

Statistica Sinica 8(1998), 445-465

THE EXAMINATION OF RESIDUAL PLOTS

Chih-Ling Tsai, Zongwu Cai and Xizhi Wu

*University of California, Southwest Missouri State University
and Nankai University*

Abstract: Linear and squared residual plots are proposed to assess nonlinearity and heteroscedasticity in regression diagnostics. It is shown that linear residual plots are useful for diagnosing nonlinearity and squared residual plots are powerful for detecting nonconstant variance. A paradigm for the graphical interpretation of residual plots is presented.

Key words and phrases: Heteroscedasticity, leverages, nonlinearity, outliers.

1. Introduction

Over the last three decades, residual plots (plots of residuals versus either the corresponding fitted values or explanatory variables) have been widely used to detect model inadequacies in regression diagnostics (see Anscombe (1961), Draper and Smith (1966), Atkinson (1985), Carroll and Ruppert (1988), Chatterjee and Hadi (1988) and Cook and Weisberg (1982, 1994)). This type of plot is often referred to as a “linear residual plot” since its y-axis is a linear function of the residual. In general, a null linear residual plot shows that there are no obvious defects in the model, a curved plot indicates nonlinearity, and a fan-shaped or double-bow pattern indicates nonconstant variance (see Weisberg (1985), and Montgomery and Peck (1992)). However, Cook (1994) recently presented examples which show that linear residual plots can be misleading, and may not be sufficiently powerful by themselves to detect nonlinearity or heteroscedasticity. This provides us with a motivation to develop a different perspective for the interpretation of linear regression residual plots. While the primary assumptions for linear regression modelling are normality, linearity, homoscedasticity, and independence, in this paper we will focus only on the use of residual plots to examine those for homoscedasticity and linearity.

Cook and Weisberg (1983) proposed both a graphical method and a score test to improve the assessment of nonconstant variance. The informal graphic method they suggested plots squared studentized residuals versus the first derivative of the weighting variables or the fitted values. Since the y-axis of this plot is a function of the squared residual, we refer to it as a “squared residual plot”. A wedge-shaped pattern in a squared residual plot is taken as evidence that

the variance depends on the quantity plotted on the abscissa; however, a wedge-shaped pattern is not always present when variance is nonconstant. Furthermore, the first derivative of the weighting function is not always nonzero for all cases of nonconstant variance, and this plot cannot reflect the weighting function directly. Thus, we propose an alternative squared residual plot which does in fact directly reflect the weighting function. This alternative plot can be viewed as a complement to Cook and Weisberg's plot, and for best data analysis results the two should both be used, together with the formal score test, to assess heteroscedasticity.

In order to detect nonlinearity, it is usual to plot residuals against either fitted values or explanatory variables. However, because Cook (1994, Examples 7.1 and 7.2) has shown that this type of plot may provide misleading information when fitted values are used, we therefore suggest using the linear residual plot (residuals versus explanatory variables case) for the detection of nonlinearity, and we provide theoretical justification to support this method.

One of the most challenging problems in assessing nonlinearity of a regression function arises when covariates are dependent on each other. Cook (1993) generalized the partial residual plot (Ezekiel (1924)) to obtain the CERES (Combined Conditional Expectations and RESiduals) plot. However, the CERES plot sometimes does not clearly reveal nonlinear components of a regression function, and so analogously we have generalized the linear residual plot to obtain the CLRES (Conditional Linear RESiduals) plot. But since neither the partial residual nor linear residual plots from which these plots are derived are perfect in their assessment of nonlinearity, we recommend that CERES and CLRES plots be used together to detect nonlinearity.

The aim of this paper is to provide a systematic way to interpret residual plots when evaluating heteroscedasticity and nonlinearity in regression analysis. This does not imply that we have a single graphical recipe which can identify all possible patterns of residual plots resulting from nonconstant variance or nonlinearity, but we can provide guidelines. Based on both theoretical justification and the analysis of examples, we will show that squared residual plots are superior to linear residual plots for assessing nonconstant variance. By contrast, linear residual plots are most appropriate for examining nonlinearity.

For the case when the true model involves nonconstant variance, but the fitted model assumes both linearity and constant variance, we will begin by obtaining the first and the second moments of residuals for given fitted values or explanatory variables (Section 2). Linear and squared residual plots are also investigated through analytical examples. Section 3 examines linear and squared residual plots when the true model includes nonlinear components, but the fitted model is linear with constant variance. Section 4 presents three analytical examples that illustrate the impact of sample size, the strength of the weighting function and nonlinearity, high leverage points, and the effect of outliers and

interdependent covariates on the pattern of residual plots. In addition, two examples are given to elucidate the interpretation of residual plots: the Speed-Braking Distance example (Ezekiel and Fox (1959), p. 45), and the Land Rent example (Cook and Weisberg (1994), p. 90). Section 5 gives concluding remarks.

2. Residual Plots for Examining Heteroscedasticity

For linear regression models, Cook and Weisberg (1983) obtained a score test to diagnose the assumption of homoscedasticity, and in addition they suggested using their squared residual plot to examine the heteroscedasticity. However, neither the score test nor Cook and Weisberg's residual plot can characterize the weighting function directly. To illustrate this, in this section we apply linear, squared, and Cook and Weisberg's squared residual plots to assess heteroscedasticity. While linear residual plots can characterize the square root of the weighting function, squared residual plots directly identify the weighting function itself. However, although it cannot reveal the weighting function directly, Cook and Weisberg's squared residual plot can be used to diagnose heteroscedasticity. In practice, both the formal score test and informal graphical residual plots should be used as counterparts in diagnosing heteroscedasticity (see Examples 4.3 and 4.4). Since the focus of this paper is on graphical plots, we will not investigate the properties of the score test. Interested readers may refer to an article which compares score tests for heteroscedasticity written by Lyon and Tsai (1996). In the section that follows, we introduce true model structures with nonconstant variance, fitted model structures assuming constant variance, and the three residual plots mentioned above that result for the fitted models.

2.1. Model structure and moments of residuals

The usual linear regression model can be represented as

$$\begin{aligned} Y &= X\beta + e, \\ e &\sim N(0, \sigma^2 I), \end{aligned} \tag{1}$$

where $Y = (y_1, \dots, y_n)'$, $e = (e_1, \dots, e_n)'$, $X = (x_1', \dots, x_n')'$, $x_i = (x_{i1}, \dots, x_{ip})'$ is a known vector for $i = 1, \dots, n$, β is a $p \times 1$ vector of unknown parameters, I is an $n \times n$ identity matrix, and σ^2 is an unknown constant. Based on model (1), the ordinary least squares estimator of β is $\hat{\beta} = (X'X)^{-1}X'Y$ and the unbiased estimator of σ^2 is $\hat{\sigma}^2 = Y'(I - H)Y/(n - p)$, where $H = X(X'X)^{-1}X'$. The fitted values and residuals that result are $\hat{Y} = (\hat{y}_1, \dots, \hat{y}_n)' = HY$ and $\hat{e} = (\hat{e}_1, \dots, \hat{e}_n)' = (I - H)Y$, respectively.

In regression analysis, it is not unusual for the underlying error variance to be nonconstant. In other words, the true error covariance matrix is

$$\sigma_0^2 W, \tag{2}$$

where σ_0^2 is the true scale parameter, W is an $n \times n$ diagonal matrix with i th entry $w_i = w(z_i, \delta)$, $z_i' = (z_{i1}, \dots, z_{iq})$, the i th row of an $n \times q$ known matrix Z , and δ is a $q \times 1$ vector.

If we fit the data with model (1), but in fact the underlying error covariance matrix is (2), then

$$E(\hat{e}_i | z_i) = 0, \quad (3)$$

$$\text{Var}(\hat{e}_i | z_i) = E(\hat{e}_i^2 | z_i) = \left\{ (1 - h_{ii})^2 w_i + \sum_{j \neq i} h_{ij}^2 w_j \right\} \sigma_0^2, \quad (4)$$

and

$$\text{Var}(\hat{e}_i^2 | z_i) = 2\{\text{Var}(\hat{e}_i | z_i)\}^2, \quad i = 1, \dots, n. \quad (5)$$

The vertical bar in the expression $(\cdot | \cdot)$ stands for the word ‘‘given’’, but it also means ‘‘conditioning on’’ if the quantity after the vertical bar represents specific values of random variables. Also, $h_{ij} = x_i'(X'X)^{-1}x_j$, $h_{ii} = x_i'(X'X)^{-1}x_i$, and h_{ii} is the i th leverage of H . We can further show that (\hat{e}_i, \hat{y}_i) is distributed as a bivariate normal distribution:

$$N\left(\begin{pmatrix} 0 \\ x_i'\beta \end{pmatrix}, \begin{pmatrix} (1 - h_{ii})^2 w_i + \sum_{j \neq i} h_{ij}^2 w_j & h_{ii} w_i - \sum_j h_{ij}^2 w_j \\ h_{ii} w_i - \sum_j h_{ij}^2 w_j & \sum_j h_{ij}^2 w_j \end{pmatrix} \sigma_0^2\right).$$

Hence,

$$E(\hat{e}_i | \hat{y}_i) = \left(\frac{h_{ii} w_i}{\sum_j h_{ij}^2 w_j} - 1 \right) (\hat{y}_i - x_i'\beta), \quad (6)$$

$$\text{Var}(\hat{e}_i | \hat{y}_i) = (1 - a_i) w_i \sigma_0^2, \quad (7)$$

$$E(\hat{e}_i^2 | \hat{y}_i) = \text{Var}(\hat{e}_i | \hat{y}_i) + \{E(\hat{e}_i | \hat{y}_i)\}^2, \quad (8)$$

and

$$\text{Var}(\hat{e}_i^2 | \hat{y}_i) = 2\{\text{Var}(\hat{e}_i | \hat{y}_i)\}^2, \quad (9)$$

where

$$a_i = \frac{h_{ii}^2 w_i}{\sum_j h_{ij}^2 w_j}. \quad (10)$$

Since $E(\hat{e}_i^2 | z_i)$ depends on $(1 - h_{ii})$ even when $W = I$, Cook and Weisberg (1983) suggest replacing \hat{e}_i with $b_i = \hat{e}_i / \sqrt{1 - h_{ii}}$. In this paper, the term ‘‘linear residual plot’’ will refer to the plot of a linear function of \hat{e}_i versus a function of \hat{y}_i (or z_i). The term ‘‘squared residual plot’’ will refer to a plot of a squared function of \hat{e}_i versus a function of \hat{y}_i (or z_i).

2.2. Linear residual plots

If it is suspected that the variance is a function of the weighting variable z_i , a conventional graphical method is to plot b_i (or \hat{e}_i) versus z_i . As can be seen

from equations (3) and (4), the mean of $b_i|z_i$ is zero. This implies that the shape of this plot is only determined by the variance,

$$\begin{aligned} \text{Var}(b_i|z_i) &= E(b_i^2|z_i) \\ &= \left\{ (1 - h_{ii})w_i + \sum_{j \neq i} h_{ij}^2 w_j / (1 - h_{ii}) \right\} \sigma_0^2. \end{aligned} \tag{11}$$

In order to view the weighting function w more clearly, we suggest removing the factor $(1 - h_{ii})$ from the first term in braces of equation (11). We do this by plotting $\hat{e}_{(i)} = b_i/\sqrt{1 - h_{ii}}$ versus z_i , where $\hat{e}_{(i)} = y_i - x_i'\hat{\beta}_{(i)}$, $\hat{\beta}_{(i)}$ is the least squares estimator with the i th case excluded, and $\hat{e}_{(i)}$ is Cook and Weisberg's (1982), p. 33 "predicted residual."

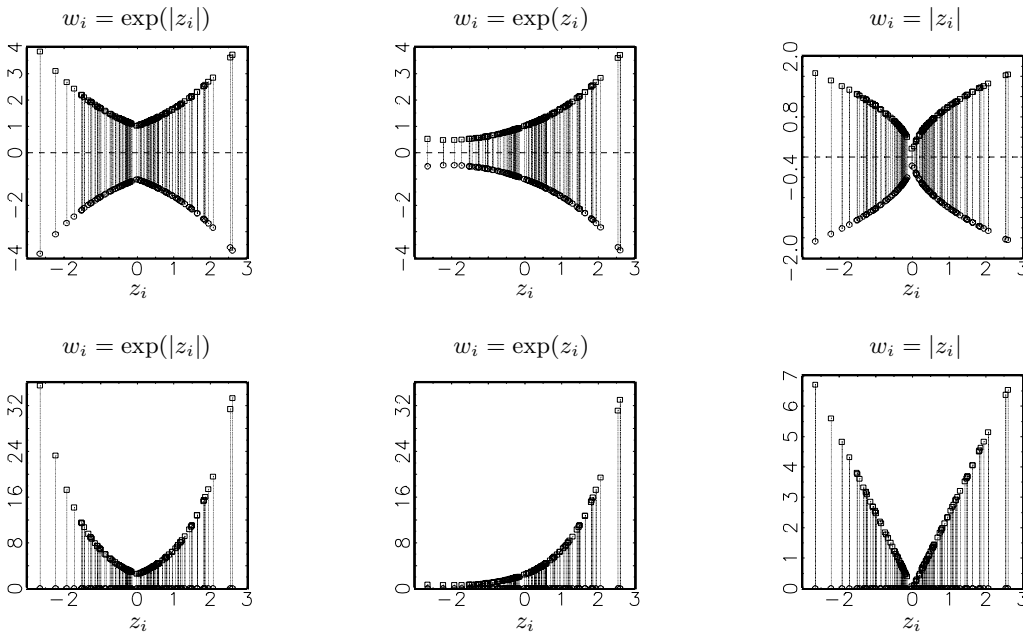


Figure 1. Linear (row 1) and squared (row 2) residual plots for assessing heteroscedasticity. The labels on the Y-axes of the linear and squared residual plots are $(-s.e.(\hat{e}_{(i)}|z_i), s.e.(\hat{e}_{(i)}|z_i))$ and $(0, E(\hat{e}_{(i)}^2|z_i) + s.e.(\hat{e}_{(i)}^2|z_i))$, respectively.

Figure 1 (row 1) presents linear residual plots obtained by plotting $0 + s.e.(\hat{e}_{(i)}|z_i)$ and $0 - s.e.(\hat{e}_{(i)}|z_i)$ versus z_i for three different weighting factors: $w_i = \exp(|z_i|)$, $w_i = \exp(z_i)$, and $w_i = |z_i|$. The standard error is denoted by $s.e.$, $z_i (i = 1, \dots, 100)$ are randomly generated from $N(0, 1)$, $\sigma_0 = 1$, $h_{ii} = x_i'(X'X)^{-1}x_i$, $x_i = (1, z_i)'$, and $X = (x_1, \dots, x_n)'$. The top and the bottom of each plot correspond to $0 + s.e.(\hat{e}_{(i)}|z_i)$ and $0 - s.e.(\hat{e}_{(i)}|z_i)$, respectively. Linear

residual plots appear to be symmetric about zero in the presence of nonconstant variance.

The second term in braces of equation (11) is $\sum_{j \neq i} (h_{jj(i)} - h_{jj})w_j$, where $h_{jj(i)}$ is the j th leverage after x_i is deleted from X . If this term is ignored, then the shape of the linear residual plot is determined by $\sqrt{w_i}$. We have found little difference in the plots that result between ignoring this term, or keeping it and using the entire term in braces in (11).

If the variance is suspected to be a function of expected response, then equations (6) and (7) indicate that the shape of the linear residual plot (\hat{e}_i versus \hat{y}_i) is determined by $E(\hat{e}_i|\hat{y}_i) \pm s.e.(\hat{e}_i|\hat{y}_i)$. The conditional mean, $E(\hat{e}_i|\hat{y}_i)$, is not completely known since $x'_i\beta$ is unknown (see equation (6)). If $x'_i\beta$ is replaced by $x'_i\hat{\beta}$, then the shape of the linear residual plot can be approximated by plotting $0 \pm s.e.(\hat{e}_i|\hat{y}_i)$ versus \hat{y}_i . In addition, if a_i in equation (7) is ignored, then the shape of the plot will be a function of $\sqrt{w_i}$.

2.3. Squared residual plots

Since linear residual plots cannot identify the weighting function directly, we recommend plotting the squared residual, $\hat{e}_{(i)}^2$, versus z_i to detect heteroscedasticity. The pattern of this plot is determined by $E(\hat{e}_{(i)}^2|z_i) + s.e.(\hat{e}_{(i)}^2|z_i)$, where $s.e.(\hat{e}_{(i)}^2|z_i) = \sqrt{2}\text{Var}(\hat{e}_{(i)}|z_i)$. The lower bound of this plot is zero, as $\hat{e}_{(i)}^2$ is a nonnegative value. Figure 1 (row 2) displays three squared residual plots which correspond to the weighting functions $\exp(|z_i|)$, $\exp(z_i)$, and $|z_i|$, respectively. If the second term of (11) is ignored, then $s.e.(\hat{e}_{(i)}^2|z_i) = \sqrt{2}w_i$. Hence, the pattern of the squared residual plot can identify the weighting function w_i directly, and is better suited to assess the adequacy of the nonconstant variance assumption than linear residual plots. Note that $E(\hat{e}_{(i)}^2|z_i) = 1/(1 - h_{ii})$ when $W = I$. Since $0 \leq h_{ii} \leq 1$, and it is non-increasing in n for fixed p , the factor h_{ii} usually does not play a significant role in assessing the function form of W , even when $W = I$.

If the variance is suspected to be a function of the expected response, we recommend plotting \hat{e}_i^2 versus \hat{y}_i . The shape of this plot is determined by $E(\hat{e}_i^2|\hat{y}_i) + s.e.(\hat{e}_i^2|\hat{y}_i)$, where $E(\hat{e}_i^2|\hat{y}_i)$ and $s.e.(\hat{e}_i^2|\hat{y}_i)$ are defined in equations (8) and (9). If we again replace $x'_i\beta$ (see equation (6)) by $x'_i\hat{\beta}$, and ignore the component a_i (see equation (7)), the squared residual plot will characterize the shape of the weighting function w_i (see equations (7) and (9)). As above, we therefore prefer the squared residual plot to the linear residual plot for the elucidation of heteroscedasticity.

2.4. Cook and Weisberg's squared residual plots

Cook and Weisberg (1983) proposed replacing $w(z_i, \delta)$ in equation (11) by

$w(z_i, \delta^*) + (\delta - \delta^*)\dot{w}_i$, where $\dot{w}_i = \left. \frac{\partial w(z_i, \delta)}{\partial \delta} \right|_{\delta=\delta^*}$ and $w(z_i, \delta^*) = 1$. This yields

$$E\left(b_i^2/\sigma_0^2|z_i\right) \simeq 1 + (\delta - \delta^*)\left\{(1 - h_{ii})\dot{w}_i + \sum_{j \neq i} \dot{w}_j h_{ij}^2/(1 - h_{ii})\right\}, \quad (12)$$

which is Cook and Weisberg's (1983) equation (16). Thus, if the second term in braces of (12) is ignored, the shape of the squared residual plot of b_i^2/σ_0^2 versus $(1 - h_{ii})\dot{w}_i$ will be characterized by $\{1 + (\delta - \delta^*)(1 - h_{ii})\dot{w}_i\} + s.e.(b_i^2/\sigma_0^2|z_i)$, where $s.e.(b_i^2/\sigma_0^2|z_i) = \sqrt{2}(1 - h_{ii})w_i/\sigma_0^2 = \sqrt{2}(1 - h_{ii})(1 + (\delta - \delta^*)\dot{w}_i)/\sigma_0^2$ (see equations (5) and (11)). The lower bound of this plot is zero, since b_i^2/σ_0^2 is a nonnegative value. This implies that the resulting squared residual plot will show a wedge-shaped pattern in the case of nonconstant variance. This is the main difference between Cook and Weisberg's squared residual plot and the plot discussed in Section 2.3. While Cook and Weisberg's plot has a wedge shape when nonconstant variance is present, the squared residual plot in Section 2.3 directly identifies the weighting function w_i . If \dot{w}_i is zero, then Cook and Weisberg's squared residual plot cannot be determined. If this is the case, we suggest plotting b_i^2/σ_0^2 versus $(1 - h_{ii})\ddot{w}_i$, where $\ddot{w}_i = \left. \frac{\partial^2 w(z_i, \delta)}{\partial \delta \partial \delta'} \right|_{\delta=\delta^*}$. As above, the plot will still have a wedge-shaped pattern (See Example 2, Section 4).

In practice there may be additional complications. Outliers may appear in the data set that mask the pattern of squared or Cook and Weisberg's squared residual plots. There are two possible remedies for this problem: the first is to adopt Carroll and Ruppert's (1988), p. 30 suggestion to plot the function of residuals instead of the square of the residuals. Using this concept, we propose plotting $\log(1 + (\hat{e}_{(i)}/\hat{\sigma})^2)$ versus x_i . If $(\hat{e}_{(i)}/\hat{\sigma})^2$ is small, then $\log(1 + (\hat{e}_{(i)}/\hat{\sigma})^2) \simeq (\hat{e}_{(i)}/\hat{\sigma})^2$. Hence, this log transformation will usually maintain the original shape of the squared residual plot after shrinking outliers. In other words, this transformation can prevent outliers from obscuring a nonconstant variance pattern (see Example 1, Section 4). The second possibility is to use a statistical software package to interactively rescale the plot to the bulk of the data, effectively deleting the outliers from the plot. For example, Data Desk (1995) has this capability. Due to limitations in our computing facility, we are unable to illustrate this method in this paper.

When both the true and fitted models have constant variance, equations (4) through (9) indicate that both the linear and squared residual plots will not reveal any pattern. We refer to these as null plots.

3. Residual Plots for Assessing Nonlinearity

In this section, we consider the mean function of the fitted and the underlying true models to be

$$X\beta + Z\gamma \text{ and } X\alpha + g(Z), \quad (13)$$

respectively, where X and β are defined as in equation (1), α is a $p \times 1$ vector, and the quantities Z and γ are adopted from equation (2) with dimension $q = 1$. Furthermore, we assume that the errors for the fitted and true models e_i , $i = (1, \dots, n)$, are independent and identically distributed as $N(0, \sigma^2)$, and $N(0, \sigma_0^2)$, respectively. Note that Z in Section 2 is used in reference to the weighting function, but here is used in reference to the mean function. In practice, it may not be known if there are problems with nonlinearity or heteroscedasticity with respect to any particular covariate; hence, we use the same notation (Z) for examining residual plots.

In subsections 3.1 and 3.2, we study the linear, partial, and squared residual plots where both X and Z are fixed, and discuss how to extend our results to the case where both variables X and Z are random, at the end of subsection 3.2. Also, in subsection 3.3, we examine Cook's (1993) partial residual plot and propose a complementary plot for use in assessing nonlinearity.

3.1. Linear residual plots

After straightforward computations, we can show that

$$E(\hat{e}_{(i)}|z_i) = g(z_i) - \sum_{j \neq i} h_{ij}g(z_j)/(1 - h_{ii}), \quad (14)$$

$$\text{Var}(\hat{e}_{(i)}|z_i) = \sigma_0^2/(1 - h_{ii}), \quad (15)$$

$$E(\hat{e}_{(i)}^2|z_i) = \sigma_0^2/(1 - h_{ii}) + \{E(\hat{e}_{(i)}|z_i)\}^2, \quad (16)$$

and

$$\text{Var}(\hat{e}_{(i)}^2|z_i) = 2\sigma_0^4/(1 - h_{ii})^2, \quad \text{and} \quad i = 1, \dots, n, \quad (17)$$

where $\hat{e}_{(i)}$ and h_{ii} are defined as in Section 2, with the exception that the mean function of the fitted model, $X\beta$, is replaced by $X\beta + Z\gamma$. Specifically, $\hat{e} = (\hat{e}_1, \dots, \hat{e}_n)' = (I - H)Y$, $\hat{e}_{(i)} = \hat{e}_i/(1 - h_{ii})$, where h_{ii} is the i th leverage of H , $H = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ and $\tilde{X} = (X, Z)$. In addition, by following the techniques used in Section 2.1 to derive equation (6) we can show that

$$E(\hat{e}_{(i)}|\hat{y}_i) = E(\hat{e}_{(i)}|z_i). \quad (18)$$

If we suspect that the mean function is nonlinear in z , then the linear residual plot of $\hat{e}_{(i)}$ versus z_i will usually reveal the nonlinear form, since the pattern of the linear residual plot is determined by $E(\hat{e}_{(i)}|z_i) \pm s.e.(\hat{e}_{(i)}|z_i)$. This means that the resulting linear residual plots are not symmetric about zero, unlike the linear residual plots in Figure 1. If the second term of equation (14) is ignored, then the pattern of the linear residual plot depends on $g(z_i)$. We have found

little difference in the plots that result between using this quantity and using the entire term in (14).

If $g(Z)$ is not a function of $X\beta$, then equations (14) and (18) indicate that the plot of $\hat{e}_{(i)}$ (or \hat{e}_i) versus \hat{y}_i might not reveal a nonlinear pattern even though the true mean function includes the nonlinear component, $g(Z)$. Cook's Example 7.1 (1994) illustrates this point. In practice, the linear residual plot may not clearly characterize the nonlinear function (see Mansfield and Conerly (1987) and Berk and Booth (1995)). If this is the case, there are a few possible alternative plots that can be used to examine nonlinearity. In this paper we consider only one of these, the partial residual plot (see Ezekiel (1924), Larsen and McCleary (1972), and Mansfield and Conerly (1987)), which is constructed by plotting

$$\hat{e} + Z\hat{\gamma} \quad \text{versus} \quad Z, \quad (19)$$

where $\hat{e} = Y - X\hat{\beta} - Z\hat{\gamma}$. There are two reasons for choosing to examine the partial residual plot as well as the linear residual plot:

- (1) It has the ability to detect nonlinearity (see Mansfield and Conerly (1987)).
- (2) It has recently been generalized and explored by Cook (1993), which leads us to adopt his approach to the exploration of the linear residual plot (Section 3.3), and also leads to a plot which complements Cook's.

We usually do not know beforehand whether the residual or partial residual plot will provide a more accurate characterization of the function g . If both plots show the same pattern, we can comfortably draw conclusions as to the nature of g . Otherwise, formal tests for nonlinearity (see Su and Wei (1991), Eubank and Hart (1992), and Azzalini and Bowman (1993)) should be considered.

3.2. Squared residual plots

Based on equations (16) and (17), the pattern of squared residual plots ($\hat{e}_{(i)}^2$ versus z_i) is determined by $E(\hat{e}_{(i)}^2|z_i) + s.e.(\hat{e}_{(i)}^2|z_i)$. This does not directly reflect the pattern of the nonlinear component $g(z_i)$, but reflects the squared pattern of $g(z_i)$. Hence, the squared residual plot can supplement the linear residual plot in identifying the nonlinear pattern, but overall we recommend using the linear residual plot to assess nonlinearity since it directly identifies the form of the nonlinear function.

We have discussed linear and partial residual plots for the case where both variables X and Z are fixed. However, if both variables are random, we need to adopt the assumption that $E(e|X, Z) = 0$ (Cook (1993)) in order to evaluate residual plots. Under this assumption, equations (14) and (15) will be valid when z_i (on the right hand side of the $(\cdot|)$ expression) is replaced by X and Z , and hence the resulting linear and partial residual plots can be used to assess

nonlinearity. Analogous modifications can be made to obtain residual plots when one variable X or Z is random and the other is fixed.

3.3. Conditional linear residual plots

We now consider linear and partial residual plots when both X and Z are random variables. We first review Cook's version of the partial residual plot, and then propose a modified version of the linear residual plot which can be used as a complement to Cook's plot. Continue to assume that $E(e|X, Z) = 0$.

For the case where g is not a linear function, Cook (1993) generalized partial residual plots (see equation (19)) to obtain CERES plots, which can be used to identify nonlinearity in g . Specifically, Cook's CERES plot is constructed by fitting the model $Y = X\beta + m(Z)\gamma^* + \text{error}$, and then plotting $\hat{e}^* + m(Z)\hat{\gamma}^*$ versus Z , where $m(Z) = E(X|Z) - E(X)$, $\hat{e}^* = Y - X\hat{\beta} - m(Z)\hat{\gamma}^*$, and $\hat{\beta}$ and $\hat{\gamma}^*$ are the least squares estimators of β and γ^* , respectively. In practice, $m(Z)$ is often unknown. Cook suggested using the *lowess* (locally weighted scatterplot smoother) smoothing technique to estimate $m(Z)$. Useful references on this topic can be found in Cleveland (1979), Härdle (1990), p. 192 and Cook and Weisberg (1994), p. 31.

When both variables X and Z are fixed, Mansfield and Conerly (1987) used a few examples to illustrate the strong and weak performance of residual and partial residual plots. For instance, their Figure 3 shows that the partial residual plot performs better than the residual plot with respect to nonlinearity, whereas their Figure 5 indicates that the residual plot performs better than the partial residual plot. Recently, Berk and Booth (1995), Figure 9 have illustrated that Cook's CERES plot may not always be able to detect nonlinearity. Because of this finding and the fact that Cook's CERES plot is an extension of the partial residual plot, we suggest making the natural extension of the linear residual plot to obtain CLRES plots (Conditional Linear RESiduals) by plotting \hat{e}^* versus Z .

Cook (1993) applied his Lemma 4.1 to obtain the population version of the CERES plot, $g(Z) - E(g(Z)) + e$ versus Z , assuming that $E(X) = E(Z) = 0$. To study the property of CLRES plot, we first let $(\tilde{\beta}', \tilde{\gamma}^{*'}) = \arg \min R(\beta', \gamma^{*'})$, where $R(\beta', \gamma^{*'}) = E\{(Y - X\beta' - m(Z)\gamma^{*'})'(Y - X\beta' - m(Z)\gamma^{*'})/n\}$, and the expectation is computed with respect to the joint distribution of Y and (X, Z) . Then, adopting Cook's (1993), pages 353 and 355 results and arguments, we can show that $(\hat{\beta}', \hat{\gamma}^{*'})$ is the Fisher consistent estimator of $(\tilde{\beta}', \tilde{\gamma}^{*'})$, and the resulting population version of the CLRES plot is $g(Z) - E(g(Z)) - m(Z)\tilde{\gamma}^{*'} + e$ versus Z . Now, we address the reason that the CLRES plot is a useful adaptation: suppose that the component $m(Z)\tilde{\gamma}^{*'}$ dominates the function $g(Z)$ over the given range of Z , but does not reflect the true relationship between $g(Z)$ and Z . If this is the case, then removing the overlapping effect $m(Z)\tilde{\gamma}^{*'}$ from $g(Z)$ (i.e. $g(Z) - m(Z)\tilde{\gamma}^{*'}$)

will better portray the correct form of g . This allows the CLRES plot to perform better than the CERES plot in identifying the true relationship between $g(Z)$ and Z . From the discussion above, we conclude that the respective role of the CLRES and CERES plots parallel those of the residual and partial residual plots; that is, the CLRES plots complement Cook's CERES plots, and both should be used to obtain the best assessment of nonlinearity when $E(X|Z)$ is not linear in Z . An analytical example (Example 3) is given in Section 4 to illustrate the combined use of CERES and CLRES plots.

The residuals $\hat{e}_{(i)}$, b_i , and \hat{e}_i discussed thus far in Sections 2 and 3 are not unit free. Subsequently, we will divide them by $\hat{\sigma}$ for the purposes of illustrating the process of data analysis. The quantity $b_i/\hat{\sigma}$ is conventionally called the studentized residual, which we denote by r_i . We can also view r_i as a replacement for b_i/σ_0 in Section 2.4, since σ_0 is not known in practice.

4. Examples

Three analytical examples are presented here to illustrate the effects of: (1) sample size, (2) the strength of the weighting function and nonlinearity, high leverages, and outliers, and (3) dependent covariates on the patterns of linear and squared residual plots. Two practical examples are also given. Our purpose in presenting these examples is to demonstrate the use of linear and squared residual plots to diagnose nonlinearity and nonconstant variance problems, and then to use this diagnostic information to find a reasonable model to describe the given data set. However, this does not imply our final model is the best choice.

4.1. Nonconstant variance

We present two examples with which to study the performance of linear and squared residual plots in assessing nonconstant variance. Both examples are fitted using the simple linear model, $E(y_i) = \beta_0 + \beta_1 x_i$, where β_0 and β_1 are unknown parameters. The fitted models are estimated with ordinary least squares.

Example 1. Consider the simple linear model,

$$y_i = 1 + x_i + e_i \quad (i = 1, \dots, 50), \quad (20)$$

where values for x_i were chosen randomly from $N(0, 1)$. Errors were simulated from the normal distribution, with mean zero, variance $w(x_i, \delta)\sigma_0^2$, and $\sigma_0^2 = 1$. The weighting functions $w(x_i, \delta)$ are $\exp(|x_i|)$, $\exp(x_i)$ and $|x_i|$. The linear and squared residual plots (rows 1 and 2, respectively, in Figure 2) clearly show nonconstant variance patterns when compared to the corresponding plots in Figure 1 (rows 1 and 2).

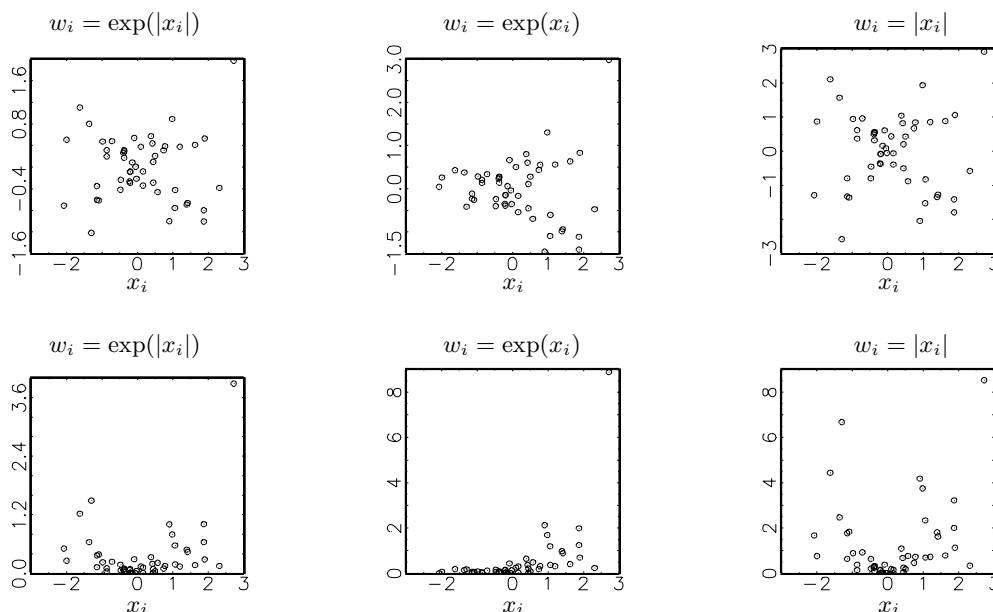


Figure 2. The linear and squared residual plots for assessing the heteroscedasticity. The labels on the Y-axes of the linear and squared residual plots are $\hat{e}_{(i)}/\hat{\sigma}$ and $(\hat{e}_{(i)}/\hat{\sigma})^2$, respectively.

To assess the impact of high leverage points, we simulated the x values from contaminated normal distributions: $95\%N(0, 1) + 5\%N(0, 9)$ and $85\%N(0, 1) + 15\%N(0, 9)$. Here the mean function is the same as that in equation (20), the weighting function is $\exp(|x_i|)$, and the sample size $n = 60$. Under these circumstances, Figure 3 shows that linear residual plots do not exhibit a clear right-opening megaphone pattern, which would indicate nonconstant variance, nor do the squared residual plots $((\hat{e}_{(i)}/\hat{\sigma})^2$ versus x_i) have a parabolic pattern (see Figure 1, row 2 and column 1). Neither do Cook and Weisberg's residual plots clearly show a wedge-shaped pattern. In other words, high leverages and outliers have caused masking effects on the nonconstant variance patterns. Close observation of the residual plots reveal a few outliers. As described in Section 2, we plot $\log(1 + (\hat{e}_{(i)}/\hat{\sigma})^2)$ versus x_i to overcome this difficulty. The resulting transformed plots (last column of Figure 3) now show clear patterns when we compare them to the corresponding plots in Figure 1 (see row 2 and column 1). We can also apply this log transformation to Cook and Weisberg's residual plot by plotting $\log(1 + r_i^2)$ versus $(1 - h_{ii})|x_i|$. The resulting transformed plots (Figure 3, column 3) show a slight improvement over those plots that have not been transformed (Figure 3, column 2). In general, the log-transformation operator can be used to amend the deficiency in squared residual plots caused by outliers.

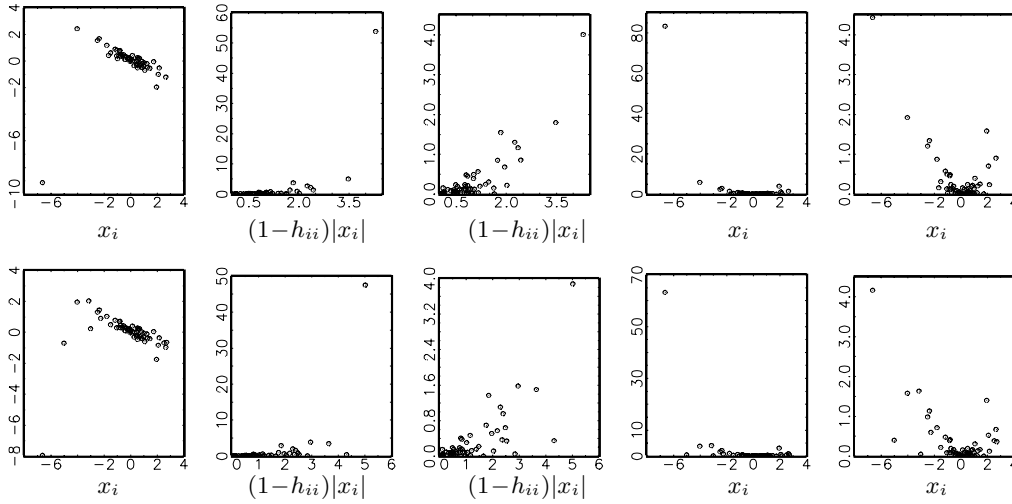


Figure 3. The linear and squared residual plots for assessing the heteroscedasticity as X is simulated from $95\%N(0, 1)+5\%N(0, 9)$ (row 1) and $85\%N(0, 1)+15\%N(0, 9)$ (row 2), respectively. The labels on the Y-axes of plots in each column correspond to $\hat{e}_{(i)}/\hat{\sigma}$, r_i^2 , $\log(1+r_i^2)$, $(\hat{e}_{(i)}/\hat{\sigma})^2$ and $\log(1+(\hat{e}_{(i)}/\hat{\sigma})^2)$, respectively. The weighting function is $w_i = \exp(|x_i|)$.

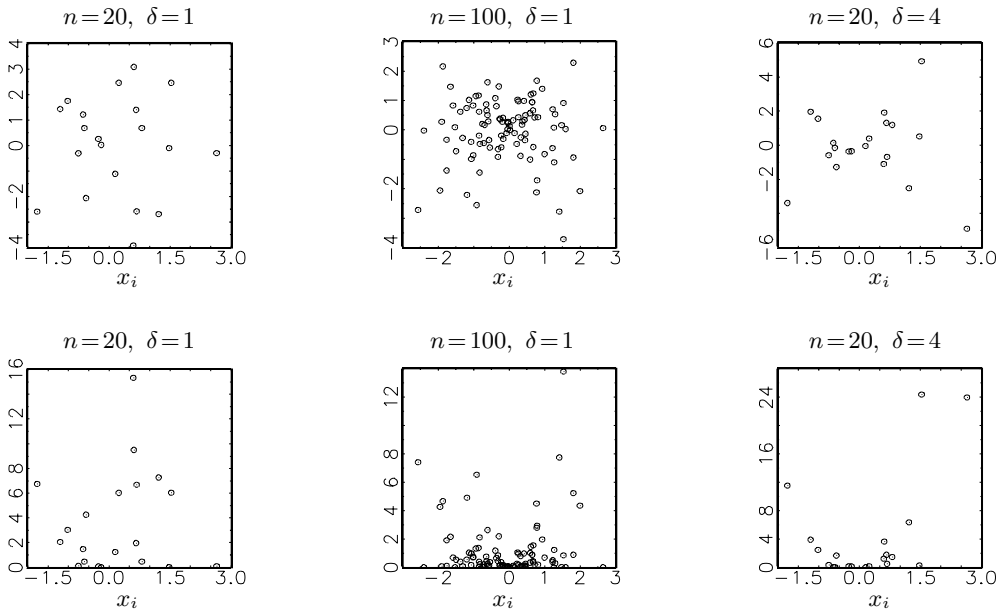


Figure 4. The linear (row 1) and squared (row 2) residual plots for assessing the heteroscedasticity as $w_i = |x_i|^\delta$, $n = 20, 100$ and $\delta = 1$ and 4 . The labels on the Y-axes of the linear and squared residual plots are $\hat{e}_{(i)}/\hat{\sigma}$ and $(\hat{e}_{(i)}/\hat{\sigma})^2$, respectively.

We now examine the effect of sample size and the strength of the weighting function on residual plots. Figure 4 presents the linear and squared residual plots that result when the true model is as given in equation (20) and the variance $w(x_i, \delta)\sigma_0^2 = |x_i|^\delta$. As the sample size increases or the weighting parameter δ gets large, both the linear and squared residual plots more closely resemble the corresponding patterns in Figure 1. This is not surprising, since the leverage given in (11) decreases in n , and the effect of nonconstant variance, $|x_i|^\delta$ increases in δ .

Example 2. Consider the regression model,

$$y_i = 1 + x_i + e_i \quad (i = 1, \dots, 100),$$

where x_i and e_i were chosen randomly from $U(-2, 2)$ and $N(0, w_i)$, respectively, and $w_i = w(x_i, \delta) = 4|x_i|^\delta / (1 + |x_i|^\delta)^2$ with $w(x_i, 0) = 1$. Since $\dot{w}_i = \dot{w}(x_i, 0) = 0$, Cook and Weisberg's squared residual plot cannot be executed directly; an alternative procedure is to plot the squared residual $(b_i/\hat{\sigma})^2$ versus $(1 - h_{ii})\ddot{w}_i$. Figure 5 presents the squared residual plots for $\delta = 2$. From these, it is apparent that Cook and Weisberg's squared residual plot shows the wedge-shaped pattern characteristic of nonconstant variance, and the squared residual plot reflects the weighting function w_i .

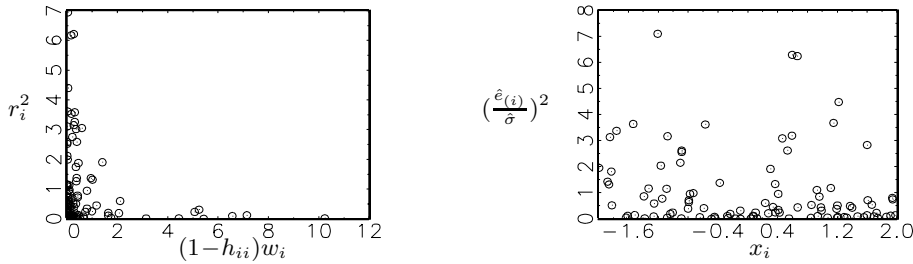


Figure 5. The squared linear residual plots for assessing the heteroscedasticity as $w_i = 4x_i^2 / (1 + x_i^2)^2$.

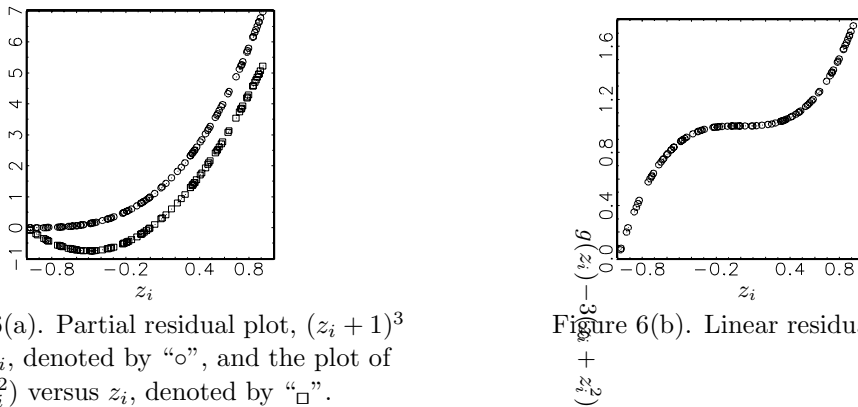


Figure 6(a). Partial residual plot, $(z_i + 1)^3$ versus z_i , denoted by “o”, and the plot of $3(z_i + z_i^2)$ versus z_i , denoted by “□”.

Figure 6(b). Linear residual plot.

4.2. Nonlinearity

Example 3. Let z_i be generated from $U(-1, 1)$ for $i = 1, \dots, 50$. In this case, the function $g(Z) = (Z + 1)^3$ versus Z can only exhibit an exponential pattern (see Figure 6(a)). This is primarily because the linear and quadratic terms for $(Z + 1)^3$, $3(Z + Z^2)$, dominate the function over the range of $(-1, 1)$ for Z , and hence $g(Z)$ and $3(Z + Z^2)$ show similar exponential patterns (see Figure 6(a)).

Now assume that $g(Z)$ is defined as above and $m(Z)\tilde{\gamma}^* = 3(Z + Z^2)$. This implies that the overlap between $g(Z)$ and $m(Z)\tilde{\gamma}^*$ is $3(Z + Z^2)$, which has a stronger impact on g than the cubic term. Therefore, we can see that the resulting CLRES plot (ignoring the constant $E(g(Z)) = 4$) in the population, $Z^3 + 1$ versus Z , can accurately characterize the function g as cubic (see Figure 6(b)). Thus, if we are only interested in finding the relationship between the $g(Z)$ and Z over the observed range of Z , then the exponential function indicated by the CERES plot should be considered. However, if we are interested in finding the true relationship between the $g(Z)$ and Z , then the cubic function identified by the CLRES plot should be considered.

4.3. Speed-braking distance example

This data set is taken from Ezekiel and Fox (1959), p. 45, and was also analyzed by Sen and Srivastava (1990), p. 112. The explanatory variable (x) is automobile speed, and the response (y) is the distance required to come to a standstill after braking. Note that Sen and Srivastava ignored the data point $(x, y) = (22, 35)$ and recoded the data $(x, y) = (7, 6)$ to $(x, y) = (7, 7)$ in their data set. These data modifications also appear in Weisberg's book (1985), p. 161.

Since the stopping distance is zero when speed is zero, we propose an initial model

$$y_i = \beta_1 x_i + e_i \quad (i = 1, \dots, 63), \quad (21)$$

where e_i is randomly distributed from $N(0, \sigma^2)$. Figure 7 (row 1) presents the scatter plot of y_i versus x_i , and the linear, squared, and Cook and Weisberg's squared residual plots resulting from fitting the data with model (21). No clear nonlinear pattern is evident in the scatterplot. However, the curved pattern in the linear residual plot, as well as its asymmetry about zero, indicates that a nonlinear function may provide a better fit. Furthermore, the squared and Cook and Weisberg's squared residual plots reveal nonconstant variance. At this point we can choose either to add nonlinear components into model (21) or to fit the data assuming nonconstant variance. We first try adding a nonlinear component into model (21).

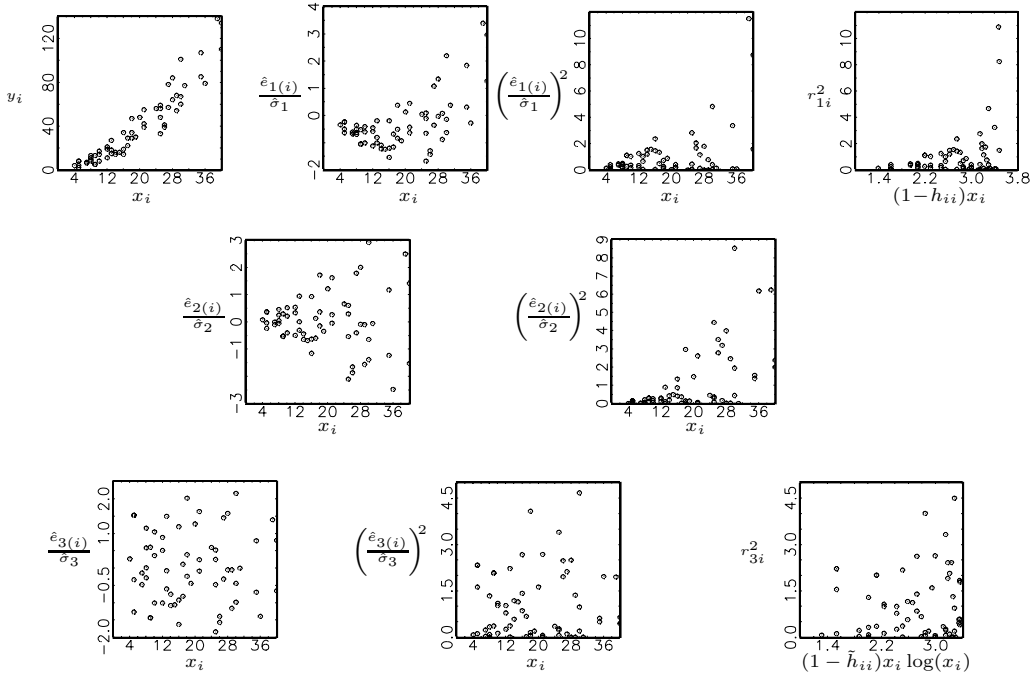


Figure 7. The scatter, linear and squared residual plots for the Speed Braking Distance Data. The $\hat{e}_{1(i)}/\hat{\sigma}_1$ (and r_{1i}), $\hat{e}_{2(i)}/\hat{\sigma}_2$, and $\hat{e}_{3(i)}/\hat{\sigma}_3$ (and r_{3i}) correspond to the scaled predicted residuals (and the studentized residuals) by fitting model (21), model (22) with constant variance, and model (22), respectively. The h_{ii} and \hat{h}_{ii} are the leverages of hat matrices H and \hat{H} , where H and \hat{H} are computed from matrices (x_i) and (x_i, x_i^2) , respectively.

Observe that the linear residual plot shows a parabolic pattern, and thus we refit the data by adding x^2 to model (21). Figure 7 (row 2) presents the resulting linear and squared residual plots. The linear residual plot shows a right-opening megaphone pattern starting at positive values of x , and is symmetric about zero. In addition, the squared residual plot shows a wedge-shaped pattern. Both indicate possible nonconstant variance. Comparing both the linear and squared residual plots with the corresponding patterns in Figure 1, we find that the weighting function w_i could either be in the family of $|x_i|^\delta$ or in the family of $\exp(\delta x_i)$. For simplicity, we choose $w_i = x_i^2$. A score test (Cook and Weisberg (1983)) with a p-value of less than 0.01 rejects the null hypothesis ($H_0 : \delta = 0$) for the weighting function $w_i = |x_i|^\delta$. The linear, squared, and Cook and Weisberg's squared residual plots that result from the weighted model are given in Figure 7 (row 3). None of them now show clear patterns. We can thus conclude that the model we have developed is appropriate to illustrate the relationship between braking distance and speed:

$$y_i = \beta_1 x_i + \beta_2 x_i^2 + e_i \quad (i = 1, \dots, 63), \tag{22}$$

where e_i is randomly distributed from $N(0, x_i^2 \sigma^2)$. This model was also recommended by Hald (1952), p. 570. It proposes that the distance to stop from the moment the brakes are applied is proportional to the square of the speed, and the variation of the distance is also proportional to the speed.

An alternative analytical approach for this data set would be to first find an appropriate weighting function after examining the residual plots from the least squares fit of model (21) (Figure 7, row 1). Then introduce appropriate nonlinear components by examining the residual plots from a weighted least squares fit. For this data, either approach will lead to the model given in (22).

4.4. Land rent example

Cook and Weisberg (1994), p. 90 presented this example in their Exercises. This data studies the variation in rent paid for agricultural land planted with alfalfa in 1977. The response variable (y) is the average rent per acre planted with alfalfa, and the three explanatory variables are average rent paid for all tillable land (x_1), density of dairy cows (x_2), and the proportion of farmland in the county used as pasture (x_3).

We first fit the data with the linear regression model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i \quad (i = 1, \dots, 67), \quad (23)$$

where e_i is a random variable distributed as $N(0, \sigma^2)$. The resulting linear and squared residual plots are presented in Figure 8 ((a),(b), and (c)) and Figure 8 ((d), (e), and (f)), respectively. They indicate a quadratic relationship for variables y and x_2 , and nonconstant variance for both variables x_1 and x_3 . A comparison of the squared residual plots with the plots in Figure 1 suggests that the weighting function is exponential. Also, because the p-values for testing $H_0 : \beta_1 = 0$, $H_0 : \beta_2 = 0$ and $H_0 : \beta_3 = 0$ are 0.000, 0.000 and 0.356, respectively, we can conclude that variable x_3 does not play a significant role in explaining variation in y . Thus, we refit the data without x_3 but including x_2^2 with the weighted regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{22} x_{i2}^2 + e_i \quad (i = 1, \dots, 67), \quad (24)$$

where e_i are randomly distributed from $N(0, \exp(\delta_1 x_{i1} + \delta_2 x_{i3}) \sigma^2)$. The p-value of the score test for testing the null hypothesis $H_0 : \delta_1 = \delta_2 = 0$ is less than 0.01, and hence we reject the assumption of constant variance. After fitting model (24), the resulting plots of linear and squared residuals show no clear pattern, but do show three outliers: cases 19, 36, and 67. After removing these three points and refitting the remaining data with model (24), the weighted regression function is

$$\hat{y}_i = -6.911 + 0.884 x_{i1} + 0.709 x_{i2} - 0.007 x_{i2}^2, \quad (25)$$

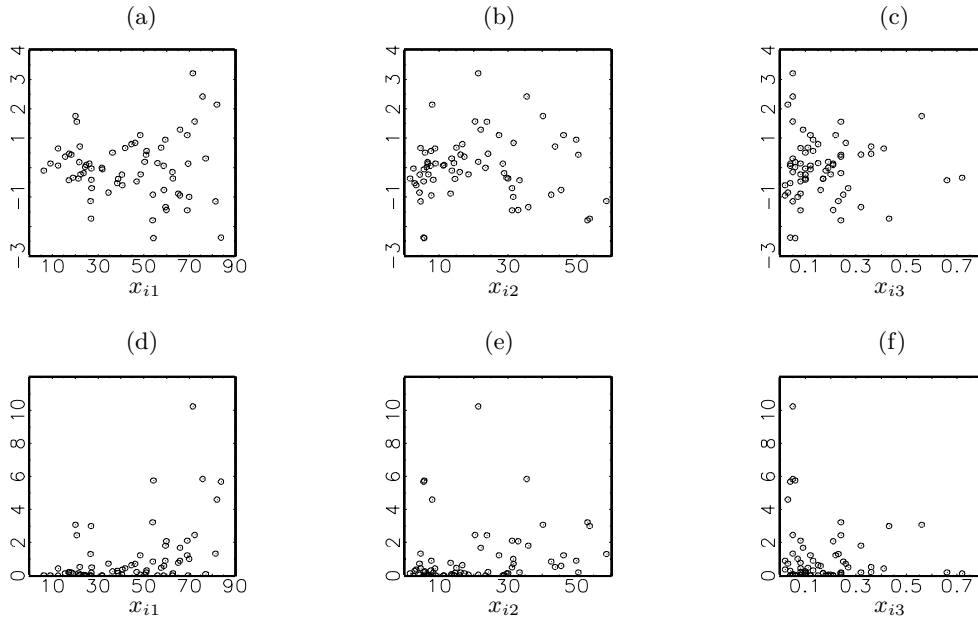


Figure 8. (a)-(f) The linear (row 1) and squared residual (row 2) plots for Land rent Data. The scaled predicted residuals are calculated by fitting the data with model (23). The labels on the Y-axes of the linear and squared residual plots are $\hat{e}_{(i)}/\hat{\sigma}$ and $(\hat{e}_{(i)}/\hat{\sigma})^2$, respectively.

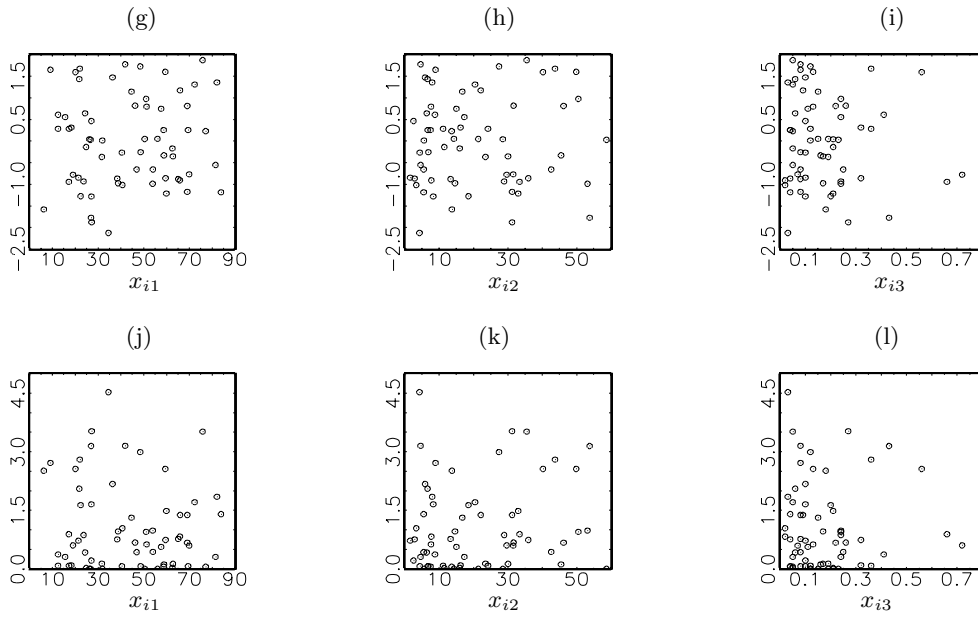


Figure 8. (g)-(l) The linear (row 1) and squared residual (row 2) plots for Land rent Data. The scaled predicted residuals are calculated by fitting the data with model (25). The labels on the Y-axes of the linear and squared residual plots are $\hat{e}_{(i)}/\hat{\sigma}$ and $(\hat{e}_{(i)}/\hat{\sigma})^2$, respectively.

with the estimated weights $\hat{w}_i = \exp(0.068x_{i1} + 5.488x_{i3})$ for $i = 1, \dots, 64$. Figure 8 ((g), (h), and (i)) and Figure 8 ((j), (k), and (l)) present resulting linear and squared residual plots, respectively. They do not exhibit any clear pattern. The coefficient of determination is 0.933, and all p-values for testing regression coefficients are less than 0.01. Equation (25) shows that the average rent per acre planted with alfalfa increases as the average rent paid for all tillable land increases. It also indicates that the average rent per acre planted with alfalfa first increases to the maximum amount, and then drops as the number of dairy cows per square mile increases. This finding is sensible. We can conclude that the use of linear and squared residual plots leads to an adequate model for describing the data.

5. Discussion

This article has introduced and discussed various linear and squared residual plots. From both a theoretical and empirical standpoint, the linear residual plot is useful for assessing nonlinearity. By contrast, the squared residual plot is preferred to detect nonconstant variance. In practice, we suggest using both Cook and Weisberg's and our squared residual plots (Section 2.3) to assess heteroscedasticity, since these plots complement each other (especially when the sample size is small or when high leverages or outliers exist).

In data analysis, we conventionally examine the residual plot, \hat{e}_i versus \hat{y}_i (or r_i versus \hat{y}_i), to assess nonlinearity or heteroscedasticity. Based on our theoretical and analytical results, this plot may not be sufficient by itself to diagnose nonconstant variance or nonlinearity, and it sometimes may provide misleading information. This plot is useful for assessing heteroscedasticity only when the variance is known to be proportional to the mean function, and the mean function contains more than one explanatory variable. In addition, the conventional residual plot is useful for examining nonlinearity only when the nonlinear component $g(Z)$, given in equation (13), is a function of $X\beta$. For all but these special cases, the other linear or squared residual plots discussed in Sections 2 and 3 should be considered when evaluating nonlinearity or heteroscedasticity.

If the true model has nonlinear components with nonconstant error variances and independent covariates, then we can adapt the techniques in Sections 2, 3.1, and 3.2 to obtain conditional means and variances for residuals, calculated through the least squares fits. Since these results involve both nonlinear and weighting functions, the interpretation of residual plots can get complicated. In this case we propose using a sequential procedure, as we did in Sections 4.3 and 4.4, to form an appropriate final model. In other words, resolve nonlinearity (or nonconstant variance) first by adding nonlinear components into the model (or by fitting the data with weighted least squares). Then analyze the resulting

residual plots and amend any deficiencies caused by nonconstant variance (or nonlinearity). In conclusion, we suggest that both linear and squared residual plots should be used routinely in data analysis.

If the true model has nonlinear components with nonconstant error variances and interdependent covariates, then we propose first using CERES and CLRES plots to identify nonlinear components. Then add the nonlinear components into the model and examine the squared residual plots, discussed in Section 2, and use them to amend any deficiency caused by nonconstant variance. The above suggestions are heuristic, and certainly need to be further explored. This is under investigation.

Acknowledgement

The first author was supported in part by grant DMS-95-10511 from the National Science Foundation. The authors thank the Associate Editor and referees for their helpful suggestions for improvements to this paper, and Dennis R. Cook, Anthony Davison, and Michelle C. Pallas for their comments on earlier versions of the manuscript.

References

- Anscombe, F. J. (1961). Examination of residuals. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **I**, 1-36.
- Atkinson, A. C. (1985). *Plots, Transformations and Regression*. Oxford University Press, Oxford.
- Azzalini, A. and Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc. Ser. B* **55**, 549-557.
- Berk, K. N. and Booth, D. E. (1995). Seeing a curve in multiple regression. *Technometrics* **37**, 385-398.
- Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall, New York.
- Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*. Wiley, New York.
- Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829-836.
- Cook, R. D. (1993). Exploring partial residual plots. *Technometrics* **35**, 351-362.
- Cook, R. D. (1994). On the interpretation of regression plots. *J. Amer. Statist. Assoc.* **89**, 177-189.
- Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.
- Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70**, 1-10.
- Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics*. Wiley, New York.
- Data Desk (1995). Version 5, Data Description Inc., New York.
- Draper, N. R. and Smith, H. (1966). *Applied Regression Analysis, 1st Ed.* Wiley, New York.

- Eubank, R. L. and Hart, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* **20**, 1412-1425.
- Ezekiel, M. (1924). A method for handling curvilinear correlation for any number of variables. *J. Amer. Statist. Assoc.* **19**, 431-453.
- Ezekiel, M. and Fox, K. A. (1959). *Methods of Correlation and Regression Analysis, 3rd Ed.* Wiley, New York.
- Hald, A. (1952). *Statistical Theory With Engineering Applications.* Wiley, New York.
- Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.
- Larsen, W. A. and McCleary, S. J. (1972). The use of partial residual plots in regression analysis. *Technometrics* **14**, 781-789.
- Lyon, J. and Tsai, C. L. (1996). A comparison of tests for heteroscedasticity. *J. Roy. Statist. Soc. Ser D* **45**, 337-349.
- Mansfield, R. E. And Conerly, M. D. (1987). Diagnostic value of residual and partial residual plots. *Amer. Statist.* **41**, 107-116.
- Montgomery D. C. and Peck, E. A. (1992). *Introduction to Linear Regression Analysis*, 2nd Ed. Wiley, New York.
- Sen, A. and Srivastava, M. (1990). *Regression Analysis (Theory, Methods, and Applications).* Springer-Verlag, New York.
- Su, J. Q. and Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *J. Amer. Statist. Assoc.* **86**, 420-426.
- Weisberg, S. (1985). *Applied Linear Regression*, 2nd Ed. Wiley, New York.

Graduate School of Management, University of California, Davis, CA 95616, U.S.A.

E-mail: cltsai@ucdavis.edu

Department of Mathematics, Southwest Missouri State University, Springfield, MO 65804, U.S.A.

E-mail: cai@math.smsu.edu

Department of Mathematics, Nankai University, Tianjin, China.

E-mail: xzwu@sun.nankai.edu.cn

(Received April 1995; accepted April 1997)