UNIVERSITY OF ALBERTA

Iterative Weighted Least Squares Estimators in Heteroscedastic Linear Models with Asymmetric Error Distribution

BY

QIUHUA(JILL) JIANG ©

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

EDMONTON, ALBERTA
FALL 1994

ISBN 0-315-95053-6

UNIVERSITY OF ALBERTA
RELEASE FORM

NAME OF AUTHOR: Qiuhua(Jill) Jiang

TITLE OF THESIS: **Iterative Weighted Least Squares Estimators in Heteroscedastic Linear Models with Asymmetric Error Distribution**

DEGREE FOR WHICH THESIS WAS PRESENTED: Master of Science

YEAR THE DEGREE GRANTED: 1994

Qiuhua Jiang

(Signed)

Permanent Address:

Department of Statistics and Applied
Probability
University of Alberta
Edmonton, Alberta T6G 2G1

Date: Aug. 22. 1994

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled

Iterative Weighted Least Squares Estimators in Heteroscedastic Linear Models with Asymmetric Error Distribution

submitted by **Qiuhua(Jill) Jiang** in partial fulfillment of the requirements for the degree of **Master of Science**.

Peter Hooper

Peter M. Hooper (Supervisor)

Edit Gombay

Edit Gombay

J D Jobson

J. David Jobson

Date: Aug 22, 94

To:

Jimmy

Gina

Abstract

In this thesis we consider a heteroscedastic linear model in which observations are arranged in groups with constant variance. We study the bias of an iterative weighted least squares estimator when the error distribution is asymmetric. We use asymptotic approximations and Monte Carlo simulations to investigate how the bias is related to the degree of skewness, the degree of heteroscedasticity, the group size, and the models for the means and variances.

Acknowledgements

I wish to thank my supervisor, Dr. Hooper, with my deepest appreciation, for his invaluable guidance and friendly support.

I also wish to thank Dr. Jobson and Dr. Gombay for their careful consideration and helpful suggestions.

I thank the Department of Statistics and Applied Probability for providing a congenial research environment. I am also grateful to Dr. Hooper and University of Alberta for their financial support.

Contents

Chapter 1

Introduction

We consider a heteroscedastic linear model with responses arranged in k groups and variances constant within each group:

$$y_{ij} = \mathbf{x}_{ij}^T\beta + e_{ij} \quad j = 1, ..., n_i, \quad i = 1, ..., k, \tag{1.1}$$

where $\mathbf{x}_{ij}$ are known $p \times 1$ vectors, β is an unknown $p \times 1$ vector of parameters, and the errors e_{ij} have mean 0 and unknown variance σ_i^2. We assume groups are determined by replicates, so that $\mathbf{x}_{ij} = \mathbf{x}_i$. Our main interest is to estimate the unknown regression parameter β.

One could ignore the heteroscedasticity and use ordinary least squares (OLS) estimators or M- estimators to estimate β (Huber 1981), but such methods are not efficient. To improve efficiency, many authors have recommended weighted least squares (WLS) estimators. We introduce vector notation for (1.1):

$$\mathbf{y}_i = \mathbf{1}_{n_i}\mathbf{x}_i^T\beta + \mathbf{e}_i. \tag{1.2}$$

Setting $\mathbf{X}_i = \mathbf{1}_{n_i}\mathbf{x}_i^T$, we have

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{e}_i. \tag{1.3}$$

An expression for a WLS estimator is

$$\hat{\beta} = (\sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{X}_i)^{-1} \sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{y}_i. \tag{1.4}$$

When the errors are normally distributed and the variances σ_i^2 are known, then the optimal weigh s are $\hat{w}_i = 1/\sigma_i^2$. When the σ_i^2 are unknown, we can obtain estimates $\hat{\sigma}_i^2$ of the variances and apply WLS using weights $\hat{w}_i = 1/\hat{\sigma}_i^2$.

There are two general approaches for estimating the variances. One approach models the variance as a function of the mean and/or the covariates. For example, the variance may be modeled as

$$\log(\sigma_i) = \theta_0 + \theta_1 \log(\mathbf{x}_i^T \beta).$$

This is called the power of the mean model (Carroll and Ruppert, 1988). Box and Hill (1974), Jobson and Fuller (1980) and Carroll and Ruppert (1982) modeled the variance as

$$\sigma_i^2 = H(\mathbf{x}_i, \theta), \text{ or } H(\mathbf{x}_i^T \beta, \theta), \; H \text{ known}.$$

Initial estimates $\hat{\theta}$, $\hat{\beta}$ were used to obtain variance estimates $\hat{\sigma_i^2} = H(\mathbf{x}_i, \hat{\theta})$ or $H(\mathbf{x}_i^T \hat{\beta}$, $\hat{\theta})$, which were then used to obtain a WLS estimate of β. This process can be iterated, with new estimates $\hat{\theta}$, $\hat{\beta}$ at each step. Jobson and Fuller (1980), and Carroll and Ruppert (1982) showed that, as long as the preliminary estimators for the parameters of the variance function are consistent, all estimators of β obtained this way are asymptotically equivalent to WLS estimators with optimal weights. Carroll (1982) modeled the variance as

$$\sigma_i^2 = H(\mathbf{x}_i), \text{ or } H(\mathbf{x}_i^T \beta), \; H \text{ unknown but smooth},$$

and used nonparametric regression to estimate H. He obtained asymptotic results similar to those for the parametric variance model. We refer to this general approach to variance estimation as model-based, where "model" refer to a model for the variances.

A second general approach is to use model-free variance estimates. The simplest choice for $\hat{\sigma}_i^2$ is the sample variance

$$s_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \quad n_i \geq 2. \tag{1.5}$$

Bement and Williams (1969) assumed normally distributed errors, used sample variances (1.5), and constructed approximations, as $n_i \to \infty$, for the exact covariance matrix of the resulting weighted least squares estimator. They did not discuss asymptotic distributions as $k \to \infty$ with fixed n_i. Carroll and Cline (1988) showed that weights based on sample variances yield highly inefficient estimates for β when the number of replicates is small. For example, when $n_i = 2$ the estimates are inconsistent. Better weights are obtained from variance estimators of the form

$$v_i(\tilde{\beta}) = n_i^{-1} \sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_i^T \tilde{\beta})^2, \tag{1.6}$$

where $\tilde{\beta}$ is an estimate of β. Fuller and Rao (1978) used the OLS estimator as $\tilde{\beta}$ in (1.6). Assuming normally distributed errors and regularity conditions, they showed that the estimate is consistent and asymptotically normal. Carroll and Cline (1988) and Shao (1989) showed that these conclusions hold under weaker assumptions. Shao (1992) considered a modification of $\hat{\beta}$ that replaces the $\hat{v}_i(\tilde{\beta})$ with empirical Bayes estimates of the variances.

These two-step estimators perform well when the variances are not too dispersed, but are less efficient under more severe heteroscedasticity. This is due to the reduced efficiency of the initial estimator $\tilde{\beta}$ (usually the OLS estimator) resulting in poorer variance estimators. Under severe heteroscedasticity, efficiency can be improved by iterating the procedure: Replace $\tilde{\beta}$ by $\hat{\beta}$ in (1.5) to obtain new weights, then compute a new WLS estimator $\hat{\beta}$. Continuing iteration until $\hat{\beta}$ converges yields the normal theory maximum likelihood estimator (MLE). Cochran (1937) and Neyman and Scott (1948) studied the MLE for the common mean problem ($p = 1, \mathbf{x}_i = 1$) and derived the limiting distribution as $k \to \infty$ for fixed $n = n_i \geq 3$. Sometimes it is not desirable to continue iteration until $\hat{\beta}$ converges. Carroll and Cline (1988) showed that under mild or moderate heteroscedasticity, Fuller and Rao's two step estimator is better than the MLE; under severe heteroscedasticity the MLE is better. As a compromise, one might consider an estimator based on , say 3 or 4 iterations. Shao and Chen (1993) derived a consistent estimator for the optimal number of iterations.

The weight $1/\hat{v}_i$ blows up as $\hat{v}_i$ approaches 0. Hooper (1993) suggested that the poor performance of the MLE was caused by a feedback effect resulting in too much weight being concentrated on a small number of groups. He studied fully iterative WLS estimators with bounded weights of the form

$$\hat{w}_i = \frac{n_i + \hat{\gamma}}{n_i\hat{v}_i + \hat{\gamma}\hat{\tau}_i}, \tag{1.7}$$

where $\hat{\tau}_i$ is a model-based variance estimate and γ is a parameter determining the amount of heteroscedasticity not accounted for by the $\hat{\tau}_i$. The inverse weight $1/\hat{w}_i$ is a weighted average of the model-free variance estimate $\hat{v}_i$ and the model-based variance estimate $\hat{\tau}_i$. The iterative weighted least squares (IWLS) estimator with weights (1.7) is asym ptotically optimal under normally distributed errors and a Bayesian model for the variances:

$$\frac{1}{\sigma_i^2} \sim \frac{1}{\tau_i\gamma}\chi^2_\gamma. \tag{1.8}$$

Hooper applied the method of moments to $\log(v_i)$ and derived estimators for γ and τ_i under a log-linear model:

$$\log \tau_i = \mathbf{u}_i^T\eta, \tag{1.9}$$

where η is an unknown $r \times 1$ parameter vector and $\mathbf{u}_i$ is an $r \times 1$ parameter vector determined by $\mathbf{x_i}$ and/or $\mathbf{x_i^T}\beta$. The simplest example of (1.8) is the constant-variance model $\log(\tau_i) = \eta$. A second example is the power of the mean model $\log(\tau_i) = \eta_1 + \eta_2\log(\mathbf{x}_i^T\beta)$. The parameters γ and τ_i in (1.7) can be estimated by the following formulas:

$$\sigma^2_{LG}(\hat{\gamma}) = \min[\max\{s_z^2 - (k-r)^{-1}\sum(1-h_{ii})\sigma^2_{LG}(n_i), \sigma^2_{LG}(\gamma_{ub})\}, \sigma^2_{LG}(\gamma_{lb})] \tag{1.10}$$

$$\hat{\eta} = (U^T U)^{-1} U^T (Z + \mu_{LG}(\hat{\gamma}) 1_k), \quad (1.11)$$

$$\hat{\tau}_i = \exp(u_i^T \hat{\eta}), \quad (1.12)$$

where $0 < \gamma_{lb} < \gamma_{ub} < \infty$ are a priori lower and upper bounds on γ, $Z = (z_1, \cdots, z_k)^T$, $z_i = \log(v_i) - \mu_{LG}(n_i)$, $U = (u_1, \cdots, u_k)^T$, $H = (h_{ij}) = U(U^T U)^{-1} U^T$, $s_z^2 = (k - r)^{-1} Z^T (I_k - H) Z$, $\mu_{LG}(a) = E\{\log g\}$, $\sigma_{LG}^2(a) = \text{var}\{\log g\}$, and $g \sim \frac{1}{a}\chi_a^2$. Hooper (1993) found in simulation studies that the estimator performs well in a wide variety of situations, ranging from homoscedasticity to severe heteroscedasticity. The estimator is consistent provided the error distribution is symmetric.

Sometimes there is no compelling reason to assume symmetry. Even when symmetry seems a plausible assumption, recognition of the approximate nature of statistical models leaves us with only an optimistic hope that any departures from this assumption are small enough to be ignored. It is important to consider the effect of asymmetry on the estimates.

One possible remedy for an asymmetric error distrbution is a transformation. Box and Cox (1964) proposed choosing a transformation from the power transformation family by the method of maximum likelihood . They applied the transformation to the response in an attempt to achieve a simple linear model, homoscedastic errors, and normally distributed errors. If we have a physical model, say $\mathbf{y}_i = f(\mathbf{x}_i, \beta)$, and we transform only the response, then we destroy the original relationship between response and regression function. Carroll and Ruppert (1984) suggested transformation on both sides (the response and the regression function simultaneously) to reduce skewness and heteroscedasticity. They showed that the MLE for β is consistent and asymptotically normal when $\sigma_i \to 0$. Ruppert and Aldershof (1989) used an M-estimator, rather than the MLE, to obtain a consistent estimator without requiring that $\sigma_i \to 0$. Carroll (1979) suggested that, when the error is asymmetric and an appropriate transformation is hard to find, then a robust M-estimator provides a reasonable solution. He introduced consistent M-estimators for the common mean model and the simple linear regression model.

Usually assumptions about the error distribution, such as zero mean, normality, or symmetry, determine the methods required for consistent estimation of parameters. That is, in order to get a consistent estimate, we have to use different methods for different error distributions. Welsh, Carroll and Ruppert (1994) addressed this problem. They considered two extended models, one assuming additive errors and the other non-additive errors, which include most of the heteroscedastic regression models in current use. They developed methods of estimation appropriate for each type of model and investigated issues of consistency and adaptability (does it matter asymptotically whether certain nuisance parameters are known or estimated ?).

In this thesis we consider a heteroscedastic linear model (1.3), where each group has the same number of replicates ($n_i = n$, for all i). We study the performance of the IWLS estimator with weight function (1.7). We are primarily concerned with the effect of the group size n and the skewness of the error distribution on the bias of the IWLS estimator. In chapter 2, we derive approximations for the asymptotic bias and show that, when τ_i is a constant, the slope parameters are estimated consistently. In general, the bias can be approximated by a product of two terms: a scalar depending only on n and the distribution of the standardized errors $e_{ij}/\sqrt{\tau_i}$, and a vector depending only on the τ_i and $\mathbf{x}_i$. We calculate values for both terms under specific model assumptions. In chapter 3, we report the results of a Monte Carlo study of the bias in finite samples.

Chapter 2

Asymptotic Bias

In this chapter, we study the asymptotic bias for the IWLS estimator $\hat{\beta}$ defined by weights (1.7). Under regularity conditions, $\hat{\beta}$ converges in probability as $k \to \infty$. We will denote the limiting vector by $\beta + \delta$ and will refer to δ as the asymptotic bias of $\hat{\beta}$. In Section 2.1 we derive an equation defining δ. In Section 2.2 we use a Taylor approximation to express the asymptotic bias as a product of a scalar term ξ_0 and a vector term $\mathbf{a}$, and show the asymptotic bias for the slope parameters is zero when τ_i is a constant. In Section 2.3 we use Monte Carlo simulation to study the scalar ξ_0. In Section 2.4 we give an example to show how τ_i affects the vector $\mathbf{a}$.

2.1 An equation for the Asymptotic Bias

In deriving an equation for the asymptotic bias, we follow Hooper (1993) in assuming that $\{(\mathbf{y}_i, \mathbf{X}_i, \mathbf{e}_i), i = 1, 2, \cdots\}$ is a sequence of independent and identically distributed random vectors. This is a convenient device for describing the limiting behavior of $\mathbf{X}_i$. Our results can be applied to applications where the $\mathbf{X}_i$ are non-random. Our assumptions primarily involve conditional distributions given $\mathbf{X}_i$. In the derivation, we assume that certain estimators converge in probability. We do not specify the regularity conditions needed for this convergence.

The IWLS estimator is

$$\begin{aligned}\hat{\beta} &= \left(\sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{y}_i \\ &= \beta + \left(\sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{e}_i. \qquad (2.1)\end{aligned}$$

Put

$$\begin{aligned}\hat{\delta} &= \hat{\beta} - \beta \qquad (2.2) \\ &= \left(\frac{1}{k}\sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \left(\frac{1}{k}\sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{e}_i\right). \qquad (2.3)\end{aligned}$$

The estimated weight function is

$$\hat{w}_i = \frac{n + \hat{\gamma}}{n\hat{v}_i + \hat{\gamma}\hat{\tau}_i} \qquad (2.4)$$

with

$$\hat{v}_i = \frac{1}{n}\sum_{j=1}^{n}\left(y_{ij} - \mathbf{x}_i^T \hat{\beta}\right)^2 = \frac{1}{n}\sum_{j=1}^{n}\left(e_{ij} - \mathbf{x}_i^T \hat{\delta}\right)^2. \qquad (2.5)$$

Write $\hat{\tau}_i = \tau(\mathbf{x}_i, \hat{\beta}, \hat{\eta})$. Under regularity conditions, $\hat{\eta}$ and $\hat{\gamma}$ converge in probability, say $\hat{\eta} \rightarrow_p \eta$ and $\hat{\gamma} \rightarrow_p \gamma$. As stated previously, we assume $\hat{\beta} \rightarrow_p \beta + \delta$ and hence $\hat{\delta} \rightarrow_p \delta$. Put

$$\begin{aligned}\tilde{\tau}_i &= \tau\left(\mathbf{x}_i, \beta + \delta, \eta\right), \qquad (2.6) \\ \tilde{v}_i &= \frac{1}{n}\sum\left(e_{ij} - \mathbf{x}_i^T \delta\right)^2, \qquad (2.7) \\ \tilde{w}_i &= \frac{n + \gamma}{n\tilde{v}_i + \gamma\tilde{\tau}_i}, \qquad (2.8) \\ \tilde{\delta} &= \left(\frac{1}{k}\sum_{i=1}^{k} \tilde{w}_i \mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \left(\frac{1}{k}\sum_{i=1}^{k} \tilde{w}_i \mathbf{X}_i^T \mathbf{e}_i\right). \qquad (2.9)\end{aligned}$$

Under regularity conditions, $\hat{\delta} - \tilde{\delta} \rightarrow_p 0$. Since $(\tilde{w}_i, \mathbf{X}_i, \mathbf{e}_i)$ is a sequence of independent and identically distributed random vectors, we have

$$\frac{1}{k}\sum_{i=1}^{k} \tilde{w}_i \mathbf{X}_i^T \mathbf{X}_i \rightarrow_p E\left\{\tilde{w}_1 \mathbf{X}_1^T \mathbf{X}_1\right\}, \qquad (2.10)$$

$$\frac{1}{k}\sum_{i=1}^{k} \tilde{w}_i \mathbf{X}_i^T \mathbf{e}_i \rightarrow_p E\left\{\tilde{w}_1 \mathbf{X}_1^T \mathbf{e}_1\right\}. \qquad (2.11)$$

Combining the above, we have

$$\delta = \left(E\left\{\tilde{w}_1\mathbf{X}_1^T\mathbf{X}_1\right\}\right)^{-1} E\left\{\tilde{w}_1\mathbf{X}_1^T\mathbf{e}_1\right\}.$$

Thus δ is a solution of the equation

$$E\left\{\tilde{w}_1\mathbf{X}_1^T(\mathbf{e}_1 - \mathbf{X}_1\delta)\right\} = 0. \tag{2.12}$$

Substituting $\mathbf{X}_1 = \mathbf{1}_n\mathbf{x}_1^T$ and (2.7) yields

$$(n+\gamma)E\left\{\frac{\sum_{j=1}^n(e_{1j} - \mathbf{x}_1^T\delta)}{\gamma\tilde{\tau}_1 + \sum_{j=1}^n(e_{1j} - \mathbf{x}_1^T\delta)^2}\mathbf{x}_1\right\} = 0. \tag{2.13}$$

Put

$$\mathbf{t} = \tilde{\tau}_1^{-1/2}\mathbf{x}_1, \tag{2.14}$$

$$h_j = \tilde{\tau}_1^{-1/2}e_{1j}. \tag{2.15}$$

The asymptotic bias δ is a solution of the equation

$$E\left\{\frac{\sum_{j=1}^n(h_j - \mathbf{t}^T\delta)}{\gamma + \sum_{j=1}^n(h_j - \mathbf{t}^T\delta)^2}\mathbf{t}\right\} = 0. \tag{2.16}$$

2.2 Approximations

In this section we introduce several approximations for the asymptotic bias. A key assumption throughout is that the distribution of the standardized errors is independent of the explanatory variables; that is , setting $\tau_i = \tau(\mathbf{x}_i, \beta, \eta)$, we assume that

$$\tau_i^{-1/2}\mathbf{e}_i \text{ and } \mathbf{x}_i \text{ are independent.} \tag{2.17}$$

An interpretation of this assumption is that the relationship between the error distribution and the explanatory variables is determined completely by the model $\tau(\mathbf{x}_i, \beta, \eta)$ for the variance.

We expect that the asymptotic bias will have relatively little effect on the model-based variance estimates and so approximate $\tilde{\tau}_i = \tau(\mathbf{x}_i, \beta + \delta, \eta)$ by $\tau_i = \tau(\mathbf{x}_i, \beta, \eta)$ in our derivation. We therefore consider a solution δ of (2.16) with

$$h_j = \tau_1^{-1/2} e_{1j} \text{ and } \mathbf{t} = \tau_1^{-1/2}\mathbf{x}_1. \tag{2.18}$$

This approximation is unnecessary under the constant variance model where, by definition, we have $\tilde{\tau}_i = \tau_i = \tau_1$ constant. Assumptions (2.17) and (2.18) imply that $(h_1, \cdots, h_n)$ and $\mathbf{t}$ are independent. Note that we do not assume that $h_1, \cdots, h_n$ are independent.

Suppose (2.18) holds and there exists a constant vector $\mathbf{a} \epsilon R^p$ such that $\mathbf{a}^T\mathbf{t} = 1$ with probability one. We can then take $\delta = \xi\mathbf{a}$ with $\xi \epsilon R^1$ and (2.16) factors into a product of two expectations:

$$E\left\{\frac{\sum_j (h_j - \xi)}{\gamma + \sum_j (h_j - \xi)^2}\right\} E\{\mathbf{t}\} = 0.$$

We assume that $E\{\mathbf{t}\} \neq 0$, which will always be the case when the model for the means includes a constant term. The scalar ξ can then be obtained by solving

$$E\left\{\frac{\sum_j (h_j - \xi)}{\gamma + \sum_j (h_j - \xi)^2}\right\} = 0. \tag{2.19}$$

The solution to (2.19) can be approximated by using a first order Taylor approximation. Put

$$f(\xi) = E\left\{\frac{\sum_j (h_j - \xi)}{\gamma + \sum_j (h_j - \xi)^2}\right\}. \tag{2.20}$$

For ξ close to zero, we have

$$f(\xi) \approx f(0) + f'(0)\xi,$$

and the solution to (2.19) is approximated by

$$\begin{aligned}\xi_0 &= \frac{-f(0)}{f'(0)} \\ &= \frac{E\left\{\frac{\sum_j h_j}{\gamma + \sum_j h_j^2}\right\}}{E\left\{\frac{n}{\gamma + \sum_j h_j^2} - 2\left(\frac{\sum_j h_j}{\gamma + \sum_j h_j^2}\right)^2\right\}}.\end{aligned} \tag{2.21}$$

Thus δ is approximated by $\xi_0\mathbf{a}$.

The approximation above can be applied under the constant variance model: $\tau_i = \tau_1$ constant for all i. We assume that the model for the means includes a constant term. More precisely, suppose $x_{i_1} = 1$ for all i; i.e., β_1 is the intercept parameter and $\beta_2, \cdots, \beta_p$ are slope parameters. We may then take $\mathbf{a} = (\sqrt{\tau_1}, 0, \cdots, 0)$. This confirms a result in Hooper (1993) that estimates of slope parameters are consistent under the constant variance model. A corresponding result for M-estimators ($n = 1$) is described by Carroll and Welsh (1988). We note that this is an exact result, not a approximation, since $\tau_i = \tilde{\tau}_i$ in this situation. The asymptotic bias for the intercept can be approximated by $\xi_0\sqrt{\tau_1}$.

When the variance model is not constant, so τ_i varies from group to group, it is not possible in general to find a constant vector $\mathbf{a}\epsilon R^p$ so that $\mathbf{a}^T\mathbf{t} = 1$. It is, however, still possible to use a linear expansion to approximate the asymptotic bias as a product of two expectations. Define f: $R^p \rightarrow R^p$ by

$$f(\delta) = E\left\{\frac{\sum_j(h_j - \mathbf{t}^T\delta)}{\gamma + \sum_j(h_j - \mathbf{t}^T\delta)^2}\mathbf{t}\right\}. \tag{2.22}$$

For δ close to zero, we have

$$f(\delta) \approx f(0) + f'(0)\delta$$

Now

$$\begin{aligned} f(0) &= E\left\{\frac{\sum_j h_j}{\gamma + \sum_j h_j^2}\right\} E\{\mathbf{t}\}, \\ f'(0) &= -E\left\{\frac{n}{\gamma + \sum_j h_j^2} - 2\left(\frac{\sum_j h_j}{\gamma + \sum_j h_j^2}\right)^2\right\} E\left\{\mathbf{t}\mathbf{t}^T\right\}. \end{aligned}$$

Thus δ can be approximated by

$$-[f'(0)]^{-1}f(0) = \xi_0\mathbf{a},$$

where ξ_0 is given by (2.21) and

$$\mathbf{a} = \left[E\left\{\mathbf{t}\mathbf{t}^T\right\}\right]^{-1} E\{\mathbf{t}\}. \tag{2.23}$$

We note that $\mathbf{a}$ in (2.23) agrees with our previous result $\mathbf{a} = (\sqrt{\tau_1}, 0, \cdots, 0)^T$ under the constant-variance model. In this case $\mathbf{t} = (\mathbf{t}_1, \cdots, \mathbf{t}_p)^T$ with $\mathbf{t}_1 = 1/\sqrt{\tau_1}$ constant. Now

$$E\left\{\mathbf{t}\mathbf{t}^T\right\}\begin{bmatrix} \sqrt{\tau_1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = E\{\mathbf{t}\}.$$

When τ_i varies from group to group, slope parameters are not necessarily consistent. Our results show that the asymptotic bias can be approximated by a product of two terms: ξ_0 depending on the group size n and the standardized error distribution $\mathbf{e}_i/\sqrt{\tau_i}$, and $\mathbf{a}$ depending on the explanatory variables $\mathbf{x}_i$ and the variance model τ_i. We give some numerical results on ξ_0 in the next section and on $\mathbf{a}$ in Section 2.4.

2.3 Evaluation of ξ_0

In this section we evaluate ξ_0 under a particular model for the standardized error distribution. Let $g_1, \cdots, g_n$ be independent Gamma$(\alpha, 1)$ random variables with shape parameter α and scale parameter 1; i.e., g_i has density

$$\{\Gamma(\alpha)\}^{-1} g^{\alpha-1} \exp(-g), g > 0.$$

Note that g_i has mean α and variance α. Let u be a Gamma$(\frac{\gamma}{2}, \frac{2}{\gamma})$ random variable; i.e., $u \sim (1/\gamma)\chi^2_\gamma$. We assume that u and $g_1, \cdots, g_n$ are independent. We model the standardized errors $h_j = e_{1j}/\sqrt{\tau_1}$ by

$$h_j = \frac{g_j - \alpha}{\sqrt{\alpha}\sqrt{u}}.$$

In terms of our original model (1.2), the variance σ_1^2 is given by

$$\sigma_1^2 = \tau_1/u,$$

and hence

$$\frac{1}{\sigma_i^2} \sim \frac{1}{\tau_1 \gamma} \chi^2_\gamma.$$

The weights (1.7) are asymptotically optimal under this Bayesian model for the variances. The parameter γ reflects the degree of heteroscedasticity in the σ_i^2 that is not accounted for by the variance model. In this sense, γ can be viewed as an overdispersion parameter. Smaller values of γ correspond to more severe heteroscedasticity.

The Gamma shape parameter α controls the degree of skewness in the error distribution, with smaller α determining greater skewness to the right. We have $(g_i - \alpha)/\sqrt{\alpha} \rightarrow_d N(0,1)$, as $\alpha \rightarrow \infty$.

We used Monte Carlo simulation, with 10,000 replicates of $(h_1, \cdots, h_n)$, to approximate the expectations defining ξ_0 in (2.21). Recall that ξ_0 is based on a linear approximation to either a scalar-valued function f in (2.20) or a vector-valued function f in (2.22). To examine the validity of this approximation, we plotted the scalar-valued f for several values of α and γ. The plots in Figure 2.1 suggest that f is almost linear near zero. Notice that the intercept is further from zero when γ and α are small.

Values for ξ_0 are given in Table 2.1 to Table 2.6 for various values of α, γ, and n. Monte Carlo standard errors for ξ_0 are given in parentheses. The results are summarized in Figure 2.2 by plots of ξ_0 versus $\ln(\alpha n)$ for different values of γ. The plots show that $|\xi_0|$ is a decreasing function of α, γ, and n. The analysis of variance in Table 2.7 shows that, as α, γ, and n vary, most of the variation in ξ_0 can be expressed as a function of γ and $n\alpha$. We do not have a complete explanation for this empirical observation but we expect this is related to the fact that $\sum_{j=1}^n g_j \sim \text{Gamma}(n\alpha, 1)$.

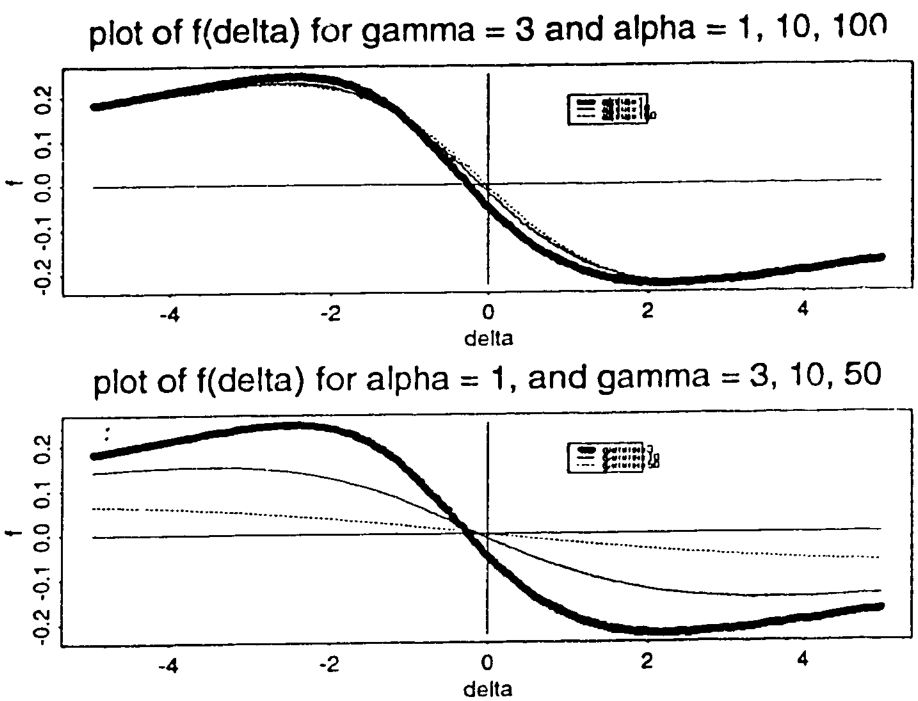

Figure 2.1: f plots

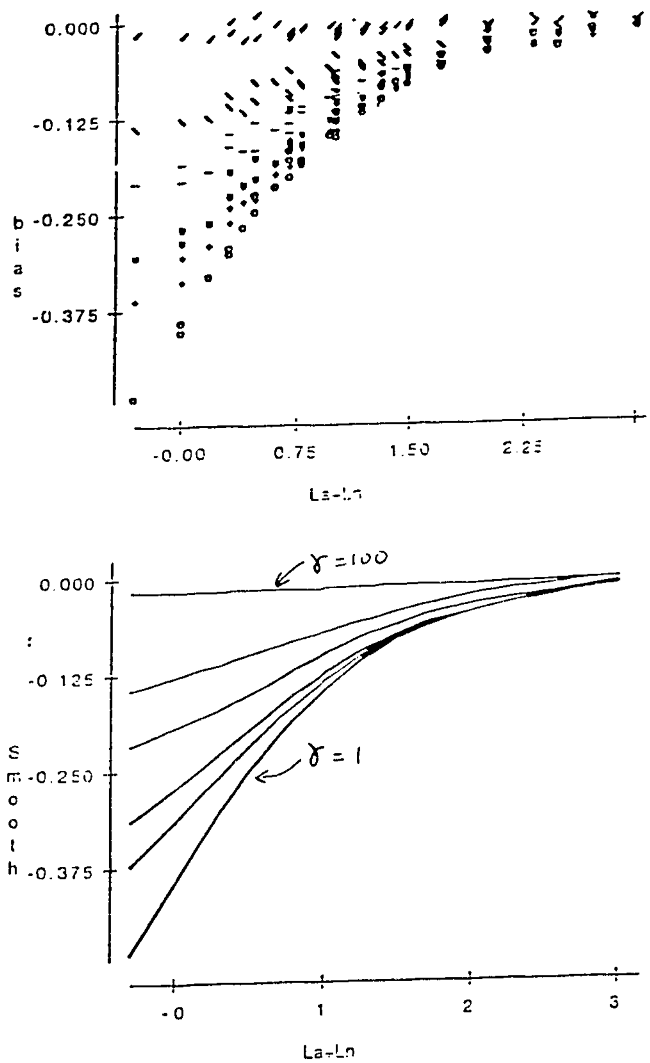

Figure 2.2: Plot of ξ_0 versus $\ln(n\alpha)$

Table 2.1: ξ_0 when $\gamma = 1$

	α						
n	0.5	1	2	3	5	10	100
1	-.491	-.405	-.297	-.228	-.183	-.133	-.032
	(.013)	(.014)	(.015)	(.014)	(.014)	(.014)	(.014)
2	-.391	-.304	-.219	-.183	-.155	-.106	-.024
	(.010)	(.010)	(.010)	(.010)	(.009)	(.009)	(.009)
3	-.334	-.250	-.188	-.152	-.123	-.086	-.040
	(.008)	(.008)	(.008)	(.007)	(.007)	(.007)	(.007)
5	-.269	-.207	-.147	-.114	-.084	-.063	-.017
	(.007)	(.006)	(.006)	(.005)	(.005)	(.005)	(.005)
10	-.172	-.119	-.089	-.067	-.050	-.044	-.014
	(.004)	(.004)	(.004)	(.004)	(.003)	(.004)	(.003)

Table 2.2: ξ_0 when $\gamma = 2$

	α						
n	0.5	1	2	3	5	10	100
1	-.363	-.306	-.243	-.206	-.158	-.092	-.018
	(.010)	(.011)	(.012)	(.012)	(.013)	(.013)	(.013)
2	-.338	-.263	-.201	-.163	-.144	-.081	-.034
	(.008)	(.008)	(.009)	(.009)	(.009)	(.009)	(.009)
3	-.292	-.234	-.167	-.144	-.114	-.078	-.025
	(.007)	(.007)	(.007)	(.007)	(.007)	(.007)	(.007)
5	-.236	-.190	-.132	-.115	-.090	-.063	-.027
	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)
10	-.164	-.120	-.084	-.072	-.060	-.038	-.010
	(.004)	(.004)	(.004)	(.004)	(.004)	(.003)	(.003)

Table 2.3: ξ_0 when $\gamma = 3$

	α						
n	0.5	1	2	3	5	10	100
1	-.307	-.286	-.197	-.178	-.114	-.108	-.030
	(.009)	(.010)	(.012)	(.012)	(.012)	(.012)	(.012)
2	-.269	-.228	-.187	-.149	-.121	-.071	-.036
	(.007)	(.008)	(.008)	(.008)	(.008)	(.008)	(.008)
3	-.262	-.207	-.162	-.129	-.103	-.077	-.022
	(.006)	(.006)	(.007)	(.007)	(.007)	(.007)	(.007)
5	-.213	-.166	-.125	-.103	-.090	-.058	-.020
	(.005)	(.005)	(.007)	(.005)	(.005)	(.005)	(.005)
10	-.159	-.121	-.083	-.063	-.058	-.041	-.009
	(.004)	(.004)	(.003)	(.003)	(.003)	(.003)	(.003)

Table 2.4: ξ_0 when $\gamma = 5$

	α						
n	0.5	1	2	3	5	10	100
1	-.212	-.186	-.148	-.133	-.119	-.067	-.019
	(.008)	(.010)	(.010)	(.011)	(.011)	(.011)	(.012)
2	-.208	-.164	-.143	-.113	-.091	-.055	-.020
	(.006)	(.007)	(.008)	(.008)	(.008)	(.008)	(.008)
3	-.197	-.169	-.121	-.101	-.080	-.062	-.035
	(.005)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)
5	-.168	-.147	-.104	-.082	-.067	-.056	-.016
	(.004)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)
10	-.136	-.099	-.072	-.061	-.047	-.033	-.008
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)

Table 2.5: ξ_0 when $\gamma = 10$

	α						
n	0.5	1	2	3	5	10	100
1	-.140	-.131	-.097	-.081	-.067	-.047	-.012
	(.008)	(.009)	(.010)	(.010)	(.010)	(.011)	(.011)
2	-.131	-.112	-.082	-.084	-.074	-.048	-.003
	(.006)	(.006)	(.007)	(.007)	(.008)	(.007)	(.008)
3	-.127	-.105	-.085	-.064	-.063	-.039	-.015
	(.005)	(.005)	(.006)	(.006)	(.006)	(.006)	(.006)
5	-.119	-.096	-.078	-.059	-.049	-.036	-.003
	(.004)	(.004)	(.005)	(.005)	(.005)	(.005)	(.005)
10	-.102	-.081	-.059	-.046	-.036	-.031	-.010
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)

Table 2.6: ξ_0 when $\gamma = 100$

	α						
n	0.5	1	2	3	5	10	100
1	-.015	-.018	.000	.007	-.014	-.022	-.011
	(.009)	(.010)	(.010)	(.010)	(.010)	(.010)	(.010)
2	-.016	-.015	-.006	-.013	-.010	-.009	-.010
	(.007)	(.007)	(.007)	(.007)	(.007)	(.007)	(.007)
3	-.023	-.022	-.011	-.008	-.013	-.007	-.007
	(.005)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)
5	-.024	-.016	-.014	-.016	-.009	-.002	-.002
	(.004)	(.004)	(.004)	(.004)	(.005)	(.004)	(.004)
10	-.023	-.018	-.016	-.013	-.008	-.012	-.006
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)

Note: Standard errors are in the parentheses

Table 2.7: Analysis of Variance for ξ_0

Source	df	Sum of Squares	Mean Square	F-ratio	Prob
$\log(n\alpha)$	22	0.858	0.041	241.51	$\leq .0001$
γ	5	0.532	0.106	628.73	$\leq .0001$
$\gamma \times \log(n\alpha)$	105	0.310	0.0029	17.437	$\leq .0001$
Error	78	0.[illegible]9	0.000169		
Total	209	1.7325			

2.4 Evaluation of a

In this section we give an example of how the variance model τ can affect the asymptotic bias . We consider a simple linear regression model

$$y_{ij} = \beta_0 + \beta_1 x_i + e_{ij},$$

so $\mathbf{x}_i = (1, x_i)^T$ and $\beta = (\beta_0, \beta_1)^T$. We adopt a power-of-mean model for the variance

$$\log(\tau_i) = \eta_0 + \eta_1 \log(\beta_0 + \beta_1 x_i),$$

and fix $\eta_0 = 0$, $\eta_1 = 2$, and $\beta_0 = 1$. Recall that the asymptotic bias is approximated by $\xi_0 \mathbf{a}$, where

$$\mathbf{a} = \left[E\left\{\mathbf{t}\mathbf{t}^T\right\}\right]^{-1} E\left\{\mathbf{t}\right\},$$

with

$$\begin{aligned} \mathbf{t} &= \frac{1}{\sqrt{\tau_1}} \begin{bmatrix} 1 \\ x_1 \end{bmatrix} \\ &= \frac{1}{1 + \beta_1 x_1} \begin{bmatrix} 1 \\ x_1 \end{bmatrix}. \end{aligned}$$

We used Monte Carlo simulation, with 10,000 replication of x_1, to evalulate $\mathbf{a}$ for several values of β_1 when $x_1 \sim \text{Uniform}(0, 1)$. The results, given in Table 2.8, suggest

Table 2.8: Values for the vector a

β_0	β_1	$\mathbf{a}^T$	
1	0	1.000	0.000
1	1	0.999	0.000
1	2	0.916	0.271
1	3	0.659	1.486
1	5	0.670	2.489
1	10	0.578	5.562

that (i) the asymptotic bias for the intercept β_0 decreases as β_1 increases; (ii) the asymptotic bias for the slope β_1 is zero when $\beta_1 = 0$ and increases with β_1. This is reasonable in view of the fact that the variance τ_i is constant when $\beta_1 = 0$ and τ_i becomes increasingly variable as β_1 increases.

Chapter 3

Bias for Finite Samples

In this chapter we prove that, when τ is constant, the IWLS estimator $\hat{\beta}$ is equivariant and hence its bias is the same for all β. We then describe several Monte Carlo studies for the common mean model and the simple linear regression model for finite samples.

3.1 Equivariance of $\hat{\beta}$ when τ is constant

Consider the model

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{e}_i, \quad i = 1, \cdots, k,$$

and the IWLS estimator

$$\hat{\beta} = \left(\sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \sum_{i=1}^{k} \hat{w}_i \mathbf{X}_i^T \mathbf{y}_i,$$

where

$$\hat{w}_i = \frac{n + \hat{\gamma}}{||\mathbf{y}_i - \mathbf{X}_i\hat{\beta}||^2 + \hat{\gamma}\hat{\tau}}, \tag{3.1}$$

with $\hat{\tau}_i = \hat{\tau}$ the same for all $i = 1, \cdots, k$. We assume that $\hat{\gamma}$ and $\hat{\tau}$ are functions of the residuals $\mathbf{y}_i - \mathbf{X}_i\hat{\beta}$. Now put

$$\zeta = \beta + \delta,$$

and consider estimation of ζ based on data

$$\tilde{\mathbf{y}}_i = \mathbf{X}_i\zeta + \mathbf{e}_i = \mathbf{y}_i + \mathbf{X}_i\delta, \quad i = 1, \cdots, k.$$

We claim that the IWLS estimator $\hat{\zeta}$ satisfies

$$\hat{\zeta} = \hat{\beta} + \delta.$$

To prove this, first consider the estimator

$$\tilde{\zeta} = \left(\sum_1^k \hat{w}_i \mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \sum_{i=1}^k \hat{w}_i \mathbf{X}_i^T \tilde{\mathbf{y}}_i,$$

where $\hat{w}_i$ is given by (3.1). We then have

$$\tilde{\zeta} = \hat{\beta} + \delta,$$

and so

$$\mathbf{y}_i - \mathbf{X}_i\hat{\beta} = \tilde{\mathbf{y}}_i - \mathbf{X}_i\tilde{\zeta}.$$

It follows that the $\hat{w}_i$ are also the weights defining the IWLS estimator $\hat{\zeta}$; i.e., $\tilde{\zeta} = \hat{\zeta}$. Thus we have

$$\hat{\zeta} - \zeta = \hat{\beta} - \beta.$$

Taking expectations shows that the bias $E\{\hat{\beta}\} - \beta$ is the same for all $\beta \epsilon R^p$.

3.2 The Common Mean Model

Consider the model

$$y_{ij} = \beta + e_{ij}, \quad j = 1, \cdots, n, \; i = 1, \cdots, k.$$

The observations are assumed to have the same mean β but different variances in different groups. We assume that the errors are given by

$$e_{ij} = \frac{(g_{ij} - \alpha)\sqrt{\tau}}{\sqrt{\alpha}\sqrt{u_i}},$$

where the g_{ij} and u_i are all independent with $g_{ij} \sim \text{Gamma}(\alpha, 1)$ and $u_i \sim \text{Gamma}\left(\frac{\gamma}{2}, \frac{2}{\gamma}\right)$. In this model we have $\tau_i = \tau = \exp\{\eta\}$, the same for all groups.

We describe a Monte Carlo study to investigate how the bias of the IWLS estimator $\hat{\beta}$ varies with n, γ, and α. Without loss generality (see Section 3.1) we fix $\beta = 0$ and $\eta = 0$. We also fix the total number of observations $nk = 24$. We obtained bias estimates for all combinations of the following values of n, γ, and α:

- $n \in \{1, 2, 3, 4, 6, 8\}$,
- $\gamma \in \{1, 2, 3, 5, 10, 100\}$,
- $\alpha \in \{0.5, 1, 2, 3, 5, 10, 100\}$.

Each bias estimate was based on 10,000 replicates. For each replicate, the IWLS estimate $\hat{\beta}$ was obtained by applying Algorithm 1 in Hooper (1993). We used the OLS estimator as the preliminary estimator $\tilde{\beta}$, then calculated $\hat{\gamma}$, $\hat{\tau}$, $\hat{w}_i$, and $\hat{\beta}$ by (1.9)-(1.11), (1.7) and (1.4), then replaced $\tilde{\beta}$ by $\hat{\beta}$, and repeated the steps until $\hat{\beta}$ converged.

The results are shown in Table 3.1 - Table 3.6. Monte Carlo standard errors are in parentheses. The results are summarized by plots of bias versus $\log(n\alpha)$ for each γ in Figure 3.1. The analysis of variance in Table 3.7 shows that, as α, γ, and n vary, most of the variation can be expressed as a function of γ and $n\alpha$. The results are similar to those for the asymptotic bias ξ_0 described in Section 2.3, except that the plots reveal less variation in the bias as γ varies from 1 to 100. This is because we set bounds $1 \leq \hat{\gamma} \leq 10$ in the algorithm.

Table 3.1: The bias for the common mean model when $\gamma = 1$

n	α						
	0.5	1	2	3	5	10	100
1	-.408 (.001)	-.343 (.001)	-.264 (.002)	-.222 (.001)	-.176 (.001)	-.124 (.002)	-.039 (.002)
2	-.339 (.001)	-.281 (.001)	-.211 (.001)	-.177 (.001)	-.140 (.001)	-.099 (.001)	-.034 (.001)
3	-.291 (.001)	-.236 (.001)	-.179 (.001)	-.146 (.001)	-.115 (.001)	-.084 (.001)	-.026 (.001)
4	-.254 (.001)	-.201 (.001)	-.150 (.001)	-.124 (.001)	-.098 (.001)	-.071 (.001)	-.020 (.001)
6	-.206 (.001)	-.162 (.001)	-.117 (.001)	-.100 (.001)	-.076 (.001)	-.054 (.001)	-.017 (.001)
8	-.173 (.001)	-.133 (.001)	-.098 (.001)	-.077 (.001)	-.063 (.001)	-.045 (.001)	-.015 (.001)

Table 3.2: The bias for the common mean model when $\gamma = 2$

n	$\alpha = 2$						
	0.5	1	2	3	5	10	100
1	-.346 (.001)	-.284 (.001)	-.215 (.001)	-.179 (.001)	-.142 (.001)	-.101 (.001)	-.033 (.001)
2	-.327 (.001)	-.259 (.001)	-.191 (.001)	-.161 (.001)	-.130 (.001)	-.091 (.001)	-.028 (.001)
3	-.290 (.001)	-.226 (.001)	-.167 (.001)	-.140 (.001)	-.108 (.001)	-.080 (.001)	-.023 (.001)
4	-.258 (.001)	-.199 (.001)	-.147 (.001)	-.120 (.001)	-.094 (.001)	-.067 (.001)	-.021 (.001)
6	-.211 (.001)	-.163 (.001)	-.117 (.001)	-.096 (.001)	-.075 (.001)	-.054 (.001)	-.017 (.001)
8	-.178 (.001)	-.133 (.001)	-.099 (.001)	-.080 (.001)	-.061 (.001)	-.045 (.001)	-.014 (.001)

Table 3.3: The bias for the common mean model when $\gamma = 3$

	α						
n	0.5	1	2	3	5	10	100
1	-.285	-.233	-.177	-.150	-.120	-.084	-.027
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
2	-.303	-.230	-.169	-.139	-.110	-.080	-.027
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
3	-.279	-.213	-.152	-.123	-.098	-.068	-.022
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
4	-.252	-.189	-.137	-.111	-.089	-.061	-.021
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
6	-.208	-.156	-.112	-.091	-.073	-.050	-.017
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
8	-.177	-.132	-.093	-.076	-.061	-.041	-.015
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)

Table 3.4: The bias for the common mean model when $\gamma = 5$

	α						
n	0.5	1	2	3	5	10	100
1	-.276 (.001)	-.192 (.001)	-.148 (.001)	-.128 (.001)	-.099 (.001)	-.073 (.001)	-.024 (.001)
2	-.264 (.001)	-.194 (.001)	-.143 (.001)	-.114 (.001)	-.089 (.001)	-.066 (.001)	-.020 (.001)
3	-.262 (.001)	-.191 (.001)	-.131 (.001)	-.107 (.001)	-.083 (.001)	-.057 (.001)	-.017 (.001)
4	-.243 (.001)	-.174 (.001)	-.119 (.001)	-.097 (.001)	-.076 (.001)	-.050 (.001)	-.016 (.001)
6	-.205 (.001)	-.146 (.001)	-.104 (.001)	-.083 (.001)	-.063 (.001)	-.043 (.001)	-.013 (.001)
8	-.170 (.001)	-.127 (.001)	-.089 (.001)	-.068 (.001)	-.054 (.001)	-.038 (.001)	-.011 (.001)

Table 3.5: The bias for the common mean model when $\gamma = 10$

	α						
n	0.5	1	2	3	5	10	100
1	-.195 (.001)	-.162 (.001)	-.130 (.001)	-.108 (.001)	-.087 (.001)	-.062 (.001)	-.020 (.001)
2	-.226 (.001)	-.164 (.001)	-.117 (.001)	-.098 (.001)	-.075 (.001)	-.052 (.001)	-.016 (.001)
3	-.239 (.001)	-.163 (.001)	-.112 (.001)	-.087 (.001)	-.068 (.001)	-.047 (.001)	-.018 (.001)
4	-.227 (.001)	-.156 (.001)	-.104 (.001)	-.083 (.001)	-.063 (.001)	-.045 (.001)	-.012 (.001)
6	-.194 (.001)	-.134 (.001)	-.090 (.001)	-.072 (.001)	-.055 (.001)	-.037 (.001)	-.010 (.001)
8	-.166 (.001)	-.118 (.001)	-.077 (.001)	-.063 (.001)	-.050 (.001)	-.034 (.001)	-.010 (.001)

Table 3.6: The bias for the common mean model when $\gamma = 100$

n	α = 0.5	1	2	3	5	10	100
1	-.169 (.001)	-.143 (.001)	-.111 (.001)	-.097 (.001)	-.077 (.001)	-.055 (.001)	-.017 (.001)
2	-.192 (.001)	-.137 (.001)	-.099 (.001)	-.083 (.001)	-.064 (.001)	-.046 (.001)	-.015 (.001)
3	-.211 (.001)	-.136 (.001)	-.094 (.001)	-.074 (.001)	-.061 (.001)	-.040 (.001)	-.013 (.001)
4	-.208 (.001)	-.133 (.001)	-.088 (.001)	-.069 (.001)	-.052 (.001)	-.036 (.001)	-.012 (.001)
6	-.187 (.001)	-.119 (.001)	-.074 (.001)	-.059 (.001)	-.047 (.001)	-.032 (.001)	-.009 (.001)
8	-.158 (.001)	-.103 (.001)	-.065 (.001)	-.053 (.001)	-.042 (.001)	-.029 (.001)	-.010 (.001)

Table 3.7: Analysis of Variance for the Bias

Source	df	Sum of Squares	Mean Square	F-ratio	Prob
$\log(n\alpha)$	26	1.335	0.051	87.653	$\leq .0001$
γ	5	0.084	0.017	28.689	$\leq .0001$
$\gamma \times \log(n\alpha)$	130	0.069	0.000531	0.90584	0.6992
Error	90	0.053	0.000586		
Total	251	1.581			

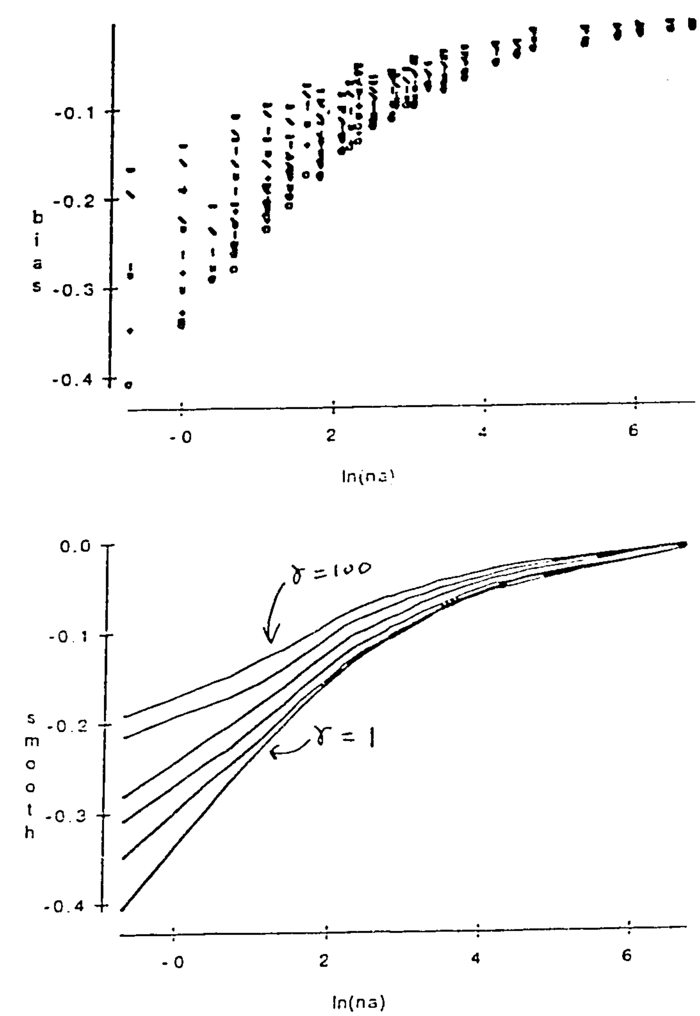

Figure 3.1: Plot of bias versus $\ln(n\alpha)$

3.3 The Simple Linear Regression Model

Consider the simple linear regression model

$$y_{ij} = \beta_0 + \beta_1 x_i + e_{ij}, \quad j = 1, \cdots, n, \quad i = 1, \cdots, k.$$

We first prove that when $\tau_i = \tau$ is constant and $\{x_i\}$ is symmetric, in the sense that $\{x_i - \bar{x} : i = 1, \cdots, k\} = \{-(x_i - \bar{x}) : i = 1, \cdots, k\}$, then the distribution of the IWLS estimator for the slope parameter is symmetric about β_1. By reparameterizing, we may assume without loss of generality that $\bar{x} = 0$. The IWLS estimator is

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \left[\sum \hat{w}_i \begin{pmatrix} 1, x_i \\ x_i, x_i^2 \end{pmatrix}\right]^{-1} \left(\sum \hat{w}_i \bar{y}_{i.} \begin{pmatrix} 1 \\ x_i \end{pmatrix}\right).$$

The bias is given by $E\{(\hat{\delta}_0, \hat{\delta}_1)^T\}$, where

$$\begin{aligned} \begin{pmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \end{pmatrix} &= \left[\sum \hat{w}_i \begin{pmatrix} 1, x_i \\ x_i, x_i^2 \end{pmatrix}\right]^{-1} \left(\sum \hat{w}_i \bar{e}_{i.} \begin{pmatrix} 1 \\ x_i \end{pmatrix}\right), \\ &= \left[\begin{pmatrix} \sum \hat{w}_i, \sum \hat{w}_i x_i \\ \sum \hat{w}_i x_i, \sum \hat{w}_i x_i^2 \end{pmatrix}\right]^{-1} \begin{pmatrix} \sum \hat{w}_i \bar{e}_{i.} \\ \sum \hat{w}_i \bar{e}_{i.} x_i \end{pmatrix}. \end{aligned} \tag{3.2}$$

Thus

$$\hat{\delta}_0 = \left(\sum \hat{w}_i \sum \hat{w}_i x_i^2 - (\sum \hat{w}_i x_i)^2\right)^{-1} \left(\sum \hat{w}_i x_i^2, -\sum \hat{w}_i x_i\right) \begin{pmatrix} \hat{w}_i \bar{e}_{i.} \\ \hat{w}_i \bar{e}_{i.} x_i \end{pmatrix}, \tag{3.3}$$

$$\hat{\delta}_1 = \left(\sum \hat{w}_i \sum \hat{w}_i x_i^2 - (\sum \hat{w}_i x_i)^2\right)^{-1} \left(-\sum \hat{w}_i x_i, \sum \hat{w}_i\right) \begin{pmatrix} \hat{w}_i \bar{e}_{i.} \\ \hat{w}_i \bar{e}_{i.} x_i \end{pmatrix}, \tag{3.4}$$

where

$$\hat{w}_i = \frac{n + \hat{\gamma}}{n\hat{v}_i + \hat{\gamma}\hat{\tau}}, \tag{3.5}$$

$$\hat{v}_i = n^{-1} \sum_{j=1}^{n} \left(e_{ij} - \hat{\delta}_0 - \hat{\delta}_1 x_i\right)^2. \tag{3.6}$$

We assume that $\hat{\gamma}$ and $\hat{\tau}$ are functions of $\{(\hat{v}_i, x_i) : i = 1, \cdots, k\}$. Now the mapping $(x_i, \bar{e}_{i.}) \rightarrow (-x_i, \bar{e}_{i.}), i = 1, \cdots, k$, produces the following effects: $\hat{\delta}_0 \rightarrow \hat{\delta}_0$, $\hat{\delta}_1 \rightarrow -\hat{\delta}_1$,

$\hat{\gamma} \rightarrow \hat{\gamma}$, $\hat{v}_i \rightarrow \hat{v}_i$, $\hat{w}_i \rightarrow \hat{w}_i$. The mapping does not affect the distribution of $\hat{\delta}_1$. Thus we have $\hat{\delta}_1 \sim -\hat{\delta}_1$.

When τ_i is not constant the bias of the $\hat{\beta}_1$ is dependent on the value of β_1. We describe Monte Carlo studies for $\beta_1 = 1$ and $\beta_1 = 2$. We adopt the same model for the errors as in Section 3.2, except the τ_i are now assumed to follow the model

$$\log(\tau_i) \;=\; \eta_0 + \eta_1 \log(\beta_0 + \beta_1 x_i).$$

We fix $\eta_0 = 0$, $\eta_1 = 2$, $\beta_0 = 1$ and the total number of observations $nk = 24$. We choose the x_i to be spread evenly across $[0, 1]$; i.e., $\{x_1, \cdots, x_k\}$ divides $[0, 1]$ into $k+1$ subintervals of equal length. For all combinations of n, α and γ in Section 3.2, we estimated the bias based on 1,000 replicates using the Algorithm 1 in Hooper (1993). The results are summarized in Figure 3.2 and Figure 3.3. These plots show that the absolute bias tends to decrease as $n\alpha$ increases, and increases as β_1 increases. This is as expected. A result that was not expected is that the bias was not related to γ, except for greater variation in bias estimates when γ is small. An examination of the estimates $\hat{\gamma}$ reveals that γ is severely underestimated in most cases. This may be due to the fact that we are attempting to estimate too many parameters $(\beta_0, \beta_1, \eta_0, \eta_1, \gamma)$ with too few data points. It may also be due in part to a technique used to deal with negative values of $\hat{\beta}_0 + \hat{\beta}_1 x_i$ when fitting the log-linear model for τ_i: we replaced $\hat{\beta}_0 + \hat{\beta}_1 x_i$ by its absolute value. In retrospect, we think a different technique would produce better results; e.g., replace $\hat{\beta}_0 + \hat{\beta}_1 x_i$ by $\max\{\epsilon, \hat{\beta}_0 + \hat{\beta}_1 x_i\}$ for some positive constant ϵ. We plan to investigate this possibility by carrying out additional simulations.

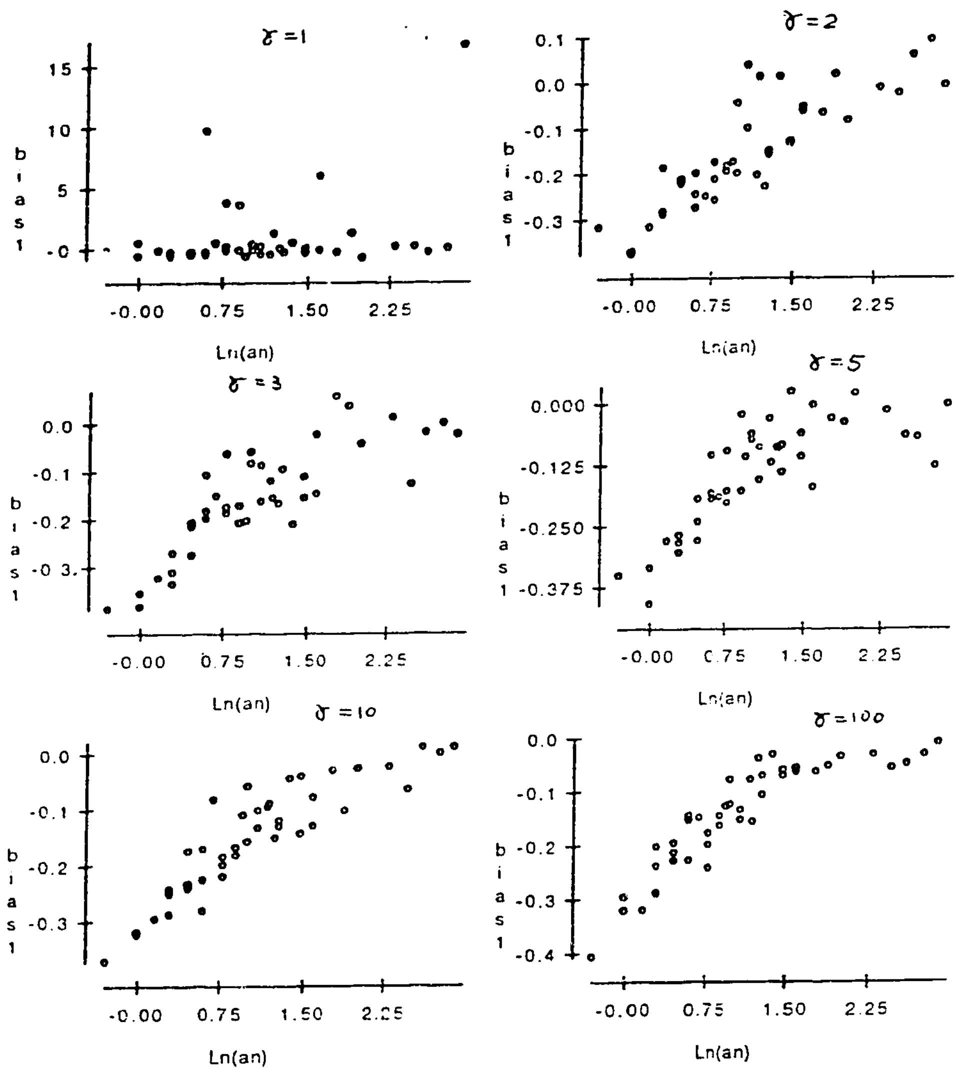

Figure 3.2: Six plots ($\gamma = 1, 2, 3, 5, 10, 100$) of bias verse $\log(n\alpha)$ for slope parameter $\beta_1 = 1$ when $\beta_1 = 1$.

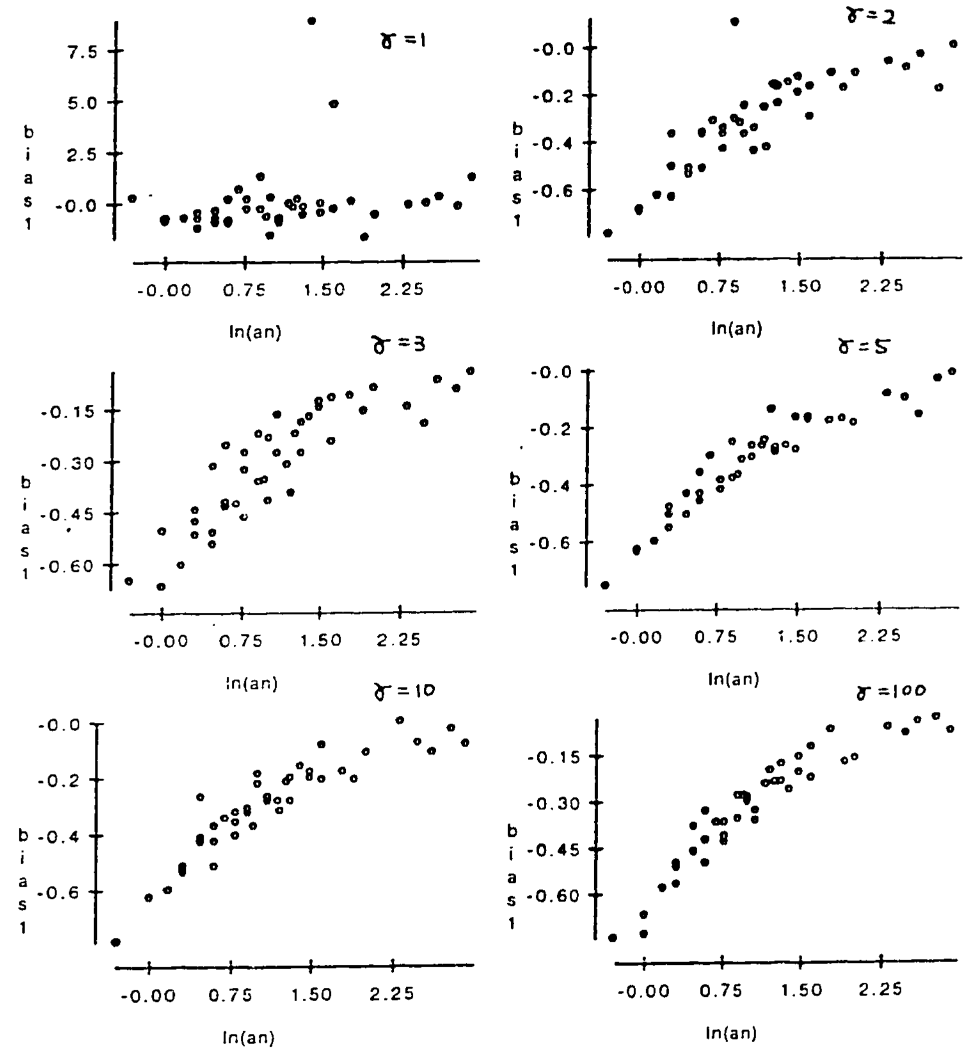

Figure 3.3: Six plots ($\gamma = 1, 2, 3, 5, 10, 100$) of bias verse $\log(n\alpha)$ for slope parameter $\beta_1 = 2$ when $\beta_1 = 2$

Chapter 4

Conclusion

For the heteroscedastic linear model with the same number of replications in each group, we studied the performance of an IWLS estimator when the error distribution is asymmetric. The weights defining the IWLS estimator were derived by Hooper (1993) assuming normal errors and a Bayesian model for the variances.

We derived an approximation for the asymptotic bias of the IWLS estimator as a product of two terms: a scalar term depending on group size and the standardized error distribution, and a vector term depending on the model for the means and the model for the variances. Under the constant-varian ce model, the slope parameters are estimated consistently. The approximation was evaluated assuming a particular family of skewed distribution. The results suggest that the magnitude of the asymptotic bias increases with the degree of skewness and heteroscedasticity, and decreases with the group size.

We also carried out a Monto Carlo study to evaluate the finite sample performance of the IWLS estimator in several examples. The results were in rough agreement with the asymptotic approximations. depends on the explanatory variables and τ_i. When τ_i is constant, we have shown that the IWLS estimator for the slope parameters are estimated consistently. The simulation study showed the absolute value for the asymptotic bias decreases as skewness

Bibliography

[1] Bement, T. R., and Williams, J. S., (1969), "Variance of Weighted Regression Estimation When Sampling Errors are Independent and Heteroscedastic," *Journal of the American Statistical Association, 64, 1369-1382.*

[2] Box, G. E. P., and Cox, D. R., (1964), "Analysis of Transformations," *Journal of the Royal Statistical Society, Ser. B, 26, 211-252.*

[3] Box, G. E. P., and Hill, W. J., (1974), "Correcting Inhomogeneity of Variance with Power Transformation Weighting," *Technometrics, 16, 385-389.*

[4] Carroll, R. J., (1979), "On Estimating Variances of Robust Estimators When Errors are Asymmetric," *Journal of the American Statistical Association, 74, 674-679.*

[5] Carroll, R. J., (1982), "Adapting for Heteroscedasticity in Linear Models," *The Annals of Statistics, 10, 1224-1233.*

[6] Carroll, R. J., and Cline, D. B. H., (1988), "An Asymptotic Theory for Weighted Least Squares with Weights Estimated by Replicates," *Biometrika, 75, 35-43.*

[7] Carroll, R. J., and Ruppert, D., (1982) "Robust Estimation in Heteroscedastic Linear Models," *The Annals of Statistics, 10, 429-441.*

[8] Carroll, R. J., and Ruppert, D., (1984), "Power Transformations when Fitting Theoretical Models to Data," *Journal of the American Statistical Association, 79, 321-328.*

[9] Carroll, R. J., and Ruppert, D., (1988), "Transformation and Weighting in Regression," *New York, Chapman and Hall.*

[10] Cochran, W. G., (1937), "Problems Arising in the Analysis of a Series of Similar Experiments," *Journal of the Royal Statistical Society, Suppl.4, 102-118.*

[11] Fuller, W. A., and Rao, J. N. K., (1978), "Estimation for Linear Regression Model with Unknown Diagonal Covariance Matrix," *The Annals of Statistics, 6, 1149-1158.*

[12] Hooper, P. M., (1993), "Iterative Weighted Least Squares Estimation in Heteroscedastic Linear Models," *Journal of the American Statistical Association, 88, 179-184.*

[13] Huber, J. D., (1981), *Robust Statistics*, Wiley, New York.

[14] Jobson, J. D., and Fuller, W. A., (1980), "Least Squares Estimation When Covariance Matrix and Parameter vector are functionally Related," *Journal of the American Statistical Association, 75, 176-181.*

[15] Neyman, J., and Scott, E. L., (1948), "Consistent Estimator Based on Partially Consistent Observations," *Econometrica, 16, 1-32.*

[16] Ruppert, D., and Aldershof, B. (1989), "Transformation to Symmetry and Homoscedasticity ," *Journal of the American Statistical Association, 84, 437-446.*

[17] Shao, J., (1989), "Asymptotic Distribution of Weighted Least Squares Estimator," *The Annals of Statistics, 41, 365-382.*

[18] Shao, J., (1992), "Emprical Bayes Estimation of Heteroscedastic Variances," *Statistica Sinica, 2, 495-518.*

[19] Shao, J., and Chen, J. H., (1993), "Iterative Weighted Least Squares Estimators," *The Annals of Statistics, 21, 1071-1092.*

[20] Welsh, A. H., Carroll, R. J., and Ruppert, D., (1994), "Fitting Heteroscedastic Regression Models," *Journal of the American Statistical Association, 89, 100-116.*