The Statistician (1999)
48, *Part* 4, *pp.* 495–506

A degrees-of-freedom approximation for a *t*-statistic with heterogeneous variance

Stuart R. Lipsitz and Joseph G. Ibrahim

Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, USA

and Michael Parzen

University of Chicago, USA

[Received November 1998]

Summary. Ordinary least squares is a commonly used method for obtaining parameter estimates for a linear model. When the variance is heterogeneous, the ordinary least squares estimate is unbiased, but the ordinary least squares variance estimate may be biased. Various jackknife estimators of variance have been shown to be consistent for the asymptotic variance of the ordinary least squares estimator of the parameters. In small samples, simulations have shown that the coverage probabilities of confidence intervals using the jackknife variance estimators are often low. We propose a finite sample degrees-of-freedom approximation to use with the *t*-distribution, which, in simulations, appears to raise the coverage probabilities closer to the nominal level. We also discuss an 'estimated generalized least squares estimator', in which we model and estimate the variance, and use the estimated variance in a weighted regression. Even when the variance is modelled correctly, the coverage can be low in small samples, and, in terms of the coverage probability, one appears better off using ordinary least squares with a robust variance and our degrees-of-freedom approximation. An example with real data is given to demonstrate how potentially misleading conclusions may arise when the degrees-of-freedom approximation is not used.

Keywords: Homogeneous variance; Linear model; Ordinary least squares

1. Introduction

Linear regression with ordinary least squares is used extensively in statistics. Often, although assumed homogeneous, the underlying variance of the response is heterogeneous. When using ordinary least squares, if the linear model is correctly specified, and the error variance of the response is heterogeneous, the parameter estimates are unbiased and consistent (under weak regularity conditions). However, the ordinary least squares variance estimators (assuming homogeneous error variance) are usually biased and inconsistent. Alternatively, various jackknife estimators of variance have been shown to be consistent for the asymptotic variance of the ordinary least squares estimator of the parameters (Hinkley, 1977; MacKinnon and White, 1985; Wu, 1986).

Heterogeneous variance occurs, for example, when the variance of the response increases or decreases when a covariate value increases (this covariate may or may not be in the model for the response); it also occurs when the mean and variance are linked, such as a linear model for binary

Address for correspondence: Joseph G. Ibrahim, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA.
E-mail: ibrahim@jimmy.harvard.edu

0039–0526/99/48495

data. In some situations, a transformation of the response and covariates can be used to formulate a regression model with constant variance. Unfortunately, the interpretation of the model is often ruined with such a transformation. Therefore, methods that preserve the linearity of the model for the outcome as a function of covariates, but also model the variance as a function of covariates, have been developed. Methods for obtaining an appropriate variance model as a function of covariates have been developed by Cook and Weisberg (1983). After finding the appropriate variance model and estimating the variance, weighted least squares can then be used to estimate the regression parameters of the mean, in which an individual's observation is weighted by the inverse of the estimated variance. Unfortunately, when the sample size is small, even if the variance is modelled correctly, estimates of the variance parameters can be highly unstable, leading to confidence intervals for the regression parameters that have low coverage. Thus, because of the problems that can occur when estimating the variance in small samples, we propose the use of ordinary least squares estimation with a robust variance, and a degrees-of-freedom correction. The proposed degrees-of-freedom correction gives better confidence interval coverage than do parametric models for the variance function. In addition, the degrees-of-freedom approximation and the parametric variance function methods give similar coverages for large samples.

In an excellent review of methods to use with heterogeneous variance, Wu (1986) posed the model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \tag{1}$$

where Y_i is the response, x_i is a fixed covariate and the ϵ_i are independently distributed as $N(0, 0.5x_i)$. To explore finite sample properties of the jackknife, Wu fixed the sample size at 12, with the 12 covariate values equal to 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8 and 10. In these small samples, his simulations found that the coverage probabilities of confidence intervals using the jackknife variance estimators were often low. In Section 3 we discuss a finite sample degrees-of-freedom approximation to use with the t-distribution, which, in simulations, appears to raise the coverage probabilities closer to the nominal level. Section 2 discusses the notation and models, and reviews the jackknife variance estimator, which gives consistent variance estimators when the variance is heterogeneous. Section 3 discusses the degrees-of-freedom approximation. Section 4 discusses a generalized least squares method for estimating the heterogeneous variances, and then using these to re-estimate the regression parameters. Section 5 presents simulation results, comparing the ordinary least squares estimate with robust variance and the degrees-of-freedom correction with the estimated generalized least squares estimate.

To illustrate the degrees-of-freedom correction, in Section 6, we consider the gasoline vapour data set of Cook and Weisberg (1982). This data set has been analysed in detail by Cook and Weisberg (1982, 1983) and Weisberg (1985) with respect to the issue of non-constant variance. The data set contains information on the amount of gasoline vapour given off into the atmosphere when filling a tank, under various temperature and pressure conditions. In particular, the data set consists of four predictors $x_1, \ldots, x_4$ and a response variable Y, where Y denotes the amount of gasoline vapour given off, x_1 is the initial temperature of the tank, x_2 is the temperature of the gasoline dispensed, x_3 is the initial vapour pressure in the tank and x_4 is the vapour pressure of the gasoline dispensed. The data set contains 32 observations. The linear regression model for these data is given by

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_2 x_{i3} + \beta_4 x_{i4} + \epsilon_i.$$

Cook and Weisberg (1983) used these data to develop diagnostic tools for checking whether the variance function has a particular parametric form. In particular, through their proposed score statistic, they found that the error variance is a function of (x_1, x_4). Our main aim here is to demonstrate the potential effect of the degrees-of-freedom approximation on the t-tests when using the jackknife variance estimator.

2. Notation and review

We have a random sample of n individuals, in which the ith individual has response Y_i and a $P \times 1$ covariate vector $\mathbf{x}_i = (x_{i1}, \ldots, x_{iP})'$. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ be the $n \times 1$ vector of responses and $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ be the $n \times P$ design matrix. The response Y_i is assumed to be normal with first two moments

$$E(Y_i) = \mathbf{x}_i'\boldsymbol{\beta},$$

$$\text{var}(Y_i) = \sigma_i^2.$$

If σ_i^2 is unknown or is difficult to model, an attractive method for estimating $\boldsymbol{\beta}$ is ordinary least squares. The ordinary least squares estimator of $\boldsymbol{\beta}$, which is unbiased, is

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = (X'X)^{-1}X'\mathbf{Y} = \left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \sum_{i=1}^{n} \mathbf{x}_i y_i.$$

However,

$$\text{var}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) = (X'X)^{-1}X' \,\text{var}(\mathbf{Y})X(X'X)^{-1} = \left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'\sigma_i^2\right)\left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'\right)^{-1}. \tag{2}$$

The ordinary least squares estimator $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ is consistent for $\boldsymbol{\beta}$ and $n^{1/2}(\hat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})$ is asymptotically multivariate normal with mean vector $\mathbf{0}$ and covariance matrix given by n times expression (2). In equation (2), $\text{var}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) = \sigma^2(X'X)^{-1}$ if $\sigma_i^2 = \sigma^2$ for all i. If $\sigma_i^2 \neq \sigma^2$ for some i, then the usual ordinary least squares variance estimate may be biased.

Several researchers have shown that equation (2) can be consistently estimated by various jackknife variance estimators of the form

$$(X'X)^{-1}X' \,\text{diag}\left(\frac{e_i^2}{w_i}\right)X(X'X)^{-1} = \left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'\frac{e_i^2}{w_i}\right)\left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'\right)^{-1}, \tag{3}$$

where $e_i = Y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{\text{OLS}}$ is the residual, w_i is a known positive weight that converges to 1 as $n \to \infty$ and $\text{diag}(a_i)$ is a diagonal matrix with $(a_1, \ldots, a_n)$ on the main diagonal. In particular, White (1980) let $w_i = 1$ for all i, and Horn *et al.* (1975) let $w_i = 1 - h_i$, where $h_i = \mathbf{x}_i'(X'X)^{-1}\mathbf{x}_i$ is the leverage. Horn *et al.* (1975) justified their weights by noting that, if $\sigma_i^2 = \sigma^2$, then $E(e_i^2) = \sigma^2(1 - h_i)$ and their estimator of $\text{var}(\hat{\boldsymbol{\beta}})$ is unbiased. Also, dividing e_i^2 by $1 - h_i$ tends to inflate the squared residual to approximate σ_i^2 more closely. When the variance is heterogeneous, simulations have shown (MacKinnon and White, 1985) that setting $w_i = 1 - h_i$ has better finite sample properties than setting $w_i = 1$. If we let $\hat{\beta}_p$ be the pth element of $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ and $\widehat{\text{var}}(\hat{\beta}_p)$ be the pth diagonal element of expression (3) with $w_i = 1 - h_i$, then, to test H_0: $\beta_p = 0$, Wu (1986) assumed that

$$T = \frac{\hat{\beta}_p - \beta_p}{\sqrt{\widehat{\text{var}}(\hat{\beta}_p)}} \tag{4}$$

follows a t-distribution with $n - P$ degrees of freedom. However, Wu's (1986) simulations with small n found that the coverage probabilities using this t-approximation were often low. In this paper, we propose to use a t-statistic for equation (4), but with a modified degrees of freedom derived from the Satterthwaite (1946) approximation.

3. The degrees-of-freedom approximation

To use the Satterthwaite approximation, we rewrite equation (4) as

$$T = \frac{\hat{\beta}_p - \beta_p}{\surd\widehat{\mathrm{var}}(\hat{\beta}_p)} = \frac{Z_p}{\surd(Q_p/f_p)}, \tag{5}$$

where

$$Z_p = \frac{\hat{\beta}_p - \beta_p}{\surd E\{\widehat{\mathrm{var}}(\hat{\beta}_p)\}}$$

is approximately a normal (0, 1) random variable and

$$Q_p = \frac{f_p\,\widehat{\mathrm{var}}(\hat{\beta}_p)}{E\{\widehat{\mathrm{var}}(\hat{\beta}_p)\}} \tag{6}$$

is an approximate χ^2 random variable with f_p degrees of freedom. Here, we also assume that $\widehat{\mathrm{var}}(\hat{\beta}_p)$ is approximately unbiased so that $E\{\widehat{\mathrm{var}}(\hat{\beta}_p)\} = \mathrm{var}(\hat{\beta}_p)$. In the Satterthwaite approximation, the degrees of freedom f_p are chosen so that Q_p has the same first two moments as a χ^2 random variable, i.e. so that the variance of Q_p is twice its mean. Satterthwaite (1946) showed that the appropriate f_p is

$$f_p = \frac{2[E\{\widehat{\mathrm{var}}(\hat{\beta}_p)\}]^2}{\mathrm{var}\{\widehat{\mathrm{var}}(\hat{\beta}_p)\}}. \tag{7}$$

If we substitute f_p in equation (6), it is easy to see that the variance of Q_p is twice its mean. Then, we shall assume that equation (5), and equivalently equation (4), follows a t-distribution with f_p degrees of freedom. We note here that f_p can be different for the different β_p.

Thus, to use the Satterthwaite approximation, we need to estimate $E\{\widehat{\mathrm{var}}(\hat{\beta}_p)\}$ and $\mathrm{var}\{\widehat{\mathrm{var}}(\hat{\beta}_p)\}$ in equation (7). An obvious estimate of $E\{\widehat{\mathrm{var}}(\hat{\beta}_p)\}$ is $\widehat{\mathrm{var}}(\hat{\beta}_p)$. Thus, we now only need an estimate of $\mathrm{var}\{\widehat{\mathrm{var}}(\hat{\beta}_p)\}$ to estimate f_p in equation (7). To estimate $\mathrm{var}\{\widehat{\mathrm{var}}(\hat{\beta}_p)\}$, we first rewrite $\hat{\beta}_p$ and $\widehat{\mathrm{var}}(\hat{\beta}_p)$ in simpler form. We write

$$\hat{\beta}_p = \mathbf{c}_p\mathbf{Y} = \sum_{i=1}^{n} c_{pi}y_i,$$

where $\mathbf{c}_p = (c_{p1}, \ldots, c_{pn})$ is the pth row of $(X'X)^{-1}X'$. Then,

$$\mathrm{var}(\hat{\beta}_p) = \sum_{i=1}^{n} c_{pi}^2\sigma_i^2,$$

which we estimate by

$$\widehat{\mathrm{var}}(\hat{\beta}_p) = \sum_{i=1}^{n} \frac{c_{pi}^2 e_i^2}{1 - h_i} = \sum_{i=1}^{n} \left\{\frac{c_{pi}e_i}{\surd(1 - h_i)}\right\}^2. \tag{8}$$

If we let $W = \mathrm{diag}\{(1 - h_i)^{-1/2}\}$, i.e. a diagonal matrix with elements $(1 - h_i)^{-1/2}$ on the diagonal, and $\mathbf{e} = (e_1, \ldots, e_n)'$, we can rewrite equation (8) as

$$\widehat{\text{var}}(\hat{\beta}_p) = (\mathbf{c}_p W\mathbf{e})'(\mathbf{c}_p W\mathbf{e}) = \mathbf{e}'(W\mathbf{c}_p'\mathbf{c}_p W)\mathbf{e} = \mathbf{e}'A_p\mathbf{e}, \tag{9}$$

where $A_p = W\mathbf{c}_p'\mathbf{c}_p W$. Equation (9) is just a quadratic form in $\mathbf{e}$.

Thus, to obtain $\text{var}\{\widehat{\text{var}}(\hat{\beta}_p)\}$, we can use multivariate normal theory results (Searle, 1982) to obtain the variance of the quadratic form $\mathbf{e}'A_p\mathbf{e}$,

$$\text{var}\{\widehat{\text{var}}(\hat{\beta}_p)\} = \text{var}(\mathbf{e}'A_p\mathbf{e}) = 2\,\text{tr}[\{A_p\,\text{var}(\mathbf{e})\}^2] + 4\,E(\mathbf{e}')A_p\,\text{var}(\mathbf{e})A_p\,E(\mathbf{e}), \tag{10}$$

where tr[·] denotes the trace of a matrix. Then, to obtain the variance of $\widehat{\text{var}}(\hat{\beta}_p)$, we compute the mean and variance of $\mathbf{e}$ and substitute them in equation (10). Using results from linear models, it is easily shown that

$$E(\mathbf{e}) = E(\mathbf{Y} - X\hat{\boldsymbol{\beta}}_{\text{OLS}}) = 0$$

and

$$\text{var}(\mathbf{e}) = (I_n - H)\,\text{diag}(\sigma_i^2)(I_n - H),$$

where I_n is an $n \times n$ identity matrix and $H = X(X'X)^{-1}X'$ is the hat matrix. Plugging these results into equation (10), we obtain

$$\text{var}\{\widehat{\text{var}}(\hat{\beta}_p)\} = 2\,\text{tr}[\{A_p(I_n - H)\,\text{diag}(\sigma_i^2)(I_n - H)\}^2]. \tag{11}$$

After a little algebra, it can be shown that equation (11) can be rewritten as

$$\text{var}\{\widehat{\text{var}}(\hat{\beta}_p)\} = 2\,\text{tr}(\{(I_n - H)A_p(I_n - H)\}[\{(I_n - H)A_p(I_n - H)\}\,\#\boldsymbol{\Sigma}]), \tag{12}$$

where the symbol # represents element multiplication (the *ij*th element of the new matrix is the product of the *ij*th elements of the matrices that you are 'multiplying'). Also, $\boldsymbol{\Sigma}$ has *i*th diagonal elements equal to σ_i^4 and *ij*th off-diagonal elements equal to $\sigma_i^2\sigma_j^2$. Thus, we need estimates of σ_i^4 and $\sigma_i^2\sigma_j^2$. We propose 'almost unbiased' estimates of these two by noting that, if $\sigma_i^2 = \sigma^2$, then

$$E(e_i^4) = 3\sigma^4(1 - h_i)^2$$

and

$$E(e_i^2e_j^2) = \text{var}(Y_i)\,\text{var}(Y_j)\{2h_{ij}^2 + (1 - h_i)(1 - h_j)\},$$

where h_{ij} is the *ij*th element of the hat matrix H. Thus, we propose to estimate $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\Sigma}}$, which has *i*th diagonal element equal to

$$\hat{\sigma}_i^4 = \frac{e_i^4}{3(1 - h_i)^2} \tag{13}$$

and *ij*th off-diagonal element equal to

$$\hat{\sigma}_i^2\hat{\sigma}_j^2 = \frac{e_i^2e_j^2}{2h_{ij}^2 + (1 - h_i)(1 - h_j)}. \tag{14}$$

As $n \to \infty$, $h_{ij} \to 0$ and $h_i \to 0$, and equations (13) and (14) are asymptotically unbiased for σ_i^4 and $\sigma_i^2\sigma_j^2$ respectively.

Thus, we propose to estimate the degrees of freedom from the Satterthwaite approximation by

$$\hat{f}_p = \frac{2\,\widehat{\text{var}}(\hat{\beta}_p)^2}{2\,\text{tr}(\{(I_n - H)A_p(I_n - H)\}[\{(I_n - H)A_p(I_n - H)\}\,\#\hat{\mathbf{\Sigma}}])}$$

$$= \frac{\widehat{\text{var}}(\hat{\beta}_p)^2}{\text{tr}(\{(I_n - H)A_p(I_n - H)\}[\{(I_n - H)A_p(I_n - H)\}\,\#\hat{\mathbf{\Sigma}}])}. \qquad (15)$$

4. Generalized least squares

If $\text{var}(Y_i) = \sigma_i^2$ is known, then the optimal linear estimation procedure is to estimate $\boldsymbol{\beta}$ via generalized least squares, i.e.

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\mathbf{x}_i\mathbf{x}_i'\right)^{-1} \sum_{i=1}^{n} \frac{1}{\sigma_i^2}\mathbf{x}_i y_i, \qquad (16)$$

which is unbiased. Furthermore (Mardia *et al.*, 1979), $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ is consistent for $\boldsymbol{\beta}$ and $n^{1/2}(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta})$ is asymptotically multivariate normal with mean vector $\mathbf{0}$ and covariance matrix given by n times

$$\text{var}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}. \qquad (17)$$

Since this generalized least squares problem can be transformed to ordinary least squares by dividing Y_i and $\mathbf{x}_i$ by σ_i,

$$T = \frac{\hat{\beta}_p - \beta_p}{\surd\widehat{\text{var}}(\hat{\beta}_p)} \qquad (18)$$

follows a t-distribution with $n - P$ degrees of freedom under the null hypothesis H_0: $\beta_p = 0$.

Unfortunately, we do not know σ_i^2. An estimated generalized least squares estimator of $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}_{\text{EGLS}}$, obtained by replacing σ_i^2 in equation (16) with a consistent estimator, say $\hat{\sigma}_i^2$, will have the same asymptotic properties as $\hat{\boldsymbol{\beta}}_{\text{GLS}}$. The variance as a function of covariates is often modelled as

$$\sigma_i^2 = g(\mathbf{u}_i'\boldsymbol{\alpha}), \qquad (19)$$

where $g(\cdot)$ is the identity function $g(\mathbf{u}_i'\boldsymbol{\alpha}) = \mathbf{u}_i'\boldsymbol{\alpha}$ or the exponential function $g(\mathbf{u}_i'\boldsymbol{\alpha}) = \exp(\mathbf{u}_i'\boldsymbol{\alpha})$ and $\mathbf{u}_i$ is some function of the covariates (often, $\mathbf{u}_i$ equals $\mathbf{x}_i$, but this is not necessary).

By modelling the variance structure as in equation (19), we hope to obtain an estimator which is more efficient than ordinary least squares. The mean of the asymptotic distribution of $e_i^2 = (Y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{\text{OLS}})^2$ is σ_i^2, i.e. $e_i^2 = \sigma_i^2 + r_i$ (with r_i having asymptotic mean 0). This suggests a set of estimating equations for σ_i^2 of the form

$$\sum_{i=1}^{n} \mathbf{u}_i'\{e_i^2 - g(\mathbf{u}_i'\boldsymbol{\alpha})\} = \mathbf{0}. \qquad (20)$$

If $g(\cdot)$ is the linear function, then the solution to equations (20) is the ordinary least squares estimate, i.e.

$$\hat{\boldsymbol{\alpha}} = \left(\sum_{i=1}^{n} \mathbf{u}_i\mathbf{u}_i'\right)^{-1} \sum_{i=1}^{n} \mathbf{u}_i e_i^2.$$

If $g(\cdot)$ is the exponential function, then the solution to equations (20) can be obtained by using any generalized linear models program (such as SAS procedure `Genmod` (SAS Institute, 1993)) with Poisson error structure and a log-link. Then $\boldsymbol{\beta}$ is re-estimated by equation (16) with σ_i^2 replaced by $g(\mathbf{u}_i'\hat{\boldsymbol{\alpha}})$. This is a three-stage technique in which stage 1 is ordinary least squares for $\boldsymbol{\beta}$, stage 2 is ordinary least squares or a generalized linear model for $\boldsymbol{\alpha}$ and stage 3 is estimated generalized least squares for $\boldsymbol{\beta}$ using $\hat{\sigma}_i$ from stage 2. In our analyses, we iterate among the three stages until a convergence criterion is met for $\hat{\boldsymbol{\beta}}_{\mathrm{EGLS}}$ and $\hat{\boldsymbol{\alpha}}$. The three-stage estimator and the fully iterated estimator have the same asymptotic properties.

Using the same type of asymptotic expansions as in Prentice (1988), assuming that equation (19) is correctly specified, $\hat{\boldsymbol{\beta}}_{\mathrm{EGLS}}$ is consistent and has the asymptotic covariance matrix consistently estimated by

$$\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}_{\mathrm{EGLS}}) = \left(\sum_{i=1}^{n} \frac{1}{\hat{\sigma}_i^2} \mathbf{x}_i \mathbf{x}_i' \right)^{-1}. \tag{21}$$

The theory on estimated generalized least squares (Maddala, 1971; De Gruttola *et al.*, 1987) proposes treating the estimates $\hat{\sigma}_i^2$ as if they are known (and equal to σ_i^2) when making inferences about $\boldsymbol{\beta}$. As such, the ratio of an element of $\hat{\boldsymbol{\beta}}_{\mathrm{EGLS}}$ to its estimated standard error is taken to have the same distribution as the ratio of an element of $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ to its estimated standard error, which is a t-distribution with $n - P$ degrees of freedom. Unfortunately, when the sample size is small, the estimate of σ_i^2 based on equations (20) can be variable and, assuming that it is known, and the resulting t-statistic has $n - P$ degrees of freedom, can result in low coverage, as seen in the simulations in the next section.

5. Simulation

We performed simulations to compare coverage probabilities using the following:

(a) ordinary least squares with $\hat{\sigma}_i^2 = \hat{\sigma}^2$ and $n - P$ degrees of freedom;
(b) ordinary least squares with variance estimator (3), $\hat{\sigma}_i^2 = e_i^2/(1 - h_i)$ and $n - P$ degrees of freedom;
(c) ordinary least squares with variance estimator (3), $\hat{\sigma}_i^2 = e_i^2/(1 - h_i)$ and $\hat{f}_p$ degrees of freedom;
(d) estimated generalized least squares with $\hat{\sigma}_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_i$ and $n - P$ degrees of freedom.

The linear regression model considered was

$$Y_i = 0.4x_i - 0.25x_i^2 + \epsilon_i, \tag{22}$$

where the x_i are fixed, and the ϵ_i are independent. We performed sets of simulations when both $\epsilon_i \sim N(0, x_i)$ and $\epsilon_i \sim N(0, 1)$. As such, the estimated variance when using estimated generalized least squares, $\hat{\sigma}_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_i$, is correctly specified (actually overspecified) in both cases. To explore the finite sample properties of the coverage probabilities, we considered sample sizes of 12, 24 and 48, with 12 distinct covariate values equal to 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8 and 10. For $n = 24$, we doubled these values, and for $n = 48$ we quadrupled them.

We are most interested in coverage probabilities for confidence intervals. Specifically, we look at 95% confidence intervals of the form

$$\hat{\beta}_{pb} \pm t_{0.025,\mathrm{df}_p} v_{pb}^{1/2},$$

where $\hat{\beta}_{pb}$ is an estimate of the pth element of $\boldsymbol{\beta}$ from the bth simulation and v_{pb} is the estimate of its variance using one of the methods discussed. Also, $t_{0.025,\mathrm{df}_p}$ is the 97.5th percentile from a t-distribution with df_p degrees of freedom. With 1825 replications, using large sample normal theory, the proportion of confidence intervals containing the true value should be in the closed interval [94, 95] if the true coverage probability is 95%. Also of interest are the average lengths of the confidence intervals and the average degrees of freedom. The results are given in Tables 1 and 2.

Table 1 contains the simulations under heterogeneous variance. When wrongly assuming constant variance ($\sigma_i^2 = \sigma^2$), we see that the coverage probabilities do not appear to depend very much on the sample size. For β_0, the coverage is high (about 98%), for β_1, the coverage is just a little low (about 93.5%) and, for β_2, it is low (about 90%). When using the robust variance estimator with df $= n - P$, the coverage for the intercept is apparently correct, for β_1, the coverage is slightly low and, for β_2, it is also low. As n increases, the coverage probabilities for β_1 and β_2 increase closer to the nominal level, though still low for β_2. The degrees-of-freedom correction f_p appears to make the coverage for β_1 and β_2 not significantly different from the nominal level. Note that the lengths of the confidence intervals are larger when using our degrees-of-freedom approximation. Further, the average degrees of freedom are much smaller using our degrees-of-freedom approximation. This is particularly true for β_2. When $n = 48$, the average of

Table 1. Heterogeneous variance—coverage probabilities and average lengths of confidence intervals from simulations with $(\beta_0, \beta_1, \beta_2) = (0, 0.4, -0.25)$ and var$(Y|x) = x$

Parameter	*n*	*Results from the following methods*:			
		Ordinary least squares, $\hat\sigma_i^2 = \hat\sigma^2$, $df_p = n - P$	*Ordinary least squares*, $\hat\sigma_i^2 = e_i^2/(1 - h_i)$,		*Estimated generalized least squares*, $\hat\sigma_i^2 = \hat\alpha_0 + \hat\alpha_1 x_i$, $df_p = n - P$
			$df_p = n - P$	$df_p = \hat f_p$	
β_0	12	97.5†	95.1	96.2	95.2
		4.45‡	3.76	4.12	3.86
		9.0§	9.0	4.5	9.0
	24	98.3	95.4	95.9	94.7
		3.00	2.51	2.62	2.51
		21.0	21.0	14.2	21.0
	48	98.8	95.3	95.5	94.6
		2.10	1.77	1.69	1.65
		45.0	45.0	28.7	45.0
β_1	12	93.6	93.2	95.5	93.2
		2.01	1.99	2.20	1.88
		9.0	9.0	6.9	9.0
	24	93.5	93.9	94.8	94.2
		1.35	1.37	1.46	1.31
		21.0	21.0	13.1	21.0
	48	93.7	93.9	94.9	94.9
		0.95	0.99	0.94	0.88
		45.0	45.0	23.4	45.0
β_2	12	91.0	90.5	94.6	90.8
		0.18	0.19	0.59	0.18
		9.0	9.0	5.7	9.0
	24	90.9	92.2	93.9	93.1
		0.14	0.14	0.15	0.13
		21.0	21.0	10.0	21.0
	48	90.1	92.9	94.5	94.7
		0.087	0.098	0.110	0.091
		45.0	45.0	16.2	45.0

†The first entry in each cell is the coverage probability.

‡The second entry in each cell is the average length of the confidence interval.

§The third entry in each cell is the average degrees of freedom.

Table 2. Homogeneous variance—coverage probabilities and average lengths of confidence intervals from simulations with $(\beta_0, \beta_1, \beta_2) = (0, 0.4, -0.25)$ and $\mathrm{var}(Y|x) = 1$

Parameter	*n*	*Results from the following methods*:			
		Ordinary least squares, $\hat{\sigma}_i^2 = \hat{\sigma}^2$, $df_p = n - P$	*Ordinary least squares,* $\hat{\sigma}_i^2 = e_i^2/(1 - h_i)$,		*Estimated generalized least squares,* $\hat{\sigma}_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_i$, $df_p = n - P$
			$df_p = n - P$	$df_p = \hat{f}_p$	
β_0	12	95.5†	92.5	95.1	92.9
		2.22‡	2.15	2.32	2.25
		9.0§	9.0	5.8	9.0
	24	94.1	93.2	94.4	94.4
		1.46	1.43	1.51	1.48
		21.0	21.0	12.5	21.0
	48	94.8	94.5	95.1	94.9
		1.01	1.00	1.03	1.00
		45.0	45.0	23.8	45.0
β_1	12	94.5	93.5	95.3	93.6
		1.00	0.98	1.05	1.01
		9.0	9.0	6.9	9.0
	24	95.4	93.5	94.6	93.6
		0.66	0.65	0.71	0.68
		21.0	21.0	14.7	21.0
	48	94.6	94.2	95.1	94.4
		0.45	0.45	0.65	0.48
		45.0	45.0	29.1	45.0
β_2	12	95.3	92.8	95.7	92.3
		0.92	0.089	0.10	0.094
		9.0	9.0	6.1	9.0
	24	94.5	93.1	93.9	93.6
		0.061	0.059	0.063	0.060
		21.0	21.0	12.9	21.0
	48	94.1	93.8	94.5	94.9
		0.042	0.042	0.043	0.041
		45.0	45.0	24.5	45.0

†The first entry in each cell is the coverage probability.
‡The second entry in each cell is the average length of the confidence interval.
§The third entry in each cell is the average degrees of freedom.

the degrees of freedom $(\hat{f}_p)$ is only 16.2. When using $\hat{\boldsymbol{\beta}}_{\mathrm{EGLS}}$, the coverage is low (about 91%) for β_2 when $n = 12$ and slightly low (about 93%) for β_2 when $n = 24$. Thus, even when the variance is modelled correctly, the coverage can be low in small samples, and, in terms of coverage probability, one is better off using ordinary least squares with a robust variance and our degrees-of-freedom approximation. Fig. 1 gives a plot of the coverage probabilities as a function of n for β_2 (based on more simulations than those given in Table 1). As stated above, we see that using ordinary least squares with a robust variance estimate and our degrees-of-freedom approximation has the best coverage.

Looking at Table 2, in which the true model has homogeneous variance, we see that the usual ordinary least squares method gives correct coverage probabilities, as expected. The coverage probability using the robust variance estimator with $n - P$ degrees of freedom appears low, although becoming closer to the nominal level as n increases. Finally, our degrees-of-freedom approximation tends to increase the coverage to the nominal level. When using $\hat{\boldsymbol{\beta}}_{\mathrm{EGLS}}$, the coverage is slightly low (about 92%) for β_3 when $n = 12$.

Because of the broad range of possible data configurations, it is difficult to draw definitive

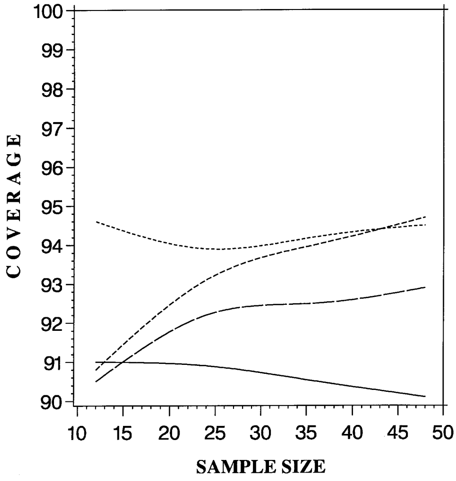

Fig. 1. Coverage probabilities as a function of n for β_2 from simulations with $(\beta_0, \beta_1, \beta_2) = (0, 0.4, -0.25)$ and $\mathrm{var}(Y|x) = x$: ··········, ordinary least squares with $\hat{\sigma}_i^2 = e_i^2/(1 - h_i)$ and $\mathrm{df}_p = \hat{f}_p$; - - - - - -, estimated generalized least squares; – – – –, ordinary least squares with $\hat{\sigma}_i^2 = e_i^2/(1 - h_i)$ and $\mathrm{df}_p = n - P$; ———, ordinary least squares with $\hat{\sigma}_i^2 = \hat{\sigma}^2$ and $\mathrm{df}_p = n - P$

conclusions from a simulation. We can only make general suggestions. We have looked at a simple case to study the properties of the degrees-of-freedom approximation. We encourage using our degrees-of-freedom approximation $\hat{f}_p$ since, when n is large, it will give the same coverage as the usual degrees of freedom $n - P$ and, when n is small, it appears more appropriate in our limited simulations and can increase the coverage probability by up to 5%. Further, it apparently has a better coverage probability in small samples than does the estimated generalized least squares estimate with correctly modelled variance. When analysing a data set in which the error variance appears heterogeneous, we suggest that the data analyst should calculate our proposed degrees of freedom $\hat{f}_p$ and, if $\hat{f}_p \leqslant 30$, we suggest that our degrees-of-freedom approximation should be used.

6. Example

Gasoline vapours are a significant source of air pollution. One source of gasoline vapours occurs when gasoline is pumped into a tank and vapours are forced into the atmosphere. A laboratory experiment was conducted to determine the amount of gasoline vapours given off into the atmosphere under different temperature and vapour conditions (Cook and Weisberg, 1982). The model of interest is

$$E(Y_i|x_{i1}, x_{i2}, x_{i3}, x_{i4}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4},$$

where Y_i is the amount of vapour given off, x_{i1} is the initial temperature of the tank, x_{i2} is the temperature of the gasoline dispensed, x_{i3} is the initial vapour pressure in the tank and x_{i4} is the vapour pressure of the gasoline dispensed. Through their proposed score statistic, Cook and Weisberg (1982) found that the error variance is of the form

$$\mathrm{var}(Y_i|x_{i1}, x_{i2}, x_{i3}, x_{i4}) = \exp(\alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i4}).$$

Tables 3–6 give the t-tests and p-values obtained using the following respective variance estimators:

Table 3. Ordinary least squares estimates of variance

Parameter	*Degrees of freedom for error*	*Estimate*	*Standard error*	*t-statistic*	*p-value*
β_0	27	0.7663	2.0214	0.379	0.7076
β_1	27	0.0151	0.0903	0.167	0.8683
β_2	27	0.1992	0.0709	2.808	0.0092
β_3	27	−4.8195	2.9318	−1.644	0.1118
β_4	27	9.1439	2.9497	3.100	0.0045

Table 4. Robust estimates of variance with $n - P$ degrees of freedom

Parameter	*Degrees of freedom for error*	*Estimate*	*Standard error*	*t-statistic*	*p-value*
β_0	27	0.7663	1.7390	0.4406	0.6630
β_1	27	0.0151	0.0923	0.1637	0.8711
β_2	27	0.1992	0.0573	3.4783	0.0017
β_3	27	−4.8195	3.9144	−1.2312	0.2288
β_4	27	9.1439	3.9841	2.2950	0.0297

Table 5. Robust estimates of variance with proposed degrees-of-freedom correction

Parameter	*Degrees of freedom for error*	*Estimate*	*Standard error*	*t-statistic*	*p-value*
β_0	13.628	0.7663	1.7390	0.4406	0.6664
β_1	16.947	0.0151	0.0923	0.1637	0.8718
β_2	10.282	0.1992	0.0573	3.4783	0.0057
β_3	8.143	−4.8195	3.9144	−1.2312	0.2526
β_4	5.789	9.1439	3.9841	2.2950	0.0631

Table 6. Estimated generalized least squares estimates with $\hat{\sigma}_i^2 = \exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{i1} + \hat{\alpha}_2 x_{i4})$

Parameter	*Degrees of freedom for error*	*Estimate*	*Standard error*	*t-statistic*	*p-value*
β_0	27	−0.2610	1.6747	−0.1558	0.8773
β_1	27	0.0219	0.0759	0.2884	0.7752
β_2	27	0.1863	0.0569	3.2725	0.0029
β_3	27	−4.7927	2.9221	−1.6401	0.1126
β_4	27	9.4411	2.9246	3.2282	0.0033

(a) ordinary least squares with $\hat{\sigma}_i^2 = \hat{\sigma}^2$ and $n - P$ degrees of freedom;
(b) ordinary least squares with variance estimator (3), $\hat{\sigma}_i^2 = e_i^2/(1 - h_i)$ and $n - P$ degrees of freedom;
(c) ordinary least squares with variance estimator (3), $\hat{\sigma}_i^2 = e_i^2/(1 - h_i)$ and $\hat{f}_p$ degrees of freedom;
(d) estimated generalized least squares with $\hat{\sigma}_i^2 = \exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{i1} + \hat{\alpha}_2 x_{i4})$ and $n - P$ degrees of freedom.

From Tables 3–6, we observe several interesting results. First, we see how the degrees of freedom change dramatically from Table 4 to Table 5, especially for β_4. Secondly, using the usual ordinary least squares method with homogeneous variance and $n - P$ degrees of freedom, we see from Table 3 that the p-value for β_4 is highly significant (p-value 0.0045) and becomes less significant (Table 4) when the robust estimate of variance is used (p-value 0.0297). However, when the degrees-of-freedom correction is used (Table 5), we see that the p-value for β_4 becomes non-significant at the 5% level (p-value 0.0631). Further, using estimated generalized least squares, we obtain results that are similar to the usual ordinary least squares results (p-value 0.0033).

Thus, the effect of x_{i4} is reduced greatly when using ordinary least squares with robust variance and the degrees-of-freedom approximation $\hat{f}_p$, and we see how potentially different conclusions can be reached when the degrees-of-freedom approximation is used. This example agrees with the simulations in Section 4: the ordinary least squares variance and robust variance with degrees of freedom $n - P$ and estimated generalized least squares can be anticonservative in small samples.

References

Cook, R. D. and Weisberg, S. (1982) *Residuals and Influence in Regression*. London: Chapman and Hall.

———(1983) Diagnostics for heteroscedasticity in regression. *Biometrika*, **70**, 1–10.

De Gruttola, V., Ware, J. H. and Louis, T. A. (1987) Influence analysis of generalized least squares estimators. *J. Am. Statist. Ass.*, **82**, 911–917.

Hinkley, D. V. (1977) Jackknifing in unbalanced situations. *Technometrics*, **19**, 285–292.

Horn, S. D., Horn, R. A. and Duncan, D. B. (1975) Estimating heteroscedastic variances in linear models. *J. Am. Statist. Ass.*, **70**, 380–385.

MacKinnon, J. G. and White, H. (1985) Some heteroscedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econometr.*, **29**, 305–325.

Maddala, G. S. (1971) Generalized least squares with an estimated variance covariance matrix. *Econometrica*, **39**, 23–33.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) *Multivariate Analysis*. New York: Academic Press.

Prentice, R. L. (1988) Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033–1048.

SAS Institute (1993) The GENMOD procedure, release 6.09. *Technical Report P-243, SAS/STAT Software*. SAS Institute, Cary.

Satterthwaite, F. E. (1946) An approximate distribution of estimates of variance components. *Biometr. Bull.*, **2**, 110–114.

Searle, S. R. (1982) *Matrix Algebra Useful for Statistics*. New York: Wiley.

Weisberg, S. (1985) *Applied Linear Regression*, 2nd edn. New York: Wiley.

White, H. (1980) A heteroscedasticity-consistent covariance matrix and a direct test for heteroscedasticity. *Econometrica*, **48**, 817–838.

Wu, C. F. J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.*, **14**, 1261–1295.