Testing Regression Models with Residuals as Data

by

Xia Hua

Bachelor of Science, University of Hong Kong, June 2006

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Author ....
Department of Mathematics
April 22, 2010

Certified by ....
Richard M. Dudley
Professor of Mathematics
Thesis Supervisor

Accepted by ....
Bjorn Poonen
Chairman, Departmental Committee on Graduate Students

Testing Regression Models with Residuals as Data

by

Xia Hua

Submitted to the Department of Mathematics
on April 22, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In polynomial regression $Y_i = \sum_{j=0}^{k} a_j X_i^j + \varepsilon_i,\ i = 1, ..., n$ where (X_i, Y_i) are observed and a_i need to be estimated, it is often assumed the errors ε_i are i.i.d. $N(0, \sigma^2)$ for some $\sigma \geq 0$. In this thesis, I developed a residual based test, the turning point test for residuals, which tests the hypothesis that the kth order polynomial regression holds with ε_j i.i.d. $N(0, \sigma^2)$ while the alternative can simply be the negation or be more specific, e.g., polynomial regression with order higher than k. This test extends the rather well known turning point test and requires approximation of residuals by errors for large n. The simple linear regression model, namely $k = 1$, will be studied in most detail. It is proved that the expected absolute difference of numbers of turning points in the errors and residuals cannot become large and under mild conditions becomes small at given rates for large n. The power of the test is then compared with another residual based test, the convexity point test, using simulations. The turning point test is shown to be more powerful against quadratic alternatives.

Thesis Supervisor: Richard M. Dudley
Title: Professor of Mathematics

Acknowledgments

The author is grateful to Professor Richard Dudley for supervising her thesis. She would also like to thank Professor Daniel Stroock and Professor Lie Wang for their helpful comments and suggestions.

Contents

List of Figures

List of Tables

Chapter 1

Turning Point Test for Residuals

1.1 Introduction

In a regression model, say $Y_i = f(X_i) + \varepsilon_i$, where (X_i, Y_i) are observed and f is an unknown regression function, possibly of a given parametric form, the errors ε_i may satisfy what we call the "weak" assumption that they are orthogonal with mean 0 and the same variance, and often the further "strong" assumption that they are i.i.d. $N(0, \sigma^2)$ for some $\sigma \geq 0$. In this thesis, we consider general including polynomial regression of degrees $2, 3, \ldots$, but in most detail the simple linear regression model, namely $f(x) = a + bx$ for unknown parameters a and b, under the strong assumption on the errors. When a and b are estimated via least squares (equivalent to maximum likelihood) by $\widehat{a}$ and $\widehat{b}$ respectively, we then get the *residuals* $\widehat{\varepsilon}_i := Y_i - a_i - \widehat{b}X_i$. We would like to test the hypothesis H_0 that the simple linear regression model holds with ε_i i.i.d. $N(0, \sigma^2)$. There is a rather well known turning point test for whether variables are i.i.d. with a continuous distribution. Here ε_i are unobserved but for large enough n, if the model holds, $\widehat{a}$ and $\widehat{b}$ will be close enough to the true a and b so that $\widehat{\varepsilon}_i$ will have approximately the properties of ε_i.

Specifically, if T_n is the number of turning points in the i.i.d. errors ε_i, then T_n has standard deviation of order $\sqrt{n}$. Under H_0, if $\widehat{T}_n$ is the number of turning points in the residuals $\widehat{\varepsilon}_i$, we show that for arbitrary design points, $E(|T_n - \widehat{T}_n|) < 1$ for all $n \geq 3$ (Theorem 4) and $E(|T_n - \widehat{T}_n|)$ is not necessarily $o_p(1)$ (Proposition

5). Moreover we show that if the design points are not too irregularly spaced then $E(|T_n - \widehat{T}_n|) = O(n^{-1/2})$ (Theorem 6) and is not $o(n^{-1/2})$ if the design points are equally spaced (Proposition 8). If the design points themselves are i.i.d. normal we show that $E(|T_n - \widehat{T}_n|) = O((\log n/n)^{1/2})$ (Theorem 7) while simulations indicate that no bound of smaller order holds (Section 1.5.2).

1.2 Turning points in i.i.d. data

The turning point test, in its basic form, is a test of the hypothesis H_0 that variables $V_1, ..., V_n$ are i.i.d. with a continuous distribution. So, it's a nonparametric test. For $j = 2, ..., n-1$, let $I_j = 1$ if $V_j < \min(V_{j-1}, V_{j+1})$ or $V_j > \max(V_{j-1}, V_{j+1})$, otherwise $I_j = 0$, also if $j = 1$ or n. Say a *turning point* occurs at j if and only if $I_j = 1$, in other words, the sequence turns from increasing to decreasing, or from decreasing to increasing, at j.

Random variables $Y_1, Y_2, ...,$ are said to be *m-dependent* if for each $k = 1, 2, ...,$ the set $Y_1, ..., Y_k$ of random variables is independent of the set of all Y_i for $i > k + m$. It follows from this that for any $j \neq i$, Y_j and Y_i are independent if $|i - j| > m$ but may be dependent if $|j - i| \leq m$, hence the name m-dependent. The I_j are 2-dependent since for any $k \geq 1$, $I_1, ..., I_k$ depend only on $V_1, \ldots, V_{k+1}$, and $\{I_{k+3}, I_{k+4}, ...\}$ depend only on V_j for $j \geq k+2$. Thus I_i and I_j are independent if $|j - i| \geq 3$. There are central limit theorems for m-dependent random variables which will apply to the I_j under H_0 since they are uniformly bounded and for $2 \leq j \leq n-1$ they are identically distributed. Berk (1973) gives an extended theorem for triangular arrays of random variables and gives the earlier references in the reference list. It will follow that $T_n := \sum_{j=2}^{n-1} I_j$, the number of turning points, has an asymptotically normal distribution as $n \to \infty$. After we find its mean and variance under H_0, we can can thus use T_n as a test statistic, rejecting H_0 if T_n is too many standard deviations away from its mean.

There is code in R for the (two-sided) turning point test, assuming the normal approximation is valid, turning.point.test.R.

If V_j actually behave not as i.i.d. variables but have a pattern such as $f(j) + \delta_j$ where δ_j are small random variables in relation to differences $f(j) - f(j-1)$ of the non-random smooth function f, then there will tend to be too few turning points in the V_j. If there are too many turning points then V_j change direction even more often than do i.i.d. variables, which we will see have turning points at about 2/3 of all values of j. This may suggest that successive V_j are negatively correlated. To detect either kind of departure from i.i.d. behavior, a two-sided test can be done. If the number of turning points in regression residuals is too small, it can indicate that the linear model is wrong, e.g., that the degree in polynomial regression is too low. For residuals, that kind of alternative seems more of interest, so we propose a one-sided test.

It has been shown (Stuart, 1954) that the turning point test is not actually very efficient as a test of the hypothesis that $V_1, ..., V_n$ are exactly i.i.d. Specifically, against the alternative regression model hypothesis that $V_j = \alpha + \beta j + \varepsilon_j$ where ε_j are i.i.d. $N(0, \sigma^2)$ for some σ, with $\beta \neq 0$, the usual estimator $\widehat{\beta}$ of the slope β provides a test statistic compared to which the turning point statistic has asymptotically 0 relative efficiency (Stuart, 1954, pp. 153-154).

But, if the regression model does hold with i.i.d. errors ε_j (not necessarily normal, but having mean 0 and finite variance) then the residuals $\widehat{\varepsilon}_j$ in the regression will be approximately i.i.d., for n large enough. For the residuals, the estimated slope $\widehat{\beta}$ will be exactly 0. Kendall, Stuart and Ord (1983, pp. 430-436) consider time series which beside a trend might have seasonal variations, although here we're concerned just with trend. They say (p. 430): "When seasonal variation and trend have been removed from the data we are left with a series [of residuals] which will present, in general, fluctuations of a more or less regular kind." They consider tests for whether these fluctuations are random, and first of all among "most suitable" tests, the turning point test (pp. 431-436).

First, the properties of the turning point test for actual i.i.d. V_j will be developed, then, we'll look at properties for residuals. The following fact gives the mean and variance of T_n under H_0 for $n \geq 4$. (For $n = 3$ the variance is 2/9.) These are well known and appear, for example, in the R code turning.point.test. A proof is given in Dudley and Hua (2009).

Theorem 1. *Under the hypothesis H_0: $V_1, ..., V_n$ are i.i.d. with a continuous distribution, and $n \geq 3$, we have $ET_n = 2(n-2)/3$, and for $n \geq 4$, $Var(T_n) = (16n-29)/90$.*

It is possible to calculate the exact distribution of T_n using the following combinatorial results (Stanley, 2008). Let $w = (a_1, ..., a_n)$ be a permutation of $1, ..., n$. Let $as(w)$ be the length of the longest alternating subsequence of w where a sequence $(b_1, ..., b_k)$ is alternating if $b_1 > b_2 < b_3 ....$ A relation between this and $T_n(w)$, the number of turning points in w, is that if $a_1 < a_2$ then $as(w) = T_n(w) + 1$ whereas if $a_1 > a_2$ then $as(w) = T_n(w) + 2$. Also, of those w with a given value of T_n, exactly half have $a_1 < a_2$. Consequently, if we let $t_k(n)$ be the number of permutations of n with k turning points for $k = 0, 1, ..., n-2$ and $a_k(n)$ be the number of permutations of n whose longest alternating subsequence has length k, then we have

$$\begin{aligned} t_{n-2}(n) &= 2a_n(n), \\ t_k(n) &= 2a_{k+2}(n) - t_{k+1}(n), \ k = 0, 1, ..., n-3. \end{aligned} \tag{1.1}$$

The distribution of T_n is

$$\Pr(T_n = k) = \frac{t_k(n)}{n!}, \ k = 0, 1, ..., n-2. \tag{1.2}$$

Let $b_k(n) = \sum_{j=1}^{k} a_j(n)$. We have the following result from Stanley (2008, Corollary 3.1) giving explicit formulas for $a_k(n)$ and $b_k(n)$.

Theorem 2. **[Stanley, 2008]** *For any positive integer n we have*

$$\begin{aligned} b_k(n) &= \frac{1}{2^{k-1}} \sum_{r+2s \leq k, r \equiv k (mod\, 2)} (-2)^s \binom{k-s}{(k+r)/2} \binom{n}{s} r^n, \quad 1 \leq k \leq n, \\ a_1(n) &= 1, \ a_k(n) = b_k(n) - b_{k-1}(n), \ 2 \leq k \leq n. \end{aligned} \tag{1.3}$$

Based on (1.1) and (1.3), we can compute numerical values of $t_k(n)$, then by (1.2) evaluate the distribution of T_n numerically. In Section 1.5, we will give tables of critical values and probabilities of T_n for n up to 50.

1.3 Regression models

Suppose given a model in which $Y_j = g(X_j|\theta)+\varepsilon_j$, where (X_j, Y_j) are observed for $j = 1, ..., n$, X_j are non-random design points, X_j and Y_j are real-valued, and $g(x|\theta)$ is a regression function with a parameter θ to be estimated from the data, with an estimate $\widehat{\theta} = \widehat{\theta}_n$. In a classical model, ε_j are assumed to be i.i.d. $N(0, \sigma^2)$ for some unknown σ. The ε_j are called *errors*, although they may result from random variation rather than measurement errors per se. The observable quantities $\widehat{\varepsilon}_j = Y_j - g(X_j|\widehat{\theta})$, which are estimates of the unobserved ε_j, are called *residuals*. For consistent estimation, by definition, if a true $\theta = \theta_0$ exists, $\widehat{\theta}_n$ will approach it in probability as n becomes large. Supposing that $g(x|\theta)$ is continuous with respect to θ, $g(X_j|\widehat{\theta}_n)$ will approximately equal $g(X_j|\theta_0)$, so $\widehat{\varepsilon}_j$ will approximately equal ε_j. Thus, approximately, we can apply a turning point test to the $\widehat{\varepsilon}_j$ to see if the model assumption of ε_j i.i.d. is valid. In order for turning points, or any patterns in residuals $\widehat{\varepsilon}_j$, to be meaningful, we can and do assume that $X_1 \leq X_2 \leq \cdots \leq X_n$.

For a linear model, suppose we have $m \geq 2$ functions f_i, $i = 1, \ldots, m$, linearly independent on $\{X_1, \ldots, X_n\}$, which implies there are at least m distinct values among the X_j. Then $\theta = (\theta_1, \ldots, \theta_m)^T \in \mathbb{R}^m$ and $g(x|\theta) \equiv \sum_{i=1}^m \theta_i f_i(x)$. Let M be the $n \times m$ matrix with $M_{ji} = f_i(X_j)$, so that M has full rank m. Then the least-squares and maximum likelihood estimate of θ is unique and is given by

$$\widehat{\theta} = (M^T M)^{-1} M^T Y \tag{1.4}$$

where $Y = (Y_1, ..., Y_n)^T$. Let $H := M(M^T M)^{-1} M^T$. Note that H is a symmetric matrix and $H^2 = H$, $HM = M$. This means H is an orthogonal projection that projects any n dimensional column vector onto the subspace spanned by the columns of M, which are linearly independent. The residuals are

$$\widehat{\varepsilon} = Y - M\widehat{\theta} = (I-H)Y = (I-H)(M\theta + \varepsilon) = (I-H)\varepsilon. \tag{1.5}$$

By (1.5), the residuals are linear transformations of ε so they follow a multivariate normal distribution with expectation $E[\widehat{\varepsilon}] = E[(I-H)\varepsilon] = 0$ and covariance

$$\mathrm{Cov}(\widehat{\varepsilon}) = \sigma^2 (I-H)(I-H)^T = \sigma^2(I-H). \tag{1.6}$$

The cross covariance between ε and $\widehat{\varepsilon}$ is

$$\mathrm{Cov}(\varepsilon, \widehat{\varepsilon}) = \sigma^2 (I-H)^T = \sigma^2(I-H). \tag{1.7}$$

Since $HM_{.i} = M_{.i}$ for any column $M_{.i} = (f_i(X_1), \ldots, f_i(X_n))^T$, we have

$$(f_i(X_1), \ldots, f_i(X_n))H = (H(f_i(X_1), \ldots, f_i(X_n))^T)^T = (f_i(X_1), \ldots, f_i(X_n)).$$

Then by (1.5), for $1 \leq i \leq m$,

$$\sum_{j=1}^{n} \widehat{\varepsilon}_j f_i(X_j) = (f_i(X_1), \ldots, f_i(X_n))\widehat{\varepsilon} = (f_i(X_1), \ldots, f_i(X_n))(I-H)\varepsilon = 0. \tag{1.8}$$

For $2 \leq i \leq n$, let $\delta_i = \varepsilon_i - \varepsilon_{i-1}$ and $\widehat{\delta}_i = \widehat{\varepsilon}_i - \widehat{\varepsilon}_{i-1}$. Also define $\delta := (\delta_2, ..., \delta_n)^T$ and $\widehat{\delta} := (\widehat{\delta}_2, ..., \widehat{\delta}_n)^T$. Then $\delta = \Lambda\varepsilon$ and $\widehat{\delta} = \Lambda\widehat{\varepsilon} = \Lambda(I-H)\varepsilon$ where

$$\Lambda = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & & \\ \vdots & & \ddots & \ddots & \\ 0 & \ldots & 0 & -1 & 1 \end{pmatrix}_{(n-1)\times n} \tag{1.9}$$

and the indices of rows are $2, ..., n$. We have

$$\begin{pmatrix} \delta \\ \widehat{\delta} \end{pmatrix} = \begin{pmatrix} \Lambda \\ \Lambda(I-H) \end{pmatrix} \varepsilon.$$

Hence $(\delta_2, ...\delta_n, \widehat{\delta}_2, ..., \widehat{\delta}_n)^T = (\delta^T, \widehat{\delta}^T)^T$ follow a multivariate normal distribution with $E[\delta] = E[\widehat{\delta}] = 0$ and

$$\text{Cov}(\delta) = \sigma^2 \Lambda\Lambda^T, \tag{1.10}$$

$$\text{Cov}(\widehat{\delta}) = \sigma^2 \Lambda(I - H)(\Lambda(I - H))^T = \sigma^2 \Lambda(I - H)\Lambda^T = \sigma^2(\Lambda\Lambda^T - \Lambda H \Lambda^T), \tag{1.11}$$

$$\text{Cov}(\delta, \widehat{\delta}) = \sigma^2 \Lambda(\Lambda(I - H))^T = \sigma^2 \Lambda(I - H)\Lambda^T = \sigma^2(\Lambda\Lambda^T - \Lambda H \Lambda^T). \tag{1.12}$$

We can calculate

$$\Lambda\Lambda^T = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{pmatrix}_{(n-1)\times(n-1)}, \tag{1.13}$$

where indices of both rows and columns are $2, \dots, n$. Let $\widehat{T}_n$ be the number of turning points in the residuals $\widehat{\varepsilon}_j$ for a given n.

Proposition 3. *Suppose given a linear model with i.i.d. normal errors ε_j, such that $Var(\widehat{\delta}_i) > 0$ for $i = 2, \dots, n$. Then the correlation ρ_i of δ_i and $\widehat{\delta}_i$ is well-defined with $\rho_i > 0$ and setting*

$$\alpha_i = 2(1 - \rho_i^2), \tag{1.14}$$

we have

$$\rho_i := \frac{Cov(\delta_i, \widehat{\delta}_i)}{\sqrt{Var(\delta_i)\, Var(\widehat{\delta}_i)}} = \sqrt{1 - \frac{\alpha_i}{2}}, \tag{1.15}$$

$$E[|\widehat{T}_n - T_n|] \leq \sum_{i=2}^{n-1} \frac{\arcsin\sqrt{1 - \rho_i^2}}{\pi} + \sum_{i=2}^{n-1} \frac{\arcsin\sqrt{1 - \rho_{i+1}^2}}{\pi}, \tag{1.16}$$

and

$$\sum_{i=2}^{n-1} \arcsin\sqrt{1 - \rho_i^2} \leq \sum_{i=2}^{n-1} \frac{\sqrt{1 - \rho_i^2}}{\rho_i} = \sum_{i=2}^{n-1} \sqrt{\frac{\alpha_i}{2 - \alpha_i}}. \tag{1.17}$$

Proof. If $\mathrm{Var}(\widehat{\delta}_i) > 0$ then since $\mathrm{Var}(\delta_i) = 2\sigma^2 > 0$ always, the correlation ρ_i is well-defined. By (1.11) and (1.12), we have $\rho_i > 0$. Relation (1.15) then follows immediately from (1.14). The quadrant probabilities are known and easily seen to be

$$\Pr(\delta_i > 0, \widehat{\delta}_i > 0) = \Pr(\delta_i < 0, \widehat{\delta}_i < 0) = \frac{1}{4} + \frac{\arcsin \rho_i}{2\pi}. \tag{1.18}$$

For $i = 2, ..., n-1$, let $\hat{I}_i = 1$ if $\widehat{\varepsilon}_j$ have a turning point at i and 0 otherwise. Let us also define $A_i = \{\delta_i \widehat{\delta}_i > 0\} = \{\delta_i > 0, \widehat{\delta}_i > 0\} \bigcup \{\delta_i < 0, \widehat{\delta}_i < 0\}$. Then $A_i \cap A_{i+1}$ implies that $\widehat{\varepsilon}_j$ has a turning point at i iff ε_j does. It follows that

$$\begin{aligned}
\Pr(I_i = \hat{I}_i) &\geq \Pr(A_i \cap A_{i+1}) \\
&\geq \Pr(A_i) + \Pr(A_{i+1}) - 1 \\
&= \Pr(\delta_i > 0, \widehat{\delta}_i > 0) + \Pr(\delta_i < 0, \widehat{\delta}_i < 0) \\
&\quad + \Pr(\delta_{i+1} > 0, \widehat{\delta}_{i+1} > 0) + \Pr(\delta_{i+1} < 0, \widehat{\delta}_{i+1} < 0) - 1 \\
&= \frac{\arcsin \rho_i}{\pi} + \frac{\arcsin \rho_{i+1}}{\pi}.
\end{aligned} \tag{1.19}$$

Therefore we have

$$\begin{aligned}
E[|\widehat{T}_n - T_n|] &= E\left[\left|\sum_{i=2}^{n-1}(\hat{I}_i - I_i)\right|\right] \\
&\leq \sum_{i=2}^{n-1} E[|\hat{I}_i - I_i|] \\
&= \sum_{i=2}^{n-1}(1 - \Pr(I_i = \hat{I}_i)) \\
&\leq \sum_{i=2}^{n-1}\left(\frac{1}{2} - \frac{\arcsin \rho_i}{\pi} + \frac{1}{2} - \frac{\arcsin \rho_{i+1}}{\pi}\right) \\
&= \sum_{i=2}^{n-1} \frac{\arccos \rho_i}{\pi} + \sum_{i=2}^{n-1} \frac{\arccos \rho_{i+1}}{\pi} \\
&= \sum_{i=2}^{n-1} \frac{\arcsin \sqrt{1-\rho_i^2}}{\pi} + \sum_{i=2}^{n-1} \frac{\arcsin \sqrt{1-\rho_{i+1}^2}}{\pi},
\end{aligned} \tag{1.20}$$

giving (1.16). Note that $\frac{d}{dx}\arcsin(x) = 1/\sqrt{1-x^2} \leq 1/\rho_i$ for all $0 \leq x \leq \sqrt{1-\rho_i^2}$

so $\arcsin(x) \le \frac{1}{\rho_i}x$ for all $0 \le x \le \sqrt{1-\rho_i^2}$. Hence (1.17) holds, completing the proof. □

We will mainly consider the case of polynomial regression $Y_j = \beta_0 + \beta_1 X_j + \cdots + \beta_k X_j^k + \varepsilon_j$, $k \in \mathbf{N}$. Here $m = k+1$, $f_i(x) \equiv x^{i-1}$ for $i = 1, \ldots, m$, $\theta = \beta := (\beta_0, \beta_1, ..., \beta_k)^T$, and

$$M = \tilde{X} = \begin{pmatrix} 1 & X_1 & \ldots & X_1^k \\ 1 & X_2 & \ldots & X_2^k \\ 1 & \vdots & \vdots & \vdots \\ 1 & X_n & \ldots & X_n^k \end{pmatrix}. \tag{1.21}$$

In section 1.4, we show that (1.8) implies that for any kth order regression there will be at least k turning points among the residuals if the design points are all different.

1.3.1 Simple linear regression models

First we derive some asymptotic results for simple linear regression models. For simple linear regression where $k = 1$, the estimates for β_0 and β_1 are

$$\widehat{\beta}_1 = \frac{\sum_{j=1}^n (X_j - \overline{X}) Y_j}{S_X^2} = \beta_1 + \frac{\sum_{j=1}^n (X_j - \overline{X})\varepsilon_j}{S_X^2}, \quad \beta_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}, \tag{1.22}$$

where

$$S_X^2 := \sum_{j=1}^n (X_j - \overline{X})^2 > 0 \tag{1.23}$$

since the X_j are not all equal (there are at least $m \ge 2$ distinct X_j, here $m = 2$). We calculate the (r, s)th entry of the matrix H to be

$$H_{rs} = \frac{1}{n} + \frac{(X_r - \overline{X})(X_s - \overline{X})}{S_X^2}. \tag{1.24}$$

Hence by (1.5) and (1.22),

$$\varepsilon_j - \widehat{\varepsilon}_j = \sum_{l=1}^{n} H_{jl}\varepsilon_l = \overline{\varepsilon} + (X_j - \overline{X})(\widehat{\beta}_1 - \beta_1). \tag{1.25}$$

Let $d_j := X_j - X_{j-1}$ for $2 \leq j \leq n$. By (1.25)

$$\delta_j - \widehat{\delta}_j = (\varepsilon_j - \widehat{\varepsilon}_j) - (\varepsilon_{j-1} - \widehat{\varepsilon}_{j-1}) = d_j(\widehat{\beta}_1 - \beta_1). \tag{1.26}$$

From (1.22) and since X_j are non-random it follows that $\beta_1 - \widehat{\beta}_1$ has a $N(0, \sigma^2/S_X^2)$ distribution. Thus if n is large, $|d_j(\beta_1 - \widehat{\beta}_1)|$ is probably small, of order $1/n^{3/2}$, if d_j is of typical order $\sqrt{S_X^2/n^3}$, but not so small if d_j is larger. Such large values of d_j can occur for relatively few values of j. If d_j and d_{j+1} are not so large, then with large probability δ_j and $\widehat{\delta}_j$ will have the same sign and so will δ_{j+1} and $\widehat{\delta}_{j+1}$, which implies that ε_i and $\widehat{\varepsilon}_i$ will either both or neither have a turning point at $i = j$.

Here is a more precise formulation, which shows that $E|\widehat{T}_n - T_n| < 1$ for all $n \geq 3$ and any design points.

Theorem 4. *In simple linear regression with ε_j i.i.d. $N(0, \sigma^2)$, for all $n \geq 3$,*

$$E[|\widehat{T}_n - T_n|] \leq \frac{2}{\pi}\sqrt{\frac{n-1}{n-2}},$$

so that as $n \to \infty$,

$$E[|\widehat{T}_n - T_n|] \leq \frac{2}{\pi} + O\left(\frac{1}{n}\right).$$

Remark. For each n, T_n depends only on the ε_j, not on the design points. The distribution of T_n can be found for each n exactly for n up to values for which the normal approximation works well. The distribution of $\widehat{T}_n$ does depend on the design points, but as will be shown, $E|\widehat{T}_n - T_n|/\sqrt{n} = O(1/\sqrt{n})$ as $n \to \infty$ (Theorem 4) with faster rates for reasonably well behaved design points (Theorems 6 and 7).

Proof. To apply (1.10), (1.11), and (1.12) to simple linear regression, with (1.24) we can also calculate

$$(\Lambda H \Lambda^T)_{uv} = d_u d_v / S_X^2, \tag{1.27}$$

$u, v = 2, \ldots, n$. Therefore

$$\operatorname{Var}(\widehat{\delta}_i) = (2 - d_i^2/S_X^2)\sigma^2 \tag{1.28}$$

and

$$\operatorname{Cov}(\delta_i, \widehat{\delta}_i) = (2 - d_i^2/S_X^2)\sigma^2 \tag{1.29}$$

for any $2 \leq i \leq n$ and relation (1.15) holds with $\alpha_i := \alpha_i^{(s)} = d_i^2/S_X^2$ (the superscript (s) indicating "simple" regression). For any $i = 2, ..., n$, if $\overline{X} \notin (X_{i-1}, X_i)$ then $d_i \leq \max(|X_{i-1} - \overline{X}|, |X_i - \overline{X}|) \leq \sqrt{S_X^2}$. If $\overline{X} \in (X_{i-1}, X_i)$ for some $2 < i < n$, then

$$\begin{aligned} d_i^2 &= (X_i - \overline{X} + \overline{X} - X_{i-1})^2 \\ &\leq 2\left((X_i - \overline{X})^2 + (\overline{X} - X_{i-1})^2\right) \leq \sum_{j=i-2}^{i+1} (X_j - \overline{X})^2 \leq S_X^2. \end{aligned} \tag{1.30}$$

If $\overline{X} \in (X_1, X_2)$, then letting $u := \overline{X} - X_1 > 0$ and $v := X_2 - \overline{X} > 0$ we have

$$S_X^2 \geq u^2 + (n-1)v^2 \geq (n-1)(u+v)^2/n = (n-1)d_2^2/n, \tag{1.31}$$

which follows from $(u - (n-1)v)^2 \geq 0$. Similarly we have $S_X^2 \geq \frac{n-1}{n} d_n^2$ if $\overline{X} \in (X_{n-1}, X_n)$. Therefore (1.17) gives

$$\sum_{i=2}^{n-1} \arcsin \sqrt{1 - \rho_i^2} \leq \sqrt{\frac{n-1}{n-2}} \sum_{i=2}^{n-1} \sqrt{\alpha_i}, \tag{1.32}$$

and we also have

$$\sum_{i=2}^{n-1} \arcsin \sqrt{1 - \rho_{i+1}^2} \leq \sqrt{\frac{n-1}{n-2}} \sum_{i=2}^{n-1} \sqrt{\alpha_{i+1}}. \tag{1.33}$$

Therefore by (1.20)

$$\begin{aligned} E[|\widehat{T}_n - T_n|] &\leq \sum_{i=2}^{n-1} \frac{\arcsin\sqrt{1-\rho_i^2}}{\pi} + \sum_{i=2}^{n-1} \frac{\arcsin\sqrt{1-\rho_{i+1}^2}}{\pi} \\ &\leq \frac{1}{\pi}\sqrt{\frac{n-1}{n-2}}\left(\sum_{i=2}^{n-1}\sqrt{\alpha_i} + \sum_{i=2}^{n-1}\sqrt{\alpha_{i+1}}\right). \end{aligned} \tag{1.34}$$

However

$$\begin{aligned} & \sum_{i=2}^{n-1}\sqrt{\alpha_i} + \sum_{i=2}^{n-1}\sqrt{\alpha_{i+1}} \qquad (1.35) \\ = \ & \frac{1}{\sqrt{S_X^2}}\sum_{i=2}^{n-1}(d_i + d_{i+1}) \\ = \ & \frac{1}{\sqrt{S_X^2}}(X_{n-1} - X_1 + X_n - X_2) \\ \leq \ & \frac{1}{\sqrt{S_X^2}}(|X_1 - \overline{X}| + |X_2 - \overline{X}| + |X_{n-1} - \overline{X}| + |X_n - \overline{X}|) \\ \leq \ & \frac{1}{\sqrt{S_X^2}} 2\sqrt{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + (X_{n-1} - \overline{X})^2 + (X_n - \overline{X})^2} \\ \leq \ & 2. \end{aligned}$$

Therefore (1.34) gives

$$E[|\widehat{T}_n - T_n|] \leq \frac{1}{\pi}\sqrt{\frac{n-1}{n-2}}\left(\sum_{i=2}^{n-1}\sqrt{\alpha_i} + \sum_{i=2}^{n-1}\sqrt{\alpha_{i+1}}\right) \leq \frac{2}{\pi}\sqrt{\frac{n-1}{n-2}} = \frac{2}{\pi} + O\left(\frac{1}{n}\right).$$

□

We will show in the next proposition that the upper bound $O(1)$ given by Theorem 4 cannot be replaced by $o(1)$ for general design points. However, for most practical cases where the design points are not too unequally spaced, or if they come from random normal samples, the bound can be greatly improved, as will be shown in Theorems 6 and 7 respectively.

Proposition 5. *In simple linear regression with ε_j i.i.d. $N(0, \sigma^2)$, for each $n \geq 3$*

there are design points x_i, $i = 1, 2, ..., n$, such that as $n \to \infty$,

$$E[|\widehat{T}_n - T_n|] \geq \frac{1}{\pi\sqrt{2}} + O\left(\frac{1}{n}\right).$$

Proof. Let us consider the following design points. Let $x_1 = -1$, $x_2 = 0$, and $d_i = x_i - x_{i-1} = 2/(n-1)(n-2)$ for $3 \leq i \leq n$. Therefore the majority of the design points are clustered around zero with only one outlier x_1. It is easy to obtain that $\overline{x} = 0$ and $S_x^2 = 1 + \frac{2(2n-3)}{3(n-1)(n-2)} \geq 1$. Thus as $n \to \infty$ we have $S_x^2 = 1 + O(1/n)$, $\alpha_2 = 1 + O(1/n)$ and $\alpha_i = O(1/n^4)$ for $i = 3, ..., n$. Using the same notations as in Theorem 4 we have

$$\begin{aligned}
& E[|\widehat{T}_n - T_n|] \\
= \; & E[|\sum_{i=2}^{n-1}(\hat{I}_i - I_i)|] \\
\geq \; & E[|\hat{I}_2 - I_2|] - \sum_{i=3}^{n-1} E[|\hat{I}_i - I_i|] \\
\geq \; & E[|\hat{I}_2 - I_2|] - \frac{1}{\pi}\sum_{i=3}^{n-1}\left(\arcsin\sqrt{1-\rho_i^2} + \arcsin\sqrt{1-\rho_{i+1}^2}\right) \\
\geq \; & E[|\hat{I}_2 - I_2|] - \frac{1}{\pi}\sqrt{\frac{n-1}{n-2}}\sum_{i=3}^{n-1}(\sqrt{\alpha_i} + \sqrt{\alpha_{i+1}}) \\
= \; & E[|\hat{I}_2 - I_2|] - \frac{1}{\pi\sqrt{S_x^2}}\sqrt{\frac{n-1}{n-2}}\sum_{i=3}^{n-1}(d_i + d_{i+1}) \\
\geq \; & E[|\hat{I}_2 - I_2|] - \frac{4(n-3)}{\pi(n-1)(n-2)}\sqrt{\frac{n-1}{n-2}} \\
= \; & \Pr(|\hat{I}_2 - I_2| = 1) - \frac{4(n-3)}{\pi(n-1)(n-2)}\sqrt{\frac{n-1}{n-2}}. \qquad (1.36)
\end{aligned}$$

Note that $\delta_2\widehat{\delta}_2 < 0$ and $\delta_3\widehat{\delta}_3 > 0$ imply that $\hat{I}_2 \neq I_2$. Using the quadrant probabilities of bivariate normal distributions (1.18) we have

$$
\begin{aligned}
\Pr(|\hat{I}_2 - I_2| = 1) &\geq \Pr(\delta_2\widehat{\delta}_2 < 0, \delta_3\widehat{\delta}_3 > 0) \\
&\geq \Pr(\delta_2\widehat{\delta}_2 < 0) + \Pr(\delta_3\widehat{\delta}_3 > 0) - 1 \\
&= \frac{1}{2} - \frac{\arcsin\rho_2}{\pi} + \frac{1}{2} + \frac{\arcsin\rho_3}{\pi} - 1 \\
&= \frac{1}{\pi}(\arcsin\rho_3 - \arcsin\rho_2).
\end{aligned}
$$

Note that

$$
\begin{aligned}
\sin(\arcsin\rho_3 - \arcsin\rho_2) &= \rho_3\sqrt{1-\rho_2^2} - \rho_2\sqrt{1-\rho_3^2} \\
&= \sqrt{\left(1-\frac{\alpha_3}{2}\right)\frac{\alpha_2}{2}} - \sqrt{\left(1-\frac{\alpha_2}{2}\right)\frac{\alpha_3}{2}} \\
&= \frac{1}{\sqrt{2}} + O\left(\frac{1}{n}\right),
\end{aligned}
$$

which also implies $0 \leq \arcsin\rho_3 - \arcsin\rho_2 \leq \pi/2$ for n large enough. But $x \geq \sin(x)$ for $0 \leq x \leq \pi/2$. Therefore

$$
\begin{aligned}
\Pr(|\hat{I}_2 - I_2| = 1) &\geq \frac{1}{\pi}(\arcsin\rho_3 - \arcsin\rho_2) \\
&\geq \frac{1}{\pi}\sin(\arcsin\rho_3 - \arcsin\rho_2) \\
&= \frac{1}{\pi\sqrt{2}} + O\left(\frac{1}{n}\right).
\end{aligned}
$$

and it follows from (1.36) that

$$
E[|\widehat{T}_n - T_n|] \geq \frac{1}{\pi\sqrt{2}} + O\left(\frac{1}{n}\right).
$$

□

Theorem 6. *In simple linear regression with ε_j i.i.d. $N(0,\sigma^2)$, let $\widehat{T}_n$ be the number of turning points in $\widehat{\varepsilon}_j$ for a given n. Let $d_i := X_i - X_{i-1}$ for $2 \leq i \leq n$ and $d_m = \min_{2\leq i\leq n} d_i$, $d_M = \max_{2\leq i\leq n} d_i$. If $d_M/d_m \leq \gamma$ for some constant γ, then for*

$n \geq 3$

$$E[|\widehat{T}_n - T_n|] \leq \frac{4\sqrt{3}\gamma}{\pi}\ n^{-0.5}$$

and

$$|E\widehat{T}_n - \frac{2}{3}(n-2)|\ \leq \frac{4\sqrt{3}\gamma}{\pi}\ n^{-0.5} = O(n^{-0.5}).$$

In particular, if the design points $X_j,\ j = 1, 2, ..., n$ *are equally spaced, then* $E[|\widehat{T}_n - T_n|] \leq \frac{4\sqrt{3}}{\pi}\ n^{-0.5}$ *and* $|E\widehat{T}_n - \frac{2}{3}(n-2)|\ \leq \frac{4\sqrt{3}}{\pi}\ n^{-0.5}$.

Proof. Let k be the index such that $X_k \leq \overline{X} < X_{k+1}$ and $u = \overline{X} - X_k$, $v = X_{k+1} - \overline{X}$. We have

$$\begin{aligned} S_X^2 = \sum_{j=1}^{n}(X_j - \overline{X})^2 &\geq \sum_{j=1}^{k}((j-1)d_m + u)^2 + \sum_{j=1}^{n-k}((j-1)d_m + v)^2 \\ &\geq d_m^2 \sum_{j=1}^{k-1} j^2 + d_m^2 \sum_{j=1}^{n-k-1} j^2 + u^2 + v^2 \\ &\geq d_m^2 f(k), \end{aligned}$$

where $f(k) = \frac{1}{6}(k-1)k(2k-1) + \frac{1}{6}(n-k-1)(n-k)(2n-2k-1) + \frac{1}{2}$. Note that $f'(x) = 0$ at $x = n/2$ which minimizes f on $[0, n-1]$. Therefore

$$S_X^2 \geq d_m^2 f(n/2) = \left(\frac{1}{12}n(n-1)(n-2) + 1/2\right) d_m^2.$$

Then (1.34) gives

$$\begin{aligned} E[|\widehat{T}_n - T_n|] &\leq \frac{1}{\pi}\sqrt{\frac{n-1}{n-2}}(\sum_{i=2}^{n-1}\sqrt{\alpha_i} + \sum_{i=2}^{n-1}\sqrt{\alpha_{i+1}}) \\ &= \frac{1}{\pi\sqrt{S_X^2}}\sqrt{\frac{n-1}{n-2}}\sum_{i=2}^{n-1}(d_i + d_{i+1}) \\ &\leq \frac{2(n-2)d_M}{\pi d_m\sqrt{1/2 + n(n-1)(n-2)/12}}\sqrt{\frac{n-1}{n-2}} \\ &\leq \frac{4\sqrt{3}\gamma}{\pi} n^{-0.5} = O(n^{-0.5}). \end{aligned}$$

The other statement follows directly. □

Theorem 7. *In simple linear regression with ε_j i.i.d. $N(0,\sigma^2)$, let $\widehat{T}_n$ be the number of turning points in $\widehat{\varepsilon}_j$ for a given n. If the design points $X_j,\ j=1,2,...,n$ are i.i.d. samples from a normal distribution rearranged in ascending order, then for n large enough, with probability at least $1-\frac{1}{n\sqrt{2\pi\log n}}-\frac{1}{n}$,*

$$E_X[|\widehat{T}_n - T_n|] \leq \frac{8}{\pi}\frac{\sqrt{\log n}}{\sqrt{n-1-2\sqrt{(n-1)\log n}}}\sqrt{\frac{n-1}{n-2}} = O(n^{-0.5}\sqrt{\log n}),$$

and

$$\left|E_X\widehat{T}_n - \frac{2}{3}(n-2)\right| \leq \frac{8}{\pi}\frac{\sqrt{\log n}}{\sqrt{n-1-2\sqrt{(n-1)\log n}}}\sqrt{\frac{n-1}{n-2}} = O(n^{-0.5}\sqrt{\log n}),$$

where E_X denotes conditional expectation given $X_1,...,X_n$.

Proof. . Note that $\widehat{\varepsilon}_i - \widehat{\varepsilon}_{i+1} = \varepsilon_i - \varepsilon_{i+1} + \zeta_i$ where $\zeta_i := (\beta_1 - \widehat{\beta}_1)(X_i - X_{i-1})$ has a $N(0, \sigma^2 d_i^2/S_x^2)$ distribution given $X_1,...,X_n$. Since rescaling of X_i does not change d_i^2/S_x^2 and therefore the distribution of $\widehat{T}_n$, we may assume that the design points $X_j,\ j=1,2,...,n$ are reordered i.i.d. samples Y_j from a normal distribution $N(\mu,1)$. Consequently S_x^2 follows a χ^2 distribution with $n-1$ degrees of freedom.

It is easy to calculate the moment generating function $M(t)$ of $(n-1)-S_x^2$: $M(t) = e^{(n-1)t}(1+2t)^{-(n-1)/2}$ which is finite for all $t>0$. Therefore for any $0<\alpha<n-1$,

$$\begin{aligned}\Pr((n-1)-S_x^2 \geq \alpha) &= \Pr(e^{t(n-1-S_x^2)} \geq e^{t\alpha}, \forall\ t>0) \\ &\leq \inf_{t>0}\{e^{-\alpha t}E[e^{t(n-1-S_x^2)}]\} \\ &= \inf_{t>0}\{e^{(n-1-\alpha)t}(1+2t)^{-(n-1)/2}\}.\end{aligned}$$

It is easy to find that the minimum of the function $f(t) := e^{(n-1-\alpha)t}(1+2t)^{-(n-1)/2}$

is achieved at $t_0 = \frac{\alpha}{2(n-1-\alpha)}$ and

$$\begin{aligned} f(t_0) &= e^{\alpha/2}(\frac{n-1-\alpha}{n-1})^{(n-1)/2} \\ &= e^{\alpha/2}e^{\frac{n-1}{2}\log(1-\alpha/(n-1))} \\ &\leq e^{\alpha/2}e^{\frac{n-1}{2}(-\alpha/(n-1)-\frac{1}{2}\alpha^2/(n-1)^2)} \\ &= e^{-\frac{1}{4}\alpha^2/(n-1)}. \end{aligned}$$

Hence $\Pr(S_x^2 \leq (n-1)-\alpha) \leq e^{-\frac{1}{4}\alpha^2/(n-1)}$. Setting $\alpha = 2\sqrt{(n-1)\log n}$, we have $\Pr(S_x^2 \leq (n-1) - 2\sqrt{(n-1)\log n}\ ) \leq 1/n$. Since X_j are reordered i.i.d. samples Y_j with $N(\mu, 1)$ distribution,

$$\Pr(X_n - \mu \geq 2\sqrt{\log n}) \leq \sum_{j=1}^{n} \Pr(Y_j - \mu \geq 2\sqrt{\log n}) \leq \sum_{j=1}^{n} \frac{e^{-2\log n}}{2\sqrt{2\pi\log n}} = \frac{1}{2n\sqrt{2\pi\log n}}.$$

By symmetry, $\Pr(X_1 - \mu \leq -2\sqrt{\log n}) \leq 1/(2n\sqrt{2\pi\log n})$. Therefore

$$\begin{aligned} &\Pr(X_n - X_1 \leq 4\sqrt{\log n}, S_x^2 \geq (n-1) - 2\sqrt{(n-1)\log n}\ ) \\ \geq\ &\Pr(X_n - \mu \leq 2\sqrt{\log n}, X_1 - \mu \geq -2\sqrt{\log n}, S_x^2 \geq (n-1) - 2\sqrt{(n-1)\log n}\ ) \\ \geq\ &1 - \Pr(X_n - \mu \geq 2\sqrt{\log n}) - \Pr(X_1 - \mu \leq -2\sqrt{\log n}) \\ &-\Pr(S_x^2 \leq (n-1) - 2\sqrt{(n-1)\log n}\ ) \\ \geq\ &1 - \frac{1}{n\sqrt{2\pi\log n}} - \frac{1}{n}. \end{aligned}$$

Hence with probability greater than $1 - \frac{1}{n\sqrt{2\pi\log n}} - \frac{1}{n}$, we have $\sum_{i=2}^{n-1}(d_i + d_{i+1}) \leq 2(X_n - X_1) \leq 8\sqrt{S_x^2}\sqrt{\log n/(n-1-2\sqrt{(n-1)\log n})}$ and inequality (1.34) becomes

$$\begin{aligned} E_X[|\widehat{T}_n - T_n|] &\leq \frac{1}{\pi\sqrt{S_x^2}}\sqrt{\frac{n-1}{n-2}}\sum_{i=2}^{n-1}(d_i + d_{i+1}) \\ &\leq \frac{8}{\pi}\frac{\sqrt{\log n}}{\sqrt{n-1-2\sqrt{(n-1)\log n}}}\sqrt{\frac{n-1}{n-2}} = O(n^{-0.5}\sqrt{\log n}). \end{aligned}$$

For $n \geq 20$, we have

$$\begin{aligned} & \frac{\sqrt{\log n}}{\sqrt{n-1-2\sqrt{(n-1)\log n}}}\sqrt{\frac{n-1}{n-2}} \\ = & \; n^{-0.5}\sqrt{\log n}\frac{1}{\sqrt{1-\frac{1}{n}-\frac{2\sqrt{(n-1)\log n}}{n}}}\sqrt{\frac{n-1}{n-2}} \\ \leq & \; 2.33n^{-0.5}\sqrt{\log n}. \end{aligned}$$

So for $n \geq 20$ with probability greater than $1 - \frac{1}{n\sqrt{2\pi \log n}} - \frac{1}{n}$ we have

$$E_X[|\widehat{T}_n - T_n|] \;\leq\; 6n^{-0.5}\sqrt{\log n}.$$

The other statement follows directly. □

Proposition 8. *In simple linear regression with ε_j i.i.d. $N(0, \sigma^2)$, let $\widehat{T}_n$ be the number of turning points in $\widehat{\varepsilon}_j$ for a given n. If the design points x_j, $j = 1, 2, ..., n$ are equally spaced, then for all $n \geq 4$*

$$E[|\widehat{T}_n - T_n|] \geq \frac{2\sqrt{6}}{\pi}\left(\frac{1}{2} - \frac{1}{\pi}\arcsin\frac{1}{3}\right)\frac{1}{\sqrt{n}} + O(n^{-1}).$$

Proof. A crucial part of the proof of the proposition is bounding of quadrivariate normal probabilities such as $\Pr(\delta_i > 0, \widehat{\delta}_i > 0, \delta_j < 0, \widehat{\delta}_j > 0)$ for all pairs of i, j. Let d be the distance between consecutive design points and

$$\alpha := \frac{d^2}{S_X^2} = \frac{12}{n^3 - n} = \frac{12}{n^3} + O\left(\frac{1}{n^5}\right). \tag{1.37}$$

By (1.10), (1.11), (1.12), and (1.27), $\mathrm{Var}(\delta_i) = 2\sigma^2$, $\mathrm{Var}(\widehat{\delta}_i) = (2-\alpha)\sigma^2$, $\mathrm{Cov}(\delta_{i-1}, \delta_i) = -\sigma^2$ for any $3 \leq i \leq n$, $\mathrm{Cov}(\delta_i, \delta_j) = 0$ for any $2 \leq i, j \leq n$ such that $|i - j| > 1$, $\mathrm{Cov}(\delta_i, \widehat{\delta}_i) = (2 - \alpha)\sigma^2$ for any $2 \leq i \leq n$ by (1.29), $\mathrm{Cov}(\delta_{i-1}, \widehat{\delta}_i) = \mathrm{Cov}(\delta_i, \widehat{\delta}_{i-1}) = (-1 - \alpha)\sigma^2$ for any $3 \leq i \leq n$, $\mathrm{Cov}(\delta_i, \widehat{\delta}_j) = -\alpha\sigma^2$ for any $2 \leq i, j \leq n$ such that $|i - j| > 1$, $\mathrm{Cov}(\widehat{\delta}_{i-1}, \widehat{\delta}_i) = (-1 - \alpha)\sigma^2$ for any $3 \leq i \leq n$ and $\mathrm{Cov}(\widehat{\delta}_i, \widehat{\delta}_j) = -\alpha\sigma^2$ for any $2 \leq i, j \leq n$ such that $|i-j| > 1$. Correlations can be calculated easily from the covariances and we will denote the correlation between $\delta_i(\widehat{\delta}_i), \delta_j(\widehat{\delta}_j)$ by $\rho(\delta_i(\widehat{\delta}_i), \delta_j(\widehat{\delta}_j))$.

First we find a lower bound for the probability $\Pr(\delta_i\widehat{\delta}_i < 0$ for exactly one index $2 \leq i \leq n)$. Note that $\delta_i\widehat{\delta}_i < 0$ for exactly one index i does not necessarily imply that $|T_n - \widehat{T}_n| > 0$. If $i = 2$ or n, it is easy to see this implies $|T_n - \widehat{T}_n| = 1$. If $2 < i < n$ and if $\delta_{i-1}\delta_{i+1} > 0$ this implies that $|T_n - \widehat{T}_n| = 2$. However if $\delta_{i-1}\delta_{i+1} < 0$ then both errors and residuals will have exactly one turning point at i or $i-1$ and if the turning point occurs at i $(i-1)$ for the errors it will occur at $i-1$ (i) for the residuals. Consequently we will have $|T_n - \widehat{T}_n| = 0$.

To bound $\Pr(\delta_i\widehat{\delta}_i < 0$ for exactly one index $2 \leq i \leq n)$ let us recall a Bonferroni inequality. Let us define events $B_i = \{\delta_i\widehat{\delta}_i < 0\}$ and $q_i = \Pr(B_i)$ for $2 \leq i \leq n$. We also define $q_{ij} = \Pr(B_i \cap B_j)$ for $2 \leq i, j \leq n$. Finally let $Q_1 = \sum_{i=2}^n q_i$ and $Q_2 = \sum_{i=2}^{n-1}\sum_{j=i+1}^n q_{ij}$. We have the following inequality (Bonferroni (1936); also Galambos and Simonelli (1996, p. 12)):

$$\Pr\left(\bigcup_{i=2}^n B_i\right) \geq Q_1 - Q_2. \tag{1.38}$$

Note that

$$\Pr(\delta_i\widehat{\delta}_i < 0 \text{ for exactly one index } 2 \leq i \leq n) = \Pr\left(\bigcup_{i=2}^n B_i\right) - \Pr\left(\bigcup_{i=2}^{n-1}\bigcup_{j=i+1}^n B_i \cap B_j\right),$$

and we have

$$\Pr\left(\bigcup_{i=2}^{n-1}\bigcup_{j=i+1}^n B_i \cap B_j\right) \leq \sum_{i=2}^{n-1}\sum_{j=i+1}^n \Pr(B_i \cap B_j) = Q_2.$$

Therefore we obtain the following inequality,

$$\Pr(\delta_i\widehat{\delta}_i < 0 \text{ for exactly one index } 2 \leq i \leq n) \geq Q_1 - 2Q_2. \tag{1.39}$$

To calculate Q_1, note that

$$q_i = 1/2 - \frac{1}{\pi}\arcsin\rho_{i,\hat{i}} = 1/2 - \frac{1}{\pi}\arcsin\sqrt{1-\frac{\alpha}{2}} = \frac{1}{\pi}\arcsin\sqrt{\frac{\alpha}{2}}. \tag{1.40}$$

Hence

$$Q_1 = \frac{n-1}{\pi}\arcsin\sqrt{\frac{\alpha}{2}}. \tag{1.41}$$

Next we find an upper bound for Q_2 and this takes a few steps. Due to symmetry among the correlations, $\Pr(\delta_i > 0, \widehat{\delta}_i < 0, \delta_j < 0, \widehat{\delta}_j > 0)$ attain only two values, one for pairs i, j with $|i-j| = 1$ and the other for pairs i, j with $|i-j| > 1$. We first calculate an upper bound for $\Pr(\delta_{i-1} > 0, \widehat{\delta}_{i-1} < 0, \delta_i < 0, \widehat{\delta}_i > 0)$. For convenience, let us write $z_1^{(1)} = -\delta_{i-1}/(\sqrt{2}\sigma)$, $z_2^{(1)} = \widehat{\delta}_{i-1}/(\sqrt{2-\alpha}\sigma)$, $z_3^{(1)} = \delta_i/(\sqrt{2}\sigma)$, and $z_4^{(1)} = -\widehat{\delta}_i/(\sqrt{2-\alpha}\sigma)$. Let Σ_1 be the covariance matrix and $R_1 = (\rho_{kl}^{(1)})$ the correlation matrix of $z_i^{(1)}$. Since $z_i^{(1)}$ are normalized, $\Sigma_1 = R_1$ in our case. Then we have $\rho_{12}^{(1)} = \rho_{34}^{(1)} = a := -\sqrt{1-\alpha/2}$, $\rho_{13}^{(1)} = 1/2$, $\rho_{24}^{(1)} = b := (1+\alpha)/(2-\alpha)$ and $\rho_{14}^{(1)} = \rho_{23}^{(1)} = -(1+\alpha)/\sqrt{2(2-\alpha)} = ab$. Under these correlations, we can obtain an upper bound for the orthant probability $\Phi(\mathbf{0}, R_1) := \Pr(z_1^{(1)} < 0, z_2^{(1)} < 0, z_3^{(1)} < 0, z_4^{(1)} < 0)$ from another orthant probability $\Phi(\mathbf{0}, R')$ with correlation matrix $R' = (\rho'_{kl})$ such that $\rho'_{13} = \rho'_{31} = b$ and $\rho'_{kl} = \rho_{kl}^{(1)}$ otherwise. To see that R' is indeed a correlation matrix, we find the Cholesky decomposition of the matrix $R = (R_{ij})_{1\le i,j\le 4}$ with $R_{ij} = \rho_{ij}^{(1)}$ for $(i,j) \neq (1,3)$ and $(3,1)$ and $R_{13} = R_{31} = \rho_{13}$ for $1/2 \le \rho_{13} \le b$. We can write $R = AA^T$ for

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a & \sqrt{1-a^2} & 0 & 0 \\ \rho_{13} & \frac{a(b-\rho_{13})}{\sqrt{1-a^2}} & \sqrt{1-\rho_{13}^2 - \frac{a^2(b-\rho_{13})^2}{1-a^2}} & 0 \\ ab & b\sqrt{1-a^2} & \frac{a(1-b^2)}{\sqrt{1-\rho_{13}^2-\frac{a^2(b-\rho_{13})^2}{1-a^2}}} & \sqrt{(1-b^2)(1-\frac{a^2(1-b^2)(1-a^2)}{(1-a^2)(1-\rho_{13}^2)-a^2(b-\rho_{13})^2})} \end{pmatrix}.$$

It is straightforward to check that A is a real matrix for $\rho_{13} = b$ and thus by the fact that the class of correlation matrices is convex, R is a correlation matrix for any

$1/2 \le \rho_{13} \le b$. If $\mathbf{w} = (w_1, w_2, w_3, w_4)$ have i.i.d standard normal components, then $A\mathbf{w}^T$ has distribution $N(\mathbf{0}, R)$. For a vector v, let $v < 0$ mean that all coordinates $v_j < 0$. The spherical polyhedron determined by $A\mathbf{w}^T < 0$ in the unit sphere has its boundary vary continuously as ρ_{13} varies between $1/2$ and b. Therefore $\Phi(0, R)$, being the volume of the polyhedron under the normalized orthogonally invariant measure on the sphere, is a continuous function of ρ_{13} for $1/2 \le \rho_{13} \le b$. Note that for $1/2 < \rho_{13} \le b$ and $n \ge 4$, the correlation matrix R satisfies $\det R = (1 - b^2)[(1-a^2)^2 - (\rho_{13} - a^2 b)^2] > 0$. Plackett (1954, (3)) showed that for a nonsingular correlation matrix R and multivariate normal density $\phi(z, R)$ with all variates having unit variance,

$$\frac{\partial}{\partial \rho_{kl}} \phi(z, R) = \frac{\partial^2}{\partial z_k \partial z_l} \phi(z, R). \tag{1.42}$$

Therefore $\Phi(0, R)$ is a smooth function of ρ_{13} for $1/2 < \rho_{13} \le b$. Hence $\Phi(0, R_1)$ and $\Phi(0, R')$ are connected by the following:

$$\Phi(0, R_1) = \Phi(0, R') - \int_{1/2}^{\frac{1+\alpha}{2-\alpha}} \frac{\partial \Phi(\mathbf{0}, R)}{\partial \rho_{13}} d\rho_{13}. \tag{1.43}$$

R' has the property that $\rho'_{12} = \rho'_{34} = a$, $\rho'_{13} = \rho'_{24} = b$, $\rho'_{14} = \rho'_{23} = ab$. Cheng (1969, (2.18)) calculated $\Phi(\mathbf{0}, R')$ to be

$$\begin{aligned} \Phi(\mathbf{0}, R') &= \frac{1}{16} + \frac{1}{4\pi}(\arcsin a + \arcsin b + \arcsin ab) \\ &\quad + \frac{1}{4\pi^2}\left((\arcsin a)^2 + (\arcsin b)^2 - (\arcsin ab)^2\right). \end{aligned} \tag{1.44}$$

Using the series expansion

$$\arcsin(x+h) = \arcsin x + \frac{h}{\sqrt{1-x^2}} + \frac{xh^2}{2(1-x^2)^{3/2}} + \frac{(1+2x^2)h^3}{6(1-x^2)^{5/2}} + O(h^4) \tag{1.45}$$

for $|x|<1$ and $|h|<1-|x|$ we have

$$\arcsin b = \arcsin\frac{1}{2} + \frac{2}{\sqrt{3}}\frac{3\alpha}{2(2-\alpha)} + O\left(\left(\frac{3\alpha}{2(2-\alpha)}\right)^2\right) = \frac{\pi}{6} + \frac{\sqrt{3}\alpha}{2} + O(\alpha^2),$$

and

$$\begin{aligned}\arcsin ab &= -\arcsin\frac{1+\alpha}{\sqrt{2(2-\alpha)}}\\ &= -\arcsin\left(\frac{1}{2}+\frac{5}{8}\alpha+O(\alpha^2)\right)\\ &= -\frac{\pi}{6}-\frac{5\sqrt{3}}{12}\alpha+O(\alpha^2).\end{aligned}$$

We also have

$$\begin{aligned}\arcsin a &= -\arcsin\sqrt{1-\frac{\alpha}{2}}\\ &= -\frac{\pi}{2}+\arcsin\sqrt{\frac{\alpha}{2}},\end{aligned}$$

and thus

$$(\arcsin a)^2 = \frac{\pi^2}{4} - \pi\arcsin\sqrt{\frac{\alpha}{2}} + \left(\arcsin\sqrt{\frac{\alpha}{2}}\right)^2. \tag{1.46}$$

Therefore (1.44) gives

$$\begin{aligned}\Phi(\mathbf{0},R') = \frac{1}{16} + \frac{1}{4\pi}\left(-\frac{\pi}{2}+\arcsin\sqrt{\frac{\alpha}{2}}+\arcsin b+\arcsin ab\right) \qquad (1.47)\\ +\frac{1}{4\pi^2}\left(\frac{\pi^2}{4}-\pi\arcsin\sqrt{\frac{\alpha}{2}}+\left(\arcsin\sqrt{\frac{\alpha}{2}}\right)^2+(\arcsin b)^2-(\arcsin ab)^2\right)\\ = \frac{1}{4\pi}(\arcsin b+\arcsin ab)\left(1+\frac{1}{\pi}(\arcsin b-\arcsin ab)\right)+\frac{1}{4\pi^2}\left(\arcsin\sqrt{\frac{\alpha}{2}}\right)^2\\ = \frac{1}{4\pi}\left(\frac{\sqrt{3}}{12}\alpha+O(\alpha^2)\right)\left(1+\frac{1}{\pi}\left(\frac{\pi}{3}+\frac{11\sqrt{3}}{12}\alpha+O(\alpha^2)\right)\right)+\frac{\alpha}{8\pi^2}+O(\alpha^2)\\ = \left(\frac{\sqrt{3}}{36\pi}+\frac{1}{8\pi^2}\right)\alpha+O(\alpha^2).\end{aligned}$$

To bound $\int_{1/2}^{\frac{1+\alpha}{2-\alpha}} \frac{\partial\Phi(\mathbf{0},R)}{\partial\rho_{13}}d\rho_{13}$, recall Plackett's (1954, (4), (5), (6)) reduction formula for any nonsingular correlation matrix R:

$$\frac{\partial\Phi(\mathbf{0},R)}{\partial\rho_{34}} = \phi(\mathbf{0},\rho_{34})\Phi(\mathbf{0},C_{\overline{12}}), \tag{1.48}$$

where $\phi(\mathbf{0},\rho_{34})$ is the marginal density of z_3, z_4 at $(0,0)$ and $\Phi(\mathbf{0},C_{\overline{12}})$ is the conditional probability that z_1, z_2 are less than 0 given $z_3 = z_4 = 0$. Here $C_{\overline{12}}$ is the covariance matrix of the conditional distribution of z_1, z_2 given $z_3 = z_4 = 0$. More specifically if we write the original covariance matrix as Σ as a block matrix of 2 by 2 submatrices:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

then

$$C_{\overline{12}} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \tag{1.49}$$

We will use this formula for $1/2 < \rho_{13} \le (1+\alpha)/(2-\alpha)$ where $\det R > 0$. In our case

$$\phi(\mathbf{0},\rho_{13}) = \frac{1}{2\pi\sqrt{1-\rho_{13}^2}},$$

and

$$C_{\overline{24}} = \begin{pmatrix} 1 - \frac{a^2(1-2\rho_{13}b+b^2)}{1-\rho_{13}^2} & b - \frac{a^2(2b-\rho_{13}-\rho_{13}b^2)}{1-\rho_{13}^2} \\ b - \frac{a^2(2b-\rho_{13}-\rho_{13}b^2)}{1-\rho_{13}^2} & 1 - \frac{a^2(1-2\rho_{13}b+b^2)}{1-\rho_{13}^2} \end{pmatrix}.$$

Consequently using (1.18) we have

$$\begin{aligned} \Phi(\mathbf{0},C_{\overline{24}}) &= \frac{1}{4} + \frac{1}{2\pi}\arcsin\left(\frac{b - \frac{a^2(2b-\rho_{13}-\rho_{13}b^2)}{1-\rho_{13}^2}}{1 - \frac{a^2(1-2\rho_{13}b+b^2)}{1-\rho_{13}^2}}\right) \\ &= \frac{1}{4} + \frac{1}{2\pi}\arcsin\left(b + \frac{(1-b^2)a^2(\rho_{13}-b)}{-(\rho_{13}-a^2b)^2 + (a^2b^2-1)(a^2-1)}\right). \end{aligned}$$

Note that $1/2 < a^2b = (1+\alpha)/2 < b$ and so for $\frac{1}{2} < \rho_{13} \le b$, $-(\rho_{13}-a^2b)^2 + (a^2b^2-1)(a^2-1) \ge \min\{-(b-a^2b)^2+(a^2b^2-1)(a^2-1), -(1/2-a^2b)^2+(a^2b^2-1)(a^2-1)\} > 0$. Note that the term $b + \frac{(1-b^2)a^2(\rho_{13}-b)}{-(\rho_{13}-a^2b)^2+(a^2b^2-1)(a^2-1)}$ is monotonically increasing with ρ_{13}

for $\frac{1}{2} < \rho_{13} \leq b$ and it is equal to -1 for $\rho_{13} = \frac{1}{2}$ and b for $\rho_{13} = b$. We settle for the trivial bound:

$$\begin{aligned}
&\int_{1/2}^{\frac{1+\alpha}{2-\alpha}} \frac{\partial \Phi(\mathbf{0}, R)}{\partial \rho_{13}} d\rho_{13} \\
\geq\ & \int_{1/2}^{\frac{1+\alpha}{2-\alpha}} \frac{1}{2\pi\sqrt{1-\rho_{13}^2}} \left(\frac{1}{4} + \frac{1}{2\pi}\arcsin(-1)\right) du \\
=\ & 0.
\end{aligned}$$

Combining with (1.43) and (1.47) we have

$$\Phi(0, R_1) \leq \Phi(0, R') = \left(\frac{\sqrt{3}}{36\pi} + \frac{1}{8\pi^2}\right)\alpha + O(\alpha^2). \tag{1.50}$$

Next we bound $\Pr(\delta_{i-1} > 0, \widehat{\delta}_{i-1} < 0, \delta_i > 0, \widehat{\delta}_i < 0)$. Let us write $z_1^{(2)} = -\delta_{i-1}/(\sqrt{2}\sigma)$, $z_2^{(2)} = \widehat{\delta}_{i-1}/(\sqrt{2-\alpha}\sigma)$, $z_3^{(2)} = -\delta_i/(\sqrt{2}\sigma)$, $z_4^{(2)} = \widehat{\delta}_i/(\sqrt{2-\alpha}\sigma)$ and let $R_2 = (\rho_{kl}^{(2)})$ be the correlation matrix of the $z_j^{(2)}$. We have $\rho_{12}^{(2)} = \rho_{34}^{(2)} = -\sqrt{1-\alpha/2}$, $\rho_{14}^{(2)} = \rho_{23}^{(2)} = (1+\alpha)/\sqrt{2(2-\alpha)}$, $\rho_{13}^{(2)} = -1/2$ and $\rho_{24}^{(2)} = -(1+\alpha)/(2-\alpha)$. As before let us consider another orthant probability $\Phi(\mathbf{0}, R'')$ with correlation matrix $R'' = (\rho''_{kl})$ such that $\rho''_{13} = \rho''_{31} = -(1+\alpha)/(2-\alpha)$ and $\rho''_{kl} = \rho_{kl}^{(2)}$ otherwise.

To see that R'' is indeed a correlation matrix, we can again use the Cholesky decomposition of the matrix $R = (R_{ij})_{1\leq i,j\leq 4}$ defined as: $R_{ij} = \rho_{ij}^{(2)}$ for $(i,j) \neq (1,3)$ and $(3,1)$ and $R_{13} = R_{31} = \rho_{13}$ for $-b \leq \rho_{13} \leq -1/2$. The lower triangular matrix in the Cholesky decomposition of R is obtained by replacing b by $-b$ in A. Arguments similar to those following the expression for A show that R is a correlation matrix for $-b \leq \rho_{13} \leq -1/2$ and $\Phi(0, R)$ is continuous on $[-b, -1/2]$. The determinant of R is $\det R = (1-b^2)[(1-a^2)^2 - (\rho_{13} + a^2 b)^2] > 0$ for $-b \leq \rho_{13} < -1/2$, so (1.42) holds, which implies that $\Phi(0, R)$ is a smooth function of ρ_{13} for $\rho_{13} \in [-b, -1/2)$. Thus an analogue of (1.43) holds, namely

$$\Phi(0, R_2) = \Phi(0, R'') + \int_{-\frac{1+\alpha}{2-\alpha}}^{-1/2} \frac{\partial \Phi(\mathbf{0}, R)}{\partial \rho_{13}} d\rho_{13}.$$

R'' has the property that $\rho''_{12} = \rho''_{34} = a$, $\rho''_{13} = \rho''_{24} = -b$, $\rho''_{14} = \rho''_{23} = -ab$ and the orthant probability is therefore, similarly to (1.47) but with b replaced by $-b$, and using (1.46),

$$\begin{aligned}
\Phi(\mathbf{0}, R'') &= \frac{1}{16} + \frac{1}{4\pi}\left(-\frac{\pi}{2} + \arcsin\sqrt{\frac{\alpha}{2}} - \arcsin b - \arcsin ab\right) \\
&+ \frac{1}{4\pi^2}\left(\frac{\pi^2}{4} - \pi\arcsin\sqrt{\frac{\alpha}{2}} + \left(\arcsin\sqrt{\frac{\alpha}{2}}\right)^2 + (\arcsin b)^2 - (\arcsin ab)^2\right) \\
&= -\frac{1}{4\pi}(\arcsin b + \arcsin ab)\left(1 - \frac{1}{\pi}(\arcsin b - \arcsin ab)\right) + \frac{1}{4\pi^2}\left(\arcsin\sqrt{\frac{\alpha}{2}}\right)^2 \\
&= -\frac{1}{4\pi}\left(\frac{\sqrt{3}}{12}\alpha + O(\alpha^2)\right)\left(1 - \frac{1}{\pi}\left(\frac{\pi}{3} + \frac{11\sqrt{3}}{12}\alpha + O(\alpha^2)\right)\right) + \frac{\alpha}{8\pi^2} + O(\alpha^2) \\
&= \left(-\frac{\sqrt{3}}{72\pi} + \frac{1}{8\pi^2}\right)\alpha + O(\alpha^2).
\end{aligned}$$

Using (1.48) we have

$$\begin{aligned}
\int_{-\frac{1+\alpha}{2-\alpha}}^{-1/2} \frac{\partial \Phi(\mathbf{0}, R)}{\partial \rho_{13}} d\rho_{13} &= \int_{-\frac{1+\alpha}{2-\alpha}}^{-1/2} \phi(\mathbf{0}, \rho_{13})\Phi(\mathbf{0}, C_{\overline{24}}) d\rho_{13} \\
&\le \int_{-\frac{1+\alpha}{2-\alpha}}^{-1/2} \frac{1}{2\pi\sqrt{1-\rho_{13}^2}} d\rho_{13} \\
&= \frac{1}{2\pi}\left(\arcsin b - \arcsin\frac{1}{2}\right) \\
&= \frac{1}{2\pi}\left(\frac{\sqrt{3}\alpha}{2}\right) + O(\alpha^2).
\end{aligned}$$

As in (1.43) we have

$$\begin{aligned}
& \Pr(\delta_{i-1} > 0, \widehat{\delta}_{i-1} < 0, \delta_i > 0, \widehat{\delta}_i < 0) \qquad (1.51) \\
&= \Phi(\mathbf{0}, R'') + \int_{-\frac{1+\alpha}{2-\alpha}}^{-1/2} \frac{\partial \Phi(\mathbf{0}, R)}{\partial \rho_{13}} d\rho_{13} \\
&\le \left(-\frac{\sqrt{3}}{72\pi} + \frac{1}{8\pi^2}\right)\alpha + \frac{1}{2\pi}\left(\frac{\sqrt{3}\alpha}{2}\right) + O(\alpha^2) \\
&= \left(\frac{17\sqrt{3}}{72\pi} + \frac{1}{8\pi^2}\right)\alpha + O(\alpha^2).
\end{aligned}$$

By symmetry we can conclude from (1.50) and (1.51) that

$$\begin{aligned}\Pr(B_{i-1} \cap B_i) &\leq 2\left(\left(\frac{\sqrt{3}}{36\pi} + \frac{1}{8\pi^2}\right)\alpha + \left(\frac{17\sqrt{3}}{72\pi} + \frac{1}{8\pi^2}\right)\alpha\right) + O(\alpha^2) \quad (1.52)\\ &= \left(\frac{19\sqrt{3}}{36\pi} + \frac{1}{2\pi^2}\right)\alpha + O(\alpha^2).\end{aligned}$$

Next we bound $\Pr(\delta_i > 0, \widehat{\delta}_i < 0, \delta_j < 0, \widehat{\delta}_j > 0)$ for $|i-j| > 1$. Let us write $z_1^{(3)} = -\delta_i/(\sqrt{2}\sigma)$, $z_2^{(3)} = \widehat{\delta}_{i-1}/(\sqrt{2-\alpha}\sigma)$, $z_3^{(3)} = \delta_j/(\sqrt{2}\sigma)$, $z_4^{(3)} = -\widehat{\delta}_j/(\sqrt{2-\alpha}\sigma)$ and let $R_3 = (\rho_{kl}^{(3)})$ be the correlation matrix. We have $\rho_{12}^{(3)} = \rho_{34}^{(3)} = -\sqrt{1-\alpha/2}$, $\rho_{14}^{(3)} = \rho_{23}^{(3)} = -\alpha/\sqrt{2(2-\alpha)}$, $\rho_{13}^{(3)} = 0$ and $\rho_{24}^{(3)} = \alpha/(2-\alpha)$. As before let us consider another orthant probability $\Phi(\mathbf{0}, R''')$ with correlation matrix $R''' = (\rho_{kl}''')$ such that $\rho_{13}''' = \rho_{31}''' = \alpha/(2-\alpha) = c$ and $\rho_{kl}''' = \rho_{kl}^{(3)}$ otherwise.

To see that R''' is indeed a correlation matrix, we can again use the Cholesky decomposition of the matrix $R = (R_{ij})_{1\leq i,j\leq 4}$ defined as: $R_{ij} = \rho_{ij}^{(3)}$ for $(i,j) \neq (1,3)$ and $(3,1)$ and $R_{13} = R_{31} = \rho_{13}$ for $0 \leq \rho_{13} \leq c$. The lower triangular matrix in the Cholesky decomposition of R is obtained from A by putting c in place of b. Similar arguments to those following the expression for A show that R is a correlation matrix for $0 \leq \rho_{13} \leq c$ and $\Phi(0, R)$ is continuous on $[0, c]$. The determinant of R is $\det R = (1-c^2)[(1-a^2)^2 - (\rho_{13} - a^2c)^2] > 0$ for $0 < \rho_{13} \leq c$ so (1.42) holds which implies that $\Phi(0, R)$ is smooth for $\rho_{13} \in (0, c]$. Thus an analogue of (1.43) holds, namely

$$\Phi(0, R_3) = \Phi(0, R''') - \int_0^c \frac{\partial \Phi(\mathbf{0}, R)}{\partial \rho_{13}} d\rho_{13}. \qquad (1.53)$$

R''' has the property that $\rho_{12}''' = \rho_{34}''' = a$, $\rho_{13}''' = \rho_{24}''' = c$, $\rho_{14}''' = \rho_{23}''' = ac$. Note that

$$\arcsin c = \frac{\alpha}{2-\alpha} + O\left(\left(\frac{\alpha}{2-\alpha}\right)^3\right) = \frac{\alpha}{2} + \frac{\alpha^2}{4} + O(\alpha^3),$$

and

$$\arcsin ac = -\frac{\alpha}{\sqrt{2(2-\alpha)}} + O\left(\left(\frac{\alpha}{\sqrt{2(2-\alpha)}}\right)^3\right) = -\frac{\alpha}{2} - \frac{\alpha^2}{8} + O(\alpha^3).$$

Therefore the orthant probability $\Phi(\mathbf{0}, R''')$ is by (1.44) as in (1.47)

$$\begin{aligned}\Phi(\mathbf{0}, R''') &= \frac{1}{16} + \frac{1}{4\pi}\left(-\frac{\pi}{2} + \arcsin\sqrt{\frac{\alpha}{2}} + \arcsin c + \arcsin ac\right) \\ &+ \frac{1}{4\pi^2}\left(\frac{\pi^2}{4} - \pi \arcsin\sqrt{\frac{\alpha}{2}} + (\arcsin\sqrt{\frac{\alpha}{2}})^2 + (\arcsin c)^2 - (\arcsin ac)^2\right) \\ &= \frac{1}{4\pi}(\arcsin c + \arcsin ac)\left(1 + \frac{1}{\pi}(\arcsin c - \arcsin ac)\right) + \frac{1}{4\pi^2}\left(\arcsin\sqrt{\frac{\alpha}{2}}\right)^2 \\ &= \frac{1}{4\pi}\left(\frac{\alpha^2}{8} + O(\alpha^3)\right)(1 + \frac{1}{\pi}(\alpha + \frac{3}{8}\alpha^2 + O(\alpha^3))) + \frac{\alpha}{8\pi^2} + O(\alpha^2) \\ &= \frac{1}{8\pi^2}\alpha + O(\alpha^2).\end{aligned}$$

Using (1.48) we have

$$\int_0^c \frac{\partial\Phi(\mathbf{0}, R)}{\partial\rho_{13}} d\rho_{13} = \int_0^c \phi(\mathbf{0}, \rho_{13})\Phi(\mathbf{0}, C_{\overline{24}}) d\rho_{13} \geq 0.$$

Thus by (1.53) we have

$$\begin{aligned}&\Pr(\delta_i > 0, \widehat{\delta}_i < 0, \delta_j < 0, \widehat{\delta}_j > 0) \\ &= \Phi(\mathbf{0}, R''') - \int_0^c \frac{\partial\Phi(\mathbf{0}, R)}{\partial\rho_{13}} d\rho_{13} \\ &\leq \frac{1}{8\pi^2}\alpha + O(\alpha^2).\end{aligned} \tag{1.54}$$

Finally we find an upper bound for $\Pr(\delta_i > 0, \widehat{\delta}_i < 0, \delta_j > 0, \widehat{\delta}_j < 0)$ for $|i - j| > 1$. Let $z_1^{(4)} = -\delta_i/(\sqrt{2}\sigma)$, $z_2^{(4)} = \widehat{\delta}_{i-1}/(\sqrt{2-\alpha}\sigma)$, $z_3^{(4)} = -\delta_j/(\sqrt{2}\sigma)$, $z_4^{(4)} = \widehat{\delta}_j/(\sqrt{2-\alpha}\sigma)$ and let $R_4 = (\rho_{kl}^{(4)})$ be the correlation matrix. We have $\rho_{12}^{(4)} = \rho_{34}^{(4)} = -\sqrt{1-\alpha/2}$, $\rho_{14}^{(4)} = \rho_{23}^{(4)} = \alpha/\sqrt{2(2-\alpha)}$, $\rho_{13}^{(4)} = 0$ and $\rho_{24}^{(4)} = -\alpha/(2-\alpha)$. As before let us consider another orthant probability $\Phi(\mathbf{0}, R'''')$ with correlation matrix $R'''' = (\rho_{kl}'''')$ such that $\rho_{13}'''' = \rho_{31}'''' = -\alpha/(2-\alpha) = -c$ and $\rho_{kl}'''' = \rho_{kl}^{(4)}$ otherwise.

To see that R'''' is indeed a correlation matrix, we can again use the Cholesky decomposition of the matrix $R = (R_{ij})_{1\leq i,j\leq 4}$ defined as: $R_{ij} = \rho_{ij}^{(4)}$ for $(i,j) \neq (1,3)$ and $(3,1)$ and $R_{13} = R_{31} = \rho_{13}$ for $-c \leq \rho_{13} \leq 0$. The lower triangular matrix in the Cholesky decomposition of R is obtained from the matrix A by putting $-c$ in place of b. Similar arguments as those following the expression for A show that R is a correlation matrix for $-c \leq \rho_{13} \leq 0$ and $\Phi(0, R)$ is continuous on $[-c, 0]$. The determinant of R is $\det R = (1-c^2)[(1-a^2)^2 - (\rho_{13} + a^2 c)^2] > 0$ for $-c \leq \rho_{13} < 0$ so (1.42) holds which implies that $\Phi(0, R)$ is smooth for $\rho_{13} \in ([-c, 0)$. Thus an analogue of (1.43) holds, namely

$$\Phi(0, R_3) = \Phi(0, R''') + \int_{-c}^{0} \frac{\partial \Phi(\mathbf{0}, R)}{\partial \rho_{13}} d\rho_{13}. \tag{1.55}$$

R'''' has the property that $\rho''''_{12} = \rho''''_{34} = a$, $\rho''''_{13} = \rho''''_{24} = -c$, $\rho''''_{14} = \rho''''_{23} = -ac$. Therefore the orthant probability $\Phi(\mathbf{0}, R'''')$ is, again by (1.44) as in (1.47),

$$\begin{aligned}
\Phi(\mathbf{0}, R'''') &= \frac{1}{16} + \frac{1}{4\pi}\left(-\frac{\pi}{2} + \arcsin\sqrt{\frac{\alpha}{2}} - \arcsin c - \arcsin ac\right) \\
&+ \frac{1}{4\pi^2}\left(\frac{\pi^2}{4} - \pi \arcsin\sqrt{\frac{\alpha}{2}} + \left(\arcsin\sqrt{\frac{\alpha}{2}}\right)^2 + (\arcsin c)^2 - (\arcsin ac)^2\right) \\
&= -\frac{1}{4\pi}(\arcsin c + \arcsin ac)\left(1 - \frac{1}{\pi}(\arcsin c - \arcsin ac)\right) + \frac{1}{4\pi^2}\left(\arcsin\sqrt{\frac{\alpha}{2}}\right)^2 \\
&= -\frac{1}{4\pi}\left(\frac{\alpha^2}{8} + O(\alpha^3)\right)\left(1 - \frac{1}{\pi}\left(\alpha + \frac{3}{8}\alpha^2 + O(\alpha^3)\right)\right) + \frac{\alpha}{8\pi^2} + O(\alpha^2) \\
&= \frac{1}{8\pi^2}\alpha + O(\alpha^2).
\end{aligned}$$

By (1.48),

$$\begin{aligned}
&\int_{-\frac{\alpha}{2-\alpha}}^{0} \frac{\partial \Phi(\mathbf{0}, R)}{\partial \rho_{13}} d\rho_{13} \\
&= \int_{-\frac{\alpha}{2-\alpha}}^{0} \phi(\mathbf{0}, \rho_{13}) \Phi(\mathbf{0}, C_{\overline{24}}) d\rho_{13} \\
&\leq \int_{-\frac{\alpha}{2-\alpha}}^{0} \frac{1}{2\pi\sqrt{1-\rho_{13}^2}} d\rho_{13} \\
&= \frac{1}{2\pi} \arcsin\frac{\alpha}{2-\alpha} \\
&= \frac{1}{2\pi}\left(\frac{\alpha}{2} + \frac{\alpha^2}{4}\right) + O(\alpha^3).
\end{aligned}$$

Hence by (1.55) we have

$$\begin{aligned}
& \Pr(\delta_i > 0, \widehat{\delta}_i < 0, \delta_j > 0, \widehat{\delta}_j < 0) \\
= \ & \Phi(\mathbf{0}, R'''') + \int_{-\frac{\alpha}{2-\alpha}}^{0} \frac{\partial \Phi(\mathbf{0}, R)}{\partial \rho_{13}} d\rho_{13} \\
\leq \ & \left(\frac{1}{8\pi^2} + \frac{1}{4\pi}\right) \alpha + O(\alpha^2).
\end{aligned} \tag{1.56}$$

By symmetry we can conclude that for $|i-j| > 1$,

$$\begin{aligned}
\Pr(B_i \cap B_j) & \leq 2\left(\frac{1}{8\pi^2}\alpha + \left(\frac{1}{8\pi^2} + \frac{1}{4\pi}\right)\alpha\right) + O(\alpha^2) \\
& = \left(\frac{1}{2\pi^2} + \frac{1}{2\pi}\right)\alpha + O(\alpha^2).
\end{aligned} \tag{1.57}$$

Recall that Q_1 and Q_2 were defined before (1.38) and recall (1.37). We then have

$$\begin{aligned}
Q_2 & \leq (n-2)\left(\frac{19\sqrt{3}}{36\pi} + \frac{1}{2\pi^2}\right)\alpha + \frac{(n-2)(n-3)}{2}\left(\frac{1}{2\pi^2} + \frac{1}{2\pi}\right)\alpha + \\
& \quad \frac{(n-2)(n-1)}{2} O(\alpha^2) \\
& = \left(\frac{19\sqrt{3}}{36\pi} + \frac{1}{2\pi^2}\right)\frac{12}{n^2} + \left(\frac{1}{2\pi^2} + \frac{1}{2\pi}\right)\left(\frac{6}{n} - \frac{30}{n^2}\right) + O(n^{-3}) \\
& = \left(\frac{3}{\pi^2} + \frac{3}{\pi}\right)\frac{1}{n} + \left(\frac{19\sqrt{3}}{3\pi} - \frac{15}{\pi} - \frac{9}{\pi^2}\right)\frac{1}{n^2} + O(n^{-3}).
\end{aligned}$$

By (1.41), (1.45), and (1.37),

$$Q_1 = \frac{n-1}{\pi} \arcsin\sqrt{\frac{\alpha}{2}} = \frac{\sqrt{6}}{\pi}\left(\frac{1}{\sqrt{n}} - \frac{1}{n\sqrt{n}}\right) + o(n^{-2}).$$

Therefore by (1.39)

$$\begin{aligned}
& \Pr(\delta_i \widehat{\delta}_i < 0 \text{ for exactly one index } 2 \leq i \leq n) \\
\geq \ & Q_1 - 2Q_2 \\
= \ & \frac{n-1}{\pi} \arcsin\sqrt{\frac{\alpha}{2}} - \left(\frac{6}{\pi^2} + \frac{6}{\pi}\right)\frac{1}{n} + O(n^{-2}).
\end{aligned}$$

Since $\Pr(\delta_2\widehat{\delta}_2 < 0 \text{ and } \delta_i\widehat{\delta}_i > 0\ \forall i \neq 2) \leq \Pr(\delta_2\widehat{\delta}_2 < 0) = \frac{1}{\pi}\arcsin\sqrt{\frac{\alpha}{2}}$ by (1.40) and similarly $\Pr(\delta_n\widehat{\delta}_n < 0 \text{ and } \delta_i\widehat{\delta}_i > 0\ \forall i \neq n) \leq \frac{1}{\pi}\arcsin\sqrt{\frac{\alpha}{2}}$ we have

$$\begin{aligned} & \Pr(\delta_j\widehat{\delta}_j < 0 \text{ for some } j = 3, ..., n-1 \text{ and } \delta_i\widehat{\delta}_i > 0\ \forall i \neq j) \\ \geq\ & \frac{n-3}{\pi}\arcsin\sqrt{\frac{\alpha}{2}} - \left(\frac{6}{\pi^2} + \frac{6}{\pi}\right)\frac{1}{n} + O(n^{-2}). \end{aligned} \tag{1.58}$$

We mentioned before that $\delta_i\widehat{\delta}_i < 0$ for exactly one index i does not necessarily imply that $|T_n - \widehat{T}_n| > 0$. If $\delta_i\widehat{\delta}_i < 0$ for exactly one index $2 < i < n$ and $\delta_{i-1}\delta_{i+1} < 0$ then both errors and residuals will have exactly one turning point at i or $i-1$ and if the turning point occurs at i $(i-1)$ for the errors it will occur at $i-1$ (i) for the residuals. Consequently we will have $|T_n - \widehat{T}_n| = 0$. We now calculate an upper bound for the probability of such cases. First we notice the following trivial inequality:

$$\begin{aligned} & \Pr(\delta_2 > 0, \delta_3 < 0, \widehat{\delta}_3 > 0, \delta_4 < 0 \text{ and } \delta_i\widehat{\delta}_i > 0 \text{ for all } i \neq 3) \\ \leq\ & \Pr(\delta_2 > 0, \delta_3 < 0, \widehat{\delta}_3 > 0, \delta_4 < 0). \end{aligned}$$

We now show $\Pr(\delta_2 > 0, \delta_3 < 0, \widehat{\delta}_3 > 0, \delta_4 < 0) \leq \frac{1}{2\pi}(\frac{1}{4} + \frac{1}{2\pi}\arcsin\frac{1}{3})\arcsin\sqrt{\alpha/2} + O(\alpha^{3/2})$. Let us set $z_1 = -\delta_2/(\sqrt{2}\sigma)$, $z_2 = \delta_3/(\sqrt{2}\sigma)$, $z_3 = -\widehat{\delta}_3/(\sqrt{2-\alpha}\sigma)$, $z_4 = \delta_4/(\sqrt{2}\sigma)$, so z_i are normalized. Let $R_c = (\rho_{kl})$ be the correlation matrix. Then $\rho_{12} = 1/2$, $\rho_{13} = -(1+\alpha)/\sqrt{2(2-\alpha)}$, $\rho_{14} = 0$, $\rho_{23} = -\sqrt{1-\alpha/2}$, $\rho_{24} = -1/2$, $\rho_{34} = (1+\alpha)/\sqrt{2(2-\alpha)}$. Let $R(u)$ be the matrix with entries as in R_c but with variable u in place of α. From (1.37), it is easy to see that $0 < \alpha \leq 1/5$ for all $n \geq 4$. For $0 \leq u \leq 1/5$, the Cholesky decomposition of $R(u) = A(u)A(u)^T$ gives

$$A(u) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 \\ -\frac{1+u}{\sqrt{2(2-u)}} & \frac{\sqrt{3}(u-1)}{\sqrt{2(2-u)}} & \sqrt{\frac{u(1-2u)}{2-u}} & 0 \\ 0 & -\frac{1}{\sqrt{3}} & \sqrt{\frac{2u}{1-2u}} & \sqrt{\frac{2}{3} - \frac{2u}{1-2u}} \end{pmatrix}.$$

$A(u)$ is a real matrix for $0 \leq u \leq 1/5$ and therefore $R(u)$ is a correlation matrix for

$0 \le u \le 1/5$. As u approaches zero, $R(u)$ will approach the matrix R_c^0 entry-wise where

$$R_c^0 := \begin{pmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & 1 & -1 & -\frac{1}{2} \\ -\frac{1}{2} & -1 & 1 & \frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}.$$

If $\mathbf{w} = (w_1, w_2, w_3, w_4)$ have i.i.d standard normal components, then $A(u)\mathbf{w}^T$ has distribution $N(\mathbf{0}, R(u))$. We can see the spherical polyhedron in the unit sphere determined by $A\mathbf{w}^T < 0$ has its boundary vary continuously with $0 \le u \le \frac{1}{5}$. Therefore $\Phi(0, R(u))$, being the volume of the polyhedron under the normalized orthogonally invariant measure on the sphere, is a continuous function of u for $0 \le u \le \frac{1}{5}$. Furthermore $\det R(u) = \frac{u-5u^2}{2(2-u)} > 0$ for $0 < u < 1/5$. Therefore $\Phi(0, R(u))$ is a smooth function of u for $0 < u < 1/5$ and we have

$$\Phi(0, R_c) = \Phi(0, R_c^0) + \int_0^\alpha \frac{d}{du}\Phi(0, R(u))du = \int_0^\alpha \frac{d}{du}\Phi(0, R(u))du.$$

The chain rule implies

$$\frac{d}{du}\Phi(0, R(u)) = \frac{\partial\Phi(0, R)}{\rho_{13}}\frac{d\rho_{13}}{du} + \frac{\partial\Phi(0, R)}{\rho_{23}}\frac{d\rho_{23}}{du} + \frac{\partial\Phi(0, R)}{\rho_{34}}\frac{d\rho_{34}}{du}.$$

Let us set $q = (1+u)/\sqrt{2(2-u)}$ and $a = -\sqrt{1-u/2}$. Using (1.49), we can calculate the covariance matrix of the conditional distribution of z_1, z_4 given $z_2 = z_3 = 0$ to be

$$C_{\overline{14}} = \begin{pmatrix} 1 - \frac{1/4+aq+q^2}{1-a^2} & \frac{1/4+aq+q^2}{1-a^2} \\ \frac{1/4+aq+q^2}{1-a^2} & 1 - \frac{1/4+aq+q^2}{1-a^2} \end{pmatrix}.$$

Similarly we have

$$C_{\overline{24}} = \begin{pmatrix} 1 - \frac{1/4+aq+a^2}{1-q^2} & -\frac{1}{2} - \frac{2aq+q^2}{2(1-q^2)} \\ -\frac{1}{2} - \frac{2aq+q^2}{2(1-q^2)} & \frac{1-2q^2}{1-q^2} \end{pmatrix},$$

and

$$C_{\overline{12}} = \begin{pmatrix} \frac{1-2q^2}{1-q^2} & \frac{1}{2} + \frac{2aq+q^2}{2(1-q^2)} \\ \frac{1}{2} + \frac{2aq+q^2}{2(1-q^2)} & 1 - \frac{1/4+aq+a^2}{1-q^2} \end{pmatrix}.$$

Using Plackett's reduction formula (1.48) and (1.45) we have

$$\begin{aligned}
\frac{\partial\Phi(0,R)}{\rho_{23}} &= \phi(0,\rho_{23})\Phi(\mathbf{0},C_{\overline{14}}) \\
&= \frac{1}{2\pi\sqrt{1-a^2}}\left(\frac{1}{4}+\frac{1}{2\pi}\arcsin\left(\left(\frac{1/4+aq+q^2}{1-a^2}\right)/\left(1-\frac{1/4+aq+q^2}{1-a^2}\right)\right)\right) \\
&= \frac{1}{2\pi\sqrt{u/2}}\left(\frac{1}{4}+\frac{1}{2\pi}\arcsin\frac{1+4u}{3-6u}\right) \\
&= \frac{1}{2\pi\sqrt{u/2}}\left(\frac{1}{4}+\frac{1}{2\pi}\left(\arcsin\left(\frac{1}{3}\right)+\frac{3u}{\sqrt{2}}+O(u^2)\right)\right),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\Phi(0,R)}{\rho_{13}} &= \phi(0,\rho_{13})\Phi(\mathbf{0},C_{\overline{24}}) \\
&= \frac{1}{2\pi\sqrt{1-q^2}}\left(\frac{1}{4}+\frac{1}{2\pi}\arcsin\frac{-\frac{1}{2}-\frac{2aq+q^2}{2(1-q^2)}}{\frac{\sqrt{(1-2q^2)(3/4-a^2-aq-q^2)}}{1-q^2}}\right) \\
&= \frac{1}{2\pi\sqrt{1-q^2}}\left(\frac{1}{4}+\frac{1}{2\pi}\arcsin\frac{(2-u)\sqrt{u}}{\sqrt{(1-3u-u^2)(3-6u)}}\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial\Phi(0,R)}{\rho_{34}} &= \phi(0,\rho_{34})\Phi(\mathbf{0},C_{\overline{12}}) \\
&= \frac{1}{2\pi\sqrt{1-q^2}}\left(\frac{1}{4}-\frac{1}{2\pi}\arcsin\frac{-\frac{1}{2}-\frac{2aq+q^2}{2(1-q^2)}}{\frac{\sqrt{(1-2q^2)(3/4-a^2-aq-q^2)}}{1-q^2}}\right) \\
&= \frac{1}{2\pi\sqrt{1-q^2}}\left(\frac{1}{4}-\frac{1}{2\pi}\arcsin\frac{(2-u)\sqrt{u}}{\sqrt{(1-3u-u^2)(3-6u)}}\right).
\end{aligned}$$

We also have

$$\frac{d\rho_{23}}{du} = \frac{1}{4\sqrt{1-u/2}},$$
$$\frac{d\rho_{13}}{du} = -\frac{d\rho_{34}}{du} = -\left((1+u)(2(2-u))^{-3/2} + (2(2-u))^{-1/2}\right).$$

Therefore

$$\begin{aligned}
&\frac{\partial\Phi(0,R)}{\rho_{23}}\frac{d\rho_{23}}{du} \\
=\ & \frac{1}{2\pi\sqrt{u/2}}\left(\frac{1}{4}+\frac{1}{2\pi}\left(\arcsin\left(\frac{1}{3}\right)+\frac{3u}{\sqrt{2}}+O(u^2)\right)\right)\frac{1}{4\sqrt{1-u/2}} \\
=\ & \frac{1}{2\pi\sqrt{u/2}}\left(\frac{1}{4\sqrt{1-u/2}}\right)\left(\frac{1}{4}+\frac{1}{2\pi}\arcsin\frac{1}{3}\right)+\frac{3\sqrt{u}}{16\pi^2}+O(u\sqrt{u}).
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial\Phi(0,R)}{\rho_{13}}\frac{d\rho_{13}}{du}+\frac{\partial\Phi(0,R)}{\rho_{34}}\frac{d\rho_{34}}{du} \\
=\ & -\frac{((1+u)(2(2-u))^{-3/2}+(2(2-u))^{-1/2})}{2\pi^2\sqrt{1-q^2}}\arcsin\frac{(2-u)\sqrt{u}}{\sqrt{(1-3u-u^2)(3-6u)}} \\
=\ & -\frac{5\sqrt{u}}{12\pi^2}+O(u\sqrt{u}).
\end{aligned}$$

Consequently

$$\begin{aligned}
\Phi(0,R_c) &= \int_0^\alpha \frac{d}{du}\Phi(0,R(u))du \\
&= \int_0^\alpha \left[\frac{1}{2\pi\sqrt{u/2}}\left(\frac{1}{4\sqrt{1-u/2}}\right)\left(\frac{1}{4}+\frac{1}{2\pi}\arcsin\frac{1}{3}\right)+\frac{3\sqrt{u}}{16\pi^2}-\frac{5\sqrt{u}}{12\pi^2}\right]du \\
&\quad +O(\alpha^{5/2}) \\
&= \frac{1}{2\pi}\left(\frac{1}{4}+\frac{1}{2\pi}\arcsin\frac{1}{3}\right)\arcsin\sqrt{\frac{\alpha}{2}}-\frac{11}{72\pi^2}\alpha^{\frac{3}{2}}+O(\alpha^{5/2}).
\end{aligned}$$

Next note that $\Pr(\delta_4 > 0, \delta_3 < 0, \widehat{\delta}_3 > 0, \delta_2 < 0) = \Phi(0, R_c)$ due to the fact that if we let $z_1 = -\delta_4/(\sqrt{2}\sigma)$, $z_2 = \delta_3/(\sqrt{2}\sigma)$, $z_3 = -\widehat{\delta}_3/(\sqrt{2-\alpha}\sigma)$,$z_4 = \delta_2/(\sqrt{2}\sigma)$, the correlation matrix for z_i is R_c. We also have $\Pr(\delta_2 < 0, \delta_3 > 0, \widehat{\delta}_3 < 0, \delta_4 > 0) =$

$\Pr(\delta_2 > 0, \delta_3 < 0, \widehat{\delta}_3 > 0, \delta_4 < 0)$ and $\Pr(\delta_4 < 0, \delta_3 > 0, \widehat{\delta}_3 < 0, \delta_2 > 0) = \Pr(\delta_4 > 0, \delta_3 < 0, \widehat{\delta}_3 > 0, \delta_2 < 0)$. Therefore $\Pr(\delta_3\widehat{\delta}_3 < 0,\ \delta_2\delta_4 < 0) = 4\Phi(0, R_c)$.

Note that in the calculation above if we replace δ_3 with δ_j, δ_2 with δ_{j-1} and δ_4 with δ_{j+1} for any $2 < j < n$ we will obtain the same correlation matrices and thus the same quadrivariate probabilities. Consequently $\Pr(\delta_j\widehat{\delta}_j < 0,\ \delta_{j-1}\delta_{j+1} < 0) = 4\Phi(0, R_c)$ for any $2 < j < n$. Therefore for any $2 < j < n$, by (1.37)

$$\begin{aligned}
& \Pr(\delta_j\widehat{\delta}_j < 0,\ \delta_{j-1}\delta_{j+1} < 0,\ \delta_i\widehat{\delta}_i > 0\ \forall i \neq j) \\
\leq\ & 4\Phi(0, R_c) \\
\leq\ & \frac{2}{\pi}\left(\frac{1}{4} + \frac{1}{2\pi}\arcsin\frac{1}{3}\right)\arcsin\sqrt{\frac{\alpha}{2}} - \frac{11}{18\pi^2}\alpha^{\frac{3}{2}} + O(\alpha^{5/2}) \\
=\ & \frac{2}{\pi}\left(\frac{1}{4} + \frac{1}{2\pi}\arcsin\frac{1}{3}\right)\arcsin\sqrt{\frac{\alpha}{2}} + O(n^{-3})
\end{aligned}$$

and consequently by (1.58)

$$\begin{aligned}
& \Pr(\delta_j\widehat{\delta}_j < 0 \text{ for some } j = 3, ..., n-1 \text{ and } \delta_{j-1}\delta_{j+1} > 0,\ \delta_i\widehat{\delta}_i > 0\ \forall i \neq j) \\
\geq\ & \frac{n-3}{\pi}\arcsin\sqrt{\frac{\alpha}{2}} - \left(\frac{6}{\pi^2} + \frac{6}{\pi}\right)\frac{1}{n} - \frac{2(n-3)}{\pi}\left(\frac{1}{4} + \frac{1}{2\pi}\arcsin\frac{1}{3}\right)\arcsin\sqrt{\frac{\alpha}{2}} + O(n^{-2}) \\
=\ & \frac{n-3}{\pi}\left(\frac{1}{2} - \frac{1}{\pi}\arcsin\frac{1}{3}\right)\arcsin\sqrt{\frac{\alpha}{2}} - \left(\frac{6}{\pi^2} + \frac{6}{\pi}\right)\frac{1}{n} + O(n^{-2}).
\end{aligned}$$

Finally we obtain

$$\begin{aligned}
& E[|T_n - \widehat{T}_n|] \\
\geq\ & 2\Pr(|T_n - \widehat{T}_n| = 2) \\
\geq\ & 2\Pr(\delta_j\widehat{\delta}_j < 0 \text{ for some } j = 3, ..., n-1 \text{ and } \delta_{j-1}\delta_{j+1} > 0,\ \delta_i\widehat{\delta}_i > 0\ \forall i \neq j) \\
\geq\ & 2\left(\frac{n-3}{\pi}\left(\frac{1}{2} - \frac{1}{\pi}\arcsin\frac{1}{3}\right)\arcsin\sqrt{\frac{\alpha}{2}} - \left(\frac{6}{\pi^2} + \frac{6}{\pi}\right)\frac{1}{n}\right) + O(n^{-2}) \\
\geq\ & \frac{2(n-3)}{\pi}\left(\frac{1}{2} - \frac{1}{\pi}\arcsin\frac{1}{3}\right)\sqrt{\frac{\alpha}{2}} + O(n^{-1}) \\
=\ & \frac{2\sqrt{6}}{\pi}\left(\frac{1}{2} - \frac{1}{\pi}\arcsin\frac{1}{3}\right)\frac{1}{\sqrt{n}} + O(n^{-1}).
\end{aligned}$$

□

We have $\Pr(T_n \neq \widehat{T}_n) \leq E(|T_n - \widehat{T}_n|)$, and so for any integer $m \geq 1$,

$$|\Pr(T_n \leq m) - \Pr(\widehat{T}_n \leq m)| \;\leq\; E(|T_n - \widehat{T}_n|). \tag{1.59}$$

For some $\beta > 0$ such as 0.05 or 0.025, let $m(n, \beta)$ be the largest m such that $P(T_n \leq m) \leq \beta$ (one-sided critical value). Let $\widehat{m}(n, \beta)$ be the corresponding critical value for $\widehat{T}_n$. From Tables 1.2, 1.4 and 1.9, we can see that the critical values are the same for T_n and for $\widehat{T}_n$ for equally spaced design points for all $11 \leq n \leq 50$ except for $n = 22, 31$ in the left tail at the 0.01 level. If the design points are fairly equally spaced, by Theorem 6), $E(|T_n - \widehat{T}_n|) \leq O(n^{-0.5})$ as $n \to \infty$ so the critical values will again tend to be the same.

If the design points are reordered from a random normal sample, we have $|\Pr(\widehat{T}_n \leq \alpha) - \Pr(T_n \leq \alpha)| \leq O(n^{-0.5}\sqrt{\log n})$ with probability at least $1 - \frac{1}{n\sqrt{2\pi \log n}} - \frac{1}{n}$ by Theorem 7.

For very small n, however, there are differences in the turning point behavior of ε_j and $\widehat{\varepsilon}_j$. If $n = 3$, then ε_j have no turning point with probability 1/3, whereas $\widehat{\varepsilon}_j$ must have a turning point, as otherwise they would be monotonically increasing or decreasing, contrary to (1.8).

1.3.2 Quadratic regression models

For higher order regression models, we can derive similar bounds as we have for simple linear regression models. In this section, we will derive several asymptotic results for quadratic regression models.

Note that any linear transformation that maps X_i to $aX_i + b$ for all $1 \leq i \leq n$ for any given constants $a \neq 0, b \in \mathbf{R}$ will not change the subspace spanned by $(X_1^i, ..., X_n^i)^T$, $i = 0, 1, 2$, the three columns of $\tilde{X}$, and hence will produce the same projection matrix H. Consequently by (1.5) the residuals will be the same after any such linear transformation. Thus without loss of generality we will assume $X_1 = 0$ apart from assuming $X_1 < X_2 < \cdots < X_n$. To ensure uniqueness of the least-square estimator we also assume $n \geq 3$.

Let us define $U_i = X_i^2$ for $1 \le i \le n$. We will use the following notation along with (1.23):

$$\overline{U} := \frac{1}{n}\sum_{i=1}^{n} U_i, \quad S_U^2 := \sum_{i=1}^{n}(U_i - \overline{U})^2, \tag{1.60}$$

$$\begin{aligned} \mathrm{Cov}(X,U) &:= \sum_{i=1}^{n}(X_i - \overline{X})(U_i - \overline{U}) = \sum_{i=1}^{n}(X_i - \overline{X})U_i \\ &= \sum_{i=1}^{n}(X_i - \overline{X})(U_i - \overline{X}^2) = \sum_{i=1}^{n}(X_i - \overline{X})^2(X_i + \overline{X}) > 0, \end{aligned} \tag{1.61}$$

$$\rho_{XU} := \frac{\mathrm{Cov}(X,U)}{\sqrt{S_X^2 S_U^2}} > 0. \tag{1.62}$$

By the Cauchy-Schwartz inequality

$$\left(\sum_{i=1}^{n}(X_i - \overline{X})(U_i - \overline{U})\right)^2 \le \sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(U_i - \overline{U})^2, \tag{1.63}$$

and equality holds iff $U_i - \overline{U} \equiv C'(X_i - \overline{X})$ for some constant C'. However if $U_i - \overline{U} \equiv C'(X_i - \overline{X})$, then for $i \ge 2$, $U_i - U_{i-1} = C'(X_i - X_{i-1})$ so if $X_i > X_{i-1}$, $X_i + X_{i-1} = (U_i - U_{i-1})/(X_i - X_{i-1}) = C'$ which is impossible because there are at least three different values among the X_i. Hence we have $\rho_{XU} < 1$. Let

$$\alpha_i^{(q)} := \frac{1}{1-\rho_{XU}^2}\left\{\frac{(X_i - X_{i-1})^2}{S_X^2} - 2\rho_{XU}\frac{(X_i - X_{i-1})(U_i - U_{i-1})}{\sqrt{S_X^2 S_U^2}} + \frac{(U_i - U_{i-1})^2}{S_U^2}\right\}. \tag{1.64}$$

Proposition 9. *In quadratic linear regression with ε_j i.i.d. $N(0,\sigma^2)$ and design points $0 = X_1 < \cdots < X_n$, let $\widehat{T}_n$ be the number of turning points in $\widehat{\varepsilon}_j$ for a given n. Then for all $n \ge 5$,*

$$E[|\widehat{T}_n - T_n|] \le \frac{4}{\pi\sqrt{2-\alpha_{\max}}\sqrt{1-\rho_{XU}^2}},$$

where $\alpha_{\max} = \max_i \alpha_i^{(q)}$. In particular, if the design points are equally spaced, then

$\alpha_{\max} = O(1/n^3)$, $\rho_{XU} = \sqrt{\frac{15}{16}} + O(1/n)$ *and*

$$E[|\widehat{T}_n - T_n|] \leq \frac{4\left(4\sqrt{3}+3\sqrt{5}\right)}{\pi\sqrt{2}} \frac{1}{\sqrt{n}} + O(1/n).$$

Proof. . The proof is similar to that of Theorem 4 but H will be different for quadratic regression. It is straightforward to calculate that

$$\begin{aligned} H_{rs} &= \frac{1}{n} + \frac{S_U^2(X_r - \overline{X})(X_s - \overline{X}) + S_X^2(U_r - \overline{U})(U_s - \overline{U})}{S_X^2 S_U^2 - \operatorname{Cov}(X,U)^2} \\ &\quad - \frac{\operatorname{Cov}(X,U)\left((X_r - \overline{X})(U_s - \overline{U}) + (X_s - \overline{X})(U_r - \overline{U})\right)}{S_X^2 S_U^2 - \operatorname{Cov}(X,U)^2} \end{aligned} \tag{1.65}$$

for $1 \leq r, s \leq n$ and for Λ defined by (1.9)

$$\begin{aligned} (\Lambda H \Lambda^T)_{kl} &= \frac{(X_k - X_{k-1})(X_l - X_{l-1})S_U^2 + (U_k - U_{k-1})(U_l - U_{l-1})S_X^2}{S_X^2 S_U^2 - \operatorname{Cov}(X,U)^2} \\ &\quad - \frac{\operatorname{Cov}(X,U)\left((X_k - X_{k-1})(U_l - U_{l-1}) + (X_l - X_{l-1})(U_k - U_{k-1})\right)}{S_X^2 S_U^2 - \operatorname{Cov}(X,U)^2} \end{aligned} \tag{1.66}$$

for $2 \leq k, l \leq n$. Therefore by (1.11) and (1.12), (1.66) and a short calculation implies

$$\operatorname{Var}(\widehat{\delta}_i) = (2 - \alpha_i^{(q)})\sigma^2, \ \operatorname{Cov}(\delta_i, \widehat{\delta}_i) = (2 - \alpha_i^{(q)})\sigma^2. \tag{1.67}$$

We have $\widehat{\delta}_i = v^T\widehat{\varepsilon}$ where $v_i = 1$, $v_{i-1} = -1$, and $v_j = 0$ for $j \neq i-1, i$. By (1.5), $\widehat{\delta}_i = v^T(I - H)\varepsilon$. If $\operatorname{Var}(\widehat{\delta}_i) = 0$, then $v^T(I - H) = 0$ since ε are i.i.d. $N(0, \sigma^2)$. Consequently $v^T = v^T H = v^T M(M^T M)^{-1} M^T$ or $v = M(M^T M)^{-1} M^T v$. Note that $(M^T M)^{-1} M^T v$ is a 3 by 1 vector, say with entries a, b and c. Then $v_j = a + bX_j + cX_j^2 = 0$ for $j \neq i-1, i$, $= 1$ for $j = i$, which implies a, b, and c not all 0, and $= -1$ for $j = i-1$. Since $n \geq 5$ this gives a non-zero quadratic equation with 3 or more distinct roots, a contradiction, so $\operatorname{Var}(\widehat{\delta}_i) > 0$ for $i = 2, \ldots, n$. Moreover, $\operatorname{Cov}(\delta_i, \widehat{\delta}_i) = \operatorname{Var}(\widehat{\delta}_i) > 0$, so Proposition 3 applies. Together with (1.10) we obtain the correlation between δ_i and $\widehat{\delta}_i$:

$$\rho_i = \sqrt{1 - \alpha_i^{(q)}/2}, \tag{1.68}$$

which has the same form as (1.15). In what follows let $\alpha_i = \alpha_i^{(q)}$. By the first equation in (1.67) we have $\alpha_i < 2$. Since $\rho_i \leq 1$, (1.68) implies $\alpha_i \geq 0$. Furthermore since we assume $0 = X_1 < X_2 < \cdots < X_n$, $X_{i-1}^2 < X_i^2$ for all $2 \leq i \leq n$. Consequently by (1.61) and (1.62) we have

$$\begin{aligned}
\alpha_i &= \frac{(X_i - X_{i-1})^2 S_U^2 - 2(X_i - X_{i-1})(U_i - U_{i-1})\mathrm{Cov}(X,U) + (U_i - U_{i-1})^2 S_X^2}{S_X^2 S_U^2 - \mathrm{Cov}(X,U)^2} \\
&\leq \frac{(X_i - X_{i-1})^2 S_U^2 + (U_i - U_{i-1})^2 S_X^2}{S_X^2 S_U^2 - \mathrm{Cov}(X,U)^2} \\
&= \frac{1}{1-\rho_{XU}^2}\left(\frac{(X_i - X_{i-1})^2}{S_X^2} + \frac{(U_i - U_{i-1})^2}{S_U^2}\right). \qquad (1.69)
\end{aligned}$$

Since $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for any $a, b > 0$,

$$\sqrt{\alpha_i} \leq \frac{1}{\sqrt{1-\rho_{XU}^2}}\left(\frac{X_i - X_{i-1}}{\sqrt{S_X^2}} + \frac{U_i - U_{i-1}}{\sqrt{S_U^2}}\right). \qquad (1.70)$$

By (1.17), (1.20)and (1.69)

$$\begin{aligned}
E[|\widehat{T}_n - T_n|] &\leq \sum_{i=2}^{n-1} \frac{\arcsin\sqrt{1-\rho_i^2}}{\pi} + \sum_{i=2}^{n-1} \frac{\arcsin\sqrt{1-\rho_{i+1}^2}}{\pi} \\
&\leq \frac{1}{\pi}\sum_{i=2}^{n-1} \frac{\sqrt{1-\rho_i^2}}{\rho_i} + \frac{1}{\pi}\sum_{i=2}^{n-1} \frac{\sqrt{1-\rho_{i+1}^2}}{\rho_{i+1}} \\
&\leq \frac{1}{\pi}\sum_{i=2}^{n-1}\left(\sqrt{\frac{\alpha_i}{2-\alpha_i}} + \sum_{i=2}^{n-1}\sqrt{\frac{\alpha_{i+1}}{2-\alpha_{i+1}}}\right) \\
&\leq \frac{1}{\pi\sqrt{2-\alpha_{\max}}}\left(\sum_{i=2}^{n-1}\sqrt{\alpha_i} + \sum_{i=2}^{n-1}\sqrt{\alpha_{i+1}}\right) \\
&\leq \frac{1}{\pi\sqrt{2-\alpha_{\max}}\sqrt{1-\rho_{XU}^2}}\left(\sum_{i=2}^{n-1}\frac{(X_i - X_{i-1})}{\sqrt{S_X^2}} + \sum_{i=2}^{n-1}\frac{(X_{i+1} - X_i)}{\sqrt{S_X^2}}\right) \\
&\quad + \frac{1}{\pi\sqrt{2-\alpha_{\max}}\sqrt{1-\rho_{XU}^2}}\left(\sum_{i=2}^{n-1}\frac{(U_i - U_{i-1})}{\sqrt{S_U^2}} + \sum_{i=2}^{n-1}\frac{(U_{i+1} - U_i)}{\sqrt{S_U^2}}\right) \\
&= \frac{1}{\pi\sqrt{2-\alpha_{\max}}\sqrt{1-\rho_{XU}^2}}\left(\frac{X_{n-1} - X_1 + X_n - X_2}{\sqrt{S_X^2}}\right) \\
&\quad + \frac{1}{\pi\sqrt{2-\alpha_{\max}}\sqrt{1-\rho_{XU}^2}}\left(\frac{U_{n-1} - U_1 + U_n - U_2}{\sqrt{S_U^2}}\right). \qquad (1.71)
\end{aligned}$$

By the last three inequalities in (1.35)

$$\frac{X_{n-1} - X_1 + X_n - X_2}{\sqrt{S_X^2}} \leq 2, \quad \frac{U_{n-1} - U_1 + U_n - U_2}{\sqrt{S_U^2}} \leq 2.$$

Hence (1.71) gives

$$E[|\widehat{T}_n - T_n|] \leq \frac{4}{\pi\sqrt{2 - \alpha_{\max}}\sqrt{1 - \rho_{XU}^2}}, \tag{1.72}$$

proving the first statement in Proposition 9.

If the design points are equally spaced with distance d between consecutive points, it is easy to calculate that

$$S_X^2 = \frac{d^2}{12}(n^3 - n), \; S_U^2 = \frac{d^4}{180}(n-1)n(2n-1)(8n^2 - 3n - 11), \tag{1.73}$$

and

$$\text{Cov}(X, U) = \frac{d^3}{12}(n-1)^2 n(n+1). \tag{1.74}$$

Hence

$$\rho_{XU} = \sqrt{\frac{15(n-1)^2}{(16n^2 - 30n + 11)}} = \sqrt{\frac{15}{16}} + O(1/n), \tag{1.75}$$

and by (1.69)

$$\begin{aligned}
&\alpha_i \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (1.76)\\
= \; & \frac{(X_i - X_{i-1})^2 S_U^2 - 2(X_i - X_{i-1})(U_i - U_{i-1})\text{Cov}(X,U) + (U_i - U_{i-1})^2 S_X^2}{S_X^2 S_U^2 - \text{Cov}(X,U)^2} \\
= \; & \frac{\frac{d^6}{180}(n-1)n(2n-1)(8n^2 - 13n - 11)}{\frac{d^6}{12\times 180}(n-2)(n-1)^2 n^2 (2n-1)(8n^2 - 13n - 11) - \frac{d^6}{144}(n-1)^4 n^2 (n+1)^2} \\
& + \frac{-2(2i-3)\frac{d^6}{12}(n-1)^2 n(n+1) + (2i-3)^2 \frac{d^6}{12}(n^3 - n)}{\frac{d^6}{12\times 180}(n-2)(n-1)^2 n^2 (2n-1)(8n^2 - 13n - 11) - \frac{d^6}{144}(n-1)^4 n^2 (n+1)^2} \\
= \; & \frac{O(n^5)}{(\frac{1}{135} - \frac{1}{144})n^8 + O(n^7)} \\
= \; & O(n^{-3}).
\end{aligned}$$

Therefore (1.71) gives

$$
\begin{aligned}
& E[|\widehat{T}_n - T_n|] && (1.77)\\
\leq\ & \frac{1}{\pi\sqrt{2-\alpha_{\max}}\sqrt{1-\rho_{XU}^2}}\left(\frac{2(n-2)d}{\sqrt{\frac{d^2}{12}(n^3-n)}}\right)\\
& + \frac{1}{\pi\sqrt{2-\alpha_{\max}}\sqrt{1-\rho_{XU}^2}}\left(\frac{\left((n-2)^2+(n-1)^2-1\right)d^2}{\sqrt{\frac{d^4}{180}(n-1)n(2n-1)(8n^2-13n-11)}}\right)\\
\leq\ & \frac{4\left(4\sqrt{3}+3\sqrt{5}\right)}{\pi\sqrt{2}}\frac{1}{\sqrt{n}} + O(1/n).
\end{aligned}
$$

□

1.4 Minimal number of turning points in residuals

In this section, we show that for any kth order polynomial regression there will be at least k turning points among the residuals if $X_1 < X_2 < \cdots < X_n$. The proof given here does not require the errors to be i.i.d normal. In this section we will only assume the residuals are not all zero (and therefore we must assume $n > k+1$).

Let's define $(j, ..., s)$ to be a tied turning point of the residuals $r_i = \widehat{\varepsilon}_i$ if $r_{j-1} < r_j = ... = r_s > r_{s+1}$ or if $r_{j-1} > r_j = ... = r_s < r_{s+1}$. This is an extension of the previous definition of a turning point. For convenience, in this section we will simply call a tied turning point a turning point and all appearances of the latter refer to the extended definition. Let us also say that there is a change of sign in the residuals if $r_{j-1} > r_j = 0 = ... = r_s > r_{s+1}$ or $r_{j-1} < r_j = 0 = ... = r_s < r_{s+1}$ in addition to the cases where we would already say that, $r_j > 0 > r_{j+1}$ or $r_j < 0 < r_{j+1}$.

Theorem 10. *For any kth order polynomial regression with design points $X_1 < X_2 < \cdots < X_n$ and $n > k+1$, if the residuals are not all zero, then there are at least k (tied) turning points among them.*

Proof. . We first show there is at least one turning point among the residuals for simple linear regression. By equations (1.8) we have

$$\sum_{i=1}^{n} \widehat{\varepsilon}_i = 0 = \sum_{i=1}^{n} X_i \widehat{\varepsilon}_i. \tag{1.78}$$

Let

$$\Delta_1 = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_n \end{pmatrix}.$$

We can rewrite equations (1.78) in a matrix form:

$$\Delta_1 \begin{pmatrix} \widehat{\varepsilon}_1 \\ \dots \\ \widehat{\varepsilon}_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Suppose there are no turning points among the residuals. Then $\widehat{\varepsilon}_i$ change signs at most once. Together with the first equality in (1.78) this means $\widehat{\varepsilon}_i$ change signs exactly once. Without loss of generality, we may assume $\widehat{\varepsilon}_1 < 0$. Let j be the index with $1 < j \leq n$ such that $\widehat{\varepsilon}_i < 0$ for all $i < j$ and $\widehat{\varepsilon}_i \geq 0$ for all $i \geq j$. We apply a row operation to the matrix Δ_1 by taking $-X_j$ times the first row and adding it to the second row:

$$\begin{pmatrix} 1 & \dots & 1 & 1 & 1 & \dots & 1 \\ X_1 - X_j & \dots & X_{j-1} - X_j & 0 & X_{j+1} - X_j & \dots & X_n - X_j \end{pmatrix}.$$

Note that the first $j-1$ terms in the second row are < 0 and the last $n-j$ terms are ≥ 0. Consequently we have

$$(X_1 - X_j, \dots, X_{j-1} - X_j, 0, X_{j+1} - X_j, \dots, X_n - X_j) \begin{pmatrix} \widehat{\varepsilon}_1 \\ \dots \\ \widehat{\varepsilon}_n \end{pmatrix} \geq (X_1 - X_j)\widehat{\varepsilon}_1 > 0,$$

hence contradicting (1.78). Therefore in simple linear regression we have at least one turning point among the residuals. Furthermore what we actually proved is a stronger necessary condition: the residuals change signs at least twice.

Let $\Delta_m = \tilde{X}^T$ for $\tilde{X}$ defined by (1.21) with $m = k$, an $m+1$ by n matrix. Then by (1.8) for $f_i(x) \equiv x^{i-1}$, $i = 1, \ldots, m+1$, the residuals from an mth order regression satisfy

$$\Delta_m \begin{pmatrix} \widehat{\varepsilon}_1 \\ \cdots \\ \widehat{\varepsilon}_n \end{pmatrix} = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}. \tag{1.79}$$

It will be shown by induction on m, using only equation (1.79), that the residuals change signs at least $m+1$ times and thus have at least m turning points. We have proved this for $m = 1$. For $m \geq 2$, suppose this is true for $m-1$. Then Δ_{m-1}, being the first m rows of Δ_m, satisfies

$$\Delta_{m-1} \begin{pmatrix} \widehat{\varepsilon}_1 \\ \cdots \\ \widehat{\varepsilon}_n \end{pmatrix} = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}. \tag{1.80}$$

By induction assumption, (1.80) implies the residuals have at least m changes of signs. Suppose the residuals change signs exactly m times and let $1 \leq i_1 < \ldots < i_s < \ldots < i_m < i_{m+1} \leq n$ be indices such that for $s = 1, \ldots, m$, $\widehat{\varepsilon}_{i_s}\widehat{\varepsilon}_{i_{s+1}} < 0$, and for $i_s < l \leq i_{s+1}$, $\widehat{\varepsilon}_l$ are either zero or have the same sign as $\widehat{\varepsilon}_{i_{s+1}}$; for $1 \leq l \leq i_1$, $\widehat{\varepsilon}_l$ are either zero or have the same sign as $\widehat{\varepsilon}_{i_1}$; and for $i_{m+1} < l \leq n$, $\widehat{\varepsilon}_l$ are either zero or have the same sign as $\widehat{\varepsilon}_{i_{m+1}}$. Without loss of generality let us assume $\widehat{\varepsilon}_{i_1} < 0$. We will show that Δ_m can be transformed by row operations into

$$\begin{pmatrix} 1 & \cdots & 1 \\ X_1 - X_{i_1} & \cdots & X_n - X_{i_1} \\ \cdots & \cdots & \cdots \\ \prod_{j=1}^{s}(X_1 - X_{i_j}) & \cdots & \prod_{j=1}^{s}(X_n - X_{i_j}) \\ \cdots & \cdots & \cdots \\ \prod_{j=1}^{m}(X_1 - X_{i_j}) & \cdots & \prod_{j=1}^{m}(X_n - X_{i_j}) \end{pmatrix}.$$

We have $(\Delta_m)_{ij} = X_j^{i-1}$, $i = 1, \ldots, m+1$. The linear span of the rows of Δ_m consists of all polynomials of degree at most m evaluated at X_j for $j = 1, \ldots, n$. The rows of the above matrix are such polynomials, of respective degrees $0, 1, \ldots, m$, at X_j, $j = 1, \ldots, n$, and are monic (have leading coefficient 1), so this matrix and Δ_m are equivalent via row operations.

Since $n > m + 1$, there is at least one $t = 1, \ldots, n$ such that either $1 \leq t < i_1$, or $i_s < t < i_{s+1}$ for some s, or $i_{m+1} < t \leq n$. For any such t, if $i_s < t < i_{s+1}$ for some $1 \leq s \leq m$ then $\text{sgn}(\prod_{j=1}^{m}(X_t - X_{i_j})) = (-1)^{m-s}$ and $\text{sgn}(\widehat{\varepsilon}_t) = (-1)^{s+1}$ if $\widehat{\varepsilon}_t \neq 0$ so $\text{sgn}(\prod_{j=1}^{m}(X_t - X_{i_j})\widehat{\varepsilon}_t) = (-1)^{m+1}$. If $t > i_{m+1}$ or $t < i_1$, then it is obvious that $\text{sgn}(\prod_{j=1}^{m}(X_t - X_{i_j})\widehat{\varepsilon}_t) = (-1)^{m+1}$ if $\widehat{\varepsilon}_t \neq 0$. Denote the last row of the last matrix by R_m. To $R_m\widehat{\varepsilon}$, terms with $X_r = X_{i_j}$ for some $j = 1, \ldots, m$ make 0 contribution. Other r with $\widehat{\varepsilon}_r = 0$ also make 0 contribution, or if $\widehat{\varepsilon}_r \neq 0$ they make a contribution with sign $(-1)^{m+1}$. There is at least one such r with $\widehat{\varepsilon}_r \neq 0$, namely $r = i_{m+1}$. Thus

$$\text{sgn}\left(R_m \begin{pmatrix} \widehat{\varepsilon}_1 \\ \cdots \\ \widehat{\varepsilon}_n \end{pmatrix} \right) = (-1)^{m+1},$$

and hence

$$R_m \begin{pmatrix} \widehat{\varepsilon}_1 \\ \cdots \\ \widehat{\varepsilon}_n \end{pmatrix} \neq 0,$$

which contradicts (1.8) for $f_i \equiv x^{i-1}$, $i = 1, \ldots, m+1$. Therefore the residuals from an mth order polynomial regression change signs at least $m + 1$ times and thus have at least m turning points, proving the theorem. □

1.5 Distributions of turning point statistics by exact calculations and simulations

Consider the simple linear regression model $Y_j = a + bX_j + \varepsilon_j$ where the errors ε_j are assumed to be i.i.d. $N(0, \sigma^2)$ for some unknown σ. Suppose X_j, $j = 1, 2, ..., n$ are equally spaced design points. Let T_n be the number of turning points of the errors and $\widehat{T}_n$ the number of turning points of the residuals.

1.5.1 Exact distributions

As we showed in Section 1.2, we can evaluate the distribution of T_n numerically using (1.1), (1.2), and (1.3). In this section, we will consider both two-sided and one-sided tests and tabulate critical values and critical probabilities associated with 0.05 and 0.01 level tests.

For $n = 7, 8$, our numerical evaluation shows $P_0(T_7 = 5) \doteq 0.1079$ and $P_0(T_8 = 6) \doteq 0.0687$. Therefore for $n = 7, 8$ the upper 0.05 and 0.01 quantiles do not exist and both 0.05 and 0.01 level tests are one-sided. For $n = 9, 10$ we obtained $P_0(T_9 = 7) \doteq 0.0437$ and $P_0(T_{10} = 8) \doteq 0.0278$ so the tests are one-sided at the 0.01 level. For $n = 9$ at the 0.05 level since $P_0(T_9 \leq 1) \doteq 0.0014$ and $P_0(T_9 \leq 2) \doteq 0.0257$ we have two options. The first option is to reject the null hypothesis if $T_n \leq 2$ while the second option is to reject if $T_n \leq 1$ or $T_n = 7$. Here we give some critical probabilities for $n = 7, 8, 9, 10$ for use in one-sided tests at the 0.05 and 0.01 levels.

Table 1.1: Critical Values of T_n for One-Sided 0.05 and 0.01 Level Tests

n	$P_0(T_n \leq 1)$	$P_0(T_n \leq 2)$	$P_0(T_n \leq 3)$	$P_0(T_n = n-2)$	$\alpha = 0.05$	$\alpha = 0.01$
7	0.0250	0.1909	0.5583	0.1079	$T_n \leq 1$	$T_n \leq 0$*
8	0.0063	0.0749	0.3124	0.0687	$T_n \leq 1$	$T_n \leq 1$
9	0.0014	0.0257	0.1500	0.0437	$T_n \leq 2$ or $T_n = 7$*	$T_n \leq 1$
10	0.0003	0.0079	0.0633	0.0278	$T_n \leq 2$ or $T_n = 8$*	$T_n \leq 2$
*$P_0(T_7 = 0) = 0.0004$, $P_0(T_{10} = 7) = 0.1393$						

For $11 \leq n \leq 50$, let $k_0 := k_0(n)$ be the largest k such that $P_0(T_n \leq k) \leq 0.025$ and $k_1 := k_1(n)$ the smallest k such that $P_0(T_n \geq k) \leq 0.025$. let $l_0 := l_0(n)$ be the largest l such that $P_0(T_n \leq l) \leq 0.005$ and $l_1 := l_1(n)$ the smallest l such that $P_0(T_n \geq l) \leq 0.005$. In other words, k_0, k_1, l_0, l_1 are the critical values of T_n for two-sided 0.05 and 0.01 level tests with symmetric critical regions. Similarly for one-sided tests, we define $k_0^1 := k_0(n)$ be the largest k such that $P_0(T_n \leq k) \leq 0.05$ and $k_1^1 := k_1(n)$ the smallest k such that $P_0(T_n \geq k) \leq 0.05$. let $l_0^1 := l_0(n)$ be the largest l such that $P_0(T_n \leq l) \leq 0.01$ and $l_1^1 := l_1(n)$ the smallest l such that $P_0(T_n \geq l) \leq 0.01$.

We define $F_n^l(j) := P_0(T_n \leq j)$ and $F_n^r(j) := P_0(T_n \geq j)$. Further, let Y_n be a normal random variable with the same mean $2(n-2)/3$ and the same variance $(16n-29)/90$ as T_n under H_0. For any integer $j \geq 0$, normal approximations to $F_n^l(j)$ and $F_n^r(j)$ with correction for continuity is $G_n^l(j) := P(Y_n \leq j + \frac{1}{2})$ and $G_n^r(j) := P(Y_n \geq j - \frac{1}{2})$.

The following tables give k_0, k_1, l_0, l_1, k_0^1, k_1^1, l_0^1, l_1^1 and the corresponding probabilities, found by exact calculations. In conservative two-sided tests of H_0 at the 0.05 (or 0.01) level, one would reject H_0 if if $T_n \leq k_0$ (or l_0) or if $T_n \geq k_1$ (or l_1), respectively. The tables also give $F_n^l(k_0+1)$ and $F_n^r(k_1-1)$ both to justify that k_0 and k_1 have the given values and to allow H_0 to be rejected in a non-conservative test, if the user so chooses (in advance), in case the larger probabilities are closer to 0.025 than the conservative probabilities are. Analogous options are available for $\alpha = 0.01$ and for one-sided tests for each of the two values of α.

We can see that for $n \geq 11$, the normal approximations for $F_n^l(k_0)$ and $F_n^l(l_0)$ with correction for continuity work well. They give, except for $n = 11$ as shown in bold in Table 1.2, for $\alpha = 0.05$ and 0.01, the correct values of k_0 and l_0 respectively for the two-sided test and the corresponding quantities with superscript 1 for one-sided tests. In fact for $n \geq 21$ the differences between the cumulative probabilities and their normal approximations are smaller than 0.0004 and for $n \geq 30$ they are smaller than 0.0003 in the lower left tail.

On the contrary, the approximation of F_n^r by G_n^r is not so good as G_n^r tend to be larger than F_n^r. The numbers in boldface indicate that the the normal approximation would erroneously replace k_0 (or l_0) by $k_0 - 1$ (or $l_0 - 1$) or k_1 (or l_1) by $k_1 + 1$ (or $l_1 + 1$) in a conservative two-sided test. However, the normal approximation would never erroneously give a choice of $k_1 - 1$ (or $l_1 - 1$) in place of k_1 (or l_1). The same pattern is observed for one-sided tests. In Tables 1.2 through 1.5, except for $n = 11$, the bold face discrepancy cases are all in the upper tail. Thus if a lower one-sided test is used, not only is it more powerful, but this error issue is avoided.

Table 1.2: Critical Values of T_n and Probabilities for Two-Sided 0.05 Level Tests

n	k_0	$F_n^l(k_0)$	$F_n^l(k_0+1)$	$G_n^l(k_0)$	$G_n^l(k_0+1)$	k_1	$F_n^r(k_1)$	$F_n^r(k_1-1)$	$G_n^r(k_1)$	$G_n^r(k_1-1)$
11	3	.0239	.1196	**.0252**	.1203	9	.0177	.1177	**.0252**	.1203
12	3	.0082	.0529	.0093	.0537	10	.0113	.0821	.0176	.0866
13	4	.0213	.0964	.0223	.0968	11	.0072	.0568	.0124	.0622
14	4	.0079	.0441	.0087	.0447	12	.0046	.0391	.0087	.0447
15	5	.0186	.0782	.0193	.0785	13	.0029	.0267	.0061	.0321
16	5	.0072	.0367	.0079	.0372	13	.0182	.0828	.0231	.0862
17	6	.0160	.0638	.0166	.0641	14	.0123	.0600	.0166	.0641
18	6	.0065	.0306	.0070	.0310	15	.0083	.0431	.0119	.0474
19	7	.0137	.0523	.0142	.0525	16	.0056	.0308	.0086	.0350
20	7	.0058	.0255	.0062	.0258	16	.0218	.0793	**.0258**	.0822
21	8	.0117	.0431	.0120	.0432	17	.0154	.0591	.0190	.0625
22	9	.0213	.0674	.0215	.0674	18	.0108	.0437	.0139	.0473
23	9	.0099	.0356	.0102	.0357	19	.0076	.0321	.0102	.0357
24	10	.0177	.0554	.0180	.0554	19	.0235	.0742	**.0268**	.0768
25	10	.0084	.0294	.0086	.0295	20	.0170	.0564	.0201	.0594
26	11	.0148	.0458	.0150	.0457	21	.0123	.0426	.0150	.0457
27	12	.0244	.0674	.0245	.0673	22	.0088	.0320	.0112	.0350
28	12	.0124	.0379	.0125	.0378	22	.0238	.0686	**.0267**	.0711
29	13	.0203	.0558	.0203	.0557	23	.0177	.0530	.0203	.0557
30	13	.0104	.0314	.0105	.0313	24	.0130	.0407	.0154	.0434
31	14	.0169	.0463	.0169	.0462	25	.0095	.0310	.0117	.0337
32	14	.0087	.0261	.0088	.0260	25	.0235	.0631	**.0260**	.0654
33	15	.0141	.0385	.0141	.0384	26	.0177	.0493	.0201	.0518
34	16	.0217	.0547	.0217	.0545	27	.0132	.0383	.0154	.0408
35	16	.0117	.0321	.0118	.0320	28	.0099	.0295	.0118	.0320
36	17	.0181	.0457	.0181	.0455	28	.0227	.0578	.0250	.0600
37	17	.0098	.0268	.0098	.0267	29	.0173	.0456	.0194	.0479
38	18	.0151	.0382	.0151	.0380	30	.0131	.0357	.0151	.0380
39	19	.0223	.0528	.0222	.0526	31	.0099	.0279	.0116	.0301
40	19	.0126	.0320	.0126	.0318	31	.0216	.0528	.0237	.0549
41	20	.0187	.0443	.0186	.0441	32	.0167	.0420	.0186	.0441
42	20	.0105	.0268	.0105	.0266	33	.0128	.0332	.0145	.0353
43	21	.0156	.0373	.0156	.0370	34	.0098	.0261	.0113	.0281
44	22	.0224	.0504	.0223	.0502	34	.0204	.0482	.0223	.0502
45	22	.0131	.0313	.0130	.0311	35	.0159	.0386	.0176	.0406
46	23	.0188	.0426	.0187	.0423	36	.0123	.0307	.0139	.0326
47	23	.0110	.0263	.0109	.0262	36	.0243	.0544	**.0262**	.0562
48	24	.0158	.0359	.0157	.0357	37	.0191	.0440	.0209	.0458
49	25	.0222	.0479	.0220	.0476	38	.0150	.0354	.0166	.0372
50	25	.0133	.0303	.0132	.0301	39	.0117	.0283	.0132	.0301

Table 1.3: Critical Values of T_n and Probabilities for One-Sided 0.05 Level Tests

	Lower one-sided tests					Upper one-sided tests				
n	k_0^1	$F_n^l(k_0^1)$	$F_n^l(k_0^1+1)$	$G_n^l(k_0^1)$	$G_n^l(k_0^1+1)$	k_1^1	$F_n^r(k_1^1)$	$F_n^r(k_1^1-1)$	$G_n^r(k_1^1)$	$G_n^r(k_1^1-1)$
11	3	.0239	.1196	.0252	.1203	9	.0177	.1177	.0252	.1203
12	3	.0082	.0529	.0093	.0537	10	.0113	.0821	.0176	.0866
13	4	.0213	.0964	.0223	.0968	11	.0072	.0568	.0124	.0622
14	5	.0441	.1534	.0447	.1541	11	.0391	.1536	.0447	.1541
15	5	.0186	.0782	.0193	.0785	12	.0267	.1134	.0321	.1156
16	6	.0367	.1238	.0372	.1242	13	.0182	.0828	.0231	.0862
17	6	.0160	.0638	.0166	.0641	14	.0123	.0600	.0166	.0641
18	7	.0306	.1006	.0310	.1008	14	.0431	.1389	.0474	.1399
19	7	.0137	.0523	.0142	.0525	15	.0308	.1055	.0350	.1076
20	8	.0255	.0821	.0258	.0822	16	.0218	.0793	.0258	.0822
21	9	.0431	.1202	.0432	.1204	17	.0154	.0591	.0190	.0625
22	9	.0213	.0674	.0215	.0674	17	.0437	.1251	.0473	.1264
23	10	.0356	.0988	.0357	.0988	18	.0321	.0968	.0357	.0988
24	10	.0177	.0554	.0180	.0554	19	.0235	.0742	.0268	.0768
25	11	.0294	.0815	.0295	.0814	20	.0170	.0564	.0201	.0594
26	12	.0458	.1139	.0457	.1140	20	.0426	.1125	.0457	.1140
27	12	.0244	.0674	.0245	.0673	21	.0320	.0882	.0350	.0903
28	13	.0379	.0946	.0378	.0946	22	.0238	.0686	.0267	.0711
29	13	.0203	.0558	.0203	.0557	23	.0177	.0530	.0203	.0557
30	14	.0314	.0787	.0313	.0786	23	.0407	.1012	.0434	.1028
31	15	.0463	.1068	.0462	.1068	24	.0310	.0802	.0337	.0822
32	15	.0261	.0656	.0260	.0654	25	.0235	.0631	.0260	.0654
33	16	.0385	.0895	.0384	.0893	25	.0493	.1132	**.0518**	.1144
34	16	.0217	.0547	.0217	.0545	26	.0383	.0911	.0408	.0928
35	17	.0321	.0750	.0320	.0748	27	.0295	.0728	.0320	.0748
36	18	.0457	.0996	.0455	.0995	28	.0227	.0578	.0250	.0600
37	18	.0268	.0629	.0267	.0627	28	.0456	.1014	.0479	.1027
38	19	.0382	.0840	.0380	.0838	29	.0357	.0821	.0380	.0838
39	19	.0223	.0528	.0222	.0526	30	.0279	.0661	.0301	.0680
40	20	.0320	.0708	.0318	.0706	31	.0216	.0528	.0237	.0549
41	21	.0443	.0926	.0441	.0924	31	.0420	.0910	.0441	.0924
42	21	.0268	.0598	.0266	.0595	32	.0332	.0741	.0353	.0758
43	22	.0373	.0785	.0370	.0783	33	.0261	.0599	.0281	.0618
44	22	.0224	.0504	.0223	.0502	33	.0482	.0994	**.0502**	.1006
45	23	.0313	.0666	.0311	.0663	34	.0386	.0818	.0406	.0833
46	24	.0426	.0859	.0423	.0857	35	.0307	.0669	.0326	.0686
47	24	.0263	.0565	.0262	.0562	36	.0243	.0544	.0262	.0562
48	25	.0359	.0732	.0357	.0730	36	.0440	.0892	.0458	.0905
49	26	.0479	.0930	.0476	.0928	37	.0354	.0736	.0372	.0751
50	26	.0303	.0624	.0301	.0621	38	.0283	.0604	.0301	.0621

Table 1.4: Critical Values of T_n and Probabilities for 0.01 Level Tests

n	l_0	$F_n^l(l_0)$	$F_n^l(l_0+1)$	$G_n^l(l_0)$	$G_n^l(l_0+1)$	l_1	$F_n^r(l_1)$	$F_n^r(l_1-1)$	$G_n^r(l_1)$	$G_n^r(l_1-1)$
11*	2	.0022	.0239	.0031	.0252	NA	NA	NA	NA	NA
12*	2	.0005	.0082	.0010	.0093	NA	NA	NA	NA	NA
13*	3	.0026	.0213	.0033	.0223	NA	NA	NA	NA	NA
14	3	.0007	.0079	.0011	.0087	12	.0046	.0391	**.0087**	.0447
15	4	.0027	.0186	.0033	.0193	13	.0029	.0267	**.0061**	.0321
16	4	.0009	.0072	.0012	.0079	14	.0019	.0182	.0044	.0231
17	5	.0026	.0160	.0031	.0166	15	.0012	.0123	.0031	.0166
18	5	.0009	.0065	.0012	.0070	16	.0008	.0083	.0022	.0119
19	6	.0025	.0137	.0028	.0142	17	.0005	.0056	.0016	.0086
20	6	.0009	.0058	.0011	.0062	17	.0038	.0218	.0062	.0258
21	7	.0023	.0117	.0026	.0120	18	.0025	.0154	.0044	.0190
22	7	.0009	.0050	.0010	.0054	19	.0017	.0108	.0032	.0139
23	8	.0021	.0099	.0023	.0102	20	.0011	.0076	.0023	.0102
24	9	.0044	.0177	.0046	.0180	21	.0007	.0053	.0017	.0075
25	9	.0018	.0084	.0020	.0086	21	.0037	.0170	**.0055**	.0201
26	10	.0038	.0148	.0040	.0150	22	.0025	.0123	.0040	.0150
27	10	.0016	.0071	.0018	.0073	23	.0017	.0088	.0029	.0112
28	11	.0033	.0124	.0034	.0125	24	.0012	.0063	.0021	.0083
29	11	.0014	.0060	.0016	.0062	24	.0045	.0177	**.0062**	.0203
30	12	.0028	.0104	.0029	.0105	25	.0032	.0130	.0046	.0154
31	12	.0012	.0051	.0014	.0052	26	.0023	.0095	.0034	.0117
32	13	.0024	.0087	.0025	.0088	27	.0016	.0070	.0025	.0088
33	14	.0043	.0141	.0044	.0141	27	.0011	.0051	**.0066**	.0201
34	14	.0020	.0073	.0021	.0074	28	.0037	.0132	.0050	.0154
35	15	.0036	.0117	.0037	.0118	29	.0026	.0099	.0037	.0118
36	15	.0017	.0061	.0018	.0062	30	.0019	.0073	.0028	.0090
37	16	.0031	.0098	.0031	.0098	31	.0014	.0054	.0021	.0068
38	16	.0015	.0051	.0016	.0052	31	.0040	.0131	**.0052**	.0151
39	17	.0026	.0082	.0027	.0082	32	.0029	.0099	.0039	.0116
40	18	.0043	.0126	.0044	.0126	33	.0021	.0075	.0030	.0090
41	18	.0022	.0069	.0022	.0069	34	.0015	.0056	.0022	.0069
42	19	.0036	.0105	.0037	.0105	34	.0042	.0128	**.0053**	.0145
43	19	.0019	.0058	.0019	.0058	35	.0031	.0098	.0040	.0113
44	20	.0030	.0088	.0031	.0088	36	.0023	.0074	.0031	.0088
45	21	.0048	.0131	.0048	.0130	37	.0017	.0056	.0023	.0068
46	21	.0026	.0074	.0026	.0074	37	.0042	.0123	**.0053**	.0139
47	22	.0041	.0110	.0041	.0109	38	.0032	.0095	.0041	.0109
48	22	.0022	.0062	.0022	.0062	39	.0024	.0073	.0031	.0085
49	23	.0034	.0092	.0034	.0092	40	.0018	.0056	.0024	.0067
50	23	.0018	.0052	.0018	.0052	40	.0042	.0117	**.0052**	.0132

*For n=11, $P_0(T_{11}=9)=0.0177$, reject H_0 if $T_{11} \leq 2$

*For n=12, $P_0(T_{12} \leq 4)=0.0529$, $P_0(T_{12}=10)=0.0113$, reject H_0 if $T_{12} \leq 3$

*For n=13, $P_0(T_{13}=11)=0.0072$, $P_0(T_{13}=10)=0.0497$ reject H_0 if $T_{13} \leq 3$ or $T_{13}=11$

Table 1.5: Critical Values of T_n and Probabilities for One-Sided 0.01 Level Tests

	Lower one-sided tests					Upper one-sided tests				
n	l_0^1	$F_n^l(l_0^1)$	$F_n^l(l_0^1+1)$	$G_n^l(l_0^1)$	$G_n^l(l_0^1+1)$	l_1^1	$F_n^r(l_1^1)$	$F_n^r(l_1^1-1)$	$G_n^r(l_1^1)$	$G_n^r(l_1^1-1)$
11	2	.0022	.0239	.0031	.0252	NA	NA	NA	NA	NA
12	3	.0082	.0529	.0093	.0537	NA	NA	NA	NA	NA
13	3	.0026	.0213	.0033	.0223	11	.0072	.0568	**.0124**	.0622
14	4	.0079	.0441	.0087	.0447	12	.0046	.0391	.0087	.0447
15	4	.0027	.0186	.0033	.0193	13	.0029	.0267	.0061	.0321
16	5	.0072	.0367	.0079	.0372	14	.0019	.0182	.0044	.0231
17	5	.0026	.0160	.0031	.0166	15	.0012	.0123	.0031	.0166
18	6	.0065	.0306	.0070	.0310	15	.0083	.0431	**.0119**	.0474
19	6	.0025	.0137	.0028	.0142	16	.0056	.0308	.0086	.0350
20	7	.0058	.0255	.0062	.0258	17	.0038	.0218	.0062	.0258
21	7	.0023	.0117	.0026	.0120	18	.0025	.0154	.0044	.0190
22	8	.0050	.0213	.0054	.0215	19	.0017	.0108	.0032	.0139
23	9	.0099	.0356	.0102	.0357	19	.0076	.0321	**.0102**	.0357
24	9	.0044	.0177	.0046	.0180	20	.0053	.0235	.0075	.0268
25	10	.0084	.0294	.0086	.0295	21	.0037	.0170	.0055	.0201
26	10	.0038	.0148	.0040	.0150	22	.0025	.0123	.0040	.0150
27	11	.0071	.0244	.0073	.0245	22	.0088	.0320	**.0112**	.0350
28	11	.0033	.0124	.0034	.0125	23	.0063	.0238	.0083	.0267
29	12	.0060	.0203	.0062	.0203	24	.0045	.0177	.0062	.0203
30	12	.0028	.0104	.0029	.0105	25	.0032	.0130	.0046	.0154
31	13	.0051	.0169	.0052	.0169	25	.0095	.0310	**.0117**	.0337
32	14	.0087	.0261	.0088	.0260	26	.0070	.0235	.0088	.0260
33	14	.0043	.0141	.0044	.0141	27	.0051	.0177	.0066	.0201
34	15	.0073	.0217	.0074	.0217	28	.0037	.0132	.0050	.0154
35	15	.0036	.0117	.0037	.0118	28	.0099	.0295	**.0118**	.0320
36	16	.0061	.0181	.0062	.0181	29	.0073	.0227	.0090	.0250
37	17	.0098	.0268	.0098	.0267	30	.0054	.0173	.0068	.0194
38	17	.0051	.0151	.0052	.0151	31	.0040	.0131	.0052	.0151
39	18	.0082	.0223	.0082	.0222	31	.0099	.0279	**.0116**	.0301
40	18	.0043	.0126	.0044	.0126	32	.0075	.0216	.0090	.0237
41	19	.0069	.0187	.0069	.0186	33	.0056	.0167	.0069	.0186
42	19	.0036	.0105	.0037	.0105	34	.0042	.0128	.0053	.0145
43	20	.0058	.0156	.0058	.0156	34	.0098	.0261	**.0113**	.0281
44	21	.0088	.0224	.0088	.0223	35	.0074	.0204	.0088	.0223
45	21	.0048	.0131	.0048	.0130	36	.0056	.0159	.0068	.0176
46	22	.0074	.0188	.0074	.0187	37	.0042	.0123	.0053	.0139
47	22	.0041	.0110	.0041	.0109	37	.0095	.0243	**.0109**	.0262
48	23	.0062	.0158	.0062	.0157	38	.0073	.0191	.0085	.0209
49	24	.0092	.0222	.0092	.0220	39	.0056	.0150	.0067	.0166
50	24	.0052	.0133	.0052	.0132	40	.0042	.0117	.0052	.0132

*For n=11, $P_0(T_{11} = 9) = 0.0177$, upper one-sided 0.01 level tests do not exist.
*For n=12, $P_0(T_{12} = 10) = 0.0113$, upper one-sided 0.01 level tests do not exist.

1.5.2 Distributions from simulations

Here we will give simulation results for $\widehat{T}_n$ for use in two-sided tests of H_0 at levels $\alpha = 0.01$ or 0.05. We will first assume that design points are equally spaced. Then we consider normal design points and general design points.

Equally Spaced Design Points

We will first assume that design points are equally spaced. For each n, we run $N = 5 \cdot 10^8$ iterations. In each iteration we generate a sample of n i.i.d $N(0,1)$ points as follows. We first generate uniform $[0,1]$ random variables using the random number generator gsl_rng_ranlxs2 in the GNU Scientific Library (GSL). According to the GSL documentation, this generator is based on the algorithm developed by Lüscher (1994) and has a period of about 10^{171}. We then use the Box-Muller transform (Box and Muller (1958)) to generate $N(0,1)$ random variables. We perform simple linear regression (or quadratic regression, cubic regression) on each sample and calculate the numbers of turning points for errors and residuals. Let us denote the fraction of samples with less than or equal to i residual turning points by $\hat{p} = f_n(i)$, which is a point estimate of $p = p_{n,i} = \Pr(T_n \leq i)$ for $0 \leq i \leq n-2$. Since our goal is to decide whether to reject the null hypothesis H_0 of simple linear regression with i.i.d. normal errors, at the .01 or .05 levels, we need to know whether one-sided cumulative probabilities are larger that 0.005 or 0.025 respectively and thus the smallest probability p (or $1-p$) of interest is 0.005. We can find confidence intervals for $0.005 \leq p \leq 0.995$ as follows. Note that $N = 5 \cdot 10^8$ and $Np \geq 2.5 \cdot 10^6$, which is in a range where the normal approximation to binomial probabilities is very good so we can use the conventional plug-in confidence intervals $\hat{p} \pm \zeta\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$ where $\zeta = 1.96$ or 2.576 for a 95% or 99% interval. Quadratic intervals work better than plug-in intervals for small p in general (Brown, Cai and DasGupta (2001, 2002)) but the endpoints of the two intervals differ by at most $K/N \leq 2 \cdot 10^{-8}$, which is negligible, for $N = 5 \cdot 10^8$ because $K \leq 10$ in our cases (as simple calculations show). Since $\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} < 2.3 \cdot 10^{-5}$ for $\hat{p} \in [0,1]$, $1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} < 0.5 \cdot 10^{-4}$ which means each $\hat{p}$ is accurate to four decimal places with 95% confidence. For $0.005 \leq \hat{p} < 0.01$, $2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} < 0.5 \cdot 10^{-4}$ so $\hat{p}$ is accurate to two significant digits with 99% confidence. For quadratic and cubic regressions, we estimate the distribution of $\widehat{T}_n$ in a similar fashion but with $N = 10^6$ iterations. Consequently $2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} < 0.005$ for any $\hat{p} \in [0,1]$ so $\hat{p}$ is accurate to two decimal places with 99% confidence. For $0.01 \leq \hat{p} \leq 0.05$, $1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} < 0.5 \cdot 10^{-3}$ so $\hat{p}$ is accurate to two significant digits with 95% confidence. For $0.005 \leq \hat{p} < 0.01$, $2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} < 0.5 \cdot 10^{-3}$ so $\hat{p}$ is accurate to one significant digit with 99% confidence.

For the distribution of $\hat{T}_n$ under the null hypothesis, critical regions do not exist at the 0.05 or 0.01 levels for $n \leq 5$. The following tables give critical values of $\widehat{T}_n$ for $6 \leq n \leq 10$ and critical regions for tests at the 0.05 and 0.01 levels.

Table 1.6: Critical Values for $\widehat{T}_n$ of Simple Linear Regression at 0.05 and 0.01 Levels

n	$P_0(\widehat{T}_n \leq 1)$	$P_0(\widehat{T}_n \leq 2)$	$P_0(\widehat{T}_n \leq 3)$	$P_0(\widehat{T}_n = n-2)$	$\alpha = 0.05$	$\alpha = 0.01$
6	0.0705	0.3791	0.8180	0.1820	NA	NA
7	0.0200	0.1706	0.5410	0.1154	$\widehat{T}_n \leq 1$	NA
8	0.0049	0.0658	0.2988	0.0719	$\widehat{T}_n \leq 1$	$\widehat{T}_n \leq 1$
9**	0.0011	0.0223	0.1421	0.0456	$\widehat{T}_n \leq 2$	$\widehat{T}_n \leq 1$
10	0.0002	0.0067	0.0595	0.0287	$\widehat{T}_n \leq 2$ or $\widehat{T}_n = 8$*	$\widehat{T}_n \leq 2$

*$P_0(\widehat{T}_n = 0) = 0$ for any $n \geq 3$, $P_0(\widehat{T}_{10} = 7) = 0.1431$
**For $n = 9$, an alternative critical region is $\widehat{T}_n \leq 1$ or $\widehat{T}_n = 7$

Table 1.7: Critical Values for $\widehat{T}_n$ of Quadratic Regression at 0.05 and 0.01 Levels

n	$P_0(\widehat{T}_n \leq 2)$	$P_0(\widehat{T}_n \leq 3)$	$P_0(\widehat{T}_n = n-2)$	$\alpha = 0.05$	$\alpha = 0.01$
6	0.3247	0.7459	0.2541	NA	NA
7	0.1366	0.4540	0.1322	NA	NA
8	0.0497	0.2355	0.0878	$\widehat{T}_n \leq 2$	NA
9	0.0164	0.1070	0.0508	$\widehat{T}_n \leq 2$	NA
10**	0.0049	0.0433	0.0330	$\widehat{T}_n \leq 2$ or $\widehat{T}_n = 8$*	$\widehat{T}_n \leq 2$

*$P_0(\widehat{T}_n \leq 1) = 0$ for any $n \geq 3$, $P_0(\widehat{T}_{10} = 4) = 0.1695$, $P_0(\widehat{T}_{10} = 7) = 0.1431$
**For $n = 10$, an alternative critical region is $\widehat{T}_n \leq 3$

Table 1.8: Critical Values for $\widehat{T}_n$ of Cubic Regression at 0.05 and 0.01 Levels

n	$P_0(\widehat{T}_n \leq 3)$	$P_0(\widehat{T}_n = n-2)$	$\alpha = 0.05$	$\alpha = 0.01$
6	0.7044	0.2956	NA	NA
7	0.3999	0.2327	NA	NA
8	0.1960	0.1036	NA	NA
9	0.0841	0.0731	NA	NA
10**	0.0321	0.0337	$\widehat{T}_n \leq 3$*	NA

*$P_0(\widehat{T}_n \leq 2) = 0$ for any $n \geq 3$, $P_0(\widehat{T}_{10} = 4) = 0.1087$, $P_0(\widehat{T}_{10} = 7) = 0.1956$.
**For $n = 10$, an alternative critical region is $\widehat{T}_n = 8$.

For $11 \leq n \leq 50$, let $\hat{k_0}$ be the largest k such that $P_0(\widehat{T}_n \leq k) \leq 0.025$ and $\hat{k_1}$ the smallest k such that $P_0(\widehat{T}_n \geq k) \leq 0.025$. let $\hat{l_0}$ be the largest l such that $P_0(\widehat{T}_n \leq l) \leq 0.005$ and $\hat{l_1}$ the smallest l such that $P_0(\widehat{T}_n \geq l) \leq 0.005$. We define $\hat{F}_n^l(j) := P_0(\widehat{T}_n \leq j)$ and $\hat{F}_n^r(j) := P_0(\widehat{T}_n \geq j)$.

Table 1.9: Critical Values and Probabilities for $\widehat{T}_n$ of Simple Linear Regression

n	$\hat{l}_0$	$\hat{k}_0$	$\hat{F}^l_n(\hat{l}_0)$	$\hat{F}^l_n(\hat{l}_0+1)$	$\hat{F}^l_n(\hat{k}_0)$	$\hat{F}^l_n(\hat{k}_0+1)$	$\hat{k}_1$	$\hat{l}_1$	$\hat{F}^r_n(\hat{l}_1)$	$\hat{F}^r_n(\hat{l}_1-1)$	$\hat{F}^r_n(\hat{k}_1)$	$\hat{F}^r_n(\hat{k}_1-1)$
11	2	3	.0018	.0223	.0223	.1146	9	NA	NA	NA	.0183	.1203
12	2	3	.0005	.0077	.0077	.0505	10	NA	NA	NA	.0113	.0836
13	3	4	.0024	.0202	.0202	.0938	11	NA	NA	NA	.0075	.0579
14	3	4	.0007	.0075	.0075	.0429	12	12	.0045	.0396	.0045	.0396
15	4	5	.0026	.0180	.0180	.0765	13	13	.0030	.0271	.0030	.0271
16	4	5	.0008	.0070	.0070	.0358	13	14	.0019	.0185	.0185	.0837
17	5	6	.0025	.0156	.0156	.0629	14	15	.0011	.0124	.0124	.0605
18	5	6	.0009	.0064	.0064	.0301	15	16	.0008	.0085	.0085	.0436
19	6	7	.0024	.0134	.0134	.0516	16	17	.0005	.0057	.0057	.0311
20	6	7	.0009	.0057	.0057	.0252	16	17	.0038	.0220	.0220	.0798
21	7	8	.0023	.0115	.0115	.0427	17	18	.0025	.0155	.0155	.0594
22	8	9	.0049	.0210	.0210	.0668	18	19	.0016	.0108	.0108	.0439
23	8	9	.0020	.0098	.0098	.0352	19	20	.0011	.0076	.0076	.0323
24	9	10	.0043	.0175	.0175	.0550	19	21	.0008	.0236	.0236	.0745
25	9	10	.0018	.0083	.0083	.0292	20	21	.0038	.0172	.0172	.0568
26	10	11	.0037	.0147	.0147	.0455	21	22	.0026	.0124	.0124	.0428
27	10	12	.0015	.0069	.0241	.0669	22	23	.0021	.0092	.0092	.0324
28	11	12	.0033	.0124	.0124	.0378	22	24	.0011	.0238	.0238	.0687
29	11	13	.0014	.0202	.0202	.0556	23	24	.0045	.0177	.0177	.0531
30	12	13	.0028	.0103	.0103	.0312	24	25	.0031	.0130	.0130	.0407
31	13	14	.0050	.0168	.0168	.0461	25	26	.0023	.0096	.0096	.0310
32	13	14	.0024	.0087	.0087	.0260	25	27	.0017	.0071	.0236	.0632
33	14	15	.0043	.0141	.0141	.0384	26	27	.0050	.0176	.0176	.0492
34	14	16	.0021	.0073	.0217	.0546	27	28	.0037	.0133	.0133	.0383
35	15	16	.0036	.0117	.0117	.0320	28	29	.0027	.0099	.0099	.0296
36	15	17	.0017	.0061	.0181	.0456	28	30	.0021	.0075	.0228	.0579
37	16	17	.0030	.0097	.0097	.0266	29	31	.0015	.0056	.0175	.0458
38	16	18	.0015	.0051	.0151	.0381	30	31	.0040	.0131	.0131	.0357
39	17	19	.0025	.0081	.0222	.0526	31	32	.0032	.0102	.0102	.0281
40	18	19	.0043	.0126	.0126	.0318	31	33	.0022	.0075	.0216	.0528
41	18	20	.0022	.0069	.0187	.0443	32	34	.0013	.0053	.0164	.0417
42	19	20	.0036	.0105	.0105	.0266	33	34	.0040	.0126	.0126	.0330
43	19	21	.0018	.0057	.0156	.0372	34	35	.0030	.0097	.0097	.0260
44	20	22	.0030	.0088	.0224	.0504	34	36	.0023	.0074	.0204	.0482
45	21	22	.0048	.0131	.0131	.0312	35	37	.0018	.0057	.0159	.0386
46	21	23	.0026	.0074	.0188	.0425	36	37	.0043	.0123	.0123	.0307
47	22	23	.0040	.0109	.0109	.0263	36	38	.0034	.0097	.0245	.0546
48	22	24	.0021	.0061	.0157	.0358	37	39	.0026	.0075	.0194	.0442
49	23	25	.0034	.0092	.0222	.0478	38	40	.0018	.0056	.0150	.0354
50	23	25	.0018	.0052	.0133	.0303	39	40	.0041	.0116	.0116	.0282

*$P_0(\widehat{T}_{11} = 9) = 0.0182$, $P_0(\widehat{T}_{12} = 10) = 0.0115$, $P_0(\widehat{T}_{13} = 11) = 0.0073$, $P_0(\widehat{T}_{13} = 10) = 0.0504$

*Critical regions for n=11,12,13 at the 0.01 level are $\widehat{T}_{11} \leq 2$, $\widehat{T}_{12} \leq 3$ and $\widehat{T}_{13} \leq 3$ or $\widehat{T}_{13} = 11$.

Table 1.10: Critical Values and Probabilities for $\widehat{T}_n$ of Quadratic Regression

n	$\hat{l}_0$	$\hat{k}_0$	$\hat{F}_n^l(\hat{l}_0)$	$\hat{F}_n^l(\hat{l}_0+1)$	$\hat{F}_n^l(\hat{k}_0)$	$\hat{F}_n^l(\hat{k}_0+1)$	$\hat{k}_1$	$\hat{l}_1$	$\hat{F}_n^r(\hat{l}_1)$	$\hat{F}_n^r(\hat{l}_1-1)$	$\hat{F}_n^r(\hat{k}_1)$	$\hat{F}_n^r(\hat{k}_1-1)$
11	2	3	.0014	.0159	.0159	.1021	9	NA	NA	NA	.0199	.1328
12	2	3	.0003	.0052	.0052	.0443	10	NA	NA	NA	.0128	.0896
13	3	4	.0016	.0174	.0174	.0821	11	NA	NA	NA	.0078	.0619
14	3	4	.0004	.0063	.0063	.0371	12	NA	NA	NA	.0051	.0414
15	4	5	.0021	.0154	.0154	.0714	13	13	.0032	.0287	.0032	.0287
16	4	5	.0006	.0059	.0059	.0330	13	14	.0020	.0192	.0192	.0873
17	5	6	.0022	.0146	.0146	.0586	14	15	.0012	.0130	.0130	.0626
18	5	6	.0007	.0058	.0058	.0277	15	16	.0008	.0087	.0087	.0452
19	6	7	.0022	.0124	.0124	.0499	16	17	.0005	.0058	.0058	.0319
20	6	**8**	.0008	.0051	.0240	.0779	16	17	.0040	.0230	.0230	.0818
21	7	8	.0021	.0109	.0109	.0404	17	18	.0027	.0159	.0159	.0607
22	8	9	.0047	.0201	.0201	.0651	18	19	.0017	.0111	.0111	.0448
23	8	9	.0019	.0093	.0093	.0340	19	20	.0012	.0080	.0080	.0331
24	9	10	.0041	.0168	.0168	.0530	19	21	.0007	.0053	.0238	.0755
25	9	10	.0017	.0080	.0080	.0284	20	21	.0038	.0177	.0177	.0576
26	10	11	.0036	.0140	.0140	.0444	21	22	.0025	.0125	.0125	.0432
27	10	12	.0015	.0068	.0235	.0653	22	23	.0018	.0091	.0091	.0326
28	11	12	.0030	.0119	.0119	.0369	22	24	.0012	.0065	.0243	.0698
29	11	13	.0013	.0058	.0196	.0546	23	24	.0046	.0180	.0180	.0539
30	12	13	.0027	.0098	.0098	.0302	24	25	.0032	.0132	.0132	.0412
31	13	14	.0050	.0165	.0165	.0455	25	26	.0022	.0095	.0095	.0314
32	13	14	.0023	.0084	.0084	.0254	25	27	.0016	.0071	.0238	.0641
33	14	15	.0041	.0136	.0136	.0377	26	27	.0050	.0179	.0179	.0497
34	14	16	.0020	.0071	.0213	.0540	27	28	.0037	.0135	.0135	.0389
35	15	16	.0034	.0115	.0115	.0315	28	29	.0027	.0100	.0100	.0298
36	15	17	.0017	.0060	.0178	.0450	28	30	.0018	.0072	.0226	.0581
37	16	17	.0030	.0097	.0097	.0264	29	31	.0013	.0053	.0171	.0456
38	**17**	18	.0050	.0149	.0149	.0380	30	31	.0040	.0132	.0132	.0358
39	17	19	.0026	.0082	.0223	.0521	31	32	.0030	.0101	.0101	.0281
40	18	19	.0041	.0125	.0125	.0316	31	33	.0021	.0075	.0218	.0534
41	18	20	.0022	.0068	.0185	.0435	32	34	.0015	.0055	.0166	.0420
42	19	20	.0035	.0104	.0104	.0263	33	34	.0042	.0128	.0128	.0332
43	19	21	.0018	.0057	.0154	.0368	34	35	.0031	.0099	.0099	.0263
44	20	22	.0030	.0086	.0224	.0500	34	36	.0022	.0074	.0206	.0484
45	21	22	.0047	.0129	.0129	.0310	35	37	.0017	.0057	.0161	.0390
46	21	23	.0025	.0073	.0186	.0423	36	37	.0044	.0125	.0125	.0310
47	22	23	.0040	.0108	.0108	.0259	36	38	.0032	.0096	.0244	.0547
48	22	24	.0020	.0059	.0152	.0352	37	39	.0024	.0075	.0193	.0442
49	23	25	.0034	.0092	.0220	.0474	38	40	.0017	.0054	.0149	.0353
50	**24**	25	.0050	.0131	.0131	.0300	39	40	.0043	.0119	.0119	.0286

*$P_0(\widehat{T}_{11} = 9) = 0.0199$, $P_0(\widehat{T}_{12} = 10) = 0.0128$, $P_0(\widehat{T}_{13} = 11) = 0.0078$, $P_0(\widehat{T}_{13} = 10) = 0.0541$

*Critical regions for n=11,12,13 at the 0.01 level are $\widehat{T}_{11} \leq 2$, $\widehat{T}_{12} \leq 3$ and $\widehat{T}_{13} \leq 3$ or $\widehat{T}_{13} = 11$.

*A critical region for n=14 at the 0.01 level is $\widehat{T}_{14} \leq 3$ or $\widehat{T}_{14} \geq 12$. Alternatively, one can use $\widehat{T}_{14} \leq 4$.

Bold fonts indicate entries different for simple linear and quadratic regressions.

Table 1.11: Critical Values and Probabilities for $\widehat{T}_n$ of Cubic Regression

n	$\hat{l}_0$	$\hat{k}_0$	$\hat{F}^l_n(\hat{l}_0)$	$\hat{F}^l_n(\hat{l}_0+1)$	$\hat{F}^l_n(\hat{k}_0)$	$\hat{F}^l_n(\hat{k}_0+1)$	$\hat{k}_1$	$\hat{l}_1$	$\hat{F}^r_n(\hat{l}_1)$	$\hat{F}^r_n(\hat{l}_1-1)$	$\hat{F}^r_n(\hat{k}_1)$	$\hat{F}^r_n(\hat{k}_1-1)$
11	NA	3	NA	NA	.0113	.0637	NA	NA	NA	NA	NA	NA
12	**3**	3	.0036	.0261	.0036	.0261	10	NA	NA	NA	.0142	.1065
13	3	4	.0011	.0098	.0098	.0707	11	NA	NA	NA	.0094	.0676
14	**4**	4	.0036	.0313	.0036	.0313	12	NA	NA	NA	.0056	.0481
15	4	5	.0012	.0129	.0129	.0564	13	13	.0036	.0312	.0036	.0312
16	**5**	5	.0047	.0254	.0047	.0254	13	14	.0022	.0217	.0217	.0934
17	5	6	.0018	.0110	.0110	.0529	14	15	.0014	.0142	.0142	.0687
18	**6**	7	.0042	.0248	.0248	.0839	15	16	.0009	.0097	.0097	.0481
19	6	7	.0016	.0109	.0109	.0424	16	17	.0006	.0064	.0064	.0346
20	**7**	8	.0044	.0204	.0204	.0721	16	17	.0041	.0239	.0239	.0867
21	7	8	.0018	.0093	.0093	.0377	17	18	.0028	.0171	.0171	.0637
22	8	9	.0040	.0183	.0183	.0592	18	19	.0018	.0119	.0119	.0473
23	8	9	.0016	.0084	.0084	.0308	19	20	.0012	.0081	.0081	.0341
24	9	10	.0035	.0150	.0150	.0501	**20**	21	.0009	.0058	.0058	.0253
25	9	10	.0015	.0071	.0071	.0264	20	21	.0039	.0181	.0181	.0603
26	10	11	.0031	.0132	.0132	.0414	21	22	.0026	.0131	.0131	.0448
27	10	12	.0013	.0064	.0221	.0630	22	23	.0018	.0094	.0094	.0338
28	11	12	.0029	.0111	.0111	.0353	22	24	.0012	.0067	.0250	.0720
29	11	13	.0013	.0054	.0186	.0515	23	24	.0046	.0184	.0184	.0549
30	12	13	.0025	.0096	.0096	.0291	24	25	.0034	.0137	.0137	.0423
31	13	14	.0047	.0157	.0157	.0439	25	26	.0022	.0098	.0098	.0318
32	13	**15**	.0022	.0081	.0249	.0626	25	27	.0016	.0073	.0245	.0649
33	14	15	.0039	.0133	.0133	.0365	26	**28**	.0012	.0054	.0182	.0507
34	14	16	.0020	.0069	.0206	.0524	27	28	.0038	.0139	.0139	.0396
35	15	16	.0034	.0110	.0110	.0305	28	29	.0027	.0101	.0101	.0305
36	15	17	.0016	.0057	.0173	.0435	28	30	.0020	.0076	.0233	.0596
37	16	17	.0028	.0093	.0093	.0257	29	31	.0015	.0055	.0177	.0464
38	17	18	.0050	.0146	.0146	.0371	30	31	.0041	.0136	.0136	.0368
39	17	19	.0025	.0078	.0217	.0511	31	32	.0030	.0102	.0102	.0286
40	18	19	.0041	.0121	.0121	.0306	31	33	.0022	.0077	.0221	.0540
41	18	20	.0021	.0068	.0183	.0431	32	34	.0015	.0056	.0170	.0427
42	19	20	.0033	.0099	.0099	.0258	33	34	.0042	.0129	.0129	.0337
43	19	21	.0018	.0055	.0150	.0361	34	35	.0033	.0101	.0101	.0269
44	20	22	.0028	.0084	.0215	.0490	34	36	.0024	.0077	.0207	.0491
45	21	22	.0046	.0125	.0125	.0304	35	37	.0017	.0056	.0159	.0392
46	21	23	.0025	.0072	.0183	.0417	36	37	.0043	.0124	.0124	.0311
47	22	23	.0040	.0107	.0107	.0258	36	38	.0033	.0097	.0248	.0555
48	22	24	.0020	.0059	.0151	.0349	37	39	.0023	.0073	.0193	.0446
49	23	25	.0032	.0089	.0217	.0468	38	40	.0018	.0057	.0153	.0360
50	24	25	.0049	.0128	.0128	.0294	39	40	.0044	.0121	.0121	.0288

*$P_0(\widehat{T}_{11} = 9) = .0254$, $P_0(\widehat{T}_{13} = 10) = .0582$, $P_0(\widehat{T}_{14} = 11) = .0425$

*A critical region for n=12 at the 0.01 level is $\widehat{T}_{12} \leq 3$.

*A critical region for n=13 at the 0.01 level is $\widehat{T}_{13} \leq 4$. Alternatively one can use $\widehat{T}_{13} = 11$.

*A critical region for n=14 at the 0.01 level are $\widehat{T}_{14} \leq 4$ or $\widehat{T}_{14} \geq 12$.

Bold fonts indicate entries different for quadratic and cubic regressions.

Normal Design Points

For design points which are a random normal sample rearranged in ascending order, Theorem 7 gives an $O(n^{-0.5}\sqrt{\log n})$ upper bound. We will investigate this upper bound through simulations.

For each $n = 10, 20, ..., 100$, we generate 100 random normal samples with each sample having size n. We then rearrange each sample in ascending order to obtain a set of n design points X. For each set of design points, we perform simulation as we did with equally spaced design points but with $N = 5 \cdot 10^5$ iterations instead of $N = 5 \cdot 10^8$. In each iteration i, we generate n i.i.d. $N(0,1)$ errors and perform linear regression and calculate $T_n^{(i)}$ and $\widehat{T}_n^{(i)}$. The sample average $S_X := \frac{1}{N}\sum_i |T_n^{(i)} - \widehat{T}_n^{(i)}|$ will be an estimator of $E_X[T_n - \widehat{T}_n]$ and $S = \frac{1}{100}\sum_X S_X$ is an estimator of $E[T_n - \widehat{T}_n]$. For $n = 10, 20, ..., 100$, our simulations give $S \doteq$0.4799, 0.3870, 0.3367, 0.3037, 0.2797, 0.2612, 0.2464, 0.2340, 0.2236, 0.2146. The sample correlation between S and $(\sqrt{\log n/n})_{n=10,20,...,100}$ is 0.9996 while the sample correlation between S and $(1/\sqrt{n})_{n=10,20,...,100}$ is 0.9932. Linearly regressing S against $(\sqrt{\log n/n})_{n=10,20,...,100}$ gives MSE $= 8.69 \cdot 10^{-7}$ and 7 residual turning points (compared to a maximum possible of $10 - 2 = 8$) so the linear regression model seems a good one. Theorem 7 appears from this evidence to be sharp. The regression slope is 0.3608 and the intercept is 0.0120.

The critical values of $\widehat{T}_n$ with normal design points generally agree with those of T_n except in cases where some cumulative probabilities of T_n are very close to .025 or 0.05. From the simulations, the critical values of T_n and $\widehat{T}_n$ agree for all of 100 sets of design points at the .05 level except for $n = 12, 19, 20, 33$ and 44. For $n = 12$, the lower one-sided test for T_n has its critical value equal to 3 but with probability $\Pr(T_n \leq 4) = .0529$ close to .05. The critical value of $\widehat{T}_n$ equals 3 for 32 out of 100 sets of normal design points and equals 4 for the rest where $\Pr(\widehat{T}_n \leq 4)$ becomes less than .05. For $n = 19$, the lower one-sided test for T_n has its critical value equal to 7 with probability $\Pr(T_n \leq 8) = .0523$. The critical value of $\widehat{T}_n$ equals 7 for 98 out of 100 sets of normal design points and equals 8 for the other two. For $n = 20$, the two-sided test for T_n has a critical value equal to 7 with probability $\Pr(T_n \leq 8) = .0255$. The critical value of $\widehat{T}_n$ of the two-sided test equals 7 for 32 out of 100 sets of normal design points and equals 8 for the rest. For $n = 33$, the upper one-sided test for T_n has its critical value equal to 25 with probability $\Pr(T_n \geq 25) = .0493$ but the critical value of $\widehat{T}_n$ equals 26 for all of 100 sets of normal design points. Finally for $n = 44$, the lower one-sided test for T_n has its critical value equal to 22 with probability $\Pr(T_n \leq 23) = .0504$. The critical value of $\widehat{T}_n$ equals 22 for 78 of 100 sets of normal design points and equals 23 for the rest.

General Design Points

For non-equally spaced design points, Proposition 5 gives us theoretical evidence that the distribution of $\widehat{T}_n$ will not necessarily get close to that of T_n. From the proof of Proposition 5, we can see that the difference between $\widehat{T}_n$ and T_n is more likely to be large if most design points are very close to the mean with a few outliers positioned very far from each other and the rest of the design points. Motivated by this observation, we choose the design points to be arranged in the following manner: for even $n \geq 10$, $X_1 = 0$, $X_2 - X_1 = 1$, $X_n - X_{n-1} = 1$ and $X_i - X_{i-1} = 10^{-8}$ for $i = 3, ..., n-1$. Then we perform simulations just as what we described at the beginning of the section but we used $N = 10^5$ independent iterations instead of $N = 10^8$. In the following table, we use $\overline{T_n}$ to denote the average of T_n and $\overline{\widehat{T}_n}$ the average of $\widehat{T}_n$ over $N = 10^5$ independent iterations.

Table 1.12: Estimate of $E[T_n]$, $E[\widehat{T}_n]$ and $E[T_n - \widehat{T}_n]$ and $E[|\widehat{b}|]$

n	$\overline{T_n}$	$\overline{\widehat{T}_n}$	$\overline{\widehat{T}_n} - \overline{T_n}$	$\overline{\lvert T_n - \widehat{T}_n\rvert}$	$\overline{\lvert\widehat{b}\rvert}$
10	5.33437	5.39184	0.05747	0.29467	0.56411
20	12.00204	12.06149	0.05945	0.29439	0.56233
50	31.99161	32.05209	0.06048	0.29414	0.56340
100	65.30487	65.36367	0.05880	0.29258	0.56300
500	331.99233	332.04568	0.05335	0.29521	0.56141
1000	665.33015	665.38788	0.05773	0.29010	0.56385

Note that by Theorem 1, for $N = 10^5$ and $n = 1000$ the standard deviation of $\overline{T_n}$ is $\sqrt{\frac{16n-29}{90N}} = 0.04212$ and therefore the difference between $\overline{T_n}$ and $\overline{\widehat{T}_n}$ is only slightly larger than standard deviation of $\overline{T_n}$. It is interesting to see that $\widehat{T}_n$ tends to be larger than T_n.

We know that $\widehat{b}$ has $N(0, \sigma/S_X^2)$ distribution. With the given design points and $\sigma = 1$, $\sigma/S_X^2 \sim 1/2$. Therefore $|\widehat{b}|$ follows a half normal distribution with expectation and variance approximately equal to $\sqrt{1/\pi} = 0.5642$ and $1/2(1-2/\pi) = 0.1817$. The former is consistent with $\overline{|\widehat{b}|}$ obtained from simulations.

1.6 Examples

Example 1. In this example, we consider atmospheric densities measured by Smith et al. (1971). A rocket was launched into the upper atmosphere. Instruments and devices carried by the rocket and on the ground measured elevations of the rocket and the local air density. Data are given in the table accompanying Figure 23 of Smith et al. (1971) for elevations of m km. for $m = 33, 34, \ldots, 130$ from such an experiment conducted in 1969. We chose to analyze the data only for the $n = 48$ measurements up to the altitude of 80 km. This range of altitudes corresponds to the stratosphere and mesosphere. In the paper by Smith et al. (1971, p. 5), the authors estimated that the errors were $\pm 1\%$ for altitudes under 84 km. but became 4% or more at higher elevations. Here we plot the density vs. the altitude.

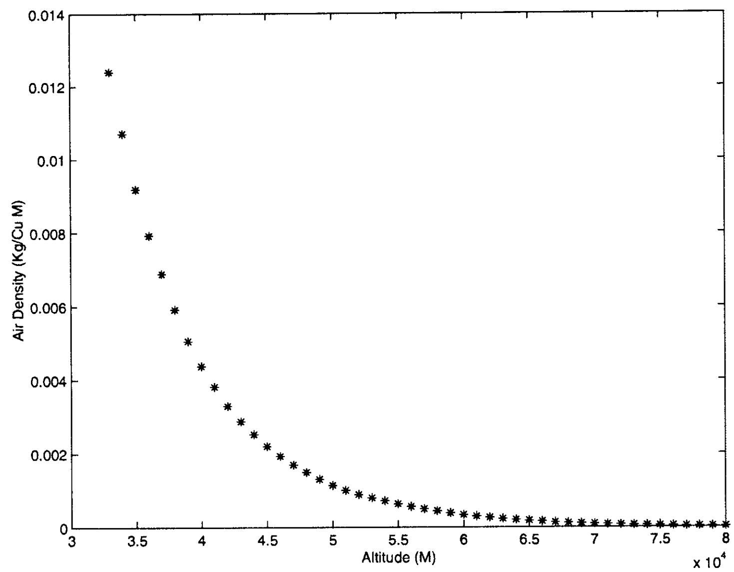

Figure 1-1: Air Density vs Altitude.

As the altitude increases, the density decreases to zero so we would like to explore the relation between log density and the altitude. We first apply linear regression to the log of the air density against the altitude. The residual sum of squares is RSS $= 0.2168$ and the mean square error is MSE $=$ RSS$/(n-2) = 0.0047$. (In general, for a kth order polynomial regression model, MSE is RSS$/(n-k-1)$ and MLE is RSS$/n$.)

The square root of MSE is then 0.0686 which is very small compared to 7.6027, the absolute value of the average log air density. Figure 1-2 also shows that the residuals are very small. However, there are 12 turning points among the residuals while for $n = 48$, the critical value is 22 for the two-sided 0.01 level turning point test and is 23 for the one-sided 0.01 level turning point test. Figure 1-3 also indicates the residuals form a cubic kind of pattern that is first convex and then concave. Therefore there is strong evidence against a linear relation between the log of the air density and the altitude.

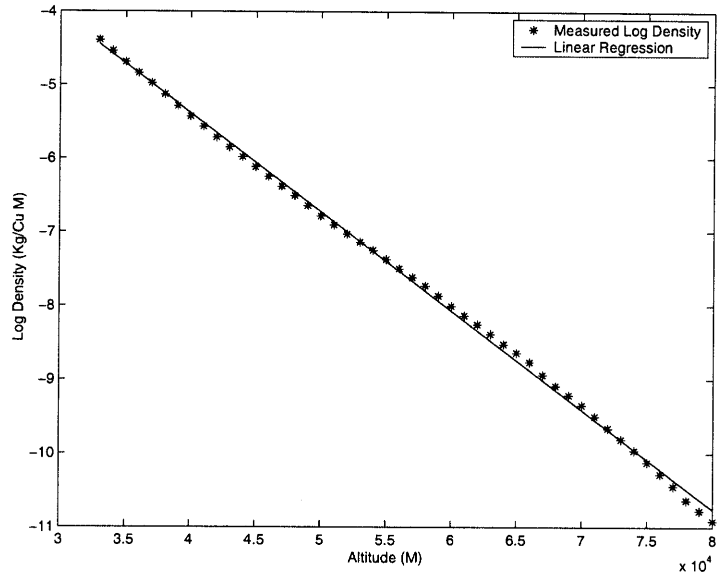

Figure 1-2: Linear Regression of Log Density vs Altitude.

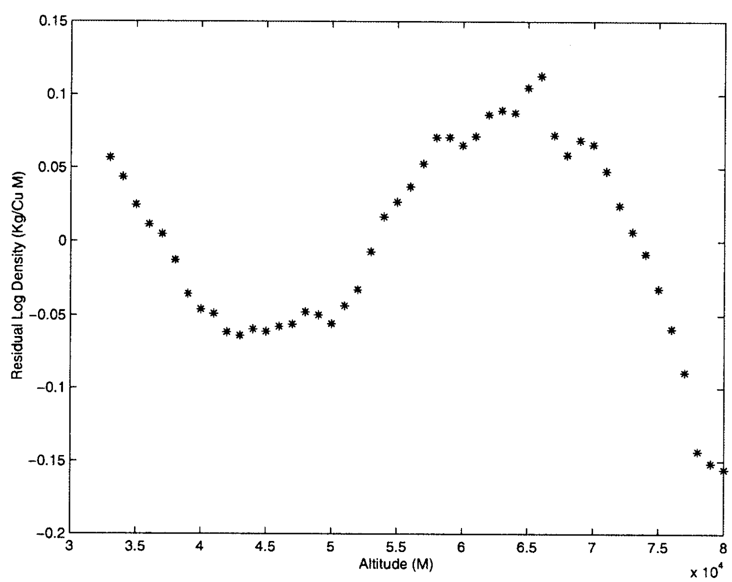

Figure 1-3: Residuals of Linear Regression of Log Density vs Altitude.

Next we apply quadratic regression to the log of the air density against the altitude. The mean square error of the quadratic regression is MSE $=$ RSS$/(n-3) =$ 0.0039 which is slightly less than that of the linear regression (Figure 1-4). We find

9 turning points among the residuals while for $n = 48$, the critical value is 22 for the two-sided 0.01 level turning point test and is 23 for the one-sided 0.01 level turning point test. Moreover, one still sees a clear cubic pattern in the residuals (Figure 1-5). Therefore there is also strong evidence against the log of the air density being a quadratic function of the altitude.

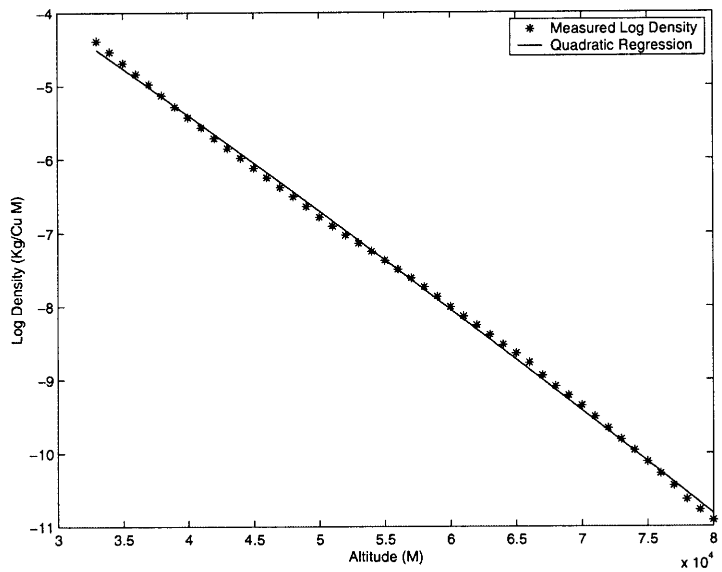

Figure 1-4: Quadratic Regression of Log Density vs Altitude.

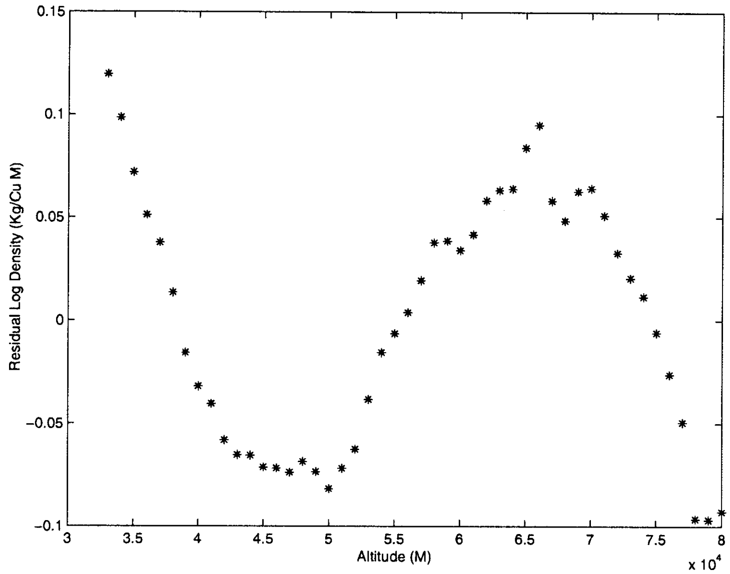

Figure 1-5: Residuals of Quadratic Regression of Log Density vs Altitude.

Then we apply cubic regression to the log of the air density against the altitude. The mean square error is MSE $= 2.0669 \cdot 10^{-4}$ which is much smaller than that of linear and quadratic regression. The residuals are very small (Figure 1-6) and the cubic pattern has disappeared in Figure 1-7, the residual plot.

However, we find 22 turning points among the residuals while for $n = 48$ the critical values is 22 for the two-sided 0.01 level turning point test and is 24 for the two-sided 0.05 level turning point test. Therefore there is also evidence against the log of the air density being a cubic function of the altitude despite the residual plot not showing a very clear pattern.

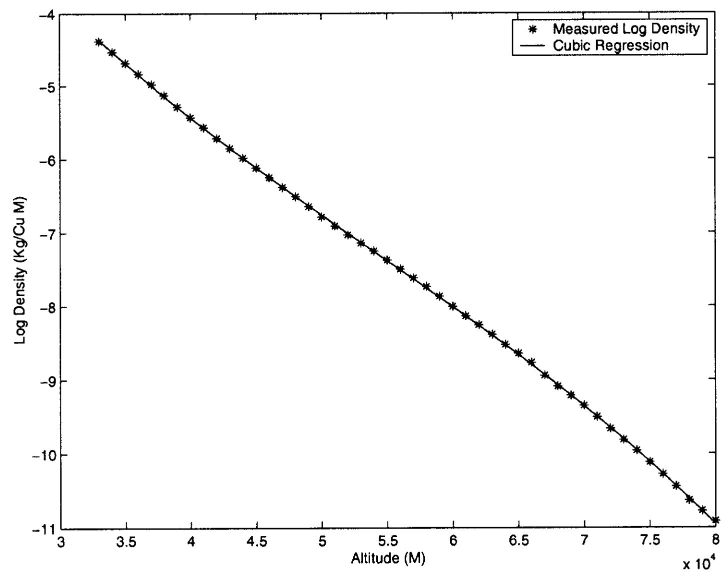

Figure 1-6: Cubic Regression of Log Density vs Altitude.

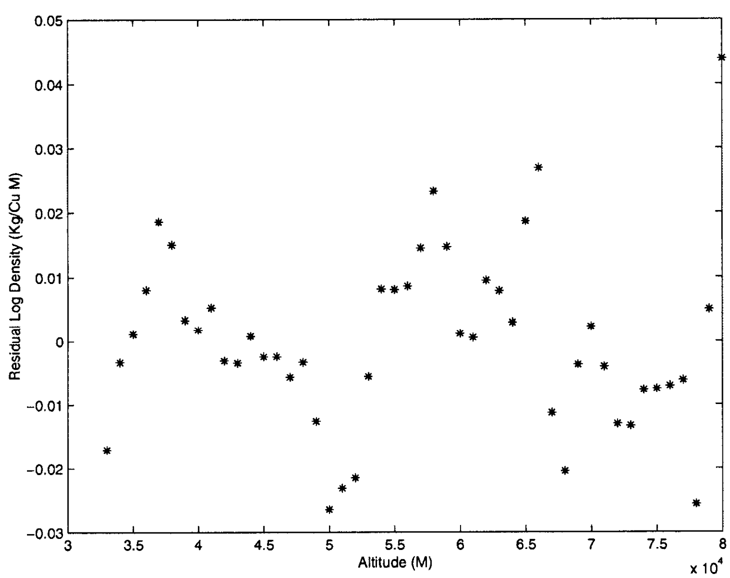

Figure 1-7: Residuals of Cubic Regression of Log Density vs Altitude.

Quartic Regression does not perform well either. The mean square error is MSE $= 2.0103 \cdot 10^{-4}$ which is only slightly less than that of the cubic regression. The number of residual turning points is 20.

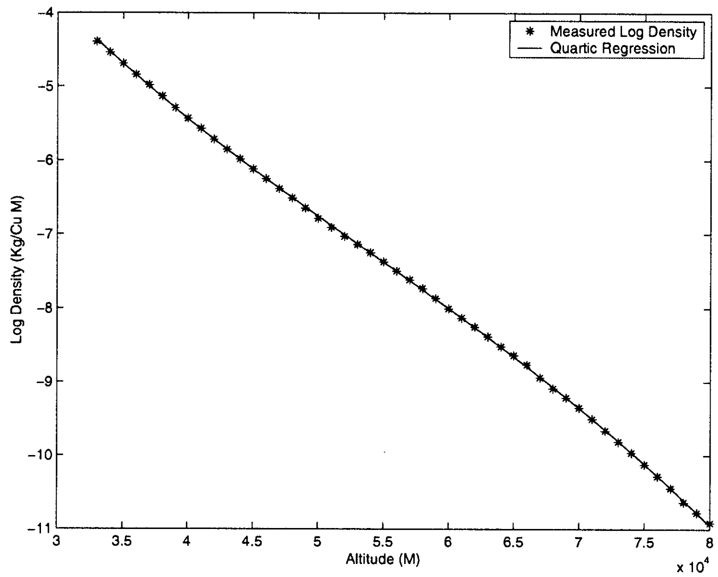

Figure 1-8: Quartic Regression of Log Density vs Altitude.

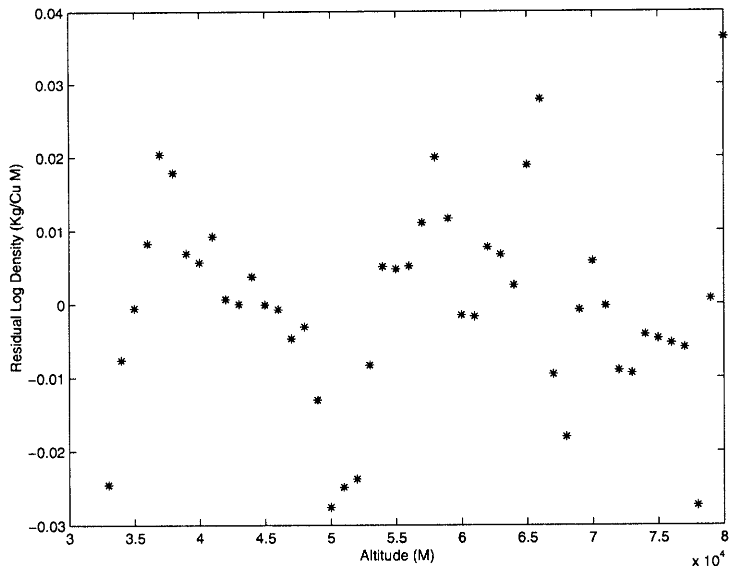

Figure 1-9: Residuals of Quartic Regression of Log Density vs Altitude.

5th order polynomial regression gives 24 residual turning points with MSE $= 1.0337 \cdot 10^{-4}$ and 6th order polynomial regression gives 26 residual turning points with MSE $= 1.0539 \cdot 10^{-5}$. Despite being close to the critical values, they will not be rejected by the 0.01 level turning point test although the 5th order model is rejected by the two-sided and one-sided 0.05 level turning point test. The 5th order polynomial regression model also gives the smallest Bayesian Information Criterion BIC $= -2\log(\text{Maximum Likelihood}) + d\log(n) = n\log(\text{MLE}) + n(1+\log(2\pi)) + d\log(n)$ where d is the number of parameters in the model and is equal to $k+2$ if we use kth order polynomial regression with unknown σ^2.

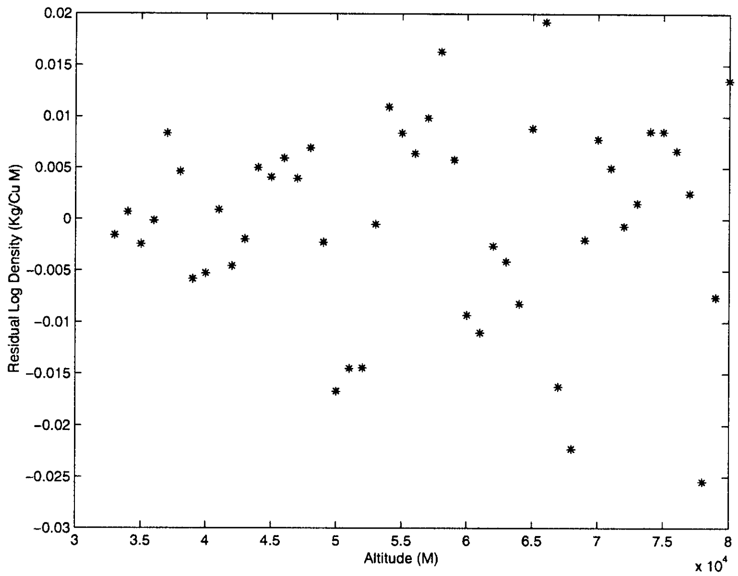

Figure 1-10: Residuals of 5th Order Polynomial Regression of Log Density vs Altitude.

Finally we summarize the results we have obtained so far in the following table.

Table 1.13: Summary

Model	$\widehat{T_n}$	MSE	MLE	BIC
Linear	12	0.0047	0.0045	-111.3724
Quadratic	9	0.0039	0.0036	-118.0933
Cubic	22	$2.0669 \cdot 10^{-4}$	$1.8947 \cdot 10^{-4}$	-255.8482
Quartic	20	$2.0103 \cdot 10^{-4}$	$1.8009 \cdot 10^{-4}$	-254.4123
5th Order	24	$1.0337 \cdot 10^{-4}$	$9.0447 \cdot 10^{-5}$	-283.5991
6th Order	26	$1.0539 \cdot 10^{-4}$	$9.0019 \cdot 10^{-5}$	-279.9559
7th Order	25	$1.0710 \cdot 10^{-4}$	$8.9247 \cdot 10^{-5}$	-276.4978
8th Order	24	$9.9605 \cdot 10^{-5}$	$8.0929 \cdot 10^{-5}$	-277.3227
9th Order	25	$8.7764 \cdot 10^{-5}$	$6.9480 \cdot 10^{-5}$	-280.7733
10th Order	26	$8.9836 \cdot 10^{-5}$	$6.9248 \cdot 10^{-5}$	-277.0624
11th Order	26	$9.2160 \cdot 10^{-5}$	$6.9120 \cdot 10^{-5}$	-273.2803

A well known atmospherical model was published by the US government in 1976 (US Standard Atmosphere). It is also the most recent version the US government has published. In this model, the Earth's atmosphere is divided into multiple layers and in each layer, the air density is modelled approximately as an exponential function in altitude. In particular, there are four zones between the altitude of 33000 M.

and 80000 M. with 47000 M., 51000 M. and 71000 M. defining the boundaries of these zones. Motivated by this model, we also tried fitting cubic splines using least square error with knots at 47000 M., 51000 M. and 71000 M. so the model has 8 parameters which includes the unknown variance of errors. We obtained MSE $= 1.1012 \cdot 10^{-4}$, MLE $= 9.4057 \cdot 10^{-5}$ and BIC $= -277.8496$. The number of residual turning points is 26.

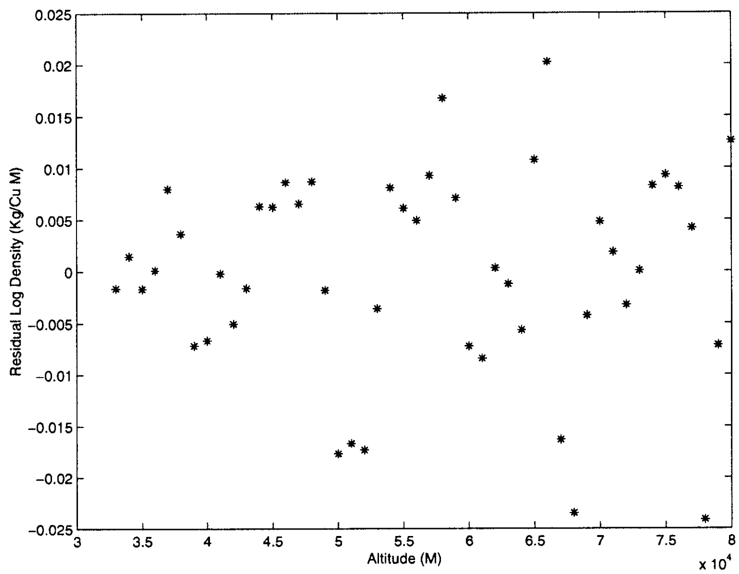

Figure 1-11: Residuals of Cubic Spline Regression of Log Density vs Altitude.

Chapter 2

Power of Turning Point and Convexity Point Tests against Quadratic Alternatives

In this chapter, we investigate the power of the one-sided and two-sided turning point tests and the (two-sided) convexity point test (Dudley and Hua, 2009) against quadratic alternatives. Throughout the chapter, by "one-sided" turning point test we mean the lower one-sided test as in Table 1.3. We will assume that the design points are equally spaced, in which case the turning point statistics of residuals and errors become close quickly (Theorem 6).

The quadratic alternatives will have the following form:

$$Y_i = aX_i^2 + \varepsilon_i, \tag{2.1}$$

where X_i are non-random and ε_i are i.i.d. $N(0, \sigma^2)$. Note that if we let $Y_i' = Y_i/\sigma$ we can write Y_i' as

$$Y_i' = \frac{a}{\sigma}X_i^2 + \frac{\varepsilon_i}{\sigma} = \frac{a}{\sigma}X_i^2 + \varepsilon_i'.$$

where ε_i' are i.i.d. $N(0, 1)$. Since Y_i' is a constant multiple of Y_i, if we apply linear regression to Y_i' the regression coefficients as well as the residuals will also be constant multiples of those obtained from linear regressions for Y_i. Hence it suffices to consider the alternatives where $\sigma = 1$. Furthermore $a(cX_i)^2 = ac^2X_i^2$ so a change of scale in X_i is equivalent to a change of scale in a while keeping X_i fixed. So we can always assume that X_i are equally spaced with distance 1 between consecutive points.

If we were to consider alternatives of the form $Y_i = aX_i^2 + \beta X_i + \gamma + \varepsilon_i$ for constants β and γ then the slope and intercept of the fitted simple linear regression would change but the changes would cancel in finding the residuals, so the residuals and their numbers of turning points or convexity points would be the same as for the model (2.1). Therefore to study the power of our tests it suffices to consider quadratic alternatives in the form of equation (2.1) with varied a and fixed X_i and $\sigma = 1$. We will simply use the numbers $1, 2, \ldots, n$ as our design points.

For fixed a and n, we run $N = 50000$ iterations. In each iteration we generate n i.i.d. $N(0,1)$ points as errors to obtain $Y_i = aX_i^2 + \varepsilon_i$. Then we perform linear regression of Y_i on X_i and obtain residuals. We calculate the number of turning points $\widehat{T}_n$ and the number of convexity points $\check{N}_n$ among the residuals. For $n \leq 50$ we can refer to tables to decide whether to reject the null or not. For $n > 50$ we can use normal approximations to find critical values. If the total number of rejections is r for the turning point test (resp. convexity point test), then $\hat{p} = r/N$ will be an estimate of the power of the test. If the power of our test is too small, say it is less than 0.05, then our test cannot statistically distinguish between the null and alternative at 0.05 level at all. For such small power, there is no need to give an accurate estimate anyway so we may focus on the case where the power $p > 0.05$ in which case $Np \geq 2500$ so we can use the plug-in confidence interval $\hat{p} \pm \zeta\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$ where $\zeta = 1.96$ or 2.576 for a 95% or 99% interval. Since $2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} < 2.52 \cdot 10^{-3}$, our estimate of each power will be accurate to the 2nd decimal place with 99% confidence.

Simulation was carried out for the three tests at 0.01 and 0.05 levels for selected n between 20 and 100 and values of a to be shown. First, the following three tables list the powers of the tests for $n = 20,\ 50,\ 100$.

Table 2.1: Power of One-sided Turning Point Test at 0.01 and 0.05 Levels

n, α	20, 0.01	20, 0.05	50, 0.01	50, 0.05	100, 0.01	100, 0.05
$a = 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$a = 0.5$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$a = 0.4$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$a = 0.3$	0.9979	0.9982	1.0000	1.0000	1.0000	1.0000
$a = 0.2$	0.9473	0.9499	1.0000	1.0000	1.0000	1.0000
$a = 0.1$	0.3934	0.4541	1.0000	1.0000	1.0000	1.0000
$a = 0.05$	0.0590	0.1110	0.9428	0.9849	1.0000	1.0000
$a = 0.04$	0.0337	0.0716	0.7633	0.8988	1.0000	1.0000
$a = 0.03$	0.0186	0.0504	0.4133	0.6350	1.0000	1.0000
$a = 0.02$	0.0098	0.0358	0.1095	0.2639	0.9916	0.9968
$a = 0.01$	0.0069	0.0278	0.0155	0.0669	0.3262	0.4865
$a = 0.005$	0.0058	0.0261	0.0065	0.0392	0.0395	0.0949

Table 2.2: Power of Two-sided Turning Point Test at 0.01 and 0.05 Levels

n, α	20, 0.01	20, 0.05	50, 0.01	50, 0.05	100, 0.01	100, 0.05
$a = 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$a = 0.5$	0.9991	1.0000	1.0000	1.0000	1.0000	1.0000
$a = 0.4$	0.9940	0.9999	1.0000	1.0000	1.0000	1.0000
$a = 0.3$	0.9563	0.9981	1.0000	1.0000	1.0000	1.0000
$a = 0.2$	0.7368	0.9485	1.0000	1.0000	1.0000	1.0000
$a = 0.1$	0.1448	0.3956	1.0000	1.0000	1.0000	1.0000
$a = 0.05$	0.0128	0.0655	0.9376	0.9817	1.0000	1.0000
$a = 0.04$	0.0079	0.0438	0.7353	0.8808	1.0000	1.0000
$a = 0.03$	0.0057	0.0317	0.3565	0.5788	1.0000	1.0000
$a = 0.02$	0.0049	0.0293	0.0758	0.2005	0.9807	0.9926
$a = 0.01$	0.0043	0.0273	0.0090	0.0423	0.2387	0.3845
$a = 0.005$	0.0046	0.0276	0.0064	0.0266	0.0234	0.0639

Table 2.3: Power of Convexity Point Test at 0.01 and 0.05 Levels

n, α	20, 0.01	20, 0.05	50, 0.01	50, 0.05	100, 0.01	100, 0.05
$a = 1$	0.8932	0.9752	0.9999	1.0000	1.0000	1.0000
$a = 0.9$	0.8254	0.9527	0.9993	0.9999	1.0000	1.0000
$a = 0.8$	0.7303	0.9102	0.9972	0.9994	1.0000	1.0000
$a = 0.7$	0.6018	0.8388	0.9858	0.9964	1.0000	1.0000
$a = 0.6$	0.4582	0.7382	0.9467	0.9822	0.9998	1.0000
$a = 0.5$	0.3058	0.5973	0.8383	0.9281	0.9962	0.9996
$a = 0.4$	0.1730	0.4335	0.6209	0.7847	0.9591	0.9926
$a = 0.3$	0.0757	0.2672	0.3306	0.5284	0.7583	0.9219
$a = 0.2$	0.0246	0.1327	0.1066	0.2332	0.3355	0.6197
$a = 0.1$	0.0062	0.0526	0.0178	0.0571	0.0529	0.1936
$a = 0.05$	0.0029	0.0362	0.0056	0.0239	0.0138	0.0737

Overall, the tables show that for fixed n and $a \leq 1$, both one-sided and two-sided turning point tests are more powerful than the convexity point test. To compare the power of the convexity point test with the one-sided turning point test, we choose n such that the probability of rejecting the null is approximately the same for both tests. Let $p_c(n)$ be the type one error of the convexity point test and $p_1(n)$ the type one error of the one-sided test. These include $n = 34, 47$ at the 0.05 level where $p_1(34) - p_c(34) = 0.0009$ and $p_1(47) - p_c(47) = 0.0007$ and $n = 31, 38$ for the 0.01 level where $p_1(31) - p_c(31) = -0.0003$ and $p_1(38) - p_c(38) = 0.0001$.

Table 2.4: Power Comparison of Convexity Point Test and One-sided Turning Point Tests at Given Levels

Convexity Point vs. One-sided Turning Point				
n, α	34, 0.05	47, 0.05	31, 0.01	38, 0.01
$a = 1.0$	.9989 1.0000	1.0000 1.0000	.9934 1.0000	.9989 1.0000
$a = 0.9$	.9967 1.0000	.9999 1.0000	.9833 1.0000	.9955 1.0000
$a = 0.8$	.9891 1.0000	.9994 1.0000	.9573 1.0000	.9855 1.0000
$a = 0.7$	.9713 1.0000	.9962 1.0000	.9031 1.0000	.9544 1.0000
$a = 0.6$	.9207 1.0000	.9843 1.0000	.8024 1.0000	.8828 1.0000
$a = 0.5$	.8167 1.0000	.9384 1.0000	.6366 1.0000	.7393 1.0000
$a = 0.4$	.6363 1.0000	.8087 1.0000	.4248 1.0000	.5189 1.0000
$a = 0.3$	.3985 1.0000	.5717 1.0000	.2204 1.0000	.2762 1.0000
$a = 0.2$	.1808 1.0000	.2772 1.0000	.0803 1.0000	.0943 1.0000
$a = 0.1$	.0547 .9912	.0831 1.0000	.0187 .9501	.0212 .9985
$a = 0.05$	.0289 .5463	.0397 .9588	.0081 .3218	.0076 .6254
$a = 0.04$	.0257 .3318	.0350 .8169	.0079 .1559	.0069 .3703
$a = 0.03$	.0211 .1696	.0320 .5185	.0067 .0631	.0069 .1506
$a = 0.02$	.0203 .0716	.0269 .2004	.0055 .0226	.0051 .0413
$a = 0.01$	.0203 .0316	.0248 .0509	.0061 .0074	.0054 .0095
$a = 0.005$	.0205 .0250	.0251 .0324	.0054 .0051	.0045 .0059

Similarly, to compare the power of the convexity point test with the two-sided turning point test, we take n such that the probability of rejecting the null is approximately the same for both tests. Let $p_2(n)$ be the type one error of the two-sided test. These include $n = 20, 40$ at the 0.05 level where $p_2(40) - p_c(40) = 0.0000$ and $p_2(20) - p_c(20) = -0.0010$, and $n = 24, 38$ for the 0.01 level where $p_2(24) - p_c(24) = 0.0001$ and $p_2(38) - p_c(38) = 0.0005$.

Table 2.5: Power Comparison of Convexity Point Test and Two-sided Turning Point Test at Given Levels

Convexity Point vs. Two-sided Turning Point				
n, α	20, 0.05	40, 0.05	24, 0.01	38, 0.01
$a = 1.0$	.9752 1.0000	.9999 1.0000	.9686 1.0000	.9989 1.0000
$a = 0.9$	.9527 1.0000	.9996 1.0000	.9389 1.0000	.9955 1.0000
$a = 0.8$	.9102 1.0000	.9980 1.0000	.8834 1.0000	.9855 1.0000
$a = 0.7$	.8388 1.0000	.9918 1.0000	.7905 1.0000	.9544 1.0000
$a = 0.6$	.7382 1.0000	.9697 1.0000	.6609 1.0000	.8828 1.0000
$a = 0.5$	.5973 1.0000	.9070 1.0000	.4901 1.0000	.7393 1.0000
$a = 0.4$	.4335 .9999	.7695 1.0000	.3082 1.0000	.5189 1.0000
$a = 0.3$	.2672 .9981	.5420 1.0000	.1531 1.0000	.2762 1.0000
$a = 0.2$	.1327 .9485	.2738 1.0000	.0567 .9951	.0943 1.0000
$a = 0.1$	.0526 .3956	.0887 .9998	.0147 .6397	.0212 .9892
$a = 0.05$	.0362 .0655	.0475 .7902	.0067 .1053	.0076 .4193
$a = 0.04$	.0322 .0438	.0419 .5480	.0061 .0519	.0069 .2036
$a = 0.03$	.0319 .0317	.0387 .2701	.0057 .0250	.0069 .0671
$a = 0.02$	.0301 .0293	.0365 .0906	.0056 .0123	.0051 .0151
$a = 0.01$	.0285 .0273	.0346 .0387	.0054 .0058	.0054 .0055
$a = 0.005$	.0273 .0276	.0332 .0340	.0051 .0049	.0045 .0055

To compare the power of the one-sided and two-sided turning point tests, we choose n in the same way. Let $p_2(n)$ be the type one error of the two-sided test. For $n = 37, 48$ at the 0.05 level we have $p_1(37) - p_2(37) = -0.0003$ and $p_1(48) - p_2(48) = 0.0010$. For $n = 29, 38$ at the 0.01 level we have $p_1(29) - p_2(29) = 0.0001$ and $p_1(38) - p_2(38) = -0.0004$.

Table 2.6: Power Comparison of One-sided and Two-sided Turning Point Tests at Given Levels

One-sided Power vs. Two-sided Power				
n, α	37, 0.05	48, 0.05	29, 0.01	38, 0.01
$a = 0.3$	1.0000 1.0000	1.0000 1.0000	1.0000 1.0000	1.0000 1.0000
$a = 0.2$	1.0000 1.0000	1.0000 1.0000	.9999 .9998	1.0000 1.0000
$a = 0.1$	.9984 .9985	1.0000 1.0000	.8456 .8344	.9985 .9892
$a = 0.05$	.6976 .6495	.9837 .9529	.2006 .1518	.6254 .4193
$a = 0.04$	.4680 .4061	.9001 .7998	.0991 .0657	.3703 .2036
$a = 0.03$	.2449 .1825	.6448 .4809	.0436 .0240	.1506 .0671
$a = 0.02$	.0996 .0633	.2790 .1656	.0172 .0087	.0413 .0151
$a = 0.01$	.0397 .0320	.0777 .0461	.0085 .0059	.0095 .0055
$a = 0.005$	.0298 .0282	.0461 .0353	.0065 .0060	.0059 .0055

By comparing the above two tables, we see that the one-sided turning point test is more powerful than the two-sided turning point test when they have similar type one errors while the convexity point test is the least powerful. In particular, this trend holds for $n = 38$, level 0.01, where all three tests have similar type one error probabilities.

We also did some simulation for $n = 1000$ but with a much smaller $N = 5000$. For both one-sided and two-sided turning point tests, the power stays close to 1 for $a \geq 0.001$ and then starts decreasing. For $a = 0.0005$, the power is about 0.18 for the 0.01 level tests and 0.36 for the 0.05 level tests. For $a = 0.0001$, the power is about 0.01 for the 0.01 level test and 0.05 for the 0.05 level test. For the convexity point test, the power stays close to 1 for $a \geq 0.1$ and then starts decreasing. For $a = 0.05$, the power is about 0.22 for the 0.01 level test and 0.43 for the 0.05 level test. For $a = 0.01$, the power is about 0.01 for the 0.01 level test and 0.05 for the 0.05 level test.

The tests have their powers increasing as n gets larger for any fixed a. For fixed n, their powers decrease as a gets smaller (there is an exception for the turning point test with $n = 20$ but the difference is at the 4th decimal place which cannot be estimated accurately with $N = 50000$). For the same a the two turning point tests always have more power. Both turning point and convexity point tests have critical values for a such that if a is larger than the critical value the power is close to 1 but if a is below the critical value the power decreases very fast.

Bibliography

[1] Berk, K. N. (1973), A central limit theorem for m-dependent random variables with unbounded m, *Ann. Probability* **1**, 352-354.

[2] Bonferroni, C. E. (1936), Teoria statistica delle classi e calcolo delle probabilità, *Pubblicazioni del R. Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **8**, 1-62.

[3] Box, G. E. P and Muller, M. E. (1958), A note on the generation of random normal deviates, *Ann. Math. Statist* **29**, 610-611.

[4] Cheng, M. C. (1969), The orthant probabilities of four Gaussian variates, *Ann. Math. Statist* **40**, 152-161.

[5] Diananda, P. H. (1955), The central limit theorem for m-dependent variables, *Proc. Camb. Philos. Soc.* **51**, 92-95.

[6] Dudley, R. M. (2002), *Real Analysis and Probability*, 2d ed., Cambridge University Press.

[7] Dudley, R. M., and Hua, X., Testing independence and normality of regression residuals via turning points and convexity points. Unpublished manuscript.

[8] Galambos, J. and Simonelli, I. (1996). *Bonferroni-type Inequalities with Applications*, Springer.

[9] Hoeffding, W., and Robbins, H. (1948), The central limit theorem for dependent random variables, *Duke Math. J.* **15**, 773-780.

[10] Kendall, M., Stuart, A., and Ord, J. K. (1983). *The Advanced Theory of Statistics*, vol. 3, *Design and Analysis, and Time-Series*, 4th ed., MacMillan, New York.

[11] Lüscher M. (1994), A portable high-quality random number generator for lattice field theory calculations, *Computer Physics Communications* **79**, 100-110.

[12] Orey, S. A. (1958), Central limit theorem for m-dependent random variables, *Duke Math. J.* **25**, 543-546.

[13] Plackett, R. L. (1954), A reduction formula for normal multivariate integrals, *Biometrika* **41**, 351-360.

[14] Smith, W. S., Theon, J. S., Casey J. F., Azcárraga A. and Horvath, J. J. (1971), Temperature, pressure, density and wind measurements in the stratosphere and mesosphere, 1969, *NASA technical report* TR R-360, available on NASA Technical Reports Server.

[15] Stanley, R. P. (2008) Longest alternating subsequences of permutations, *Michigan Math. J.* **57**, 675-687.

[16] Stuart, A. (1954). Asymptotic relative efficiencies of distribution-free tests of randomness against normal alternatives. *J. Amer. Statist. Assoc.* **49**, 147-157.

[17] U. S. Standard Atmosphere, 1976, U.S. Government Printing Office, Washington, D.C.