Corrected Maximum Likelihood Estimators in Linear Heteroskedastic Regression Models*

Gauss M. Cordeiro**

Abstract

The linear heteroskedastic regression model, for which the variance of the response is given by a suitable function of a set of linear exogenous variables, is very useful in econometric applications. We derive a simple matrix formula for the n^{-1} biases of the maximum likelihood estimators of the parameters in the variance of the response, where n is the sample size. These biases are easily obtained as a vector of regression coefficients in a simple weighted least squares regression. We use simulation to compare the uncorrected estimators with the bias-corrected ones to conclude the superiority of the corrected estimators over the uncorrected ones with regard to the normal approximation. The practical use of such biases is illustrated in two applications to real data sets.

Keywords: Bias Correction, Fisher Scoring Method, Heteroskedastic Model, Linear Regression, Maximum Likelihood, Weighted Least Squares.

JEL Codes: C4, C5.

*Submitted in October 2006. Revised in March 2008. The author is very grateful to two referees and to the Editor for helpful comments that considerably improved the paper.

**Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco (UFRPE). Rua Dom Manoel de Medeiros, s/n, 50171-900 – Recife, PE, Brazil.
E-mail: gauss@deinfo.ufrpe.br

1. Introduction

When economic data are used, *heteroskedasticity* can arise from a variety of sources, and the solution depends on the nature of the problem identified. Heteroskedastic regression models have recently gained popularity in applications for analyzing unreplicated experiments, experiments for robust design and the analysis of process data. Many authors have also considered variance modeling to obtain corrected standard errors and confidence intervals for mean parameters in regression analysis. Judge et al. (1985) proposed a variety of solutions when the heteroskedasticity is thought to be related to exogenous variables. They basically discussed the estimation of parameters in the following cases when the variance σ_i^2 of the response y_i is a function of exogenous variables: (i) $\sigma_i^2 = (z_i^T\gamma)^2$; (ii) $\sigma_i^2 = z_i^T\gamma$; (iii) $\sigma_i^2 = (x_i^T\beta)^p$; (iv) $\sigma_i^2 = \exp(z_i^T\gamma)$. In case (iv), the log of the variance is a linear function of exogenous variables which leads to a multiplicative heteroskedastic model. Clearly, there are other ways to model heteroskedasticity. If we are only interested in parameters of the mean in the presence of heteroskedasticity, then robust inference does not involve modeling the variance.

An important issue in asymptotic likelihood theory is the calculation of the second-order biases of the maximum likelihood estimators (MLEs). These estimators typically have biases of order $\mathcal{O}\left(n^{-1}\right)$ for large sample size n, which are commonly ignored in practice, the justification being that they are small when compared to the standard errors of the parameter estimators of order $\mathcal{O}\left(n^{-1/2}\right)$. For small sample sizes, however, the second-order biases can be appreciable and of the same magnitude as the corresponding standard errors. In such cases the biases should not be neglected, and including already second-order bias corrections may substantially improve the MLEs, whereas higher-order corrections are usually much more difficult to evaluate (Cordeiro and Barroso, 2007) and yield lesser improvement in comparison with second-order bias corrections.

In order to make the estimation of second-order bias corrections feasible in practical applications, corresponding formulas for their calculation need to be established for a wide range of regression models. Unfortunately, it is common in econometric applications to consider that the MLEs are unbiased and the main textbooks in econometrics usually do not give expressions or references for second-order biases of these estimators for the most important regression models. Practitioners usually ignore these biases based on the first-order asymptotic theory. To improve the accuracy of the MLEs, substantial effort has been placed upon computing the cumulants of log-likelihood derivatives which are, however, notoriously cumbersome; see, e.g. Cordeiro and McCullagh (1991) and the references therein. This paper shows how to easily compute the second-order biases of the MLEs in linear heteroskedastic regression models (without computing cumulants) from a vector of regression coefficients in a suitable weighted linear regression.

A general formula for the $\mathcal{O}\left(n^{-1}\right)$ multiparameter biases of the MLEs for a model with p parameters was given by Cox and Snell (1968) and Cordeiro and

McCullagh (1991). Some closed-form expressions for second-order biases of the MLEs in regression models are available in the literature which do not involve cumulants of log-likelihood derivatives. In this paper we obtain bias-corrected estimators for linear heteroskedastic regression models for which the variance of the response takes the form $\sigma_i^2 = h(z_i^T\gamma)$, where $h(\cdot)$ is a known function of a linear combination of exogenous variables $\tau_i = z_i^T\gamma$ for $i = 1, \ldots, n$. There are many possible ways to choose $h(\cdot)$ such as those cases (i), (ii) and (iv) mentioned before. We thus generalize the results by Cordeiro (1993), which hold only for the multiplicative heteroskedastic model (case (iv)). We propose bias-corrected estimators that have good performance in small samples as illustrated in some Monte Carlo simulation studies.

We consider the linear heteroskedastic regression models

$$\begin{aligned} E(y) &= \mu = X\beta, \quad Cov(y) = \Lambda \\ \Lambda &= diag\{\sigma_1^2, \ldots, \sigma_n^2\} \\ \sigma_i^2 &= h(z_i^T\gamma) \end{aligned} \tag{1}$$

where X is a specified $n \times p$ matrix of full rank $p < n$, $\beta = (\beta_1, \ldots, \beta_p)^T$ is a set of unknown parameters, $h(\cdot)$ is any continuous twice differentiable function, γ is a $q \times 1$ vector of unknown parameters and z_i^T is a $1 \times q$ vector of known exogenous variables. The variables $z_i^{T\prime}s$ may be functions of the regression variables in X. Let $Z^T = (z_1, \ldots, z_n)$ be a $q \times n$ matrix obtained from the $z_i'^s$ assumed to be of full rank $q < n$, and $\tau = Z\gamma$. Further, we consider that the responses $y_1, \ldots, y_n$ are normally distributed $y_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$. Throughout the paper, we assume that X and z are non-stochastic and the parameters β and γ functionally independent and then there are $p + q$ parameters to be estimated.

The paper is organized as follows. In Section 2, we show that any software with a weighted linear regression routine can be used to iteratively calculate the MLEs $\hat{\beta}$ and $\hat{\gamma}$ of the parameters β and γ. In Section 3, we show that the second-order bias of $\hat{\beta}$ is zero and provide a simple formula for the n^{-1} bias vector of $\hat{\gamma}$. Some special cases are also discussed. In Section 4, we present some simulation results for model (1) to show that the bias-corrected estimators produce smaller biases than the usual MLEs without increasing the mean squared errors. Moreover, we use the simulation results to assess the adequacy of the normal approximation for the empirical distributions of the uncorrected and corrected estimators. Finally, Section 5 gives two applications of bias correction to real data sets.

2. Maximum Likelihood Estimator

For the linear heteroskedastic regression model (1), the total log-likelihood for $\theta = (\beta^T, \gamma^T)^T$ given $y = (y_1, \ldots, y_n)^T$ can be written apart from a constant

$$l(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\log h_i - \frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-\mu_i)^2}{h_i} \tag{2}$$

where $h_i = h(z_i^T\gamma)$. We assume that the function $l = l(\theta)$ is regular with respect to all β and γ derivatives up to and including those of fourth order.

We now introduce the following total log-likelihood derivatives in which we reserve lower-case subscripts $r, s, t, \ldots$ to denote components of the β vector and upper-case subscripts $R, S, T, \ldots$ for components of the γ vector: $U_r = \partial l/\partial\beta_r$, $U_{rS} = \partial^2 l/\partial\beta_r\partial\gamma_S$, $U_{rST} = \partial^3 l/\partial\beta_r\partial\gamma_S\partial\gamma_T$, and so on. The standard notation will be adopted for the cumulants of the log-likelihood derivatives $\kappa_{rs} = E(U_{rs})$, $\kappa_{r,s} = E(U_rU_s)$, $\kappa_{rsT} = E(U_{rsT})$, etc., where all $\kappa's$ refer to a total over the sample and are, in general, of order n. We also define the derivatives of the cumulants by $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, $\kappa_{rs}^{(T)} = \partial\kappa_{rs}/\partial\gamma_T$, etc.

We can easily obtain the following second-order derivatives of the log-likelihood from (1)

$$U_r = \sum_i \frac{(y_i-\mu_i)x_{ir}}{h_i}, \quad U_{rs} = -\sum_i \frac{x_{ir}x_{is}}{h_i}$$

$$U_R = -\frac{1}{2}\sum_i \frac{h_i' z_{iR}}{h_i} + \frac{1}{2}\sum_i \frac{(y_i-\mu_i)^2 h_i' z_{iR}}{h_i^2}$$

$$U_{RS} = -\frac{1}{2}\sum_i \left(\frac{h_i'}{h_i}\right)' z_{iR}z_{iS} + \frac{1}{2}\sum_i (y_i-\mu_i)^2\left(\frac{h_i'}{h_i^2}\right)' z_{iR}z_{iS}$$

and

$$U_{rS} = -\sum_i \frac{(y_i-\mu_i)h_i' x_{ir}z_{iS}}{h_i^2}$$

where the dashes here and henceforth denote derivatives with respect to $\tau_i = z_i^T\gamma$.

The score functions for β and γ are then given by

$$U_\beta = X^T\Lambda^{-1}(y-\mu) \tag{3}$$

and

$$U_\gamma = -\frac{1}{2}Z^T\Lambda^{-1}F\,1 + \frac{1}{2}Z^T\Lambda^{-2}\,F\,v \tag{4}$$

where $\Lambda = diag\{h_i\}$ and $F = diag\{h_i'\}$ are diagonal matrices of order n, 1 is an $n \times 1$ vector of ones and $v = (v_1, \ldots, v_n)^T$ has the ith component given by $v_i = (y_i - \mu_i)^2$. Let $\kappa_{r,s} = -\kappa_{rs}$ and $\kappa_{R,S} = -\kappa_{RS}$ be typical elements of the information matrices K_β and K_γ for β and γ, respectively. We have

$K_\beta = X^T\Lambda^{-1}X$ and $K_\gamma = Z^TVZ$, where $V = diag\{\frac{h_i'^2}{2h_i^2}\}$ is a diagonal matrix of order n.

The parameters β and γ are globally orthogonal since $\kappa_{rS} = 0$ and then the estimates $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically independent due to their normality and to the block diagonal structure of the information matrix. We obtain $\hat{\beta}$ and $\hat{\gamma}$ iteratively from Fisher scoring method

$$\begin{pmatrix} \beta^{(m+1)} \\ \gamma^{(m+1)} \end{pmatrix} = \begin{pmatrix} \beta^{(m)} \\ \gamma^{(m)} \end{pmatrix} + \begin{pmatrix} K_\beta^{(m)^{-1}} & 0 \\ 0 & K_\gamma^{(m)^{-1}} \end{pmatrix} \begin{pmatrix} U_\beta^{(m)} \\ U_\gamma^{(m)} \end{pmatrix} \tag{5}$$

where U_β and U_γ are given by (3) and (4), respectively. From (5) we can write

$$\beta^{(m+1)} = (X^T\Lambda^{(m)^{-1}}X)^{-1}X^T\Lambda^{(m)^{-1}}y \tag{6}$$

and

$$\gamma^{(m+1)} = (X^TV^{(m)}Z)^{-1}Z^TV^{(m)}\,(\tau^{(m)} + \delta^{(m)}) \tag{7}$$

where

$$\delta = \frac{1}{2}\,V^{-1}\Lambda^{-1}(\Lambda^{-1}Fv - F1)$$

is an $n \times 1$ vector. Each loop through the iterative scheme given by equations (6) and (7) can be solved by iteratively weighted least squares and any software with a weighted linear regression routine can be used to calculate $\hat{\beta}$ and $\hat{\gamma}$ iteratively. Initial approximations $\beta^{(1)}$ and $\gamma^{(1)}$ for the iterative algorithm are used to evaluate $\Lambda^{(1)}$, $V^{(1)}$, $\tau^{(1)}$, $F^{(1)}$, $v^{(1)}$ and $\delta^{(1)}$ from which equations (6) and (7) can be used to obtain the next estimates $\beta^{(2)}$ and $\gamma^{(2)}$. These new values can update Λ, V, τ, F, v and δ and so the iterations continue until convergence is observed. The asymptotic covariance matrices of the estimators $\hat{\beta}$ and $\hat{\gamma}$ are just given by $\widehat{K}_\beta^{-1} = (X^T\widehat{\Lambda}^{-1}X)^{-1}$ and $\widehat{K}_\gamma^{-1} = (Z^T\widehat{V}Z)^{-1}$, respectively.

3. Biases of the Estimators of β and γ

In recent years there has been considerable interest in finding formulas for the n^{-1} biases of the MLEs in some classes of regression models. The methodology has been applied to normal nonlinear models (Cook et al., 1986), generalized linear models (Cordeiro and McCullagh, 1991), multiplicative regression models (Cordeiro, 1993), ARMA models (Cordeiro and Klein, 1994), generalized linear models with dispersion covariates (Botter and Cordeiro, 1998) and overdispersed generalized linear models (Cordeiro and Botter, 2001). All these papers give closed-form expressions for the biases of the MLEs without involving cumulants of log-likelihood derivatives. The bias approximation may be used to produce a bias-reduced estimator by subtracting the bias approximation from the MLE.

Alternatively, an examination of the form of the bias may suggest a reparameterization of the model that results is less biased estimates.

In regular parametric problems, there is an alternative bias correction scheme proposed by Firth (1993) who shows how to remove the first-order term from the asymptotic bias of these estimates by a suitable modification of the score function. In exponential families with canonical parameterization his correction scheme consists in penalizing the likelihood by the Jeffreys invariant prior. This is a preventive approach to bias adjustment which has its merits, but the connections between our results and his work are not pursued in this paper since they could be developed in future research. We should also stress that it is possible to avoid cumbersome and tedious algebra on cumulant calculations by using Efron's bootstrap; see Efron and Tibshirani (1993). We use the analytical approach here since this leads to a nice formula. Moreover, the application of the analytical bias approximation seems to generally be the most feasible procedure to use and it continues to receive attention in the literature. A simple matrix version for bias correction is developed to provide a basis for improving inference on the parameters modeling the variance.

We denote by $\kappa^{r,s} = -\kappa^{rs}$ and $\kappa^{R,S} = -\kappa^{RS}$ the corresponding elements of the inverses K_{β}^{-1} and K_{γ}^{-1} of the information matrices for β and γ, respectively. These inverse matrices are of order $O(n^{-1})$. We obtain the following cumulants needed for the second-order biases of $\hat{\beta}$ and $\hat{\gamma}$.

$$\kappa_{rst} = \kappa_{rs}^{(t)} = \kappa_{RSt} = 0, \;\; \kappa_{Rst} = \sum_i \left(\frac{h_i'}{h_i^2}\right) x_{is} x_{it} z_{iR}$$

$$\kappa_{RST} = \frac{1}{2}\sum_i \frac{(4h_i'^3 - 3h_i h_i' h_i'')}{h_i^3} z_{iR} z_{iS} z_{iT} \text{ and } \kappa_{RS}^{(T)} = -\sum_i \left(\frac{h_i'^2}{2h_i^2}\right)' z_{iR} z_{iS} z_{iT}$$

Let $B(\hat{\beta}_r)$ and $B(\hat{\gamma}_R)$ be the second-order biases of the MLEs $\hat{\beta}_r$ and $\hat{\gamma}_R$, respectively, for $r = 1, \ldots, p$ and $R = 1, \ldots, q$. From the general formulas for the n^{-1} biases of the MLEs in regular problems given by Cox and Snell (1968) and Cordeiro and McCullagh (1991), $B(\hat{\beta}_a)$ follows due to the orthogonality between β and γ as

$$B(\hat{\beta}_a) = \kappa^{ar}\kappa^{st}\left(\kappa_{rs}^{(t)} - \frac{1}{2}\kappa_{rst}\right) - \frac{1}{2}\kappa^{ar}\kappa^{ST}\kappa_{rST}$$

where Einstein summation convention is adopted with the indices varying over the corresponding parameters. All third-order cumulants in the above expression vanish and then $B(\hat{\beta}_a) = 0$, i.e., the estimator $\hat{\beta}$ has no bias to order $O(n^{-1})$. Most obviously, this is to be expected for the linear homoskedastic model. However, it is not evident that this happens for any linear heteroskedastic model (1) since the MLE $\hat{\beta}$ is obtained from the nonlinear equation (5). This surprising result could

be further investigated by computing the bias of order n^{-2} of $\hat{\beta}$ following the same lines described by Cordeiro and Barroso (2007) in the context of the generalized linear models to check whether it also vanishes. However, this will be done in a future research because of the complicated algebra involving joint cumulants of log-likelihood derivatives of fourth and fifth orders. See, formulas $(6)-(8)$ of their paper, where there are 52 terms to be computed involving these cumulants.

Using the orthogonality between β and γ, the n^{-1} bias of $\hat{\gamma}_A$ is given by

$$B(\hat{\gamma}_A) = -\frac{1}{2}\kappa^{AR}\kappa^{st}\kappa_{Rst} + \kappa^{AR}\kappa^{ST}\left(\kappa_{RS}^{(T)} - \frac{1}{2}\kappa_{RST}\right)$$

where, again, we use Einstein summation convention over the parameters. We have

$$\kappa_{RS}^{(T)} - \frac{1}{2}\kappa_{RST} = -\frac{1}{4}\sum_i \frac{h_i' h_i''}{h_i^2} z_{iR} z_{iS} z_{iT}$$

Hence, we can write the n^{-1} bias of $\hat{\gamma}_A$ as

$$\begin{aligned} B(\hat{\gamma}_A) \quad & -\frac{1}{2}\sum_{i,R} \kappa^{AR} z_{iR} \frac{h_i'}{h_i^2}\left(\sum_{s,t} \kappa^{st} x_{is} x_{it}\right) \\ & -\frac{1}{4}\sum_{i,R} \kappa^{AR} z_{iR} \frac{h_i' h_i''}{h_i^2}\left(\sum_{S,T} \kappa^{ST} z_{iS} z_{iT}\right) \end{aligned}$$

where i runs over the observations and the other indices run over the parameters.

We now define the $n \times n$ matrices $A = ZK_\gamma^{-1}Z^T = Z(Z^TVZ)^{-1}Z^T$ and $B = XK_\beta^{-1}X^T = X(X^T\Lambda^{-1}X)^{-1}X^T$. Let $A_d = diag\{a_{ii}\}$ and $B_d = diag\{b_{ii}\}$ be the diagonal matrices with the corresponding elements of A and B, respectively. Note that A and B are the asymptotic covariance matrices of $Z\hat{\gamma}$ and $X\hat{\beta}$, respectively. We then have:

$$B(\hat{\gamma}_A) = \frac{1}{2}\sum_{i,R} \kappa^{AR} \frac{h_i'}{h_i^2} z_{iR} b_{ii} + \frac{1}{4}\sum_{i,R} \kappa^{AR} \frac{h_i' h_i''}{h_i^2} z_{iR} a_{ii}$$

Finally, after some algebra, we obtain the n^{-1} bias of $\hat{\gamma}$

$$B(\hat{\gamma}) = (Z^TVZ)^{-1}Z^TV\xi \tag{8}$$

where $\xi = -V^{-1}(B_dH_1 + A_dH_2)\,1$ is an $n \times 1$ vector and $H_1 = diag\{\frac{h_i'}{2h_i^2}\}$ and $H_2 = diag\{\frac{h_i' h_i''}{4h_i^2}\}$ are diagonal matrices of order n. The bias vector $B(\hat{\gamma})$ depends only on the matrix Z and the diagonal matrices V, H_1, H_2, A_d and B_d. These matrices are functions of the heteroskedastic function h and of its first two derivatives. The

bias vector $B(\hat{\gamma})$ is simply the set of coefficients from the weighted least squares regression of ξ in the columns of Z with V as a weighting matrix. Equation (8) represents the main result of this paper.

For homoskedastic regression models, $q = 1$, $Z = 1$, $\gamma = \sigma^2$ is a scalar parameter, h is the identity function which implies $H_2 = 0$, $\Lambda = \sigma^2 I$, $V = H_1 = (2\sigma^2)^{-1} I$, $A = \frac{2\sigma^4}{n} 11^T$, where I is the identity matrix, $B = \sigma^2 X(X^T X)^{-1} X^T$ and it is easy to show from (8) that $B(\hat{\gamma}) = -p\sigma^2/n$ as expected.

We now give expressions of V, H_1 and H_2 for some heteroskedastic functions. For the heteroskedastic power function $h(\tau_i) = \tau_i^{\theta}$ with $\tau_i = z_i^T \gamma$, we obtain $V = \frac{\theta^2}{2} diag\{\tau_i^{-2}\}$, $H_1 = \frac{\theta}{2} diag\{\tau_i^{-\theta-1}\}$, and $H_2 = \frac{\theta^2(\theta-1)}{4} diag\{\tau_i^{-3}\}$ and, therefore, the weights in the regression (8) increase with θ. When the variance is the square of a linear function of exogenous variables, we have $V = 2\Lambda^{-1}$ and $H_1 = H_2 = \Lambda^{-3/2}$. When the variance is a linear function of exogenous variables, $V = H_1 = \frac{1}{2}\Lambda^{-2}$ and H_2 vanishes. For the multiplicative heteroskedastic model for which $h_i = \exp(z_i^T \gamma)$, we have $V = \frac{1}{2} I$, $H_1 = I$ and $H_2 = \frac{1}{2} I$. In this case, equation (8) reduces to Cordeiro's (1993) expression (7).

The biases of the MLEs of the variances follow by expanding $\hat{\sigma}_i^2 = h(\hat{\tau}_i)$ in Taylor series and using $B(\hat{\tau}_i) = z_i^T B(\hat{\gamma})$ and $Var(\hat{\tau}_i) = a_{ii}$

$$B(\hat{\sigma}_i^2) = h'(\tau_i)\ z_i^T B(\hat{\gamma}) + \frac{1}{2}\ h''(\tau_i)\ a_{ii}$$

We can define the bias-corrected estimator (BCE)

$$\tilde{\gamma} = \hat{\gamma} - \widehat{B}(\hat{\gamma}) \tag{9}$$

where $\widehat{B}(\hat{\gamma})$ is the value of (8) in the MLE $(\hat{\beta}^T, \hat{\gamma}^T)^T$. The BCE defined in (8) removes the term of order n^{-1} from the asymptotic bias of $\hat{\gamma}$ and is, quite generally, second-order efficient.

Finally, we do not know of any formula for the $\mathcal{O}\left(n^{-2}\right)$ covariance of the BCE for a multiparameter model, although its $\mathcal{O}\left(n^{-1}\right)$ covariance is identical to the asymptotic covariance of the MLE given by the inverse of the information matrix. For one-parameter models, Ferrari et al. (1996) compared the BCEs and the MLEs in terms of their mean squared errors showing that the bias correction could increase or decrease the variance, but no results are available in multiparameter models.

4. Simulation Results

Some Monte Carlo simulations have been developed to compare the performance of the MLE $\hat{\gamma}$ and the bias-corrected counterpart $\tilde{\gamma}$ in linear models with heteroskedastic function $h(z_i^T \gamma) = (z_i^T \gamma)^{\theta}$ for $\theta = 1$ and 2 and $h(z_i^T \gamma) = \exp(z_i^T \gamma)$. The systematic component for the models is defined by

$$\mu = \beta_1 + \beta_2 x_2 + \beta_3 x_3 \text{ and } \tau = \gamma_1 + \gamma_2 z_2 + \gamma_3 z_3$$

and the true values of the parameters for the simulations were taken as $\beta_1 = 1$, $\beta_2 = -1$, $\beta_3 = 2$, $\gamma_1 = 2$, $\gamma_2 = 1$ and $\gamma_3 = -1$. All independent explanatory variables in X and Z were obtained as random draws from a uniform distribution $U(0,1)$ and their values were held fixed throughout the simulations with equal sample sizes. The number of observations was set at $n = 20, 30$ and 40 and the simulations run-size was $10,000$ in each case. The calculations were performed using the statistical package SAS.

In each of the $10,000$ replications, we fitted the model and computed the uncorrected estimates $\hat{\beta}$ and $\hat{\gamma}$ through the procedures NLP and NLMIXED of SAS. Then, we obtained $\widehat{B}(\hat{\gamma})$ from (8) with all quantities evaluated at $(\hat{\beta}^T, \hat{\gamma}^T)^T$ and $\tilde{\gamma}$ from (9). For each n, we computed the sample means and standard errors of the uncorrected $\hat{\gamma}$ and corrected estimates $\tilde{\gamma}$ based on their values from the $10,000$ trials. Tables 1–3 give the sample means of both estimates with their respective standard errors in parentheses. We can see that the bias-corrected estimators $\tilde{\gamma}$ are closer to the true parameter values than the uncorrected estimators $\hat{\gamma}$, thus correctly signalizing the direction of the bias corrections. They also show that the biases of the estimators $\hat{\gamma}$ are negative for the three heteroskedastic functions considered. The biases are greater in magnitude for the multiplicative model and smaller for the square heteroskedastic function.

Some simulations with other models, not given here, suggest that the heteroskedastic function $h(z_i^T\gamma) = (z_i^T\gamma)^{1/2}$ reduces the magnitude of the biases of the corrected estimators $\hat{\gamma}$. The bias correction has less impact as n increases and the estimators $\tilde{\gamma}'s$ tend to have slightly smaller standard errors than $\hat{\gamma}'s$ for small-sized samples. It is common belief that bias correction entails some variance inflation. However, these simulation results show otherwise. In these cases, the bias correction can lead to substantial improvement in terms of bias and mean squared error. The simulations suggest that bias correction in linear heteroskedastic regression models can then be used to obtain improved estimators with more reliable finite sample behavior.

Table 1
Uncorrected and corrected estimators of γ parameters for the heteroskedastic function $\sigma_i^2 = (z_i^T\gamma)^2$

n	Estimate	$\gamma_1 = 2$	$\gamma_2 = 1$	$\gamma_3 = -1$
20	MLE	1.89	0.88	−1.10
		(2.71)	(1.83)	(1.79)
	BCE	1.95	0.96	−1.06
		(2.40)	(1.80)	(1.71)
30	MLE	1.93	0.89	−1.05
		(2.29)	(1.72)	(1.73)
	BCE	1.96	0.95	−1.03
		(2.05)	(1.75)	(1.70)
40	MLE	1.98	0.95	−1.04
		(1.91)	(1.68)	(1.65)
	BCE	1.99	0.97	−1.01
		(1.88)	(1.63)	(1.62)

Table 2
Uncorrected and corrected estimators of γ parameters for the heteroskedastic function $\sigma_i^2 = z_i^T\gamma$

n	Estimate	$\gamma_1 = 2$	$\gamma_2 = 1$	$\gamma_3 = -1$
20	MLE	1.85	0.84	−1.14
		(2.68)	(1.87)	(1.80)
	BCE	1.92	0.95	−1.05
		(2.39)	(1.83)	(1.68)
30	MLE	1.89	0.88	−1.09
		(2.27)	(1.76)	(1.71)
	BCE	1.94	0.97	−1.04
		(2.08)	(1.72)	(1.69)
40	MLE	1.96	0.93	−1.06
		(1.90)	(1.65)	(1.67)
	BCE	1.98	0.95	−1.02
		(1.87)	(1.60)	(1.58)

Even though econometricians need to compute point estimates and their standard errors, the interest typically lies in performing hypothesis testing inference on the regression parameters. We now show a simulation study to evaluate interval estimation based on the corrected estimators in (9). We consider only the model with heteroskedasticity function $\sigma_i^2 = \exp(z_i^T\gamma)$ and $n = 30$, although the results are similar under the other two heteroskedastic models considered here and for other sample sizes.

Table 3
Uncorrected and corrected estimators of γ parameters for the heteroskedastic function $\sigma_i^2 = \exp(z_i^T\gamma)$

n	Estimate	$\gamma_1 = 2$	$\gamma_2 = 1$	$\gamma_3 = -1$
20	MLE	1.72	0.79	−1.26
		(2.40)	(1.99)	(1.84)
	BCE	1.89	0.93	−1.11
		(2.19)	(1.87)	(1.80)
30	MLE	1.83	0.84	−1.17
		(2.20)	(1.90)	(1.69)
	BCE	1.95	0.96	−1.07
		(2.06)	(1.92)	(1.70)
40	MLE	1.91	0.87	−1.09
		(1.96)	(1.88)	(1.56)
	BCE	1.98	0.94	−1.03
		(1.85)	(1.80)	(1.54)

First, in Table 4, we give the sample mean, variance, skewness and kurtosis of the 10,000 simulated values for both (uncorrected and bias-corrected) standardized estimates $(estimate - true\,value)/standard\,error$, namely USEs and CSEs, respectively, where the common $standard\,error's$ for both estimates are equal to the square roots of the diagonal elements of the inverse information matrix $\widehat{K}_\gamma^{-1} = (Z^T\widehat{V}Z)^{-1}$. Good agreement with the normal distribution happens when these figures are, on average, close to $0, 1, 0$ and 3, respectively. The four cumulants of the CSEs are generally closer to their limiting values than the USEs, the most significant being the mean and variance, as expected.

Second, the histograms of the 10,000 uncorrected and corrected values for both standardized estimates are given in Figure 1 together with the graph of the probability density function of a standard normal distribution. We can see that the empirical distributions of the CSEs tend to be closer to the standard normal distribution than the empirical distributions of the USEs. For example, in terms of shape, the empirical distribution of $\hat{\gamma}_1$ has a more acute "peak" around the mean than the distribution of $\tilde{\gamma}_1$, i.e. a higher probability than a normally distributed variable for values near the mean.

Table 4
Comparison of the uncorrected and corrected standardized estimators for the heteroskedastic function $\sigma_i^2 = \exp(z_i^T\gamma)$ and $n = 30$

	mean		variance		skewness		kurtosis	
i	USE_i	CSE_i	USE_i	CSE_i	USE_i	CSE_i	USE_i	CSE_i
1	-0.11	-0.05	0.81	0.94	0.11	0.08	3.98	3.22
2	0.44	0.02	0.66	0.83	0.22	0.17	3.24	3.13
3	-0.09	-0.01	0.79	1.06	0.27	0.20	3.29	3.10

Further, we constructed asymptotic 95% confidence intervals given by *estimate* ∓ 1.96 *standard error* and noted that the intervals from CSEs seem to be more precise than those from the MLEs. In fact, the observed frequencies that the true parameters $\gamma_1 = 2$, $\gamma_2 = 1$ and $\gamma_3 = -1$ are within the intervals based on the

USEs were 9647, 9416 and 9428, whereas those frequencies based on the CSEs were 9514, 9532 and 9522, respectively.

To verify numerically that the distribution of the corrected estimates is generally closer to a standard normal distribution than the distribution of the uncorrected estimates, we perform a one-sample Kolmogorov-Smirnov (K-S) test on both sets of $10,000$ USEs and CSEs. In Table 5, we give the values of the K-S statistic (which is a metric on the set of distribution functions), and the corresponding $p-$values by applying the K-S test for normality on each set of $10,000$ USEs and each set of $10,000$ CSEs. From Table 5, we conclude that the K-S statistics are smaller for the CSEs than for the corresponding USEs. Then, the empirical distributions of the corrected estimates are closer to the normal distribution than the empirical distributions of the uncorrected estimates.

The arguments above reveal the superiority of the BCEs $\tilde{\gamma}_i's$ over the MLEs $\hat{\gamma}_i$ with regard to the normal approximation. Further, we should not expect a great improvement in regard to the normal approximation since we are adjusting only the first moment, yet, here, we obtained some improvement. These results show that a simple adjustment of the first moment of the MLE to order n^{-1} may yield some improvement in the normal approximation for its distribution.

Table 5
K-S statistics for comparing the USEs and CSEs for the heteroskedastic function $\sigma_i^2 = \exp(z_i^T\gamma)$ and $n = 30$

i	K-S stat. for USE_i	p-value for USE_i	K-S stat. for CSE_i	p-value for CSE_i
1	0.2134	$< 2.2 \times 10^{-16}$	0.0215	0.00018
2	0.0314	1.3×10^{-10}	0.0148	0.02713
3	0.0731	$< 2.2 \times 10^{-16}$	0.0016	0.00778

Econometricians tend to avoid estimation by maximum likelihood when they use the linear regression model because maximum likelihood requires an additional assumption about the distribution of the data. For that reason, least squares estimation is typically used. We now show some simulations (with the heteroskedastic function $\sigma_i^2 = \exp(z_i^T\gamma)$ and $n = 30$) considering that the normality assumption does not hold. In order to do this, we use formula (9) for the BCE but generate the data from three other symmetric distributions, namely Student t_4 and logistic (types I and II) distributions, to investigate how the biases of $\hat{\gamma}$ and $\tilde{\gamma}$ were sensitive to misspecification of the distribution.

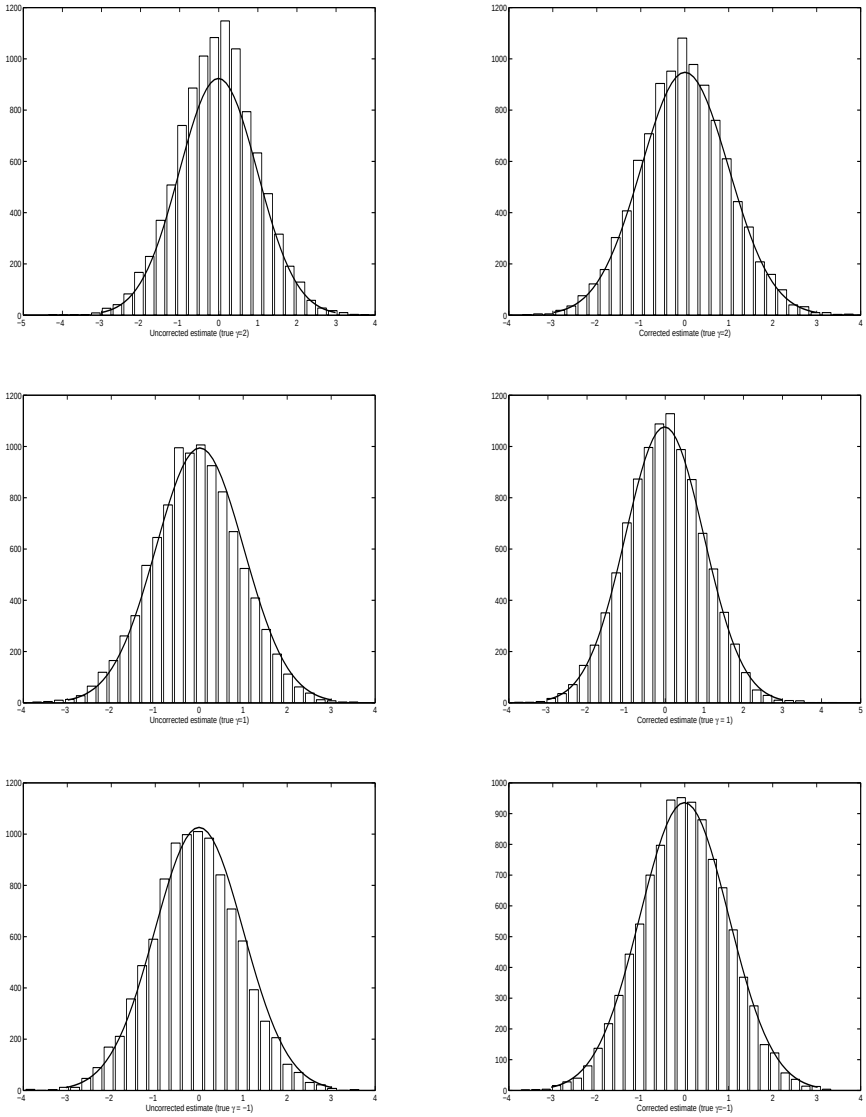

Figure 1 – Empirical distributions of the USEs and CSEs

The results are in Table 6. The figures in this table show that the BCEs are more reliable than the MLEs even under misspecification in the sense that they always displayed smaller biases. The same does not happen in relation to the mean squared errors. Although, these results may look unfamiliar, we note that the normal distribution is in fact the unnormalized saddlepoint approximation for all symmetric distributions and, therefore, asymptotic results under the normal model could in principle be applied approximately for other models provided the saddlepoint approximation holds.

Table 6
MLEs and BCEs under misspecification

Distribution	Estimate	$\gamma_1 = 2$	$\gamma_2 = 1$	$\gamma_3 = -1$
t_4	MLE	1.64	0.69	−1.38
		(2.31)	(1.60)	(2.41)
	BCE	1.70	0.81	−1.19
		(2.05)	(1.71)	(2.17)
logistic I	MLE	1.70	0.65	−1.29
		(2.20)	(1.90)	(1.69)
	BCE	1.83	0.87	−1.21
		(1.93)	(1.92)	(1.70)
logistic II	MLE	1.76	0.80	−1.23
		(2.08)	(1.95)	(1.78)
	BCE	1.81	0.93	−1.14
		(2.13)	(2.01)	(1.51)

5. Applications

In this section, we provide two applications to illustrate the bias-corrected estimators. First, we consider a data set in portfolio market (Simonoff and Tsai, 1994) that represents the monthly excess returns over the riskless rate for the Acme Cleveland Corporation (Y) and the return for the market (x) for the period January 1986—December 1990. Quite often, the response variable that represents the return of a security displays heteroskedasticity. In the original analysis of these data, they suggested a linear heteroskedastic model to predict Y_i in terms of a linear function of the return of the market x_i by considering two systematic components

$$\mu_i = \beta_1 + \beta_2 x_i \text{ and } \sigma_i^2 = \exp(\gamma_1 + \gamma_2 x_i)$$

for $i = 1, \ldots, 59$. We are interested in testing the null hypothesis of homoskedasticity $H : \gamma_2 = 0$. The uncorrected estimates (with their corresponding standard errors in parentheses) are: $\hat{\beta}_1 = 1.2534\,(0.5782)$, $\hat{\beta}_2 = -0.0057\,(0.0021)$, $\hat{\gamma}_1 = -4.4164\,(2.3781)$ and $\hat{\gamma}_2 = 8.0935\,(5.3127)$. The bias-corrected estimates are $\tilde{\gamma}_1 = -4.7694\,(2.3781)$ and $\tilde{\gamma}_2 = 10.8314\,(5.3127)$, yielding a corrected $t-$ratio of 2.0388 for testing $H : \gamma_2 = 0$ that is significant at 5%, whereas the usual t test is not.

Second, we now reanalyze the example discussed by (Ramanathan, 1993, Table 10.1) in a study in which seven variables were observed in 40 metropolitan areas. The response variable is the number (in thousands) of cable TV subscribers (Y) and the explanatory variables are the number (in thousands) of homes in the area (*home*), the *per capita* income for each television market with cable (*income*), the installation fee (fee), the monthly service charge (*charge*), the number of television signals carried by each cable system (*signals*) and the number of television signals received with good quality without cable (*wcable*).

Because Y corresponds to count data, Cysneiros and Paula (2005) used a square root transformation in order to try to obtain a common variance. Alternatively, we fitted a normal linear model with exponential heteroskedasticity and estimated the vectors β and γ using SAS. The systematic components of the model were taken as

$$\mu_i = \beta_1 + \beta_2\, home_i + \beta_3\, charge_i \text{ and } \sigma_i^2 = exp(\gamma_1 + \gamma_2\, charge_i)$$

since the coefficients of the other explanatory variables were not significant in wider fitted models. The excluded variables do not contain much information about the response variable in the analysis. The MLEs are: $\hat{\beta}_1 = 15.995085\,(10.033904)$, $\hat{\beta}_2 = 0.401087\,(0.026102)$, $\hat{\beta}_3 = -2.178738\,(1.566041)$, $\hat{\gamma}_1 = -3.353204\,(1.312739)$ and $\hat{\gamma}_2 = 1.061738\,(0.176258)$. The corrected estimates are $\tilde{\gamma}_1 = -2.249512$ (1.312739) and $\hat{\gamma}_2 = 1.234521\,(0.176258)$, changing the significance (at 5%) of γ_1 and improving the significance of γ_2.

References

Botter, D. A. & Cordeiro, G. M. (1998). Improved estimators for generalized linear models with dispersion covariates. *Journal of Statistical Computation and Simulation*, 62:91–104.

Cook, D. R., Tsai, C. L., & Wei, B. C. (1986). Bias in nonlinear regression. *Biometrika*, 73:615–623.

Cordeiro, G. M. (1993). Barlett corrections and bias correction for two heteroskedastic regression models. *Communications in Statistics – Theory and Methods*, 22:169–188.

Cordeiro, G. M. & Barroso, L. P. (2007). A third-order bias-corrected estimate in generalized linear models. *Test*, 16:76–89.

Cordeiro, G. M. & Botter, D. A. (2001). Second-order biases of maximum likelihood estimates in overdispersed generalized linear models. *Statistics and Probability Letters*, 55:269–280.

Cordeiro, G. M. & Klein, R. (1994). Bias correction in ARMA models. *Statistics and Probability Letters*, 19:169–176.

Cordeiro, G. M. & McCullagh, P. (1991). Bias correction in generalized linear models. *Journal Royal Statistical Society B*, 53:629–643.

Cox, D. R. & Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal Royal Statistical Society B*, 30:248–275.

Cysneiros, F. J. A. & Paula, G. A. (2005). Restricted methods in symmetrical linear regression models. *Computational Statistics & Data Analysis*, 49:689–708.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Ferrari, S. L. P., Botter, D. A., Cordeiro, G. M., & Cribari-Neto, F. (1996). Second- and third-order bias reduction for one-parameter family models. *Statistics and Probability Letters*, 30:339–345.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80:27–38.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1985). *The Theory and Practice of Econometrics*. John Wiley, New York.

Ramanathan, R. (1993). *Statistical Methods in Econometrics*. John Wiley, New York.

Simonoff, J. S. & Tsai, C.-H. (1994). Use of modified profile likelihood for improved tests of constancy of variance in regression. *Applied Statistics*, 43:357–370.