Journal of the Korean Statistical Society (2001), **30**: *4, pp 645–660*

Testing the Goodness of Fit of a Parametric Model via Smoothing Parameter Estimate

Choongrak Kim, Changkon Hong, and Meeseon Jeong[1]

ABSTRACT

In this paper we propose a goodness-of-fit test statistic for testing the (null) parametric model versus the (alternative) nonparametric model. Most of existing nonparametric test statistics are based on the residuals which are obtained by regressing the data to a parametric model. Our test is based on the bootstrap estimator of the probability that the smoothing parameter estimator is infinite when fitting residuals to cubic smoothing spline. Power performance of this test is investigated and is compared with many other tests. Illustrative examples based on real data sets are given.

Keywords: Generalized cross-validation; Residuals; Smoothing spline; Wild bootstrap.

1. INTRODUCTION

It is quite often to test the goodness-of-fit of the postulated model when one fits a parametric model to data. For a long time parametric goodness-of-fit test has been used for this purpose, which is an F-test using the general linear test approach. As argued by Eubank and Spiegelman (1990), parametric tests are against another postulated (alternative) model and are inconsistent against many other alternatives, especially against those which are orthogonal to the postulated alternative. This drawback of parametric approach demands a different approach; nonparametric approach. For the last decade many nonparametric tests have been suggested by Cox, et. al. (1988), Munson and Jernigan (1989), Eubank and Spiegelman (1990), Buckley (1991), Eubank and Hart (1992, 1993), and Härdle and Mammen (1993). In this paper, we introduce a new nonparametric goodness-of-fit test which has good power and ready-to-use critical value.

Consider the regression model

$$y_i = f(x_i) + g(x_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

[1]Department of Statistics, Pusan National University, Pusan, 609-735, KOREA

where $0 \le x_1 < \cdots < x_n \le 1$ are fixed design points, f has a known parametric (linear or nonlinear) form, g is unknown smooth function, and the ε_i's are independent and identically distributed random variables with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2 < \infty$. We wish to test

$$\mathrm{H}_0 : g = 0$$

Most of the existing nonparametric tests assume f is linear. But our test allows f to be linear or nonlinear. Our test has the same spirit as many other tests in the sense of using the residuals from parametric regressing y on x, and regressing the residuals nonparametrically.

In Section 2 most of the nonparametric tests and their critical values are introduced. In Section 3 we discuss the relationship between the hat matrix and the smoothing parameter, and the rationale on our test statistic. In Section 4 we study the pattern of generalized cross-validation (GCV), critical values for our test statistic, and compare power with other tests under alternatives. Illustrative examples based on real data sets are given in Section 5, and concluding remarks are given in Section 6.

2. NONPARAMETRIC TESTS : REVIEW

2.1. Test Statistics

Let $\mathbf{r} = (r_1, \ldots, r_n)'$ be residual vector from fitting the data to a parametric model specified under null hypothesis. Most test statistics are derived by nonparametric regression of $\mathbf{r}$. Nonparametric regression methods used are smoothing spline, series estimator and kernel regression. Before introducing nonparametric test statistics suggested so far, it is helpful to introduce unified notations. Let $\mathbf{Q}$ be $n \times n$ matrix, $\mathbf{M}(\lambda) = \mathbf{Q} + n\lambda\mathbf{I}$, and $\mathbf{X} = \{x_i^{j-1}\}_{i=1,\ldots,n;j=1,\ldots,m}$ be $n \times m$ matrix. If a cubic smoothing spline is used, $m = 2$. Also, define an $n \times (n-m)$ matrix $\mathbf{U}$ such that $\mathbf{U}'\mathbf{X} = \mathbf{0}$ and $\mathbf{U}'\mathbf{U} = \mathbf{I}$. Define sample Fourier coefficients by $a_j = \sqrt{2}\sum_{i=1}^{n} y_i \cos(j\pi x_i)$, $j = 1, \ldots, n-1$, and let $\hat\sigma^2$ be a $\sqrt{n}$–consistent estimator of σ^2. (For example see Rice (1984)).

Earlier, von Neumann (1941) used

$$\mathrm{VN} = \mathbf{r}'\mathbf{r}/\hat\sigma^2$$

as a test statistic.

Cox, et. al. (1988) suggested the locally most powerful (LMP) test when the null hypothesis is f is a polynomial of degree $m-1$ or less, and the alternative is

$f+g$ is "smooth" based on a Bayesian model. Let $Q(s,t)$ be the covariance kernel for $(m-1)$-fold integrated Wiener process, that is, $Q(s,t) = \{(m-1)!\}^{-2}\int_0^1 (s-u)_+^{m-1}(t-u)_+^{m-1}du$, where $t_+ = max\{0,t\}$. It rejects H_0 if

$$\mathrm{CK} = \mathbf{r}'\mathbf{Q}\mathbf{r}$$

is too large, where $\mathbf{Q} = (Q(x_i, x_j))$.

Under the same perspective as Cox, et. al. (1988), Buckley (1991) derived an LMP test

$$\mathrm{BU} = \sum_{i=1}^{n}\{\sum_{j=1}^{i} r_j\}^2/n^2\hat{\sigma}^2.$$

Eubank and Hart (1993) note that $\hat{\sigma}^2\mathrm{BU}/\sigma^2 = \mathrm{CK}$.

Similar test to CK is considered by Munson and Jernigan (1989). Let $\tilde{r}$ be the natural cubic spline interpolant to $\mathbf{r}$. They suggested

$$\mathrm{MJ} = J(\tilde{r})/\mathbf{r}'\mathbf{r},$$

where $J(h) = \int_0^1 h''(t)^2 dt$.

On the other hand, Eubank and Spiegelman (1990) suggested a test based on fitting cubic smoothing splines to $\mathbf{r}$, i.e.,

$$\mathrm{ES} = \{\hat{\mathbf{r}}'\hat{\mathbf{r}} - \hat{\sigma}^2\sum_{j=3}^{n}(1+\lambda\theta_j)^{-2}\}/\hat{\sigma}^2\{2\sum_{j=3}^{n}(1+\lambda\theta_j)^{-4}\}^{1/2},$$

where $\hat{\mathbf{r}}$ is cubic smoothing spline fit to $\mathbf{r}$ and $0 < \theta_3 \le \theta_4 \le \cdots \le \theta_n$, which are eigenvalues defined in Demmler and Reinsch(1975). Here, λ must be preassigned.

When the Fourier series estimator is used, the risk becomes

$$\frac{1}{n}\mathbf{r}'\mathbf{r} - \{\sum_{j=1}^{k} a_j^2 - \frac{2\sigma^2 k}{n}\}.$$

Eubank and Hart (1992) used

$$\mathrm{E1} = \hat{k} = \arg \max_{k\in\{0,1,2,\cdots\}}\{\sum_{j=1}^{k} a_j^2 - \frac{c_\alpha\hat{\sigma}^2 k}{n}\}$$

as a test statistic, and derived c_α so that $P(\hat{k}=0) = 1-\alpha$. Formally, E1 is given by "reject H_0 if $\hat{k} \ge 1$".

Later, instead of $\hat{k}$, Eubank and Hart (1993) proposed

$$\mathrm{E2} = \sum_{j=1}^{k} a_j^2/\hat{\sigma}^2$$

as a test statistic. Here, k can be preassigned or replaced by data-driven estimator.

In the same paper, Eubank and Hart (1993) proposed another test based on the linear smoothing spline fit to $\mathbf{r}$, i.e.,

$$\mathrm{E3} = \sum_{j=1}^{n-1} a_j^2/\hat{\sigma}^2(1+\lambda\gamma_j)^2,$$

where $\gamma_j = \{2n\sin(\frac{j\pi}{2n})\}^2$. Again, λ must be preassigned or be replaced by data-driven estimator.

On the other hand, Härdle and Mammen (1993) proposed a test statistic using the kernel regression. To be more specific, let $\hat{m}_h$ be a kernel estimator with bandwidth h and kernel K. They proposed

$$\mathrm{HM} = nh^{1/2}\int_0^1 (\hat{m}_h(x) - \mathcal{K}_h\hat{f}(x))^2 dx,$$

where $\mathcal{K}_h g(\cdot) = \sum K_h(\cdot - X_i)g(X_i)/\sum K_h(\cdot - X_i)$, $K_h(\cdot) = h^{-1}K(\cdot/h)$ and $\hat{f}(x)$ is a parametric fit under H_0. Similar to k in E2 and λ in E3, h must be preassigned or be replaced by data-driven estimator.

2.2. Critical Values

For a test statistic to be useful in practice, an appropriate critical value must be available for a given level of significance. By the generic property of nonparametric approach, it is impossible to get an exact critical value for a nonparametric test, and therefore, only asymptotic results are available by the asymptotic distribution of test statistics. Asymptotic distribution for VN, BU, E2, E3 are given by Eubank and Hart (1993), ES by Eubank and Spiegelman (1990), E1 by Eubank and Hart (1992), and HM by Härdle and Mammen (1993).

As argued by Härdle and Mammen (1993), however, convergence to the asymptotic distribution is quite slow so that it is more appropriate not to use the asymptotic critical values. In fact, as critical values most authors who suggested test statistics in Section 2.1 used the Monte Carlo approximation or bootstrap in their simulation studies on power of tests and in real data analysis. To derive critical values of the test statistics in Section 2.1, λ in ES, k in E2, λ in E3, and h in HM must be preassigned or be replaced by data-driven estimator. Therefore, critical values based on the Monte Carlo or bootstrap depend on the actual assignment. For example, ES converges to the standard normal under some regularity conditions, however, Eubank and Spiegelman (1990) suggested 2.2 by Monte Carlo study instead of 1.645 as a 95th percentile. But, based on our simulation, it was 1.8 when $\lambda = .0001$ (which was used in their numerical example). Also, Härdle and Mammen (1993) showed that the kernel density via "wild bootstrap" is more appropriate than the asymptotic normal distribution as a distribution for HM. However, the critical value depend heavily on the smoothing parameter h.

3. THE PROPOSED TEST

Let $\mathbf{P}$ be the hat matrix such that $\mathbf{r} = (\mathbf{I} - \mathbf{P})\mathbf{y}$. If H_0 is correct, the pattern of $\mathbf{r}$ is close to white noise since the residuals are evenly distributed around zero. Therefore, the nonparametric regression, for example cubic smoothing spline, of $\mathbf{r}$ will result in the large value of the smoothing parameter λ. Conversely, if H_0 is not correct, the resulting λ will be small. So, we propose a goodness-of-fit test for $H_0 : g = 0$ as the magnitude of $\hat{\lambda}$, estimate of λ. Throughout this paper, we estimate λ by the GCV criterion.

Our idea on the test statistic is closely related with the Theorem 3 in Cox, et. al. (1988). The hat matrix in smoothing spline is

$$\mathbf{H}(\lambda) = \mathbf{I} - n\lambda\mathbf{U}(\mathbf{U}'\mathbf{M}(\lambda)\mathbf{U})^{-1}\mathbf{U}' \tag{3.1}$$

when we fit nonparametrically to $\mathbf{r}$. Therefore, the fitted values are $\hat{\mathbf{r}} = \mathbf{H}(\lambda)\mathbf{r}$. Note that if H_0 is correct, we prefer $\lambda = \infty$ and the corresponding hat matrix $\mathbf{H}$ in (3.1) becomes $\mathbf{P}$. The Theorem 3 in Cox et. al. (1988) is that the GCV criterion $V(\lambda)$ has a (possibly local) minimum at $\lambda = \infty$ whenever

$$\mathbf{y}'(\mathbf{I} - \mathbf{P})\mathbf{Q}(\mathbf{I} - \mathbf{P})\mathbf{y} \le tr(\mathbf{Q})\mathbf{y}'(\mathbf{I} - \mathbf{P})\mathbf{y}/(n - m). \tag{3.2}$$

Through simulation study they showed that under H_0 the probability that the inequality (3.2) holds is approximately 0.671. However (3.2) is not a sufficient condition that GCV has $\lambda = \infty$ as the minimizer, but is a necessary condition. Hopefully, when fitting $\mathbf{r}$ to cubic smoothing spline, the minimizer of GCV occurs at $\lambda = \infty$ under H_0. However, our simulation results show that the minimizer of GCV occurs at $\lambda = \infty$ in about 58900 cases out of 100,000 replications. This property was already recognized by Wahba (1990, p.86; 581 cases out of 1000 replications) through simulation studies. That is, about 41% of the minimizer of GCV is finite even though H_0 is true. We noticed that this probability changes as the sample size n changes. (See Section 4.2 for details) This shows that $\hat{\lambda}$, the GCV estimator of λ, is too sensitive to the randomness of the error terms. The sensitivity of $\hat{\lambda}$ may cause trouble when we use $\hat{\lambda}$ itself as a test statistic. If the level of significance $\alpha < .41$, the critical value based on $\hat{\lambda}$ itself becomes meaningless. However, $\hat{\lambda}$ will be also sensitive to departures from H_0. Therefore, the probability that GCV criterion will choose $\hat{\lambda} = \infty$ as the minimizer will serve as a good criterion. So, if the probability that $\hat{\lambda} = \infty$ could be estimated, it could be a good (in the sense of power) test statistic. That is to say, a good test is not "reject H_0 if $\hat{\lambda}$ is small", but "reject H_0 if the estimated probability of $\hat{\lambda} < \infty$ is large". In order to estimate this probability we use resampling technique, and suggest our test procedure as follows;

1. Obtain B sets of data by resampling techniques such as bootstrap or jackknife. One possible method is;
   i) fit cubic smoothing spline to the original data, and get residuals.
   ii) resample B sets from residuals by bootstrap.
   iii) generate B sets of response vector by adding B sets of resampled residuals from ii) to the fitted values from i).
2. Fit B sets of data generated by step 1 to a parametric model and get residuals.
3. Fit the B sets of residuals obtained in step 2 to cubic smoothing spline and get B $\hat{\lambda}$'s by the GCV criterion.
4. Reject H_0 if LC $\geq c$, where LC(Lambda hat Count) is number of finite $\hat{\lambda}$ out of B $\hat{\lambda}$'s obtained from 3.

The critical value c will be discussed in Section 4. The bootstrap procedure in step 1 is a version of residual bootstrap.

4. SIMULATION RESULTS

In the following Monte Carlo studies, we assume X has uniform design in [0,1], i.e, $x_i = (i-1)/(n-1)$, $i = 1, \ldots, n$, and the ε_i's are *iid* $N(0, \sigma^2)$. Cubic smoothing spline is used and the GCV criterion to estimate λ. We use RKPACK (Gu, 1989) with the grid search method, and 1,000 replications are done.

4.1. GCV Pattern

We present four typical types of GCV pattern when $n = 100$ (see Figure 1). In fact, we compute GCV for $-15 \leq \log_{10} n\lambda \leq 35$ and Figure 1 shows GCV for $-5 \leq \log_{10} n\lambda \leq 10$. Minimum occurs at $\lambda = \infty$ in Figure 1(a) and (c), and at finite λ in Figure 1(b) and (d). To compute $\hat{\lambda}$ via grid search, we have to be very cautious in Figure 1(c) case. Sufficient range of λ is required to get global minimum.

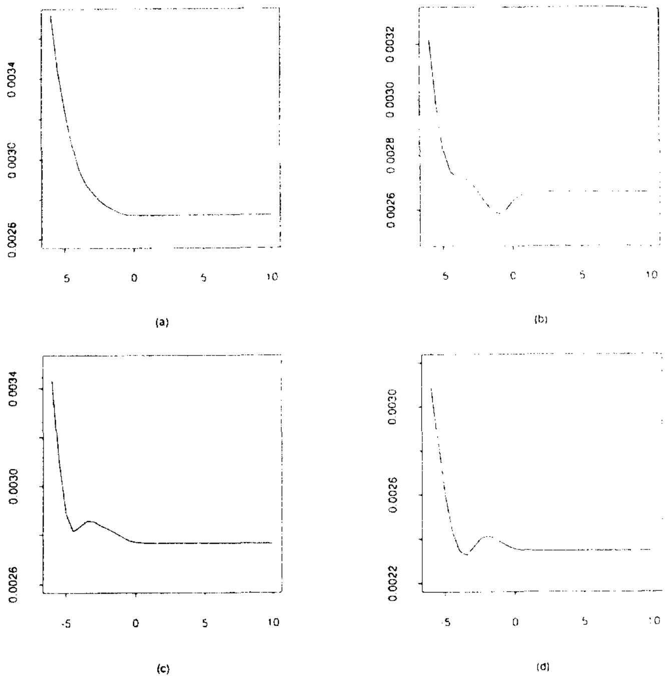

Figure 1: Four Typical Types of GCV Pattern When $n = 100$. Minimum Occurs at $\lambda = \infty$ in (a) and (c), and at Finite λ in (b) and (d).

4.2. Critical Values

To get Monte Carlo approximation to $p_n = P(\hat{\lambda} = \infty | \text{H}_0)$, we generate 10^5 samples of size n under null model. For $\sigma = 0.05, 0.1, 0.5, 1.0$, we obtain almost the same result. Scale invariant property can be easily checked by the definition of GCV. The p_n's for $n = 10$ (10) 100 (100) 500 are approximated. For $n \leq 100$, we obtain a simple formula;

$$p_n = .399 + .0949 \log_{10} n \tag{4.1}$$

But for $n \geq 100$, $p_n \simeq .59$. When B=100, $95th$ percentile of the simulated data could be simplified as

$$c_n \simeq \begin{cases} 88 & , \quad n > 100 \\ 108 - n/5 & , \quad 40 \leq n \leq 100 \\ 100 & , \quad n < 40 \end{cases} \tag{4.2}$$

4.3. Power of the Proposed Test

(i) $f(x) = \beta_0$ (constant); $n = 100$

We choose two types of models in Eubank and Hart (1993), i.e.

$g_1(x) = \beta_1\{e^{4x} - (e^4 - 1)/4\}\{\frac{e^8-1}{8} - (\frac{e^4-1}{4})^2\}^{-1/2}$

$g_2(x) = 2\beta_1\{20(x - \frac{1}{2})^3 - 3(x - \frac{1}{2})\}$

Powers are estimated for VN, BU, ES, E1, E2, E3, HM, and LC with $\beta_1 =$ .25 (.25) 1.00 and standard normal errors, and 1,000 replications are done. Critical values for VN, BU, ES, E2, E3, and HM are evaluated based on 1,000 replications to control the level and "wild bootstrap" is used for HM. Of course 1 for E1 and 88 for LC by (4.2) are used. Results are summarized in Table 1. In g_1, BU, E3, and HM are better than others, but others are not bad at all. In g_2, ES, HM, and LC are much better than others. Therefore, it seems that power performance of BU, VN, E1, E2, and E3 depend heavily on the specification of the alternative model.

(ii) $f(x) = \beta_0 + \beta_1 x$; $n = 100$

From (i), we found that ES, HM, and LC are quite powerful in both g_1 and g_2. Here, we set the null model as linear instead of constant, and choose the same alternative models as in Eubank and Spiegelman (1990).

$g_1(x) = \beta_2 x e^{-2x}$

$g_2(x) = \beta_2 x^2$

We compute powers of ES, E1, HM, and LC. We include E1 because it has ready-to-use critical value. 1,000 replications were done for $\sigma = .05, .10, .20$, and β_0 and β_1 were set as 1. Powers are estimated for $\beta_2 = .00$ (.05) 1.00. Figure 2(a), (b), (c) show power of four tests for $\sigma = .05, .10, .20$, respectively in g_1. As in (i), ES, HM, and LC perform similarly, and better than E1 for all σ. The same phenomenon occurs in g_2, too. (See Figure 3.)

Table 1. Proportion of Rejections in 1,000 Samples of Size 100 with g_1 and g_2

	β_1	1/4	2/4	3/4	4/4
	VN	.120	.613	.984	1.000
	BU	.563	.995	1.000	1.000
	ES	.131	.475	.832	.969
g_1	E1	.376	.906	1.000	1.000
	E2	.328	.940	1.000	1.000
	E3	.420	.979	1.000	1.000
	HM	.470	1.000	1.000	1.000
	LC	.140	.470	.820	.980
	VN	.087	.340	.802	.974
	BU	.045	.134	.450	.849
	ES	.346	.922	1.000	1.000
g_2	E1	.088	.288	.769	.976
	E2	.160	.643	.972	.998
	E3	.169	.701	.984	.999
	HM	.180	.750	.990	1.000
	LC	.340	.820	.990	1.000

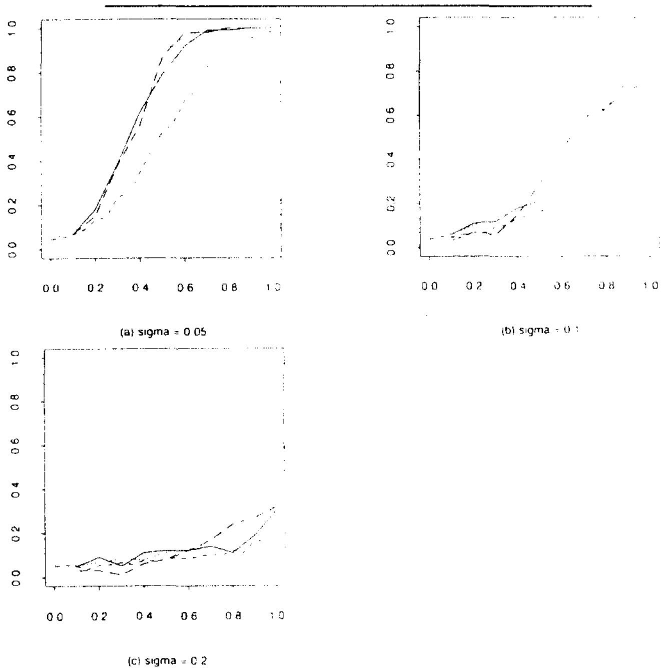

Figure 2: Empirical Powers of ES (······), E1 (- - - -), HM (– – –), and LC (——-) in 1,000 Samples of Size 100 With $\mu_1(x) = \beta_0 + \beta_1 x + \beta_2 x e^{-2x}$; (a), (b), (c) Correspond to $\sigma = .05, .10, .20$, respectively.

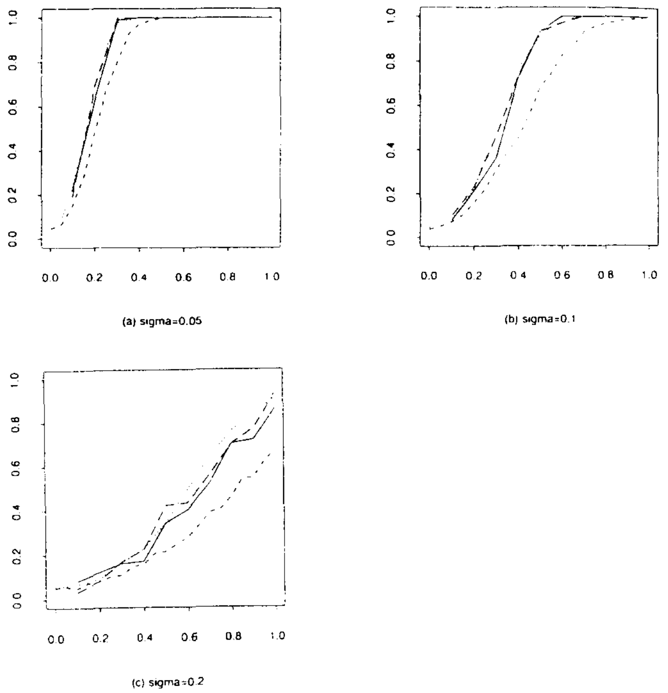

Figure 3: Empirical Powers of ES (······), E1 (- - - -), HM (– – –), and LC (——) in 1,000 Samples of Size 100 With $\mu_2(x) = \beta_0 + \beta_1 x + \beta_2 x^2$; (a), (b), (c) correspond to σ = .05, .10, .20, respectively.

4.4. Comparison with the Parametric F-test

Note that F-test is possible only when the alternative hypothesis is specifically postulated. Therefore, best power for F-test can achieve occurs when the data are generated from the postulated model. Otherwise, power of F-test will be clearly poor. To see this, we do F-test in two ways. Let $\mathrm{H}_0 : \beta_0 + \beta_1 x$ vs. $\mathrm{H}_1 : \beta_0 + \beta_1 x + \beta_2 x^2$. First, if data are generated under H_1 for $\beta_2 = 0.0\ (0.2)\ 1.0$, then power of F-test, say F1, can be computed by the noncentral F-distribution. Therefore, power of F1 is best possible that F-test can achieve. Next, we generate data from $\beta_0 + \beta_1 x + \beta_2 x e^{-2x}$ and compute power of F-test, say F2, through the Monte Carlo approximation. We compare power of F1 and F2 with LC and ES when error terms follow $N(0, \sigma^2)$ with $\sigma = .05$. As shown in Figure 4, LC is still more powerful than F1, and F1 is much better than F2. Also, F1 is slightly better than ES when $0 < \beta_2 < .10$.

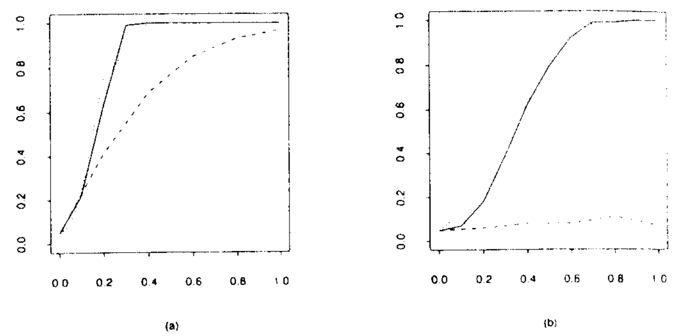

Figure 4: Empirical Powers of ES ($\cdots\cdots$), F-test (- - - -), and LC (——) in 1,000 Samples of Size 100 With $f(x) = \beta_0 + \beta_1 x$ and $g(x) = \beta_2 x^2$. In (a), Power of F-test Is Computed When Random Numbers Are Generated From $\beta_0 + \beta_1 x + \beta_2 x^2$, and in (b), Power of F-test Is Computed When Random Numbers Are Generated From $\beta_0 + \beta_1 x + \beta_2 x e^{-2x}$.

4.5. Comparision with LMP Tests

Under Bayesian model, CK and BU are LMP tests. To compare the local behavior of LC with that of CK, we use two types of data; one from the Bayesian model and the other from the (non-Bayesian) polynomial model. We can anticipate that CK outperform LC under Bayesian model.

First, under Bayesian model, we consider

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 Z(x_i) + \epsilon_i,\ i = 1, \ldots, 100,$$

where $Z(x_i)$ is the i-th component of random vector from the multivariate normal distribution with mean zero and variance-covariance matrix $\mathbf{Q}$ defined in Section 2, and ϵ_i is *iid* normal with mean zero and variance σ^2. Also, $Z(x_i)$'s and ϵ_i's are independent. β_0 and β_1 are set as 1 and σ= .05, and powers are estimated for β_2 =.02 (.02) .10 (.20) 1.0. As shown in Figure 5(a), CK is much more powerful than LC.

To see what will happen under polynomial model, we consider

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i,\ i = 1, \ldots, 100,$$

which was used in Section 4.3 (ii). Again, we set β_0 = β_1 =1, σ=.05. In this case, LC outperform CK uniformly. (See Figure 5(b)).

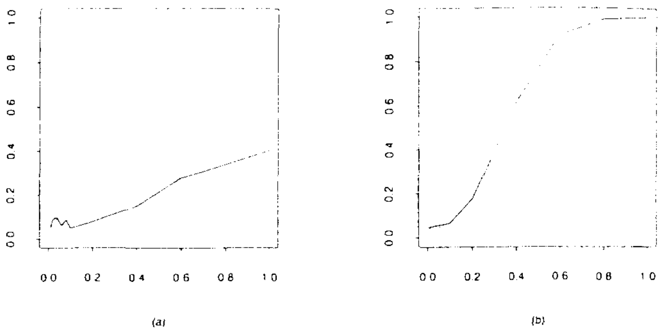

Figure 5: Empirical Powers of CK (······) and LC (——) in 1,000 Samples of Size 100. Bayesian Model and Polynomial Model are Used in (a) and (b), respectively.

5. EXAMPLES

As illustrative examples, we apply our procedure to three sets of data; one is linear and others are nonlinear under H_0.

5.1. Salary Data

First, consider the average salaries of teachers in public elementary and secondary schools in the U.S. from 1964-1965 to 1974-1975 (Gunst and Mason, 1980). This data set were used by Eubank and Spiegelman (1990) with $f(x) = \beta_0 + \beta_1 x$, and they concluded that a linear model does not seem appropriate for this data. As shown in Figure 6, it is more reasonable to set $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ by the scatter diagram. We apply ES, E1, and LC(with B=100) to $H_0 : f(x) = \beta_0 + \beta_1 x$ and $H_0 : f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$, and results are summarized in Table 2. All the three tests give the same results at $\alpha = .05$. We conclude that second order polynomial fit is enough to the Salary data. (See Figure 6)

Table 2. Test Results to Salary Data

null model	test statistic	rejection region	result
$\beta_0 + \beta_1 x$	ES = 13.38	ES > 2.2	Reject H_0
	E1 = 1	E1 $\geq$ 1	Reject H_0
	LC = 100	LC $\geq$ 100	Reject H_0
$\beta_0 + \beta_1 x + \beta_2 x^2$	ES= 0.70	ES> 2.2	Do not reject H_0
	E1 = 0	E1 $\geq$ 1	Do not reject H_0
	LC = 47	LC $\geq$ 100	Do not reject H_0

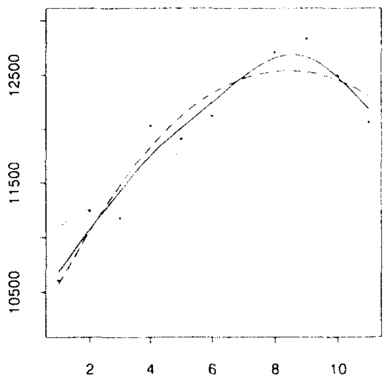

Figure 6: Fits to the Salary Data : • • •, Response; ······, Linear Fit; - - - - , Quadratic Fit; ——, Spline Fit.

5.2. Prognosis Index Data

As a nonlinear case, we adopt prognosis index data in Neter, Wasserman, and Kutner (1989, p.552). The predictor variable is number of days of hospitalization, and the response variable is index for long-term recovery. The model used was

$$y_i = \beta_0 \exp(\beta_1 x_i) + \varepsilon_i$$

under H_0. Here, $n = 15$ and LC is applied for B=100 resamples, and get LC=61. Since the rejection region from (4.2) is LC $\geq$ 100, we can not reject H_0 at $\alpha = 0.05$, and Figure 7 shows that it is hard to reject H_0.

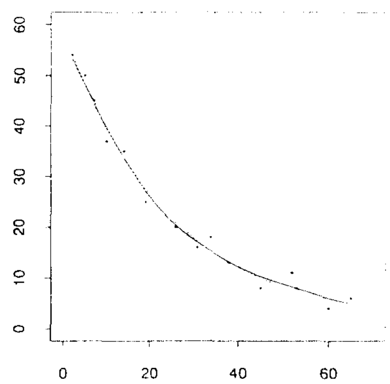

Figure 7: Fits to the Prognosis Index Data : • • •, Response; ······, Nonlinear Fit; ——, Spline Fit.

5.3. Chloride Data

Sredni (1970) analyzed data on chloride ion transport through blood cell walls using the nonlinear regression model

$$y_i = \beta_0(1 - \beta_1 e^{-\beta_2 x_i}) + \varepsilon_i, \quad i = 1, \ldots, 54 .$$

The data are listed in Bates and Watts (1988, p.276). Bates and Watts (1988) notice that residual plot show sinusoidal pattern. So, they transform the response based on AR(1), and the fit after transformation is more satisfactory than the original fit.

We apply LC to the original data, and get LC = 100 when B=100. The rejection region from (4.2) is LC $\geq$ 98, and the null hypothesis is rejected.

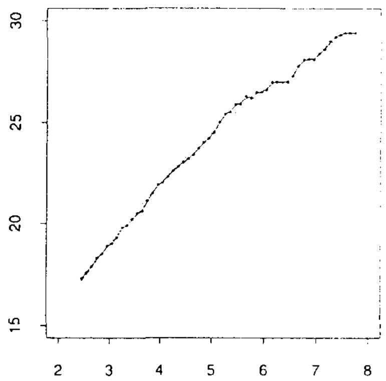

Figure 8: Fits to the Chloride Data : • • •, Response; ······, Nonlinear Fit; ——, Spline Fit.

6. CONCLUDING REMARKS

Goodness-of-fit tests are frequently employed by statisticians when they fit a parametric model to data. Parametric F-test is available if the alternative is given. Many nonparametric tests are suggested for the general alternative. All these test statistics are based on the residuals obtained from fitting the data to a parametric model. Then, they applied nonparametric fit (smoothing spline, series estimator, or kernel regression) to residuals. One disappointing story is that critical values for these statistics are not at hand because the convergence is so slow that the asymptotic distribution for those statistics cannot be used. Instead, the Monte Carlo approximation was used. However, the critical values obtained in this way cannot be expressed as a simple formula (except E1), so that they must be evaluated whenever the size of data n changes. Another drawback in these tests (except HM) is that the null model must be linear. Also, some tests like ES, E2, E3, and HM require the amount of smoothing. This can be done by arbitrary assignment or replaced by the data-driven estimator, however, it is quite dangerous as argued in Section 3.

In this paper, we propose a new test statistic (LC) based on the magnitude of the smoothing parameter estimator, and compare power of LC with others under various situations. We show that ES, HM, and LC are more powerful than others, and are robust in the sense of alternative specification. But, ES and HM contain unknown smoothing parameter. In addition, LC is more powerful than F-test when the alternative is specified. Further, our test can also be used without any modification when the null is nonlinear. One possible disadvantage of our test is using the resampling technique to compute test statistic. We believe the resampling step is not a big deal to achieve good power, ready-to-use critical values, and no restrictions in the specification of null (parametric) model.

References

Bates, D. and Watts, G. (1989), *Nonlinear Regression Analysis and Its Application*, John Wiley, New York.

Buckley, M.J. (1991), "Detecting a Smooth Signal: Optimality of Cusum Based Procedures", *Biometrika*, 78, 253-262.

Cox, D. D., Koh, E., Wahba, G., and Yandell, B. (1988), "Testing the (Parametric) Null Hypothesis in (Semiparametric) Partial and Generalized Spline Models", *The Annals of statistics*, 16, 113-119.

Demmler, A. and Reinsch, C. (1975), "Oscillation Matrices With Spline Smoothing," *Numerische Mathematik,* 24, 375-382.

Eubank, R.L. and Hart, J.D. (1992), "Testing Goodness-of-fit in Regression Via Order Selection Criterion", *The Annals of Statistics*, 20, 1412-1425.

Eubank, R.L. and Hart, J.D. (1993), "Commonality of Cusum, von Neumann and Smoothing-based Goodness-of-fit Tests", *Biometrika*, 80, 89-98.

Eubank, R.L. and Spiegelman, C.H. (1990), "Testing the Goodness of Fit of a Linear Model Via Nonparametric Regression Techniques", *Journal of the American Statistical Association*, 85, 387-392.

Gu, C. (1989), "Rkpack and Its Applications: Fitting Smoothing Spline Models", Technical Report, No. 857, Department of statistics, University of Wisconsin-Madison.

Gunst, R. and Mason, R. (1980), *Regression Analysis and Its Application: A data Oriented Approach*, New York: Marcel Dekker.

Härdle, W. and Mammen, E. (1992), "Comparing Nonparametric Regression versus Parametric Regression Fits", *The Annals of Statistics*, 21, 1926-1947.

Munson, P.J. and Jernigan, R.W. (1989), "A Cubic Spline Extension of the Durbin-Watson Test", *Biometrika*, 76, 39-47.

Neter, J., Wasserman, W., and Kutner, M.H. (1989), *Applied Linear Regression Models*, Irwin; Boston.

Neumann, J. Von (1941), "Distribution of the ratio of the mean squared successive difference to the variance", *The Annals of Mathematical Statistics*, 12, 367-395.

Rice, J. (1984), "Bandwidth Choice for Nonparametric Regression", *The Annals of Statistics*, 12, 1215-1230.

Sredni, J (1970), "Problems of Design, Estimation, and Lack of Fit in Model Building", Ph.D Thesis, University of Wisconsin-Madison.

Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia; SIAM.