Case Studies of Normal Diagnostics in Regression Using Recovered Errors

D. R. Jensen

Department of Statistics
Virginia Polytechnic Institute
Blacksburg, VA 24061

D. E. Ramirez

Department of Mathematics
University of Virginia
Charlottesville, VA 22904-4137

Abstract

Diagnostics for normal errors in regression currently utilize ordinary residuals, despite the failure of assumptions validating their use. Case studies here show that such misuse may be critical even in samples of size exceeding currently accepted guidelines. A remedy is to employ recovered errors having the required properties.

1 Introduction

We first review the role of regression diagnostics. A case study then illuminates the central issues, followed by a brief discussion of the notation and models to be considered.

1.1 Case Studies: A First Look

We consider a study on actual market returns for stocks as related to corresponding accounting rates. For each of $n = 54$ companies the mean yearly market return (Y) and the mean yearly accounting rate (X) were determined for the period 1959-1974. Simple linear regression analysis then gave the best-fitting line as reported in Myers (1990, p.16 ff.), along with the ordinary residuals. The normal probability plot for the ordinary residuals appears as Figure 1.

In testing for normality with $n = 54$, the Anderson-Darling test gave $A^2 = 1.876$ and a p-value of 0.000 to three decimals, whereas the Kolmogorov-Smirnov test gave $D = 0.192$ with approximate p-value < 0.01. For details and further references regarding these tests, see D'Agostino (1982) and Mudholkar *et al.* (1995). Note that $n = 54$ exceeds the empirical guidelines of Pierce and Gray (1982) of $n = 20$ by a factor of 2.7, so evidence against normality of the errors is unequivocal using currently accepted methods.

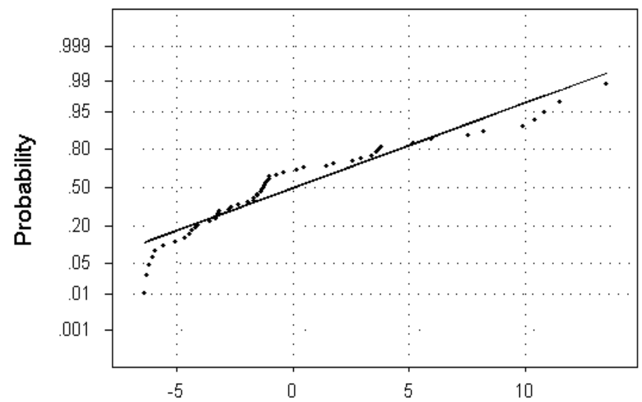

Figure 1: Probability Plot of Residuals for Accounting Data

Further insight regarding the actual error distribution is often garnered from the histogram of the ordinary residuals as depicted in Figure 2.

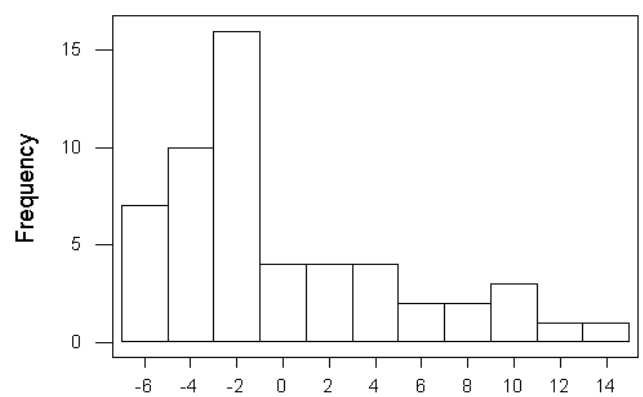

Figure 2: Histogram of Residuals for Accounting Data

To the extent that these residuals might truly reflect the unobserved errors, the evidence suggests a shifted gamma distribution as a viable candidate for the underlying errors. This, in turn, would place in serious jeopardy any reasonable prospects for using strict normal-theory inferences in the analysis.

Nonetheless, these conclusions do appear somewhat surprising in view of the fact that each Y itself is the average of 16 yearly market returns. Central limit tendencies often are observed in practice in samples of moderate size.

A close inspection of the accounting data shows that the ordinary residuals tend to be negatively correlated. Descriptive statistics, including their minimum, first, second, and third quartiles, and their maximum correlations are $\{-0.22767, -0.02651, -0.01957, -0.01160, 0.10340\}$, respectively, with a range of 0.33107, and their histogram is depicted in Figure 3. Correlations among the ordinary residuals are not negligible in this study and thus may be a problem. The corresponding array of descriptive statistics for variances of the ordinary residuals is given by $\{0.67670,\ 0.96264,\ 0.97268, 0.97943,\ 0.98138\}$, with a range of 0.30468. Heterogeneous variances thus may be an issue as well, as these matters appear not to have been studied in depth. We return to these topics subsequently, including difficulties surrounding the singularity of the joint distribution of the ordinary residuals.

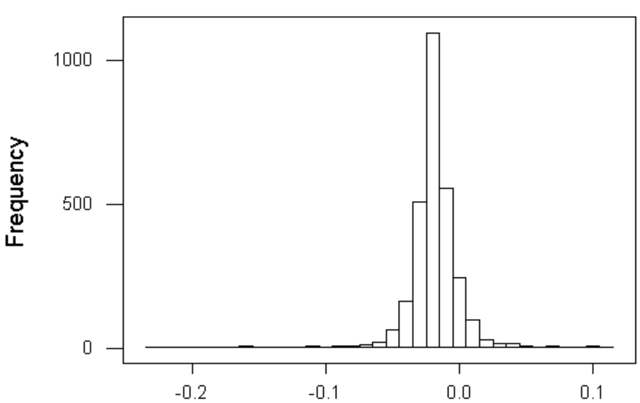

FIGURE 3: CORRELATIONS AMONG RESIDUALS FOR ACCOUNTING DATA

The aforementioned difficulties are not easily dismissed. To do so is to disclaim the importance of normality itself. In this paper, we preempt all those difficulties on reconstructing entities, called *recovered errors*, having the requisite properties. Moreover, these are recovered from the ordinary residuals using standard software that may be assimilated readily into existing regression packages. We turn next to matters of notation and details surrounding the models to be considered.

1.2 Notation

Designate by $\mathcal{L}(\mathcal{Y})$ the law of distribution of Y. In particular, $N_1(\mu, \sigma^2)$ is the one-dimensional normal distribution having the mean $E(Y) = \mu$ and variance $Var(y) = \sigma^2$. The covariance (Y_1, Y_2) is designated as $Cov(Y_1, Y_2)$.

In what follows we consider full-rank multilinear models of the type

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i \qquad (1)$$

relating the typical response Y_i to the regressors $\{X_{i1}, X_{i2}, \ldots, X_{ik}\}$ through unknown parameters $\{\beta_0, \beta_1, \ldots, \beta_k\}$, for $i = 1, 2, \ldots, n$. The ordinary residuals are given by $e_i = Y_i - \widehat{Y}_i = Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 X_{i1} + \cdots + \widehat{\beta}_k X_{ik})$ in terms of the ordinary least-squares estimators $\{\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k\}$. Conventional assumptions regarding such models include

$$\begin{aligned}
A1 &: \ E(\varepsilon_i) = 0, 1 \le i \le n;\\
A2 &: \ Var(\varepsilon_i) = \sigma^2, 1 \le i \le n;\\
A3 &: \ Cov(\varepsilon_i, \varepsilon_j) = 0, i \ne j = 1, \ldots, n;\\
A4 &: \ \mathcal{L}(\varepsilon_\rangle) = \mathcal{N}_\infty(\prime, \sigma^\in), \text{ independently}\\
&\quad \text{for } i = 1, 2, \ldots, n.
\end{aligned}$$

1.3 Basics

Consider the ordinary residuals $\{e_1, \ldots, e_n\}$, their expected values $E(e_i)$, their variances $Var(e_i)$, and their covariances $Cov(e_i, e_j)$, for $i, j = 1, 2, \ldots, n$. Under assumptions $A1 - A3$ it follows directly that $E(e_i) = 0$, whereas their variances and covariances are determined by the matrix of regressor variables. In consequence, the elements $\{e_1, \ldots, e_n\}$ are typically heteroscedastic and are correlated, as documented empirically in the case study of Section 1.1. Moreover, their joint n-dimensional distribution is singular of rank $n-k-1$, there being $k+1$ sure linear relations among them. This joint singularity is itself troubling for reasons to follow.

Using all n ordinary residuals diagnostically ignores redundancies among them, and it supports the illusion that there are n, instead of the actual $n-k-1$, effective data points. Write model (1) in block form as

$$\begin{bmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_\mathbf{1} \\ \varepsilon_\mathbf{2} \end{bmatrix}, \qquad (2)$$

with $\mathbf{X_2}$ invertible and, as usual, with $\mathbf{X}$ having as its first column the unit vector. Since $\mathbf{X}'\mathbf{e} = \mathbf{0}$, it follows with $\mathbf{e} = [\mathbf{e}_1', \mathbf{e}_2']'$ that $\mathbf{e_2} = -(\mathbf{X_1}\mathbf{X_2^{-1}})'\mathbf{e_1}$. Thus $\mathbf{e_1}$ and $\mathbf{e}$ are equally informative, and $\mathbf{e_2}$ contains no further information. In consequence, the usual graphical displays, order statistics, sample moments, the empirical distribution function (EDF), and various tests for goodness of fit are based on excessive and partially redundant data, whereas p-values depending on sample size are reported erroneously.

1.4 The Recovered Errors

Using properties of linear statistics and using $t = n-k-1$, we recover elements $\{R_1, \ldots, R_t\}$ from the ordinary residuals $\{e_1, \ldots, e_n\}$, as linear functions whose coefficients depend on the matrix projecting observations onto the error space. An algorithm is based on the spectral decomposition $\mathbf{B_n} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \xi_\mathbf{i} \mathbf{q_i q_i'} = \mathbf{QD}_\xi \mathbf{Q'}$ of $\mathbf{B_n} = \mathbf{I_n} - \mathbf{X(X'X)^{-1}X'}$, such that $\mathbf{D}_\xi = \mathbf{Diag}(\xi_\mathbf{1}, \ldots, \xi_\mathbf{n})$ contains the ordered eigenvalues and $\mathbf{Q} = [\mathbf{q_1}, \ldots, \mathbf{q_n}]$ the corresponding eigenvectors of $\mathbf{B_n}$. To these ends, let $\mathbf{R_{(n)}} = \mathbf{Q'e}$; partition $\mathbf{R_{(n)}}$ as $\mathbf{R_{(n)}} = [\mathbf{R'_{(t)}}, \mathbf{R'_{(k+1)}}]'$ with $\mathbf{R_{(t)}} \in \mathbb{R}^\mathbf{t}$ and $\mathbf{R_{(k+1)}} \in \mathbb{R}^\mathbf{k+1}$; and identify $\mathbf{R'_{(t)}} = [\mathbf{R_1}, \ldots, \mathbf{R_t}]$. Their properties are summarized in Proposition 1 (see Jensen and Ramirez (1999). We henceforth refer to $\{R_1, \ldots, R_t\}$ as the *linearly recovered errors.*

Proposition 1 *Let* $\mathbf{R_{(n)}} = [\mathbf{R'_{(t)}}, \mathbf{R'_{(k+1)}}]' = \mathbf{Q'e}$ *with* $\mathbf{Q}$ *from the spectral decomposition* $\mathbf{B_n} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \xi_\mathbf{i} \mathbf{q_i q_i'} = \mathbf{QD}_\xi \mathbf{Q'}$ *of* $\mathbf{B_n} = \mathbf{I_n} - \mathbf{X(X'X)^{-1}X'}$. *Then under assumptions* $A1-A3$ *with* $t = n-k-1$, *the linearly recovered errors* $\mathbf{R'_{(t)}} = [\mathbf{R_1}, \ldots, \mathbf{R_t}]$ *are homoscedastic, nonsingular, and uncorrelated, whereas under assumptions* $A1-A4$, *the variables* $\{R_1, \ldots, R_t\}$ *now comprise a simple random sample of size* $t = n-k-1$ *from* $N_1(0, \sigma^2)$, *independently of* $\mathbf{X}$.

1.5 Non-Uniqueness of Linearly Recovered Errors

Since the eigenvalues of the matrix $\mathbf{B_n}$ are not distinct, with unity appearing $t = n-k-1$ times, the matrix $\mathbf{Q}$ is not unique. We introduce in this section our methodology for choosing $\mathbf{Q}$. Write $\mathbf{Q'} = [\mathbf{Q_1}, \mathbf{Q_2}]'$ with $\mathbf{Q_1}$ of rank t and $\mathbf{Q_2}$ of rank $n-t$. For any orthogonal matrix $\mathbf{P}$ of order $t \times t$, the matrix $\mathbf{P'Q_1'}$ will also transform the singular errors $\mathbf{e}$ into a nonsingular vector of *recovered errors.*

We consider two examples, as shown in Table 1, which will demonstrate our methodology as well as exemplify some problems with the usual regression diagnostics. The two examples both have the same $\mathbf{x}$ values, but different $\mathbf{y}$ values.

Table 1: Case Study Examples

$\mathbf{x}$	-1	-1	-1	0	0	0	1	1	1
$\mathbf{y_1}$	.4	.4	.4	.1	.1	.1	-.2	-.2	.8
$\mathbf{y_2}$	0	0	0	0	0	0	0	0	1

For strict validity, graphics and hypothesis tests for normality both presuppose simple random sampling as noted, whereas the ordinary residuals are correlated and their joint distribution is singular, as these two examples will show.

Example 1 is concerned with $(\mathbf{x}, \mathbf{y_1})$. The Kolmogorov-Smirnov Normality Test on the residuals of Example 1 has p-value > 0.15, and so the normality assumption for this model, at this point, would be deemed to be valid using standard procedures.

Example 2 is concerned with $(\mathbf{x}, \mathbf{y_2})$. Note that the value 1 for y at $(x, y) = (1, 1)$ could be any value, say 1,000,000, by scaling. Although it is not immediately apparent, Examples 1 and 2 have identical residuals. They thus have identical studentized residuals and identical p-values for the Kolmogorov-Smirnov Normality Test which, for Example 2, also has p-value > 0.15. The normality assumption for this model also would be taken to be valid. However, one would hope that Example 2 would fail the normality test.

For a linear regression model, the *studentized residuals* r_i are given by normalizing the residuals $e_i = y_i - \widehat{y}_i$ $(1 \leq i \leq n)$ by dividing e_i by the corresponding standard deviation, replacing σ by s, as $r_i^2 = e_i^2/(s\sqrt{1-h_{ii}})$ where $h_{ii} = \mathbf{x_i'(X'X)^{-1}x_i}$ denotes the ordinary leverages. For example, Myers (1990, p. 217) suggests using the studentized residuals in residuals diagnostics. Unfortunately, these diagnostic tools are routinely misapplied using ordinary residuals, which are correlated and the joint distribution is singular. Example 2 has been constructed to have the maximum value for $r_9^2 = n-k-1 = 7$. This is an extension of Thompson's inequality $(Z_i^2 \leq n-1)$ which gives a bound on how deviant an observation can be; see, for example, Olkin (1992). The extension to linear models has been noted by Gray and Woodall (1994). The externally studentized residual t_i use for the estimate of σ the estimate s_i which is computed with the i^{th} value deleted. This diagnostic is known to be distributed as $t(n-k-2)$. Using the relationship between s and s_i (for example, Myers (1990, p. 408)), we can write

$$r_i^2 = \frac{(n-k-1)t_i^2/(n-k-2)}{(1+t_i^2/(n-k-2))} \leq n-k-1.$$

Examples 1 and 2 have been constructed to satisfy $\mathbf{Q_1'y_1} = \mathbf{Q_1'y_2}$, and so they have identical linearly recovered errors. To examine the nonuniqueness of $\mathbf{Q}$, let

$$\mathbf{P}'(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) & 0 & 0 & 0 & 0 \\ 0 & -\sin(\theta) & \cos(\theta) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then $\mathbf{P}'(\theta)\mathbf{Q_1'}$ also can be used to define *recovered errors.* Figure 4 presents the graph of the seven resulting standardized quantities as a function of θ on $-\pi \le \theta \le \pi$. For a fixed $\theta \in [-\pi, \pi]$, we standardized the seven components of $\mathbf{P}'(\theta)\mathbf{Q}'\mathbf{Y_2}$ by subtracting their mean value and dividing by their standard deviation with $n - 1 = 6$.

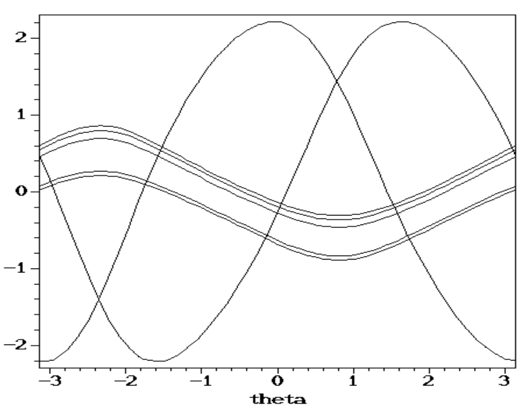

FIGURE 4: STANDARDIZED RECOVERED ERRORS USING $P'(\theta)Q_1'Y_2$

All recovered errors are in the class of Linear Unbiased Scaled (LUS) residual estimators; see Theil (1971). These estimators are not unique. Given a choice for $\mathbf{Q_1'}$, any orthogonal rotation $\mathbf{P}'\mathbf{Q_1'}$ also will produce a LUS estimator for the disturbances. Theil suggests using the matrix $\mathbf{P}'\mathbf{Q_1'}$ that minimizes the expected squared length of the transformed residuals, to be called the $BLUS$ residuals. Other criteria are required in testing for normality.

Table 2 gives the standardized linearly recovered errors for two different values of the angle θ. The first, $\theta = -2.3562$ (where two of the curves cross), has dampened the magnitudes of the standardized recovered errors with the maximum value being $R_2 = -1.4180$. The second, $\theta = 1.7530$, has been chosen to amplify the magnitudes of the standardized errors with the maximum value being $R_2 = 2.2017$. We will explain the choice for these values of θ below.

Recall that both models from Table 1 have p-value > 0.15 using the ordinary residuals together with the Kolmogorov-Smirnov Normality Test. This same test using the $n-2 = 7$ recovered residuals at $\theta = 2.3562$ has p-value 0.058, again validating normality for this representation of the recovered errors. However, when applied to the errors recovered at $\theta = 1.7530$, the Kolmogorov-Smirnov Normality Test now contraindicates normality, with p-value < 0.01. At this representation for the recovered errors, the original models in Table 1 do not satisfy the normality requirements. This example makes clear that opposite conclusions may be drawn from a given data set, depending on the representation used for recovered errors.

Table 2: Two Standardized Linearly Recovered Errors

θ	R_1	R_2	R_3
-2.3562	0.6943	-1.4180	-1.4180
1.7530	-0.2062	2.2017	-0.6498

R_4	R_5	R_6	R_7
0.8616	0.7975	0.2129	0.2697
-0.0691	-0.1216	-0.6008	-0.5542

To measure the spread of the recovered errors, the sum of the fourth powers of their deviations from the mean of the seven components of $\mathbf{P}'(\theta)\mathbf{Q}'\mathbf{Y_2}$ is considered, as plotted in Figure 5 with $-\pi \le \theta \le \pi$. For reasons to be given, we call this function of θ on $-\pi \le \theta \le \pi$ the *Kurtosis Estimate.* The graph shows a minimum at $\theta = -2.3562$, and a maximum at $\theta = -0.1822$ and $\theta = 1.7530$. Both of these maximum values for θ produce identical values for the recovered errors.

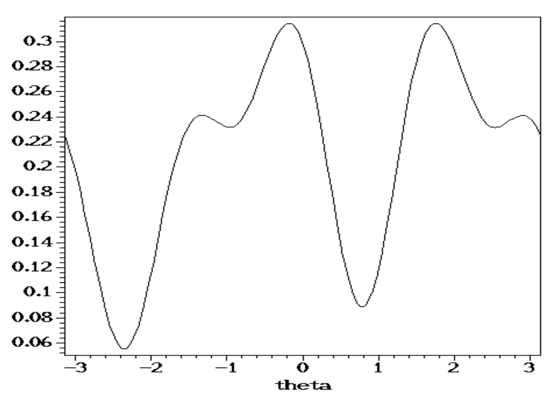

FIGURE 5: KURTOSIS ESTIMATE FOR THE RECOVERED ERRORS

In this section we have demonstrated the nonuniqueness for the linearly recovered errors, and we have shown that the standardized recovered errors can vary greatly. The methodology we have introduced is that of finding the recovered errors which have a large kurtosis estimate for reasons to be given. We show how to achieve this in the next section.

2 Kurtosis Estimates

We assume that the linearly recovered errors $\{R_1, \ldots, R_t\}$ have been determined from the matrix $\mathbf{Q}'$ as in Section 1.3. For any orthogonal matrix $\mathbf{P}$ of order $t \times t$, the matrix $\mathbf{P}'\mathbf{Q}_1' = \mathbf{A}'$ will also transform the ordinary residuals $\mathbf{e}$ into a nonsingular vector of recovered errors. The kurtosis γ_2 for the *BLUS* residuals have been given by Huang and Bolch (1974); see also Misra (1972). The result, which extends to include the *LUS* residuals, is (where we correct the typographical error)

$$\gamma_2(R_i) = 3 + (\gamma_2(\varepsilon_i) - 3)\sum_{j=1}^{n} a_{ji}^4, \text{ for } 1 \leq i \leq t.$$

The criterion that we use is to consider $\Psi(\mathbf{P}) = \sum_{\mathbf{i=1}}^{\mathbf{t}} \gamma_2(\mathbf{R_i})$ as $\mathbf{P}$ is varied, eventually choosing P so as to maximize $\Phi(\mathbf{P}) = \sum_{\mathbf{i=1}}^{\mathbf{t}} \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{a_{ji}^4}$. This maximization is available with standard software since it is equivalent to the Kaiser "raw" varimax criterion in factor analysis; see, for example, Harman (1976, p. 290). The varimax criterion, in turn, is used to find the orthogonal matrix P which will maximize

$$\frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{n} a_{ji}^4 - \frac{1}{n^2}\sum_{i=1}^{t}\left(\sum_{j=1}^{n} a_{ji}^2\right)^2,$$

where $\mathbf{A} = \mathbf{Q_1 P}$. We now note that the second term above is the constant t/n^2 since $A'A = I_t$.

2.1 The Methodology

For a given data set, first compute the ordinary leverages $h_{ii} = \mathbf{x_i}(\mathbf{X'X})^{-1}\mathbf{x_i'}$, for $1 \leq i \leq n$. We sort the data in ascending order based on the leverages. The order of the data does affect the linearly recovered errors. We find the linearly recovered errors $\mathbf{R_H} = \mathbf{Q_{H1}'Y}$ from the eigenvectors $\mathbf{Q_H}$ of $\mathbf{H_n} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$, and the linearly recovered errors $\mathbf{R_B} = \mathbf{Q_{B1}'Y}$ from the eigenvectors $\mathbf{Q_B}$ of $\mathbf{B_n} = \mathbf{I_n} - \mathbf{H_n}$. The first method transforms into zero the residuals corresponding to the $k+1$ values with low leverages, and the second transforms into zero the residuals corresponding to the $k+1$ values with high leverages. We now use a varimax rotation to find the matrices $\mathbf{P_H}$ and $\mathbf{P_B}$, that have maximized $\Phi(\cdot)$ for $\mathbf{Q_{H1}}$ and $\mathbf{Q_{B1}}$, respectively. With $\mathbf{A_H'} = \mathbf{P_H'Q_{H1}'}$ and $\mathbf{A_B'} = \mathbf{P_B'Q_{B1}'}$, we determine two sets of recovered errors $\mathbf{R_{A_H'}}$ and $\mathbf{R_{A_B'}}$, respectively, each having maximal kurtosis, which we now can test for normality without violating the important assumptions required. Note that the rotation matrices are independent of the responses Y. The Monte Carlo simulation study in Jensen and Ramirez (1999) showed that our recovered errors procedure had uniformly more power than the $BLUS's$ procedure over all cases studied. We also showed that there the order of the data caused statistical differences in the p-values for the *BLUS* residuals. No such statistical differences were seen with our recovered error residuals $\mathbf{A_H'}$ and $\mathbf{A_B'}$.

2.2 Case Studies: Revisited

We next examine a multilinear model in light of the foregoing developments. The results are rather striking. We consider a study of discoloration in canned applesauce during storage. Draper (1965) examined effects of temperature (X_1), betaine added (X_2), and storage time (X_3) on the Munsell chroma (Y) of each specimen studied. A multilinear model was determined using $n = 48$ data points. The normal probability plot of the ordinary residuals appears as Figure 6; the Anderson-Darling test, using these residuals, gives $A^2 = 0.300$ and a p-value $= 0.567$; and the Kolmogorov-Smirnov test gives $D = 0.066$ and p-value > 0.15. This example falls within the guidelines of Pierce and Gray (1982) regarding the sample size, so that nonnormality of errors is not an issue using currently accepted methods. However, the applicability of the methods themselves must be questioned for reasons given earlier.

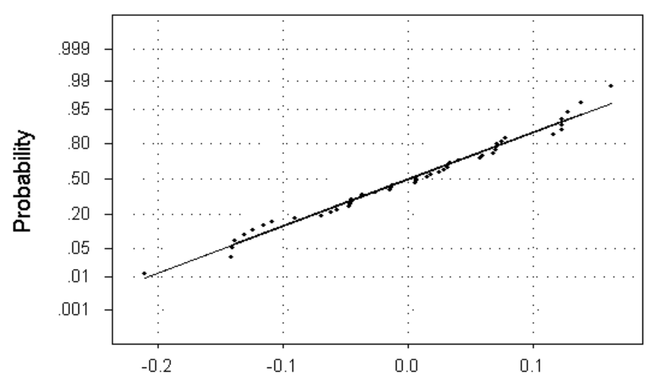

FIGURE 6: PROBABILITY PLOT OF RESIDUALS FOR APPLESAUCE DATA

Proceeding as before, we again find that the ordinary residuals tend to be negatively correlated. Accordingly, we rework the foregoing tests using $n-k-1 = 44$ linearly recovered errors instead, as prescribed in Proposition 1, to ensure their validity. The p-values from the Shapiro-Wilk test for normality using the ordinary residuals $\mathbf{e}$, and the residuals from the two rotated transformed models, $\mathbf{A}_\mathbf{H}'\mathbf{Y}$ and $\mathbf{A}_\mathbf{B}'\mathbf{Y}$, which maximize the kurtosis of the recovered errors, are 0.4568, 0.1219, and 0.0262 respectively. As expected, the rotated models are better able to reveal the nonnormality of the linearly recovered errors, with the residuals $\mathbf{R}_{\mathbf{A}_\mathbf{B}'} = \mathbf{A}_\mathbf{B}'\mathbf{Y}$ having a p-value of 0.0262.

The revised tests clearly point towards nonnormal errors, contradicting our earlier assessment based on the ordinary residuals. The normal probability plot of the recovered errors $\mathbf{R}_{\mathbf{A}_\mathbf{B}'} = \mathbf{A}_\mathbf{B}'\mathbf{Y}$ appears in Figure 7. In summary, evidence for or against normality is always suspect when based on the ordinary residuals in regression.

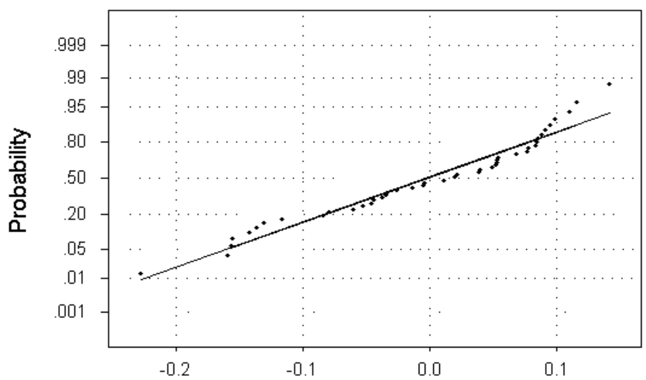

FIGURE 7: RECOVERED ERRORS FOR THE APPLESAUCE DATA

3 Other Case Studies

The accounting data set from Section 1 failed the tests for normality using the ordinary residuals. As we have demonstrated in this paper, the user should base conclusions of normality on the recovered errors. The Shapiro-Wilk test has p-value = 0.0002 for both sets of recovered errors $\mathbf{R}_{\mathbf{A}_\mathbf{H}'} = \mathbf{A}_\mathbf{H}'\mathbf{Y}$ and $\mathbf{R}_{\mathbf{A}_\mathbf{B}'} = \mathbf{A}_\mathbf{B}'\mathbf{Y}$. Thus we have shown the nonnormality of this data set without violating the assumptions of the Shapiro-Wilk test for normality.

4 Conclusions

Diagnostics for normal errors in regression are subject to misuse when applied to singular, heteroscedastic, and correlated residuals. Case studies show that such misuse may undermine substantially the intent of these diagnostics and should be discontinued. Fortunately, these difficulties may be surmounted on using recovered errors amenable to standard regression software packages.

References

Agostino, R. B. (1982). Departures from normality, tests for. *Encyclopedia of Statistical Science, Volume 2,* N. L. Johnson, S. Kotz, and C. B. Read, eds. 315-324, John Wiley, New York.

Draper, W. (1965). *Effect of Carbohydrate Degradation upon Browning and Corrosion in Canned Applesauce.* Ph.D. Dissertation Library, Virginia Polytechnic Institute, Blacksburg, VA 24061.

Gray, J. B. and Woodall, W. H. (1994), On the Maximum Size of Standardized and Internally Studentized Residuals in Regression Analysis, *The American Statistician*, **48**: 111-113.

Harman, H. (1976). *Modern Factor Analysis.* The University of Chicago Press, Chicago.

Huang, C. J. and Bolch, B. W. (1974). On the testing of regression disturbances for normality, *J. Amer. Statist. Assoc.* **69:** 330-335.

Jensen, D. R. and Ramirez, D. E. (1999). Recovered errors and normal diagnostics in regression, *Metrika* **49**: 107-119.

Misra, P. N. (1972). Relationship between Pearsonian coefficients of distributions of least squares estimators and the disturbance term, *J. Amer. Statist. Assoc.* ***67**: 662-663.*

Mudholkar, G. S., Srivastava, D. K., and Lin, C. T. (1995). Some p-variate adaptations of the Shapiro-Wilk test of normality. *Comm. Statist. A - Theory Methods* **24**: 953-985.

Myers, R. H. (1990). *Classical and Modern Regression with Applications, Second Edition,* PWS-Kent Publishing Company, Boston.

Olkin, I. (1992). A matrix formulation on how deviant an observation can be, *The American Statistician*, **46**: 205-209.

Pierce, D. A. and Gray, R. J. (1982). Testing normality of errors in regression models, *Biometrika* **69**: 233-236.

Theil, H. (1971). *Principles of Econometrics*, John Wiley and Sons, Inc., New York.