree of freedom and the noncentrality $\boldsymbol{\delta}'\boldsymbol{\delta}$. Note that $G_{r+k}(x) - G_{r+k-2}(x) = -2g_{r+k}(x)$. Then, the inversion of the Laplace transformation provides

$$\begin{aligned} P(W > x) =& 1 - G_r(x) + (a - b + c - \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{\Delta}\right))g_{r+2}(x) \\ &+ (b + \mathrm{tr}\left((\boldsymbol{A} - \boldsymbol{B} + \boldsymbol{C})\boldsymbol{\Delta}\right))g_{r+4}(x) \\ &+ (\mathrm{tr}\left((\boldsymbol{B} - \boldsymbol{D})\boldsymbol{\Delta}\right))g_{r+6}(x) + \mathrm{tr}\left(\boldsymbol{D}\boldsymbol{\Delta}\right)g_{r+8}(x) + o(N^{-1}). \end{aligned} \tag{A.8}$$

[Bartlett correction of the Wald test] To derive the Bartlett correction, we use the asymptotic expansion (A.8) under the null hypothesis, namely, $\boldsymbol{\delta} = \boldsymbol{0}$. In this case, the asymptotic expansion is expressed as

$$P(W \leq x) = F_r(x) - \frac{x}{r}\left(a - b + c + \frac{x}{r+2}b\right)f_r(x) + o(N^{-1}),$$

which also implies that

$$P\left(W \leq x\left(1 + \frac{h}{N}\right)\right) = F_r(x) + \frac{hx}{N}f_r(x) - \frac{x}{r}f_r(x)\left(a - b + c + \frac{x}{r+2}b\right) + o(N^{-1}).$$

If the second term is equal to the third term in the r.h.s. of the equality, then the second-order term $O(N^{-1})$ vanishes. Thus,

$$\frac{h}{N} = \frac{1}{r}\left(a - b + c + \frac{z}{r+2}b\right) = O(N^{-1}).$$

It is also noted that $P\left(W \leq x\left(1 + h/N\right)\right) = P\left(W\left(1 - h/N\right) \leq x\right) + o(N^{-1})$. Thus, we get the Bartlett correction of Wald statistic given by

$$W^{\mathrm{BC}} = W\left(1 - \frac{1}{r}\left(a(\widehat{\boldsymbol{\theta}}) - b(\widehat{\boldsymbol{\theta}}) + c(\widehat{\boldsymbol{\theta}})\right) - \frac{x}{r(r+2)}b(\widehat{\boldsymbol{\theta}})\right),$$

whose distribution function is approximated as

$$P(W^{\mathrm{BC}} \leq x) = F_r(x) + o(N^{-1}).$$

The power distribution of the adjusted Wald test statistic can be derived from (A.8).

[Expansion and Bartlett correction of the score test] We next derive the asymptotic expansion of the score test statistic. The moment generating function of the score test statistic can be approximated as

$$\begin{aligned} E\left[e^{tS}\right] =& E\left[e^{t\left(W - \mathbf{col}(V_{(i)})'\boldsymbol{\Lambda}\mathbf{col}(V_{(i)}) + o_p(N^{-1})\right)}\right] \\ =& E[e^{tW}] - E[te^{t\boldsymbol{x}'\boldsymbol{x}}\mathbf{col}(V_{(i)})'\boldsymbol{\Lambda}\mathbf{col}(V_{(i)})] + o(N^{-1}), \end{aligned} \tag{A.9}$$

since $\lambda \equiv \mathbf{col}(V_{(i)})'\mathbf{\Lambda}\mathbf{col}(V_{(i)}) = O_p(N^{-1})$. It is noted that $V_{(i)}$ is expressed as

$$V_{(i)} = 2\boldsymbol{u}'\boldsymbol{M}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(i)}\boldsymbol{X}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{Q}'\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1})_{(i)}\boldsymbol{X}\boldsymbol{Q}\boldsymbol{x},$$

for $\boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{H}$ and $\boldsymbol{M}_{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}$. Then, it is observed that

$$\begin{aligned}
E\left[te^{t\boldsymbol{x}'\boldsymbol{x}}\lambda\right] =& \frac{t}{2}\sum_{i,j}^{q} E\Big[e^{t\boldsymbol{x}'\boldsymbol{x}}\hat{\theta}_i^\dagger\hat{\theta}_j^\dagger(\boldsymbol{x}'\boldsymbol{B}_{(i)}\boldsymbol{x}\boldsymbol{x}'\boldsymbol{B}_{(j)}\boldsymbol{x} + 4\boldsymbol{x}'\boldsymbol{A}_{(i)}\boldsymbol{u}\boldsymbol{u}'\boldsymbol{A}_{(j)}'\boldsymbol{x})\Big] \\
=& \frac{t\gamma^{\frac{r}{2}}}{2}e^{\frac{\gamma-1}{2}\boldsymbol{\delta}'\boldsymbol{\delta}}\sum_{i,j}^{q} E\Big[\hat{\theta}_i^\dagger\hat{\theta}_j^\dagger(\tilde{\boldsymbol{x}}'\boldsymbol{B}_{(i)}\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}'\boldsymbol{B}_{(j)}\tilde{\boldsymbol{x}} + +4\tilde{\boldsymbol{x}}'\boldsymbol{A}_{(i)}\boldsymbol{\Sigma}\boldsymbol{A}_{(j)}'\tilde{\boldsymbol{x}})\Big] \\
=& \frac{t\gamma^{\frac{r}{2}}}{2}e^{\frac{\gamma-1}{2}\boldsymbol{\delta}'\boldsymbol{\delta}}\sum_{i,j}^{q} E\Big[\hat{\theta}_i^\dagger\hat{\theta}_j^\dagger(\gamma^2\mathrm{tr}\,(\boldsymbol{B}_{(i)})\mathrm{tr}\,(\boldsymbol{B}_{(j)}) + 2\gamma^3\mathrm{tr}\,(\boldsymbol{B}_{(i)}\boldsymbol{\Delta})\mathrm{tr}\,(\boldsymbol{B}_{(j)}) \\
&+ \gamma^4\mathrm{tr}\,(\boldsymbol{B}_{(i)}\boldsymbol{\Delta})\mathrm{tr}\,(\boldsymbol{B}_{(j)}\boldsymbol{\Delta}) + 2\gamma^2\mathrm{tr}\,(\boldsymbol{B}_{(i)}\boldsymbol{B}_{(j)}) + 4\gamma^3\mathrm{tr}\,(\boldsymbol{B}_{(i)}\boldsymbol{B}_{(j)}\boldsymbol{\Delta}) \\
&+ 4\gamma\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}_{(i)}'\boldsymbol{A}_{(j)}) + 4\gamma^2\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}_{(i)}'\boldsymbol{A}_{(j)}\boldsymbol{\Delta}))\Big].
\end{aligned}$$

Using the observations (A.4), (A.5), (A.6) and (A.7), we can evaluate $E\left[te^{t\boldsymbol{x}'\boldsymbol{x}}\lambda\right]$ as

$$\begin{aligned}
E\left[te^{t\boldsymbol{x}'\boldsymbol{x}}\lambda\right] =& \gamma^{\frac{r}{2}}e^{\frac{\gamma-1}{2}\boldsymbol{\delta}'\boldsymbol{\delta}}\Big\{(\gamma-1)a + (\gamma-1)(b+\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{\Delta})) + (\gamma-1)^2(b+\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{\Delta})) \\
&+ \gamma^2(\gamma-1)\mathrm{tr}\,(\boldsymbol{B}\boldsymbol{\Delta}) + \gamma^3(\gamma-1)\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{\Delta})\Big\}. \qquad \text{(A.10)}
\end{aligned}$$

Combining (A.9) and (A.10) gives

$$\begin{aligned}
E[e^{tS}] =& \gamma^{\frac{r}{2}}e^{\frac{\gamma-1}{2}\boldsymbol{\delta}'\boldsymbol{\delta}}\Big\{1 + \frac{\gamma-1}{2}(-a+c) - (\gamma-1)(b+\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{\Delta})) - \frac{(\gamma-1)^2}{2}(b+\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{\Delta})) \\
&- \frac{\gamma(\gamma^2-1)}{2}\mathrm{tr}\,(\boldsymbol{B}\boldsymbol{\Delta}) + \frac{\gamma(\gamma-1)}{2}\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{\Delta}) - \frac{\gamma^2(\gamma^2-1)}{2}\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{\Delta})\Big\} + o(N^{-1}).
\end{aligned}$$

By an inversion formula, the distribution of the score test statistic can be approximated as

$$\begin{aligned}
P(S>x) =& 1 - G_r(x) + (-a-b+c-\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{\Delta}))g_{r+2}(x) - (b+\mathrm{tr}\,((\boldsymbol{A}-\boldsymbol{B}+\boldsymbol{C})\boldsymbol{\Delta}))g_{r+4}(x) \\
&- (\mathrm{tr}\,((\boldsymbol{B}+\boldsymbol{D})\boldsymbol{\Delta}))g_{r+6}(x) - \mathrm{tr}\,(\boldsymbol{D}\boldsymbol{\Delta})g_{r+8}(x) + o(N^{-1}).
\end{aligned}$$

The Bartlett correction of the score test can be derived from the above expansion under null hypothesis. Similar to the case of the Wald test, it is seen that

$$P(S \le x) = F_r(x) - \frac{x}{r}\left(-a-b+c-\frac{x}{r+2}b\right)f_r(x) + o(N^{-1}),$$

which provides the Bartlett correction

$$S^{\mathrm{BC}} = S\left(1 - \frac{1}{r}\left(-a(\widehat{\boldsymbol{\theta}}) - b(\widehat{\boldsymbol{\theta}}) + c(\widehat{\boldsymbol{\theta}})\right) + \frac{x}{r(r+2)}b(\widehat{\boldsymbol{\theta}})\right).$$

Thus, it can be confirmed that $P(S^{\text{BC}} \leq x) = F_r(x) + o(N^{-1})$. Similar to the case of the Wald test statistic, the asymptotic power functions of the score test statistic can be derived.

[Expansion and Bartlett correction of the modified LR test] Finally, we derive the Bartlett correction of the modified LR test statistic. The generating function of the modified LR test statistic can be approximated as

$$\begin{aligned}E\left[e^{tLR}\right] =& E\left[e^{t\left(W-\lambda/2+o_p(N^{-1})\right)}\right]\\ =& E[e^{tW}] - \frac{1}{2}E[te^{t\boldsymbol{x}'\boldsymbol{x}}\lambda] + o(N^{-1}).\end{aligned}$$

Combining (A.9) and (A.10) gives

$$\begin{aligned}E\left[e^{tLR}\right] =& \gamma^{\frac{r}{2}}e^{\frac{\gamma-1}{2}\boldsymbol{\delta}'\boldsymbol{\delta}}\Big\{1 + \frac{\gamma-1}{2}c - \frac{\gamma-1}{2}(b + \operatorname{tr}(\boldsymbol{A\Delta}))\\ &- \frac{\gamma(\gamma-1)}{2}\operatorname{tr}(\boldsymbol{B\Delta}) + \frac{\gamma(\gamma-1)}{2}\operatorname{tr}(\boldsymbol{C\Delta}) - \frac{\gamma^2(\gamma-1)}{2}\operatorname{tr}(\boldsymbol{D\Delta})\Big\} + o(N^{-1}),\end{aligned}$$

so that an inversion formula yields the expansion

$$\begin{aligned}P(LR > x) =& 1 - G_r(x) + (-b + c - \operatorname{tr}(\boldsymbol{A\Delta}))g_{r+2}(x)\\ &+ \operatorname{tr}((-\boldsymbol{B} + \boldsymbol{C})\boldsymbol{\Delta})g_{r+4}(x) - \operatorname{tr}(\boldsymbol{D\Delta})g_{r+6}(x) + o(N^{-1}).\end{aligned}$$

From the argument under null hypothesis, it follows that the Bartlett correction is given by

$$LR^{\text{BC}} = LR\left(1 - \frac{1}{r}\left(-b(\widehat{\boldsymbol{\theta}}) + c(\widehat{\boldsymbol{\theta}})\right)\right),$$

and we can confirm that $P(LR^{\text{BC}} \leq x) = F_r(x) + o(N^{-1})$. Similar to the case of the Wald test statistic, the asymptotic power functions of the modified LR test statistics can be derived.

Acknowledgments. This research of the second author was supported in part by Grant-in-Aid for Scientific Research Nos. 19200020 and 21540114 from Japan Society for the Promotion of Science.

References

[1] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A*, **160**, 268-282.

[2] Datta, G. S., Rao, J. N. K. and Smith, D. D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, **92**, 1, 183-196.

[3] Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of American Statistical Association*, **74**, 269-277.

[4] 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.

[5] Kubokawa, T. (2011a). On measuring uncertainty of small area estimators with higher order accuracy. *Journal of the Japan Statistical Society*, **41**, 2, 93-119.

[6] Kubokawa, T. (2011b). Conditional and unconditional methods for selecting variables in linear mixed models. *Journal of Multivariate Analysis*, **102**, 641-660.

[7] Kubokawa, T. and Erdembat, N. (2011). On testing linear hypothesis in a nested error regression model. *Communications in Statistics-Theory and Methods*, **39**, 1552-1562.

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

[9] 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.

[10] 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

[11] Rayner, R. K. (1990). Bartlett's correction and the bootstrap in normal linear regression models. *Economics Letters*, **33**, 255-258.

[12] Rothenberg, T. J. (1984). Hypothesis testing in linear models when the error covariance matrix is nonscalar. *Econometrica*, **52**, 4, 827-842.

[13] Srivastava, M.S. and Kubokawa, T. (2010). Conditional information criteria for selecting variables in linear mixed models. *Journal of Multivariate Analysis*, **101**, 1970-1980.

[14] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, **9**, 1135-1151.

[15] Vaida, F., and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, **92**, 351-370.

[16] 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.