Testing Normality in Designs With Many Parameters

Richard A. LOCKHART

Department of Mathematics and Statistics
Simon Fraser University
Burnaby, BC V5A 1S6, Canada

Chandanie W. PERERA

Department of Mathematics and Computer Science
The Open University of Sri Lanka
Nawala, Nugegoda, Sri Lanka

Goodness-of-fit tests are proposed for the assumption of normality of random errors in experimental designs where the variance of the response may vary with the levels of the covariates. The exact distribution of standardized residuals is used to make the probability integral transform for use in tests based on the empirical distribution function. A different mean and variance is estimated for each level of the covariate; corresponding large sample theory is provided. The proposed tests are robust to a possible misspecification of the model and permit data collected from several similar experiments to be pooled to improve the power of the test.

KEY WORDS: Anderson–Darling test; Cramér–von Mises test; Empirical distribution function; Empirical process; Goodness of fit.

1. INTRODUCTION

Testing the normality of random errors ϵ_i in the regression model, $y_i = \mu_i + \sigma_i\epsilon_i$, is needed in many applications. In situations where the mean function μ_i has a certain hypothesized form (such as $x_i\beta$ in the linear regression model) and the variance of the response σ_i is a known functional form of μ_i, the usual practice (see Pierce and Kopecky 1979) is to compute the standardized fitted residuals, $\tilde{\epsilon}_i = (y_i - \hat{\mu}_i)/\hat{\sigma}_i$, and test the hypothesis of normal errors by examining whether the $\tilde{\epsilon}_i$'s are approximately normally distributed. An important class of tests based on the empirical distribution function of the probability integral transforms u_i's of the residuals computed using the standard normal distribution Φ exists for this case. Asymptotic results for the empirical distribution function of the u_i's are available when the number of parameters p is fixed as n grows (see Stephens 1976) or grows slightly (see Mammen 1996; Chen and Lockhart 2001).

These large-sample results can provide poor approximations in several contexts. First, when nonlinear response functions are estimated, the fitted residuals need not be normally distributed, even if the random errors are. Second, when acquisition of new data requires fitting more parameters, the fitted values $\hat{\mu}_i$ need not be consistent; in this case, the variance of $y_i - \hat{\mu}_i$ may be seriously underestimated by $\hat{\sigma}_i^2$. Third, if the assumed model for the mean or for the mean variance relation is incorrect, the test of normality becomes confounded with a test of model specification.

All these problems may be addressed when the experiment is structured with replicate observations, that is, when there are several observations y_{ij}, $j = 1, \ldots, n_i$, for each value x_i of the covariates. In this case, we propose to fit different means and variances at each level of the covariate and to compute the exact probability integral transforms, u_i's, using the true distribution of the resulting residuals. The assumption of normality can then be tested by examining whether the resulting transforms have Uniform[0, 1] distributions. We provide asymptotic results for the empirical distribution function of the u_i's in this case.

Because the proposed tests allow fitting different mean functions at different levels of the covariate, they possess the advantage that the data collected from several similar experiments can be pooled to improve the power of the test. Thermoluminescence sedimentary dating provides a good example of such an application. In thermoluminescence sedimentary dating, several subsamples are prepared from a core drilled from a sedimentary deposit such as a sand that is to be dated. One of two possible pretreatments is applied to each subsample; the subsample is then exposed to a dose, d, of gamma radiation. For each combination of dose and pretreatment, a small number of subsamples are prepared. Each subsample is heated and the amount, y, of light given off (thermoluminescence) is measured. Different nonlinear mean functions (or the same mean function but with different parameter values) relating the amount of thermoluminescence, y, to the amount of radiation, d, to which a subsample is exposed are fitted for different cores. The fitted mean functions are usually nonlinear and depend on at least three parameters. Moreover, only 15–20 observations are available to estimate these parameters for a given core.

With such small datasets and nonlinear mean functions the asymptotic approximations cited previously cannot be expected to be very good. To obtain a larger dataset to test normality, it is necessary to collect a new core and fit new parameter values. Thus, the number of parameters fitted will grow linearly with the total sample size and the parameter estimates themselves will not be consistent. In turn, it will not be the case that the fitted residuals $y - \hat{\mu}$ are asymptotically close to the underlying residuals $\sigma\epsilon$. As a consequence, the sizes of the usual tests of normality based on normal probability integral transforms of the residuals may be substantially different from the corresponding nominal levels.

In this article we overcome these problems by fitting, for each combination, x_i, of dose and pretreatment, separate mean and variance estimates. We compute the true distribution of the residuals from this model fit. We use this true distribution to make an exact probability integral transform of the residuals to produce a set of observations that will be Uniform[0, 1] if the null hypothesis of normal errors holds. The empirical process of

TECHNOMETRICS, AUGUST 2006, VOL. 48, NO. 3
DOI 10.1198/004017006000000147

these exact transforms has a covariance function that we compute exactly. In large samples, moreover, the empirical process in question will be approximately Gaussian. This permits us to give tests of the hypothesis of normal errors. Furthermore, fitting different means and variances for each covariate combination ensures that the tests proposed in this article will have the correct level, as tests of the hypothesis of normal errors, even if the original nonlinear response models are incorrect.

The rest of this article is organized as follows. In Section 2 we propose tests based on the Anderson–Darling statistic and the Cramér–von Mises statistic computed from the exact probability integral transforms of the fitted residuals and discuss computation of the test statistics. Section 3 proposes an approximate p value for testing the assumption of normality based on each of the proposed statistics. In Section 4 we describe the results of a simulation study carried out to assess the performance of the suggested tests in finite samples. Monte Carlo critical points that can be used in cases with equal number of replicates at each level of the covariate are also offered in Section 4. The tests we propose are valid in the more general context of unequal number of replicates as well. We have developed software using S–PLUS to compute the test statistics and to produce approximate p values for assessing normality in unbalanced designs. In Section 5 we use published thermoluminescence dating test data to illustrate the method. The weak convergence result for the empirical process of the transformed residuals that justifies the use of the approximate p values in large samples is outlined in the Appendix.

2. COMPUTATION OF THE TEST STATISTICS

We now focus on the general case of fitting different means at different levels of the covariate. We begin by introducing the notation and presenting the true distribution of the residuals needed in computing the test statistics.

Consider the model $y_{ij} = \mu_i + \sigma_i \epsilon_{ij}$, where we let $i = 1, \ldots, k$ denote the level of the covariate and j denote the replicate. Suppose n_i observations are available at the ith level of the covariate. We assume that each $n_i \geq 3$; because there are two parameters to fit for each i, levels i with $n_i = 1$ or $n_i = 2$ do not provide information about the normal assumption. Let $\hat{\mu}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$ be the least squares estimate for μ_i and $\hat{\epsilon}'_{ij} = Y_{ij} - \hat{\mu}_i$. We study the standardized fitted residuals $\hat{\epsilon}_{ij} = (Y_{ij} - \hat{\mu}_i)/\hat{\sigma}_i$, where $\hat{\sigma}_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2/n_i$. Let $G_{n_i}(\cdot)$ be the true distribution of $\hat{\epsilon}_{ij}$ when the random errors are normally distributed; $G_{n_i}(\cdot)$ depends on n_i but not on μ_i, σ_i, or j. Let $\nu_i = n_i - 1$ and $\hat{\tau}_{ij} = \hat{\epsilon}_{ij}\sqrt{(\nu_i - 1)/(\nu_i - \hat{\epsilon}_{ij}^2)}$. According to Beckman and Trussell (1974), when the ϵ_{ij} follow a standard normal distribution, the variates $\hat{\tau}_{ij}$ follow a univariate Student-t distribution with degrees of freedom $\nu_i - 1$. The exact probability integral transforms of the $\hat{\epsilon}_{ij}$ are, therefore, given by

$$u_{ij} = G_{n_i}(\hat{\epsilon}_{ij}) = t_{\nu_i - 1}\left(\hat{\epsilon}_{ij}\sqrt{\frac{\nu_i - 1}{\nu_i - \hat{\epsilon}_{ij}^2}}\right).$$

We now outline the procedure for computing the test statistics.

1. For each level i of the covariate, estimate μ_i and σ_i^2 using $\hat{\mu}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$ and $\hat{\sigma}_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2/n_i$. Retain only levels i with $n_i \geq 3$.
2. Compute the standardized fitted residuals $\hat{\epsilon}_{ij} = (Y_{ij} - \hat{\mu}_i)/\hat{\sigma}_i$.
3. Compute the exact probability integral transforms $u_{ij} = G_{n_i}(\hat{\epsilon}_{ij}) = t_{\nu_i - 1}(\hat{\epsilon}_{ij}\sqrt{(\nu_i - 1)/(\nu_i - \hat{\epsilon}_{ij}^2)})$, where t_ν denotes the Student-t distribution function on ν degrees of freedom.
4. Let $z_1 < \cdots < z_n$ be the probability integral transforms, u_{ij}, sorted into increasing order, where $n = \sum n_i$ is the total number of observations.
5. Compute the Cramér–von Mises statistic

$$W_n^2 = \sum_{l=1}^{n} \left\{ z_l - \frac{2l-1}{2n} \right\}^2 + \frac{1}{12n}$$

or the Anderson–Darling statistic

$$A_n^2 = -n - \frac{1}{n}\sum_{l=1}^{n} \{(2l-1)\ln z_l + (2n+1-2l)\ln(1-z_l)\}.$$

3. COMPUTATION OF AN APPROXIMATE p VALUE

In this section we describe the computation of approximate p values for the test statistics by two methods. The first method is Monte Carlo based; the second is a large-sample approximation.

3.1 Monte Carlo p Values

In each cell, we are fitting a location–scale model. As a result, the distribution of our test statistics does not depend on the unknown values of μ_i or σ_i. We may then compute a p value by a simple (though perhaps somewhat time consuming) Monte Carlo method. Pick some large number of replicates M. Generate for each m from 1 to M a set of independent N(0, 1) variables, ϵ_{ij}^*, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$. From the ϵ_{ij}^*, compute the values w_m^* or a_m^* of the Cramér–von Mises or Anderson–Darling statistics, respectively. The desired approximate p value is simply the fraction of values of w_m^* or a_m^* that exceed the observed value of the corresponding statistic for the data at hand. The limit, as $M \to \infty$, of the p value obtained is an exact p value; that is, it has exactly a Uniform[0, 1] distribution on the null. For fixed M, the p value obtained is uniformly distributed on the numbers $0/M, 1/M, \ldots, M/M$ if the null hypothesis is correct.

Use of a small value of M produces an approximation to the exact p value obtained in the limit $M \to \infty$. For a correct p value of .05, the Monte Carlo standard error is around .007 when $M = 1{,}000$; this figure might usefully be compared with the approximation error in the large-sample approximation suggested in the next section.

3.2 Large-Sample Approximate p Values

An alternative to the Monte Carlo method is provided by large-sample approximation. The approximation is based on the

representation of the statistics as

$$W^2 = \int_0^1 W_n^2(t)\,dt,$$

$$A^2 = \int_0^1 \frac{W_n^2(t)}{t(1-t)}\,dt,$$

where W_n is the empirical process $W_n(s) = n^{-1/2}\sum_{l=1}^n \{I[z_i \le s] - s\}$. Large-sample distribution theory, presented in the Appendix, shows that these statistics may be treated, in large samples, as if they had the same law as

$$\sum_{i=1}^{\infty} \lambda_i \chi_i^2,$$

where the χ_i^2's denote a set of independent chi-squared random variables each on 1 degree of freedom and the λ_i's are eigenvalues of an integral equation $\int_0^1 \alpha_n(s,t)f(t)\,dt = \lambda f(s)$. For the statistic W^2, the kernel $\alpha_n(s,t)$ is the covariance function of the process W_n, namely,

$$\alpha_n(s,t) = \min(s,t) - st + \frac{1}{n}\sum_i n_i(n_i-1)\left\{G_{2,n_i}\left(G_{n_i}^{-1}(s), G_{n_i}^{-1}(t)\right) - st\right\},$$

where G_r^{-1} is the inverse function of the true distribution of a standardized residual in a cell with r observations and $G_{2,r}$ is the joint distribution of two such standardized residuals. That is,

$$G_{2,n_i}(x,y) = P(\hat{\epsilon}_{ij} \le x, \hat{\epsilon}_{ij'} \le y)$$

for any two distinct observations $j \ne j'$ in cell i.

An approximate p value for testing the assumption of normality using W^2 can, thus, be computed as $P(\sum_{i=1}^m \lambda_i \chi_i^2 \ge w)$, where w denotes the value of the test statistic and the λ_i's are numerical estimates for the largest m eigenvalues of the covariance kernel $\alpha_n(s,t)$ for a suitable value of m (we usually use $m = 100$). Once the λ_i's have been calculated, this probability can be computed by following Imhof's (1961) method of numerical Fourier inversion of the characteristic function of a linear combination of chi-squares.

Approximations for the required eigenvalues can be computed as (see Lockhart, O'Reilly, and Stephens 1986) eigenvalues $\lambda_1, \ldots, \lambda_m$ of the matrix $\mathbf{Q}$ whose elements are $Q(i,j) = \alpha_n(s_i, s_j)/m$, where $s_i = (i - .5)/m$ for $i = 1, \ldots, m$.

For the Anderson–Darling statistic, the covariance $\alpha_n(s,t)$ in the foregoing must be replaced by

$$\alpha_{A,n}(s,t) = \frac{\alpha_n(s,t)}{\sqrt{s(1-s)t(1-t)}}.$$

It remains to show how to compute the joint distribution function $G_{2,n}$. This is presented in the next section.

3.3 Joint Distribution Function of Two Fitted Residuals

In this section we show how to compute $G_{2,n}$, the joint cumulative distribution function of two residuals $\hat{\epsilon}_i$ and $\hat{\epsilon}_j$, for an iid sample of size n. Our calculations use results of Ellenberg (1973), who provided the joint density of the standardized residuals for the linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta}$ is a p-dimensional vector of unknown parameters and $\mathbf{X}$ is fixed and of full rank. Let $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. Let $\hat{u}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$, where $\mathbf{x}_i$ is the ith row of $\mathbf{X}$ and $\hat{\boldsymbol{\beta}}$ is the least squares estimate for $\boldsymbol{\beta}$. Let m_{ii} be the ith diagonal element of $\mathbf{M}$ and $\zeta_i = \hat{u}_i/\sqrt{m_{ii}\sum_{i=1}^n \hat{u}_i^2}$. Let $\rho = m_{ij}/\sqrt{m_{ii}m_{jj}}$ and $\gamma = (n-p-2)/2$. For $n > p+2$, Ellenberg (1973) gave the joint density of ζ_i and ζ_j as

$$g(u,v;\rho,\gamma) = \frac{\gamma}{\pi}\frac{1}{\sqrt{1-\rho^2}}\left\{1 - \frac{u^2 - 2\rho uv + v^2}{1-\rho^2}\right\}^{\gamma-1},$$

over the region $u^2 - 2\rho uv + v^2 \le 1 - \rho^2$.

We now follow closely the work of Dunnett and Sobel (1954) to derive a computing formula for the joint cumulative distribution function $G(\cdot,\cdot;\rho,\gamma)$ corresponding to $g(\cdot,\cdot;\rho,\gamma)$, that is, for

$$G(\cdot,\cdot;\rho,\gamma) = \int_{-\infty}^h \int_{-\infty}^k g(u,v;\rho,\gamma)\,du\,dv.$$

Note that g is a density for all $-1 < \rho < 1$ and $\gamma > 0$. Here, as in Ellenberg (1973), however, we need formulas only for γ an integer multiple of $1/2$. Our goodness-of-fit application needs only the special case where $\mathbf{X}$ is a column of n ones, $p = 1$, $\rho = -1/(n-1)$, $\gamma = (n-3)/2$, and $\hat{\epsilon}_i = \zeta_i\sqrt{n-1}$, whence

$$G_{2,n}(u,v) = G\left(\frac{u}{\sqrt{n-1}}, \frac{v}{\sqrt{n-1}}; -\frac{1}{n-1}, \frac{n-3}{2}\right). \quad (1)$$

Our formula is a recursion in γ. For γ an integer multiple of $1/2$, we give an explicit form for the starting value for the recursion but not for general γ. Note that if $n = p+2$ the joint distribution of ζ_i and ζ_j is singular; we provide a separate evaluation of $G(\cdot,\cdot;\rho,0)$ at the end of this section, which is needed for the case $n = 3$.

First, note that for nonpositive values of h and k, the joint cumulative distribution function $G(\cdot,\cdot;\rho,\gamma)$ can be obtained from integrals for positive values of h and k as follows:

$$G(h,k;\rho,\gamma) = \begin{cases} G(h,1;\rho,\gamma) - G(h,|k|;-\rho,\gamma), \\ \qquad h \ge 0 \text{ and } k < 0 \\ G(1,k;\rho,\gamma) - G(|h|,k;-\rho,\gamma), \\ \qquad h < 0 \text{ and } k \ge 0 \\ 1 - G(1,|k|;-\rho,\gamma) \\ \quad - G(|h|,1;-\rho,\gamma) + G(|h|,|k|;\rho,\gamma), \\ \qquad h < 0 \text{ and } k < 0 \\ \dfrac{1}{4} + \dfrac{1}{2\pi}\arctan\left(\dfrac{\rho}{\sqrt{1-\rho^2}}\right), \\ \qquad h = 0 \text{ and } k = 0. \end{cases}$$

Thus, it suffices to compute the joint distribution function for positive values of h and k.

If U_1, U_2 have joint density g, then

$$G(h,k;\rho,\gamma) = 1 - P(U_1 > h, U_1 > hU_2/k) - P(U_2 > k, U_2 > kU_1/h).$$

If we put $H_\gamma(h,k,\rho) = P(U_1 > h, U_1 > hU_2/k)$, then, by symmetry,

$$G(h,k;\rho,\gamma) = 1 - H_\gamma(h,k,\rho) - H_\gamma(k,h,\rho).$$

Define new variables R and Θ by $U_1 = R\cos\Theta$ and $(U_2 - \rho U_1)/\sqrt{1-\rho^2} = R\sin\Theta$, where $0 \le R \le 1$ and $-\pi < \Theta \le \pi$. It is then elementary algebra to check that

$$H_\gamma(h,k,\rho) = P\{R\cos\Theta \ge h, \theta_\ell(h) \le \tan\Theta \le \theta_u(h,k,\rho)\},$$

where

$$\theta_\ell(h) = -\arccos(h)$$

and

$$\theta_u(h,k,\rho) = \max\left[\min\left\{\arccos(h), \arctan\left[\frac{k-\rho h}{h\sqrt{1-\rho^2}}\right]\right\}, \theta_\ell(h)\right].$$

Here the arccosine takes values in $[0,\pi/2]$ and the arctangent in $(-\pi/2,\pi/2)$. Notice that if

$$x(h,k,\rho) \equiv \frac{k-\rho h}{\sqrt{(1-h^2)(1-\rho^2)}} < -1,$$

then

$$H_\gamma(h,k,\rho) = 0.$$

For the remainder of this calculation, we assume $x(h,k,\rho) \ge -1$.

The joint density of R and Θ may be seen to be $f(r,\theta) = \gamma r(1-r^2)^{\gamma-1}/\pi$ over $r^2 \le 1$ and $-\pi < \theta \le \pi$. Thus,

$$\begin{aligned} H_\gamma(h,k,\rho) &= \frac{\gamma}{\pi}\int_{\theta_\ell(h)}^{\theta_u(h,k,\rho)}\int_{h\sec\theta}^{1} r(1-r^2)^{\gamma-1}\,dr\,d\theta \\ &= \frac{1}{2\pi}\int_{\theta_\ell(h)}^{\theta_u(h,k,\rho)} (1-h^2\sec^2\theta)^\gamma\,d\theta. \end{aligned}$$

It can be shown, by writing $(1-h^2\sec^2\theta)^\gamma = (1-h^2\sec^2\theta)^{\gamma-1}(1-h^2\sec^2\theta)$ and using the identity $\sec^2\theta = 1+\tan^2\theta$ and the fact $d\tan\theta/d\theta = \sec^2\theta$, that $H_\gamma(h,k,\rho)$ satisfies the recurrence formula

$$\begin{aligned} H_\gamma(h,k,\rho) = H_{\gamma-1}(h,k,\rho) &- \frac{h(1-h^2)^{\gamma-1/2}\Gamma(\gamma)}{4\sqrt{\pi}\Gamma(\gamma+1/2)} \\ &\times\left\{1+\operatorname{sgn}(k-\rho h) I_{z(h,k,\rho)}\left(\frac{1}{2},\gamma\right)\right\}, \quad (2) \end{aligned}$$

where $z(h,k,\rho) = \min\{1, (k-\rho h)^2/\{(1-h^2)(1-\rho^2)\}\} = \min\{1, x^2(h,k,\rho)\}$ and $I_y(p,q)$ is the beta(p,q) distribution function given by

$$I_y(p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\int_0^y t^{p-1}(1-t)^{q-1}\,dt. \quad (3)$$

The recurrence formula (2) leads immediately to a recursion formula for $G(h,k;\rho,\gamma)$. When $\gamma = (n-3)/2$ and n is odd, γ is an integer and our recursion becomes

$$\begin{aligned} G(h,k;\rho,\gamma) &= G(h,k;\rho,0) - \sum_{j=1}^{\gamma}\{H_j(h,k,\rho) - H_{j-1}(h,k,\rho) \\ &\qquad + H_j(k,h,\rho) - H_{j-1}(k,h,\rho)\}, \end{aligned}$$

where, letting $\gamma \to 0$ in the definition of H_γ, we find

$$G(h,k;\rho,0) = 1 - \frac{\theta_u(h,k,\rho) - \theta_\ell(h) + \theta_u(k,h,\rho) - \theta_\ell(k)}{2\pi}.$$

For n even, we get

$$\begin{aligned} G(h,k;\rho,\gamma) &= G(h,k;\rho,1/2) \\ &\quad - \sum_{j=1}^{\gamma-1/2}\{H_{j+1/2}(h,k,\rho) - H_{j-1/2}(h,k,\rho) \\ &\qquad + H_{j+1/2}(k,h,\rho) - H_{j-1/2}(k,h,\rho)\}. \end{aligned}$$

The recursion starts with

$$G(h,k;\rho,1/2) = 1 - H_{1/2}(h,k,\rho) - H_{1/2}(k,h,\rho).$$

Adopting the shorthand

$$y(h,k,\rho) = \frac{z(h,k,\rho)(1+h^2) - h^2}{2\sqrt{z(h,k,\rho)(1-z(h,k,\rho))}h},$$

we find by direct integration that

$$H_{1/2}(h,k,\rho) = \begin{cases} \dfrac{1-h}{2}, & x(h,k,\rho) > 1 \\ \dfrac{3(1-h)}{8} + \dfrac{\arctan\{y(h,k,\rho)\} - h\arcsin(2z(h,k,\rho)-1)}{4\pi}, & 0 \le x(h,k,\rho) \le 1 \\ \dfrac{1-h}{8} - \dfrac{\arctan\{y(h,k,\rho)\} - h\arcsin(2z(h,k,\rho)-1)}{4\pi}, & -1 < x(h,k,\rho) \le 0 \\ 0, & x(h,k,\rho) \le -1. \end{cases}$$

For $x(h,k,\rho) = 0$, this reduces to $(1-h)/4$.

Formulas given in Dunnett and Sobel (1954) can be used to evaluate the incomplete beta functions:

$$\begin{aligned} I_x\left(\frac{1}{2}, j+\frac{1}{2}\right) &= \frac{2}{\pi}\arctan\sqrt{\frac{x}{1-x}} + \frac{2}{\pi}\sqrt{x(1-x)}\sum_{i=0}^{j-1}\frac{4^i(i!)^2}{(2i+1)!}(1-x)^i \end{aligned}$$

and

$$I_x\left(\frac{1}{2}, j\right) = \sqrt{x}\sum_{i=0}^{j-1}\frac{(2i)!}{4^i(i!)^2}(1-x)^i.$$

Note that some care is needed to avoid numerical difficulties near $h=0$ or $k=0$. As $h \to 0$ with $k > 0$, we find $\theta(h,k) \to \pi/2$ and $\theta(k,h) \to \min\{\arccos(k), \arctan(-\rho/\sqrt{1-\rho^2})\}$. Values of $x(h,k,\rho)$ and $z(h,k,\rho)$ behave well for small h or k, but when $h \to 0$ with $k > 0$ the function $y(h,k,\rho)$ converges to $+\infty$. When h is very close to 0, these limits should be used in calculations.

4. MONTE CARLO STUDY

In this section we describe the results of a Monte Carlo study that justifies the use of asymptotic critical points in obtaining a p value as described in the previous section. Moreover, we offer some comparisons of the suggested tests (based on the exact probability integral transforms) with those based on normal probability integral transforms that are valid in the case of a single population. The results presented in this section are all based on 10,000 simulations.

Table 1 presents Monte Carlo critical points together with corresponding asymptotic critical points for various values of k, the number of levels of the factor, and m, the number of observations at each level of the factor, where we assume $n_1 = \cdots = n_k = m$.

The results presented in Table 1 clearly show that the Monte Carlo critical points are well approximated by the asymptotic critical points even in samples of size as small as 5. Even with three replicates, the asymptotic critical points appear to provide reasonably good approximations.

For the case of fitting a mean to a sample from a single population, one can expect the residuals to be approximately normally distributed provided the random errors are normally distributed. Thus, for this case, we compared the performance of the tests based on the true distribution (labeled G) with those based on normal probability integral transforms (labeled Φ). Table 2 presents the powers of the two tests using Monte Carlo critical points for a variety of alternatives.

The results presented in Table 2 clearly show that both kinds of tests perform fairly well for skewed alternatives. For symmetric alternatives, the tests based on the true distribution appear to work well for heavy-tailed alternatives, whereas the tests based on the normal distribution function appear to be more sensitive to light-tailed alternatives.

We also examined the powers of the proposed tests obtained by pooling several populations with small numbers of replicates for each. For this case, the tests based on the normal probability

Table 2. Power Comparisons: Single Population

			Power of the test			
			Using Φ		Using G	
Distribution	n	k	W^2	A^2	W^2	A^2
χ^2_1	10	1	.6544	.6878	.6519	.6680
	20	1	.9503	.9681	.9496	.9658
Exponential	10	1	.3729	.4013	.3736	.3980
	20	1	.7206	.7775	.7180	.7790
Lognormal	10	1	.5371	.5632	.5382	.5641
	20	1	.8776	.9041	.8763	.9055
Uniform	20	1	.1358	.1683	.1263	.1392
Cauchy	10	1	.6048	.6042	.6145	.6259
	20	1	.8852	.8864	.8876	.8959
Laplace	10	1	.1500	.1520	.1563	.1739
	20	1	.2603	.2732	.2682	.2952
Beta(2, 2)	20	1	.0521	.0563	.0488	.0441
t_2	10	1	.2980	.3772	.2983	.3252
	20	1	.5099	.5305	.5165	.5532
t_3	10	1	.1762	.1835	.1813	.2071
	20	1	.3016	.3275	.3072	.3496
Gamma(1, 2)	10	1	.1988	.2150	.1996	.2223
	20	1	.4107	.4636	.4087	.4685

NOTE: Powers are based on 10,000 iid samples of size n from the distributions listed. All tests are at level $\alpha = .05$. For the Uniform and Beta(2, 2) distributions and $n = 10$, the powers are negligibly different from α.

Table 1. Monte Carlo Critical Points for Finite n With Exact Asymptotic Points

		Percentage points for											
		W^2						A^2					
		Upper tail probability (percent)											
m	k	15	10	5	2.5	1.0	.5	15	10	5	2.5	1.0	.5
3	10	.093	.114	.151	.185	.232	.274	.741	.880	1.142	1.424	1.780	2.031
	20	.094	.115	.151	.191	.240	.273	.747	.882	1.151	1.425	1.810	2.075
	∞	.095	.116	.154	.194	.248	.290	.745	.894	1.161	1.442	1.825	2.122
4	10	.085	.101	.132	.159	.196	.234	.648	.769	.995	1.216	1.544	1.783
	20	.086	.102	.130	.158	.197	.229	.652	.765	.991	1.208	1.491	1.761
	∞	.085	.101	.129	.157	.197	.228	.648	.763	.970	1.188	1.485	1.715
5	10	.085	.100	.127	.152	.190	.219	.620	.725	.915	1.125	1.434	1.632
	20	.087	.100	.127	.156	.185	.215	.622	.726	.911	1.125	1.370	1.594
	∞	.085	.099	.123	.148	.182	.209	.614	.712	.886	1.066	1.314	1.505
7	10	.085	.099	.124	.148	.176	.196	.584	.668	.822	.975	1.166	1.306
	20	.087	.101	.124	.145	.180	.199	.587	.674	.826	.968	1.194	1.386
	∞	.087	.100	.123	.146	.177	.201	.587	.671	.818	.968	1.172	1.329
10	1	.087	.099	.120	.140	.173	.190	.569	.651	.778	.920	1.129	1.270
	5	.088	.101	.125	.149	.183	.207	.582	.658	.782	.939	1.139	1.291
	10	.088	.101	.126	.151	.181	.201	.573	.655	.797	.949	1.122	1.280
	20	.088	.100	.123	.144	.173	.194	.568	.646	.783	.910	1.100	1.238
	∞	.088	.101	.124	.146	.177	.201	.575	.653	.787	.923	1.106	1.247
20	1	.090	.103	.125	.145	.176	.203	.569	.643	.763	.887	1.073	1.172
	5	.092	.105	.129	.150	.178	.201	.578	.647	.780	.902	1.041	1.194
	10	.090	.102	.125	.145	.176	.204	.564	.636	.753	.870	1.048	1.207
	∞	.089	.102	.125	.148	.178	.201	.566	.639	.765	.892	1.061	1.191
30	1	.090	.102	.125	.149	.180	.204	.563	.631	.758	.898	1.074	1.200
	5	.090	.104	.128	.152	.184	.203	.564	.646	.769	.904	1.082	1.200
	10	.090	.103	.126	.148	.175	.198	.566	.639	.760	.879	1.042	1.172
	∞	.090	.103	.125	.148	.178	.201	.564	.636	.760	.885	1.051	1.179
∞		.091	.105	.127	.150	.181	.204	.556	.627	.749	.872	1.035	1.159

NOTE: There are m observations at each of k levels of some factor.

Table 3. Power Comparisons: Multiple Populations

		Power of the test					
		$k = 10$		$k = 20$		$k = 30$	
Distribution	*n*	W^2	A^2	W^2	A^2	W^2	A^2
χ^2_1	3	.5000	.5481	.7834	.8158	.9222	.9402
	5	.9757	.9793	.9998	1.0000	1.0000	1.0000
Exponential	3	.2590	.2737	.4603	.4718	.6185	.6325
	5	.7785	.7907	.9713	.9763	.9983	.9987
Lognormal	3	.3545	.3780	.5965	.6198	.7711	.7860
	5	.9030	.9136	.9962	.9964	1.0000	1.0000
Uniform	3	.0590	.0632	.0622	.0670	.0659	.0700
	5	.0713	.0573	.1085	.0846	.1645	.1273
	10	.3839	.3988	.7444	.7762	.9173	.9315
Cauchy	3	.0889	.1236	.0975	.1406	.1167	.1733
	5	.3792	.6211	.6079	.8545	.8000	.9512
Laplace	3	.0534	.0558	.0538	.0554	.0592	.0599
	5	.0974	.1136	.1092	.1336	.1384	.1746
Beta(2, 2)	3	.0512	.0538	.0519	.0515	.0556	.0573
	5	.0425	.0347	.0526	.0435	.0643	.0562
	10	.1034	.1014	.2106	.2239	.3390	.3560
Gamma(2)	3	.1384	.1422	.2397	.2400	.3368	.3369
	5	.4623	.4775	.7605	.7803	.9193	.9315
	10	.8889	.8689	.9964	.9967	1.0000	1.0000
t_2	3	.0607	.0682	.0649	.0700	.0632	.0680
	5	.1473	.2027	.1854	.2778	.2487	.3871
	10	.7420	.8012	.9496	.9676	.9916	.9953

NOTE: Powers are based on 10,000 replications of selecting k samples of size n from the distributions listed. All tests are at level $\alpha = .05$.

transforms do not have the correct size so we only report the results obtained using the true distribution G. Table 3 presents the results where we have let k denote the number of populations and n denote the number of observations for each population.

The results presented in Table 3 indicate that by pooling data belonging to several populations the powers of the tests can be improved substantially. With skewed alternatives, even with three replicates, one can expect satisfactory power provided the number of populations exceeds 10. In many applications, the number of replicates does not exceed 5. Thus, we expect that the tests proposed in this article to be quite useful in areas where data are available from a number of similar experiments as is the case with thermoluminescence sedimentary dating. The results also confirm the usual comparisons between W^2 and A^2, namely, that A^2 tends to have better power for long-tailed symmetric alternatives and slightly better power for skewed alternatives so that overall A^2 would be the recommended test.

5. EXAMPLE

Berger and Huntley (1989) presented datasets from experiments to date sediments using thermoluminescence; the datasets are reproduced in Tables 4 and 5. In Table 4 we give

Table 4. Photon Counts per Degree Celsius Temperature Increase in a Thermoluminescence Dating Experiment for a Glaciolacustrine Silt

Unbleached samples					*Bleached samples*				
Level	*Dose*	*Photon count*	$\hat{\epsilon}_{ij}$	*Exact PIT*	*Level*	*Dose*	*Photon count*	$\hat{\epsilon}_{ij}$	*Exact PIT*
1	0	38,671	.175	.558	5	0	20,766	−.898	.217
1	0	40,646	1.173	.891	5	0	21,393	−.180	.450
1	0	38,149	−.089	.470	5	0	22,493	1.078	.883
1	0	35,836	−1.259	.080	NA	120	31,290	—	—
2	120	65,931	−.669	.303	NA	120	33,779	—	—
2	120	67,887	1.150	.970	6	240	43,221	.471	.634
2	120	66,133	−.481	.363	6	240	43,450	.678	.700
3	240	82,496	−1.154	.009	6	240	41,427	−1.149	.033
3	240	86,708	.604	.675	7	480	51,804	−.831	.244
3	240	86,580	.550	.658	7	480	59,555	1.110	.911
4	480	110,978	−.536	.321	7	480	54,013	−.278	.423
4	480	113,807	.750	.750	NA	960	75,748	—	—
4	480	114,192	.925	.808	NA	960	76,613	—	—
4	480	109,652	−1.138	.121					
NA	960	130,373	—	—					
NA	960	137,789	—	—					

NOTE: Samples were irradiated with doses of gamma radiation from a ^{60}Co source. See the text for details of the units for dose. The first column, Level, is the index i running from 1 to $k = 17$ labeling the sets of replicates for our tests. We also report standardized residuals from $\hat{\epsilon}_{ij}$ and the values of the exact probability integral transforms $u_{ij} = G_{n_i}(\hat{\epsilon}_{ij})$ for each data point.

Table 5. Photon Counts per Degree Celsius Temperature Increase in a Thermoluminescence Dating Experiment for a Lake Silt

Unbleached samples					*Bleached samples*				
Level	*Dose*	*Photon count*	$\hat{\epsilon}_{ij}$	*Exact PIT*	*Level*	*Dose*	*Photon count*	$\hat{\epsilon}_{ij}$	*Exact PIT*
8	0	20,522.2	1.075	.858	14	0	11,814.6	.978	.821
8	0	19,373.6	−.491	.336	14	0	11,587.8	−1.021	.155
8	0	20,14.6	.555	.685	14	0	11,708.6	.043	.512
8	0	18,899.1	−1.138	.121	15	1	26,645.2	1.112	.914
9	1	50,382.5	.980	.823	15	1	26,445.2	−.288	.420
9	1	48,57.2	−1.019	.156	15	1	26,368.6	−.824	.247
9	1	49,529.5	.039	.511	16	2	41,487.1	.914	.791
10	2	77,706.6	1.126	.929	16	2	39,125.1	−1.068	.124
10	2	75,291.3	−.342	.404	16	2	40,582.5	.155	.543
10	2	74,563.8	−.784	.262	NA	4	61,532.1	—	—
11	4	111,547.5	−.040	.489	NA	4	57,023.6	—	—
11	4	113,899.1	1.019	.844	17	8	93,015.8	1.154	.987
11	4	109,461.1	−.979	.178	17	8	87,907.7	−.535	.347
12	8	164,564.9	.366	.603	17	8	87,655.2	−.619	.320
12	8	151,504.2	−1.132	.064	NA	16	107,618.3	—	—
12	8	168,042.1	.765	.731	NA	16	110,394.2	—	—
13	16	204,726.5	.796	.742					
13	16	201,964.3	.326	.591					
13	16	193,457.6	−1.122	.076					

NOTE: Samples were irradiated with doses of beta radiation from a ^{90}Sr source. See the text for details of the units for dose. The first column, Level, is the index i running from 1 to $k = 17$ labeling the sets of replicates for our tests. We also report standardized residuals from $\hat{\epsilon}_{ij}$ and the values of the exact probability integral transforms $u_{ij} = G_{n_i}(\hat{\epsilon}_{ij})$ for each data point.

dataset 1, labeled QNL84-2 by Berger and Huntley; the sediment is glaciolacustrine silt. A total of 29 samples were prepared. Of these, 13 were pretreated by optical bleaching. The samples were exposed to gamma radiation at doses (in minutes of ^{60}Co gamma radiation at 1.6 Gy/min) listed in the table. The samples were then heated and the thermoluminescence measured as recorded in the table. (Units are photon counts per degree Celsius as the temperature is raised smoothly.) Dataset 2, shown in Table 5, is for a lake sediment. For this dataset, there were 35 data points, 16 of which corresponded to pretreatment by optical bleaching. The gamma radiation is measured in kiloseconds of ^{90}Sr beta radiation at 90 Gy/ks for this dataset.

For these datasets, interest centers on fitting heteroscedastic, nonlinear models relating photon count to dose; see Berger, Lockhart, and Kuo (1987) for details of these model fits. Generally, the models have the form

$$y_{ij} = g(d_{ij}, \theta)(1 + \sigma\epsilon_{ij}),$$

where it is assumed that the ϵ_{ij} are independent standard normal variates. A common example is

$$g(d, \theta) = \alpha\{1 - \exp(-\beta(d + \gamma))\},$$

where the values of α and β, at least, depend on whether the observation is bleached or not. For each of our two datasets, there are then six parameters (or seven if, as is fairly commonly done, we allow σ to be different for bleached and unbleached data).

It is useful to test the assumption of normal errors because the behavior of some of the fitting methods used depends on the quality of this normal approximation and because diagnostic statistics have behaviors that depend on the assumption of normality. This assumption could be tested by fitting the nonlinear regression model in question, extracting standardized residuals, and hoping these would be approximately normal. No exact distribution theory for the residuals is available; it would be necessary to assess the extent to which they were approximately normal. Moreover, the hypothesis of normal errors might be rejected even if it were correct because the model itself might be wrong. Consequently, we will apply the tests proposed here fitting separate means and variances at each level of the covariates.

Altogether, there are 22 different combinations of covariate levels (two treatments times five dose levels for dataset 1 and two treatments times six dose levels for dataset 2). It will be seen in Tables 4 and 5 that for 5 of these 22 combinations there were only 2 replicates; these 10 observations do not provide information for testing fit to the normal distribution. Eliminating them leaves $k = 17$ combinations. Of these, 14 have $n_i = 3$ and 3 have $n_i = 4$. We use these 54 data points to assess the normality of the residuals using our proposed tests.

The values of $\hat{\epsilon}_{ij}$ and $u_{ij} = G_{n_i}(\hat{\epsilon}_{ij})$ from steps 1 to 3 of Section 2.1 are recorded in the tables. We sort the probability integral transforms and then compute the statistics to get $W^2_{\text{obs}} = .008875$ and $A^2_{\text{obs}} = .07979$. Finally, we compute asymptotic P values corresponding to these values using the method of Section 3, taking $m = 100$ to discretize the integral equation.

We find that the P value corresponding to the statistic W^2 is .998, whereas that corresponding to A^2 is .992. Both of these P values indicate that the data are, if anything, surprisingly too normal looking. A small Monte Carlo study confirmed that the asymptotic P values are quite accurate in this case. Thus, these large P values are not easily explained by the quality of the asymptotic approximation.

The datasets used here were published for use in testing software; we speculate that the process of selecting suitable data for such a purpose might have tended to eliminate less normal looking datasets though we do not actually think the normality of the data was directly assessed in selecting the sets.

6. DISCUSSION

In testing the normality of random errors in regression models, the usual practice is to compute the probability integral transforms, u_i, of the fitted residuals using the normal distribution and to test whether the resulting u_i's follow a uniform distribution. When the fitted models are nonlinear with many parameters, there is no guarantee that the fitted residuals are normally distributed even if the random errors are. Thus, the actual sizes of the tests that use normal probability integral transforms can deviate substantially from the nominal levels. We propose basing the tests on the empirical distribution of the exact probability integral transforms of residuals.

The performance of the proposed tests in finite samples is examined by Monte Carlo simulations. The study shows that Monte Carlo critical points are well approximated by the asymptotic critical points when the number of replicates at each level is as small as 5, thus justifying their use in realistic samples. The approximation is not so poor, even with three replicates at each level. When the number of replicates exceeds 10, the critical points of the proposed tests approach the asymptotic points of the tests based on normal probability integral transforms.

The powers of the proposed tests were compared with those based on the normal probability integral transforms for the case of fitting a single response mean. The proposed tests are found to be more sensitive to alternatives with heavy tails, whereas the latter was found to be more sensitive to alternatives with light tails. Both tests show considerable power against skewed alternatives. The proposed tests have the advantage that the data collected from several similar experiments can be used to improve the power of the test. Furthermore, they are robust to a possible misspecification of the model.

APPENDIX: LARGE–SAMPLE DISTRIBUTION THEORY

We begin by showing that the process W_n is approximately Gaussian with mean 0 and covariance function as given in Section 3. To be precise, we have the following theorem.

Theorem 1. Let $n = \sum_{i=1}^{k} n_i$ be the total number of observations and let $z_1, \ldots, z_n$ be the ordered probability integral transforms. Assume there is an integer N with $n_i \le N \ \forall\, i,\ n$. Let $\zeta_{r,n} = \#\{i : n_i = r\}/k$. Furthermore, assume that $\zeta_{r,n} \to \zeta_r$. Then the process $W_n(t)$ converges weakly in $D[0,1]$ to a Gaussian process W with mean 0 and covariance function $\alpha(s,t)$ given by

$$\alpha(s,t) = \min(s,t) - st + \sum_{r=3}^{N} \zeta_r r(r-1)\big[G_{2,r}\{G_r^{-1}(s), G_r^{-1}(t)\} - st\big].$$

Let k denote the number of distinct levels of the covariate and let n_i denote the number of replicates at each level. Let $n = \sum_{i=1}^{k} n_i$ be the total number of observations. We prove the weak convergence of the process

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \big[I\big\{G_{n_i}(\hat{\epsilon}_{ij}) \le t\big\} - t\big]$$

for the case of an equal number of replicates at each level of the covariate (i.e., $n_i = m$ for all i). We fix m and let $k \to \infty$. In this case, the process W_n can be rewritten as

$$W_n(t) = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} W_{nj}(t),$$

where each $W_{nj}(t) = k^{-1/2} \sum_{i=1}^{k} [I(u_{ij} \le t) - t]$ and $u_{ij} = G_m(\hat{\epsilon}_{ij})$. For each fixed j, the variables $u_{1j}, \ldots, u_{kj}$ are, under the null hypothesis, iid Uniform$[0,1]$ variables and so each W_{nj} converges weakly in $D[0,1]$ to a Brownian bridge, that is, a Gaussian process W_j with mean 0 and covariance $\min(s,t) - st$. Therefore, W_{nj} is tight in $D[0,1]$. This, in turn, implies that, for each j, there is a compact $K_j \subset D[0,1]$ such that $P(W_{nj} \in K_j) \ge 1 - \epsilon/m$ for any $\epsilon > 0$. Because K_j is compact in $D[0,1]$, $K = K_1 \times \cdots \times K_m$ is compact in $(D[0,1])^m$ and $P((W_{n1}, \ldots, W_{nm}) \in K) \ge 1 - \epsilon$. Because ϵ is arbitrary, it follows that the process $(W_{n1}, \ldots, W_{nm})$ is tight in $(D[0,1])^m$.

Now consider $0 \le t_1 < \cdots < t_r \le 1$. The matrix M_k whose ljth entry is $W_{nj}(t_l)$ can be written as $\sum_{i=1}^{k} Q_i/\sqrt{k}$, where the matrices Q_i's are iid and Q_i has ljth entry $I(u_{ij} \le t_l) - t_l$. Each Q_i has mean 0 and so M_k converges in distribution by the usual central limit theorem to a Gaussian matrix $\mathbf{M}$ with $E(\mathbf{M}) = \mathbf{0}$ and

$$\begin{aligned}\operatorname{cov}(M_{lj}, M_{l'j'}) &= \operatorname{cov}\{I(u_{ij} \le t_l), I(u_{ij'} \le t_{l'})\} \\ &= G\big\{G^{-1}(t_l), G^{-1}(t_{l'}), \rho_{jj'}, (m-3)/2\big\} - t_l t_{l'},\end{aligned}$$

where $\rho_{jj'} = 1$ if $j = j'$ and $-1/(m-1)$ if $j \ne j'$.

Thus, $(W_{n1}, \ldots, W_{nm})$ converges weakly in $(D[0,1])^m$ to a Gaussian process $(W_1, \ldots, W_m)$ with mean 0 and $\operatorname{cov}\{W_j(t_l), W_{j'}(t_{l'})\} = G\{G^{-1}(t_l), G^{-1}(t_{l'}), \rho_{jj'}, (m-3)/2\} - t_l t_{l'}$.

Because each W_j is in $C[0,1]$ (each is a Brownian bridge), it follows that $W_n = m^{-1/2} \sum_{j=1}^{m} W_{nj}$ converges weakly in $D[0,1]$ to $W = m^{-1/2} \sum_{j=1}^{m} W_j$, which is a mean-zero Gaussian process with covariance

$$\begin{aligned}\alpha(s,t) &= \operatorname{cov}(W(s), W(t)) \\ &= \frac{1}{m} \sum_{j=1}^{m} \sum_{j'=1}^{m} \operatorname{cov}\{W_j(s), W_{j'}(t)\} \\ &= \min(s,t) - st + \frac{1}{m} \sum_{j \ne j'} \big[G_{2,m}\{G^{-1}(s), G^{-1}(t)\} - st\big].\end{aligned}$$

This proves the theorem.

It is well known (see, for instance, Stephens 1986) that if W is a mean-zero Gaussian process with covariance α_n, then $\int_0^1 W^2(t)\,dt$ has the same law as

$$\sum_{i=1}^{\infty} \lambda_i \chi_i^2,$$

where the χ_i^2's denote a set of independent chi-squared random variables each on 1 degree of freedom and the λ_i's are eigenvalues of the integral equation $\int_0^1 \alpha_n(s,t) f(t)\,dt = \lambda f(s)$. With the theorem, this shows that under the null hypothesis the p values computed in Section 3 are asymptotically uniformly distributed under the null hypothesis.

[Received August 2002. Revised September 2004.]

REFERENCES

Beckman, R. J., and Trussell, H. J. (1974), "The Distribution of an Arbitrary Studentized Residual and the Effects of Updating in Multiple Regression," *Journal of the American Statistical Association*, 69, 199–201.

Berger, G. W., and Huntley, D. J. (1989), "Test Data for Exponential Fits," *Ancient TL*, 7, 43–46.

Berger, G. W., Lockhart, R. A., and Kuo, J. (1987), "Regression and Error Analysis Applied to the Dose Response Curves in Thermoluminescence Dating," *Nuclear Tracks and Radiation Measurements*, 13, 177–184.

Chen, G., and Lockhart, R. A. (2001), "Weak Convergence of the Empirical Process of Residuals in Linear Models With Many Parameters," *The Annals of Statistics*, 29, 748–762.

Dunnett, C. W., and Sobel, M. (1954), "A Bivariate Generalization of Student's *t*-Distribution, With Tables for Certain Special Cases," *Biometrika*, 41, 153–169.

Ellenberg, J. H. (1973), "The Joint Distribution of the Standardized Least Squares Residuals From a General Linear Regression," *Journal of the American Statistical Association*, 68, 941–943.

Imhof, J. P. (1961), "Computing the Distribution of the Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.

Lockhart, R. A., O'Reilly, F. J., and Stephens, M. A. (1986), "Tests of Fit Based on Normalised Spacings," *Journal of the Royal Statistical Society*, Ser. B, 48, 344–352.

Mammen, E. (1996), "Empirical Process of Residuals for High Dimension Linear Models," *The Annals of Statistics*, 24, 307–335.

Pierce, D. A., and Kopecky, K. J. (1979), "Testing Goodness of Fit for the Distribution of Errors in Regression Models," *Biometrika*, 66, 1–5.

Stephens, M. A. (1976), "Asymptotic Results for Goodness of Fit Statistics With Unknown Parameters," *The Annals of Statistics*, 4, 357–369.

——— (1986), "Tests Base on EDF Statistics," in *Goodness of Fit Techniques*, eds. R. B. D'Agostino and M. A. Stephens, New York: Marcel Dekker, Chap. 4.