

RESIDUALS INFLUENCE AND WEIGHTING
IN ESTIMATION OF REGRESSION PARAMETERS

BY

OMBUI THOMAS MAGETO

A DISSERTATION SUBMITTED TO THE

KENYATTA UNIVERSITY, KENYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE

DEGREE OF MASTER OF SCIENCE [STATISTICS]

IN THE DEPARTMENT OF MATHEMATICS

JULY 1997

SUPERVISOR: DR. T. K. KIBUA

Ombui, Thomas Mageto
*Residuals influence
and weighting in*



2000/258305

KENYATTA UNIVERSITY LIBRARY

DEPARTMENT OF MATHEMATICS
KENYATTA UNIVERSITY
P. O. BOX 49844, NAIROBI

TO MY PARENTS

JOYCE KWAMBOKA

ANDREW OMBUI

CONTENTS

Table of Contents	1
List of Figures	3
List of Tables	4
Abstract	5
1 INTRODUCTION	6
1.1 Introduction.....	6
2 PARAMETRIC APPROACH	9
2.1 Model.....	9
2.2 Selection of Model.....	10
2.3 Estimation of parameter θ	21
2.3.1 Pseudo likelihood estimation of variance functions.....	21
2.3.2 Restricted maximum likelihood.....	24
2.3.3 Least squares on squared residuals.....	26
2.3.4 Least squares on absolute residuals.....	28
2.3.5 Modified maximum likelihood.....	29
2.3.6 Extended quasi likelihood.....	33
2.3.7 Logarithm of absolute residuals.....	34
2.3.8 Rodbard and Frazier.....	35
2.3.9 Maximum likelihood.....	35
2.3.10 Sadler and Smith.....	36
3 EMPIRICAL STUDY	38
3.1 Introduction.....	38
3.2 A Study data.....	39
3.3 Working models.....	40
3.3.1 Parametric estimation of β	47
3.3.2 Pseudo likelihood estimation of parameter θ	51

3.3.3 Restricted maximum likelihood estimation of θ	51
3.3.4 Least squares on squared residuals estimation of θ	51
3.3.5 Least squares on absolute residuals estimation of θ	53
3.3.6 Modified maximum likelihood equation for $\hat{\theta}$	54
3.3.7 Extended quasi likelihood estimation of θ	56
3.3.8 Rodbard and Frazier.....	57
3.3.9 Sadler and Smith.....	57
3.3.10 Algorithm.....	57

4 NONPARAMETRIC APPROACH 60

4.1 Introduction.....	60
4.2 Kernel smoothing.....	61
4.3 Example.....	63
4.4 Conclusion.....	75

BIBLIOGRAPHY 76

LIST OF FIGURES

2.1a	Null plot.....	12
2.1b	Right-opening megaphone.....	13
2.1c	Left-opening megaphone.....	14
2.1d	Double outward bow.....	15
2.1e	Nonlinearity.....	16
2.1f	Nonlinearity.....	17
2.1g	Nonlinearity and nonconstant variance.....	18
2.1h	Nonlinearity and nonconstant variance.....	19
3.1a	Scatter plot of study data using one set of replicate.....	41
3.1b	Line plot of study data involving both replications.....	42
3.2	Scatter plot of residuals and predictors of study data.....	46
3.3	Algorithm flow diagram.....	58
4.1	A plot of Yield against time.....	70
4.2	A plot of residuals against time	71
4.3	A plot of parametric variance estimates against time.....	72
4.4	A scatter plot of squared residuals against time.....	73
4.5	A smooth curve of nonparametric variance estimates.....	74

LIST OF TABLES

3.1	A study data.....	39
3.2	Predictor values and residuals of study data.....	44
3.3	Results of analysis on the study data.....	59
4.1	Variables and estimates for Yield data.....	63

ABSTRACT

We consider analysis and fitting models of the regression data in the fields which exhibit nonconstant variances which are referred to as regression models. We focus on various approaches and procedures used in estimating the variances in such models as a way of estimating the regression parameters. Chapter 1 comprise the introduction to the subject. In the first part, chapter 2 and 3 purely discusses the parametric approach which is commonly used due to its outstanding features of simplicity in computation, compatibility with model assumptions and for its mathematical convenience. The procedures are fully formulated in chapter 2 and the empirical study using the same procedures is covered in chapter 3.

In the second part, chapter 4 discusses nonparametric method. The central problems of interest are the choice of the smoothing methods, choice of the Kernel and bandwidth. In chapter 4 we illustrate both parametric and nonparametric methods in a practical situation. A contrast and the conclusion has been done in the same chapter.

All the computation has been done in Splus programming language. Table formats and other organization matters are comfortably done in Microsoft office [Word] while graphics, figure representation and analysis are computer drawn in Microsoft office [Excel].

Chapter 1

INTRODUCTION

1.1 Introduction

Regression analysis is a statistical technique for investigating the relationship between variables. Application of regression are numerous and occur in almost every field, including engineering, the physical sciences, economics, management, biological sciences and the social sciences. In fact, regression analysis may be the most widely used statistical technique. It provides a technique for establishing a functional relationship between an N -vector variable $Y = (y_1, y_2, \dots, y_N)'$ and a p -variate variable $X = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p)$. Where $\underline{x}_1 = (x_{11}, x_{21}, \dots, x_{N1})'$, $\underline{x}_2 = (x_{12}, x_{22}, \dots, x_{N2})'$, \dots , $\underline{x}_p = (x_{1p}, x_{2p}, \dots, x_{Np})'$, x_{ij} is the element in the i th row and j th column. The relationship is expressed in a mathematical model form connecting the response Y and one or more predictor variables $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p$.

Consider a general multiple linear regression model, where more than one regressor is involved. Then

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i \quad i = 1, 2, \dots, N \quad (1.1)$$

Where $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ is the p -vector of unknown regression parameter, e_i are the error

variables with mean zero and variance σ_i^2 . The equation (1.1) is often a theoretical model, but it may also be an empirical model that seems to work well in practice especially in a linear regression model. Once we have determined β then the system will be completely specified. The salient problem of regression analysis is to estimate the unknown parameter $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ in the regression model. In most cases the observed data never fit a model exactly, hence β cannot be determined exactly. To take care of such situation, we observe the response Y on the predictor X only with error.

Over the years the primary interest has been the inference about the unknown regression parameter $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$. Classical methods such as the least squares, weighted least squares, maximum likelihood and Ridge regression estimation have been widely used. Least squares method has been used over the years. When the form of the distribution of the errors is not known, the method of maximum likelihood can be applied. Weisberg [1980], referred to this method as weighted least squares. Some assumptions made in the classical methods are that, the model is correctly specified on average, the errors are normal, identical and independently distributed. However if any of these assumptions is violated the estimates obtained under the classical assumptions are not good. We hope to obtain better estimate of $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$, if we investigate and incorporate the information about the variance estimates of the errors as they are needed for better understanding of the variability of the data.

In homoscedastic regression models, the variance is constant. However, in heteroscedastic regression models, the variance is not constant. Often, as in the case with the

mean, the heteroscedasticity is believed to be in functional form which is referred to as the variance function. We try to understand the structure of the variances as a function of predictors.

There are two methods of estimating the variance function. The parametric method and the nonparametric method. Parametric variance function estimation may be defined as a type of regression problem in which we visualise variance as a function of estimable quantities. Thus the heteroscedasticity is modelled as a function of the regression and other structural parameters. This function is therefore completely known, specified up to these unknown parameters. Estimation of these parameters is what entails the parametric method.

In nonparametric method, the heteroscedasticity is regarded as of unknown form. Thus the variance function is completely unknown but assumed to be smooth. We will be interested in studying both methods for estimating variances in heteroscedastic regression models.

$$Var(Y) = \sigma^2 g(x, \theta) \quad (2.2)$$

This is a general parametric model for the variance. Here, σ^2 is an unknown scale

Chapter 2

PARAMETRIC APPROACH

2.1 Model

Linear models play an important role in both applied and theoretical work. The behaviour of the model should be consistent with known physical, biological or mathematical laws as far as possible especially its limiting behaviour. We have already considered the mean linear model (1.1). To solve the problem of heterogeneity we consider a heteroscedastic regression model for observable data Y :

$$\text{Var}(y_i) = \sigma^2 g^2(z_i, \mu_i(\beta), \theta) \quad (2.1)$$

This is a general parametric model for the variance. Here, σ is an unknown scale parameter, θ is unknown $r \times 1$ vector of parameters, g is the variance function expressing the heteroscedasticity, $\{z_i\}$ are the known vectors, possibly $\{x_i\}$ and $\mu_i(\beta)$ is the mean.

If $z_i = x_i$, it implies that the variance is a function of the predictors. The variance can also depend on the known mean $\mu_i(\beta)$ or on the estimated mean response $\mu_i(\hat{\beta})$. We estimate $g^2(z_i, \mu_i(\beta), \theta)$ using an estimate of θ . This approach of unknown parameter model is flexible unlike the model with a constant coefficient of variation which requires a prior

assumption about the mean function. In our study, g is taken to be known and to satisfy appropriate smoothness conditions. The problem of estimating model (2.1) is solved once an estimate of θ is obtained and σ is obtained once θ and β are known.

Popular methods in practice which are widely used to estimate θ include:

- (i) Maximum likelihood
- (ii) Pseudo likelihood
- (iii) Weighted squared residuals
- (iv) Weighted absolute residuals
- (v) Logarithm method
- (vi) Restricted maximum likelihood
- (vii) Rodbard and Frazier
- (viii) Sadler and Smith
- (ix) Modified maximum likelihood
- (x) Extended quasi likelihood

2.2 Selection of model

Graphical techniques are employed for detecting and understanding heteroscedastic variability which will simplify the problem of deciding on the nature of the model to be fitted. The most widely used diagnostic for heterogeneity is the unweighted least squares residual plot.

A residual is defined as the algebraic difference between an observation y_i and the predicted $\hat{y}_i = f(x_i, \hat{\beta})$. Predicted values $\hat{y}_i = f(x_i, \hat{\beta})$ from an unweighted fit are plotted along the horizontal axis, while the residuals from the fit $(y_i - \hat{y}_i)$ are plotted along the vertical axis. Residuals carry important information concerning the appropriateness assumptions. Analysis may include informal graphics to display general features of the residuals as well as the formal tests to detect specific departures from underlying assumptions. Such formal and informal procedures are complementary and both have a place in residual analysis. Let us denote the ordinary residuals by r_i , then

$$r_i = y_i - \hat{y}_i, \quad i=1, 2, \dots, N \quad (2.2)$$

Other types of residuals can be obtained by making several transformations over the ordinary residuals. They include studentized residuals in which we obtain a set of residuals that have null distributions that are independent of the scale parameters. We also have predicted residuals, uncorrelated residuals and recursive residuals. While it would be straight forward to consider these residuals, it would be sufficient to employ the ordinary residuals as they have a simple structure that can be calculated without difficulty. It is essential to study a variety of plots or graphs of the original data (the independent variables x_1, x_2, \dots, x_p and the dependent variable Y), derived statistics (especially the ordinary residuals r_i and the fitted values \hat{y}_i) and occasionally, other variables such as time or case number. Apart from using these plots for statistical analysis, they also contribute in providing an easy way of understanding a complex problem and allowing an examination of the data as an aggregate while clearly displaying the behaviour of individual cases.

In simple regression, the plot of residuals r_i verses \hat{y}_i is adequate to provide relevant information without use of extra information such as time or spatial ordering of cases. In residual plotting, we are primarily interested in the shape of the plot, and often the values on the axes are not important. We have quite a variety of patterns as shown in Figure 2.1.

A plot of residuals against the corresponding values \hat{y}_i is helpful for identifying and revealing several common types of mean model inadequacies. In Figure 2.1a, the residuals are contained in a horizontal band, then it implies that the mean model so far used is correct, in other words there are no obvious mean model defects. The patterns in Figures 2.1b, 2.1c, 2.1d, 2.1e, 2.1f, 2.1g and 2.1h actually shows mean model deficiencies. They indicate that the residual variance is a systematic function of the quantity plotted on the horizontal axis. In

Figure 2.1a: Null plot

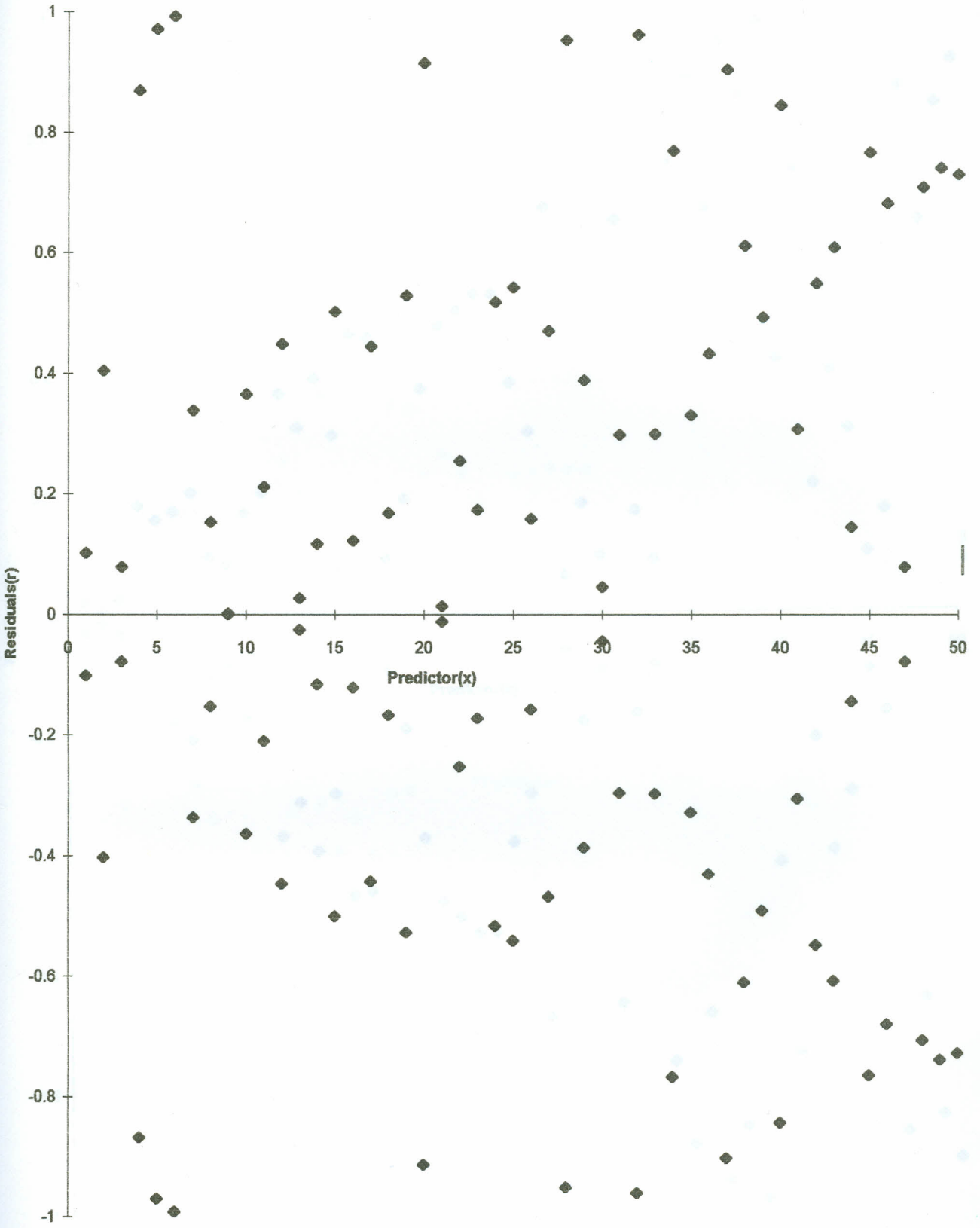


Figure 2.1b: Right-opening megaphone

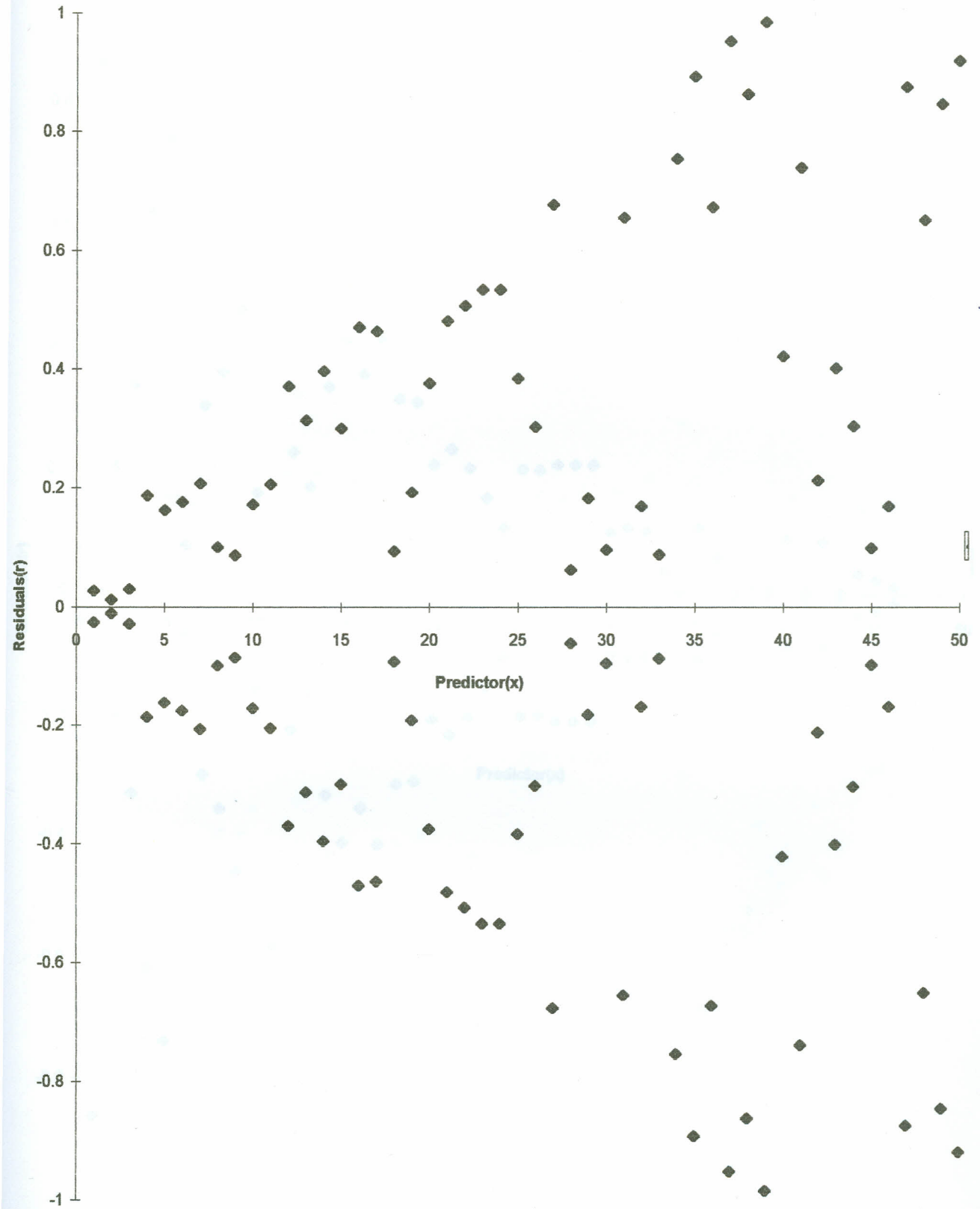


Figure 2.1c: Left-opening megaphone

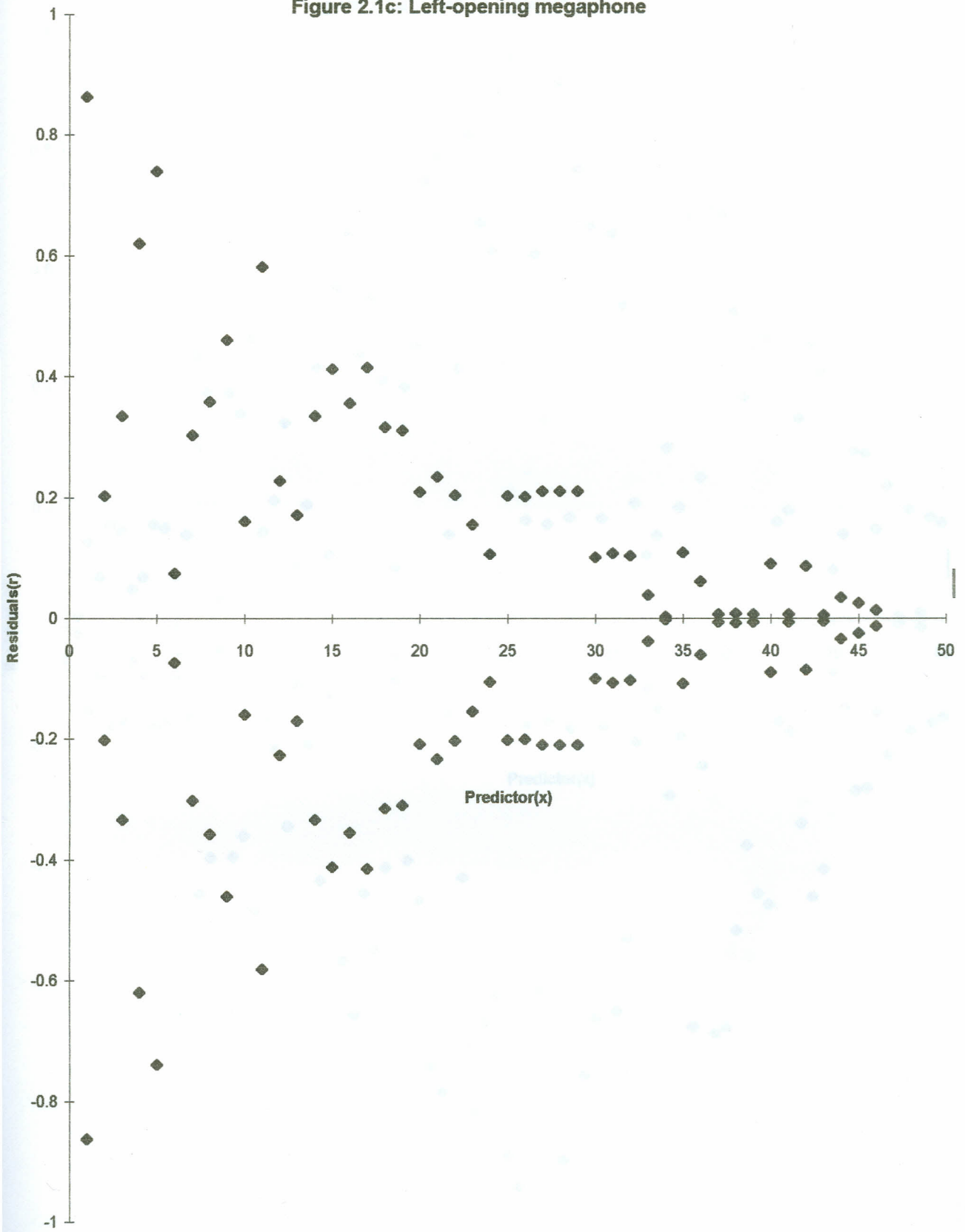


Figure 2.1d: Double outward bow

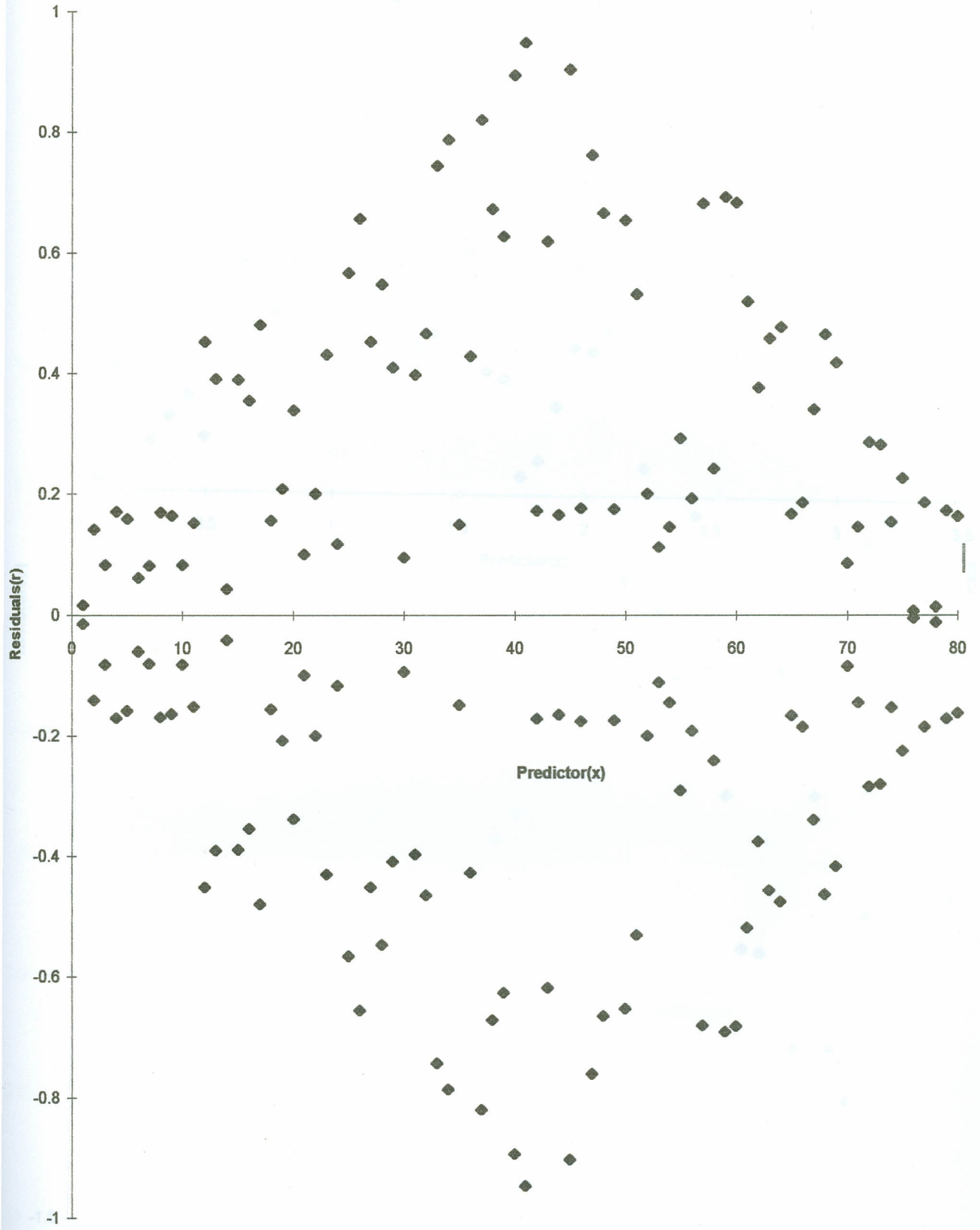


Figure 2.1e: Nonlinearity

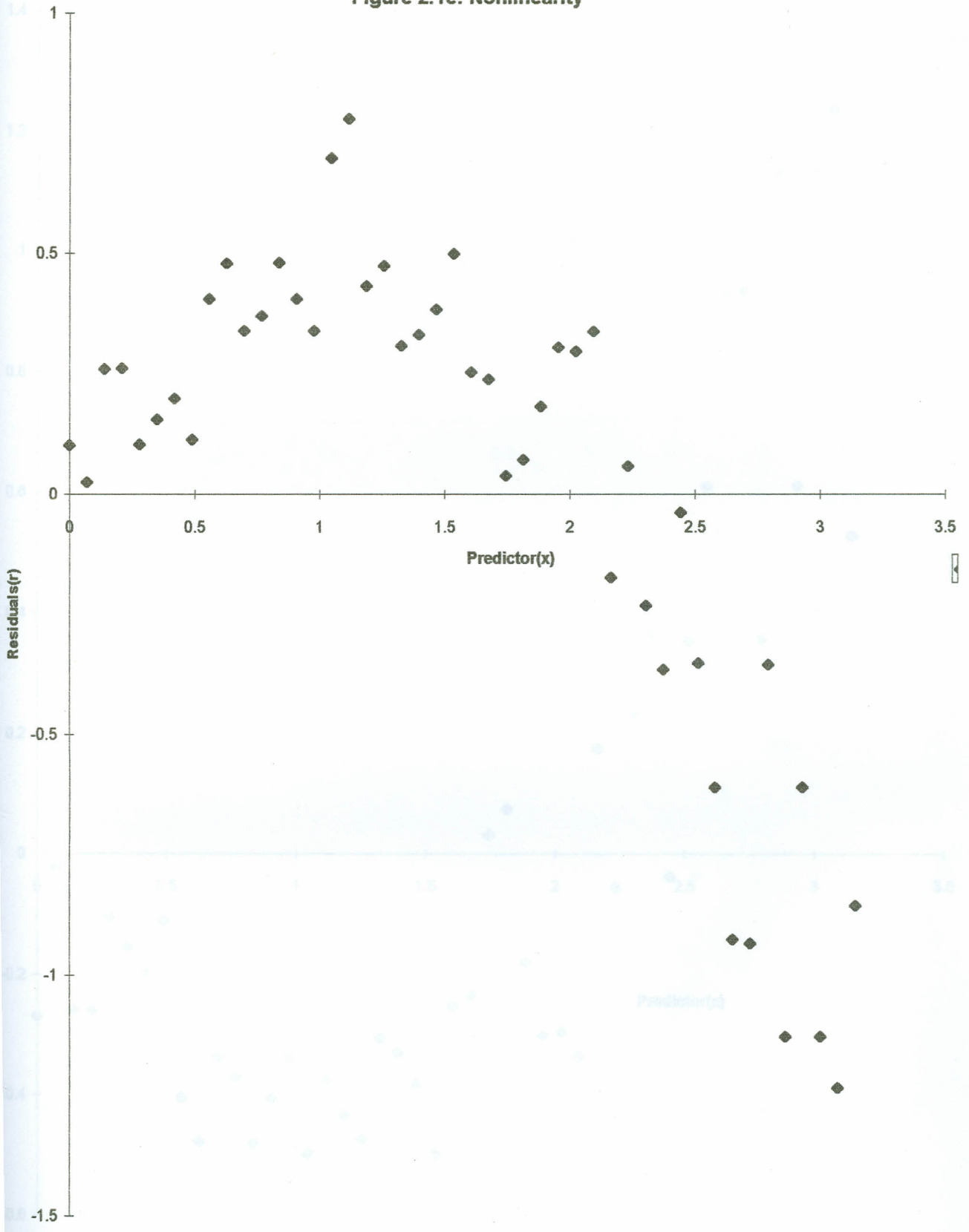


Figure 2.1f: Nonlinearity

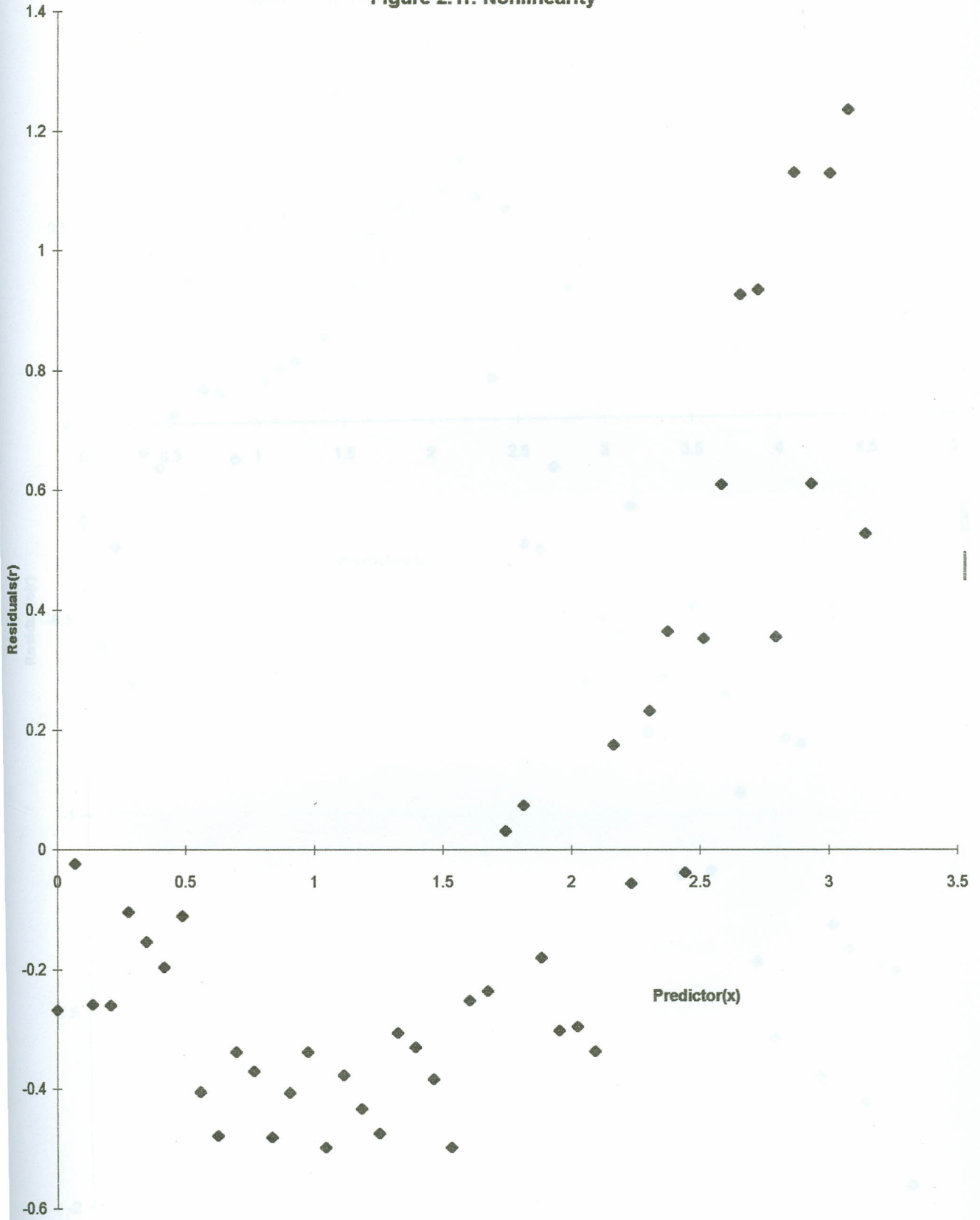


Figure 2.1g: Nonlinearity and nonconstant variance

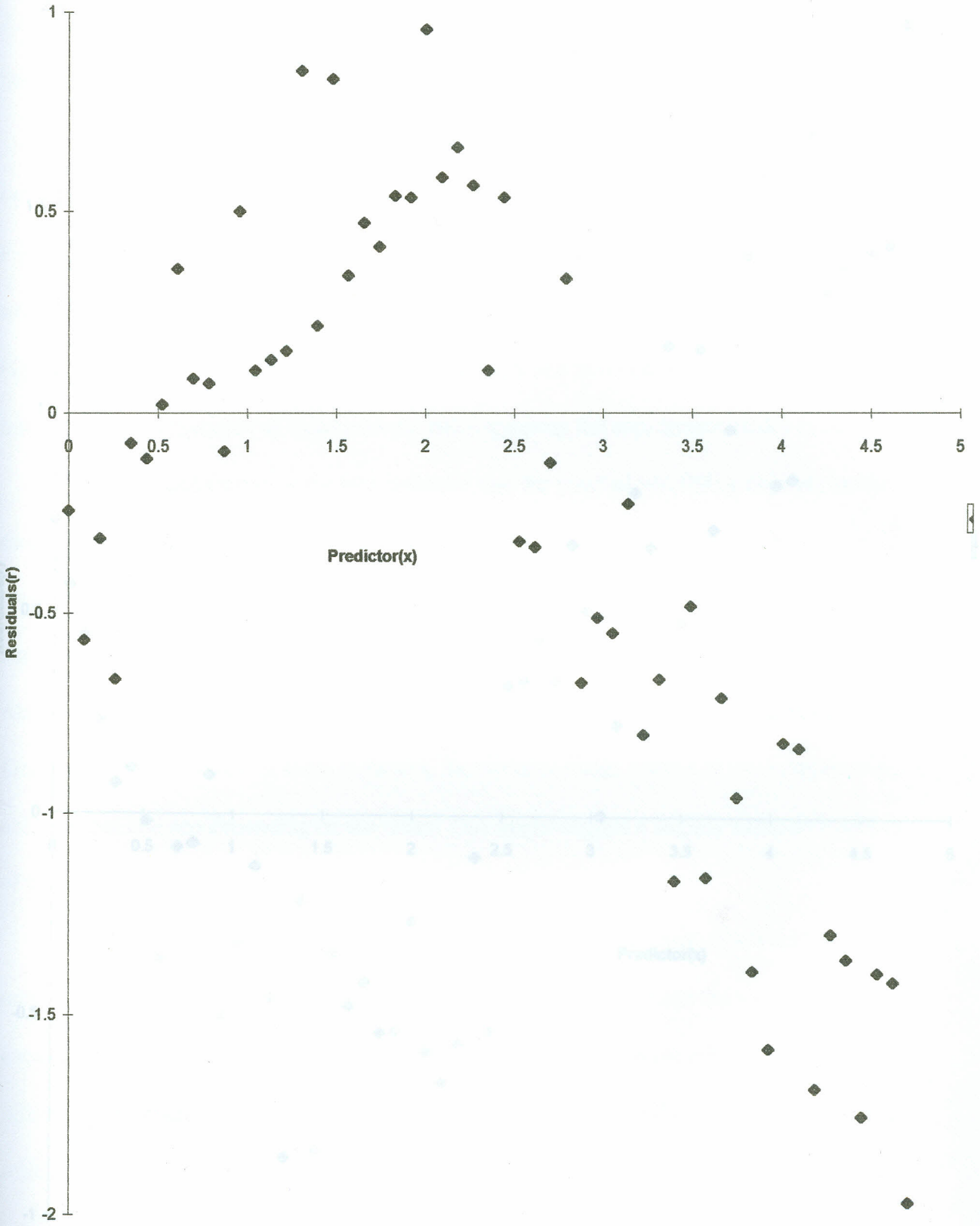
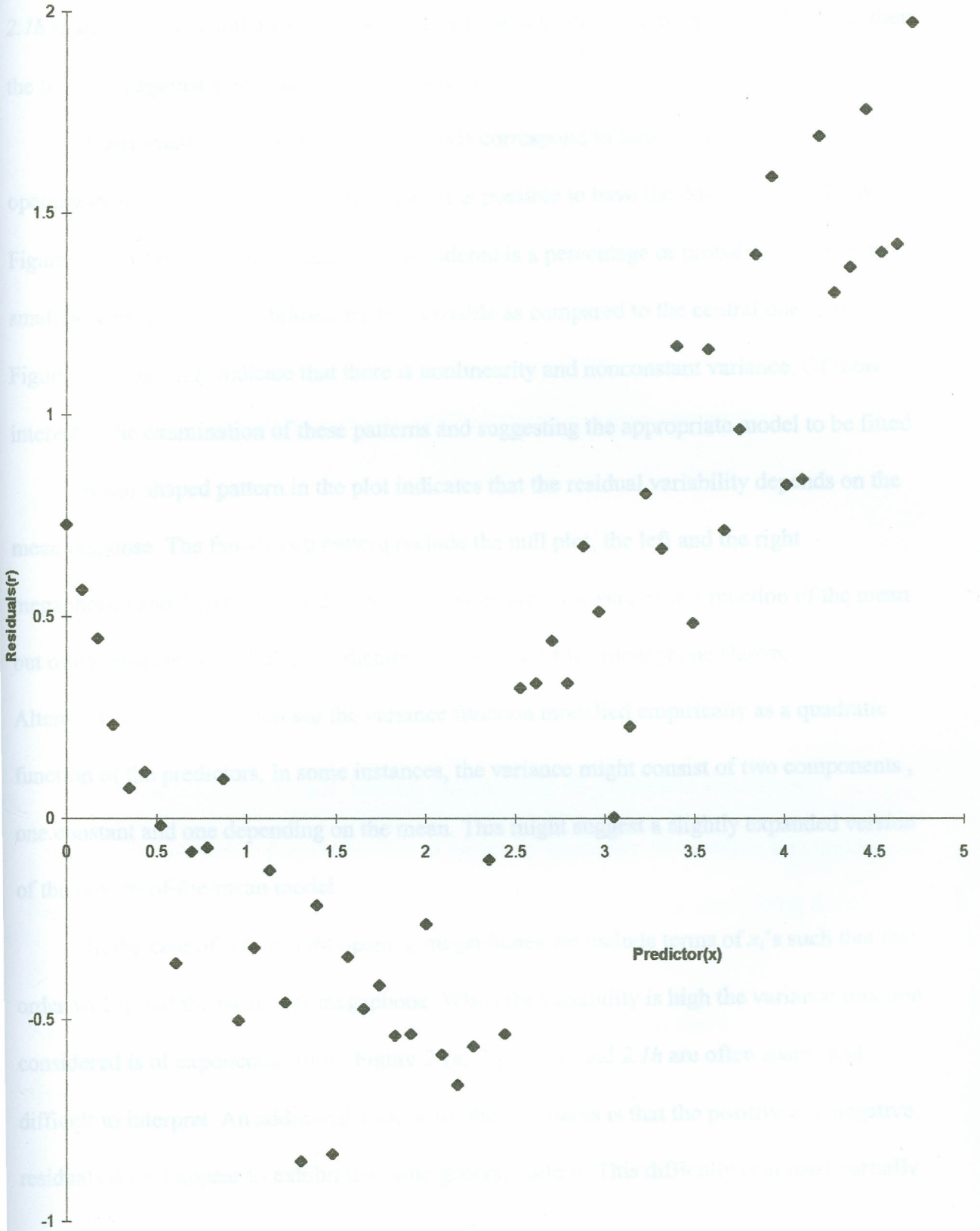


Figure 2.1h: Nonlinearity and nonconstant variance



particular for the situation described above, the right-opening megaphone pattern of Figure 2.1b indicates that residual variance is strongly related to any good predictor of Y , say X , then the plot of r_i against x_i should reveal this as well.

When small values of the horizontal axis correspond to large variability then a left-opening megaphone, Figure 2.1c will occur. It is possible to have the double outward bow, Figure 2.1d, which arise if the response considered is a percentage or probability. Large or small percentages or probabilities are less variable as compared to the central one. The Figures 2.1e and 2.1f indicate that there is nonlinearity and nonconstant variance. Of more interest is the examination of these patterns and suggesting the appropriate model to be fitted.

A fan-shaped pattern in the plot indicates that the residual variability depends on the mean response. The fan-shaped pattern include the null plot, the left and the right megaphones and double outward bow. So far what we are aware of is a function of the mean but other terms to be included are dictated by the type of the megaphone shown.

Alternatively, one will often see the variance function modelled empirically as a quadratic function of the predictors. In some instances, the variance might consist of two components, one constant and one depending on the mean. This might suggest a slightly expanded version of the power- of-the-mean model.

In the case of left or right opening megaphones we include terms of x_i 's such that the order will reveal the particular megaphone. When the variability is high the variance function considered is of exponential form. Figure 2.1e, 2.1f, 2.1g, and 2.1h are often sparse and difficult to interpret. An additional feature for these patterns is that the positive and negative residuals do not appear to exhibit the same general pattern. This difficulty is at least partially

removed by plotting squared residuals verses predictors or predicted values and thus visually doubling the sample size. The squaring will cause scaling problems in particular for problems with moderately large residuals and in such instances we use the absolute residuals.

There is no additional assumption made about the distribution of the y_i other than that of the form of the first two moments. The key point of this discussion is that when heterogeneity of the variance is present, how well one models and estimates the variances will have substantial impact on the prediction and calibration based on the estimated mean response since the form of the intervals depends on the form of the variance function. Thus misspecification of the variance model could cause some serious catastrophe.

2.3 Estimation of parameter θ

The proposed possible procedures for estimating the parameter θ will be discussed fully in this section.

2.3.1 Pseudo likelihood estimation of variance functions

In the method of Pseudo likelihood we do not make any distribution assumptions but basically rely on the mean model (1.1) and variance model (2.1). In determining the estimates of θ , we assume that the regression parameter β is known or fixed.

Assuming normality conditions hold, we estimate θ by maximum likelihood.

Consider the density function of a normal distribution

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

The likelihood for the normal distribution is then

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

We can write this as

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

Taking logarithms of both sides we obtain

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

Let $l = \log L(\beta, \theta, \sigma^2/Y)$

Differentiate the log-likelihood with respect to β and equate to zero.

Thus maximizing in β we get

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

for given estimates of θ and σ .

But $\frac{\delta l}{\delta \beta} = 0$

Then

$$\sum_{i=1}^N \left[\frac{[y_i - f(x_i, \beta)] \frac{\delta}{\delta \beta} f(x_i, \beta)}{\hat{\sigma}^2 g^2(z_i, \mu_i(\beta), \hat{\theta})} \right] = 0 \quad (2.3)$$

Write the weighted residuals as

$$R_i = \frac{y_i - f(x_i, \hat{\beta})}{g(z_i, \mu_i(\hat{\beta}), \theta)}$$

Then

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

Take derivatives with respect to σ^2 and equate to zero

Thus

$$\frac{\delta l}{\delta \sigma^2} = -\frac{N}{2\sigma^2} + \sum_{i=1}^N \left[\frac{R_i^2}{2\sigma^4} \right]$$

But $\frac{\delta l}{\delta \sigma^2} = 0$

Then

$$-\frac{N}{2\sigma^2} + \sum_{i=1}^N \left[\frac{R_i^2}{2\sigma^4} \right] = 0$$

And

$$\hat{\sigma}^2 = \sum_{i=1}^N \left[\frac{R_i^2}{N} \right] \quad (2.4)$$

Now we differentiate l' with respect to θ and equate to zero

Thus

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

But $\frac{\delta l'}{\delta \theta} = 0$

Then

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

Let

$$\frac{\frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{g(z_i, \mu_i(\hat{\beta}), \theta)} = S\theta_i$$

Giving us

$$-NS\theta_i + \sum_{i=1}^N \left[\frac{R_i^2 S\theta_i}{\hat{\sigma}^2} \right] = 0$$

Then

$$-N\hat{\sigma}^2 S\theta_i + \sum_{i=1}^N [R_i^2 S\theta_i] = 0$$

Which can be written as

$$\sum_{i=1}^N [(R_i^2 - \hat{\sigma}^2) S\theta_i] = 0 \quad (2.5)$$

We use equation (2.5) to solve for $\hat{\theta}$. Note that $\hat{\theta}$ cannot be expressed explicitly and hence an algorithm process is employed.

2.3.2 Restricted maximum likelihood

Evidently from the discussion in section (2.3.1) the Pseudo-likelihood is less efficient as it takes no account of the loss in degrees of freedom that results from estimating β . We can adjust this by substituting the denominator N by $N-p$ in equation (2.4). Where p is the number of regression parameters.

Thus we write

$$\hat{\sigma}^2 = \sum_{i=1}^N \left[\frac{R_i^2}{N-p} \right] \quad (2.6)$$

Further adjustment is needed to account for leverage. In linear regression models the well-known measure of leverage is given by the diagonal elements h_{ii} of the "hat" matrix

$$H = Q(Q^T Q)^{-1} Q^T \quad (2.7)$$

Where Q is the $N \times p$ matrix whose i th row is given by

$$\frac{f_{\beta}(x_i, \beta)'}{g(z_i, \mu_i(\hat{\beta}), \hat{\theta})}$$

And

$$f_{\beta}(x_i, \beta) = \frac{\delta}{\delta \beta} \{f(x_i, \beta)\}$$

From equation (2.5), we write

$$\sum_{i=1}^N R_i^2 S\theta_i = \sum_{i=1}^N \hat{\sigma}^2 S\theta_i (1 - h_{ii})$$

Which can be written as

$$\sum_{i=1}^N \frac{[y_i - f(x_i, \hat{\beta})]^2}{\hat{\sigma}^2 g^2(z_i, \mu_i(\hat{\beta}), \theta)} S\theta_i = \sum_{i=1}^N S\theta_i (1 - h_{ii})$$

Simplifying to

$$\sum_{i=1}^N R_i^2 S\theta_i - \hat{\sigma}^2 \sum_{i=1}^N S\theta_i (1 - h_{ii}) = 0 \quad (2.8)$$

2.3.3 Least squares on squared residuals

The residuals are defined as

$$y_i - f(x_i, \hat{\beta}_*)$$

Where $\hat{\beta}_*$ is the current estimate of β . The procedure is based on the squared residuals

$$[y_i - f(x_i, \hat{\beta}_*)]^2$$

The criterion is of more help for a nonlinear regression problem where the responses are squared residuals. The motivating idea for this procedure is that the expectation of squared residuals is approximately the variance.

Thus

$$E[y_i - f(x_i, \hat{\beta}_*)]^2 \cong \sigma^2 g^2(z_i, \mu_i(\hat{\beta}_*), \theta)$$

Next we determine the estimate of θ by minimizing in θ and σ the expression

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

We realize that for a normal distribution the squared residuals have a variance proportional to

$$\sigma^4 g^4(z_i, \mu_i(\beta), \theta)$$

The unweighted least-squares estimates have the expectation of squared residuals different from variance due to least squares estimate of β by a factor depending on the i th leverage value from the unweighted least squares fit. With this, we resort to the choice of the weighted least squares. We therefore Minimize σ and θ expression

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

where $\hat{\theta}_*$ is a preliminary estimate of θ .

Differentiating with respect to θ and equating to zero we get

$$\sum_{i=1}^N \frac{[y_i - f(x_i, \hat{\beta}_*)]^2 - \hat{\sigma}^2 g^2(z_i, \mu_i(\hat{\beta}_*), \theta)}{g^4(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} g(z_i, \mu_i(\hat{\beta}_*), \theta) \frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}_*), \theta) = 0 \quad (2.9)$$

and differentiating with respect to σ and equating to zero we obtain

$$\sum_{i=1}^N \frac{[y_i - f(x_i, \hat{\beta}_*)]^2 - \sigma^2 g^2(z_i, \mu_i(\hat{\beta}_*), \theta)}{g^4(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} \sigma g^2(z_i, \mu_i(\hat{\beta}_*), \theta) = 0 \quad (9.10)$$

Solve (2.9) and (2.10) to obtain $\hat{\theta}$ and $\hat{\sigma}$ respectively.

To account for leverage, we minimize in θ and σ the expression

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

Thus differentiating with respect to θ and equating to zero gives

$$\sum_{i=1}^N \frac{[y_i - f(x_i, \hat{\beta}_*)]^2 - \hat{\sigma}^2 (1 - \hat{h}_{ii}) g^2(z_i, \mu_i(\hat{\beta}_*), \theta)}{(1 - \hat{h}_{ii})^2 g^4(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} g(z_i, \mu_i(\hat{\beta}_*), \theta) \frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}_*), \theta) = 0 \quad (2.11)$$

and for σ we obtain

$$\sum_{i=1}^N \left[\frac{[y_i - f(x_i, \hat{\beta}_*)]^2 - \sigma^2 (1 - \hat{h}_{ii}) g^2(z_i, \mu_i(\hat{\beta}_*), \hat{\theta})}{(1 - \hat{h}_{ii})^2 g^4(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} \sigma g^2(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}) \right] = 0 \quad (2.12)$$

The estimates of θ and σ are obtained through solving equations (2.11) and (2.12)

respectively.

2.3.4 Least squares on absolute residuals

Squared residuals are hard to interpret as they are long tailed. What this means is that wild outliers are noticeable if we employ the plotting technique. This has led many authors to propose using absolute residuals to estimate θ . Assume that

$$E|y_i - f(x_i, \hat{\beta})| = \sigma g(z_i, \mu_i(\hat{\beta}), \theta)$$

This condition is only true if and only if the errors are identically and independently distributed. One obtains the estimator $\hat{\theta}$ by minimizing in θ and σ the expression

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

For consistency it will be appropriate to use the weighted version by minimizing the expression

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

Where $\hat{\theta}_*$ is a preliminary estimate of θ . Minimizing in θ we obtain

$$\sum_{i=1}^N \frac{\left\{ |y_i - f(x_i, \hat{\beta}_*)| - \hat{\sigma} g(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}) \right\} \frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}_*), \hat{\theta})}{g^2(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} = 0 \quad (2.13)$$

and minimizing in σ we get

$$\sum_{i=1}^N \frac{\left\{ |y_i - f(x_i, \hat{\beta}_*)| - \hat{\sigma} g(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}) \right\} g(z_i, \mu_i(\hat{\beta}_*), \hat{\theta})}{g^2(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} = 0 \quad (2.14)$$

The estimate of σ from equation (2.14) is used to solve for $\hat{\theta}$ in equation (2.13). We presumably could modify this approach to account for leverage by minimizing in θ and σ the expression

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

Thus differentiating with respect to σ and equating to zero we get

$$\sum_{i=1}^N \frac{\left\{ \left| y_i - f(x_i, \hat{\beta}_*) \right| - \sigma(1 - \hat{h}_{ii})g(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}) \right\} g(z_i, \mu_i(\hat{\beta}_*), \hat{\theta})}{(1 - \hat{h}_{ii})g^2(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} = 0 \quad (2.15)$$

and for θ we obtain

$$\sum_{i=1}^N \frac{\left\{ \left| y_i - f(x_i, \hat{\beta}_*) \right| - \hat{\sigma}(1 - \hat{h}_{ii})g(z_i, \mu_i(\hat{\beta}_*), \theta) \right\} \frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}_*), \theta)}{(1 - \hat{h}_{ii})g^2(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} = 0 \quad (2.16)$$

2.3.5 Modified maximum likelihood

In many assay problems the experiments replicate the response at each value of the predictor. The predictor levels are (x_1, x_2, \dots, x_N) and at each value x_i we observe n_i replicated responses $(y_{1i}, y_{2i}), i = 1, 2, \dots, N$. This phenomenon is made clear if we assume that the scheme is similar to obtaining the samples from a large population of size N . From this population we select the samples of sizes n_i , the possible number of samples is M . For equal samples $N = nM$. Let y_{ij} be the j th unit from i th sample while \bar{y}_i is the sample mean at predictor x_i . Where $i = 1, 2, \dots, M, j = 1, 2, \dots, n_i$. The usual maximum likelihood estimate of σ is biased, but of course it can be made unbiased by dividing the corrected sum of squares

by the degrees of freedom rather than the sample size. Simply we replace the term $\sigma^{-n_i/2}$ by $\sigma^{-(n_i-1)/2}$. Further the information from all possible samples is necessary, it is instructive to employ pooling technique. In achieving the end estimate of θ we have to use the modified maximum likelihood expression. we make the simplifying assumption that the number of replicates is constant across groups.

Then

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

Taking logarithms both sides we get

$$l = -\left(\frac{n-1}{2}\right) \sum_{i=1}^M \log \left[2\pi\sigma^2 g^2(z_i, \mu_i(\beta), \theta) \right] - \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \mu_i(\beta)]^2}{2\sigma^2 g^2(z_i, \mu_i(\beta), \theta)} \right]$$

Differentiate with respect to σ^2 and equate to zero to obtain

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

But $\frac{\partial l}{\partial \sigma^2} = 0$

Thus

$$-\left(\frac{n-1}{2}\right) \sum_{i=1}^M \frac{1}{\sigma^2} + \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \mu_i(\hat{\beta})]^2}{2\sigma^4 g^2(z_i, \mu_i(\hat{\beta}), \hat{\theta})} \right] = 0$$

Rewrite this as

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

and gives

$$\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^M (n-1)} \left[\sum_{i=1}^M \sum_{j=1}^n \frac{[y_{ij} - \mu_i(\hat{\beta})]^2}{g^2(z_i, \mu_i(\hat{\beta}), \hat{\theta})} \right] \quad (2.17)$$

If we replace $\mu_i(\beta)$ by the sample mean, then (2.17) becomes

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^n \left[\frac{(y_{ij} - \bar{y}_i)^2}{g^2(z_i, \mu_i(\hat{\beta}), \hat{\theta})} \right] \quad (2.18)$$

Note that, often the number of replicates is rather small, in such a case the resulting weighted least-squares estimator can be a disaster. However, with no other information available, it is not surprising to estimate β using unweighted least squares. We now proceed to obtain

$$\frac{\delta l}{\delta \theta} = -(n-1) \sum_{i=1}^M \left[\frac{\frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{g(z_i, \mu_i(\hat{\beta}), \theta)} \right] + \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \mu_i(\hat{\beta})]^2 \frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{\hat{\sigma}^2 g^3(z_i, \mu_i(\hat{\beta}), \theta)} \right]$$

But $\frac{\delta l}{\delta \theta} = 0$

Giving us

$$\sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \mu_i(\hat{\beta})]^2 \frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{\hat{\sigma}^2 g^3(z_i, \mu_i(\hat{\beta}), \theta)} \right] - (n-1) \sum_{i=1}^M \left[\frac{\frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{g(z_i, \mu_i(\hat{\beta}), \theta)} \right] = 0 \quad (2.19)$$

The modified maximum likelihood estimate for β is through the expression

$$l = -\left(\frac{n-1}{2}\right) \sum_{i=1}^M \log(2\pi) - \left(\frac{n-1}{2}\right) \sum_{i=1}^M \log(\sigma^2) - \left(\frac{n-1}{2}\right) \sum_{i=1}^M \log\left(g^2(z_i, \mu_i(\beta), \theta)\right) \\ - \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \mu_i(\beta)]^2}{2\sigma^2 g^2(z_i, \mu_i(\beta), \theta)} \right]$$

Thus

$$\frac{\delta l}{\delta \beta} = -(n-1) \sum_{i=1}^M \left[\frac{\frac{\delta}{\delta \beta} g(z_i, \mu_i(\beta), \hat{\theta})}{g(z_i, \mu_i(\beta), \hat{\theta})} \right] + \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \mu_i(\beta)] \frac{\delta}{\delta \beta} [y_{ij} - \mu_i(\beta)]}{\hat{\sigma}^2 g^2(z_i, \mu_i(\beta), \hat{\theta})} \right] \\ + \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \mu_i(\beta)]^2 \frac{\delta}{\delta \beta} g(z_i, \mu_i(\beta), \hat{\theta})}{\hat{\sigma}^2 g^3(z_i, \mu_i(\beta), \hat{\theta})} \right]$$

But $\frac{\delta l}{\delta \beta} = 0$

Then

$$\sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \mu_i(\beta)]^2 \frac{\delta}{\delta \beta} g(z_i, \mu_i(\beta), \hat{\theta})}{\hat{\sigma}^2 g^3(z_i, \mu_i(\beta), \hat{\theta})} \right] + \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \mu_i(\beta)] \frac{\delta}{\delta \beta} [y_{ij} - \mu_i(\beta)]}{\hat{\sigma}^2 g^2(z_i, \mu_i(\beta), \hat{\theta})} \right] \\ - \sum_{i=1}^M \left[\frac{\frac{\delta}{\delta \beta} g(z_i, \mu_i(\beta), \hat{\theta})}{g(z_i, \mu_i(\beta), \hat{\theta})} \right] = 0 \quad (2.20)$$

The estimates σ , θ and β are easily determined on solving the equations (2.18), (2.19) and (2.20) respectively.

2.3.6 Extended quasi likelihood

In the family of distributions where we have additional properties, for instance homoscedastic normal, gamma, and inverse Gaussian distributions as special cases the idea of extended quasi likelihood procedure will be of most importance. For given θ and σ , β is to be estimated by generalized least squares, while for given β and θ the estimate of σ is determined. The extended quasi likelihood for the entire data is

$$Q^+ = -\frac{1}{2} \sum_{i=1}^N \left[\log(2\pi\sigma^2 g^2(z_i, \mu_i(\beta), \theta)) - \frac{2}{\sigma^2} \int_{y_i}^{\mu_i(\beta)} \frac{(y_i - \mu)}{g^2(z_i, \mu_i(\beta), \theta)} d\mu \right]$$

Estimate the parameters by jointly maximizing the extended quasi likelihood expression in θ , β and σ^2 . Differentiate with respect to θ , then $\hat{\theta}$ is the solution, assuming it exists, to the equation of first order derivative.

Define

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

Letting $\frac{\delta Q^+}{\delta \theta} = 0$

leads us to

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

Taking derivatives of Q^+ with respect to σ^2 gives us $\hat{\sigma}^2$ as the solution to the equation

$$\sum_{i=1}^N \left[\frac{1}{\sigma^2} + \frac{2}{\sigma^4} \int_{y_i}^{\mu_i(\hat{\beta})} \frac{(y_i - \mu)}{g^2(z_i, \mu_i(\hat{\beta}), \hat{\theta})} d\mu \right] = 0 \quad (2.22)$$

whereas derivatives of Q^+ with respect to β gives us $\hat{\beta}$ as the solution to the equation

$$\sum_{i=1}^N \frac{(y_i - \mu_i(\beta))}{g^2(z_i, \mu_i(\beta), \hat{\theta})} \frac{\delta}{\delta \beta} \mu_i(\beta) = 0 \quad (2.23)$$

2.3.7 Logarithm of absolute residuals

Suppose that the errors are independent and identically distributed. We pursue the idea that the logarithm of the absolute residuals has approximate expectation $\log\{\sigma g(z_i, \mu_i(\beta), \theta)\}$.

That is

$$E \log |y_i - f(x_i, \hat{\beta})| = \log\{\sigma g(z_i, \mu_i(\beta), \theta)\} \quad (2.24)$$

We regress the logarithm of the absolute residuals on the logarithm of their approximate values. This approach has one advantage due to the fact that calculations are easy since only one program is required. However, there are practical problems which arise especially when the residuals are very close to zero, in which case logarithms of those residuals cause large outliers. To avoid this difficulty we simply eliminate the residuals which are close to zero although another question which remain to be addressed is; how many do we eliminate?. As much as we perform this operation, the solution will be dictated by the population size.

Taking logarithms results in a linear model with θ as the slope and $\log(\sigma)$ as the intercept.

2.3.8 Rodbard and Frazier

This is a regression based method which is a close analog to the logarithm method except that we regress the sample standard deviation S_i on the logarithm of the observed sample mean \bar{y}_i . This scheme has two major weaknesses, one being that we are using the sample means instead of the true means, obviously this is an errors-in variables problem. The second problem arises when the number of replicates is not constant across groups. For the moment, the second problem is easy to tackle as we only need to make the number of replicates constant across groups. The interpretation of the regression remains the same as discussed for the logarithm method.

2.3.9 Maximum likelihood

This is a common procedure and basically it is familiar. Estimation of θ and β in the parametric model of (1.1) and (2.1) is possible by joint maximum likelihood estimation. We do not look at this method in detail as it is equivalent to other methods under study depending on the assumptions and conditions for the models used and distributions under consideration. In particular, when the variance function does not depend on the mean function, maximum likelihood is equivalent to weighted least squares methods based on the squared residuals. For well defined distributions, estimation of the unknown parameters by regression methods is similar to maximum likelihood estimation. It implies that once we have estimated the parameter θ using the other methods, using the maximum likelihood is an overlap. Which leaves us to suggest that depending on the constraints available like known mean, considering particular distribution and using estimates obtained from methods already discussed provide estimates which are easily obtained from the alternative methods.

Since we are interested in the normal distribution in our study, to the best of our knowledge estimates of regression methods are equivalent to the estimates of maximum likelihood procedure.

2.3.10 Sadler and Smith

The procedure is based on the assumption of independence. Estimation of θ is by joint maximization in $(M+r+1)$ parameters $\sigma^2, \theta, \mu_1, \mu_2, \dots, \mu_M$ of the expression

$$L(\beta, \sigma^2, \theta / Y) = \prod_{i=1}^M \left\{ [2\pi\sigma^2 g^2(z_i, \mu_i(\beta), \theta)]^{-\frac{(n-1)}{2}} \exp \left[-\sum_{j=1}^n \frac{[y_{ij} - \mu_i(\beta)]^2}{2\sigma^2 g^2(z_i, \mu_i(\beta), \theta)} \right] \right\} \quad (2.25)$$

with $\mu_i(\beta)$ replaced by the sample mean \bar{y}_i . Taking logarithms to both sides we get

$$l = -\left(\frac{n-1}{2}\right) \sum_{i=1}^M \left\{ \log[2\pi\sigma^2 g^2(z_i, \mu_i(\beta), \theta)] \right\} - \sum_{i=1}^M \sum_{j=1}^n \frac{[y_{ij} - \bar{y}_i]^2}{2\sigma^2 g^2(z_i, \mu_i(\beta), \theta)}$$

and taking derivatives with respect to σ^2 we obtain

$$\frac{\delta l}{\delta \sigma^2} = -\left(\frac{n-1}{2}\right) \sum_{i=1}^M \frac{1}{\sigma^2} + \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \bar{y}_i]^2}{2\sigma^4 g^2(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} \right]$$

Let $\frac{\delta l}{\delta \sigma^2} = 0$

then

$$\frac{1}{2\sigma^2} \sum_{i=1}^M (n-1) = \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \bar{y}_i]^2}{2\sigma^4 g^2(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} \right]$$

solving to

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \bar{y}_i]^2}{g^2(z_i, \mu_i(\hat{\beta}_*), \hat{\theta}_*)} \right]}{\sum_{i=1}^M (n-1)} \quad (2.26)$$

Next, minimize the loglikelihood in θ to get

$$\frac{\delta l}{\delta \theta} = -(n-1) \sum_{i=1}^M \left[\frac{\frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}_*), \theta)}{g(z_i, \mu_i(\hat{\beta}_*), \theta)} \right] + \sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \bar{y}_i]^2 \frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}_*), \theta)}{\hat{\sigma}^2 g^3(z_i, \mu_i(\hat{\beta}_*), \theta)} \right]$$

Now let $\frac{\delta l}{\delta \theta} = 0$

giving us

$$\sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - \bar{y}_i]^2 \frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}_*), \theta)}{\hat{\sigma}^2 g^3(z_i, \mu_i(\hat{\beta}_*), \theta)} \right] - (n-1) \sum_{i=1}^M \left[\frac{\frac{\delta}{\delta \theta_i} g(z_i, \mu_i(\hat{\beta}_*), \theta)}{g(z_i, \mu_i(\hat{\beta}_*), \theta)} \right] = 0 \quad (2.27)$$

Estimate of β is determined as in the modified maximum likelihood.

Chapter 3

EMPIRICAL STUDY

3.1 Introduction

Perhaps the most important criterion by which such theoretical illustration as are reported in chapter 2 must be evaluated is their ability to predict the actual performance of estimation strategies in practical problems. A good statistician selects his model class carefully paying attention to the type and structure of the data. There are two possible techniques used in selection of the model which are termed as either informal or formal. Informal technique rely upon the human mind and eye to detect the pattern. The argument is that if we detect the pattern in the scatter plot we find a better model. The practical problem is that any finite set of raw means or residuals can be made to yield some kind of pattern if we look hard enough, so that we guard against over-interpretation. Formal methods rely on embedding the current model in a wider class that includes all parameters.

Numerical computation and graphical analysis is done using **S-PLUS PROGRAMMING LANGUAGE**. Our objective more concretely is to understand fully the information contained in the output other than the operation of the program itself .

3.2 A study data

We shall use secondary data, see Jobson [1980]. Mainly we are interested in data which is equally spaced. In this data, the $\{x_i\}$ have this characteristic with equal spacing of 0.1 between 0 and 4.0 for 40 cases.

TABLE 3.1

x_i	y_{i1}	y_{i2}	x_i	y_{i1}	y_{i2}
0.1	9.04	5.40	2.1	18.05	9.10
0.2	13.16	5.30	2.2	21.42	11.80
0.3	12.50	7.70	2.3	22.91	17.90
0.4	14.20	1.70	2.4	22.53	23.10
0.5	16.93	12.10	2.5	10.27	7.00
0.6	19.27	26.00	2.6	14.41	12.10
0.7	9.56	2.70	2.7	20.31	16.70
0.8	14.55	7.40	2.8	22.05	13.90
0.9	15.64	5.20	2.9	25.96	29.10
1.0	16.74	10.00	3.0	26.85	30.00
1.1	17.33	9.60	3.1	14.84	3.80
1.2	17.80	12.90	3.2	19.31	11.50
1.3	11.04	4.00	3.3	23.72	16.90
1.4	15.48	7.30	3.4	22.70	16.50
1.5	16.94	10.50	3.5	32.90	19.50
1.6	19.38	7.70	3.6	17.05	18.00
1.7	27.54	6.80	3.7	22.25	15.30
1.8	23.03	12.30	3.8	20.16	15.10
1.9	23.70	0.00	3.9	28.84	10.20
2.0	15.33	8.60	4.0	36.19	20.90

3.3 Working models

We examine the scatter plot of the 40 observations of $\{y_i\}$ on $\{x_i\}$ as in Figure 3.1a.

The data are clearly heteroscedastic. Further simple smooth plots of $\{y_{i1}\}$ and $\{y_{i2}\}$ on $\{x_i\}$ are given in Figure 3.1b. We consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i=1, 2, \dots, N \quad (3.1)$$

Observe the residual plots against the predicted values as well as each of the predictors.

Certainly this technique is affected by outliers or non-normal error distributions. We require the residuals denoted by r_i in order to obtain a scatter plot. But we realize that we need to estimate the parameters β_0 and β_1 in the regression model. The method at our disposal is the least squares method. Minimize the error sum of squares (*ESS*).

Where

$$ESS = \sum_{I=1}^N (Y_I - \beta_0 - \beta_1 X_I)^2 \quad (3.2)$$

One method for carrying out this minimization is to differentiate (3.2) with respect to β_0 and β_1 , set the derivatives to zero, and solve the resulting equations for estimates. Carrying out

this plan, we differentiate with respect to β_0 to obtain

$$\frac{\delta}{\delta \hat{\beta}_0} ESS = -2 \sum_{I=1}^N (Y_I - \hat{\beta}_0 - \hat{\beta}_1 X_I)$$

Figure 3.1a: A scatter plot of study data using one set of replicate

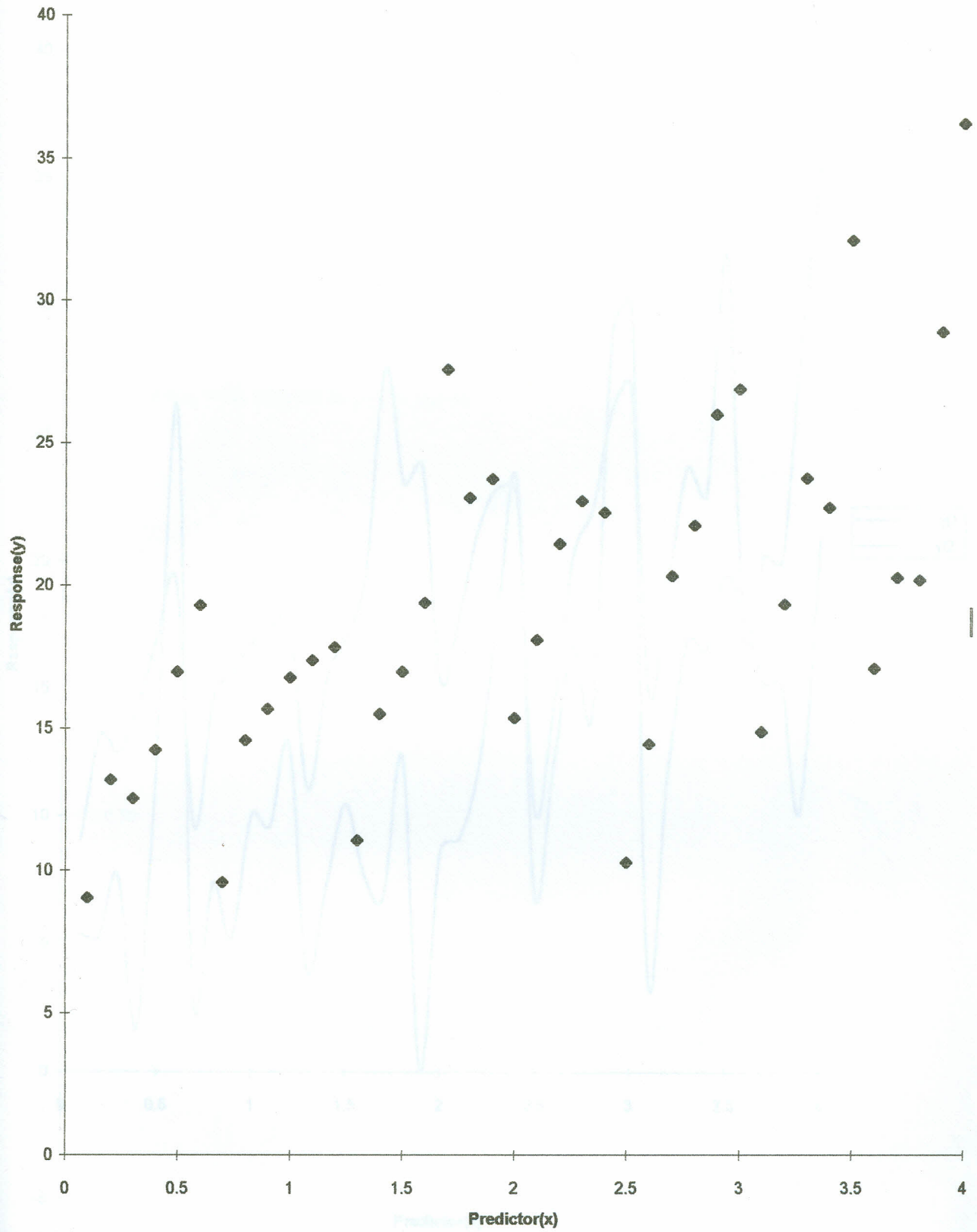
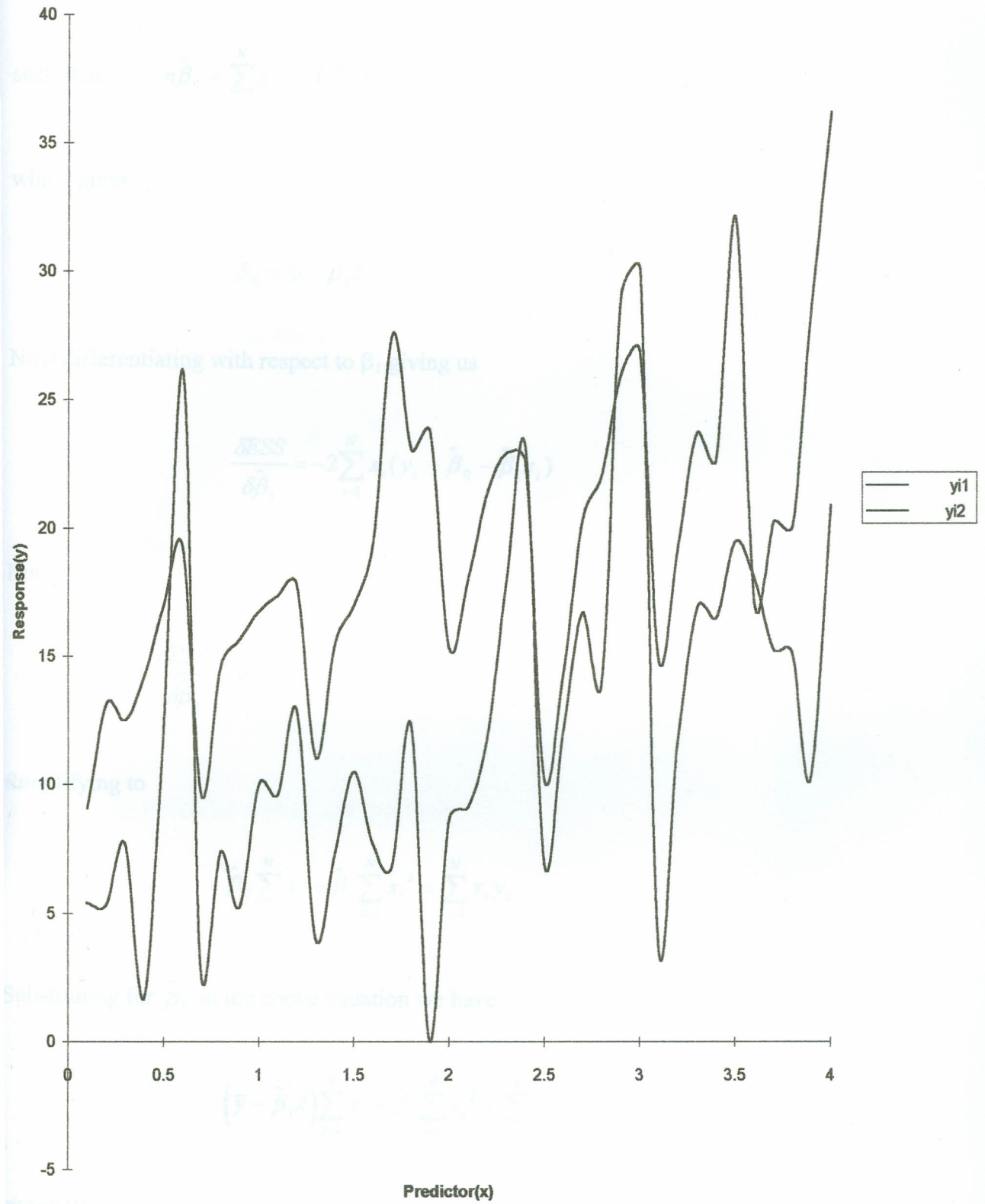


Figure 3.1b: Line plot of study data involving both replicates



But
$$\frac{\partial ESS}{\partial \hat{\beta}_0} = 0$$

such that
$$n\hat{\beta}_0 = \sum_{i=1}^N y_i - \hat{\beta}_1 \sum_{i=1}^N x_i$$

which gives

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3.3)$$

Now differentiating with respect to β_1 giving us

$$\frac{\partial ESS}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^N x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

But

$$\frac{\partial ESS}{\partial \hat{\beta}_1} = 0$$

Simplifying to

$$\hat{\beta}_0 \sum_{i=1}^N x_i + \hat{\beta}_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i y_i$$

Substituting for $\hat{\beta}_0$ in the above equation we have

$$(\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^N x_i + \hat{\beta}_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i y_i$$

giving us

$$\hat{\beta}_1 \left[\sum_{i=1}^N x_i^2 - N\bar{x}^2 \right] = \sum_{i=1}^N x_i y_i - N\bar{x}\bar{y}$$

Or

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N x_i y_i - N\bar{x}\bar{y}}{\sum_{i=1}^N x_i^2 - N\bar{x}^2}$$

Alternatively we can write this as

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

such that the predicted value \hat{y}_i is given by

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

We then compute for residuals as recorded in Table 3.2

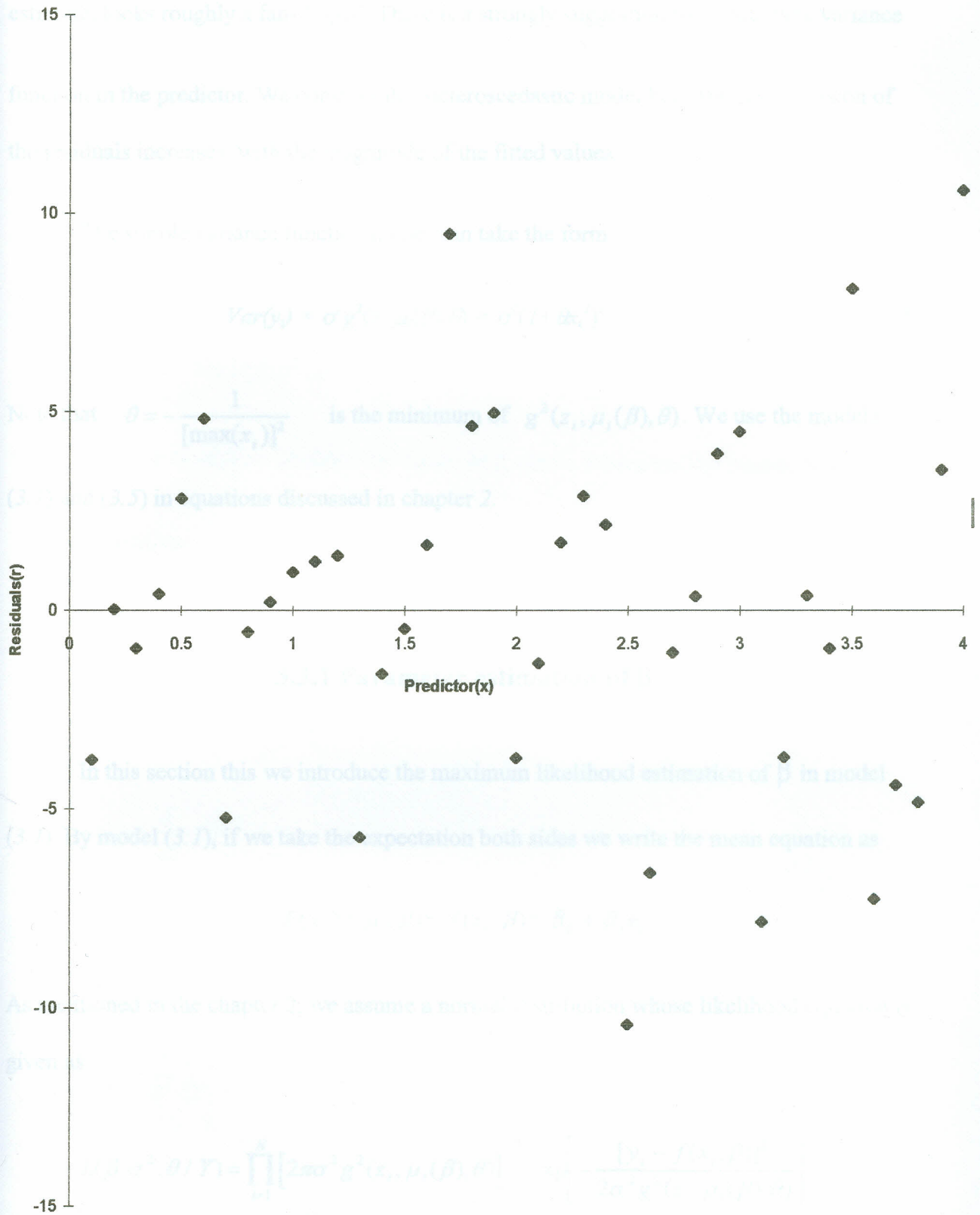
TABLE 3.2: Predictor values and residuals

x_i	r_i	x_i	r_i
0.1	-3.77	2.1	-1.34
0.2	0.02	2.2	1.70
0.3	-0.97	2.3	2.86
0.4	0.40	2.4	2.15

TABLE: 3.2 (cont.)

0.5	2.80	2.5	-10.44
0.6	4.81	2.6	-6.63
0.7	-5.22	2.7	-1.06
0.8	-0.56	2.8	0.35
0.9	0.20	2.9	3.94
1.0	0.97	3.0	4.50
1.1	1.23	3.1	-7.84
1.2	1.37	3.2	-3.70
1.3	-5.72	3.3	0.38
1.4	-1.61	3.4	-0.97
1.5	-0.48	3.5	8.09
1.6	1.63	3.6	-7.28
1.7	9.47	3.7	-4.41
1.8	4.63	3.8	-4.83
1.9	4.97	3.9	3.52
2.0	-3.73	4.0	10.54

Figure 3.2: A scatter plot of residuals and predictors of study data



An examination of the scatter plot in Figure 3.2 of residuals on $\{x_i\}$ by our eyeball estimate, looks roughly a fan-shaped. There is a strongly suggestion for a quadratic variance function in the predictor. We consider this heteroscedastic model because the dispersion of the residuals increases with the magnitude of the fitted values.

The simple variance function model can take the form

$$\text{Var}(y_i) = \sigma^2 g^2(z_i, \mu_i(\beta), \theta) = \sigma^2(1 + \theta x_i^2)^2 \quad (3.5)$$

Note that $\theta = -\frac{1}{[\max(x_i)]^2}$ is the minimum of $g^2(z_i, \mu_i(\beta), \theta)$. We use the models

(3.1) and (3.5) in equations discussed in chapter 2.

3.3.1 Parameter estimation of β

In this section this we introduce the maximum likelihood estimation of β in model

(3.1). By model (3.1), if we take the expectation both sides we write the mean equation as

$$E(y_i) = \mu_i(\beta) = f(x_i, \beta) = \beta_0 + \beta_1 x_i$$

As mentioned in the chapter 2, we assume a normal distribution whose likelihood equation is given as

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

The loglikelihood for the data of the normal distribution is then

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

Let

$$l = \log L(\beta, \sigma^2, \theta / Y)$$

Given σ^2, θ ; the maximum likelihood estimate of β can be determined by maximizing over β .

Taking derivatives of l with respect to β giving us

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

Where $\frac{\partial l}{\partial \beta} = 0$

Such that

$$\frac{1}{\sigma^2} \sum_{i=1}^N \left[\frac{[y_i - f(x_i, \beta)] \frac{\delta}{\delta \beta} f(x_i, \beta)}{\sigma^2 g^2(z_i, \mu_i(\beta), \theta)} \right] = 0 \quad (3.6)$$

By model (3.1) and (3.5) we construct suitable mean and variance equations defined as

$$f(x_i, \beta) = \beta_0 + \beta_1 x_i$$

$$g^2(z_i, \mu_i(\beta), \theta) = (1 + \theta x_i^2)^2$$

Now, using these mean and variance functions, we write (3.6) as

$$\frac{1}{\hat{\sigma}^2} \sum_{i=1}^N \frac{[y_i - (\beta_0 + \beta_1 x_i)] \frac{\delta}{\delta \beta_j} [\beta_0 + \beta_1 x_i]}{[1 + \hat{\theta} x_i^2]^2} = 0 \quad j=0, 1$$

which can be expressed in the form

$$\sum_{i=1}^N \left[\frac{\frac{\delta}{\delta \beta_0} [\beta_0 + \beta_1 x_i]}{\frac{\delta}{\delta \beta_1} [\beta_0 + \beta_1 x_i]} \right] \frac{[y_i - (\beta_0 + \beta_1 x_i)]}{[1 + \hat{\theta} x_i^2]^2} = 0$$

Simplifying to the expression

$$\sum_{i=1}^N \left[\frac{1}{x_i} \right] \frac{[y_i - (\beta_0 + \beta_1 x_i)]}{[1 + \hat{\theta} x_i^2]^2} = 0$$

This gives two equations. The first being

$$\sum_{i=1}^N \left[\frac{[y_i - (\beta_0 + \beta_1 x_i)]}{[1 + \hat{\theta} x_i^2]^2} \right] = 0 \quad (3.7)$$

Such that

$$\sum_{i=1}^N \frac{y_i}{[1 + \hat{\theta} x_i^2]^2} = \sum_{i=1}^N \frac{[\beta_0 + \beta_1 x_i]}{[1 + \hat{\theta} x_i^2]^2}$$

Or

$$\sum_{i=1}^N \frac{y_i}{[1 + \hat{\theta} x_i^2]^2} = \sum_{i=1}^N \frac{\beta_0}{[1 + \hat{\theta} x_i^2]^2} + \sum_{i=1}^N \frac{\beta_1 x_i}{[1 + \hat{\theta} x_i^2]^2} \quad (3.8)$$

The second equation is

$$\sum_{i=1}^N \frac{y_i x_i - (\beta_0 x_i + \beta_1 x_i^2)}{[1 + \hat{\theta} x_i^2]^2} = 0 \quad (3.9)$$

or

$$\sum_{i=1}^N \frac{y_i x_i}{[1 + \hat{\theta} x_i^2]^2} = \sum_{i=1}^N \frac{\beta_0 x_i}{[1 + \hat{\theta} x_i^2]^2} + \sum_{i=1}^N \frac{\beta_1 x_i^2}{[1 + \hat{\theta} x_i^2]^2} \quad (3.10)$$

Equation (3.8) can be written as

$$\begin{bmatrix} \sum_{i=1}^N \frac{1}{[1 + \hat{\theta} x_i^2]^2} & \sum_{i=1}^N \frac{x_i}{[1 + \hat{\theta} x_i^2]^2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \sum_{i=1}^N \frac{y_i}{[1 + \hat{\theta} x_i^2]^2}$$

Similarly from equation (3.10) we have

$$\begin{bmatrix} \sum_{i=1}^N \frac{x_i}{[1 + \hat{\theta} x_i^2]^2} & \sum_{i=1}^N \frac{x_i^2}{[1 + \hat{\theta} x_i^2]^2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \sum_{i=1}^N \frac{y_i x_i}{[1 + \hat{\theta} x_i^2]^2}$$

These expressions can be given in matrix form as

$$\begin{bmatrix} \sum_{i=1}^N \frac{1}{[1 + \hat{\theta} x_i^2]^2} & \sum_{i=1}^N \frac{x_i}{[1 + \hat{\theta} x_i^2]^2} \\ \sum_{i=1}^N \frac{x_i}{[1 + \hat{\theta} x_i^2]^2} & \sum_{i=1}^N \frac{x_i^2}{[1 + \hat{\theta} x_i^2]^2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \frac{y_i}{[1 + \hat{\theta} x_i^2]^2} \\ \sum_{i=1}^N \frac{y_i x_i}{[1 + \hat{\theta} x_i^2]^2} \end{bmatrix}$$

The general matrix form for solving these equations is therefore now given by the expression

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \frac{1}{[1 + \hat{\theta} x_i^2]^2} & \sum_{i=1}^N \frac{x_i}{[1 + \hat{\theta} x_i^2]^2} \\ \sum_{i=1}^N \frac{x_i}{[1 + \hat{\theta} x_i^2]^2} & \sum_{i=1}^N \frac{x_i^2}{[1 + \hat{\theta} x_i^2]^2} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^N \frac{y_i}{[1 + \hat{\theta} x_i^2]^2} \\ \sum_{i=1}^N \frac{y_i x_i}{[1 + \hat{\theta} x_i^2]^2} \end{bmatrix} \quad (3.11)$$

3.3.2 Pseudo likelihood estimation of parameter θ

It follows from equation (2.5) that the estimate of θ can be achieved by solving the

expression

$$\sum_{i=1}^N \left[\frac{y_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i]}{1 + \hat{\alpha}_i^2} \right]^2 \frac{x_i^2}{1 + \hat{\alpha}_i^2} - \frac{1}{N} \sum_{i=1}^N \left[\frac{y_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i]}{1 + \hat{\alpha}_i^2} \right]^2 \sum_{i=1}^N \left[\frac{x_i^2}{1 + \hat{\alpha}_i^2} \right] = 0 \quad (3.12)$$

Equation (2.4) becomes

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N \left[\frac{y_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i]}{1 + \hat{\alpha}_i^2} \right]^2 \quad (3.13)$$

3.3.3 Restricted maximum likelihood estimation of θ

As discussed in section (2.3.2), we solve

$$\sum_{i=1}^N \left[\frac{y_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i]}{1 + \hat{\alpha}_i^2} \right]^2 \frac{x_i^2}{1 + \hat{\alpha}_i^2} - \frac{1}{N-2} \sum_{i=1}^N \left[\frac{y_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i]}{1 + \hat{\alpha}_i^2} \right]^2 \sum_{i=1}^N \left[\frac{x_i^2 (1 - \hat{h}_{ii})}{1 + \hat{\alpha}_i^2} \right] = 0 \quad (3.14)$$

Where

$$Q = \left[\frac{1}{1 + \hat{\alpha}_i^2} \quad \frac{x_i}{1 + \hat{\alpha}_i^2} \right] \quad (3.15)$$

We estimate σ as in equation (3.13).

3.3.4 Least squares on squared residuals estimation of θ

Leverage Not Corrected

From equation (2.9) solve for θ as

$$\sum_{i=1}^N \frac{\left[\left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2 - \hat{\sigma}^2 (1 + \hat{\alpha} x_i^2)^2 \right] [1 + \hat{\alpha} x_i^2] (x_i^2)}{\left[1 + \hat{\theta}_* x_i^2 \right]^4} = 0 \quad (3.16)$$

and from equation (2.10), we solve for σ as

$$\sum_{i=1}^N \frac{\left[\left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2 - \sigma^2 (1 + \hat{\alpha} x_i^2)^2 \right] [1 + \hat{\alpha} x_i^2]^2}{\left[1 + \hat{\theta}_* x_i^2 \right]^4} = 0$$

If $\hat{\sigma}$ is the maximum likelihood estimate of σ then

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N \left[\frac{\left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2 [1 + \hat{\alpha} x_i^2]^2}{\left[1 + \hat{\theta}_* x_i^2 \right]^4} \right]}{\sum_{i=1}^N \left[\frac{1 + \hat{\alpha} x_i^2}{1 + \hat{\theta}_* x_i^2} \right]^4} \quad (3.17)$$

Leverage Corrected

We obtain θ from

$$\sum_{i=1}^N \frac{\left[\left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2 - \hat{\sigma}^2 (1 - \hat{h}_{ii}) (1 + \hat{\alpha} x_i^2)^2 \right] (1 + \hat{\alpha} x_i^2) (x_i^2)}{\left[1 - \hat{h}_{ii} \right]^2 \left[1 + \hat{\theta}_* x_i^2 \right]^4} = 0 \quad (3.18)$$

and σ from

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N \left[\frac{\left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2 [1 + \hat{\alpha} x_i^2]^2}{\left[1 - \hat{h}_{ii} \right]^2 \left[1 + \hat{\theta}_* x_i^2 \right]^4} \right]}{\sum_{i=1}^N \left[\frac{(1 + \hat{\alpha} x_i^2)^4}{\left[1 + \hat{\theta}_* x_i^2 \right]^4 (1 - \hat{h}_{ii})} \right]} \quad (3.19)$$

3.3.5 Least squares on absolute residuals estimate of θ

Leverage Not Corrected

From equation (2.13) we solve for θ as

$$\sum_{i=1}^N \frac{\left[\left| y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right| - \hat{\sigma}(1 + \hat{\alpha}_i^2) \right] [x_i^2]}{[1 + \hat{\theta}_* x_i^2]^2} = 0 \quad (3.20)$$

and for σ as

$$\sum_{i=1}^N \frac{\left[\left| y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right| - \sigma(1 + \hat{\alpha}_i^2) \right] [1 + \hat{\alpha}_i^2]}{[1 + \hat{\theta}_* x_i^2]^2} = 0$$

which we can write as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N \left[\frac{\left[\left| y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right| \right] [1 + \hat{\alpha}_i^2]}{[1 + \hat{\theta}_* x_i^2]^2} \right]}{\sum_{i=1}^N \left[\frac{1 + \hat{\alpha}_i^2}{1 + \hat{\theta}_* x_i^2} \right]^2} \quad (3.21)$$

Leverage Corrected

From equation (3.20) we solve for using

$$\sum_{i=1}^N \frac{\left[\left| y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right| - \hat{\sigma}(1 - \hat{h}_{ii})(1 + \hat{\alpha}_i^2) \right] [x_i^2]}{(1 - \hat{h}_{ii})[1 + \hat{\theta}_* x_i^2]^2} = 0 \quad (3.22)$$

and from equation (2.14) we solve for σ as

$$\sum_{i=1}^N \frac{\left[\left| y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right| - \sigma(1 - \hat{h}_{ii})(1 + \hat{\alpha}_i^2) \right] [1 + \hat{\alpha}_i^2]}{(1 - \hat{h}_{ii})[1 + \hat{\theta}_* x_i^2]^2} = 0$$

which can be given as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N \left[\frac{[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)] [1 + \hat{\alpha} x_i^2]}{(1 - \hat{h}_{ii}) [1 + \hat{\theta}_* x_i^2]^2} \right]}{\sum_{i=1}^N \left[\frac{1 + \hat{\alpha} x_i^2}{1 + \hat{\theta}_* x_i^2} \right]^2} \quad (3.23)$$

3.3.6 Modified maximum likelihood equation for $\hat{\theta}$

From equation (2.17) we obtain

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2}{[1 + \hat{\alpha} x_i^2]^2} \right]}{\sum_{i=1}^M (n-1)} \quad (3.24)$$

and from equation (2.19) we obtain $\hat{\theta}$ as

$$\sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1)]^2 [x_i^2]}{\hat{\sigma}^2 [1 + \hat{\alpha} x_i^2]^3} \right] - (n-1) \sum_{i=1}^M \left[\frac{x_i^2}{1 + \hat{\alpha} x_i^2} \right] = 0 \quad (3.25)$$

From equation (2.20) we obtain $\hat{\beta}$ as

$$\sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1)] \begin{bmatrix} 1 \\ x_i \end{bmatrix}}{[1 + \hat{\alpha} x_i^2]^2} \right] = 0$$

Which can be split into two expressions as

$$\sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]}{[1 + \hat{\alpha} x_i^2]^2} \right] = 0$$

and

$$\sum_{i=1}^M \sum_{j=1}^n \left[\frac{[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1)(x_i)](x_i)}{[1 + \hat{\alpha}_i^2]^2} \right] = 0$$

Thus

$$\hat{\beta}_0 \sum_{i=1}^M \frac{1}{[1 + \hat{\alpha}_i^2]^2} + \hat{\beta}_1 \sum_{i=1}^M \frac{x_i}{[1 + \hat{\alpha}_i^2]^2} = \frac{1}{n} \sum_{i=1}^M \sum_{j=1}^n \frac{y_{ij}}{[1 + \hat{\alpha}_i^2]^2}$$

and

$$\hat{\beta}_0 \sum_{i=1}^M \frac{x_i}{[1 + \hat{\alpha}_i^2]^2} + \hat{\beta}_1 \sum_{i=1}^M \frac{x_i^2}{[1 + \hat{\alpha}_i^2]^2} = \frac{1}{n} \sum_{i=1}^M \sum_{j=1}^n \frac{y_{ij} x_i}{[1 + \hat{\alpha}_i^2]^2}$$

simplifying to

$$\begin{bmatrix} \sum_{i=1}^M \frac{1}{[1 + \hat{\alpha}_i^2]^2} & \sum_{i=1}^M \frac{x_i}{[1 + \hat{\alpha}_i^2]^2} \\ \sum_{i=1}^M \frac{x_i}{[1 + \hat{\alpha}_i^2]^2} & \sum_{i=1}^M \frac{x_i^2}{[1 + \hat{\alpha}_i^2]^2} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^M \sum_{j=1}^n \frac{y_{ij}}{[1 + \hat{\alpha}_i^2]^2} \\ \frac{1}{n} \sum_{i=1}^M \sum_{j=1}^n \frac{y_{ij} x_i}{[1 + \hat{\alpha}_i^2]^2} \end{bmatrix}$$

Hence

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^M \frac{1}{[1 + \hat{\alpha}_i^2]^2} & \sum_{i=1}^M \frac{x_i}{[1 + \hat{\alpha}_i^2]^2} \\ \sum_{i=1}^M \frac{x_i}{[1 + \hat{\alpha}_i^2]^2} & \sum_{i=1}^M \frac{x_i^2}{[1 + \hat{\alpha}_i^2]^2} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^M \sum_{j=1}^n \frac{\bar{y}_i}{[1 + \hat{\alpha}_i^2]^2} \\ \sum_{i=1}^M \sum_{j=1}^n \frac{x_i \bar{y}_i}{[1 + \hat{\alpha}_i^2]^2} \end{bmatrix} \quad (3.26)$$

Note that this is for equal replications. For unequal replications we, replace n by n_i . Thus

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^M \sum_{j=1}^{n_i} \frac{[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2}{[1 + \hat{\alpha}_i^2]^2}}{\sum_{i=1}^M (n_i - 1)} \quad (3.27)$$

while for θ the estimate is

$$\sum_{i=1}^M \frac{\sum_{j=1}^{n_i} [y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 x_i^2}{\hat{\sigma}^2 [1 + \hat{\alpha}_i^2]^3} - \sum_{i=1}^M \frac{[n_i - 1] x_i^2}{[1 + \hat{\alpha}_i^2]} = 0 \quad (3.28)$$

We realize that the actual mean (μ_i) is unknown and therefore it is replaced by its estimate, the sample mean \bar{y}_i .

3.3.7 Extended quasi likelihood estimation of θ

From equation (2.21) we obtain $\hat{\theta}$ as

$$\sum_{i=1}^N \left[\frac{2x_i^2}{[1 + \hat{\alpha}_i^2]} + \frac{4}{\hat{\sigma}^2} \int_{y_i}^{\mu_i(\hat{\beta})} \frac{[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)] x_i^2 \delta\mu}{[1 + \hat{\alpha}_i^2]^3} \right] = 0$$

which becomes

$$\sum_{i=1}^N \left[\frac{x_i^2}{[1 + \hat{\alpha}_i^2]} + \frac{2y_i [\hat{\beta}_0 + \hat{\beta}_1 x_i] x_i^2}{\hat{\sigma}^2 [1 + \hat{\alpha}_i^2]^3} - \frac{2y_i^2 x_i^2}{\hat{\sigma}^2 [1 + \hat{\alpha}_i^2]^3} - \frac{x_i^2 [\hat{\beta}_0 + \hat{\beta}_1 x_i]^2}{\hat{\sigma}^2 [1 + \hat{\alpha}_i^2]^3} + \frac{x_i^2 y_i^2}{\hat{\sigma}^2 [1 + \hat{\alpha}_i^2]^3} \right] = 0$$

Thus

$$\sum_{i=1}^N \left[\frac{\hat{\sigma}^2 x_i^2 [1 + \hat{\alpha}_i^2] + 2y_i [\hat{\beta}_0 + \hat{\beta}_1 x_i] x_i^2 - y_i^2 x_i^2 - x_i^2 [\hat{\beta}_0 + \hat{\beta}_1 x_i]^2}{\hat{\sigma}^2 [1 + \hat{\alpha}_i^2]^3} \right] = 0 \quad (3.29)$$

Now from equation (2.22), σ is given by

$$\sum_{i=1}^N \left[\frac{1}{\sigma^2} + \frac{2}{\sigma^4} \left[\frac{2y_i [\hat{\beta}_0 + \hat{\beta}_1 x_i] - [\hat{\beta}_0 + \hat{\beta}_1 x_i]^2 - [2y_i^2 - y_i^2]}{2[1 + \hat{\alpha}_i^2]^2} \right] \right] = 0$$

which gives

$$\sum_{i=1}^N \left[\frac{\sigma^2 [1 + \hat{\alpha}_i^2] + 2y_i [\hat{\beta}_0 + \hat{\beta}_1 x_i] - y_i^2 - [\hat{\beta}_0 + \hat{\beta}_1 x_i]^2}{[1 + \hat{\alpha}_i^2]^2} \right] = 0$$

Hence

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N \left[\frac{y_i^2 + [\hat{\beta}_0 + \hat{\beta}_1 x_i]^2 - 2y_i[\hat{\beta}_0 + \hat{\beta}_1 x_i]}{[1 + \hat{\theta} x_i^2]^2} \right] \quad (3.30)$$

3.3.8 Rodbard and Frazier

This procedure involves graphical techniques only, where we obtain a simple linear regression of the logarithm of the sample standard deviation on the logarithm of the sample mean.

3.3.9 Sadler and Smith

From equation (2.26) we solve for σ

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^n \frac{[y_{ij} - \bar{y}_i]^2}{[1 + \hat{\theta} x_i^2]^2} \quad (3.31)$$

and from equation (2.27) we solve for $\hat{\theta}$ as

$$\sum_{i=1}^M \sum_{j=1}^n \frac{[y_{ij} - \bar{y}_i] x_i^2}{\hat{\sigma} [1 + \hat{\theta} x_i^2]^3} - (n-1) \sum_{i=1}^M \left[\frac{x_i^2}{[1 + \hat{\theta} x_i^2]} \right] = 0 \quad (3.32)$$

Note that the estimate of β is determined as in the modified likelihood procedure.

3.3.10 Algorithm

The parameter θ is estimated iteratively. The algorithm is as follows.

Step 1: Start with preliminary estimate of parameter θ as zero. Determine the estimate of β .

Step 2: Estimate σ using the current estimates of β and θ .

Step 3: Update the estimate of θ using the current β and σ .

Step 4: Repeat steps 1-3 until there is minimal or no change at all.

The steps may be made clear in a flow diagram in Figure 3.3 shown below.

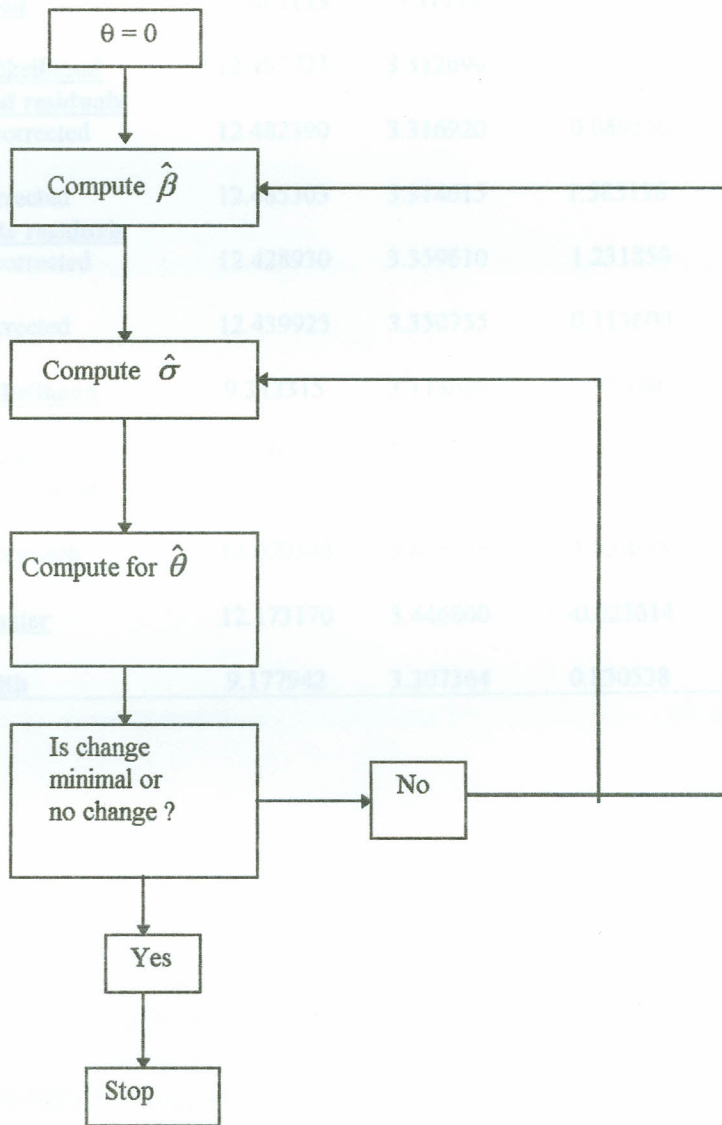


Fig 3.3 Algorithm flow diagram.

TABLE 3.3**Results of the analysis on the study data for Figure 3.1 assuming (3.5).**

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\theta}$	$\hat{\sigma}^2$
<u>Pseudo likelihood</u>	12.482133	3.317125	0.089680	8.788860
<u>Restricted maximum likelihood</u>	12.487723	3.312699	1.932678	9.441409
<u>Least squares on squared residuals</u>				
Leverage not corrected	12.482390	3.316920	0.089550	8.797012
Leverage corrected	12.485303	3.314615	1.585116	9.300803
<u>Least squares on absolute residuals</u>				
Leverage not corrected	12.428930	3.359610	1.231850	4.511729
Leverage corrected	12.439925	3.350755	0.113600	5.163838
<u>Modified maximum likelihood</u>	9.312315	3.113629	0.502096	8.881123
<u>Extended quasi likelihood</u>	12.480790	3.318189	0.090456	8.746938
<u>Logarithm of absolute residuals</u>				
Regression approach	12.070340	3.686545	0.401075	0.627825
<u>Rodbard and Frazier</u>	12.173170	3.446800	-0.023614	8.790574
<u>Sadler and Smith</u>	9.177942	3.307364	0.830538	3.198697

Chapter 4

NONPARAMETRIC APPROACH

4.1 INTRODUCTION

Here, the variance function is completely unspecified. The nonparametric approach in estimating a regression curve has its main purposes which include

- (a) Provides a versatile method of exploring a general relationship between two variables.
- (b) Gives predictions of observations yet to be made without reference to a fixed parametric model.
- (c) Provides a tool for finding spurious observations by studying the influence of isolated points.
- (d) Constitutes a flexible method of substituting for missing values or interpolating between adjacent x -values.

By the nature of flexibility the nonparametric method is helpful in a preliminary and exploratory statistical analysis of a data set.

We specifically explore the possibility of situations where nothing is known about the variance function in heteroscedastic regression problems except that it is a smooth function of

the design or mean response. Let us denote the variance function by $H(x_i, \hat{\beta})$ or $H(c_i)$. In estimating the variance function we use residuals. We define residuals as in equation (2.2).

Thus the expectation of squared residuals gives the estimate of the variance function given by

$$Er_i^2 = E[y_i - f(x_i, \hat{\beta})]^2 \cong H(x_i, \hat{\beta}) \quad (4.1)$$

We can also have the model in the design alone which is defined as

$$Var(y_i) = \sigma_i = H(c_i), \quad H(.) \text{ unknown} \quad (4.2)$$

Where $\{c_i\}$ is a set of identically independently distributed random variable independent of $\{e_i\}$. In practical studies, to achieve the smooth conditions we use large data sets with the help of graphical enhancements and smoothing techniques.

The smoothing methods commonly used include

1. Kernel smoothing
2. Spline smoothing
3. K-nearest-neighbor smoothing
4. Median smoothing
5. Convolution smoothing
6. The regressogram smoothing

We will use Kernel smoothing to portray the approach.

4.2 Kernel smoothing

A variety of Kernel functions have been studied. However both practical and theoretical considerations limit the choice of these Kernels. Given the bivariate data (x_i, y_i) , $i=1, 2, \dots, N$, the smoothed values \hat{z}_i produced by a Kernel smoother function $K(x)$ can be

given as

$$\hat{z}_i = \frac{\sum_{j=1}^N K\left(\frac{x_i - x_j}{b}\right) z_j}{\sum_{j=1}^N K\left(\frac{x_i - x_j}{b}\right)}, \quad 0 \leq x_i, x_j \leq 1, \quad i=1, 2, \dots, N \quad (4.3)$$

Nadaraya [1964] and Watson [1964].

A Kernel function $K(x)$ typically has the following properties.

(a) $K(x) \geq 0$ for all x

(b) $\int_{-\infty}^{\infty} K(x) dx = 1$

(c) $K(x) = K(-x)$ for all x

(d) $\int_{-\infty}^{\infty} xK(x) dx = 0$

(e) $\int_{-\infty}^{\infty} x^2 K(x) dx = K_2 \neq 0$

(f) $\int_{-\infty}^{\infty} [K(x)]^2 dx < \infty$.

The observations considered are in a small neighborhood of x since Y -observations far away from x will have, in general, very different mean values. We shall use the Gaussian Kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right] \quad |x| < 1 \quad (4.4)$$

In obtaining the smooth curve the selection of the Kernel function is not enough but rather the consideration of the bandwidth abbreviated by b is equally important. Several procedures for obtaining the bandwidth have been studied. Our choice is trial and error method where we make several plots and select the bandwidth which outperform the rest.

4.3 Example

An economist wanted to study the Yield of treasury bonds in successive months over a 21-year period from [1970] through [1990]. He wanted to construct a regression model that relates time $\{x_i\}$ and $\{y_i\}$ of treasury bonds. We will use these data to illustrate numerically the procedures discussed in the previous sections. The list of variables and estimates are provided in Table 4.1 where

- x_i Time
- y_i Yield
- r_i Residuals (Resids)
- r_i^2 Residuals squared (Rs. Sq.)
- $\hat{\sigma}_{iP}^2$ Parametric variance estimates (P. V. E.) of pseudo likelihood procedure.
- $\hat{\sigma}_{iNP}^2$ Nonparametric variance estimates (N. V. E)

Table 4.1 Variables and estimates for Yield data.

Time	Yield	Resids	Rs.Sq	P.V.E.	N.V.E	Time	Yield	Resids	Rs.Sq	P.V.E.	N.V.E
(x_i)	(y_i)	(r_i)	(r_i^2)	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$	(x_i)	(y_i)	(r_i)	(r_i^2)	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$
0.004	2.22	0.273	0.0748	1.0000	0.1548	0.032	1.98	-0.140	0.0195	1.0002	0.1733
0.008	2.23	0.259	0.0669	1.0000	0.1573	0.036	1.94	-0.204	0.0417	1.0002	0.1759
0.012	2.22	0.224	0.0502	1.0000	0.1601	0.040	1.79	-0.379	0.1437	1.0002	0.1785
0.016	2.20	0.179	0.0321	1.0000	0.1628	0.044	1.74	-0.454	0.2059	1.0003	0.1811
0.020	2.09	0.045	0.0020	1.0001	0.1654	0.048	1.86	-0.358	0.1285	1.0003	0.1837
0.024	1.97	-0.100	0.0100	1.0001	0.1681	0.052	1.78	-0.463	0.2145	1.0004	0.1862
0.028	2.03	-0.065	0.0042	1.0001	0.1707	0.056	1.72	-0.548	0.3002	1.0005	0.1888

Table 4.1 (continued).

x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$	x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$
0.060	1.79	-0.503	0.2526	1.0005	0.1914	0.155	3.17	0.284	0.0807	1.0037	0.2980
0.063	1.82	-0.497	0.2473	1.0006	0.1939	0.159	3.09	0.179	0.0322	1.0039	0.3068
0.067	1.89	-0.452	0.2043	1.0007	0.1965	0.163	3.02	0.085	0.0072	1.0041	0.3160
0.071	1.99	-0.377	0.1420	1.0008	0.1991	0.167	2.99	0.030	0.0009	1.0043	0.3257
0.075	1.89	-0.501	0.2515	1.0009	0.2017	0.171	2.97	-0.015	0.0002	1.0045	0.3359
0.079	1.83	-0.586	0.3436	1.0010	0.2045	0.175	2.94	-0.069	0.0048	1.0047	0.3464
0.083	1.71	-0.731	0.5342	1.0011	0.2072	0.179	2.84	-0.194	0.0377	1.0049	0.3574
0.087	1.70	-0.766	0.5862	1.0012	0.2101	0.183	2.85	-0.209	0.0436	1.0051	0.3687
0.091	1.97	-0.520	0.2708	1.0013	0.2131	0.187	2.86	-0.223	0.0500	1.0054	0.3804
0.095	2.21	-0.305	0.0931	1.0014	0.2163	0.190	2.89	-0.218	0.0477	1.0056	0.3924
0.099	2.36	-0.180	0.0323	1.0015	0.2196	0.194	2.93	-0.203	0.0412	1.0058	0.4047
0.103	2.41	-0.155	0.0239	1.0016	0.2230	0.198	2.93	-0.228	0.0519	1.0061	0.4171
0.107	2.92	0.331	0.1094	1.0018	0.2268	0.202	2.87	-0.312	0.0976	1.0063	0.4298
0.111	3.15	0.536	0.2873	1.0019	0.2307	0.206	2.82	-0.387	0.1499	1.0066	0.4426
0.115	3.26	0.621	0.3861	1.0020	0.2350	0.210	2.63	-0.602	0.3623	1.0068	0.4555
0.119	3.51	0.847	0.7168	1.0022	0.2395	0.214	2.33	-0.927	0.8587	1.0071	0.4684
0.123	3.48	0.792	0.6271	1.0023	0.2444	0.218	2.22	-1.061	1.1265	1.0073	0.4813
0.127	3.16	0.447	0.2000	1.0025	0.2496	0.222	2.15	-1.156	1.3365	1.0076	0.4942
0.131	3.01	0.272	0.0742	1.0026	0.2552	0.226	2.28	-1.058	1.1042	1.0079	0.5070
0.135	2.97	0.208	0.0432	1.0028	0.2612	0.230	2.28	-1.076	1.1567	1.0082	0.5197
0.139	2.88	0.093	0.0086	1.0030	0.2677	0.234	2.06	-1.320	1.7430	1.0084	0.5322
0.143	2.91	0.098	0.0097	1.0031	0.2746	0.238	2.54	-0.865	0.7481	1.0087	0.5446
0.147	3.45	0.614	0.3765	1.0033	0.2819	0.242	2.29	-1.140	1.2988	1.0090	0.5567
0.151	3.29	0.429	0.1839	1.0035	0.2897	0.246	2.66	-0.794	0.6310	1.0093	0.5685

Table 4.1 (Continued)

x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$	x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$
0.250	3.03	-0.449	0.2017	1.0096	0.5800	0.345	4.79	0.718	0.5150	1.0184	0.7204
0.254	3.17	-0.334	0.1114	1.0099	0.5914	0.349	4.83	0.733	0.5372	1.0188	0.7180
0.258	3.83	0.301	0.0909	1.0102	0.6023	0.353	4.80	0.678	0.4600	1.0193	0.7147
0.262	3.99	0.437	0.1907	1.0106	0.6129	0.357	4.83	0.683	0.4672	1.0197	0.7106
0.266	4.11	0.532	0.2831	1.0109	0.6232	0.361	4.77	0.599	0.3585	1.0201	0.7057
0.270	4.51	0.907	0.8232	1.0112	0.6330	0.365	4.80	0.604	0.3649	1.0206	0.6999
0.274	4.66	1.032	1.0662	1.0115	0.6425	0.369	5.38	1.159	1.3441	1.0210	0.6932
0.278	4.37	0.718	0.5153	1.0119	0.6515	0.373	6.18	1.935	3.7428	1.0215	0.6858
0.282	4.45	0.773	0.5978	1.0122	0.6601	0.377	6.02	1.750	3.0622	1.0219	0.6776
0.286	4.58	0.878	0.7716	1.0126	0.6683	0.381	5.91	1.615	2.6088	1.0224	0.6687
0.290	4.58	0.854	0.7288	1.0129	0.6760	0.385	5.66	1.340	1.7969	1.0229	0.6591
0.294	4.76	1.009	1.0181	1.0133	0.6832	0.389	5.42	1.076	1.1572	1.0234	0.6489
0.298	4.89	1.114	1.2416	1.0137	0.6898	0.393	5.06	0.691	0.4775	1.0238	0.6381
0.302	4.65	0.850	0.7217	1.0140	0.6960	0.397	4.70	0.306	0.0938	1.0243	0.6269
0.306	4.51	0.685	0.4690	1.0144	0.7016	0.401	4.73	0.312	0.0971	1.0248	0.6152
0.310	4.65	0.800	0.6402	1.0148	0.7066	0.405	4.64	0.197	0.0388	1.0253	0.6031
0.313	4.52	0.645	0.4165	1.0152	0.7110	0.409	4.62	0.152	0.0232	1.0258	0.5908
0.317	4.52	0.621	0.3852	1.0155	0.7147	0.413	4.48	-0.013	0.0002	1.0263	0.5783
0.321	4.57	0.646	0.4173	1.0159	0.7178	0.417	4.43	-0.087	0.0076	1.0268	0.5658
0.325	4.65	0.701	0.4917	1.0163	0.7214	0.421	4.33	-0.212	0.0449	1.0274	0.5532
0.329	4.74	0.767	0.5876	1.0167	0.7218	0.425	4.32	-0.247	0.0609	1.0279	0.5407
0.333	5.10	1.102	1.2140	1.0171	0.7226	0.429	4.30	-0.291	0.0849	1.0284	0.5283
0.337	5.00	0.977	0.9547	1.0176	0.7227	0.433	4.26	-0.356	0.1268	1.0289	0.5162
0.341	4.74	0.692	0.4794	1.0180	0.7220	0.437	4.02	-0.621	0.3855	1.0295	0.5044

Table 4.1 (continued).

x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$	x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$
0.440	4.06	-0.606	0.3667	1.0300	0.4930	0.536	5.44	0.181	0.0328	1.0446	0.4620
0.444	4.08	-0.610	0.3725	1.0306	0.4821	0.540	5.32	0.036	0.0013	1.0452	0.4731
0.448	4.09	-0.625	0.3907	1.0311	0.4717	0.544	5.21	-0.098	0.0097	1.0459	0.4852
0.452	4.14	-0.600	0.3597	1.0317	0.4619	0.548	5.47	0.137	0.0188	1.0466	0.4981
0.456	4.15	-0.614	0.3776	1.0322	0.4528	0.552	5.96	0.602	0.3628	1.0473	0.5120
0.460	4.20	-0.589	0.3471	1.0328	0.4444	0.556	6.50	1.118	1.2490	1.0480	0.5266
0.464	4.30	-0.514	0.2641	1.0334	0.4368	0.560	6.48	1.073	1.1510	1.0487	0.5420
0.468	4.26	-0.579	0.3348	1.0340	0.4300	0.563	6.00	0.568	0.3228	1.0494	0.5582
0.472	4.15	-0.713	0.5089	1.0345	0.4240	0.567	5.83	0.373	0.1394	1.0501	0.5749
0.476	4.27	-0.618	0.3820	1.0351	0.4189	0.571	5.91	0.429	0.1838	1.0508	0.5921
0.480	4.69	-0.223	0.0496	1.0357	0.4147	0.575	5.98	0.474	0.2247	1.0515	0.6098
0.484	4.72	-0.218	0.0473	1.0363	0.4115	0.579	5.91	0.379	0.1438	1.0522	0.6278
0.488	4.92	-0.042	0.0018	1.0369	0.4092	0.583	5.64	0.085	0.0071	1.0530	0.6460
0.492	5.10	0.113	0.0128	1.0375	0.4079	0.587	5.49	-0.090	0.0081	1.0539	0.6642
0.496	5.20	0.188	0.0355	1.0382	0.4076	0.591	5.43	-0.175	0.0306	1.0544	0.6824
0.500	5.56	0.524	0.2742	1.0388	0.4084	0.595	5.33	-0.300	0.0898	1.0552	0.7005
0.504	6.08	1.019	1.0382	1.0394	0.4102	0.599	5.22	-0.434	0.1886	1.0559	0.7181
0.508	6.13	1.044	1.0903	1.0400	0.4130	0.603	5.03	-0.649	0.4213	1.0567	0.7353
0.512	6.09	0.979	0.9594	1.0407	0.4169	0.607	4.74	-0.964	0.9289	1.0574	0.7519
0.516	5.99	0.855	0.7306	1.0413	0.4218	0.611	4.55	-1.178	1.3889	1.0582	0.7677
0.520	5.58	0.420	0.1764	1.0419	0.4278	0.615	4.68	-1.107	1.1518	1.0590	0.7825
0.524	5.59	0.405	0.1643	1.0426	0.4348	0.619	4.53	-1.248	1.5573	1.0597	0.7963
0.528	5.42	0.211	0.0444	1.0432	0.4429	0.623	4.67	-1.133	1.2829	1.0605	0.8089
0.532	5.30	0.066	0.0043	1.0439	0.4520	0.627	4.81	-1.017	1.0350	1.0613	0.8201

Table 4.1 (Continued)

x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$	x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{iP}^2$	$\hat{\sigma}_{iNP}^2$
0.631	4.98	-0.872	0.7605	1.0621	0.8298	0.722	6.52	0.099	0.0099	1.0817	0.5858
0.635	5.00	-0.877	0.7688	1.0629	0.8380	0.726	6.60	0.155	0.0239	1.0826	0.5624
0.639	4.94	-0.962	0.9245	1.0637	0.8444	0.730	6.78	0.310	0.0961	1.0836	0.5392
0.643	4.84	-1.086	1.1799	1.0645	0.8491	0.734	6.79	0.295	0.0871	1.0845	0.5162
0.647	4.76	-1.191	1.4184	1.0653	0.8519	0.738	6.83	0.311	0.0964	1.0854	0.4937
0.651	4.67	-1.306	1.7048	1.0661	0.8529	0.742	6.91	0.366	0.1338	1.0864	0.4717
0.655	4.51	-1.490	2.2213	1.0669	0.8518	0.746	6.93	0.361	0.1304	1.0873	0.4505
0.659	4.42	-1.605	2.5764	1.0678	0.8488	0.750	6.65	0.056	0.0032	1.0883	0.4301
0.663	4.53	-1.520	2.3099	1.0686	0.8439	0.754	6.53	-0.088	0.0078	1.0892	0.4106
0.667	4.70	-1.375	1.8894	1.0694	0.8370	0.758	6.50	-0.143	0.0205	1.0902	0.3922
0.671	4.75	-1.349	1.8205	1.0703	0.8282	0.762	6.69	0.022	0.0005	1.0912	0.3748
0.675	4.90	-1.224	1.4981	1.0711	0.8175	0.766	6.58	-0.113	0.0127	1.0921	0.3587
0.679	5.06	-1.089	1.1853	1.0720	0.8051	0.770	6.42	-0.297	0.0883	1.0931	0.3438
0.683	4.99	-1.183	1.4005	1.0728	0.7909	0.774	6.79	0.048	0.0023	1.0941	0.3302
0.687	4.96	-1.238	1.5333	1.0737	0.7752	0.778	6.82	0.053	0.0028	1.0951	0.3179
0.690	5.03	-1.193	1.4229	1.0746	0.7580	0.782	6.76	-0.031	0.0010	1.0961	0.3069
0.694	5.22	-1.028	1.0559	1.0754	0.7395	0.786	6.88	0.064	0.0041	1.0971	0.2974
0.698	5.47	-0.802	0.6437	1.0763	0.7198	0.790	7.22	0.379	0.1438	1.0981	0.2892
0.702	5.45	-0.847	0.7174	1.0772	0.6991	0.794	7.41	0.544	0.2964	1.0991	0.2824
0.706	5.48	-0.842	0.7085	1.0781	0.6775	0.798	7.27	0.380	0.1442	1.1001	0.2769
0.710	5.57	-0.776	0.6029	1.0790	0.6552	0.802	7.03	0.115	0.0132	1.1011	0.2728
0.714	6.33	-0.041	0.0017	1.0799	0.6324	0.806	7.09	0.150	0.0226	1.1022	0.2701
0.718	6.67	0.274	0.0751	1.0808	0.6092	0.810	7.18	0.216	0.0465	1.1032	0.2686

Table 4.1 (Continued)

x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{IP}^2$	$\hat{\sigma}_{iNP}^2$	x_i	y_i	r_i	r_i^2	$\hat{\sigma}_{IP}^2$	$\hat{\sigma}_{iNP}^2$
0.813	6.69	-0.299	0.0895	1.1042	0.2683	0.909	8.10	0.518	0.2679	1.1309	0.4916
0.817	6.50	-0.514	0.2640	1.1053	0.2693	0.913	8.18	0.573	0.3282	1.1320	0.5029
0.821	6.46	-0.579	0.3347	1.1063	0.2715	0.917	8.52	0.888	0.7889	1.1332	0.5137
0.825	6.35	-0.713	0.5088	1.1074	0.2748	0.921	8.56	0.903	0.8162	1.1344	0.5241
0.829	6.31	-0.778	0.6053	1.1084	0.2792	0.925	9.00	1.319	1.7391	1.1356	0.5339
0.833	6.41	-0.703	0.4938	1.1095	0.2846	0.929	9.34	1.634	2.6700	1.1368	0.5432
0.837	6.60	-0.537	0.2889	1.1106	0.2909	0.933	9.04	1.039	1.7143	1.1380	0.5518
0.841	6.57	-0.592	0.3507	1.1117	0.2982	0.937	9.08	1.325	1.7545	1.1393	0.5598
0.845	6.59	-0.597	0.3563	1.1128	0.3063	0.940	9.14	1.360	1.8492	1.1405	0.5672
0.849	6.80	-0.412	0.1694	1.1138	0.3152	0.944	8.99	1.185	1.4046	1.1417	0.5738
0.853	7.16	-0.076	0.0058	1.1149	0.3248	0.948	8.96	1.130	1.2779	1.1429	0.5798
0.857	7.51	0.249	0.0620	1.1160	0.3350	0.952	8.86	1.006	1.0115	1.1442	0.5851
0.861	7.52	0.234	0.0549	1.1171	0.3458	0.956	8.79	0.911	0.8299	1.1454	0.5896
0.865	7.40	0.090	0.0080	1.1183	0.3571	0.960	8.62	0.716	0.5131	1.1467	0.5935
0.869	7.48	0.145	0.0210	1.1194	0.3687	0.964	8.29	0.362	0.1307	1.1479	0.5966
0.873	7.42	0.060	0.0036	1.1205	0.3808	0.968	8.05	0.097	0.0094	1.1492	0.5991
0.877	7.53	0.145	0.0211	1.1216	0.3930	0.972	8.00	0.022	0.0005	1.1505	0.6009
0.881	7.75	0.341	0.1160	1.1228	0.4055	0.976	7.89	-0.113	0.0127	1.1517	0.6020
0.885	7.80	0.366	0.1339	1.1239	0.4181	0.980	7.48	-0.547	0.2996	1.1530	0.6025
0.889	7.63	0.171	0.0293	1.1250	0.4307	0.984	7.31	-0.742	0.5506	1.1543	0.6024
0.893	7.51	0.026	0.0007	1.1262	0.4432	0.988	7.42	-0.657	0.4313	1.1556	0.6017
0.897	7.49	-0.018	0.0003	1.1274	0.4557	0.992	7.51	-0.591	0.3498	1.1569	0.6004
0.901	7.64	0.107	0.0115	1.1285	0.4679	0.996	7.71	-0.416	0.1732	1.1582	0.5985
0.905	7.92	0.362	0.1313	1.1297	0.4799	1.000	7.99	-0.161	0.0259	1.1595	0.5962

Parametric approach: A plot of the yield against time in Figure 4.1 suggests linear relationship between the response variables and predictors. Further, we examine a plot of residuals against the predictors as in Figure 4.2 which shows some heterogeneity, we fit a quadratic variance function. Assuming the mean model and variance model of (3.1) and (3.5), the pseudo likelihood procedure gives the estimates as

$$\hat{\beta}_0 = 1.927486$$

$$\hat{\beta}_1 = 6.217056$$

$$\hat{\theta} = 0.076802$$

$$\hat{\sigma}^2 = 0.465633$$

Then the variance estimates for the parametric approach using pseudo likelihood procedure will be determined using the equation (3.5) which are recorded in Table 4.1 (column 5 and 11). The plot of these variance estimates against the predictors is shown in Figure 4.3.

Nonparametric approach: We plot the residuals squared against the predictors in Figure 4.4 from which we obtain a smooth curve shown in Figure 4.5. The curve has been drawn using the bandwidth $b=0.18$. The respective residuals squared on this curve gives the nonparametric variance estimates for each predictor value as recorded in Table 4.1 (column 6 and 12).

Figure 4.1: A plot of Yield against time

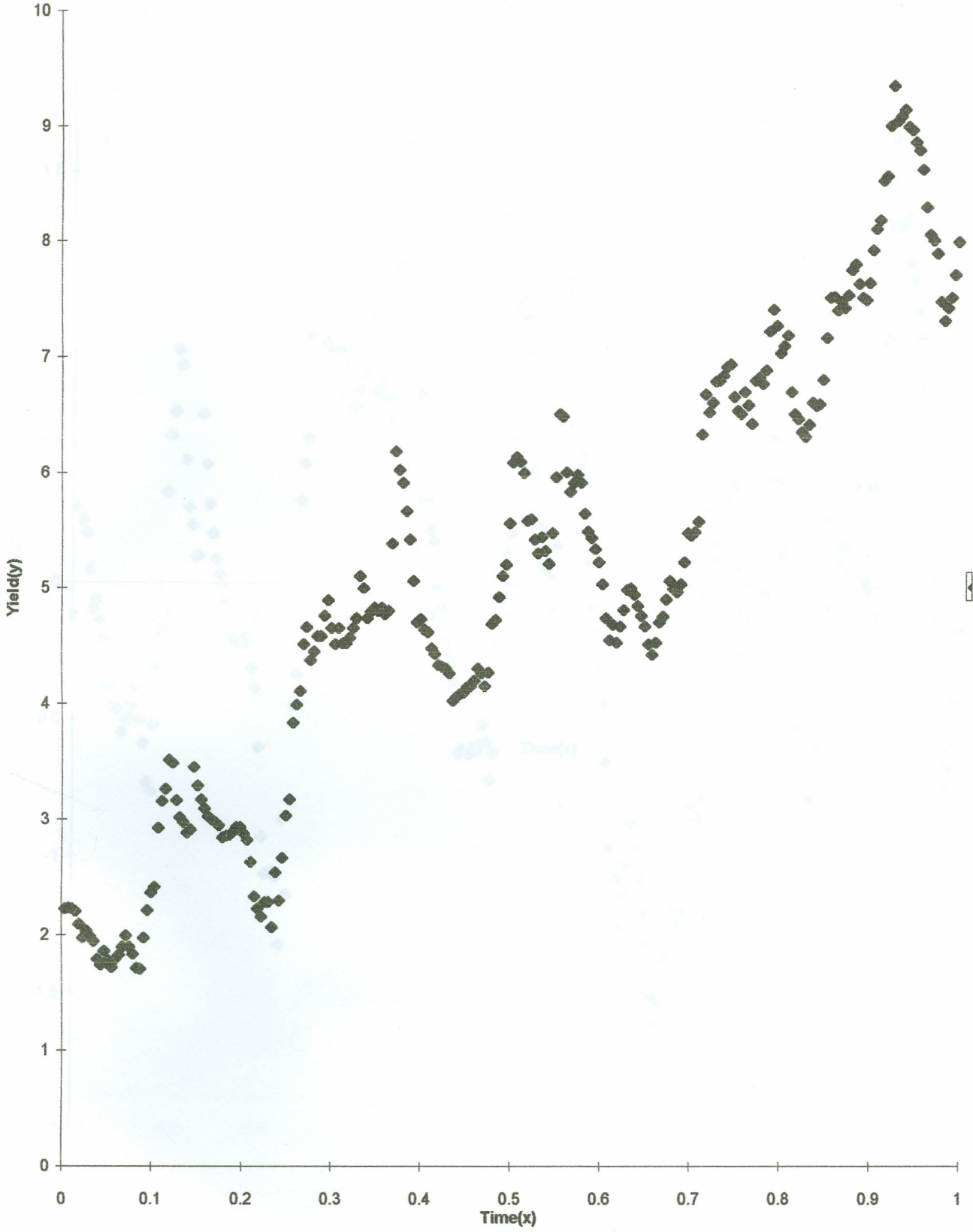


Figure 4.2: A scatter plot of residuals against time

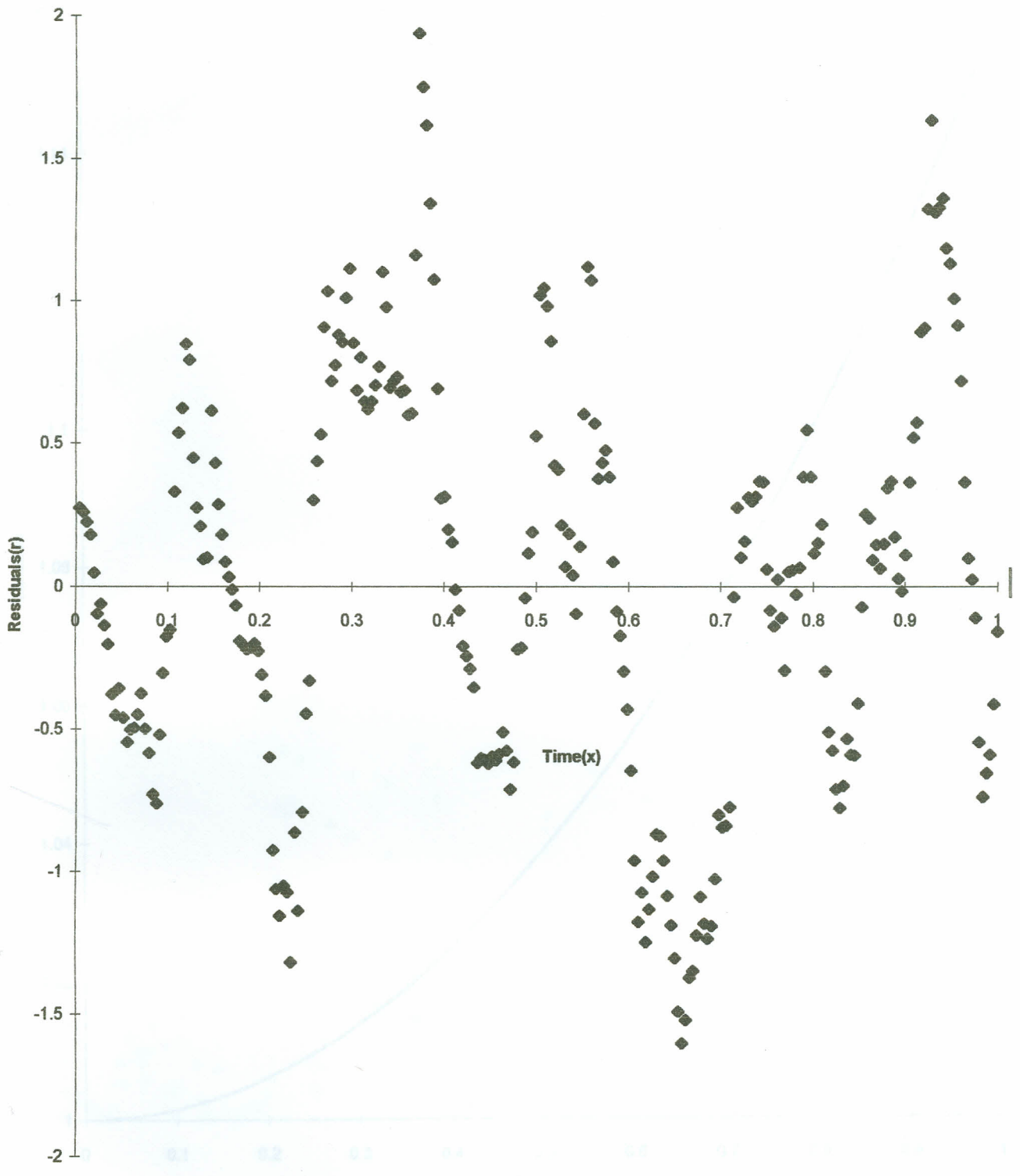


Figure 4.3: A plot of parametric variance estimates against time

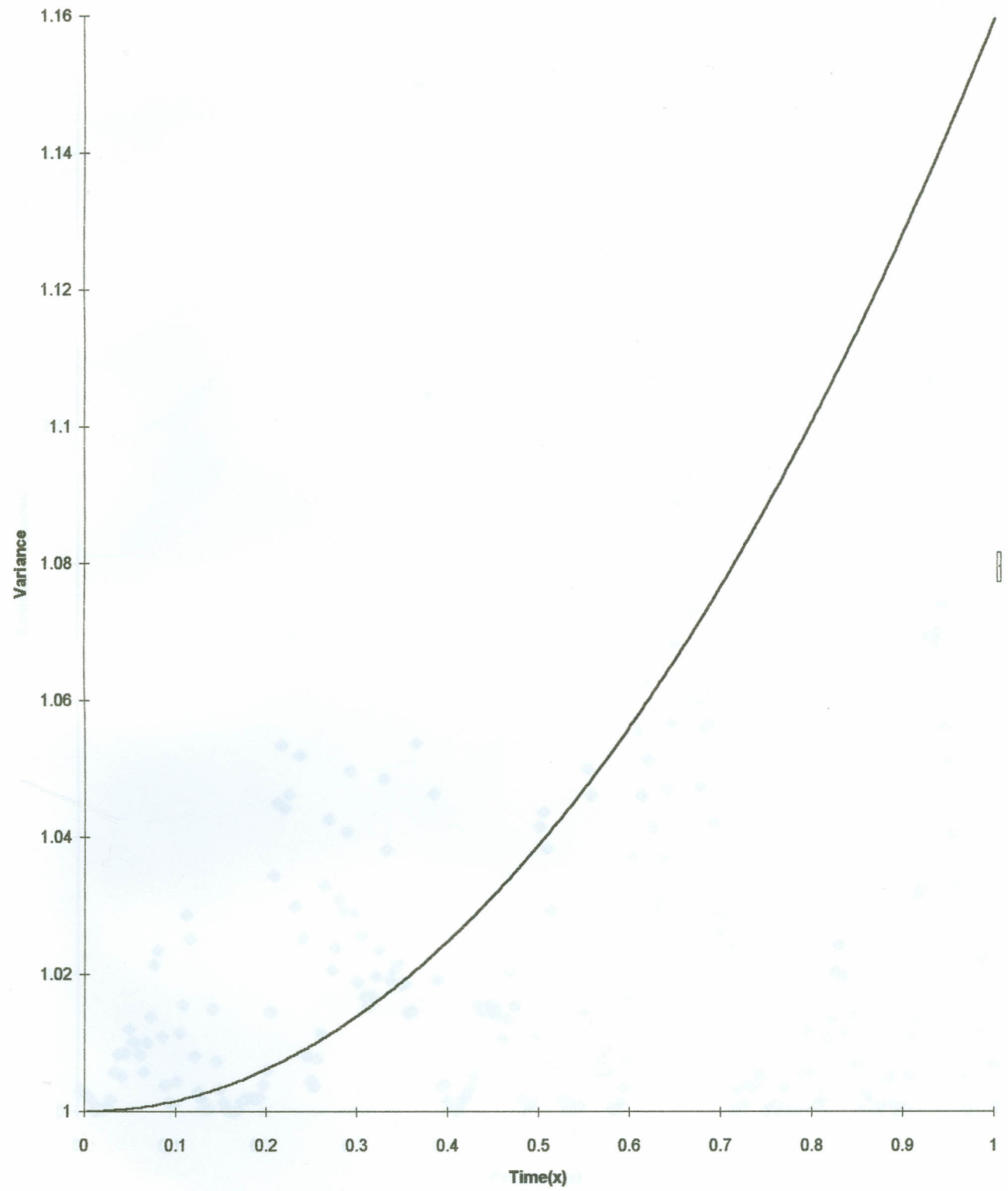


Figure 4.4: A scatter plot of squared residuals against time

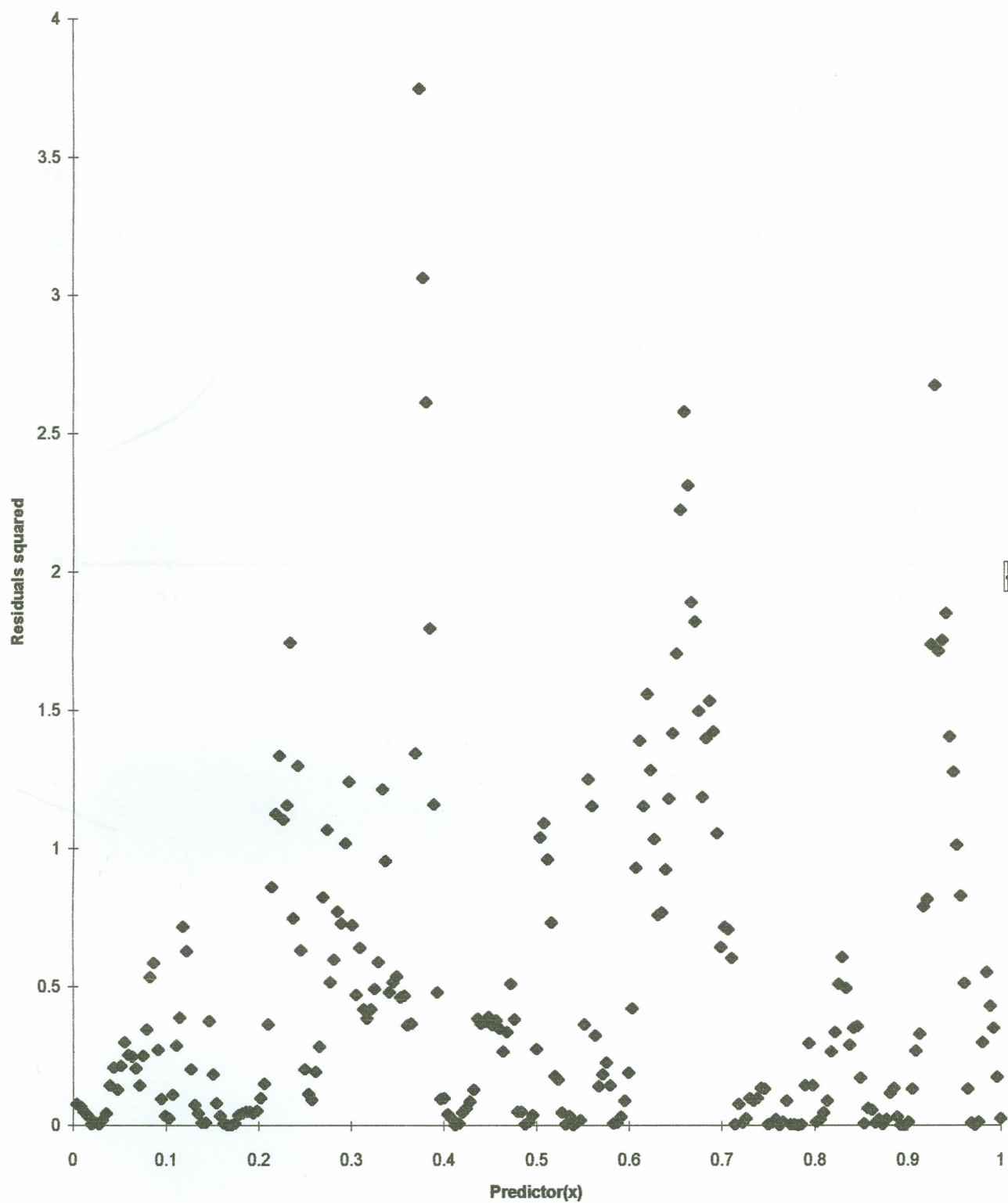
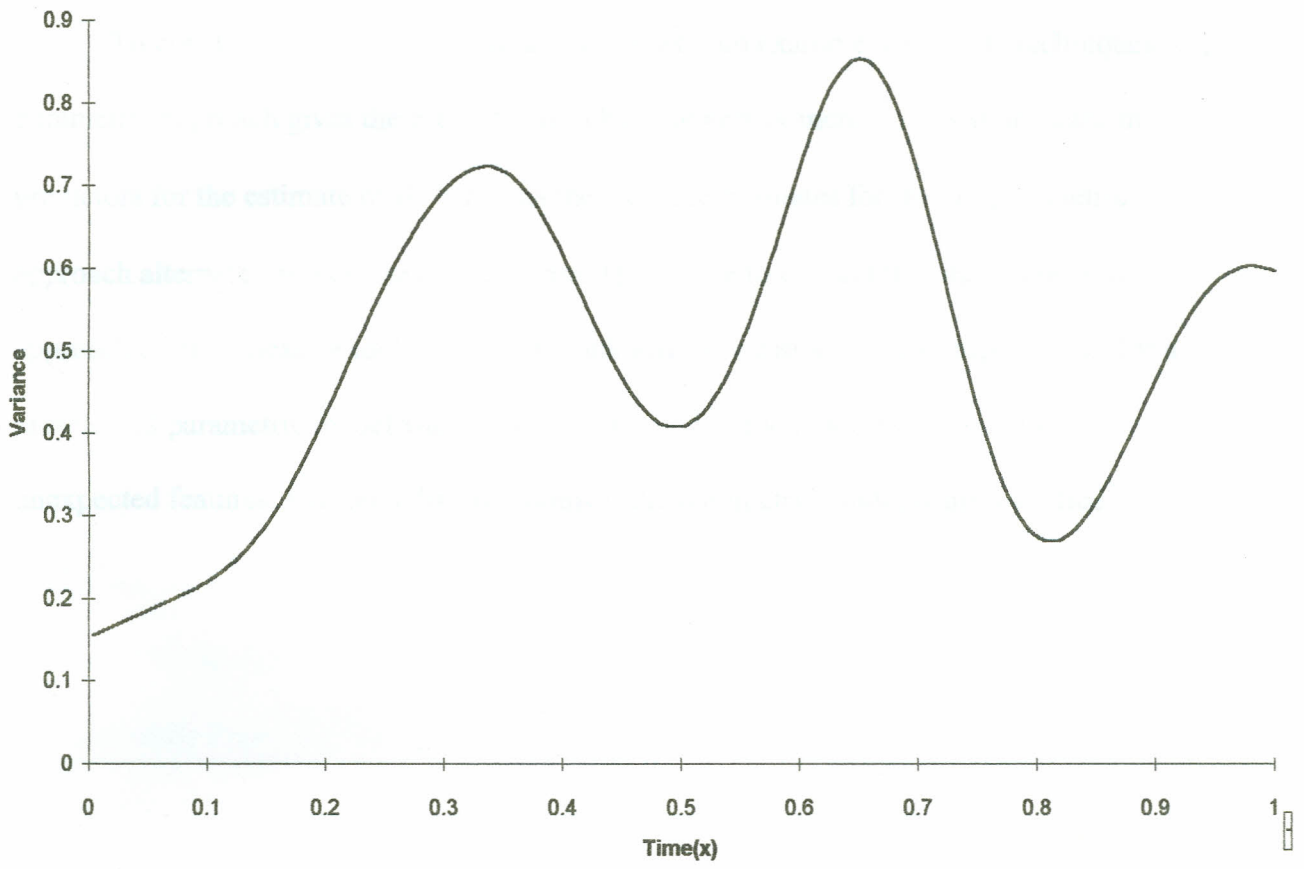


Figure 4.5: A smooth curve of nonparametric variance estimates



4.4 Conclusion:

To conclude the discussion, we use the illustrated example using both techniques. The parametric approach gives the estimates which are generally increasing with increase in predictors for the estimate of θ . Whereas the variance estimates for the nonparametric approach alternate for increase in predictors. This is due to the fact that the nonparametric approach offers a flexible tool in analyzing unknown regression relationships. Unlike for a preselected parametric model which might be limited or too low-dimensional to fit unexpected features. Further a disaster looms if the parametric model is misspecified.

BIBLIOGRAPHY

- [1] Carroll, R.J. [1982]. Adapting for heteroscedasticity in linear models. *Annals of statistics* 10, 1224-1233.
- [2] Carroll, R.J. and Ruppert, D. [1982]. Robust estimation in heteroscedastic linear models. *Annals of statistics* 10, 429-441.
- [3] Carroll, R.J. and Ruppert, D. [1987]. Variance function estimation. *Journal of the American Statistical Association* 82, 1079-1091.
- [4] Carroll, R.J. and Ruppert, D. [1988]. *Transformations and Weighting in Regression*. Chapman and Hall, London.
- [5] Haerdle, W. [1993]. *Applied Nonparametric Regression*, Cambridge, U. K. , Cambridge University Press.
- [6] 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.
- [7] Silverman, B. W. [1986]. *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- [8] Weisberg, S. [1985]. *Applied linear regression*. Wiley, New York.