Journal of the Korean Statistical Society (1998), **27**: *3, pp 305–318*

Testing for Lack of Fit via the Generalized Neyman Smooth Test

GeungHee Lee[1]

ABSTRACT

Smoothing tests based on an L_2 error between a truncated Fourier series estimator and a true function have shown good powers for a wide class of alternatives. These tests have the same form of the Neyman smooth test whose performance depends on the selected order, a basis, the form of estimators. We construct flexible data driven Neyman smooth tests by changing a basis, combining model selection criteria and different series estimators. A simulation study shows that the generalized Neyman smooth test with the best basis provides good power for a wider class of alternatives compared with other data driven Neyman smooth tests based on a fixed form of estimator, a fixed basis and a fixed criterion.

Keywords: L_2 error test; Model selection; Fourier series; Wavelets; Best basis

1. INTRODUCTION

Suppose that we have data $y_1, \ldots, y_n$ from the following model:

$$y_i = f(\frac{i-0.5}{n}) + \epsilon_i \qquad i = 1, \ldots, n, \tag{1.1}$$

where $\epsilon_1, \ldots \epsilon_n$ are *i.i.d.* normal random variables with mean zero and finite known variance σ^2 and f is a function defined on [0,1]. Having selected a model, we should check the model for lack of fit. Under the model in (1.1), the lack-of-fit hypothesis for the residuals y_i can be expressed as follows:

$$H_0 \ : \ f \text{ is constant vs. } H_a \ : \ f \text{ is not constant.} \tag{1.2}$$

A number of parametric tests can be used to detect specified departures from H_0. We call these parametric tests directional tests. For a wider class of alternatives, tests using the idea of nonparametric function estimation have been developed.

[1]The Bank of Korea, 110, 3-Ga, Namdaemun-Ro, Jung-Gu, Seoul, Korea.

If a possible alternative function f is smooth enough, the Fourier coefficients decay quickly to zero as the frequency becomes high. In this case, the following truncated series estimator, usually provides a reasonable estimate of $f(x)$:

$$\hat{f}_1(x) = \hat{a}_0 + 2\sum_{j=1}^{k} \hat{a}_j \cos(\pi j x), \tag{1.3}$$

where $\hat{a}_j = \frac{1}{n}\sum_{i=1}^{n} \cos(\frac{i-0.5}{n})y_i$ and k is a fixed constant. If f is a constant, an L_2 error can be estimated by $\int_0^1 (\hat{f}_1(x;k) - \bar{Y})^2 dx$, which becomes $2\sum_{j=1}^{k} \hat{a}_j^2$ by Parseval's relation. Considering the variance of $\hat{a}_j$, we can obtain a standard form for the L_2 error, $\frac{2n}{\sigma^2}\sum_{j=1}^{k} \hat{a}_j^2$. The L_2 error between the truncated series estimator and a constant has the same form as a Neyman smooth statistic,

$$T_{Neyman} = \frac{2n}{\sigma^2}\sum_{j=1}^{k} \hat{a}_j^2. \tag{1.4}$$

The performance of the test depends on the selected order k and the set of basis functions used for estimating f.

If we choose the wrong k, however, the result is low power for T_{Neyman}. A subjective choice of k often gives poor results. It is often difficult to detect high order behavior subjectively, i.e., by eye. We need to find an objective and data-adaptive rule to select an order k. The Neyman smooth test based on data-adaptive k is called the data-driven Neyman smooth test. If our data comes under H_0 in (1.2), the order k is selected as zero because of the property of data-driven criteria; otherwise, the order k is greater than zero. This order selection based on data-driven criteria enlarges the alternative class. Adding the order selection to the Neyman smooth test makes the test more powerful for a broader class. Kuchibhatla and Hart (1996) and Ledwina (1994) construct data driven Neyman smooth tests using the AIC(Akaike Information Criterion) and BIC(Beysian Information Criterion) type automatic order selection criteria to select k in $\hat{f}_1$, respectively. Since each criterion has its own merits and demerits, it is desirable to choose a suitable model selection criterion.

To represent f in (1.1) efficiently, we have to choose the suitable basis instead of the fixed basis. If the data have periodic behavior, sine and cosine series estimators provide an efficient means of explaining the data. On the other hand, Haar wavelets give a good expression for blocky data. By choosing the best basis, we can improve the performance of T_{Neyman}.

In this paper, we modify the data driven Neyman smooth test in order to improve its performance for a wide class of alternatives using different bases, such as wavelets, and different criteria. By selecting the best basis and combining AIC and BIC, we can generalize Neyman smooth tests which perform well for a wider class of alternatives compared with other data driven Neyman smooth tests based on a fixed basis or a fixed criterion.

To combine AIC, BIC and a threshold, we construct a general series estimator and a new type of data driven Neyman smooth test in Section 2. In Section 3, we apply a different basis - wavelets - to data driven Neyman smooth tests. In Section 4, we improve the test by selecting the best basis from a library that consists of several bases. We compare several Neyman smooth tests by the simulation study in Section 5. Section 6 gives the conclusion.

2. DATA DRIVEN NEYMAN SMOOTH TEST WITH A GENERAL SERIES ESTIMATOR

Often we are mainly interested in low frequency behavior such as a trend. To check these low frequency departures, traditional nonparametric tests such as the Kolmogorov Smirnov and Cramér-Von Mises tests may be used. These tests suffer from poor performance for high frequency alternatives. Fan (1996) points out these problems and suggests data driven Neyman smooth tests based on a thresholded estimator. His tests perform well for a dominant frequency alternative but have some restriction in detecting a low frequency alternative with a moderate size. To solve this problem, Lee and Hart (1998) suggest a hybrid test based on an L_2 error which uses AIC criterion and a threshold.

It is always difficult to choose a suitable model selection criterion. When we focus on AIC and BIC, AIC has good small sample properties but is not consistent, whereas BIC is consistent but has a slow convergence rate. The orders $\hat{k}_{AIC}$ and $\hat{k}_{BIC}$ based on AIC and BIC are chosen as the maximzer of the following criterion with $C = 2$ and $C = \log n$, respectively:

$$r(k) = \begin{cases} 0, & \text{for } k = 0 \\ \sum_{j=1}^{k} \hat{a}_j^2 - C\frac{k}{2n}\sigma^2, & \text{for } k = 0, 1, \dots, n-1. \end{cases} \tag{2.1}$$

We can construct a more general estimator $\hat{f}_2$ using the consistency of BIC and the tendency that AIC chooses an overfitted model as follows:

$$\hat{f}_2(x) = \begin{cases} \hat{a}_0 + 2\sum_{j=1}^{\hat{k}_{BIC}} \hat{a}_j \cos(\pi j x) + 2\hat{g}_1(x), & \text{if } \hat{k}_{BIC} \geq \hat{k}_{AIC} \\ \hat{a}_0 + 2\sum_{j=1}^{\hat{k}_{BIC}} \hat{a}_j \cos(\pi j x) + 2\hat{g}_2(x), & \text{if } \hat{k}_{BIC} < \hat{k}_{AIC}, \end{cases} \tag{2.2}$$

where $\hat{g}_1(x) = \sum_{j=\hat{k}_{BIC}+1}^{n-1} \hat{a}_j \, I(\frac{|\sqrt{2n}\hat{a}_j|}{\sigma} > \delta_1)\cos(\pi jx)$ and $\hat{g}_2(x) = \sum_{j=\hat{k}_{BIC}+1}^{\hat{k}_{AIC}} \hat{a}_j \, I(\frac{|\sqrt{2n}\hat{a}_j|}{\sigma} > \delta_2)\cos(\pi jx) + \sum_{j=\hat{k}_{AIC}+1}^{n-1} \hat{a}_j I(\frac{|\sqrt{2n}\hat{a}_j|}{\sigma} > \delta_1)\cos(\pi jx)$.
We use $\sqrt{(2+\gamma_1)\log n}$ and $\sqrt{(2+\gamma_2)\log n}$ as δ_1 and δ_2, respectively and γ_1 and γ_2 are positive constants such that $\gamma_1 > \gamma_2$. Based on the L_2 difference between $\hat{f}_2$ and a constant, a data driven Neyman smooth test statistic is constructed as follows:

$$T_{GN} = \begin{cases} \frac{2n}{\sigma^2}[\sum_{j=1}^{\hat{k}_{BIC}} \hat{a}_j^2 + Tg_1], & \text{if } \hat{k}_{BIC} \geq \hat{k}_{AIC} \\ \frac{2n}{\sigma^2}[\sum_{j=1}^{\hat{k}_{BIC}} \hat{a}_j^2 + Tg_2], & \text{if } \hat{k}_{BIC} < \hat{k}_{AIC}, \end{cases} \tag{2.3}$$

where $Tg_1 = \sum_{j=\hat{k}_{BIC}+1}^{n-1} \hat{a}_j^2 I(\frac{|\sqrt{2n}\hat{a}_j|}{\sigma} > \delta_1)$ and $Tg_2 = \sum_{j=\hat{k}_{BIC}+1}^{\hat{k}_{AIC}} \hat{a}_j^2 I(\frac{|\sqrt{2n}\hat{a}_j|}{\sigma} > \delta_2) + \sum_{j=\hat{k}_{AIC}+1}^{n-1} \hat{a}_j^2 \, I(\frac{|\sqrt{2n}\hat{a}_j|}{\sigma} > \delta_1)$. If we have the data with $n \geq 8$, $\hat{k}_{AIC} \geq \hat{k}_{BIC}$ so that $\hat{f}_2$ and T_{GN} can be expressed as the simpler form. If we choose 0 as $\hat{k}_{BIC}$, we set $\hat{k}_{BIC}$ to 1. The limiting distribution and consistency of T_{GN} are given in Theorems 2.1 and 2.2.

Theorem 2.1. *Under the model (1.1) with a constant f and $\delta_1 = \sqrt{(2+\gamma_1)\log n}$ $\delta_2 = \sqrt{(2+\gamma_2)\log n}$ for positive finite constants γ_1 and γ_2 such that $\gamma_1 > \gamma_2$, the statistic T_{GN} in (2.3) converges in distribution to a χ_1^2 random variable as $n \to \infty$.*

Theorem 2.2. *Suppose that f in (1.1) is such that, for some j, $\lim_{n\to\infty} Pr(|\hat{a}_j| \geq A) = 1$ for some $A > 0$. Then, the test of (1.2) based on T_{GN} is consistent.*

These proofs are given in Appendix.

3. DATA DRIVEN NEYMAN SMOOTH TEST WITH WAVELETS

If the underlying function contains nonsmooth local behavior, Fourier series estimators have difficulty in explaining such behavior. To overcome this problem, we may consider wavelets $\psi_{j,k}$, an orthonormal basis analogous to bases used in Fourier analysis and the following series estimator:

$$\hat{f}_3(x) = \hat{c}_{0,0}\phi(x) + \sum_{j=0}^{j_0-1}\sum_{k=0}^{2^j-1} \hat{d}_{j,k}\psi_{j,k}(x) + \sum_{j=j_0}^{M-1}\sum_{k=0}^{2^j-1} \hat{d}_{j,k}\psi_{j,k}(x) I_{j,k}, \tag{3.1}$$

where ϕ is the scaling function, $M = \log_2 n$, $\hat{c}_{0,0} = \frac{1}{n}\sum_{i=1}^n y_i\phi(\frac{i}{n})$, $\hat{d}_{j,k} = \frac{1}{n}\sum_{i=1}^n y_i\psi_{j,k}(\frac{i}{n})$ and $I_{j,k} = I(\frac{|\sqrt{n}\hat{d}_{j,k}|}{\sigma} > \delta)$. In this vein, Fan (1996) proposes data driven Neyman smooth tests based on wavelet thresholding with fixed j_0 as follows:

$$T_H = \frac{n}{\sigma^2}\sum_{j=0}^{j_0-1}\sum_{k=0}^{2^j-1}\hat{d}_{j,k}^2 + \frac{n}{\sigma^2}\sum_{j=j_0}^{M-1}\sum_{k=0}^{2^j-1}\hat{d}_{j,k}^2 I(\frac{|\sqrt{n}\hat{d}_{j,k}|}{\sigma} > \delta_H), \tag{3.2}$$

where $\delta_H = \sqrt{2\log n/a_n}$, and $a_n = (\log_n)^2$. The normalized tests of T_H are distributed with standard normal distribution under the H_0 in (1.2) (Fan, 1996, Theorem 2.3).

His test cannot detect low frequency behavior well due to a fixed j_0. To solve this problem, it is desirable to use data adaptive j_0. Lee (1997) chooses smoothing parameters j_0 and a threshold δ based on a risk estimator, simultaneously, which estimators provide better fits than those with fixed j_0. Considering wavelet series estimators with a data adaptive j_0, new data driven Neyman smooth tests are constructed as follows:

1. Choose the order $\hat{j}_{AIC}$ and $\hat{j}_{BIC}$ that maximize the following criterion:

$$r(j) = \sum_{i=0}^{j}\sum_{k=0}^{2^i-1}\hat{d}_{i,k}^2 - C\frac{\sigma^2}{n}(2^{j+1}-1) \qquad \text{for } j = -1, 0, \dots, M-1,$$

   where we use $C = 2$ and $C = \log n$ for j_{AIC} and j_{BIC}, respectively and $r(-1) = 0$.

2. If $\hat{j}_{BIC} = -1$, set $\hat{j}_{BIC}$ to 0. Based on the selected $\hat{j}_{AIC}$ and $\hat{j}_{BIC}$, construct a data driven Neyman smooth test as follows:

$$T_{WB} = \begin{cases} \frac{n}{\sigma^2}[\sum_{j=0}^{\hat{j}_{BIC}}\sum_{k=0}^{2^j-1}\hat{d}_{j,k}^2 + Tw_1], & \text{if } \hat{j}_{BIC} \geq \hat{j}_{AIC} \\ \frac{n}{\sigma^2}[\sum_{j=0}^{\hat{j}_{BIC}}\sum_{k=0}^{2^j-1}\hat{d}_{j,k}^2 + Tw_2], & \text{if } \hat{j}_{BIC} < \hat{j}_{AIC}, \end{cases} \tag{3.3}$$

   where $Tw_1 = \sum_{j=\hat{j}_{BIC}+1}^{M-1}\sum_{k=0}^{2^j-1}\hat{d}_{j,k}^2 I(\frac{|\sqrt{2n}\hat{d}_{j,k}|}{\sigma} > \delta_1)$ and $Tw_2 = \sum_{j=\hat{j}_{BIC}+1}^{\hat{j}_{AIC}}\sum_{k=0}^{2^j-1}\hat{d}_{j,k}^2 I(\frac{|\sqrt{2n}\hat{d}_{j,k}|}{\sigma} > \delta_2) + \sum_{j=\hat{j}_{AIC}+1}^{M-1}\sum_{k=0}^{2^j-1}\hat{d}_{j,k}^2 I(\frac{|\sqrt{2n}\hat{d}_{j,k}|}{\sigma} > \delta_1)$.

In this paper, we consider Daubechie wavelets with reguarity $p > 2$ and Haar wavelets. For properties of wavelet coefficients $\hat{d}_{j,k}$ such as their asymptotic normality, see Chapter 3 of Lee (1997). The limiting distribution and consistency of T_{WB} are given in Theorems 3.1 and 3.2.

Theorem 3.1. *Under the assumptions of Theorem 2.1, the statistic* T_{WB} *in (3.3) converges in distribution to a* χ_1^2 *random variable as* $n \to \infty$.

Theorem 3.2. *Suppose that* f *in (1.1) is such that, for some* j *and* k, $\lim_{n\to\infty} Pr(|\hat{d}_{j,k}| \geq A) = 1$ *for some* $A > 0$. *Then, a test of (1.2) based on* T_{WB} *is consistent.*

Using the same proof as for Theorem 2.1, the asymptotic distribution of this test under H_0 can also be derived as χ_1^2. Consistency of this test under H_a can be proved in the similar way to Theorem 2.2. Spokoniny (1996) proposes an adaptive test using wavelets with an adaptive j_0 in the minimax sense.

4. GENERLIZED NEYMAN SMOOTH TEST WITH BEST BASIS

Until now, we consider a fixed basis such as cosine series or Daubechies wavelets with fixed regularity. In this section, we try to find the best basis from the data instead of the fixed basis. To represent the data efficiently, we have to choose the suitable basis. To give more flexibility to choose a best basis, we may consider a library approach. A library of orthonormal bases can be defined as a collection of orthonormal bases.

We assume that a library consists of L bases $\{\xi^1, \xi^2, \cdots, \xi^L\}$ where each basis ξ^k is an orthonormal basis such as Fourier series or Daubechies wavelets with several regularities. Each basis ξ^k has n-elements $\xi^k = \{\xi_1^k, \xi_2^k, \cdots, \xi_n^k\}$. For each ξ^k, a data driven Neyman smooth test T_{ξ^k} can be constructed with the same form as (6). This test is asymtotically distributed as a χ_1^2 random variable under H_0 in (2) and consistent under H_a in (2).

Coifman and Majid (1993) propose a rule for selecting a single best basis among bases in the library based on Shannon entropy which is the popular measure to evaluate the efficiency of bases in terms of data compression. Entropy $E(x)$ is given by

$$E(x) = -\sum_i x_i \log x_i, \qquad \text{for } x_i \geq 0 \text{ and } \sum_i x_i = 1.$$

The basis which minimizes Shannon entropy can represent the data efficiently and is called the best basis. In the case of the cosine series, we calculate entropy using $x_i = \frac{\hat{a}_i^2}{\sum_{j=1}^n \hat{a}_j^2}$ where $\hat{a}_j = \frac{1}{n}\sum_{i=1}^n y_i \cos(\pi j \frac{i-0.5}{n})$. If a function is well approximated by a single basis, this best basis based on entropy represents a function efficiently; otherwise, the basis based on this selection has a restriction in explaining a function. To overcome this problem, Chen and Donoho (1994) suggest a new algorithm – *Basis Pursuit*. In this section, we focus on selecting the single best basis based on entropy and assume that the underlying function can be expressed as the truncated series form based on the selected basis. As the number of bases increases, the probability of selecting a wrong basis increases. Thus, it is necessary to restrict the number of bases as much as possible based on a prior knowledge for the data. We can construct a data driven Neyman smooth test T_{BB} using the best basis and the general series estimator $\hat{f}_2$ as follows:

1. Construct a library based on data.
2. Choose the best basis from the library based on entropy.
3. For the selected basis, a data driven Neyman smooth test statistic based on $\hat{f}_2$ (T_{BB}) is obtained.

We call this test the generalized data driven Neyman smooth test. Under H_0, this test statistic has an asymptotically χ_1^2 distribution regardless of bases, due to BIC, while under H_a, this test performs better than the test with fixed basis due to the best basis selection algorithm. The limiting distribution and consistency of T_{BB} are given in Theorems 4.1 and 4.2.

Theorem 4.1. *Under the assumptions of Theorem 2.1, the statistic T_{BB} converges in distribution to a χ_1^2 random variable as $n \to \infty$.*

Theorem 4.2. *Suppose that f in (1.1) is such that $\lim_{n\to\infty} Pr(|\hat{\xi}_j^k| \geq A) = 1$, $\hat{\xi}_j^k = \frac{1}{n}\sum_{i=1}^n \xi_j^k(\frac{i}{n})y_i$ for some $A > 0, k, j$. Then, a test of (1.2) based on T_{BB} is consistent.*

These proofs are given in Appendix.

5. SIMULATION STUDY

A simulation study was conducted to compare several data driven Neyman smooth tests (Table 5.1). The data driven Neyman smooth test T_{KH} is proposed by Kuchibhatla and Hart (1995). As thresholds for the other tests, we used $\delta_1 = \sqrt{3\log n}$ and $\delta_2 = \sqrt{2.1\log n}$. For T_{WB} and T_H, we used Daubechies wavelets with regularity 6. In the case of T_H, we used a fixed $j_0 = 0$ in $\hat{f}_3$ in (3.2). For T_{BB}, we constructed a library with three bases – Haar wavelets,

Table 5.1: Structure of data driven Neyman smooth tests

Name	Basis	Criterion
T_{KH}	Fourier	AIC
T_{GN}	Fourier	AIC, BIC and Threshold
T_H	Wavelets	Threshold
T_{WB}	Wavelets	AIC, BIC and Threshold
T_{BB}	Library	AIC, BIC and Threshold

Daubechies wavelets with regularity 6 and cosine series. To obtain the the critical values of the tests, 10,000 sample sets with the sample size 128 were generated from a standard normal distribution. Table 5.2 shows the empirical percentile of 5 data driven Neyman smooth tests.

Table 5.2: Empirical percentile of data driven Neyman smooth tests under H_0

Percentile	T_{KH}	T_{GN}	T_H	T_{WB}	T_{BB}
99	30.357	17.016	82.181	16.544	18.732
	(0.995)	(0.306)	(0.878)	(0.242)	(0.279)
98	22.645	14.842	76.820	13.932	16.528
	(0.475)	(0.430)	(0.668)	(0.693)	(0.210)
97	19.285	12.029	73.381	8.144	15.176
	(0.447)	(0.400)	(0.425)	(0.846)	(0.200)
96	16.712	9.753	70.667	5.973	12.716
	(0.415)	(0.495)	(0.400)	(0.279)	(0.554)
95	14.863	7.563	68.800	5.069	10.740
	(0.381)	(0.230)	(0.332)	(0.159)	(0.667)
90	9.002	4.118	62.432	3.169	4.173
	(0.221)	(0.127)	(0.372)	(0.066)	(0.111)
75	2.693	1.606	52.235	1.481	1.603
	(0.103)	(0.030)	(0.212)	(0.031)	(0.033)

Note: Standard errors are given in parentheses.

To compare the power of the five tests, we generated 1,000 samples from the following 3 underlying functions with standard normal errors and different amplitudes:

$$f(x) = r(\frac{e^{5x}}{1+e^{5x}} - 1) \quad (5.1)$$

$$f(x) = r\cos(80\pi x) \quad (5.2)$$

$$f(x) = \begin{cases} 0, & \text{if } x \leq \frac{54}{128} \text{ or } x > \frac{55}{128} \\ r, & \text{if } \frac{54}{128} < x \leq \frac{55}{128} \end{cases} \quad (5.3)$$

The results of the simulation study are shown in Figures 5.1.- 5.3. The range of standard errors for empirical power is from 0.0 to 0.016. We took the Type I error probability to be 0.05. T_{GN} performs better than T_{KH} for low, high frequency data. For high frequency data, T_{KH} has low power due to the failure of the order selection by AIC. T_{WB} is superior to T_H due to a data adaptive j_0 for low frequency data. For the spike data, T_{WB} and T_H work better than data driven Neyman smooth tests based on Fourier series (T_{KH} and T_{GN}), because wavelets better detect local behavior. For low, or high frequency data, T_{GN} is better than T_{WB} since these functions are well approximated by Fourier series. The performance of T_{GN} and T_{WB} depends on the alternative function, while T_{BB} has stable power for all alternatives, because the best basis algorithm chooses a basis properly. Since the best basis algorithm selects Haar wavelets for the spike data, T_{BB} is superior to the other tests. When the variance σ^2 in (1) is unknown, we can use the consistent variance estimator based on the difference for a continuous function f (Rice, 1984; Hall, Kay and Titterington, 1990). For the test with wavelets, robust variance estimators of Donoho and Johnstone (1994) can be considered.

6. CONCLUSION

In this paper, we generalized data driven Neyman smooth tests to obtain good power for a broad class of alternatives. Data driven Neyman smooth tests depend on 3 components - the selected smoothing parameters, the type of the estimator and the selected basis. We proposed several data driven Neyman smooth tests by combining the model selection criterion, the form of estimators and the best basis, The simulation study showed that the generalized data driven Neyman smooth test performed well for a broader class of alternatives than other data driven Neyman smooth tests. The proposed tests would be extended via better best basis algorithms and model selection criteria.

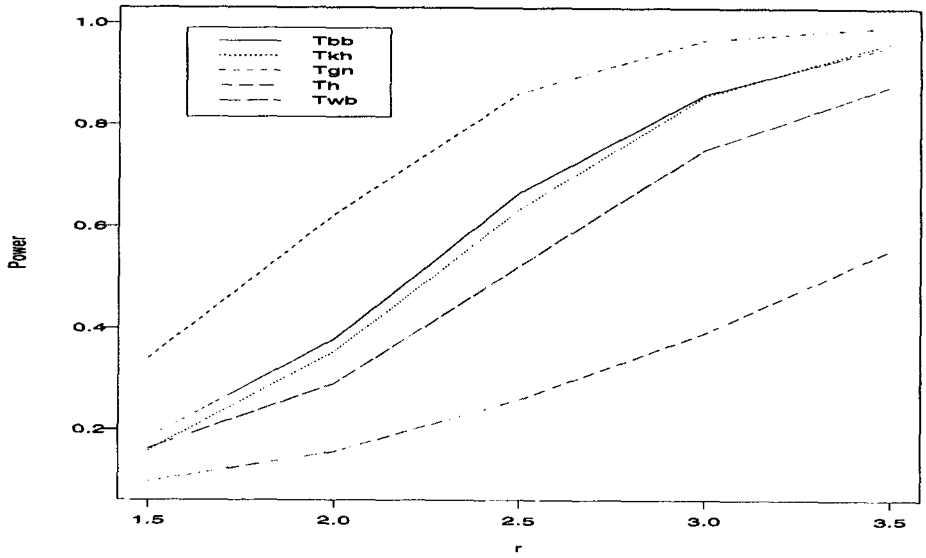

Figure 5.1: Empirical power of 5 tests for $f = r(\frac{e^{5x}}{1+e^{5x}} - 1)$

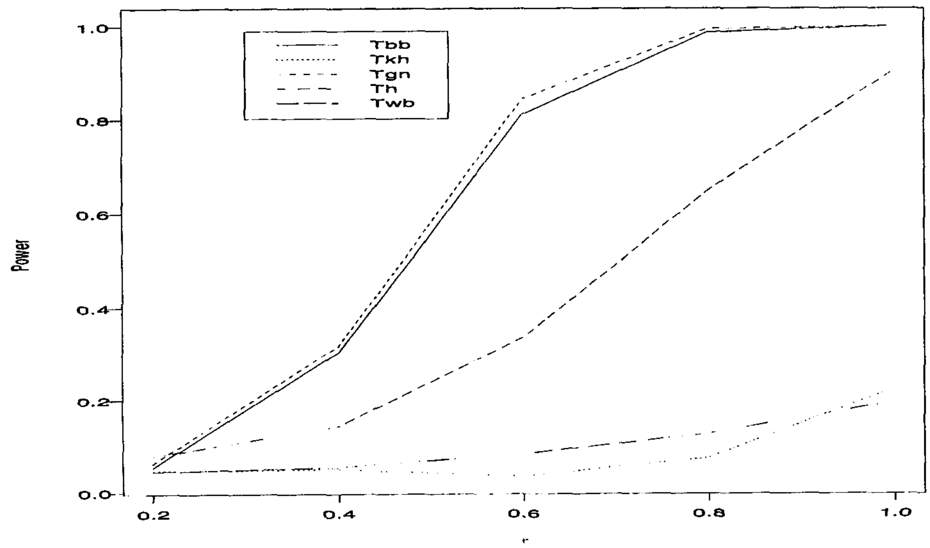

Figure 5.2: Empirical power of 5 tests for $f = r\cos(80\pi x)$

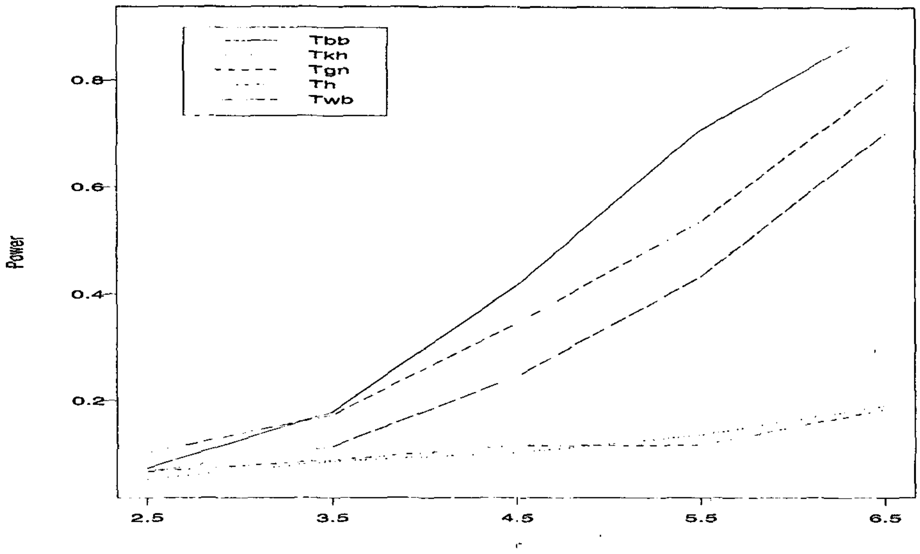

Figure 5.3: Empirical power of 5 tests for spike data

ACKNOWLEDGMENTS

The author is grateful to Jeffrey D. Hart, an associate editor, and a referee for valuable comments and discussion.

APPENDIX

Proof of Theorem 2.1. To prove Theorem 2.1, we need the following lemmas A.1 and A.2 which come from Hart (1997) and Fan (1996), respectively. The data driven Neyman smooth test based on BIC, T_L is given by

$$T_L = \begin{cases} \frac{2n}{\sigma^2}\sum_{j=1}^{\hat{k}_{BIC}} \hat{a}_j^2 & \text{if } \hat{k}_{BIC} \geq 1 \\ \frac{2n}{\sigma^2}\hat{a}_1^2 & \text{if } \hat{k}_{BIC} = 0. \end{cases}$$

Lemma A.1. *Under the model (1.1) with a constant f, the statistic T_L converges in distribution to a χ_1^2 random variable as $n \to \infty$.*

Lemma A.2. *Assume that X is normally distributed with mean zero and variance one. Let δ be a chosen threshold and $Y = X^2 I(|X| > \delta)$. Then*

$$EY^k = \sqrt{2/\pi}\delta^{2k-1}\{1 + (2k-1)\delta^{-2} + O(\delta^{-4})\}\exp(-\delta^2/2).$$

When $\delta = \sqrt{(2+\gamma)\log n}$ for a positive finite constant γ, $EY = O(\delta \exp(-\delta^2/2))$ $= O(\frac{\sqrt{\log n}}{n^{1+\gamma/2}})$.

Take $\sigma^2 = 1$ without loss of generality. T_{GN} can be expressed as the sum of T_L and Tg_1 or Tg_2, where $Tg_1 = 2n\sum_{j=\hat{k}_{BIC}+1}^{n-1} \hat{a}_j^2 I(\sqrt{2n}|\hat{a}_j| > \delta_2)$ and $Tg_2 = 2n\sum_{j=\hat{k}_{BIC}+1}^{\hat{k}_{AIC}} \hat{a}_j^2 I(\sqrt{2n}|\hat{a}_j| > \delta_2) + 2n\sum_{j=\hat{k}_{BIC}+1}^{n-1} \hat{a}_j^2 I(\sqrt{2n}|\hat{a}_j| > \delta_1)$. The $\sqrt{2n}\hat{a}_j$'s are *i.i.d.* as Normal(0,1) under H_0. By Lemma A.1, T_L converges in distribution to χ_1^2. By Markov's Theorem and Lemma A.2, for any $\epsilon > 0$

$$
\begin{aligned}
P(Tg_1 > \epsilon) &\leq \frac{E(Tg_1)}{\epsilon} \\
&\leq \frac{\sum_{j=1}^{n-1} E(2n\hat{a}_j^2 I(\sqrt{2n}|\hat{a}_j| > \delta_1))}{\epsilon} \\
&= \frac{n-1}{\epsilon} O(\frac{\sqrt{\log n}}{n^{1+\gamma_1/2}}) \\
&= \frac{1}{\epsilon} O(\frac{\sqrt{\log n}}{n^{\gamma_1/2}}) \to 0.
\end{aligned}
$$

Similary, $P(Tg_2 > \epsilon) \to 0$. Therefore, by Slutsky's Theorem, T_{GN} converges in distribution to χ_1^2.

Proof of Theorem 2.2. For Theorem 2.2, we need the following Lemma A.3 whose proof is given in Hart (1997). □

Lemma A.3. *Suppose that we have f in (1.1) such that for some j, $\lim_{n\to\infty}$ $Pr(|\hat{a}_j| \geq A) = 1$ for some $A > 0$. Then, T_L is consistent.*

The rejection region of a test based on T_{GN} is asymptotically the same as that of T_L by Theorem 1. If C_α is the asymtotic critical value, then

$$P(T_{GN} \geq C_\alpha) \geq P(T_L \geq C_\alpha) \to 1.$$

Thus, the power of T_{GN} tends to 1 as $n \to \infty$ under any alternative satisfying the assumptions of Theorem 2.2.

Proof of Theorem 4.1. Assume that a library consists of L bases and take $\sigma^2 = 1$ without loss of generality. Let $E_i = \{\xi^i | Entropy(\xi^i) \leq Entropy(\xi^j), i \neq j\}$. Then T_{BB} can be expressed as follows:

$$T_{BB} = \sum_{i=1}^{L} T_{\xi^i} I(E_i)$$

where I is the indicator function. For any $x > 0$,

$$\begin{aligned} P(T_{BB} \leq x) &= \sum_{k=1}^{L} P(\sum_{i=1}^{L} T_{\xi^i} I(E_i) \leq x | E_k) P(E_k) \\ &= \sum_{k=1}^{L} P(T_{\xi^k} \leq x) P(E_k) \end{aligned}$$

Since the limiting distribution of T_{ξ^k} under H_0 in (2) is χ_1^2 for each k, T_{BB} converges in distribution to a χ_1^2 random variable.

Proof of Theorem 4.2. Assume that C_α is the asymtotic critical value of T_{BB}.

$$\begin{aligned} P(T_{BB} \geq C_\alpha) &= \sum_{k=1}^{L} P(\sum_{i=1}^{L} T_{\xi^i} I(E_i) \geq C_\alpha | E_k) P(E_k) \\ &= \sum_{k=1}^{L} P(T_{\xi^k} \geq C_\alpha) P(E_k) \end{aligned}$$

Since T_{ξ^k} is consistent for all k, T_{BB} is also consistent under any alternative satisfying the assumptions of Theorem 6. □

REFERENCES

Chen, S. and Donoho, D. (1994). Automatic decomposition by basis pursuit. Technical Report Department of Statistics, Stanford University.

Coifman, R. and Majid, F. (1993). Adapted waveform analysis and denoising. In Y. Meyer and R. Roques (eds.), *Progress in Wavelet Analysis and Applications*, pp. 63 – 76. Gif-sur-Yvette Cedex, France: Editions Frontieres.

Donoho, D. and Johnstone, I. (1994). "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, **81**, 425-455.

Fan, J. (1996). "Test of significance based on wavelet thresholding and Neyman's truncation," *Journal of the American Statistical Association*, **91**, 674–688.

Hart, J. D. (1997). *Nonparametric Smoothing and Its Applications in Lack-of-Fit Testing,* Springer-Verlag, New York.

Hall, P., Kat., T. W. and Titterington, D. M. (1990). "Asymptotically optimal difference-based estimation of variance in nonparametric regression," *Biometrika*, **77**, 521-528.

Kuchibhatla, M. and Hart, J. D. (1996). "Smoothing-based lack-of-fit tests: Variations on a theme," *Journal of Nonparametric Statistics*, **7**, 1-22.

Ledwina, T. (1994). "Data driven version of Neyman's smooth test of fit," *Journal of the American Statistical Association*, **89**, 1000–1005.

Lee, G. H. (1997). A Statistical Wavelet Approach to Model Selection and Data Driven Neyman Smooth Tests, Ph.D. dissertation, Texas A&M University.

Lee, G. H. and Hart, J. D. (1998). "An L_2 Error Test with Order Selection and Thresholding," *Statistics & Probability Letters*, **39**, 61-72.

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

Spokoiny, V.G. (1996). "Adaptive hypothesis testing using wavelets," *Annals of Statistics*, **24**, 2477-2498.