Global Validation of Linear Model Assumptions

E. A. Peña* E. H. Slate†

December 13, 2003

Abstract

A test for globally testing the four assumptions of the linear model is proposed. The test can be viewed as a Neyman's smooth test and it only relies on the residual vector. The components of the global test statistic could be utilized to gain insights into which assumptions have been violated if the global procedure indicates that there is a breakdown in at least one of the four assumptions. The procedure could be used in conjunction with the usual graphical methods, and it is simple enough to be implemented by beginning statistics students. The procedure is demonstrated by analyzing data sets that have been used in previous works dealing with model diagnostics, and a real data set pertaining to end-of-trading-day share values of the College Retirement and Equities Funds Growth and Stock accounts. Simulation results are presented indicating the sensitivity of the procedure in detecting model violations under a variety of situations.

Key Words and Phrases: Deletion statistics; Model diagnostics and validation; Neyman smooth test; Outlier detection; Score test.

*E. Peña is Professor in the Department of Statistics, University of South Carolina, Columbia, SC 29208. His e-mail address is `pena@stat.sc.edu`. He acknowledges the research support provided by NSF Grant DMS 0102870, NIH Grant GM056182, NIH COBRE Grant RR17698, and the USC/MUSC Collaborative Research Program.

†E. Slate is Associate Professor in the Department of Biometry and Epidemiology, Medical University of South Carolina, Charleston, SC 29425. Her e-mail address is `slateeh@musc.edu`. She acknowledges the research support provided by NIH Grant CA077789, NIH COBRE Grant RR17696, DAMD Grant 17-02-1-0138 and the USC/MUSC Collaborative Research Program.

AMS Subject Classification: Primary: 62J20 Secondary: 62J05, 62H15

1 Linear Model and its Assumptions

One of the most important models in Statistics is the linear model which postulates a relationship between an observable $n \times 1$ response vector $\mathbf{Y}$ and an observable $n \times p$ matrix $\mathbf{X}$ which could be constituted by the design matrix and/or the values of predictor variables. In the linear model the relationship between $\mathbf{Y}$ and $\mathbf{X}$ is given by

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\epsilon, \tag{1}$$

where β is a $p \times 1$ vector of unknown coefficients, σ is an unknown scale parameter, and ϵ is an $n \times 1$ vector of unobservable error variables. Furthermore, it is assumed that, conditionally on $\mathbf{X}$, ϵ has a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}$, the $n \times n$ identity matrix. This distributional assumption, together with the linear link specification in (1) are usually enumerated as the following four distinct assumptions.

(A1) (Linearity) $\mathbf{E}\{Y_i|\mathbf{X}\} = \mathbf{x}_i\beta$, where $\mathbf{x}_i$ is the ith row of $\mathbf{X}$;

(A2) (Homoscedasticity) $\mathbf{Var}\{Y_i|\mathbf{X}\} = \sigma^2, (i = 1, 2, \ldots, n)$;

(A3) (Uncorrelatedness) $\mathbf{Cov}\{Y_i, Y_j|\mathbf{X}\} = 0, (i \neq j)$; and

(A4) (Normality) $Y_i|\mathbf{X}, (i = 1, 2, \ldots, n)$, have normal distributions.

Assumptions (A3) and (A4) are equivalent to the assumption that $Y_i|\mathbf{X}, (i = 1, 2, \ldots, n)$, are independent normal random variables. Without loss of generality, we assume that $\mathbf{X}$ is of full rank and with $n > p$, that is, $\text{rank}(\mathbf{X}) = p$. It is well-known that under (A1)-(A4), the maximum likelihood (ML) estimators of β and σ^2 are given, respectively, by

$$\mathbf{b} = \hat{\beta} = (\mathbf{X}^{\mathrm{t}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{t}}\mathbf{Y}; \tag{2}$$

$$s^2 = \hat{\sigma}^2 = \frac{1}{n}\mathbf{Y}^{\mathrm{t}}(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}, \tag{3}$$

where $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^{\mathrm{t}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{t}}$ is the projection operator on the linear subspace generated by the columns of $\mathbf{X}$. This matrix is usually denoted by $\mathbf{H}$. It is well-known that the estimator $\mathbf{b}$ in (2) is also the least-squares (LS) estimator of β. The usual procedures for constructing confidence ellipsoids/intervals and for testing hypotheses for β and σ^2 were also developed under the assumptions (A1)-(A4), and consequently, the validity of these inferential procedures rely to a great extent on the validity of (A1)-(A4). The consequences of the breakdown of any of these four assumptions are well-known, and possible remedial measures such as the use of variable transformations, weighted regression, incorporating additional predictor variables and, if need be, the adoption of nonparametric methods, have also been discussed in numerous research papers. The assessment of whether assumptions (A1)-(A4) are satisfied, based on the data $(\mathbf{Y}, \mathbf{X})$, has therefore received considerable attention and has led to numerous procedures. Most of these procedures typically involve the use of the observed $n \times 1$ vector of standardized residuals $\mathbf{R}$, defined

via

$$\mathbf{R} = \frac{\mathbf{Y} - \mathbf{X}\mathbf{b}}{s} = \frac{(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}}{s}. \tag{4}$$

There are other types of residuals that have been used in model validation and diagnostics such as Theil's (1965) best linear unbiased scalar covariance residuals, referred to as BLUS residuals; as well as recursive or sequential residuals, see for instance Kianifard and Swallow's (1996) review paper. In this paper we restrict our attention to the use of the (ordinary) residuals $\mathbf{R}$.

Early important works dealing with the use of $\mathbf{R}$ in assessing the model assumptions are those by Tukey (1949) in the context of testing for nonadditivity (assessing A1); Durbin and Watson's (1950, 1951) test for serial correlation (assessing A3); Anscombe (1961) and Anscombe and Tukey (1963) for checking normality (assumption A4) and homoscedasticity (assumption A2). Many of these residual-based methods for validating the assumptions (A1)-(A4) are summarized and discussed in the excellent monographs of Cook and Weisberg (1982) and Atkinson (1985); indeed, the references in these two monographs serve as excellent resources for the literature in this area. It should be pointed out that the computation of $\mathbf{R}$ involves the reuse of the data $(\mathbf{Y}, \mathbf{X})$ as this is also used for estimating β and σ^2, and in statistics utilized in testing hypotheses for β and σ^2. As a consequence of this reuse of the data $(\mathbf{Y}, \mathbf{X})$, the residual vector generally obtains a more complicated distributional structure; in particular, the residuals are not independent even if (A1)-(A4) hold, in contrast to the independence of the 'true' residual vector $\epsilon = (\mathbf{Y} - \mathbf{X}\beta)/\sigma$. The impact, especially the non-negligible change on

the distributional properties of the residuals *even* in large samples, by the substitution of estimators for unknown parameters to obtain the residuals have been duly noted in research papers, cf., Durbin and Watson (1950), Anscombe and Tukey (1963), Theil (1965), and Atkinson (1985, p. 24). However, in practical settings, especially in the assessment of (A1)-(A4) through graphical methods, the impact of the aforementioned substitution is mostly ignored.

Existing methods for checking the validity of (A1)-(A4) can be classified into two types: graphical methods and formal significance testing methods. Graphical methods, which we usually teach in our elementary courses because of ease and convenience by virtue of the fact that most statistical softwares automatically generate the 'appropriate' plots, usually involve a combination of plots of the residuals $\mathbf{R}$ with respect to fitted values, functions of both included and omitted predictor variables, and time sequence, as well as their normal probability plots and histograms/boxplots, cf., Cook and Weisberg (1982), Atkinson (1985), and Cook (1998). These graphical methods are also utilized to detect outlying and/or influential observations, though from a mathematical viewpoint, these latter issues can be subsumed into the assessment of (A1)-(A4). Graphical methods are certainly convenient to use, especially so with the availability of statistical softwares that easily generate a multitude of plots; however, its use in assessing (A1)-(A4) is also highly subjective, and it could be quite misleading owing to the effect of the substitution of estimators for the unknown parameters. Furthermore,

a particular plot is used to assess a particular assumption, but sometimes it is not clear how combinations of violations of (A1)-(A4) could impact the behavior of the resulting plot. Thus, it will be beneficial to the practitioner if we could augment these residual plots with a value that could serve as a measure of the degree in which the assumptions (A1)-(A4) are violated.

Formal significance tests for (A1)-(A4) involve testing the null hypothesis (H_0) versus the alternative alternative (H_1) where

$$\begin{array}{ll} H_0: & \text{Assumptions (A1)-(A4) all hold;} \\ H_1: & \text{At least one of (A1)-(A4) does not hold.} \end{array} \tag{5}$$

The typical structure of such a test is to define a statistic $S(\mathbf{R})$ whose sampling distribution is known under H_0 and such that departures from H_0 will manifest in terms of larger values of $S(\mathbf{R})$. Given an observed residual vector $\mathbf{R} = \mathbf{r}$, one calculates the p-value via

$$p = \mathbf{P}\{S(\mathbf{R}) > S(\mathbf{r})|H_0\}, \tag{6}$$

and the decision to reject H_0 is based on the magnitude of p. However, the current state of the art is that these formal significance tests are typically tests for a specific assumption, so they are not simultaneous or global tests for the four assumptions (A1)-(A4). For instance, there are tests for the normality assumption (cf., Anscombe and Tukey (1963)); there are tests for link mis-specifications (cf., Tukey (1949)); there are tests for heterogeneity of variances (Cook and Weisberg (1983); Bickel (1978), and Anscombe (1961)); and there are tests for the uncorrelatedness or independence of the error components (cf., Durbin and Watson (1950, 1951), Theil and Nagar (1961)).

See also Kianifard and Swallow (1996) for significance testing procedures for the different assumptions which utilizes recursive residuals. The difficulty with these tests is that each may not be able to detect departures from those assumptions that the particular test is not specific for, and the impact of a violation of another assumption on this test is usually not apparent. Another potential problem is that when a specific test indicates a violation, it might be due to the violation of another assumption which also affects this specific test. For example, a test for normality could also be drastically affected by a mis-specified link function or dependent error components. One may decide to perform tests for each of the different assumptions, but this will lead to an increase in the Type I error probability when the results of these tests are combined. There is therefore a need to have a global test for all the assumptions (A1)-(A4) which controls the Type I error rate and which could be used if the user does not have an idea of which particular assumption may be violated. If such a test indicates that at least one of the assumptions is not satisfied, then directional tests may be used to determine the assumptions that have been violated. Knowing the particular assumption that has been violated is important for instituting appropriate remedial measures.

In this paper we propose such a global test. The test is based on the residual vector, and it could be viewed as a Neyman (1937) smooth test (cf., Thomas and Pierce (1979) and Rayner and Best (1986, 1989)), hence it possesses local optimality properties. Furthermore, the components of this global test could be utilized as directional tests for determining the assump-

tions that have been violated. As alluded to earlier, statistics that are functions of the residual vector $\mathbf{R}$ generally possess complicated distributional properties. Consequently, the distributional aspects for the proposed global test is asymptotic. In principle, for small sample sizes, computer-intensive methods, such as bootstrapping or Monte Carlo methods, may be employed to determine p-values.

The remainder of this paper is organized as follows. Section 2 will describe and discuss the global and the component statistics. 'Deletion' statistics obtained by excluding an observation from the analysis will also be described. Their utility for assessing outlying and influential observations will be discussed. The theoretical justification of the global procedure will be presented in Section 3 where it will be derived as a Neyman smooth test. The asymptotic normality and the asymptotic independence of the components will be established in this section. The sensitivity of the procedures were examined through computer simulations. The results of this simulation study are summarized in Section 4. Simulated levels and powers of the test under different situations depicting violations of the model assumptions are presented. Section 5 will contain an application of the procedure to the Forbes data set; an analysis of Ruppert and Carroll's (1980) water salinity data; an illustration using a textile data from Box and Cox (1964); and an application to a real data set consisting of end-of-trading-day values of the College Retirement Equities Funds (CREF) Growth and Stock Accounts. The first three are well-known data sets which have been used in papers and

monographs dealing with model validation and diagnostics.

2 Global Procedure and Component Statistics

Henceforth, we assume that the matrix $\mathbf{X}$ has as its first column the $n \times 1$ vector $\mathbf{1} = (1, 1, \ldots, 1)^{\mathrm{t}}$. This is hardly a restrictive assumption as this simply means that we are incorporating an intercept term in model (1), which is typically the case. Recalling that the ith component of the residual vector is

$$R_i = \frac{Y_i - \hat{Y}_i}{s}, \; i = 1, 2, \ldots, n, \tag{7}$$

where $\hat{Y}_i = \mathbf{x}_i \mathbf{b}$ is the ith fitted or predicted value, the first three component statistics are as follows:

$$\hat{S}_1^2 = \left\{ \frac{1}{\sqrt{6n}} \sum_{i=1}^{n} R_i^3 \right\}^2; \tag{8}$$

$$\hat{S}_2^2 = \left\{ \frac{1}{\sqrt{24n}} \sum_{i=1}^{n} [R_i^4 - 3] \right\}^2; \tag{9}$$

$$\hat{S}_3^2 = \frac{\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 R_i \right\}^2}{(\hat{\Omega} - \mathbf{b}^{\mathrm{t}} \hat{\mathbf{\Sigma}}_X \mathbf{b} - \hat{\Gamma} \hat{\Sigma}_X^{-1} \hat{\Gamma}^{\mathrm{t}})}, \tag{10}$$

where, with $\bar{\mathbf{z}} = \frac{1}{n} \mathbf{1}^{\mathrm{t}} \mathbf{Z}$ for an $n \times q$ matrix $\mathbf{Z}$, we define

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^4, \quad \hat{\Sigma}_X = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{t}} (\mathbf{x}_i - \bar{\mathbf{x}}), \quad \text{and} \quad \hat{\Gamma} = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 (\mathbf{x}_i - \bar{\mathbf{x}}). \tag{11}$$

The fourth component statistic requires a user-supplied $n \times 1$ vector $\mathbf{V}$, which by default is set to be the time sequence $\mathbf{V} = (1, 2, \ldots, n)^{\mathrm{t}}$. It is defined via

$$\hat{S}_4^2 = \left\{ \frac{1}{\sqrt{2\hat{\sigma}_V^2 n}} \sum_{i=1}^{n} (V_i - \bar{V})(R_i^2 - 1) \right\}^2, \tag{12}$$

with $\hat{\sigma}_V^2 = \frac{1}{n} \sum_{i=1}^{n} (V_i - \bar{V})^2$. The global test statistic is defined as

$$\hat{G}_4^2 = \hat{S}_1^2 + \hat{S}_2^2 + \hat{S}_3^2 + \hat{S}_4^2. \tag{13}$$

Versions related to the statistics $\hat{S}_i^2, (i = 1, 2, 4)$, have been considered for significance testing purposes in earlier papers. For instance, statistics related to $\hat{S}_1^2$ and $\hat{S}_2^2$ have appeared in Anscombe and Tukey (1963), and a statistic related to $\hat{S}_4^2$ has been considered by Cook and Weisberg (1983), Bickel (1978), and Anscombe (1961) in the context of testing for heteroscedasticity. Perhaps, one of the main contributions of the present paper is combining these different directional statistics in a global statistic and determining its properties, albeit asymptotic properties. We will see in ensuing sections that this combined global statistic could serve as an omnibus statistic for *globally* testing all the assumptions of the linear model.

For large n, which for application purposes will be understood to mean that $n - p \geq 30$, the global test for the hypotheses H_0 versus H_1 in (5) at an asymptotic significance level of α is:

$$\textbf{Global Test: } \text{Reject } H_0 \text{ if } \hat{G}_4^2 > \chi_{4;\alpha}^2, \tag{14}$$

where $\chi_{k;\alpha}^2$ is the $100(1-\alpha)$th percentile of a central chi-squared distribution with degrees-of-freedom k. If the test in (14) leads to the rejection of H_0,

the component statistics $\hat{S}_1^2$, $\hat{S}_2^2$, $\hat{S}_3^2$, and $\hat{S}_4^2$ could be examined by comparing their values to $\chi^2_{1;\alpha}$ to get an indication of which particular assumption or assumptions have been violated. The following are rough guidelines in interpreting the values of these component statistics, with these guidelines suggested by the theoretical considerations to be presented in Section 3 and the simulation results in Section 4.

(i) Skewed error distributions will usually be indicated by large values of the statistic $\hat{S}_1^2$;

(ii) Deviations from the normal distribution kurtosis of the true error distribution will be generally revealed by large values of statistic $\hat{S}_2^2$;

(iii) The use of a misspecified link function or the absence of other predictor variables in the model will mostly be detected by large values of the statistic $\hat{S}_3^2$;

(iv) The presence of heteroscedastic errors and/or dependent errors will typically manifest in large values of the statistic $\hat{S}_4^2$; and

(v) Simultaneous violations of at least two of the assumptions (A1)-(A4) will be manifested by large values of several of these component statistics.

A potentially useful procedure for detecting outlying and influential observations, and which could be implemented in conjunction with the global and directional tests, is to adopt the well-known idea of 'deletion' statistics,

which reflect the change in values of statistics after the deletion of an observation. For a statistic T, denote by $T[i]$ the value of the statistic after the ith observation is deleted. We will be interested in the quantities

$$\Delta\hat{G}_4^2[i] = \left[\frac{\hat{G}_4^2[i] - \hat{G}_4^2}{\hat{G}_4^2}\right] \times 100, \quad i = 1, 2, \ldots, n, \tag{15}$$

which represent the percent relative change in the value of the global statistic $\hat{G}_4^2$ after the deletion of the ith observation. The idea is that an observation with a large absolute value of $\Delta\hat{G}_4^2[i]$ is either an outlier or has large influence. The values of $\Delta\hat{G}_4^2[i]$ could conveniently be plotted with respect to the time sequence to graphically assess their values.

Related to the statistic in (15) is to compute the p-values after the deletion of each of the observations, that is,

$$p[i] = \mathbf{P}\{\hat{G}_2^4[i] > \hat{g}_2^4[i] | H_0\}, \quad i = 1, 2, \ldots, n, \tag{16}$$

where $\hat{g}_2^4[i]$ is the observed value of the global statistic after deletion of the ith observation. The evaluation of this probability could be performed using the (approximate) chi-squared distribution with 4 degrees-of-freedom. The idea is if $p[i]$ is quite different from the other $p[j]$'s, this will be indicative that the ith observation is either an outlier or an influential observation. A plot of the $p[i]$-values with respect to the time-sequence will graphically aid in assessing their values.

Clearly, this deletion idea could be extended to each of the component statistics; however, it appears that for economy of information, the deletion statistics pertaining to the global statistic will suffice for data-analytic

purposes. We mention that our `S-Plus` program which implements the procedures have the option of computing the above quantities for each of the component statistics.

The computation of these 'deletion' statistics can be efficiently performed by solely utilizing quantities obtained by fitting the model using the full data. These computational formulas arise as a consequence of the important matrix inversion formula (cf., Atkinson (1985), formula (2.2.1)) which states that for a nonsingular matrix $\mathbf{A}$ and matrices $\mathbf{U}$ and $\mathbf{V}$,

$$(\mathbf{A} - \mathbf{U}\mathbf{V}^{\mathrm{t}})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I} - \mathbf{V}^{\mathrm{t}}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^{\mathrm{t}}\mathbf{A}^{-1}. \tag{17}$$

Denote by h_{ij} the (i,j)th element of the projection matrix $\mathbf{P}_X = \mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{t}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{t}}$, that is, $h_{ij} = \mathbf{x}_i(\mathbf{X}^{\mathrm{t}}\mathbf{X})^{-1}\mathbf{x}_j^{\mathrm{t}}$. Then, for $i = 1, 2, \ldots, n$ (cf., the formulas (2.2.8) and (2.2.9) in Atkinson (1985)),

$$\mathbf{b}[i] = \mathbf{b} - \frac{(\mathbf{X}^{\mathrm{t}}\mathbf{X})^{-1}\mathbf{x}_i^{\mathrm{t}} s R_i}{1 - h_{ii}} \quad \text{and} \quad s^2[i] = \frac{ns^2}{n-1}\left[1 + \frac{R_i^2}{n(1-h_{ii})}\right]. \tag{18}$$

Consequently, the standardized residual associated with the jth observation when the ith observation is excluded from the analysis is

$$R_j[i] = \frac{Y_j - \hat{Y}_j[i]}{s[i]} = \frac{\sqrt{n-1}\left(R_j + \frac{h_{ji}}{1-h_{ii}}R_i\right)}{\sqrt{n + \frac{1}{1-h_{ii}}R_i^2}} \tag{19}$$

since $Y_j - \hat{Y}_j[i] = s\left(R_j + \frac{h_{ji}}{1-h_{ii}}R_i\right)$. We also need formulas for $\hat{\Omega}[i]$, $\hat{\mathbf{\Sigma}}_X[i]$ and its inverse, $\hat{\Gamma}[i]$, and $\hat{\sigma}_V^2[i]$ which solely use quantities obtained from the full data model fitting. It is immediate to show that

$$\hat{\sigma}_V^2[i] = \frac{n}{n-1}\left[\hat{\sigma}_V^2 - \frac{(V_i - \bar{V})^2}{n}\right]. \tag{20}$$

On the other hand, because $\hat{Y}_j[i] = \hat{Y}_j - \frac{h_{ji}}{1-h_{ii}} sR_i$ and $\bar{Y}[i] = \frac{n}{n-1}\left(\bar{Y} - \frac{Y_i}{n}\right)$, then

$$\hat{Y}_j[i] - \bar{Y}[i] = (\hat{Y}_j - \bar{Y}) + \frac{(Y_i - \bar{Y})}{n-1} - \frac{h_{ji}}{1-h_{ii}} sR_i. \tag{21}$$

Using the above expression, we may compute $\hat{\Omega}[i]$ according to

$$\hat{\Omega}[i] = \frac{1}{n-1} \sum_{j=1;\ j\neq i}^{n} \left(\hat{Y}_j[i] - \bar{Y}[i]\right)^4. \tag{22}$$

Analogous arguments lead to

$$\hat{\mathbf{\Sigma}}_X[i] = \frac{n}{n-1}\left[\hat{\mathbf{\Sigma}}_X - \frac{(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{t}}(\mathbf{x}_i - \bar{\mathbf{x}})}{n}\right], \tag{23}$$

from which upon applying (17), we obtain

$$\hat{\mathbf{\Sigma}}_X^{-1}[i] = \frac{n-1}{n}\left\{\hat{\mathbf{\Sigma}}_X^{-1} + \frac{1}{n}\hat{\mathbf{\Sigma}}_X^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{t}}\left[\mathbf{I} - \frac{1}{n}(\mathbf{x}_i - \bar{\mathbf{x}})\hat{\mathbf{\Sigma}}_X^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{t}}\right](\mathbf{x}_i - \bar{\mathbf{x}})\hat{\mathbf{\Sigma}}_X^{-1}\right\}.$$

Finally, by using (21) and the obvious identity $\mathbf{x}_j - \bar{\mathbf{x}}[i] = (\mathbf{x}_j - \bar{\mathbf{x}}) + (\mathbf{x}_i - \bar{\mathbf{x}})/(n-1)$, we may compute $\hat{\Gamma}[i]$ according to

$$\hat{\Gamma}[i] = \frac{1}{n-1} \sum_{j=1;\ j\neq i}^{n} \left(\hat{Y}_j[i] - \bar{Y}[i]\right)^2 (\mathbf{x}_j - \bar{\mathbf{x}}[i]). \tag{24}$$

3 Theoretical Interludes

From (1), the vector of 'true' residuals

$$\mathbf{R}^0 \equiv \mathbf{R}^0(\sigma^2, \beta) = \frac{\mathbf{Y} - \mathbf{X}\beta}{\sigma} \tag{25}$$

is equal-in-distribution to the error vector ϵ. If H_0 holds, then the density function of $\mathbf{R}^0$ is

$$f_{\mathbf{R}^0}(\mathbf{r}^0) = \prod_{i=1}^{n} \phi(r_i^0), \tag{26}$$

where $\phi(\cdot)$ is the standard normal density function. Following Neyman's (1937) idea of constructing a 'smooth' test (cf., Thomas and Pierce (1979) and Rayner and Best (1989)), we embed $f_{\mathbf{R}^0}(\mathbf{r}^0)$ into a class of density functions, indexed by $\theta = (\theta_1, \theta_2, \ldots, \theta_6)^{\mathrm{t}}$, whose members are of form

$$f_{\mathbf{R}^0}(\mathbf{r}^0|\theta) = C(\theta) f_{\mathbf{R}^0}(\mathbf{r}^0) \exp\{\theta^{\mathrm{t}} \mathbf{Q}(\mathbf{r}^0)\}, \tag{27}$$

where

$$\mathbf{Q}(\mathbf{r}^0) = \sum_{i=1}^{n} \left[r_i^0, \quad (r_i^0)^2 - 1, \quad (r_i^0)^3, \quad (r_i^0)^4 - 3, \quad [(\mathbf{x}_i - \bar{\mathbf{x}})\beta]^2 r_i^0, \quad (v_i - \bar{v})[(r_i^0)^2 - 1] \right]^{\mathrm{t}}. \tag{28}$$

The function $C(\theta)$ in (27) is a proportionality constant which makes $f_{\mathbf{R}^0}(\mathbf{r}^0|\theta)$ a density function. Also, as mentioned in Section 2, the vector $\mathbf{V}$ is a user-supplied vector. Notice that in the embedding class, the null hypothesis density function obtains when $\theta = \mathbf{0}$.

Let us first consider the case where β and σ^2 are known, so $\mathbf{R}^0 = \mathbf{R}^0(\sigma^2, \beta)$ is observable. Within the embedding class of density functions specified by (27), the score test for $H_0^* : \ \theta = \mathbf{0}$ versus $H_1^* : \ \theta \neq \mathbf{0}$ is easily developed. Indeed, it is straightforward to see that the score test statistic at $\theta = \mathbf{0}$ equals

$$\mathbf{U}(\theta = \mathbf{0}, \sigma^2, \beta) = \mathbf{Q}(\mathbf{R}^0; \sigma^2, \beta). \tag{29}$$

Since under H_0, $R_i^0, (i = 1, 2, \ldots, n)$, are i.i.d. standard normal variables, then for any positive integer k, $\mathbf{E}\{[R_i^0]^{2k+1}\} = 0$ and $\mathbf{E}\{[R_i^0]^{2k}\} = \prod_{j=1}^{k}(2j -$

1), so it is immediate that the covariance matrix of $\frac{1}{\sqrt{n}}\mathbf{Q}(\mathbf{R}^0;\sigma^2,\beta)$ is

$$\mathbf{\Sigma}_{11}^{(n)}(\sigma^2,\beta)=\begin{bmatrix} 1 & 0 & 3 & 0 & \frac{1}{n}T_2(\beta) & 0 \\ 0 & 2 & 0 & 12 & 0 & 0 \\ 3 & 0 & 15 & 0 & \frac{3}{n}T_2(\beta) & 0 \\ 0 & 12 & 0 & 96 & 0 & 0 \\ \frac{1}{n}T_2(\beta) & 0 & \frac{3}{n}T_2(\beta) & 0 & \frac{1}{n}T_4(\beta) & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{n}\sum_{i=1}^n(V_i-\bar{V})^2 \end{bmatrix},$$

where $T_k(\beta)=\sum_{i=1}^n[(\mathbf{x}_i-\bar{\mathbf{x}})\beta]^k,\ k=2,3,4$. If, as $n\to\infty$, the following conditions are satisfied:

(a) There exists a nonsingular $p\times p$ matrix $\mathbf{\Sigma}_X$ such that $\frac{1}{n}T_2(\beta)\xrightarrow{\text{pr}}\beta^{\text{t}}\mathbf{\Sigma}_X\beta$;

(b) There exists a function $\Omega(\beta)$ such that $\frac{1}{n}T_4(\beta)\xrightarrow{\text{pr}}\Omega(\beta)$;

(c) There exists a $\sigma_V^2\in(0,\infty)$ such that $\frac{1}{n}\sum_{i=1}^n(V_i-\bar{V})^2\xrightarrow{\text{pr}}\sigma_V^2$;

(d) $\{\max_{1\le i\le n}[(\mathbf{x}_i-\bar{\mathbf{x}})\beta]^4\}/T_4(\beta)=o_p(1)$; and

(e) $\left\{\max_{1\le i\le n}(V_i-\bar{V})^2\right\}/\left\{\sum_{i=1}^n(V_i-\bar{V})^2\right\}=o_p(1)$;

then it follows from the Lindeberg-Feller Central Limit Theorem (CLT) that, under H_0,

$$\frac{1}{\sqrt{n}}\mathbf{Q}(\mathbf{R}^0;\sigma^2,\beta)\xrightarrow{\text{d}}N\left(\mathbf{0},\mathbf{\Sigma}_{11}(\sigma^2,\beta)\right), \tag{30}$$

where

$$\mathbf{\Sigma}_{11}(\sigma^2,\beta)=\begin{bmatrix} 1 & 0 & 3 & 0 & \beta^{\text{t}}\mathbf{\Sigma}_X\beta & 0 \\ 0 & 2 & 0 & 12 & 0 & 0 \\ 3 & 0 & 15 & 0 & 3\beta^{\text{t}}\mathbf{\Sigma}_X\beta & 0 \\ 0 & 12 & 0 & 96 & 0 & 0 \\ \beta^{\text{t}}\mathbf{\Sigma}_X\beta & 0 & 3\beta^{\text{t}}\mathbf{\Sigma}_X\beta & 0 & \Omega(\beta) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\sigma_V^2 \end{bmatrix}. \tag{31}$$

In this situation where β and σ^2 are assumed known, notice the asymptotic dependence of the components Q_1, Q_3 and Q_5; as well Q_2 and Q_4. For this situation, an asymptotic α-level score test for $H_0^* : \ \theta = \mathbf{0}$ versus $H_1^* : \ \theta \neq \mathbf{0}$ rejects H_0^* whenever

$$\frac{1}{n}\mathbf{Q}(\mathbf{R}^0; \sigma^2, \beta)^{\mathrm{t}}[\boldsymbol{\Sigma}_{11}^{(n)}(\sigma^2, \beta)]^{-}\mathbf{Q}(\mathbf{R}^0; \sigma^2, \beta) \geq \chi^2_{5;\alpha}. \tag{32}$$

However, since σ^2 and β are unknown, then neither $\mathbf{R}^0$ nor $\boldsymbol{\Sigma}_{11}^{(n)}$ are observable. There is therefore a need to plug-in estimators for σ^2 and β in $\mathbf{R}^0(\sigma^2, \beta)$, and by substituting the ML estimators s^2 and $\mathbf{b}$ in (3) and (2), respectively, we obtain the (estimated) residual vector $\mathbf{R} = \mathbf{R}^0(s^2, \mathbf{b})$ given in (4). To develop a test based on $\mathbf{R}$, we need the asymptotic distribution of $\mathbf{Q}(\mathbf{R}; s, \mathbf{b})$ under H_0. Towards this goal, observe that the ML estimating equations for σ^2 and β that give rise to s^2 and $\mathbf{b}$ are

$$A(\mathbf{R}^0(\sigma^2, \beta); \sigma^2, \beta) \equiv \mathbf{R}^0(\sigma^2, \beta)^{\mathrm{t}}\mathbf{R}^0(\sigma^2, \beta) - n = 0; \tag{33}$$

$$\mathbf{B}(\mathbf{R}^0(\sigma^2, \beta); \sigma^2, \beta) \equiv \sigma\mathbf{X}^{\mathrm{t}}\mathbf{R}^0(\sigma^2, \beta) = 0. \tag{34}$$

Augmenting the vector $\mathbf{Q}$ with A and $\mathbf{B}$, then by invoking the Lindeberg-Feller CLT we find that, under H_0 plus the conditions guaranteeing asymptotic normality of $\mathbf{Q}(\mathbf{R}^0(\sigma^2, \beta); \sigma^2, \beta)$ which were enumerated earlier,

$$\frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{Q}(\mathbf{R}^0(\sigma^2, \beta); \sigma^2, \beta) \\ A(\mathbf{R}^0(\sigma^2, \beta); \sigma^2, \beta) \\ \mathbf{B}(\mathbf{R}^0(\sigma^2, \beta); \sigma^2, \beta) \end{bmatrix} \xrightarrow{\mathrm{d}} N\left(\mathbf{0}, \boldsymbol{\Xi}(\sigma^2, \beta)\right), \tag{35}$$

where

$$\boldsymbol{\Xi}(\sigma^2, \beta) = \begin{bmatrix} \boldsymbol{\Sigma}_{11}(\sigma^2, \beta) & \boldsymbol{\Sigma}_{12}(\sigma^2, \beta) \\ \boldsymbol{\Sigma}_{12}(\sigma^2, \beta)^{\mathrm{t}} & \boldsymbol{\Sigma}_{22}(\sigma^2, \beta) \end{bmatrix},$$

with

$$\mathbf{\Sigma}_{12}(\sigma^2,\beta) = \begin{bmatrix} 0 & \sigma\mu_X \\ 2 & 0 \\ 0 & 3\sigma\mu_X \\ 12 & 0 \\ 0 & \sigma[\Gamma(\beta) + (\beta^{\mathrm{t}}\mathbf{\Sigma}_X\beta)\mu_X] \\ 0 & 0 \end{bmatrix};$$

$$\mathbf{\Sigma}_{22}(\sigma^2,\beta) = \begin{bmatrix} 2 & \mathbf{0} \\ \mathbf{0} & \sigma^2(\mathbf{\Sigma}_X + \mu_X^{\mathrm{t}}\mu_X) \end{bmatrix};$$

and with μ_X and $\Gamma(\beta)$ defined according to

$$\frac{1}{n}\bar{\mathbf{x}} \xrightarrow{\mathrm{pr}} \mu_X \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}[(\mathbf{x}_i - \bar{\mathbf{x}}]\beta]^2(\mathbf{x}_i - \bar{\mathbf{x}}) \xrightarrow{\mathrm{pr}} \Gamma(\beta). \tag{36}$$

By virtue of (33) and (34), when s^2 and $\mathbf{b}$ are substituted for σ^2 and β, respectively, then the last two components in the augmented vector are both equal to zero. Consequently, it follows by multivariate normal theory, or it could be established more formally by relying on Pierce's (1982) result, that

$$\frac{1}{\sqrt{n}}\mathbf{Q}(R^0(s^2,\mathbf{b}); s^2, \mathbf{b}) = \frac{1}{\sqrt{n}}\mathbf{Q}(\mathbf{R}; s^2, \mathbf{b}) \xrightarrow{\mathrm{d}} N\left(\mathbf{0}, \mathbf{\Xi}_{11.2}(\sigma^2,\beta)\right) \tag{37}$$

where $\mathbf{\Xi}_{11.2}(\sigma^2,\beta) = \mathbf{\Sigma}_{11}(\sigma^2,\beta) - \mathbf{\Sigma}_{12}(\sigma^2,\beta)\mathbf{\Sigma}_{22}(\sigma^2,\beta)^{-1}\mathbf{\Sigma}_{12}(\sigma^2,\beta)^{\mathrm{t}}$. To provide a simplified form for this limiting covariance matrix, we establish the following intermediate result.

Lemma 1 *If the first column of* $\mathbf{X}$ *is* $\mathbf{1}$*, then* $\mu_X\left(\mathbf{\Sigma}_X + \mu_X^{\mathrm{t}}\mu_X\right)^{-1}\mu_X^{\mathrm{t}} = 1$.

Proof: Let $\mathbf{X} = [\mathbf{1}\ \mathbf{W}]$ so that $\mathbf{\Sigma}_X + \mu_X^{\mathrm{t}}\mu_X = \begin{bmatrix} 1 & \mu_W \\ \mu_W^{\mathrm{t}} & \mathbf{\Sigma}_W + \mu_W^{\mathrm{t}}\mu_W \end{bmatrix}$. Applying the partitioned matrix inverse theorem (cf., Anderson (1984), Th. A.3.3), we obtain

$$[\mathbf{\Sigma}_X + \mu_X^{\mathrm{t}}\mu_X]^{-1} = \begin{bmatrix} 1 + \mu_W\mathbf{\Sigma}_W^{-1}\mu_W^{\mathrm{t}} & -\mu_W\mathbf{\Sigma}_W^{-1} \\ -\mathbf{\Sigma}_W^{-1}\mu_W^{\mathrm{t}} & \mathbf{\Sigma}_W^{-1} \end{bmatrix}.$$

Since $\mu_X = (1 \ \mu_W)$, then the assertion immediately follows by (matrix) multiplication. $\|$

By straightforward multiplication, and applying Lemma 1, we obtain

$$\mathbf{\Delta}(\sigma^2, \beta) \equiv \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^{\mathrm{t}} = \begin{bmatrix} 1 & 0 & 3 & 0 & \beta^{\mathrm{t}}\mathbf{\Sigma}_X\beta & 0 \\ 0 & 2 & 0 & 12 & 0 & 0 \\ 3 & 0 & 9 & 0 & 3\beta^{\mathrm{t}}\mathbf{\Sigma}_X\beta & 0 \\ 0 & 12 & 0 & 72 & 0 & 0 \\ \beta^{\mathrm{t}}\mathbf{\Sigma}_X\beta & 0 & 3\beta^{\mathrm{t}}\mathbf{\Sigma}_X\beta & 0 & \eta(\beta) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{38}$$

with $\eta(\beta) = (\beta^{\mathrm{t}}\mathbf{\Sigma}_X\beta)^2 + \Gamma(\beta)\mathbf{\Sigma}_X^{-1}\mathbf{\Gamma}(\beta)^{\mathrm{t}}$. The matrix $\mathbf{\Delta}(\sigma^2, \beta)$ is the correction factor in the limiting covariance matrix arising from plugging-in s^2 and $\mathbf{b}$ for σ^2 and β, respectively. This factor is clearly non-negligible. Finally, from (31) and (38), a simplified form of $\mathbf{\Xi}_{11.2}$ is

$$\mathbf{\Xi}_{11.2}(\sigma^2, \beta) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 24 & 0 & 0 \\ 0 & 0 & 0 & 0 & \xi(\sigma^2, \beta) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\sigma_V^2 \end{bmatrix}, \tag{39}$$

where $\xi(\sigma^2, \beta) = \Omega(\beta) - (\beta^{\mathrm{t}}\mathbf{\Sigma}_X\beta)^2 - \mathbf{\Gamma}(\beta)\mathbf{\Sigma}_X^{-1}\mathbf{\Gamma}(\beta)^{\mathrm{t}}$. We formally state this asymptotic result as a theorem.

Theorem 1 *If assumptions (A1)-(A4) hold for the linear model in (1) with* $\mathbf{X}$ *having as its first column the vector* $\mathbf{1}$*, and if conditions (a)-(e) enumerated earlier hold, then* $n^{-1/2}\mathbf{Q}(\mathbf{R}; s^2, \mathbf{b})$ *converges in distribution to a zero-mean normal distribution with covariance matrix* $\mathbf{\Xi}_{11.2}$ *given in (39).*

Note the invariance of this asymptotic result to re-scaling, that is, the result is independent of σ. This is a consequence of the facts that the model is scale-invariant and the residual vector is scale-equivariant. The theorem also indicates that $Q_1(\mathbf{R}; s^2, \mathbf{b})$ and $Q_2(\mathbf{R}; s^2, \mathbf{b})$ are degenerate at zero, hardly a surprise since these quantities are the estimating functions for σ^2 and β. What is surprising, instead, is the asymptotic independence of $Q_3(\mathbf{R}; s^2, \mathbf{b})$ and $Q_5(\mathbf{R}; s^2, \mathbf{b})$, since as noted earlier, $Q_3(\mathbf{R}^0; \sigma^2, \beta)$ and $Q_5(\mathbf{R}^0; \sigma^2, \beta)$ are *not* asymptotically independent. Evidently, the process of replacing the unknown parameters by their ML estimators in the quantities $\mathbf{Q}(\mathbf{R}^0(\sigma^2, \beta); \sigma^2, \beta)$ made all the components asymptotically independent!

The quantities $\Omega(\beta)$, $\mathbf{\Sigma}_X$, and $\mathbf{\Gamma}(\beta)$ can be consistently estimated by their empirical counterparts and with β replaced by $\mathbf{b}$. Their respective estimators are those given in (11), and so we are able to obtain a consistent estimator $\hat{\mathbf{\Xi}}_{11.2}$ of $\mathbf{\Xi}_{11.2}$. The score statistic for testing $H_0^* : \theta = \mathbf{0}$ versus $H_1^* : \theta \neq \mathbf{0}$, with σ^2 and β considered as nuisance parameters, is the quadratic form of $\frac{1}{\sqrt{n}}\mathbf{Q}(\mathbf{R}; s^2, \mathbf{b})$ with quadratic matrix $\hat{\mathbf{\Xi}}_{11.2}^{-1}$. It is immediate to see that this statistic is

$$\frac{1}{n}\mathbf{Q}(\mathbf{R}; s^2, \mathbf{b})^{\mathrm{t}}\hat{\mathbf{\Xi}}_{11.2}^{-}\mathbf{Q}(\mathbf{R}; s^2, \mathbf{b}) = \hat{S}_1^2 + \hat{S}_2^2 + \hat{S}_3^2 + \hat{S}_4^2 = \hat{G}_4^2, \qquad (40)$$

where $\hat{S}_k^2, (k = 1, 2, 3, 4)$, and $\hat{G}_4^2$ are as defined in (8),(9), (10), (12), and (13), respectively. Theorem 1 therefore justifies the use of the chi-squared distribution with four degrees-of-freedom for assessing the magnitude of $\hat{G}_4^2$, as well as the one degree-of-freedom chi-squared distributions for each of the

component statistics.

4 Monte Carlo Adventures

To examine the sensitivity of the procedures for detecting different types of departures from assumptions (A1)-(A4), computer simulation studies were performed. Each set of runs, coinciding with a particular model, consisted of 2000 replications and involved sample sizes $n \in \{30, 100, 200\}$. For each set of runs, one fixed covariate sequence $x_1, x_2, \ldots, x_n$ was generated according to a standard uniform distribution. The simulation program was coded in `S-Plus`, in particular, random variates were generated using the random number generators in `S-Plus`, and for the linear model fitting, the `S-Plus` object `lm` was utilized.

The first set of runs was for the purpose of determining if the procedures achieve the pre-specified level of significance for the three sample sizes considered. The level of significance was set to 5%. For each run of this set, the response values were generated according to the model

$$Y_i = x_i + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{41}$$

where ϵ_i's were generated from a standard normal distribution. For the resulting data, $(Y_i, x_i), i = 1, 2, \ldots, n$, the model

$$Y_i = \beta_0 + \beta_1 x_i + \sigma\epsilon_i, \quad i = 1, 2, \ldots, n, \tag{42}$$

was fitted and the resulting residuals, $R_i, (i = 1, 2, \ldots, n)$, were utilized in the testing procedure. The user-supplied vector $\mathbf{V}$ was taken to be the default

time-sequence. For the 2000 replications, the percentage of rejection of H_0 was recorded. Table 1 summarizes the observed empirical levels. Except for the two values of 4.00% for $n = 30$, the observed levels are within two standard errors of 5%. It appears therefore that the asymptotic chi-square approximation is acceptable, at least for the specific model in this simulation run.

The first type of violation examined was when the error distribution is not normal. We considered several types of error distributions, broadly classified into symmetric and skewed distributions. The true model is as in (41), but with ϵ_i's having non-normal distributions. The model fitted to the data is (42). Table 2 presents the simulated powers of the tests when the error distribution is of a symmetric variety. The first four distributions are Student's t-distribution with different degrees-of-freedom. The distribution t_1 is of course the standard Cauchy distribution, so not only is (A4) violated, but even (A1) and (A2) are also violated as the moments of this distribution do not exist. The last three distributions are the standard logistic, standard double exponential, and a centered (i.e., with zero-mean) uniform whose variance is unity. By examining these simulated powers, we observe that the detection ability of the global test is quite good relative to the best directional test based on the four component statistics, with its power not significantly degraded by combining the four statistics. The best directional test is that based on the statistic $\hat{S}_2^4$, which we recall is a kurtosis-type statistic. Notice that the test based on $\hat{S}_3^2$ does not have any power for detecting this error

distribution mis-specification. It is interesting to observe that when $n = 30$ and with the uniform error distribution, the tests based on $\hat{S}_2^2$ and the global statistic have very low powers, but for large sample sizes their powers are very good. Based on other runs performed, this behavior remains invariant under a change in the variance of the uniform error distribution, as well as when the covariate vector is instead generated according to a normal distribution. As of yet we are unable to explain this puzzling phenomenon! In contrast, note that the other three tests could not detect uniform distributed errors! Also, when the degrees-of-freedom of the t-distribution increases, then the power of the tests decreases; however, this is to be expected since as the degrees-of-freedom increases, then the t distribution becomes closer to the normal distribution.

Table 3 summarizes the power results when the errors have shifted (to have mean zero) chi-squared distributions, which are right-skewed distributions. As in the symmetric error distributions, the global test performs acceptably relative to the best test among the four directional tests, with the powers slightly degraded due to combining the four directional tests, some of which do not have good power against this assumptional departure. The best directional test for this skewed class of distributions is based on $\hat{S}_1^2$, which is the skewness-type statistic. Again, the test based on $\hat{S}_3^2$ does not have any detection power for this alternative.

The next set of simulation runs concerns the situation where (A2) is violated, so that the conditional variances of the Y_i's are not equal. Two

models were considered for this purpose. The first model has variances that depend on the covariate values. Specifically, the true model is

$$Y_i = x_i + x_i^{\gamma}\epsilon_i, \quad i = 1, 2, \ldots, n, \tag{43}$$

where ϵ_i's are i.i.d. from $N(0,1)$. The fitted model was that in (42). The simulated powers for $\gamma \in \{.5, 1, 2\}$ are summarized in Table 4. The best directional test for this departure is the $\hat{S}_2^2$-test, with the global test also performing acceptably. Again, the test based on $\hat{S}_3^2$ has very low power for this heteroscedastic model. The second model for heteroscedastic variances is of form

$$Y_i = \begin{cases} x_i + \sigma_1\epsilon_i & \text{for} \quad i \leq n/2 \\ x_i + \sigma_2\epsilon_i & \text{for} \quad i > n/2 \end{cases}, \tag{44}$$

with ϵ_i's also i.i.d. from $N(0,1)$. Table 5 presents the simulated powers when model (42) is fitted for two sets of values of (σ_1, σ_2). It is interesting to observe that the global test dominates the directional tests! This could be a consequence of the fact that the directional tests based on $\hat{S}_1^2$, $\hat{S}_2^2$, and $\hat{S}_4^2$ have high powers, and the combination of these tests made the global test more powerful.

The next set of runs were for mis-specified link functions, that is, when (A1) is violated. The data analyzed were generated according to the model

$$Y_i = x_i + \beta_2 x_i^{\gamma} + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{45}$$

with ϵ_i's i.i.d. from $N(0,1)$. Model (42) was fitted to the resulting data. Table 6 provides a summary of the simulated powers of the tests for different

sets of (β_2, γ). Interestingly, the directional tests based on $\hat{S}_1^2$, $\hat{S}_2^2$, and $\hat{S}_4^2$ are not sensitive to this violation. The best directional test is that based on $\hat{S}_3^2$, with the power of the global test quite degraded relative to the $\hat{S}_3^2$-test possibly because the other three tests have no power to detect this alternative. Notice that when $\beta_2 = 1$ and $\gamma \in \{.5, 2\}$, the powers of the tests are very low. This could be a consequence of the fact that for this parameter set, the *signal-to-noise* ratio (SNR) is very low. This SNR may be measured via

$$\text{SNR} = \frac{\mathbf{E}\{\text{MSE(Fitted)}|\text{True}\} - \mathbf{E}\{\text{MSE(True)}|\text{True}\}}{\mathbf{E}\{\text{MSE(True)}|\text{True}\}}, \tag{46}$$

with the notation that $\mathbf{E}\{\text{MSE(Model A)}|\text{Model B}\}$ is the expectation of the mean-squared error when Model A is fitted with the expectation evaluated with respect to Model B. Thus, $\mathbf{E}\{\text{MSE(True)}|\text{True}\} = \sigma^2$. To obtain the relevant SNR for the mis-specified model considered in the simulation, consider the more general model given by

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{V}\omega + \sigma\epsilon \tag{47}$$

to be the true model, where $\mathbf{V}$ is of full rank with rank q.

Lemma 2 *Let* $\text{MSE} = \{\mathbf{Y}^{\text{t}}(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}\}/(n-p)$ *be the mean-squared error from fitting the model* $\mathbf{Y} = \mathbf{X}\beta + \sigma\epsilon$*, and assume model (47) holds. Then*

$$\mathbf{E}\{\text{MSE}\} = \sigma^2 + \frac{(\mathbf{V}\omega)^{\text{t}}(\mathbf{I} - \mathbf{P}_X)(\mathbf{V}\omega)}{n-p}.$$

Proof: For notational convenience, for an $n \times r$ matrix $\mathbf{A}$, $\mathbf{P}[\mathbf{A}]$ denotes the projection operator on the linear subspace generated by the columns of $\mathbf{A}$.

Then, we have

$$\mathbf{Y}^{\mathrm{t}}(\mathbf{I}-\mathbf{P}[\mathbf{X}])\mathbf{Y} = \mathbf{Y}^{\mathrm{t}}(\mathbf{I}-\mathbf{P}[\mathbf{X},\mathbf{V}])\mathbf{Y} + \mathbf{Y}^{\mathrm{t}}(\mathbf{P}[\mathbf{X},\mathbf{V}]-\mathbf{P}[\mathbf{X}])\mathbf{Y}.$$

Under model (47), the expectation of the first term on the right-hand side is $(n-p-q)\sigma^2$, and $\mathbf{E}\{\mathbf{Y}\} = \mu \equiv \mathbf{X}\beta + \mathbf{V}\omega$ and $\mathbf{Cov}\{\mathbf{Y}\} = \sigma^2\mathbf{I}$. Therefore, by projection properties,

$$\begin{aligned}
&\mathbf{E}\{\mathbf{Y}^{\mathrm{t}}(\mathbf{P}[\mathbf{X},\mathbf{V}]-\mathbf{P}[\mathbf{X}])\mathbf{Y}\} \\
&= \sigma^2\mathrm{trace}(\mathbf{P}[\mathbf{X},\mathbf{V}]-\mathbf{P}[\mathbf{X}]) + (\mathbf{X}\beta+\mathbf{V}\omega)^{\mathrm{t}}(\mathbf{P}[\mathbf{X},\mathbf{V}]-\mathbf{P}[\mathbf{X}])(\mathbf{X}\beta+\mathbf{V}\omega) \\
&= \sigma^2 q + \left\{\mu^{\mathrm{t}}\mu - \left[\mu^{\mathrm{t}}\mu - (\mathbf{V}\omega)^{\mathrm{t}}(\mathbf{I}-\mathbf{P}[\mathbf{X}])(\mathbf{V}\omega)\right]\right\} \\
&= \sigma^2 q + (\mathbf{V}\omega)^{\mathrm{t}}(\mathbf{I}-\mathbf{P}[\mathbf{X}])(\mathbf{V}\omega).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbf{E}\{\mathbf{Y}^{\mathrm{t}}(\mathbf{I}-\mathbf{P}[\mathbf{X}])\mathbf{Y}\} &= \sigma^2(n-p-q) + \sigma^2 q + (\mathbf{V}\omega)^{\mathrm{t}}(\mathbf{I}-\mathbf{P}[\mathbf{X}])(\mathbf{V}\omega) \\
&= \sigma^2(n-p) + (\mathbf{V}\omega)^{\mathrm{t}}(\mathbf{I}-\mathbf{P}[\mathbf{X}])(\mathbf{V}\omega)
\end{aligned}$$

completing the proof. ‖

Using the result in Lemma 2, when the true model is (47) and model (1) is fitted, then the SNR is

$$\mathrm{SNR} = \frac{1}{n-p}\left(\mathbf{V}\frac{\omega}{\sigma}\right)^{\mathrm{t}}(\mathbf{I}-\mathbf{P}_X)\left(\mathbf{V}\frac{\omega}{\sigma}\right). \tag{48}$$

As expected, note that SNR becomes zero whenever $\omega = \mathbf{0}$ or $\mathbf{V} = \mathbf{AX}$, i.e., when the columns of $\mathbf{V}$ are in the linear space generated by the columns of $\mathbf{X}$. Indeed, recognize that the SNR in (48) is the mean-squared error (MSE)

obtained by regressing $\mathbf{V}\omega/\sigma$ on $\mathbf{X}$. Applying this general result to the model used in the simulation, we have the correspondences

$$p = 2,\ \mathbf{X} = [\mathbf{1}\ \ \mathbf{x}]\,,\ \mathbf{V} = \mathbf{x}^\gamma \equiv (x_1^\gamma, \ldots, x_n^\gamma)^{\mathrm{t}},\ \text{and}\ \omega = \beta_2,$$

so for this model, (48) simplifies to

$$\mathrm{SNR} = \frac{1}{n-2}\left(\frac{\beta_2}{\sigma}\right)^2 \sum_{i=1}^{n}[x_i^\gamma - \mathrm{mean}(\mathbf{x}^\gamma)]^2 \left[1 - \mathrm{corr}(\mathbf{x}, \mathbf{x}^\gamma)^2\right]. \tag{49}$$

Since the x_i's were from a standard uniform distribution, for large n, we obtain the approximations

$$\frac{1}{n}\sum_{i=1}^{n}[x_i^\gamma - \mathrm{mean}(\mathbf{x}^\gamma)]^2 \approx \frac{\gamma^2}{(\gamma+1)^2(2\gamma+1)} \quad \text{and} \quad \mathrm{corr}(\mathbf{x}, \mathbf{x}^\gamma)^2 \approx \frac{3(2\gamma+1)}{(\gamma+2)^2}.$$

Using these approximations, for large n, the expression in (49) can be approximated by

$$\mathrm{SNR}(\beta_2, \gamma, \sigma) \approx \left(\frac{\beta_2}{\sigma}\right)^2 \frac{\gamma^2}{(\gamma+1)^2(2\gamma+1)}\left\{1 - \frac{3(2\gamma+1)}{(\gamma+2)^2}\right\}. \tag{50}$$

For the values of $(\beta_2, \gamma, \sigma)$ utilized in the simulation studies, we obtain $\mathrm{SNR}(1, .5, 1) \approx \frac{1}{450}$, $\mathrm{SNR}(1, 2, 1) \approx \frac{1}{180}$, $\mathrm{SNR}(3, .5, 1) \approx \frac{1}{50}$, $\mathrm{SNR}(3, 2, 1) \approx \frac{1}{20}$, $\mathrm{SNR}(5, .5, 1) \approx \frac{1}{18}$, and $\mathrm{SNR}(5, 2, 1) \approx \frac{5}{36}$. These values explain the ordering of the simulated powers for the $\hat{S}_3^2$-based test. Note in particular that $\mathrm{SNR}(5, .5, 1) \approx 1/18$ is slightly larger than $\mathrm{SNR}(3, 2, 1) \approx 1/20$, and this is reflected by the small differences in the observed powers for the $\hat{S}_3^2$-based test for these two sets of values of (β_2, γ).

The final set of simulation runs concerns violations of assumption (A3). We considered two models for generating dependent error terms. The first

model, which endows the error sequence a martingale structure, is $Y_i = x_i + \epsilon_i, i = 1, 2, \ldots, n$, where, with ϵ_j^*'s i.i.d. from $N(0, 1)$,

$$\epsilon_i = \frac{1}{\sqrt{i}} \sum_{j=1}^{i} \epsilon_j^*. \tag{51}$$

The second class of models has a Markov structure for the error sequence. In this model, the error sequence is defined according to

$$\epsilon_1 = \epsilon_1^* \quad \text{and} \quad \epsilon_i = \frac{\rho\epsilon_{i-1} + \epsilon_i^*}{\sqrt{1+\rho^2}}, \quad i = 2, \ldots, n, \tag{52}$$

with ρ being a dependence parameter. In the simulation, we performed runs for $\rho = 1$ and $\rho = 3$. Table 7 summarizes the simulated powers of the tests under these dependent error models. Observe that for the martingale error structure, the best test is the global test, with the $\hat{S}_4^2$-test also performing very well. For the Markov error structure, the best is the $\hat{S}_4^2$-test, with the global test's power also very acceptable. The tests based on $\hat{S}_1^2$ and $\hat{S}_2^2$ also has some detection abilities for this type of assumptional departure, but are not competitive with the $\hat{S}_4^2$-based test or the global test. The test based on $\hat{S}_3^2$ possesses no ability to detect this particular type of violation.

5 Illustrative Examples

Example 1: The first illustration involves Forbes' bivariate data set discussed in both Weisberg (1980, pp. 2-4) and Atkinson (1985) which contains 17 observations on the boiling point (in degrees Fahrenheit) of water at different pressures (in inches of mercury), with the different pressures arising

from the different elevations in the Alps at which the measurements were taken. As had been discussed in these monographs, there is a very strong fit between pressure and boiling point, except that an examination of the residuals reveals the unusual nature of the 12th observation. This is demonstrated by the first four plots in Figure 1, especially the third and fourth plots. Recognizing the limitation that the sample size of this data set may not be large enough to achieve good approximations for our proposed procedures, we nevertheless applied the global test procedure and obtained the value $\hat{G}_4^2 = 98.45$ with p-value approximately zero, so we conclude that at least one of (A1)-(A4) is violated. Computing the component statistics, together with their approximate p-values, we find $\hat{S}_1^2 = 28.71(p \approx 0)$; $\hat{S}_2^2 = 65.08(p \approx 0)$; $\hat{S}_3^2 = 1.90(p \approx .1675)$, and $\hat{S}_4^2 = 2.75(p \approx .0970)$. A plot of $\Delta\hat{G}_2^4[i]$ versus the observation number is provided in the fifth graph in Figure 1, and a plot of the deletion p-values $p[i]$ versus observation number is the last graph in Figure 1. Examining these plots, it is evident that observation 12 is truly unusual. In particular, notice from the last graph in Figure 1 that the p-value when observation 12 is deleted is very different from all the other p-values. This indicates that the rejection of H_0 is mainly caused by the 12th observation. When the data is re-analyzed with observation 12 deleted, we find the following values of the test statistics, together with their approximate p-values: $\hat{G}_4^2[12] = 2.54(p \approx .64)$; $\hat{S}_1^2[12] = 1.06(p \approx .30)$; $\hat{S}_2^2[12] = .26(p \approx .61)$; $\hat{S}_3^2[12] = 1.21(p \approx .27)$, and $\hat{S}_4^2[12] = .01(p \approx .90)$. An examination of the $\Delta\hat{G}_4^2[i]$ and $p[i]$ plots for this 'new' data set did not show any unusual

observations. These results indicate that by deleting the 12th observation, assumptions (A1)-(A4) become acceptable.

Example 2: The second example involves multiple regression analyses of Ruppert and Carroll's (1980) water salinity data (see Table 3 in their paper) which they used to illustrate robust regression techniques, and which was also used for illustrative purposes in Atkinson (1985). We use this data set to demonstrate the possible utility of the deletion statistics for detecting outlying and/or influential observations, as well as in model construction in conjunction with the model validation statistics. The data set consisted of 28 observations on the variables Salinity, which is the water salinity at the specified time period; LagSalinity, which is the water salinity lagged two weeks; Trend, which represents one of the six biweekly periods in March to May; and WaterFlow, which is the river discharge. The response variable is Salinity, while the predictor variables are LagSalinity, Trend, and WaterFlow. The first part of the analyses involved fitting the multiple regression model

$$\text{Salinity} = \beta_0 + \beta_1(\text{LagSalinity}) + \beta_2(\text{Trend}) + \beta_3(\text{WaterFlow}) + \sigma\epsilon. \quad (53)$$

The fitted model had $b_0 = 9.5903$, $b_1 = .7771$, $b_2 = -.0255$, and $b_3 = -.2950$. The coefficients β_0, β_1, and β_3 were significantly different from zero. The multiple R^2 was 82.64%. When we applied the model validation procedures proposed in this paper, we found the following values of the test statistics, together with their p-values: $\hat{G}_4^2 = .16(p = .9971)$, $\hat{S}_1^2 = .02(p = .87)$, $\hat{S}_2^2 = .005(p = .95)$, $\hat{S}_3^2 = 0(p = 1)$, and $\hat{S}_4^2 = .13(p = .72)$. Thus, these

values seem to indicate that the assumptions are acceptable. However, an examination of the plots of $\Delta G_4^2[i]$ and $\Delta p[i]$ versus observation number, which are provided in the first two panels in Figure 2 vividly reveals that observation number 16 is quite an unusual observation. In Atkinson (1985), the unusual nature of the 16th observation was revealed using a half-normal plot of Cook's (1977) statistic. Following Atkinson, we suppose that the value of WaterFlow for this 16th observation, which was 33.443, was actually a misprint for 23.443, and so we re-fitted the model in (53) but with 23.443 in place of 33.443. The resulting analysis yielded the estimates $b_0 = 18.35$, $b_1 = .70$, $b_2 = -.15$, and $b_3 = -.63$, with β_0, β_1, and β_3 significantly different from zero. The multiple R^2 was 89.26%. Applying the model validation procedures, we obtained $\hat{G}_4^2 = 6.66(p = .15)$, $\hat{S}_1^2 = 1.37(p = .24)$, $\hat{S}_2^2 = .02(p = .87)$, $\hat{S}_3^2 = 4.24(p = .04)$, and $\hat{S}_4^2 = 1.03(p = .31)$. Though the global statistic has p-value exceeding 10%, the p-value for $\hat{S}_3^2$ is .04, which seems to indicate that there is a mild problem in the link function (cf., see the results of the simulations pertaining to the mis-specified link function). The plots of $\Delta\hat{G}_2^4[i]$ and $\Delta p[i]$ from this model fitting, given in the third and fourth panels in Figure 2, indicate no unusual observations, except possibly for the 5th observation. Recognizing the possible problem with the link function, we follow Atkinson's (1985, p. 50) suggestion of incorporating a quadratic term of WaterFlow and we therefore fitted the model

$$\text{Salinity} = \beta_0 + \beta_1(\text{LagSalinity}) + \beta_2(\text{Trend}) + \beta_3(\text{WaterFlow}) + \beta_4(\text{WaterFlow})^2 + \sigma\epsilon. \tag{54}$$

The resulting estimates are $b_0 = 67.57$, $b_1 = .68$, $b_2 = -.25$, $b_3 = -4.58$, and $b_4 = .08$, and the multiple R^2 was 91.64% . Only β_2 did not turn out to be significantly different from zero. Applying our model validation procedures, we find $\hat{G}_4^2 = 1.69(p = .79)$, $\hat{S}_1^2 = 1.12(p = .29)$, $\hat{S}_2^2 = .03(p = .86)$, $\hat{S}_3^2 = .19(p = .66)$, and $\hat{S}_4^2 = .35(p = .55)$. The plots of $\Delta\hat{G}_4^2[i]$ and $\Delta p[i]$ versus observation number are provided in the fifth and sixth panels of Figure 2. The values of the test statistics indicate that the assumptions are viable, and the plots of $\Delta\hat{G}_4^2[i]$ and $\Delta p[i]$ do not anymore show any unusual observations, so fitting the model (54) on the 'corrected' data set have yielded an acceptable model.

Example 3: A well-known data set (see Table 4 in Box and Cox (1964)) arising from a designed textile experiment is the wool data set consisting of 27 observations on the number of cycles to failure of a worsted yarn for the 3^3 combinations of the levels of factors Length with levels of 250mm, 300mm, and 350mm; Loading Amplitude with levels of 8mm, 9mm, and 10mm; and Load with levels of 40g, 45g, and 50g. This data set was used by Box and Cox (1964) to illustrate the family of power transformations; see also Atkinson (1985, pp. 81–84). We use this data set to demonstrate how our validation statistics could be employed to assess the viability of a fitted model. If an additive model is fitted to this data set with Cycles to Failure as response variable, a test of significance of the model reveals a p-value of 1.02×10^{-6} and a multiple R^2 equal to 72.91%. Applying the model validation procedure, the global statistic took a value of $\hat{G}_4^2 = 31.91$ with p-value of

2.00×10^{-6} indicating serious problems with at least one of the assumptions (A1)-(A4). An examination of the plots of $\Delta\hat{G}_4^2[i]$ and $\Delta p[i]$ also reveal that the 19th and 20th observations are outlying and/or highly influential. These results are consistent with those of Atkinson's (1985). Next, we fitted the same additive model, but with the response being the logarithm of Cycles to Failure. Again, the model turned out to be significant with p-value of zero and multiple R^2 equal to 96.58%. Computing the model validation statistics, we obtained $\hat{G}_4^2 = 5.81(p = .21)$ and $\hat{S}_1^2 = .87(p = .35)$, $\hat{S}_2^2 = .10(p = .75)$, $\hat{S}_3^2 = .38(p = .53)$, and $\hat{S}_4^2 = 4.45(p = .0348)$. Thus, these results indicate a mild problem with the homoscedasticity assumption, but definitely also demonstrates that a logarithmic transformation on the response provides a better fit. We also fitted the additive model with response variable being the reciprocal of Cycles to Failure. As in the two previous models, the fitted model turned out to be significant with p-value of zero and multiple R^2 of 76.57%. However, the global statistic yielded $\hat{G}_4^2 = 40.34$ with an associated p-value of 3.68×10^{-8}. This demonstrates that among the three models considered, the one utilizing a logarithmic transformation resulted in a fitted model where the assumptions (A1)-(A4) maybe acceptable. Of course, these results are consistent with those of Atkinson's (1985) which were arrived at using other methods, and the fact that logarithmic transformation resulted in the best fitting model has been established in Box and Cox (1964).

Example 4: For our last example, we considered the end-of-trading-day share values of CREF's Stock and Growth Accounts consisting of successive

observations starting from January 2, 1996 until May 31, 1996. The data set was downloaded from TIAA-CREF's website

`http://www.tiaa-cref.org/financials/selection/ann-select.html`.

The downloadable data set in TIAA-CREF's website included the share values of the accounts even for non-trading days, such as Saturdays, Sundays, and holidays, so the values for these days equal those of the preceding trading day. Before the analyses were performed, these non-trading day values were removed which left a total of $n = 106$ observations. The cleaned-out data set used in this example is available upon request from the author, or can be downloaded from his website. The goal of the model fitting is to relate the values of the Stock and Growth retirement accounts for the purpose of predicting the share value of the Growth Account from the share value of the Stock Account. A bivariate plot of the data set is provided in the first panel on Figure 3. The second panel in this figure is a plot of Growth share values versus the Time Sequence. The third panel plots ΔStock versus ΔGrowth, where for $i = 1, 2, \ldots, n$,

$$\Delta\text{Growth}_{i+1} = \text{Growth}_{i+1} - \text{Growth}_i \quad \text{and} \quad \Delta\text{Stock}_{i+1} = \text{Stock}_{i+1} - \text{Stock}_i,$$

which are the first-order differences. The last panel is a plot of ΔGrowth versus Time.

We first fitted a simple linear regression model with Growth as response variable and Stock as predictor variable. The fitted model was Growth $= -12.0942 + .5178(\text{Stock})$ with coefficient of determination $R^2 = 98.8\%$ and

residual standard error of .1623 on 104 error degrees-of-freedom. A test for the significance of the model yielded a p-value of zero. We applied our procedure for the purpose of validating the linear model assumptions. The value of the global statistic was $\hat{G}_4^2 = 7.87$ with p-value of .0965. On the basis of this global test, at 5% level of significance we may conclude that the four assumptions are acceptable, though at a 10% level of significance, the result indicates a violation of at least one of the assumptions. We therefore examined the four directional statistics, whose values, together with their p-values, were: $\hat{S}_1^2 = 2.32(p = .1281)$; $\hat{S}_2^2 = .55(p = .4568)$; $\hat{S}_3^2 = 4.65(p = .0311)$, and $\hat{S}_4^2 = .35(p = .5548)$. The value of $\hat{S}_3^2$ indicates a violation in the link function. These results are supported by the residual plots in Figure 4, specifically by looking at the fourth panel which contains a plot of the time sequence and the residuals, and to a lesser extent and with a dose of subjective judgement, from the first panel, which is a plot of the residuals and the fitted values. The second and third panels are the usual histogram and normal probability plot of the residuals, and except for a slight right-skewness (of course, subjectively assessed) these seem to indicate that the normality assumption is more or less satisfied. The last two panels are plots of $\Delta\hat{G}_4^2[i]$'s and $p[i]$'s with respect to the time sequence, and these plots seem to indicate that there are no obvious outliers and/or extremely influential observations.

By virtue of the conclusions arising from the model validation procedures above and to try to eliminate the possible effect of time, we fit-

ted a simple linear regression model but with ΔGrowth as dependent variable and ΔStock as predictor variable. The resulting fitted model was $\Delta\text{Growth} = .0057 + .4760(\Delta\text{Stock})$ with a coefficient of determination equal to $R^2 = 92.86\%$ and residual standard error of .07828 based on 103 degrees of freedom. The fitted model was again found to be significant with p-value of zero. Performing the model validation, the global statistic was $\hat{G}_4^2 = 2.81$ with p-value of .59. Therefore, at a significance level of 10%, the null hypothesis that (A1)-(A4) for this model hold cannot be rejected. The directional statistics arising from this model were $\hat{S}_1^2 = .11(p = .73)$; $\hat{S}_2^2 = .0041(p = .95)$; $\hat{S}_3^2 = .17(p = .68)$, and $\hat{S}_4^2 = 2.51(p = .11)$. These values confirm that the assumptions are quite viable for the model which utilizes the first-order differences in share values, though the p-value of $\hat{S}_4^2$ is very close to .10 which may be indicative of a mild heteroscedasticity or dependent errors. Figure 5 presents the relevant plots for this model fitting. The first four panels show no unusual patterns, which is consistent with the quantitative values provided by the global and directional statistics. The last two panels, which are plots of $\Delta\hat{G}_4^2[i]$'s and $p[i]$'s with respect to time sequence, also indicate no unusual patterns, except maybe for 7th, 48th, and 54th observations which could be outliers or mildly influential observations.

References

[1] Anderson, T. (1984). *An Introduction to Multivariate Statistical Analysis*. 2nd edition. John Wiley and Sons: New York.

[2] Andrews, D. (1971). Significance Tests Based on Residuals. *Biometrika*, **58**, 139–148.

[3] Anscombe, F. (1961). Examination of residuals. *Proc. Fourth Berkeley Symp.*, **1**, 1–36.

[4] Anscombe, F. and Tukey, J. (1963). The Examination and Analysis of Residuals. *Technometrics*, **5**, 141–160.

[5] Atkinson, A. (1985). *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Clarendon Press: Oxford.

[6] Bickel, P. (1978). Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. *Annals of Statistics*, **6**, 266–291.

[7] Box, G. and Cox, D. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, A*, **143**, 383–430.

[8] Cook, R. (1977). Detection of influential observations in linear regression. *Technometrics*, **19**, 15–18.

[9] Cook, R. (1998). *Regression Graphics*. John Wiley & Sons: New York.

[10] Cook, R. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall: New York.

[11] Cook, R. and Weisberg, S. (1983). Diagnostics for Heteroscedasticity in Regression. *Biometrika*, **70**, 1–10.

[12] Durbin, J. and Watson, G. (1950). Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*, **37**, 409-428.

[13] Durbin, J. and Watson, G. (1951). Testing for Serial Correlation in Least Squares Regression: II. *Biometrika*, **38**, 159–178.

[14] Kianifard, F. and Swallow, W. (1996). A Review of the Development and Application of Recursive Residuals in Linear Models. *Journal of the American Statistical Association*, **91**, 391–400.

[15] Neyman, J. (1937). "Smooth" test for goodness of fit. *Skand. Aktuarietidskr.*, **20**, 150–199.

[16] Pierce, D. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Annals of Statistics*, **10**, 475–478.

[17] Rayner, J. and Best, D. (1986). Neyman-Type Smooth Test for Location-Scale Families. *Biometrika*, **73**, 437–446.

[18] Rayner, J. and Best, D. (1989). *Smooth Tests of Goodness of Fit*. Oxford University Press: New York.

[19] Ruppert, D. and Carroll, R. (1980). Trimmed Least Squares Estimation in the Linear Model. *Journal of the American Statistical Association*, **75**, 828–838.

[20] Theil, H. (1965). The Analysis of Disturbances in Regression Analysis. *Journal of the American Statistical Association*, **60**, 1067–1079.

[21] Theil, H. and Nagar, A. (1961). Testing the Independence of Regression Disturbances. *Journal of the American Statistical Association*, **56**, 793–806.

[22] Thomas, D. and Pierce, D. (1979). Neyman's smooth goodness-of-fit test when the hypotheses is composite. *Journal of the American Statistical Association*, **74**, 441–445.

[23] Tukey, J. (1949). One degree of freedom for nonadditivity. *Biometrics*, **5**, 232-242.

Model Characteristic	Sample Size (n)	Component Statistics				Global Statistic
		$\hat{S}_1^2$	$\hat{S}_2^2$	$\hat{S}_3^2$	$\hat{S}_4^2$	$\hat{G}_4^2$
True Model	30	4.00	4.00	5.05	5.75	5.10
	100	5.50	4.20	4.35	4.70	5.95
	200	5.70	4.60	4.40	4.05	5.75

Table 1: Achieved levels of each of the 5%-asymptotic level tests based on 2000 replications.

Error Distribution	Sample Size (n)	Component Statistics				Global Statistic
		$\hat{S}_1^2$	$\hat{S}_2^2$	$\hat{S}_3^2$	$\hat{S}_4^2$	$\hat{G}_4^2$
t_1	30	78.30	90.65	4.40	42.20	90.55
	100	94.95	100	4.15	67.00	100
	200	97.85	100	4.35	75.10	100
t_5	30	18.45	20.35	4.70	9.60	22.40
	100	34.30	57.00	4.40	13.80	54.55
	200	42.25	83.10	4.55	15.15	80.50
t_{10}	30	8.65	8.75	5.20	5.45	10.25
	100	16.70	25.45	4.90	7.55	27.35
	200	17.15	39.50	6.05	8.50	35.70
t_{20}	30	6.05	5.85	4.50	5.00	7.00
	100	8.10	11.65	5.10	5.25	12.35
	200	10.30	17.40	5.50	6.60	16.60
Logistic	30	11.80	12.60	5.90	7.20	15.05
	100	17.45	30.30	5.50	8.20	29.35
	200	20.10	52.35	4.25	9.00	47.10
Double Exp.	30	19.50	24.75	5.60	10.35	27.20
	100	35.05	73.55	5.60	14.60	70.65
	200	39.45	95.95	6.35	14.05	92.90
Centered Uniform (Variance=1)	30	4.15	0	4.55	5.10	0.10
	100	4.20	67.25	5.85	4.70	3.90
	200	5.05	100	4.45	4.25	86.40

Table 2: Achieved powers of each of the 5%-asymptotic level tests based on 2000 replications when the error distribution is of a symmetric type. t_k represents a Student's t-distribution with k degrees-of-freedom.

Error Distribution	Sample Size (n)	Component Statistics				Global Statistic
		$\hat{S}_1^2$	$\hat{S}_2^2$	$\hat{S}_3^2$	$\hat{S}_4^2$	$\hat{G}_4^2$
$\chi_1^2 - 1$	30	91.30	59.70	5.25	21.30	80.20
	100	100	98.35	5.05	31.35	100
	200	100	99.95	4.85	33.60	100
$\chi_2^2 - 2$	30	71.90	38.95	5.40	14.15	57.05
	100	100	86.90	4.25	21.05	99.80
	200	100	98.90	5.30	24.40	100
$\chi_5^2 - 5$	30	37.15	18.05	4.45	8.65	29.15
	100	96.90	54.25	4.60	11.80	87.70
	200	100	79.40	4.40	13.15	99.90
$\chi_{10}^2 - 10$	30	22.40	12.90	4.75	6.90	18.75
	100	79.70	31.50	4.95	8.60	61.00
	200	98.90	50.00	4.60	8.80	94.70

Table 3: Achieved powers of each of the 5%-asymptotic level tests based on 2000 replications when the error distribution is a shifted chi-square. χ_k^2 represents a chi-squared distribution with k degrees-of-freedom.

Value of γ	Sample Size (n)	Component Statistics				Global Statistic
		$\hat{S}_1^2$	$\hat{S}_2^2$	$\hat{S}_3^2$	$\hat{S}_4^2$	$\hat{G}_4^2$
.5	30	8.50	10.85	5.65	6.10	14.15
	100	12.00	37.40	4.70	5.55	31.55
	200	10.65	48.85	4.80	8.35	39.50
1	30	11.10	16.65	3.40	6.20	16.15
	100	21.35	78.15	5.15	21.05	72.30
	200	24.10	96.95	5.40	13.35	93.30
2	30	21.40	52.35	6.15	12.60	46.75
	100	34.10	98.95	7.00	10.05	96.45
	200	47.50	100	7.55	31.60	100

Table 4: Achieved powers of each of the 5%-asymptotic level tests based on 2000 replications when the true model is $Y_i = x_i + x_i^\gamma \epsilon_i$ with ϵ_i's i.i.d. $N(0,1)$ and the model $Y_i = \beta_0 + \beta_1 x_i + \sigma\epsilon_i$ is fitted.

Values of (σ_1, σ_2)	Sample Size (n)	Component Statistics				Global Statistic
		$\hat{S}_1^2$	$\hat{S}_2^2$	$\hat{S}_3^2$	$\hat{S}_4^2$	$\hat{G}_4^2$
$(1, .5)$	30	20.05	29.75	7.60	16.60	35.30
	100	39.45	93.45	6.05	71.50	94.90
	200	43.50	99.90	6.00	99.90	100
$(1, 2)$	30	24.80	43.20	10.65	15.40	45.35
	100	35.20	89.85	4.35	89.40	95.55
	200	43.20	99.95	5.90	97.20	99.95

Table 5: Achieved powers of each of the 5%-asymptotic level tests based on 2000 replications when the true model is $Y_i = x_i + \sigma_i \epsilon_i$ with ϵ_i's i.i.d. $N(0, 1)$ and $\sigma_i = \sigma_1$ for $i \leq n/2$ and $\sigma_i = \sigma_2$ when $i > n/2$, and the model $Y_i = \beta_0 + \beta_1 x_i + \sigma \epsilon_i$ is fitted.

Value of (β_2, γ)	Sample Size (n)	Component Statistics $\hat{S}_1^2$	$\hat{S}_2^2$	$\hat{S}_3^2$	$\hat{S}_4^2$	Global Statistic $\hat{G}_4^2$
$(1, .5)$	30	5.45	4.10	4.20	5.00	4.80
	100	4.60	5.15	6.65	5.85	6.20
	200	4.90	5.80	7.60	4.70	8.25
$(1, 2)$	30	4.60	4.15	6.00	3.90	5.25
	100	4.45	5.30	9.35	4.85	6.30
	200	5.15	4.75	10.50	4.15	6.90
$(3, .5)$	30	5.45	4.15	8.25	4.85	5.45
	100	5.90	4.00	17.45	4.90	8.70
	200	4.00	4.60	32.95	5.00	16.55
$(3, 2)$	30	4.95	3.55	12.70	5.05	5.55
	100	4.95	4.95	43.65	4.55	22.80
	200	4.25	5.50	83.35	5.45	59.90
$(5, .5)$	30	3.70	3.70	14.45	4.20	5.70
	100	5.10	4.25	51.60	4.65	27.05
	200	5.20	5.00	84.25	5.40	62.00
$(5, 2)$	30	5.45	4.10	36.55	4.60	12.70
	100	5.20	4.80	92.00	5.45	72.60
	200	5.00	4.65	99.60	5.15	97.90

Table 6: Achieved powers of each of the 5%-asymptotic level tests based on 2000 replications when the true model is $Y_i = x_i + \beta_2 x_i^\gamma + \epsilon_i$ with ϵ_i's i.i.d. $N(0, 1)$ and the model $Y_i = \beta_0 + \beta_1 x_i + \sigma\epsilon_i$ is fitted.

Error Structure	Sample Size (n)	Component Statistics				Global Statistic
		$\hat{S}_1^2$	$\hat{S}_2^2$	$\hat{S}_3^2$	$\hat{S}_4^2$	$\hat{G}_4^2$
Martingale	30	20.85	10.25	3.80	35.45	27.85
$\epsilon_i = \frac{1}{\sqrt{i}} \sum_{j=1}^{i} \epsilon_j^*$	100	50.70	33.50	3.50	70.25	72.20
ϵ_i^* $i.i.d.$ $N(0,1)$	200	63.90	50.35	5.00	79.40	87.25
Markov type	30	5.20	3.05	3.50	8.20	5.30
$\epsilon_i = \frac{1}{\sqrt{2}}(\epsilon_{i-1} + \epsilon_i^*)$	100	8.90	4.70	6.25	15.55	12.15
ϵ_i^* $i.i.d.$ $N(0,1)$	200	11.40	5.85	5.15	18.35	15.70
Markov type	30	5.45	2.90	1.85	13.30	6.85
$\epsilon_i = \frac{1}{\sqrt{10}}(3\epsilon_{i-1} + \epsilon_i^*)$	100	19.60	10.60	5.45	36.45	34.85
ϵ_i^* $i.i.d.$ $N(0,1)$	200	29.50	21.40	3.60	47.45	54.45

Table 7: Achieved powers of each of the 5%-asymptotic level tests based on 2000 replications in the presence of dependent errors.

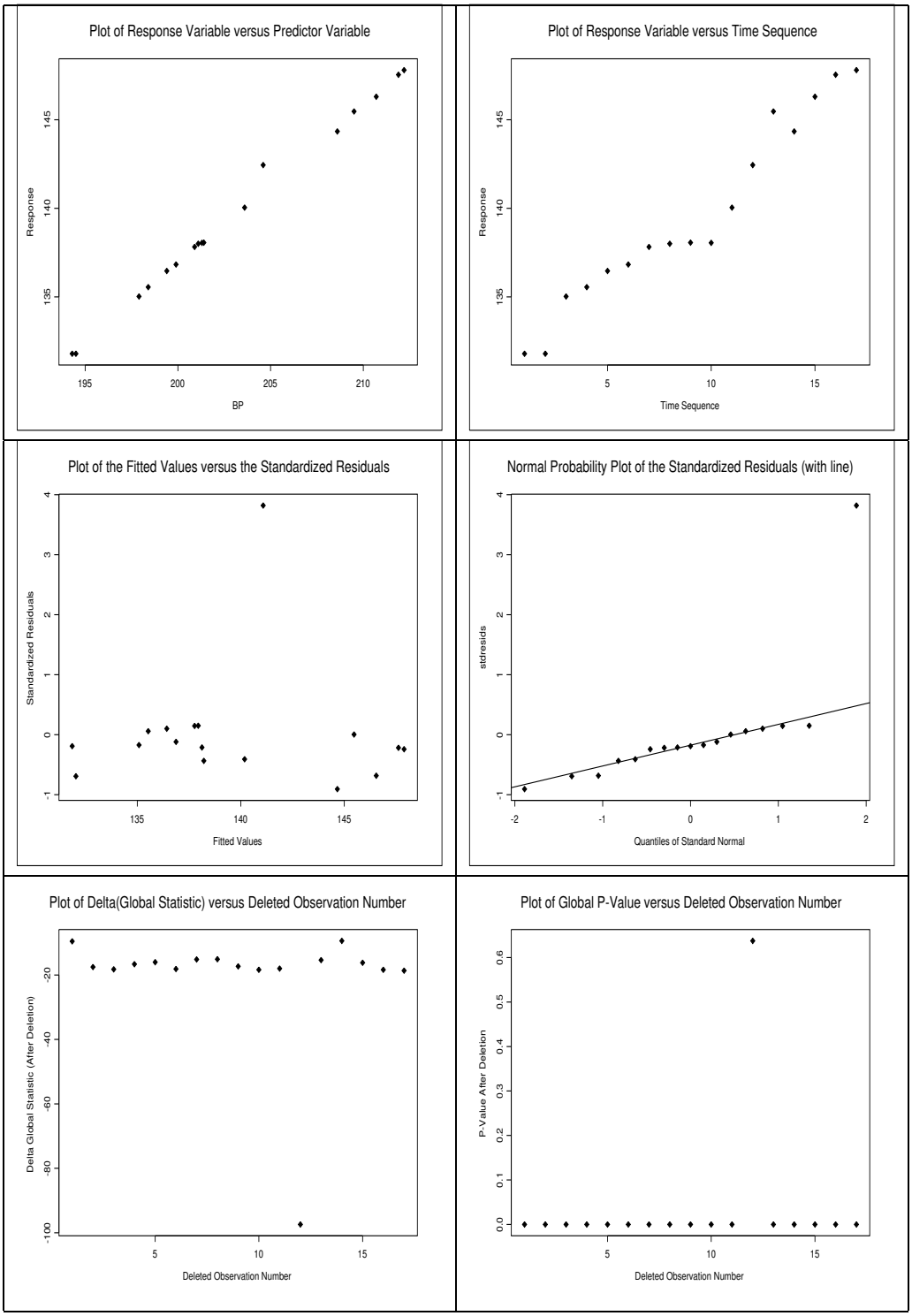

Figure 1: Plots pertaining to the analysis of Forbes' boiling point and pressure data.

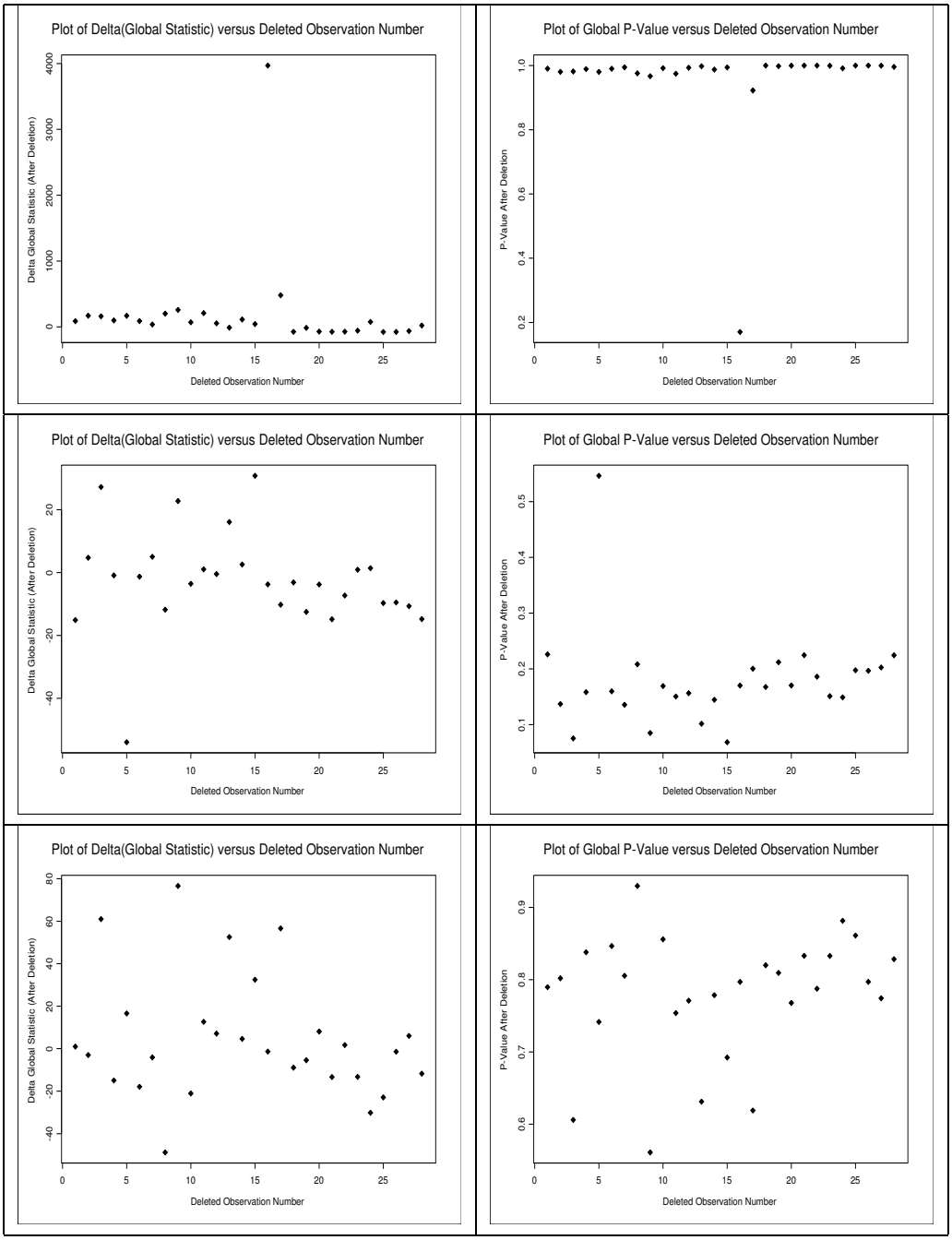

Figure 2: Plots pertaining to the analyses of the water salinity data. The first two panels are plots arising from the analysis using the original data set. The next two panels are from the analysis using the corrected data set. The last two panels are those from the corrected data set, but with the fitted model including a quadratic term of WaterFlow.

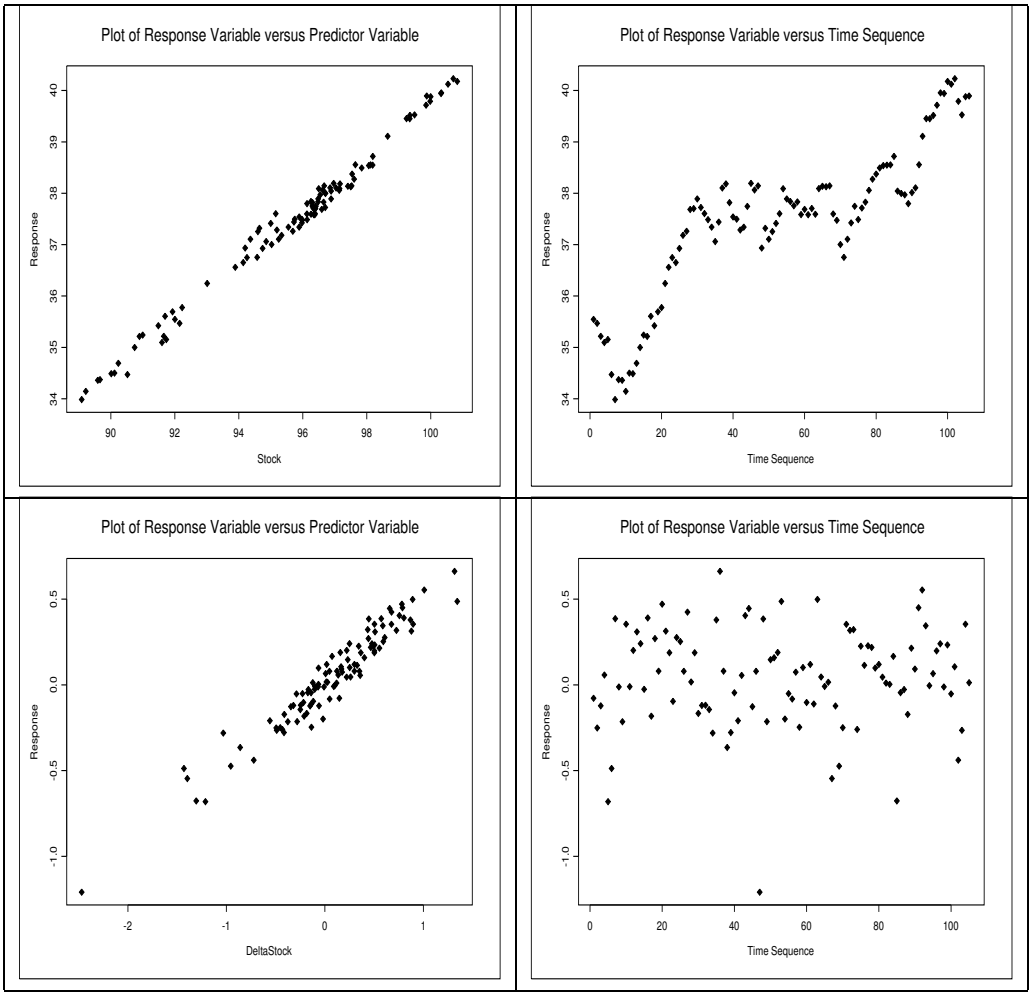

Figure 3: Plots pertaining to TIAA-CREF's Growth and Stock Accounts end-of-trading-day values for five months starting January 2, 1996. The first panel is a plot of the Stock and Growth share values, while the second panel is a time plot of the Growth share values. The third panel plots ΔGrowth versus ΔStock, while the last panel plots ΔGrowth versus Time.

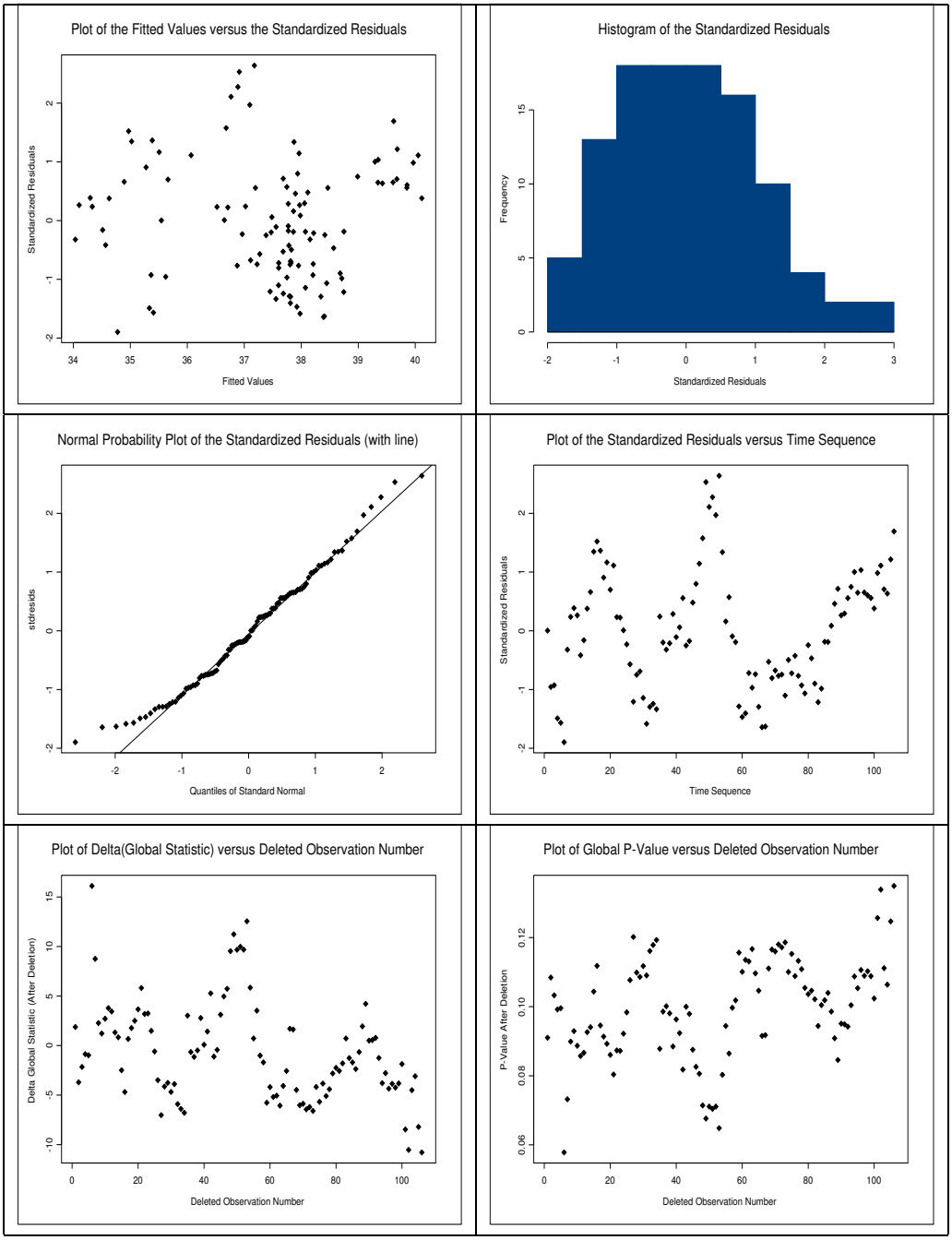

Figure 4: Plots pertaining to the analysis of TIAA-CREF's Growth and Stock Accounts upon fitting a simple linear regression model with Growth as response variable and Stock as predictor variable.

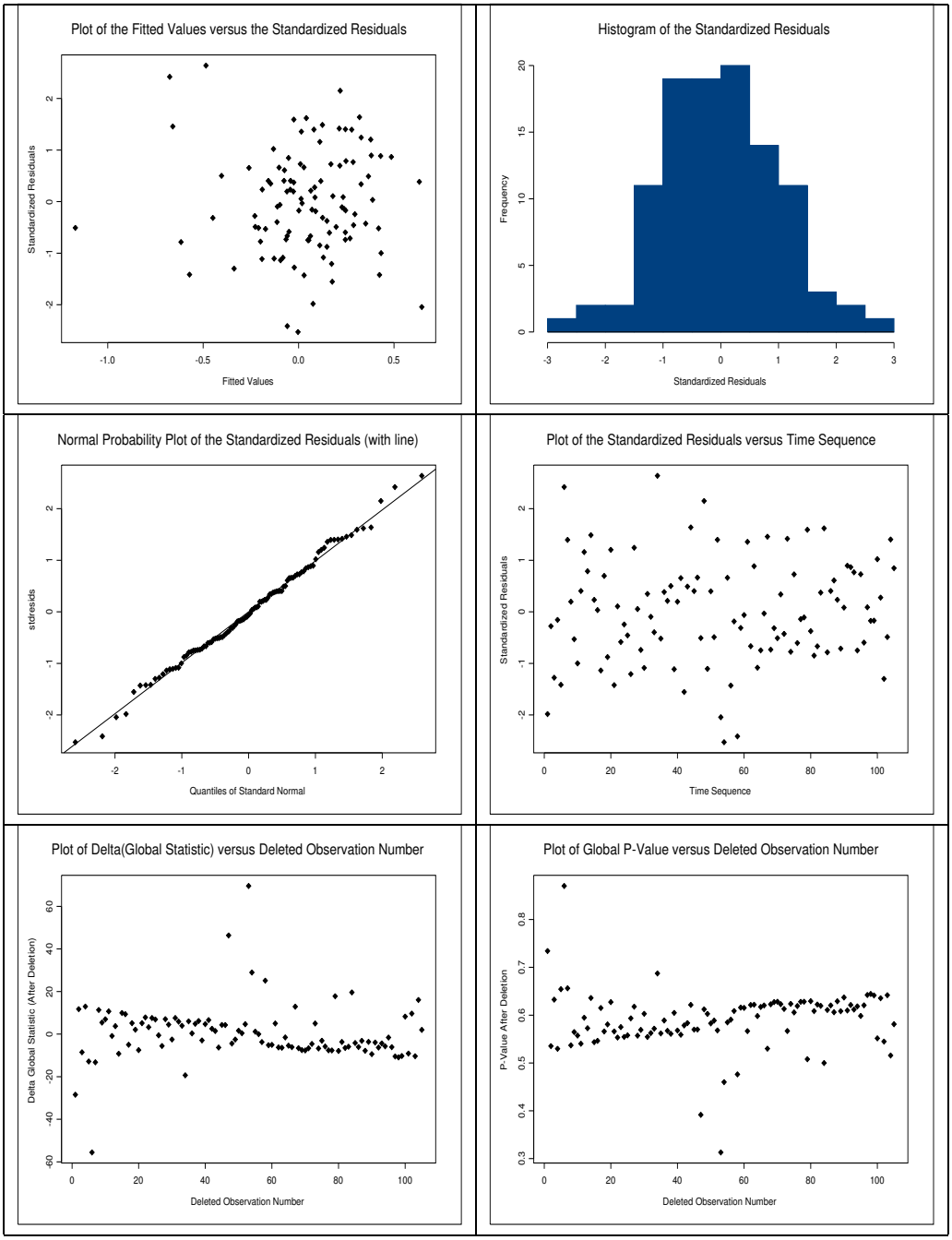

Figure 5: Plots pertaining to the analysis of TIAA-CREF's Growth and Stock Accounts upon fitting a simple regression model with ΔGrowth as dependent variable and ΔStock as predictor variable.