The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 39-46

Comparisons between Goodness-of-Fit Tests for Parametric Model via Nonparametric Fit[1)]

Choongrak KIM[2)], Changkon HONG[3)], Meeseon JEONG[4)]

Abstract

Most of existing nonparametric test statistics are based on the residuals which are obtained by regressing the data to a parametric model. In this paper we compare power of goodness-of-fit test statistics for testing the (null) parametric model versus the (alternative) nonparametric model.

1. Introduction

It is quite often to test the goodness-of-fit of the postulated model when one fits a regression 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 inconsistent against many other alternatives, especially against those which are orthogonal to the specific 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 compare power of these tests.

Consider the regression model

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

where $0 \le x_1 < \ldots < x_n \le 1$ are fixed design points, f is unknown smooth regression function, and the ε_i's are independent and identically distributed random variables with

---

1) This research is supported by the Basic Science Research Institute Program, Ministry of Education, 1996, Project Number No. 1418.
2) Associate Professor, Department of Statistics, Pusan National University, Pusan, Korea.
3) Assistant Professor, Department of Statistics, Pusan National University, Pusan, Korea.
4) Department of Statistics, Pusan National University, Pusan, Korea.

$E(\varepsilon_i)=0$ and $Var(\varepsilon_i)=\sigma^2<\infty$. We wish to test

$$H_0: f=f_0,$$

where f_0 has a parametric (linear or nonlinear) form. Most of the existing nonparametric tests assume f_0 is linear.

In Section 2 most of the nonparametric tests and their critical values are introduced. In Section 3 we compare power of these tests under alternatives. Concluding remarks are given in Section 4.

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 $\mathbf{r}$. Methods of nonparametric regression are smoothing spline, series estimator and kernel regression. Before introducing nonparametric test statistics, it is helpful to introduce unified notations. Let $\mathbf{Q}=\{Q(x_i,x_j)\}_{i,j=1,\ldots,n}$ be $n\times n$ matrix of a covariance kernel, $\mathbf{M}(\lambda)=\mathbf{Q}+n\lambda\mathbf{I}$, and $\mathbf{X}=\{x_i^{j-1}\}_{i=1,\ldots,n;j=1,\ldots,m}$ be $n\times m$ matrix. 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

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

as a test statistic.

Based on a Bayesian model, Cox, et. al. (1988) suggested the locally most powerful (LMP) test when the null parametric regression function f_0 is a polynomial of degree $m-1$ or less. It rejects H_0 if $CK=\mathbf{r}'\mathbf{Q}\mathbf{r}$ is too large.

Under the same perspective as Cox, et. al. (1988), Buckley (1991) derived an LMP test, which is based on the test statistic

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

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

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

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

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 $0 < \theta_3 \le \theta_4 \le \cdots \le \theta_n$ are eigenvalues defined in Demmler and Reinsch(1975). Here, λ must be preassigned.

When the Fourier series estimator is used, an estimate of the risk is

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

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, a test is given by " reject H_0 if $\mathrm{E1} \ge 1$ ".

Later, instead of E1, 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.,

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

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

$$\mathrm{HM} = nh^{1/2}\int_0^1 (\hat{m}_{h(x)} - K_h \hat{f}(x))^2 dx,$$

where $K_h g(\cdot) = \sum K_h(\cdot - X_i) g(X_i) / \sum K_h(\cdot - X_i)$, $K_h(\cdot) = h^{-1}K(\cdot / h)$ 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. However, by the generic property of nonparametric approach, it is impossible to get exact critical values. Fortunately, asymptotic distributions have been derived for most of the nonparametric test statistics. Asymptotic distributions 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 both their simulation studies on power of tests and 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 distribution 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 we found that the 95-th percentile 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. Simulation Results for Power Comparisons

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

(i) $f_0(x)=\beta_0(constant)$; $n = 100$

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

$$f_1(x)=f_0(x)+\beta_1\ \{e^{4x}-(e^4-1)/4\}\ \{\frac{e^8-1}{8}-(\frac{e^{4-1}}{4})^2\}^{-1/2}$$

$$f_2(x)=f_0(x)+2\beta_1\ \{20(x-\frac{1}{2})^3-3(x-\frac{1}{2})\}$$

Powers are estimated for VN, BU, ES, E1, E2, E3, and HM 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. Results are summarized in Table 1. In the case of f_1, BU, E3, and HM are better than others, but others are not bad at all. In the case of f_2, ES, and HM 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.

Table 1. Proportion of Rejections in 1,000 Samples of Size 100 for f_1 and f_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
f_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
	VN	.087	.340	.802	.974
	BU	.045	.134	.450	.849
	ES	.346	.922	1.000	1.000
f_2	E1	.088	.288	.769	.976
	E2	.160	.643	.972	.998
	E3	.169	.701	.984	.999
	HM	.180	.750	.990	1.000

(*ii*) $f_0(x)=\beta_0+\beta_1 x$; $n=100$

From (i), we found that ES and HM are quite powerful in both f_1 and f_2. Here, we set the null model as linear instead of constant, and choose the same alternative models as in Eubank and Spiegelman (1990).

$$f_1(x)=f_0(x)+\beta_2 x e^{-2x}$$

$$f_2(x)=f_0(x)+\beta_2 x^2$$

We compute powers of ES, E1, and HM. 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 1(a), (b), (c) show power of four tests for $\sigma=.05, .10, .20$, respectively in f_1. As in (i), ES, and HM perform similarly, and better than E1 for all σ. The same phenomenon occurs in f_2, too. (See Figure 2.)

4. 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 common 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 compare power of those test statistics under various situations. We show that ES and HM are more powerful than others, and are robust in the sense of alternative specification. But, ES and HM contain unknown smoothing parameter.

References

[1] Buckley, M.J. (1991). Detecting a smooth signal: Optimality of cusum based procedures, *Biometrika*, 78, 253-262.

[2] 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.

[3] Demmler, A. and Reinsch, C. (1975). Oscillation matrices with spline smoothing, *Numerische Mathematik*, 24, 375-382.

[4] 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.

[5] Eubank, R.L. and Hart, J.D. (1993). Commonality of cusum, von Neumann and smoothing-based goodness-of-fit tests, *Biometrika*, 80, 89-98.

[6] 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.

[7] Gu, C. (1989). Rkpack and its applications: Fitting smoothing spline models, Technical Report, No. 857, Department of Statistics, University of Wisconsin-Madison.

[8] Härdle, W. and Mammen, E. (1992). Comparing nonparametric regression versus parametric regression fits, *The Annals of Statistics*, 21, 1926-1947.

[9] Munson, P.J. and Jernigan, R.W. (1989). A Cubic spline extension of the Durbin-Watson test, *Biometrika*, 76, 39-47.

[10] 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.

[11] Rice, J. (1984). Bandwidth choice for nonparametric regression, *The Annals of Statistics*, 12, 1215-1230.

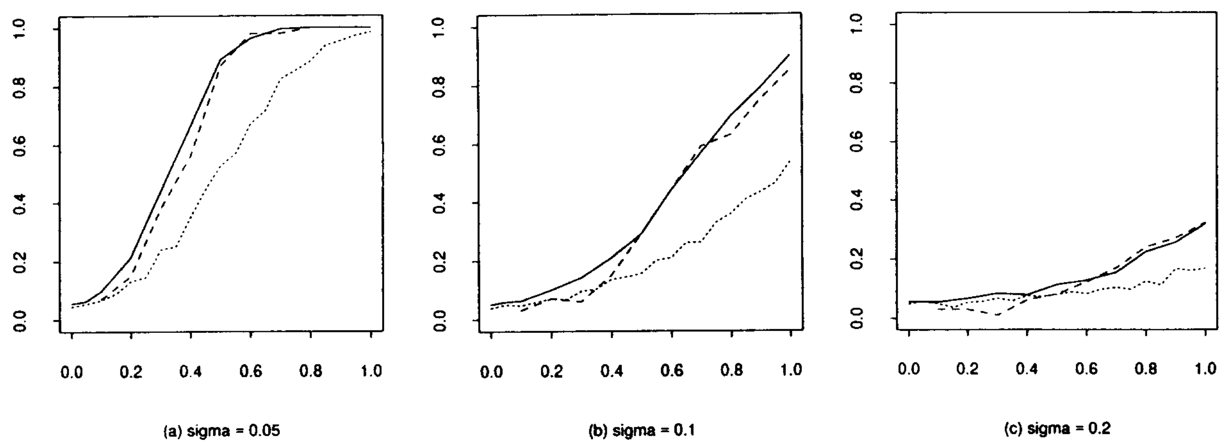

Figure 1 Empirical Powers of ES (———), E1 (⋯⋯), and HM (-----) in 1,000 Samples of Size 100 With $\mu_1(x) = \beta_0 + \beta_1 x + \beta_2 x e^{-2x}$; (a), (b), (c) Correspond to $\sigma =$ .05, .10, .20, respectively.

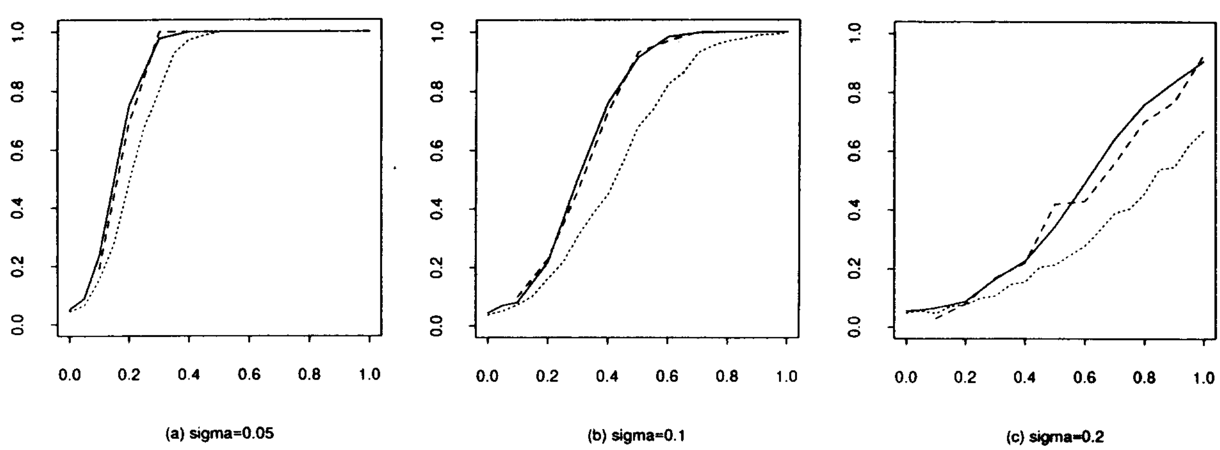

Figure 2 Empirical Powers of ES (———), E1 (⋯⋯), and HM (-----) in 1,000 Samples of Size 100 With $\mu_1(x) = \beta_0 + \beta_1 x + \beta_2 x^2$; (a), (b), (c) Correspond to $\sigma =$ .05, .10, .20, respectively.