PARAMETRIC APPROACH TO ESTIMATE VARIANCE FUNCTIONS

Titus K. Kibua
Statistics and Actuarial Science Department
Kenyatta University , Nairobi, Kenya
Email - kibua.titus@ku.ac.ke

Abstract: *Data which exhibit none constant variance is considered. Various procedures to deal with this problem are explored and their computational forms presented. Detailed investigation through empirical study to demonstrate their utility is patiently carried out using entirely simulated data. Looking at various methods of comparison, it is seen that at one time logarithm of absolute residuals, extended quasi likelihood and Rodbard and Frazier procedures display slightly better results. At other time, absolute residuals with leverage corrected do well. Again, at another time modified maximum likelihood as well as pseudo likelihood leads. Hence no procedures are very poor. For cases where replication is missing, absolute residuals may be good and modified maximum likelihood may be recommended for replicated cases.*

Key Words: *Heteroscedastic, variance, likelihood.*

1. INTRODUCTION:

Consider a general heteroscedastic regression model for observable data y_i given by

$$y_i = f(x_i, \beta) + e_i, \qquad i = 1,\ldots,n \qquad (1)$$

Where f is an unknown mean response function, e_i are uncorrelated errors with zero mean and variance σ_i^2, x_i is a p vector of predictors, β is a $p \times 1$ regression parameter and n is the sample size. The heteroscedasticity represented by non constant σ_i may be regarded as of unknown form or may be modeled as a function of the independent variable x, known factors exogenous to the model and the regression parameters. The variance function may be completely known, specified up to additional unknown parameters or completely unknown.

Under the parametric approach, the assumption is that the variances are not constant according to model (1). The problem of heterogeneity may be attacked directly by specifying models for both the mean and the variance and, in particular, a variance model with unknown parameters which must be estimated. A general parametric model for the variance can be written as

$$\sigma_i^2 = \sigma^2 g^2\left(z_i, \mu_i(\beta), \theta\right) \qquad (2)$$

Where σ is an unknown scale parameter, g is the variance function, z_i is a known vector possibly containing x_i, $\mu_i(\beta) = f(x_i, \beta)$ and θ is an unknown $r \times 1$ vector of parameters. There are a number of graphical techniques which can be used in choosing the model to be fitted by letting the data reveal themselves. The unweighted least squares residual plot is most widely used (see, for example, Weisberg 1985 and Carroll and Ruppert 1988).

If $z_i = x_i$, the variance depends on the predictors. The variance can also depend on the known mean $\mu_i(\beta)$ or on the estimated mean response $\mu_i(\hat{\beta})$.

In practice as well as for theoretical investigations, g is taken to be known and to satisfy appropriate smoothness conditions. In a model such as (2), estimation of the variances essentially reduces to the estimation of θ, since β will be estimated routinely and the final estimates of β and θ may be used to obtain a final estimate of σ. Thus investigations of the properties of variance estimators for (2) focus on properties of estimators for θ. In some applications, estimation of θ is not the only problem of interest. In chemical and biological assay problems, issues of prediction and calibration arise. In such problems the estimator of θ plays a central role. In radioimmunoassay, the statistical properties of prediction intervals and constructs such as the minimum detectable concentration are highly dependent on how one estimates θ (Raab, 1981). In engineering, quality improvement applications is an important goal to discover the source of variability. This can be obtained directly from the variance function estimate. These and many other practical examples indicate the importance of the choice of the method for estimating the variance function. In the case of model (2), the choice is defined by how we choose to estimate the variance function, g and, in particular, θ.

Many of the methods for estimating θ that have been proposed in the literature are (possibly weighted) regression methods based on functions of either absolute residuals from the current regression fit or, in the case of replications at each design point, sample standard deviations. Still other methods are joint estimation methods based on assumption about the underlying distributions in which (σ, β, θ) are in principle estimated simultaneously.

2. ESTIMATION PROCEDURES:

2.1. Pseudo likelihood procedure

Assume the data are normal and then write the likelihood as

$$L(\beta, \sigma^2, \theta / \mathrm{Y}) = \prod_{i=1}^{N} \left[2\pi\left[\sigma^2 g^2(z_i, \mu_i(\beta), \theta)\right]\right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\frac{[y_i - f(x_i, \beta)]^2}{\sigma^2 g^2(z_i, \mu_i(\beta), \theta)}\right]$$

$$= (2\pi)^{-\frac{N}{2}}\left[\sigma^2 g^2(z_i \mu_i(\beta), \theta)\right]^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\frac{\sum_{i=1}^{N}[y_i - f(x_i, \beta)]^2}{\sigma^2 g^2(z_i, \mu_i(\beta), \theta)}\right]$$

Let $\ell = \log L(\beta, \sigma^2, \theta / \mathrm{Y})$, then

$$\ell = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log \sigma^2 g^2(z_i, \mu_i(\beta), \theta) - \frac{1}{2}\frac{\sum_{i=1}^{N}[y_i - f(x_i, \beta)]^2}{\sigma^2 g^2(z_i, \mu_i(\beta), \theta)}$$

Now $\frac{\partial \ell}{\partial \beta} = 0 \;\Rightarrow\; \frac{1}{\sigma^2}\frac{\sum_{i=1}^{N}\left[[y_i - f(x_i, \beta)]\frac{\partial}{\partial \beta} f(x_i, \beta)\right]}{g^2(z_i, \mu_i(\beta), \theta)} = 0$ and

$$\frac{\sum_{i=1}^{N}\frac{\partial}{\partial \beta} f(x_i, \beta)[y_i - f(x_i, \beta)]}{g^2(z_i, \mu_i(\beta), \hat{\theta})} = 0 \qquad (3)$$

Equation (3) provides the estimate for the regression parameter β for some estimated value of θ. To obtain this estimated value of θ maximize the log likelihood $\ell(\theta, \sigma, \hat{\beta} / \mathrm{Y})$

Where $L(\hat{\beta}, \theta, \sigma^2 / \mathrm{Y}) = \prod_{i=1}^{N}\left[2\pi\sigma^2 g^2(z_i, \mu_i(\hat{\beta}), \theta)\right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\frac{[y_i - f(x_i, \hat{\beta})]^2}{\sigma^2 g^2(z_i, \mu_i(\hat{\beta}), \theta)}\right]$

$$= (2\pi)^{-\frac{N}{2}}(\sigma^2)^{-\frac{N}{2}}\left[g^2(z_i, \mu_i(\hat{\beta}), \theta)\right]^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\frac{\sum_{i=1}^{N}[y_i - f(x_i, \hat{\beta})]^2}{\sigma^2 g^2(z_i, \mu_i(\hat{\beta}), \theta)}\right]$$

Now write $\ell' = \log L(\hat{\beta}, \theta, \sigma^2 / \mathrm{Y})$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{N}{2}\log\left[g^2(z_i, \mu_i(\hat{\beta}), \theta)\right] - \frac{1}{2}\frac{\sum_{i=1}^{N}[y_i - f(x_i, \hat{\beta})]^2}{\sigma^2 g^2(z_i, \mu_i(\hat{\beta}), \theta)}$$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{N}{2}\log\left[g^2(z_i, \mu_i(\hat{\beta}), \theta)\right] - \frac{1}{2}\frac{\sum_{i=1}^{N} r_i^2}{\sigma^2}$$

Where $r_i = \frac{y_i - f(x_i, \hat{\beta})}{g(z_i, \mu_i(\hat{\beta}), \theta)}$. Then $\frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{N}\frac{r_i^2}{\sigma^4} = 0$ yields

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} r_i^2}{N} \qquad (4)$$

Next, let $\frac{\partial \ell}{\partial \theta} = 0$. Thus

$$\frac{-Ng(z_i, \mu_i(\hat{\beta}), \theta)\frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{g^2(z_i, \mu_i(\hat{\beta}), \theta)} + \frac{\sum_{i=1}^{N}[y_i - f(x_i, \hat{\beta})]^2 \frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{\hat{\sigma}^2 g^3(z_i, \mu_i(\hat{\beta}), \theta)} = 0$$

Or $\frac{-N\frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{g(z_i, \mu_i(\hat{\beta}), \theta)} + \frac{\sum_{i=1}^{N}[y_i - f(x_i, \hat{\beta})]^2 \frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{\hat{\sigma}^2 g^3(z_i, \mu_i(\hat{\beta}), \theta)} = 0$

Simplifying gives $-Nv\theta_i + \sum_{i=1}^{N}\frac{r_i^2 v\theta_i}{\hat{\sigma}^2} = 0$ or $\frac{-N\hat{\sigma}^2 v\theta_i + \sum_{i=1}^{N} r_i^2 v\theta_i}{\hat{\sigma}^2} = 0$

Where $v\theta_i = \frac{\frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{g(z_i, \mu_i(\hat{\beta}), \theta)}$. Then $\hat{\theta}$ is obtained by solving

$$\sum_{i=1}^{N}(r_i^2 - \hat{\sigma}^2)v\theta_i = 0 \qquad (5)$$

2.2. Restricted maximum likelihood procedure

Define the hat matrix $H_{n\times n} = X_*(X_*^T X_*)^{-1} X_*^T$ with diagonal elements h_{ii} where X_* is the $n \times p$ matrix where the i^{th} row is the transpose of the column vector

$\frac{\frac{\partial}{\partial \beta} f(x_i, \hat{\beta})}{g(z_i, \mu_i(\hat{\beta}), \hat{\theta})}$. The diagonal elements h_{ii} are the leverage values. Then using these leverage values and changing the divisor of (4) to $n-p$ where p is the number of regression parameters, solve for θ and σ equations

$\sum_{i=1}^{N}\frac{[y_i - f(x_i, \hat{\beta})]^2}{\sigma^2 g^2(z_i, \mu_i(\hat{\beta}), \theta)}\begin{bmatrix}1 \\ v\theta_i\end{bmatrix} = \begin{bmatrix}N-p \\ \sum_{i=1}^{N}[v\theta_i(1-h_{ii})]\end{bmatrix}$ to obtain

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} r_i^2}{N-p} \qquad (6)$$

and $\sum_{i=1}^{N}\frac{[y_i - f(x_i, \hat{\beta})]^2}{\sigma^2 g^2(z_i, \mu_i(\hat{\beta}), \theta)} v\theta_i = \sum_{i=1}^{N}[v\theta_i(1-h_{ii})]$. Finally $\hat{\theta}$ is obtained from

$$\sum_{i=1}^{N} r_i^2 v\theta_i - \hat{\sigma}^2 \sum_{i=1}^{N}\left[v\theta_i(1-h_{ii})\right] = 0 \qquad (7)$$

2.3. Least squares on squared residuals procedure

From equation (1), write the squared residuals as $\left[y_i - f\left(x_i, \hat{\beta}_*\right)\right]^2$ where $\hat{\beta}_*$ is the current estimate of β. Consider a regression problem where the responses are the squared residuals and the regression function is its approximate expectation $\sigma^2 g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)$.

Thus write $E\left[y_i - f\left(x_i, \hat{\beta}_*\right)\right]^2 \simeq \sigma^2 g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)$. Then minimize

$\sum_{i=1}^{N}\left[\left(y_i - f\left(x_i, \hat{\beta}_*\right)\right)^2 - \sigma^2 g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right]^2$ in σ and θ. However, for normally distributed data the squared residuals are themselves heteroscedastic with variance approximately proportional to $\sigma^4 g^4\left(z_i, \mu_i(\beta), \theta\right)$. Thus one is naturally led to generalized least squares, see Jobson and Fuller (1980). Hence, for generalized least squares, this suggests minimizing with respect to σ and θ the weighted least squares version

$$\sum_{i=1}^{N} \frac{\left[\left(y_i - f\left(x_i, \hat{\beta}_*\right)\right)^2 - \sigma^2 g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right]^2}{g^4\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)} \text{ obtaining}$$

$$\sum_{i=1}^{N} \frac{\left[\left(y_i - f\left(x_i, \hat{\beta}_*\right)\right)^2 - \sigma^2 g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}\right)\right]\sigma g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}\right)}{g^4\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)} \qquad (8)$$

and

$$\sum_{i=1}^{N} \frac{\left[\left(y_i - f\left(x_i, \hat{\beta}_*\right)\right)^2 - \hat{\sigma}^2 g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right] g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right) \frac{\partial}{\partial\theta_i} g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)}{g^4\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)} = 0 \qquad (9)$$

Where $\hat{\theta}_*$ is the current estimate of θ. Now solve equations (8) and (9) to get $\hat{\sigma}$ and $\hat{\theta}$ respectively. Next, to account for the effect of leverage minimize

$$\sum_{i=1}^{N} \frac{\left[\left(y_i - f\left(x_i, \hat{\beta}_*\right)\right)^2 - \sigma^2 (1-h_{ii}) g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right]^2}{\left[(1-h_{ii})^2 g^4\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)\right]} \text{ in } \sigma \text{ and } \theta \text{ obtaining}$$

$$\sum_{i=1}^{N} \frac{\left[\left(y_i - f\left(x_i, \hat{\beta}_*\right)\right)^2 - \sigma^2 (1-h_{ii})^2 g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}\right)\right] g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}\right)}{\left[(1-h_{ii})^2 g^4\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)\right]} = 0 \qquad (10)$$

and

$$\sum_{i=1}^{N} \frac{\left[\left(y_i - f\left(x_i, \hat{\beta}_*\right)\right)^2 - \hat{\sigma}^2 (1-h_{ii})^2 g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right] g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right) \frac{\partial g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)}{\partial\theta_i}}{\left[(1-h_{ii})^2 g^4\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)\right]} = 0 \qquad (11)$$

with h_{ii} as defined under Subsection (2.2). Solving equations (10) and (11) gives $\hat{\sigma}$ and $\hat{\theta}$ respectively. This is done iteratively.

2.4. Least squares on absolute residuals procedure

Write $E\left|y_i - f\left(x_i, \hat{\beta}_*\right)\right| = \sigma g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)$ leading to the minimization of

$\sum_{i=1}^{N}\left[\left|y_i - f\left(x_i, \hat{\beta}_*\right)\right| - \sigma g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right]^2$ with respect to σ and θ. However, since the residuals are to be appropriately weighted, Carroll and Ruppert(1988), suggest estimating θ by minimizing with respect to σ and θ the weighted version namely

$\sum_{i=1}^{N}\frac{\left[\left|y_i - f\left(x_i, \hat{\beta}_*\right)\right| - \sigma g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right]^2}{g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)}$. This implies that, to find $\hat{\sigma}$ solve

$$\sum_{i=1}^{N}\frac{\left[\left|y_i - f\left(x_i, \hat{\beta}_*\right)\right| - \sigma g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}\right)\right] g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}\right)}{g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)} = 0 \qquad (12)$$

and for $\hat{\theta}$ solve

$$\sum_{i=1}^{N}\frac{\left[\left|y_i - f\left(x_i, \hat{\beta}_*\right)\right| - \hat{\sigma} g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right]\hat{\sigma} g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\frac{\partial}{\partial \theta_i} g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)}{g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)} = 0$$

Which can be written as

$$\sum_{i=1}^{N}\frac{\left[\left|y_i - f\left(x_i, \hat{\beta}_*\right)\right| - \hat{\sigma} g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right] g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\frac{\partial}{\partial \theta_i} g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)}{g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)} = 0 \qquad (13)$$

Next, modify this procedure to account for the effect of leverage by minimizing with respect to σ and θ $\sum_{i=1}^{N}\frac{\left[\left|y_i - f\left(x_i, \hat{\beta}_*\right)\right| - \sigma(1-h_{ii}) g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right]^2}{\left[(1-h_{ii})^2 g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)\right]}$. Differentiate with respect to σ and θ respectively and equate to zero obtaining equations

$$\sum_{i=1}^{N}\frac{\left[\left|y_i - f\left(x_i, \hat{\beta}_*\right)\right| - \sigma(1-h_{ii}) g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}\right)\right] g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}\right)}{\left[(1-h_{ii}) g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)\right]} \qquad (14)$$

and

$$\sum_{i=1}^{N}\frac{\left[\left|y_i - f\left(x_i, \hat{\beta}_*\right)\right| - \hat{\sigma}(1-h_{ii}) g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\right] g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)\frac{\partial}{\partial \theta_i} g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)}{\left[(1-h_{ii}) g^2\left(z_i, \mu_i\left(\hat{\beta}_*\right), \hat{\theta}_*\right)\right]} = 0 \qquad (15)$$

2.5. Modified maximum likelihood procedure

This is the problem of estimating variances by pooling information from a large number of small samples. At each predictor value x_i, observe m_i replicated responses y_{ij}, $i = 1,\dots, M$ and $j = 1,\dots, m_i$. Consider the case of equal replications $m_i = m$ so that $M = nm$ is the total number of observations. Raab (1981,p. 35) suggested the modification of the standard likelihood replacing the term $\sigma^{-\frac{m_i}{2}}$ by $\sigma^{-\left(\frac{m_i-1}{2}\right)}$ and gives a number of justifications. Adopt this modified likelihood and write

$$L\left(\beta,\theta,\sigma^2/\mathrm{Y}\right)=\prod_{i=1}^{M}\left[2\pi\sigma^2 g^2\left(z_i,\mu_i(\beta),\theta\right)\right]^{-\left(\frac{m-1}{2}\right)}\exp\left[-\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\beta)\right)^2}{2\sigma^2 g^2\left(z_i,\mu_i(\beta),\theta\right)}\right].$$

Taking logarithm obtain

$$\ell=\log L\left(\beta,\theta,\sigma^2/\mathrm{Y}\right)=-\left(\frac{m-1}{2}\right)\sum_{i=1}^{M}\log\left[2\pi\sigma^2 g^2\left(z_i,\mu_i(\beta),\theta\right)\right]-\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\beta)\right)^2}{2\sigma^2 g^2\left(z_i,\mu_i(\beta),\theta\right)}$$

Now differentiate with respect to σ^2 and equate to zero, to obtain

$$\frac{\partial\ell}{\partial\sigma^2}=-\left(\frac{m-1}{2}\right)\sum_{i=1}^{M}\frac{1}{\sigma^2}+\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\beta)\right)^2}{2\sigma^4 g^2\left(z_i,\mu_i(\beta),\theta\right)}$$

giving

$$\frac{1}{2\sigma^2}\sum_{i=1}^{M}(m-1)=\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\beta)\right)^2}{2\sigma^4 g^2\left(z_i,\mu_i(\beta),\theta\right)}$$

Hence

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-u_i(\hat{\beta})\right)^2}{g^2\left(z_i,\mu_i(\hat{\beta}),\hat{\theta}\right)}}{\sum_{i=1}^{M}(m-1)} \tag{16}$$

Similarly, differentiate ℓ with respect to θ and equate to zero obtaining

$$\frac{\partial\ell}{\partial\theta}=-(m-1)\sum_{i=1}^{M}\frac{\frac{\partial}{\partial\theta_i}g\left(z_i,\mu_i(\beta),\theta\right)}{g\left(z_i,\mu_i(\beta),\theta\right)}+\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\beta)\right)^2\frac{\partial}{\partial\theta_i}g\left(z_i,\mu_i(\beta),\theta\right)}{\sigma^2 g^3\left(z_i,\mu_i(\beta),\theta\right)}$$

or

$$\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\hat{\beta})\right)^2\frac{\partial}{\partial\theta_i}g\left(z_i,\mu_i(\hat{\beta}),\theta\right)}{\hat{\sigma}^2 g^3\left(z_i,\mu_i(\hat{\beta}),\theta\right)}-(m-1)\sum_{i=1}^{M}\frac{\frac{\partial}{\partial\theta_i}g\left(z_i,\mu_i(\hat{\beta}),\theta\right)}{g\left(z_i,\mu_i(\hat{\beta}),\theta\right)}=0 \tag{17}$$

Finally differentiate ℓ with respect to β and equate to zero getting

$$\frac{\partial\ell}{\partial\beta}=-(m-1)\sum_{i=1}^{M}\frac{\frac{\partial g\left(z_i,\mu_i(\beta),\theta\right)}{\partial\beta}}{g\left(z_i,\mu_i(\beta),\theta\right)}+\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\beta)\right)}{\sigma^2 g^2\left(z_i,\mu_i(\beta),\theta\right)}$$

$$+\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\beta)\right)^2\frac{\partial g\left(z_i,\mu_i(\beta),\theta\right)}{\partial\beta}}{\sigma^2 g^3\left(z_i,\mu_i(\beta),\theta\right)}$$

$$\Rightarrow\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\beta)\right)}{\hat{\sigma}^2 g^2\left(z_i,\mu_i(\beta),\hat{\theta}\right)}$$

$$+\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{\left(y_{ij}-\mu_i(\beta)\right)^2\frac{\partial g\left(z_i,\mu_i(\beta),\hat{\theta}\right)}{\partial\beta}}{\hat{\sigma}^2 g^3\left(z_i,\mu_i(\beta),\hat{\theta}\right)}-(m-1)\sum_{i=1}^{M}\frac{\frac{\partial g\left(z_i,\mu_i(\beta),\hat{\theta}\right)}{\partial\beta}}{g\left(z_i,\mu_i(\beta),\hat{\theta}\right)}=0 \tag{18}$$

Now solve equations (16), (17) and (18) to obtain $\hat{\sigma}^2, \hat{\theta}$ and $\hat{\beta}$ respectively. Note that for the usual maximum likelihood estimate σ is biased. It is made unbiased in this case by dividing the corrected sum of squares by the degrees of freedom rather than the sample size.

2.6. Extended quasi likelihood procedure

Wedderbuen (1974) gives the definition of quasi likelihood while Nelder and Pregibon (1987) discuss the extended quasi likelihood. When θ is known and the variance function has the form (2), quasi likelihood estimation of β is a form of iterated generalized least squares. The extended quasi likelihood method is a joint estimation scheme which attempts to extend the notion of quasi likelihood to include estimation of θ. The method is based on the assumption that the data arise from a class of distributions depending on θ and involves estimation of θ by minimizing with respect to β, θ and σ^2 the extended quasi likelihood.

$Q^+ = -\frac{1}{2}\sum_{i=1}^{N}\left[\log\{2\pi\sigma^2 g^2(z_i, y_i, \theta)\} - \frac{2}{\sigma^2}\int_{y_i}^{\mu_i(\beta)} \frac{y_i - u}{g^2(z_i, \mu_i(\beta), \theta)}du\right]$ as in Davidian (1986, p.16). Differentiate Q^+ with respect to θ and σ^2 to obtain

$$\frac{\partial Q^+}{\partial \theta} = -\frac{1}{2}\sum_{i=1}^{N}\left[2\frac{\frac{\partial}{\partial\theta_i} g(z_i, y_i, \theta)}{g(z_i, y_i, \theta)} + \frac{4}{\sigma^2}\int_{y_i}^{\mu_i(\beta)}\left\{\frac{y_i - u}{g^3(z_i, \mu_i(\beta), \theta)}\right\}\left\{\frac{\partial}{\partial\theta_i} g(z_i, \mu_i(\beta), \theta)\right\}du\right]$$

and

$$\frac{\partial Q^+}{\partial \sigma^2} = -\frac{1}{2}\sum_{i=1}^{N}\left[\frac{1}{\sigma^2} + \frac{2}{\sigma^4}\int_{y_i}^{\mu_i(\beta)}\frac{y_i - u}{g^2(z_i, \mu_i(\beta), \theta)}du\right]$$

Equating the derivatives to zero gives $\hat{\theta}$ as the solution of

$$\sum_{i=1}^{N}\left[2\frac{\frac{\partial}{\partial\theta_i} g(z_i, y_i, \theta)}{g(z_i, y_i, \theta)} + \frac{4}{\hat{\sigma}^2}\int_{y_i}^{\mu_i(\beta)}\left\{\frac{y_i - u}{g^3(z_i, \mu_i(\hat{\beta}), \theta)}\right\}\left\{\frac{\partial}{\partial\theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)\right\}du\right] = 0 \qquad (19)$$

and for $\hat{\sigma}^2$ solve

$$\sum_{i=1}^{N}\left[\frac{1}{\sigma^2} + \frac{2}{\sigma^4}\int_{y_i}^{\mu_i(\hat{\beta})}\frac{y_i - u}{g^2(z_i, \mu_i(\hat{\beta}), \hat{\theta})}du\right] = 0 \qquad (20)$$

Similarly, differentiating Q^+ with respect to β gives

$$\sum_{i=1}^{N}\left(\frac{y_i - \mu_i(\beta)}{g^2(z_i, \mu_i(\beta), \hat{\theta})}\right)\left(\frac{\partial \mu_i(\beta)}{\partial \beta}\right) = 0 \qquad (21)$$

2.7. Logarithm of absolute residuals procedure

This procedure exploits the fact that $E\left|y_i - f\left(x_i, \mu_i\left(\hat{\beta}_*\right)\right)\right| \simeq \sigma g\left(z_i, \mu_i\left(\hat{\beta}_*\right), \theta\right)$ and uses a two-step estimation process. The first step consists of taking the natural logarithms of the absolute residuals $\left|y_i - f\left(x_i, \mu_i\left(\hat{\beta}_*\right)\right)\right|$. These are then regressed on $\log\left\{\sigma g\left(z, \mu\left(\hat{\beta}_*\right), \theta\right)\right\}$

thereby yielding estimates of θ as the slope and $\log\sigma^2$ as the intercept. With the assumption that the errors are independent and identically distributed, this should be approximately a homoscedastic regression. A practical problem arises if one of the residuals is very near zero, in which case taking logarithms induces a rather large and artificial outlier. To avoid this potential difficulty for fitting the variance model, Carroll and Ruppert (1988) suggest that one might wish to delete a few of the smallest absolute residuals.

2.8. Rodbard and Frazier procedure

This method uses replication as in the case of modified maximum likelihood and it is identical to the logarithm method. The idea is to avoid dependence on unweighted methods. Here the absolute residual is replaced by the sample standard deviation and $f\left(x_i, \hat{\beta}_*\right)$ in the regression function is replaced by the sample mean $\bar{y}_i$. Thus the procedure is to regress the logarithm of the sample standard deviation on the logarithm of the sample mean.

2.9. Maximum likelihood procedure

The process here is the same as in the pseudo likelihood procedure. However instead of fixing β at the current value $\hat{\beta}$ and maximizing the likelihood function in θ, one maximizes the likelihood function jointly in β and θ. Maximum likelihood assumes that the variances do not depend on the mean.

2.10. Sadler and Smith procedure

This is similar to the modified maximum likelihood procedure where one uses the sample mean $\bar{y}_i$ instead of μ_i.

3. SIMULATION STUDY:

Consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \qquad i = 1, \ldots, n \qquad (22)$$

Take a simple variance function (for example, Carroll and Ruppert, 1988, p 65)

$$\text{var}(y_i) = \sigma^2 g^2(x_i, \theta) = \sigma^2 (1 + \theta x_i^2)^2 \qquad (23)$$

Let $e_i = \varepsilon_i \sigma(1 + \theta x_i^2)$ and generate independent values of ε_i according to $\varepsilon_i \sim N(0, 1)$. Further, let the $x_i^{'}s$ be equally spaced on the interval $[-1, 1]$ with 0.1 spacing. The simulations are performed with a sample of size $n = 21$. Next, fix β_0 and β_1 to be 0 and 1 respectively. In model (23) let θ be 1 and σ^2 be 0.3. Finally in the event of replicated procedures, take $m = 2$. For each procedure, 100 simulations are carried out and the bias, root mean-squared error and root integrated mean-squared error computed as

$$\text{BIAS}[\hat{v})] = \frac{1}{100}\sum_{i=1}^{100}[\hat{v}_i - v]$$

$$\text{RMSE}[\hat{v}] = \sqrt{\frac{1}{100}\sum_{i=1}^{100}[\hat{v}_i - v]^2}$$

$$\text{RIMSE}[\hat{v}] = \sqrt{\frac{1}{100}\sum_{i=1}^{100}\left\{\frac{1}{20001}\sum_{j=1}^{20001} d_j^2\right\}_i}$$

Here, v indicates $\theta, \sigma^2, \beta_0$ and β_1 respectively, d_j is the difference between the fitted and the true function at the points $x_j = -1, \ldots, 1$ at intervals of 0.001 and subscript i denotes the i^{th} simulation. In the logarithm of absolute residuals procedure, the problem of the residuals near zero was avoided by discarding all the simulations which portrayed such a problem.

4. EMPIRICAL RESULTS:

The outputs are presented in tables one to six. Table one gives the averages, Table two gives the biases while Table three gives the root mean squared errors and Table four gives the root mean integrated squared errors for the various procedures under study. Table five gives the rankings of the procedures while table six shows the variations between the fitted and the true functions. The abbreviations PLH, RML, SR1, SR2, AR1, AR2, LAR, EQL, MML, and ROD correspond respectively to pseudo likelihood, restricted maximum likelihood, squared residuals with leverage, squared residuals without leverage, absolute residuals with leverage, absolute residuals without leverage, logarithm of absolute residuals, extended quasi likelihood, modified maximum likelihood and Rodbard.

	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$
PLH	0.83	0.32	0.02	0.95
RML	0.96	0.23	0.00	0.97
SR1	0.86	0.28	0.01	1.02
SR2	0.87	0.32	0.02	1.02
AR1	0.92	0.20	-0.01	1.01
AR2	0.98	0.23	0.02	1.02
LAR	1.14	0.38	0.01	0.99
EQL	1.16	0.27	-0.02	1.05
MML	0.93	0.49	0.01	0.97
ROD	1.15	0.34	0.02	0.96

Table 1. Average Estimates

	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$
PLH	-0.17	0.02	0.02	-0.05
RML	-0.04	-0.07	0.00	-0.03
SR1	-0.14	-0.02	0.01	0.02
SR2	-0.13	0.02	0.02	0.02
AR1	-0.08	-0.10	-0.01	0.01
AR2	-0.02	-0.07	0.02	0.02
LAR	0.14	0.08	0.01	-0.01
EQL	0.16	-0.03	-0.02	0.05
MML	-0.07	0.19	0.01	-0.03
ROD	0.15	0.04	0.02	-0.04

Table 2. Biases

	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$
PLH	0.522	0.119	0.153	0.325
RML	0.568	0.191	0.183	0.311
SR1	0.474	0.111	0.131	0.309
SR2	0.565	0.137	0.160	0.288
AR1	0.530	0.120	0.139	0.326
AR2	0.510	0.122	0.133	0.283
LAR	0.481	0.118	0.122	0.316
EQL	0.490	0.104	0.158	0.326
MML	0.455	0.228	0.111	0.233
ROD	0.489	0.124	0.136	0.310

Table 3. Root Mean Square Errors

PLH	RML	SR1	SR2	AR1	AR2	LAR	EQL	MML	ROD
0.202	0.326	0.217	0.221	0.274	0.247	0.374	0.207	0.303	0.294

Table 4. Root Mean Integrated Square Errors

		$\hat{\theta}$	
	BIAS	RMSE	RMISE
PLH	9	7	1
RML	2	10	9
SR1	6	2	3
SR2	5	9	4
AR1	4	8	6
AR2	1	6	5
LAR	6	3	10
EQL	8	5	2
MML	3	1	8
ROD	7	4	7

Table 5. Ranking

	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$
PHL	0.244	0.014	0.025	0.103
RML	0.321	0.032	0.033	0.096
SR1	0.205	0.012	0.017	0.095
SR2	0.302	0.018	0.025	0.083
AR1	0.275	0.005	0.019	0.106
AR2	0.260	0.010	0.017	0.080
LAR	0.212	0.008	0.015	0.100
EQL	0.215	0.010	0.025	0.104
MML	0.202	0.016	0.012	0.054
ROD	0.217	0.014	0.018	0.095

Table 6. Variances

5. DISCUSSION AND CONCLUSION:

With the exception of the logarithm of absolute residuals (LAR), extended quasi likelihood (EQL) and Rodbard (ROD) all the other procedures underestimate θ. Considering the absolute bias and ranking these procedures according to the three criteria, absolute bias, root mean squared error and root mean integrated squared error as in table 5 it is seen that absolute residuals with leverage corrected (AR2) does best. However for RMSE modified maximum likelihood (MML) seems to do well and for RMISE the best procedure is pseudo likelihood (PLH). Although the differences are small, squared residuals (SR1) and absolute residuals (AR2) are good at least for cases where replication is not available. Modified maximum likelihood (MML) could be recommended for replicated cases. Note that MML appear better because of the double sample size arising from the replication. Another observation is that the variances are larger than the biases for all the procedures as displayed in tables 2 and 6. No procedures are terrible.

REFERENCES:

1. Amemiya, T. (1977). A note on a heteroscedastic model. *Journal of Econometrics* 6, 365 – 370.
2. Breiman, L. and Meisel, W. S. (1976). General estimates of the intrinsic variability of data in nonlinear regression models. *Journal of the American Statistical Association* 71, 301 – 307.
3. Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *Annals of Statistics* 10, 1224 – 1233.
4. Carroll, R. J. and Ruppert, D. (1982a). A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. *Journal of the American Statistical Association* 77, 878 – 882.
5. Carroll, R. J. and Ruppert, D. (1982b). Robust estimation in heteroscedastic linear models. *Annals of Statistics* 10, 429 – 441.
6. Carroll, R. J. and Ruppert, D. (1985). Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association* 79, 321 – 328.
7. Carroll, R. J. and Ruppert, D. (1987). Diagnostics and robust estimation when transforming the regression model and the response. *Technometrics* 29, 287 – 299.

8. Carroll, R. J. and Ruppert, D. (1988). *Transformation and weighting in regression*. Chapman and Hall, London.
9. Cook, R. D. and Weisberg, S. (1992). *Residuals and Influence in Regression*. Chapman and Hall, London.
10. Davidian, M. (1990). Estimation of variance functions in assays with possibly unequal replication and nonnormal data. *Biometrika* 77, 43 – 54.
11. Davidian, M. and Carroll, R. J. (1988). An analysis of extended quasi – likelihood. *Journal of the Royal Statistical Society*. B50.
12. Dent, W. T. and Hildreth, C. (1977). Maximum likelihood estimation in random coefficient models. *Journal of the American Statistical Association* 72, 69 – 72.
13. Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. Biometrika 65, 457 – 481.
14. Egy, D. and Lahiri, K. (1979). On maximum likelihood estimation of functional forms and heteroscedasticity. *Econometrics letters* 2, 155 – 160.
15. Finney, D. J. (1964). *Statistical Methods in Biological Assay, London*: Griffin.
16. Firth, D. (1987). On the efficiency of quasi – likelihood estimation. *Biometrika* 74, 233 – 246.
17. Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Annals of Statistics* 6, 1149 – 1158.
18. Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Annals of Statistics* 9, 861 – 869.
19. Harrison, M. J. and McCabe, B. P. M. (1979). A test for Heteroscedasticity based on least squares residuals. *Journal of the American Statistical Association* 74, 494 – 500.
20. Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 79, 302 – 308.
21. Hastie, T. J. and Tibshirani, R. J. (1991). *Generalized additive models*. Chapman and Hall, London.
22. Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics* 40, 633 – 643.
23. Jobson, J. D. and Fuller, W. A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *Journal of the American Statistical Association* 75, 176 – 181.
24. Kibua. T. K. (2017). Estimation of variance in heteroscedastic data. *International Journal for Innovative Research in Multidisciplinary Field* 3, issue 10, 59 – 65.
25. Kibua. T. K. (1995). Variance function estimation. *Ph. D. Dissertation*, City University, London.
26. McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. Chapman and Hall, London.
27. Nelder, J. A. and Pregibon, D. (1987). An extended quasi likelihood function. *Biometrika* 74, 221 – 232.
28. Raab, G. M. (1981). Estimation of a variance function with application to radioimmunoassay. *Applied Statistics* 30, 32 – 40.
29. Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 77, 828 – 838.
30. Sadler, W. A. and Smith, M. H. (1985). Estimation of the response error relationship in immunoassay. *Clinical Chemistry* 31/11, 1802 – 1805.
31. Weisberg, S. (1985). *Applied Linear Regression*. Wisley, New York.