Journal of Nonparametric Statistics
Vol. 00, No. 00, March 2008, 1–19

RESEARCH ARTICLE

An ANOVA-type Nonparametric Diagnostic Test for Heteroscedastic Regression Models

Lan Wang[a*], Michael G. Akritas[b]

and Ingrid Van Keilegom[c]

[a]*School of Statistics, University of Minnesota, USA*; [b]*Department of Statistics, Pennsylvania State University, USA*; [c]*Institut de statistique, Université catholique de Louvain, Belgium*

(*v3.2 XXXXXXXX*)

For the heteroscedastic nonparametric regression model $Y_{ni} = m(x_{ni})+\sigma(x_{ni})\epsilon_{ni}$, $i = 1, ..., n$, we discuss a novel method for testing some parametric assumptions about the regression function m. The test is motivated by recent developments in the asymptotic theory for analysis of variance when the number of factor levels is large. Asymptotic normality of the test statistic is established under the null hypothesis and suitable local alternatives. The similarity of the form of the test statistic to that of the classical F-statistic in analysis of variance allows easy and fast calculation. Simulation studies demonstrate that the new test possesses satisfactory finite-sample properties.

Keywords: heteroscedastic errors; lack-of-fit tests; local alternatives, nearest neighborhood; nonparametric regression.

AMS Subject Classification: 62G08; 62G10; 62J20.

1. Introduction

Regression analysis, the cornerstone of applied statistics, often assumes that the regression function has a certain parametric form. While this is done for mathematical convenience and ease of interpretation, it is well known that model misspecification can have detrimental effects on the validity of subsequent inferences. For instance, it may result in inconsistent parameter estimates. For this reason, diagnosing the adequacy of the postulated regression function forms an important research area.

We consider the following heteroscedastic nonparametric regression model:

$$Y_{ni} = m(x_{ni}) + \sigma(x_{ni})\epsilon_{ni}, \ i = 1, ..., n, \tag{1}$$

where $m(\cdot)$ is an unknown regression function, $\sigma^2(\cdot)$ is an unknown variance function. The design points reside in a bounded interval. Without loss of generality, we may assume $x_{n1}, \ldots, x_{nn}$ form a regular sequence on the interval [0,1] in the sense

*Corresponding author. Email: lan@stat.umn.edu

ISSN: 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/1048525YYxxxxxxxx
http://www.informaworld.com

of Sacks and Ylvisaker (1970), i.e., there exists a positive continuous density $r(x)$ on [0,1] such that

$$\int_0^{x_{ni}} r(x)dx = \frac{i}{n}, \quad i = 1, ..., n. \tag{2}$$

The random errors ϵ_{ni} in (1) constitute a triangular array of row-wise independent variables with mean 0 and variance 1. Below we will omit the n in the subscript when no confusion is caused.

This paper proposes a novel method for testing the adequacy of the parametric regression function. To illustrate the central idea, we focus on testing the hypothesis of a constant regression function (or no effect hypothesis):

$$H_0 : m(x) = C \text{ for all } x, \tag{3}$$

for some unknown constant $C \in I\!R$. This hypothesis has been considered by quite a few authors, for example, Von Neumann (1941), Raz (1990), Barry and Hartigan (1990), Eubank and Hart (1992), Eubank (2000). To test for the null hypothesis of a more general parametric functional form, such as a linear regression, we may apply the same approach to residuals from the parametric fit under the null model. For more general procedures, we refer to Azzalini and Bowman (1993), González Manteiga and Cao (1993), Härdle and Mammen (1993), Fan and Li (1996), Stute (1997), the monograph of Hart (1997), Dette and Munk (1998), Aït-Sahalia, Bickel and Stoker (2001), Fan and Huang (2001), Fan, Zhang and Zhang (2001), Horowitz and Spokoiny (2001) and Munk (2002) and the references therein.

The form of the new test is motivated by recent developments in analysis of variance with large number of factor levels (Akritas and Papadatos, 2004). However, their methods can not be applied to derive the large sample distribution of the proposed test and alternative technique has to be developed. Section 2 introduces the test statistic. Section 3 discusses its asymptotic properties under the null hypothesis and suitable local alternative sequence. For the validity of the large sample results, it is required that $m(\cdot)$, $\sigma^2(\cdot)$ and $r(\cdot)$ are Lipschitz continuous. Our test statistic is asymptotically unbiased under the null hypothesis, which enables more accurate normal approximation, while many other smoothing-based tests normally have one or more bias terms and require a computing intensive procedure such as bootstrap for their applications. Simulations reported in Section 4 indicate that the new test procedure, which uses the proposed test statistic with critical points determined from its asymptotic distribution, has satisfactory finite sample properties. In fact, a key merit of the new approach is its practicality. The similarity of the form of the test statistic to that of the classical F-statistic in analysis of variance allows one to take advantage of any existing statistical software for easy and fast calculation, which makes the test especially suitable for diagnostic purpose in exploratory data analysis.

2. Test Statistic

To describe the test statistic, we first diverge briefly to introduce some recent developments in analysis of variance with large number of factor levels. Consider a balanced one-way ANOVA setting with n groups (treatment levels) and k observations per group. Let $V_{i1}, \ldots, V_{ik}$ denote the k independent observations of group

i and $N = nk$ be the total sample size. Define

$$F_n = \frac{MST}{MSE}$$

where

$$MST = \frac{k}{n-1}\sum_{i=1}^{n}(\overline{V}_{i\cdot} - \overline{V}_{\cdot\cdot})^2, \quad MSE = \frac{1}{N-n}\sum_{i=1}^{n}\sum_{j=1}^{k}(V_{ij} - \overline{V}_{i\cdot})^2. \qquad (4)$$

Then F_n is simply the classical test statistic for testing for no treatment effects. A number of authors (Boos and Brownie, 1995; Akritas and Arnold, 2000; Akritas and Papadatos, 2004) recently studied the asymptotic distribution of $\sqrt{n}(F_n - 1)$ when n converges to infinity. Note that this is different from the classical asymptotic framework where n is fixed and the group sizes tend to infinity. Since MSE tends to a constant as n tends to infinity, by Slutsky's theorem the problem reduces to studying the asymptotic distribution of $\sqrt{n}(MST - MSE)$. The basic requirement in these asymptotic results is that there are at least two observations per cell.

The idea for constructing our test statistic for lack-of-fit test in the present regression setting is to consider each distinct covariate value as a 'factor level', as proposed in Akritas (2000). Of course, we cannot use the asymptotic theory developed for the ANOVA case because typically there is only one response value associated with each covariate value. This can be remedied by considering a window W_i around each x_i consisting of the k_n nearest covariate values. That is, we augment the observed data to construct an artificial balanced one-way ANOVA with n cells, where the responses in the i-th cell are the response values corresponding to the covariate values belonging to W_i. This artificial one-way ANOVA has overlapping groups (i.e. neighboring groups have common observations), and thus the asymptotic results for one-way ANOVA with independent observations and large number of factor levels still do not apply. An alternative theory, which accommodates dependent observations, has to be developed to derive the asymptotic distribution.

In what follows, the windows W_i will be understood as sets containing the indices j of the covariate values that belong to the window of size k_n, that is

$$W_i = \left\{ j : \; |\widehat{F}_X(x_j) - \widehat{F}_X(x_i)| \le \frac{k_n - 1}{2n} \right\}, \qquad (5)$$

where $\widehat{F}_X$ is the empirical distribution function of the covariate values. Now $V_{ij}, j = 1, \ldots, k_n$, in the definition of MST and MSE in (4) will be replaced by $Y_j, \; j \in W_i$, the observations in the artificial one-way ANOVA. Under the null hypothesis (3), we expect approximately no cell effects, an intuitive test statistic for this null hypothesis therefore is $MST - MSE$. In practice, for those points at the two edges, symmetric local windows are impossible. One may use asymmetric windows for those points, which can be shown to have asymptotic negligible effects for the asymptotic distribution of the test statistic. In simulations, the approach to start the first window around the $\frac{k_n+1}{2}$th smallest covariate and the last window around the $\frac{k_n+1}{2}$th largest covariate is demonstrated to work satisfactorily.

In the next section, the asymptotic normality of the proposed test statistic is derived under two different asymptotic frameworks : (1) $n \to \infty$ and $k_n \to \infty$ at an appropriate rate; (2) $n \to \infty$ and k_n is fixed. Our test therefore does not require a consistent or direct estimator of the regression function. The next section also discusses a comparison with the simpler statistic which uses non-overlapping groups, which essentially discretizes the covariate and applies the one-way ANOVA statistic for testing equality of group means.

Our test statistic is different from that of Azzalini and Bowman (1993), which also has the form of a F-test statistic but requires a direct consistent estimator of the regression function. Under the assumption of homoscedasticity, the distribution of their test statistic is approximated by a shifted and scaled χ^2 distribution by fitting the first three moments. The test statistic of Dette and Munk (1998) can also be considered as an F-test type statistic in the broad sense. Their test statistic is the difference of two estimators of the integrated variance function, one under the parametric model and the other under nonparametric alternative. However, the computation of their test can not directly take advantage of the available ANOVA procedure in existing statistical software.

3. Main Results

The vector of all the observations in the artificial one-way ANOVA will be denoted by

$$\boldsymbol{V} = (Y_j, j \in W_1, \ldots, Y_j, j \in W_n)'. \tag{6}$$

Our test statistic $MST - MSE$ can be expressed as a quadratic form $\boldsymbol{V}'\boldsymbol{A}\boldsymbol{V}$, where

$$\boldsymbol{A} = \frac{nk_n - 1}{n(n-1)k_n(k_n-1)} \bigoplus_{i=1}^{n} \boldsymbol{J}_{k_n} - \frac{1}{n(n-1)k_n}\boldsymbol{J}_{nk_n} - \frac{1}{n(k_n-1)}\boldsymbol{I}_{nk_n}, \tag{7}$$

where $\boldsymbol{J}_d$ is a $d \times d$ matrix with all elements equal to 1, $\boldsymbol{I}_d$ is the d-dimensional identity matrix, $\bigoplus_{i=1}^{n}$ denotes the Kronecker sum or direct sum (§8.3, Schott, 2005).

We start with the result about the asymptotic equivalence, under H_0 in (3), of the quadratic form $(n/k_n)^{1/2}\boldsymbol{V}'\boldsymbol{A}\boldsymbol{V}$ with another quadratic form involving a block diagonal matrix.

Lemma 3.1: *Assume $m(x)$, $\sigma(x)$ and $r(x)$ are Lipschitz continuous, $n^{-1}k_n \to 0$, and that $E(\epsilon_i^4)$ are uniformly bounded in n and i. Then, under H_0 in (3) as $n \to \infty$,*

$$\left(\frac{n}{k_n}\right)^{1/2} [\boldsymbol{V}'\boldsymbol{A}\boldsymbol{V} - (\boldsymbol{V} - C\mathbf{1}_N)'\boldsymbol{A}_d(\boldsymbol{V} - C\mathbf{1}_N)] \xrightarrow{P} 0,$$

where $\boldsymbol{A}_d$ is the block diagonal matrix

$$\boldsymbol{A}_d = diag\{\boldsymbol{B}_1, \ldots, \boldsymbol{B}_n\}, \text{ with } \boldsymbol{B}_i = \frac{1}{n(k_n-1)}[\boldsymbol{J}_{k_n} - \boldsymbol{I}_{k_n}].$$

Note that since $\boldsymbol{A}$ is a contrast matrix, $\boldsymbol{V}'\boldsymbol{A}\boldsymbol{V} = (\boldsymbol{V} - C\mathbf{1}_N)'\boldsymbol{A}(\boldsymbol{V} - C\mathbf{1}_N)$. This equality does not hold true however for $\boldsymbol{A}_d$. This result enables us to show the asymptotic normality of the test statistic.

Theorem 3.2: *Under the assumptions of Lemma 3.1 and H_0 defined in (3)*
(1) If $k_n = k$ is fixed, then as $n \to \infty$,

$$n^{1/2}(MST - MSE) \to N\left(0, \frac{2k(2k-1)}{3(k-1)}\tau^2\right),$$

where $\tau^2 = \int_0^1 \sigma^4(x)r(x)dx$.
(2) If $n \to \infty$ *and* $k_n \to \infty$ *such that* $k_n n^{-1} \to 0$, *then with* τ^2 *defined above,*

$$\left(\frac{n}{k_n}\right)^{1/2} (MST - MSE) \to N\left(0, \frac{4}{3}\tau^2\right),$$

A simple estimator of τ^2 is a modification of Rice's (1984) estimator, see also Dette and Munk (1998),

$$\hat{\tau}^2 = \frac{1}{4(n-3)} \sum_{j=2}^{n-2} R_j^2 R_{j+2}^2, \tag{8}$$

where $R_j = Y_j - Y_{j-1}$ denote the local residuals, $j = 2, ..., n$.

The asymptotic power of a test of (3) is often investigated by considering the probability the test rejects (3) when the alternative approaches to the null at a certain rate. The form of the local alternative sequence is commonly assumed to be:

$$m(x) = C + \rho_n g(x) \quad \text{for all } x, \tag{9}$$

where ρ_n is a sequence of constants converging to zero, $g(x)$ is some function. If ρ_n converges to zero too fast, the test will not be able to distinguish the local alternative from the null; on the other hand, if ρ_n converges to zero too slow, the test will always reject the null hypothesis for large enough sample size. Many of the nonparametric tests in the literature have nontrivial power only when $\rho_n \to 0$ at a rate slower than $n^{-1/2}$. A careful examination of our test reveals this rate to be $(nk_n)^{-1/4}$. The next theorem gives the asymptotic normal distributions of the test under the local alternative sequence.

Theorem 3.3 : *Assume the conditions of Theorem* 3.2 *are satisfied. Let* $g(x)$ *be a Lipschitz continuous function on [0,1], and consider the local alternative sequence* $H_1 : m(x) = C + (nk_n)^{-1/4}g(x)$.
(1) If $k_n = k$ *is fixed, then under* H_1,

$$\left(\frac{n}{k}\right)^{1/2} (MST - MSE) \to N\left(\gamma^2, \frac{2(2k-1)}{3(k-1)}\tau^2\right),$$

where τ^2 *is defined as in Theorem* 3.2 *and* $\gamma^2 = \int_0^1 g^2(t)r(t)dt - (\int_0^1 g(t)r(t)dt)^2$. *Thus, the efficacy of the test statistic is given by*

$$\left(\frac{3(k-1)}{2(2k-1)}\right)^{1/2} \frac{\gamma^2}{\tau}.$$

(2) If $k_n \to \infty$ *such that* $n^{-3}k_n^5 = o(1)$, *then under* H_1,

$$\left(\frac{n}{k_n}\right)^{1/2} (MST - MSE) \to N\left(\gamma^2, \frac{4}{3}\tau^2\right),$$

where γ^2 *and* τ^2 *are defined as above. The efficacy of the test statistic is given by*

$$\frac{\sqrt{3}\gamma^2}{2\tau}.$$

Remark 1 : The efficacy formulas show that as k_n increases so does the efficacy. For example, the fixed k_n values of 2, 5 and 10 give efficacies of 0.707, 0.816 and 0.843, respectively. If $k_n \to \infty$, the efficacy is 0.866. Thus, when the convergence to normality is slower (which happens when $k_n \to \infty$) the efficacy is the largest. On the other hand, the rate at which $k_n \to \infty$ does not influence the efficacy, even though it does influence the rate of convergence to normality.

Remark 2 : As mentioned in the introduction, the need for considering windows around each covariate value arose from the requirement of the asymptotic theory of ANOVA with large number of factor levels to have more than one observation per group. An alternative way of satisfying this requirement is to discretize the covariate. This results in non-overlapping groups, which allows for direct application of the asymptotic theory for the ANOVA case (Akritas and Papadatos, 2004). Thus, if we consider non-overlapping groups of size k_n, and consider for simplicity the case that $k_n \to \infty$, then under the same sequence of local alternatives as in Theorem 2.3,

$$\left(\frac{n}{k_n}\right)^{1/2}(MST - MSE) \to N\left(\gamma^2, 2\tau^2\right),$$

so its efficacy is $\gamma^2/(\sqrt{2}\tau)$, implying it is about 82% as efficient as the statistic based on overlapping windows.

Remark 3 : Several tests allow $\rho_n \propto n^{-1/2}$, however these tests cannot have non-trivial power uniformly over reasonable classes of functions $g(x)$, see Horowitz and Spokoiny (2001) for more discussions. Recently, some adaptive rate optimal tests have been investigated, including Fan and Huang (2001), Horowitz and Spokoiny (2001), among others. These new tests have the property of adapting to the unknown smoothness of the alternative model and are uniformly consistent against alternatives whose distance from the null model converges to zero at the fastest possible rate. To perform the tests, computationally intensive bootstrap methods have to be used to obtain the critical values. On the other hand, the asymptotic properties tell little about what happens with moderately large sample size. In the next section, some comparison of the finite sample behaviors of our test and one such optimal test is provided. And our test is demonstrated to be quite competitive.

The goodness-of-fit test for no effect hypothesis described above can be generalized to testing whether $m(x)$ belongs to a specified parametric family. The alternative is only required to be smooth and its form needs not to be specified. More specifically, let $S_\Theta = \{f(\cdot, \theta), \theta \in \Theta\}$ be any parametric family of functions and we wish to test:

$$H_0: \quad m(x) \in S_\Theta,$$

against a general alternative. Let $\hat{\theta}$ be our "best estimator" of the true parameter θ under the null hypothesis. Under the null hypothesis, such estimators are often $\sqrt{n}$-consistent. Then we may apply the above test to the residuals: $\widehat{e}_i = Y_i - m(X_i, \widehat{\theta})$, $i = 1, ..., n$. The $\hat{e}_i$'s are not independent in general, but the dependence can be shown to be negligible in the asymptotic sense. As an example, we consider testing the null hypothesis that $m(x)$ is a simple linear function:

$$H_0: m(x) = a + bx, \tag{10}$$

for some unknown parameters a, b. Letting $\widehat{a}$ and $\widehat{b}$ denote the usual least squared estimators, and applying the above test statistic to the residuals $\widehat{e}_i = Y_i - \widehat{a} - \widehat{b}x_i$, then the asymptotic results given in Theorem 3.2 still hold. More specifically,

Theorem 3.4: *Assume the assumptions of Lemma 3.1, let MST and MSE be calculated from the hypothetical one-way ANOVA constructed by augmenting $(x_i, \widehat{e}_i)$, $i = 1, \ldots, n$. Then, under H_0 defined in (10)*
(1) If $k_n = k$ is fixed, then as $n \to \infty$,

$$n^{1/2}(MST - MSE) \to N\left(0, \frac{2k(2k-1)}{3(k-1)}\tau^2\right),$$

(2) If $n \to \infty$ and $k_n \to \infty$ such that $k_n^{3/2}n^{-1} \to 0$, then

$$\left(\frac{n}{k_n}\right)^{1/2}(MST - MSE) \to N\left(0, \frac{4}{3}\tau^2\right),$$

where τ^2 is defined as in Theorem 3.2.

4. Simulations

In this section, we investigate the finite-sample behaviors of our test under the null hypothesis and different alternative hypotheses. The test statistic is calculated using the asymptotic normality results in Theorems 3.2 and 3.4 with the asymptotic variance estimated by (8). Our simulations are based on 5,000 runs at nominal level 0.05. Random numbers are generated using Matlab.

4.1. *Level of the test*

We generate random data from a regression function that is constant zero and equally-spaced design points on the interval (0,1]: $x_i = i/n$, $i = 1, ..., n$, where $n = 60$ and 100.

We first consider three different error distributions: the standard normal, the $sinh^{-1}$-normal and the lognormal. The $sinh^{-1}$-normal distribution is symmetric with moderately large kurtosis. The lognormal distribution is highly skewed with very large kurtosis. For comparison purpose, we standardize the error distributions so that they all have mean 0 and variance 1. In Table 1, the observed proportion of rejections are reported for different local window sizes k_n. The observed levels are close to the specified nominal levels under normal distributions. Under heavier-tail errors, the observed level is a little conservative.

To investigate the effects of heteroscedastic errors on the level, we consider the following three different functional forms for $\sigma(x)$: (1) $\sigma(x) = exp(0.5x)$, (2) $\sigma(x) = 0.5 + x$, (3) $\sigma(x) = 1 + sin(x)$. The ϵ_i's in (1) are taken to be standard normal random variables. The observed proportions of rejection are reported in Table 2, which confirms that our test maintains the type I error fairly accurately.

4.2. *Power of the test*

Example 4.1. Testing for a constant regression function. In Table 3, we compare our test with a recent procedure proposed by Munk (2002). Munk's test is based on a quadratic measure of the discrepancy between the postulated parametric model

Table 1. Proportion of rejection under H_0 for homoscedastic errors.

	ε	N(0,1)	Lognormal	Sinh^{-1}-normal
n	k_n	level	level	level
60	7	0.048	0.037	0.041
	9	0.040	0.033	0.035
	11	0.039	0.027	0.037
100	9	0.054	0.037	0.046
	11	0.047	0.034	0.038
	13	0.042	0.029	0.044

Table 2. Proportion of rejection under H_0 for heteroscedastic errors.

	$\sigma(x)$	$exp(0.5x)$	$0.5 + x$	$1 + sin(x)$
n	k_n	level	level	level
60	7	0.050	0.061	0.057
	9	0.042	0.052	0.049
	11	0.034	0.043	0.043
100	9	0.053	0.055	0.053
	11	0.045	0.047	0.045
	13	0.041	0.041	0.042

and the true model, which is estimated by random Toeplitz forms. Within a general class of Toeplitz-matrices, the asymptotic efficiency of the proposed test can be maximized and lead to an optimal test in this class. Munk derived the asymptotic normality for testing general linear and nonlinear parametric assumptions. When the null hypothesis is a constant regression model, his test statistic reduces to:

$$\sqrt{n}\left[Y'D_{\alpha^*,r,n}Y - \left(\sum_{i=1}^{n} Y_i/n\right)^2\right],$$

where $D_{\alpha^*,r,n}$ is the Toeplitz band matrix with optimal weights. In Table 3, the new test (denoted by ANOVA test) is calculated for different window size k_n; Munk's test (denoted by Toeplitz test) is calculated for different band size r without any correction. Jackknife correction has been suggested by Munk, which leads to a potentially more accurate and powerful procedure. Due to the fact it is more computing intensive, it is not adopted here. The sample size n is 60. The random errors are iid standard normal. Four different regression functions are considered:

$$m(x) = 0, \tag{11}$$

$$m(x) = 2x, \tag{12}$$

$$m(x) = 64x^3(1-x)^3, \tag{13}$$

$$m(x) = (1 + sin(3\pi x))/2. \tag{14}$$

The two tests display quite competitive performance. Munk's test is more powerful for the alternative (13), while the new test seems to be more power for the alternatives (12) and (14).

Table 3. Proportion of rejection for different regression functions in Example 4.1.

m(x)		(11)	(12)	(13)	(14)
ANOVA test	kn=7	0.05	0.87	0.45	0.46
	kn=9	0.04	0.87	0.45	0.47
	kn=11	0.04	0.86	0.43	0.46
Toeplitz test	r=2	0.07	0.78	0.40	0.32
	r=4	0.05	0.83	0.52	0.36
	r=6	0.04	0.82	0.56	0.35
	r=8	0.03	0.80	0.62	0.31

Table 4. Proportion of rejection for the regression model in Example 4.2.

model	k_n	normal	mixture	extreme
null	7	0.049	0.043	0.045
	9	0.047	0.040	0.039
	11	0.040	0.040	0.038
$\tau = 1.0$	7	0.880	0.843	0.896
	9	0.914	0.875	0.922
	11	0.938	0.896	0.944
$\tau = 0.25$	7	1.000	1.000	1.000
	9	1.000	1.000	1.000
	11	1.000	1.000	1.000

Example 4.2. Testing for a simple linear regression function. In this example, we take the simulation setting of Horowitz and Spokoiny (2001). Their test is based on the distance of a kernel nonparametric estimator of the regression function and a kernel-smoothed parametric estimator. The distance is computed for a range of different values of the smoothing parameter. If the distance obtained with any one of the bandwidths is too large, then the null hypothesis is rejected. Their test has the advantage of being uniformly consistent against alternatives whose distance from the null converges to zero at the fastest possible rate.

The null model is $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where β_0 and β_1 are constants. The covariates x_i's are sampled from the $N(0, 25)$ distribution truncated at its 5th and 95th percentiles. Under the null hypothesis, $\beta_0 = \beta_1 = 1$. Three different distributions for ϵ_i's are considered: $N(0, 4)$; a mixture of normals in which ϵ_i is sampled from $N(0, 1.56)$ with probability 0.9 and from $N(0, 25)$ with probability 0.1; and the Type I extreme value distribution scaled to have variance 4. The mixture distribution is leptokurtic with variance 3.9, and the type I extreme value distribution is asymmetrical. The alternative model is:

$$Y_i = 1 + x_i + \frac{5}{\tau}\phi\left(\frac{x_i}{\tau}\right) + \epsilon_i, \tag{15}$$

where $\tau = 1$ or 0.25, $\phi(\cdot)$ is the density function of standard normal distribution. The sample size is 250. The simulation results are presented in Table 4. We observe that the power of our test is a little higher than that presented in Table 1 of Horowitz and Spokoiny (2001).

Example 4.3. We take the simulation setting of Dette and Munk (1998), where the design points are $x_i = (i - 1/2)/n$, $i = 1, \ldots, 100$, the alternative is

$$m(x) = 1 + \theta cos(10\pi x), \quad (16)$$

with $\theta = 0.00, 0.25, 0.50, 0.75, 1.00$, and the random errors are iid standard normal. Table 5 summarizes the estimated power of the proposed test for testing a constant null hypothesis at nominal levels $\alpha = 0.05$ and 0.10. Compared with the results in Table 2 of Dette and Munk (1998), the power of our test is significantly higher.

Table 5. Proportion of rejection for the regression model in Example 4.3.

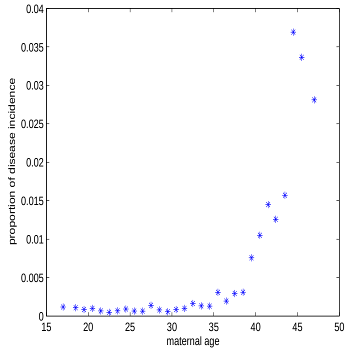
	θ	0.00	0.25	0.50	0.75	1.00
α	kn	level	power	power	power	power
	5	0.063	0.137	0.456	0.876	0.994
0.05	7	0.055	0.106	0.391	0.812	0.991
	9	0.045	0.086	0.266	0.685	0.948
	5	0.105	0.194	0.558	0.918	0.997
0.10	7	0.085	0.165	0.491	0.887	0.994
	9	0.079	0.134	0.374	0.787	0.973

The above simulations suggest that the performance of the new test is not very sensitive to the choice of the smoothing parameter k_n. How to choose a smoothing parameter to achieve the optimal power performance is still an open problem faced by all smoothing-based lack-of-fit tests; and it is very different from the problem of choosing a smoothing parameter for optimal curve fitting. The test of Horowitz and Spokoiny (2001) circumvents this problem by considering the maximum of the standardized smoothing test over a sequence of smoothing parameters.

5. Real Data Examples

Example 5.1. In the 1960's, a large scale study of the effect of maternal age on the incidence of Down's syndrome, a genetic disorder caused by an extra chromosome 21 or a part of chromosome 21 being translocated to another chromosome, was conducted at the British Columbia Health Surveillance Registry (Geyer, 1991). There are 30 different age groups, the proportion of Down's syndrome among the babies born to mothers of that age group was calculated. Of interest is to test if the incidence of Down's syndrome is affected by maternal age. Another hypothesis of interest is that the incidence of Down's syndrome only starts rising after age 30. Using age as covariate and proportion of the disease as response variable, we performed the classical F test (denoted by CF, corresponding to fit a linear regression model and test if the slope is zero) and our nonparametric test (denoted by NP). For the overall data, the p-value of the CF test is 0; the p-values of the NP test are all 0 for $k_n = 3, 5, 7$. Both the parametric and nonparametric approaches suggest that the incidence of Down's syndrome is influenced by maternal age. For the data corresponding to the first 14 age groups (up to age 30.5), the p-value of the CF test is 0.379; the p-value of the NP test is 0.908 for $k_n = 3$, 0.768 for $k_n = 5$, 0.828 for $k_n = 7$. Both methods agree that the influence of maternal age only becomes obvious after age 30.

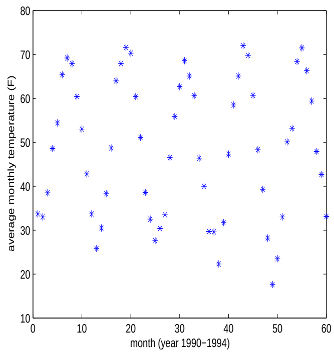
Figure 1. Scatter Plot of Proportion of Disease Incidence vs Maternal Age



Put Figure 1 about here

Example 5.2. Consider the monthly average temperature data of central Pennsylvania between 1990 and 1994 (http://climate.met.psu.edu/data/state.php). See Figure 2 for a scatter plot of the 60 data points.

Figure 2. Scatter Plot of Central PA Monthly Average Temperature



To test for the effect of time on the temperature, the classical F test gives a p-value 0.914, while the NP test gives p-values 0 for $k_n = 3, 5, 7$. The classical F test therefore has little power to detect the oscillating influence of the covariate in this case.

6. Discussion

We have presented a novel procedure for testing the hypothesis of a constant regression function which does not require consistent estimation of the regression function. Compared to nonparametric smoother based tests, our test imposes minimal smoothness assumptions, it is asymptotically unbiased under the null and has good finite sample performance. It is closely related to the classical lack-of-fit statistic used in the case of replicated observations, and to the one-way ANOVA test statistic. In fact, the only difference lies in the fact that the replications or groups are artificially generated from observations with neighboring covariate values. Thus, our statistic can be very easily computed with any software package.

The statistic belongs to the category of test statistics based on difference or ratio of two estimators of the integrated variance function $\int \sigma^2(t)r(t)dt$. One is asymptotically unbiased under the null hypothesis only, and the other is so under both the null and alternative hypotheses. Indeed, MSE is asymptotically unbiased under general nonparametric alternatives while MST is asymptotically unbiased under the null model only and asymptotically positively biased under the alternative.

A natural application of the results presented here is to test the parallelism of two curves when two data sets are observed at the same design points. Assume we observe triple variables: (x_i, Y_i, Z_i), $i = 1, ..., n$. If

$$Y_i = m_1(x_i) + \sigma_1(x_i)\epsilon_{1i},$$
$$Z_i = m_2(x_i) + \sigma_2(x_i)\epsilon_{2i},$$

then to test the hypothesis that $m_1(\cdot)$, $m_2(\cdot)$ are parallel we may apply our test to $Y_i - Z_i$, $i = 1, ..., n$. This type of problem arises frequently in practice; see Hart (1997), §9.6.

Generalization of the proposed test to random design setting is straightforward. The extension to more complicated hypotheses and to more covariates is being investigated.

References

[1] Aït-Sahalia, Bickel, P. J., and Stoker, T. M. (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, **105**, 363-412.

[2] Akritas, M. G. (2000). *Nonparametric Regression and Lack-of-fit Tests.* Unpublished lecture notes.

[3] Akritas, M. G., Arnold, S. (2000). Asymptotics for Analysis of Variance when the number of levels is large. *Journal of the American Statistical Association*, **95**, 212-226.

[4] Akritas, M. G. and Papadatos, N. (2004). Heteroscedastic one-way ANOVA and lack-of-fit test. *Journal of the American Statistical Association*, **99**, 368-382.

[5] Azzalini, A. and Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *Journal of the Royal Statistical Society*, Series B, **55**, 549-559.

[6] Barry, Daniel. and Hartigan, J.A. (1990). An omnibus test for departures from constant mean. *Annals of Statistics*, **18,** 1340–1357.

[7] Billingsley, P. (1986). *Probability and Measures* (New York:Wiley).

[8] Boos, D.D. and Brownie, C. (1995). ANOVA and rank tests when the number of treatments is large. *Statistics & Probability Letters*, **23**, 183–191.

[9] Dette, H. and Munk, A. (1998). Validation of linear regression models. *Annals of Statistics*, **26,** 778–800.

[10] Dette, H. and Munk, A. (1998). Testing heteroscedasticity in nonparametric regression. *Journal of the Royal Statistical Society*, Series B, **60,** 693–708.

[11] Eubank, R.L. Hart, J.D. (1992). Testing goodness-of-fit in regression via oder selection criteria. *Annals of Statistics*, **20,** 1412–1425.

[12] Eubank, R. L. (2000). Testing for no effect by cosine series methods. *Scandinavian Journal of Statistics*, **27,** 747–763.

[13] Fan, J. Q. and Huang, L.S. (2001). Goodness-of-Fit tests for parametric regression models. *Journal of the American Statistical Association* **96,** 640–652.

[14] Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms *Econometrica*, **64,** 865-890.

[15] Fan, J., Zhang, C.M., and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon, *Annals of Statistics*, **29,** 153-193.

[16] Geyer, C. J. (1991). Constrained maximum likelihood exemplified by isotonic convex logistic regression. *Journal of the American Statistical Association*, **86**, 717-724.

[17] González Manteiga, W. and Cao, R. (1993). Testing the hypothesis of a general linear model using nonparametric regression estimation. *Test*, **2**, 161-188.

[18] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, **21,** 1926–1947.

[19] Hart, J. (1997). *Nonparametric Smoothing and Lack-of-fit Test* (New York: Springer-Verlag.)

[20] Horowitz, J.L. and Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69,** 599–631.

[21] Munk, A. (2002). Testing the goodness of fit of parametric regression models with random Toeplitz forms. *Scandinavian Journal of Statistics*, **29(3),** 501–535.

[22] Raz, J. (1990). Testing for no effect when estimating a smooth function by nonparametric regression: a randomization approach. *Journal of American Statistical Association*, **85,** 132–138.

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

[24] Sacks, J. and Ylvisaker,D (1970). Designs for regression problems with correlated errors III. *Annals of Mathematical Statistics*, **41,** 2057–2074.

[25] Schott, J. R. (2005). *Matrix Analysis for Statistics*, 2nd ed, (New York: Wiley).

[26] Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, **25,** 613–641.

[27] von Neumann, J. (1941). Distribution of the ratio of mean squared successive difference to variance. *Annals of Mathematical Statistics*, **12,** 367–395.

Appendix: Proofs of the Theorems

In all proofs that follow, we will take k_n be odd for simplicity.

Proof of Lemma 3.1. It is easy to see that the block diagonal elements of $\boldsymbol{A}_d$ equal those of $\boldsymbol{A}$. Hence, it suffices to prove that the off diagonal blocks of $\boldsymbol{A}$ are negligible. For $i_1 \neq i_2$, every element of the block (i_1, i_2) equals $c = -\frac{1}{n(n-1)k_n}$. Hence, letting $W_i = \{j : w_{ij} \neq 0\}$, we have

$$\left(\frac{n}{k_n}\right)^{1/2} (\boldsymbol{V} - C\mathbf{1}_N)'(\boldsymbol{A} - \boldsymbol{A}_d)(\boldsymbol{V} - C\mathbf{1}_N)$$
$$= \left(\frac{n}{k_n}\right)^{1/2} c \sum_{i_1 \neq i_2} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} (Y_{j_1} - C)(Y_{j_2} - C) I(j_1 \in W_{i_1}) I(j_2 \in W_{i_2}).$$

It is easy to show that $E\left[\left(\frac{n}{k_n}\right)^{1/2} (\boldsymbol{V} - C\mathbf{1}_N)'(\boldsymbol{A} - \boldsymbol{A}_d)(\boldsymbol{V} - C\mathbf{1}_N)\right]$ is $o(1)$, while

$$\frac{n}{k_n} E[(\boldsymbol{V} - C\mathbf{1}_N)'(\boldsymbol{A} - \boldsymbol{A}_d)(\boldsymbol{V} - C\mathbf{1}_N)]^2$$
$$= \frac{n}{k_n} c^2 \sum_{i_1 \neq i_2} \sum_{i_3 \neq i_4} \sum_{j_1, \ldots, j_4 = 1}^{n} E[(Y_{j_1} - C)(Y_{j_2} - C)(Y_{j_3} - C)(Y_{j_4} - C)$$
$$\times I(j_k \in W_{i_k}, k = 1, \ldots, 4)]. \quad (17)$$

The expected value in this sum is different from zero, only if $Y_{j_1}, \ldots, Y_{j_4}$ consists of two pairs of equal observations, or $j_1 = j_2 = j_3 = j_4$. Since there are $O(n^2 k_n^4)$ terms for the former case to happen and $O(n k_n^4)$ for the latter case to happen, the order of (17) is

$$O\left(\frac{n}{k_n} \frac{1}{n^4 k_n^2} n^2 k_n^4\right) = O(k_n n^{-1}) = o(1),$$

and this finishes the proof.

Proof of Theorem 3.2. We can assume without loss of generality that $C = 0$.

$$\boldsymbol{V}' \boldsymbol{A}_d \boldsymbol{V} = \frac{1}{n(k_n - 1)} \sum_{i=1}^{n} \sum_{j_1 \neq j_2}^{n} Y_{j_1} Y_{j_2} I(j_1, j_2 \in W_i),$$

it's obvious that $E(\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}) = 0$.

$$
\begin{aligned}
E[\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}]^2 &= \frac{1}{n^2(k_n-1)^2}\sum_{i_1,i_2}^{n}\sum_{j_1\neq\ell_1}^{n}\sum_{j_2\neq\ell_2}^{n} E[Y_{j_1}Y_{\ell_1}Y_{j_2}Y_{\ell_2}]I(j_s\in W_{i_s},\ell_s\in W_{i_s}, s=1,2)\\
&= \frac{2}{n^2(k_n-1)^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j\neq\ell}^{n}\sigma^2(x_j)\sigma^2(x_\ell)I(j,\ell\in W_{i_1}\cap W_{i_2})\\
&= \frac{2}{n^2(k_n-1)^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j\neq\ell}^{n}\sigma^2(x_j)\left(\sigma^2(x_j)+O\left(\frac{k_n}{n}\right)\right)I(j,\ell\in W_{i_1}\cap W_{i_2})\\
&= \frac{2}{n^2(k_n-1)^2}\sum_{j=1}^{n}\sigma^4(x_j)\sum_{l\neq j}^{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} I(j,l\in W_{i_1}\cap W_{i_2}) + O\left(\frac{k_n^2}{n^2}\right)\\
&= \frac{2}{n^2(k_n-1)^2}\sum_{j=1}^{n}\sigma^4(x_j)2\left[1+2^2+3^2+\cdots+(k_n-1)^2\right]+O\left(\frac{k_n^2}{n^2}\right)\\
&= \frac{2}{n^2(k_n-1)^2}\frac{k_n(k_n-1)(2k_n-1)}{3}\sum_{j=1}^{n}\sigma^4(x_j)+O\left(\frac{k_n^2}{n^2}\right),
\end{aligned}
$$

where the third equality is a result of the assumptions that $\sigma^2(x)$ is Lipschitz continuous and the design density is r bounded away from zero, while the second last equality follows from the fact that if $1\leq |j_1-j_2| = s\leq k_n-1$, then they are in $(k_n-s)^2$ pairs of windows (including two identical windows) whose intersection includes both j_1 and j_2. Thus,

$$
\begin{aligned}
E\left(\sqrt{\frac{n}{k_n}}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}\right)^2 &= \frac{2(2k_n-1)}{3(k_n-1)}\frac{1}{n}\sum_{j=1}^{n}\sigma^4(x_j)+O\left(\frac{k_n}{n}\right)\\
&= \frac{2(2k_n-1)}{3(k_n-1)}\sum_{j=1}^{n}\sigma^4(x_j)\int_{x_{j-1}}^{x_j} r(t)dt + o(1)\\
&\to \begin{cases} \frac{4}{3}\tau^2, & \text{if } k_n\to\infty\\ \frac{2(2k_n-1)}{3(k_n-1)}\tau^2, & \text{if } k_n \text{ is fixed.}\end{cases}
\end{aligned}
$$

It remains to verify the asymptotic normality of $(n/k_n)^{1/2}(\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V})$. We will make use of Markov's blocking technique (see proof of Theorem 27.5 in Billingsley, 1986). Write

$$
\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V} = n^{-1}\sum_{i=1}^{n}A_i = n^{-1}S_n,
$$

where

$$
A_i = \frac{1}{k_n-1}\sum_{j_1\neq j_2}^{n} Y_{j_1}Y_{j_2}I(j_1,j_2\in W_i).
$$

We will establish the weak convergence of $(nk_n)^{-1/2}S_n$. First note that

$$E(S_n^4) = O\left(\frac{1}{k_n^4}\right) \sum_{i_1,i_2,i_3,i_4} \sum_{j_1 \neq l_1} \cdots \sum_{j_4 \neq l_4} E(Y_{j_1}Y_{l_1}...Y_{j_4}Y_{l_4}) I(j_k, l_k \in W_{i_k}, k = 1,2,3,4).$$

The nonzero terms in the sum must be one of the following forms: $E(Y_{j_1}^4 Y_{j_2}^4)$, which is of order $O(nk_n^5)$; $E(Y_{j_1}^4 Y_{j2}^2 Y_{j_3}^2)$ or $E(Y_{j_1}^3 Y_{j_2}^3 Y_{j_3}^2)$, which are both of order $O(nk_n^6)$; and $E(Y_{j_1}^2 Y_{j_2}^2 Y_{j_3}^2 Y_{j_4}^2)$, which is of order $O(n^2 k_n^6)$. Thus by the boundedness assumptions of the moments of error terms, $E(S_n^4) \leq K_1(n^2k_n^2)$, for some positive constant K_1. Next, define

$$U_{ni} = A_{(i-1)(b_n+l_n)+1} + \ldots + A_{(i-1)(b_n+l_n)+b_n}$$
$$V_{ni} = A_{(i-1)(b_n+l_n)+b_n+1} + \ldots + A_{i(b_n+l_n)},$$

$i = 1, \ldots, r_n$, where

$$b_n \sim n^{2/3} k_n^{1/3}, \quad l_n \sim k_n, \quad r_n \sim b_n^{-1} n = n^{1/3} k_n^{-1/3},$$

and assume for simplicity that n is a multiple of $b_n + l_n$. Then,

$$S_n = \sum_{i=1}^{r_n} U_{ni} + \sum_{i=1}^{r_n} V_{ni}.$$

The idea is to show that $\sum_{i=1}^{r_n} V_{ni} = o_P((nk_n)^{-1/2})$, such that by Slutsky's result, the asymptotic normality of S_n will follow from that of $\sum_{i=1}^{r_n} U_{ni}$. First consider

$$P\left((nk_n)^{-1/2}\left|\sum_{i=1}^{r_n} V_{ni}\right| \geq \varepsilon\right) \leq \sum_{i=1}^{r_n} P(|V_{ni}| \geq \varepsilon (nk_n)^{1/2} r_n^{-1})$$
$$\leq K\varepsilon^{-4}(nk_n)^{-2} r_n^5 (l_n k_n)^2 = O(n^{-2} r_n^5 k_n^2) = O(k_n^{1/3} n^{-1/3}) = o(1),$$

where the second inequality follows by Markov's inequality using the fact that $E(V_{ni}^4) \leq K(l_n k_n)^2$ (for some $K > 0$ and for all i), which follows in a similar way as for $E(S_n^4)$. Hence,

$$\sum_{i=1}^{r_n} V_{ni} = o_P((nk_n)^{-1/2}).$$

Since $U_{n1}, \ldots, U_{nr_n}$ are independent, the asymptotic normality of $\sum_{i=1}^{r_n} U_{ni}$ can be established by verifying Lyapounov's condition:

$$\sum_{i=1}^{r_n} \frac{E(U_{ni}^4)}{\left[\sum_{i=1}^{r_n} E(U_{nj}^2)\right]^2} \to 0, \text{ as } n \to \infty. \tag{18}$$

Since $E(S_n^2) = n^2 \mathrm{Var}(\boldsymbol{V}' \boldsymbol{A}_d \boldsymbol{V}) \sim nk_n$ and since $E(S_n^4) \sim (nk_n)^2$, it follows in a similar way that $E(U_{ni}^2) \sim b_n k_n$ and that $E(U_{ni}^4) \sim (b_n k_n)^2$. Hence, the order of

(18) is

$$O\left(r_n \frac{b_n^2 k_n^2}{r_n^2 b_n^2 k_n^2}\right) = O(r_n^{-1}) = o(1).$$

Hence, the asymptotic normality of $\sum_{i=1}^{r_n} U_{ni}$ follows, which finishes the proof.

Before stating the proof of Theorem 3.3, we prove a lemma which will be needed in the proof.

Lemma 6.1: *For any Lipschitz continuous function $g(x)$ on [0,1], we have*

$$k_n^{-1} \sum_{j=1}^{n} g(x_j) I(j \in W_i) - g(x_i) = O\left(\frac{k_n}{n}\right),$$

uniformly in $i = 1, \ldots, n$. In this lemma, $g(x_i)$ can be replaced with any $g(x_m)$ with $x_m \in W_i$.

Proof. Using the Lipschitz condition and the mean value theorem, the left side is less than or equal to

$$\begin{aligned} &\frac{1}{k_n} \sum_{j=1}^{n} |g(x_j) - g(x_i)| I\left[|\hat{F}_X(x_j) - \hat{F}_X(x_i)| \le \frac{k_n - 1}{2n}\right] \\ \le &\frac{M}{k_n \inf_x r(x)} \sum_{j=1}^{n} \frac{|j-i|}{n} I\left[|\hat{F}_X(x_j) - \hat{F}_X(x_i)| \le \frac{k_n - 1}{2n}\right] \\ \le &\frac{M_1}{nk_n}(2 + 4 + \cdots + 2(k_n - 1)) = O\left(\frac{k_n}{n}\right), \end{aligned}$$

where M and M_1 are some positive constants. This completes the proof.

Proof of Theorem 3.3. The proof is given for the case $k_n \to \infty$, the case k_n is fixed can be proved the same way. Let $\boldsymbol{V}$ be defined by (6) and set

$$\boldsymbol{g} = ((g(x_j), j \in W_1, \ldots, g(x_j), j \in W_n)'.$$

Thus, under $H_1 : m(x) = C + (nk_n)^{-1/4} g(x)$, we have $E(\boldsymbol{V}) = C\mathbf{1}_N + (nk_n)^{-1/4}\boldsymbol{g}$. Further denote $\boldsymbol{Z} = \boldsymbol{V} - C\mathbf{1}_N - (nk_n)^{-1/4}\boldsymbol{g}$ so that

$$\begin{aligned} MST - MSE &= \boldsymbol{V}'\boldsymbol{A}\boldsymbol{V} \\ &= \boldsymbol{Z}'\boldsymbol{A}\boldsymbol{Z} + 2(nk_n)^{-1/4}\boldsymbol{g}'\boldsymbol{A}\boldsymbol{Z} + (nk_n)^{-1/2}\boldsymbol{g}'\boldsymbol{A}\boldsymbol{g}. \qquad (19) \end{aligned}$$

We will show that

$$(nk_n)^{-1/4}\boldsymbol{g}'\boldsymbol{A}\boldsymbol{Z} = o_P(n^{-1/2}k_n^{1/2}) \qquad (20)$$

$$(nk_n)^{-1/2}\boldsymbol{g}'\boldsymbol{A}\boldsymbol{g} = n^{-1/2}k_n^{1/2}\gamma^2 + o_P(n^{-1/2}k_n^{1/2}) \qquad (21)$$

from which it will follow that (19) equals

$$\boldsymbol{Z}'\boldsymbol{A}\boldsymbol{Z} + n^{-1/2}k_n^{1/2}\gamma^2 + o_P(n^{-1/2}k_n^{1/2}).$$

Since $E(\boldsymbol{Z}) = \mathbf{0}_N$, $\boldsymbol{Z}$ satisfies the null hypothesis, hence it follows from Theorem 3.2 and Slutsky's theorem that the above converges in distribution to the stated normal distribution, concluding the proof of the theorem.

We start with verifying (20). Using (7), write

$$\begin{aligned} &\boldsymbol{g}'\boldsymbol{A}\boldsymbol{Z} \\ &= \frac{nk_n - 1}{n(n-1)k_n(k_n-1)} \sum_{i=1}^{n} \left[\sum_{j=1}^{n} g(x_j) I(j \in W_i) \right] \left[\sum_{k=1}^{n} Z_k I(k \in W_i) \right] \\ &\quad - \frac{1}{n(n-1)k_n} \left[k_n \sum_{i=1}^{n} g(x_i) \right] \left[k_n \sum_{i=1}^{n} Z_i \right] - \frac{k_n}{n(k_n-1)} \sum_{i=1}^{n} g(x_i) Z_i. \qquad (22) \end{aligned}$$

Using Lemma 6.1, the sum in the first term can be written as

$$\begin{aligned} & k_n \sum_{i=1}^{n} \left[g(x_i) + O(n^{-1}k_n) \right] \left[\sum_{k=1}^{n} Z_k I(k \in W_i) \right] \\ &\le k_n \sum_{k=1}^{n} \left[\sum_{i=1}^{n} g(x_i) I(i \in W_k) \right] Z_k + k_n^2 O(n^{-1}k_n) \sum_{k=1}^{n} |Z_k| \\ &= k_n^2 \sum_{k=1}^{n} g(x_k) Z_k + O_P(k_n^3), \end{aligned}$$

where the last equality is an application of central limit theorem for independent but not identically distributed random variables. Hence, (22) equals

$$\begin{aligned} & \frac{(nk_n-1)k_n}{n(n-1)(k_n-1)} \sum_{i=1}^{n} g(x_i) Z_i - \frac{k_n}{n(n-1)} \left[\sum_{i=1}^{n} g(x_i) \right] \left[\sum_{i=1}^{n} Z_i \right] \\ & - \frac{k_n}{n(k_n-1)} \sum_{i=1}^{n} g(x_i) Z_i + O_P(n^{-1}k_n^2) \\ &= \frac{nk_n}{n-1} \left\{ \left[n^{-1} \sum_{i=1}^{n} g(x_i) Z_i \right] - \left[n^{-1} \sum_{i=1}^{n} g(x_i) \right] \left[n^{-1} \sum_{i=1}^{n} Z_i \right] \right\} + O_P(n^{-1}k_n^2). \end{aligned}$$

Denote $D = [n^{-1} \sum_{i=1}^{n} g(x_i) Z_i] - [n^{-1} \sum_{i=1}^{n} g(x_i)][n^{-1} \sum_{i=1}^{n} Z_i]$, then it's easy to check that $E(D) = 0$, $Var(D) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2(x_i)(g(x_i) - n^{-1} \sum_{j=1}^{n} g(x_j))^2 = O(n^{-1})$, thus $D = O_p(n^{-1/2})$ and

$$\boldsymbol{g}'\boldsymbol{A}\boldsymbol{Z} = O_p(k_n n^{-1/2}) + O_P(n^{-1}k_n^2).$$

So (20) is satisfied provided $n^{-3}k_n^5 \to 0$. Similarly, we have

$$\boldsymbol{g}'\boldsymbol{A}\boldsymbol{g} = \frac{nk_n}{n-1} \left\{ \left[n^{-1} \sum_{i=1}^{n} g^2(x_i) \right] - \left[n^{-1} \sum_{i=1}^{n} g(x_i) \right]^2 \right\} + O(n^{-1}k_n^2).$$

Using the definition (2) of the design points it is straightforward to show that

$$\begin{aligned}\left|\int_0^1 u(t)r(t)dt - \frac{1}{n}\sum_{i=1}^n u(x_i)\right| &= \left|\sum_{i=1}^n \int_{x_{i-1}}^{x_i} u(t)r(t)dt - \sum_{i=1}^n u(x_i)\int_{x_{i-1}}^{x_i} r(t)dt\right| \\ &= \left|\sum_{i=1}^n \int_{x_{i-1}}^{x_i} (u(t)-u(x_i))r(t)dt\right| \\ &\le \sum_{i=1}^n \int_{x_{i-1}}^{x_i} |u(t)-u(x_i)|r(t)dt \\ &\le c\sum_{i=1}^n \int_{x_{i-1}}^{x_i} |t-x_i|^\delta r(t)dt \\ &\le \frac{c}{(n\min_p r(p))^\delta}\sum_{i=1}^n \int_{x_{i-1}}^{x_i} r(t)dt \\ &= \frac{c}{(n\min_p r(p))^\delta}\int_0^1 r(t)dt = O(n^{-\delta}),\end{aligned}$$

where we assume $x_0 = 0$, and $u(t)$ is Hölder continuous of order $\delta > 0$. Thus

$$\left[n^{-1}\sum_{i=1}^n g^2(x_i)\right] - \left[n^{-1}\sum_{i=1}^n g(x_i)\right]^2 = \int_0^1 g^2(t)r(t)dt - \left(\int_0^1 g(t)r(t)dt\right)^2 + O(n^{-\delta}),$$

and so

$$\boldsymbol{g}'\boldsymbol{A}\boldsymbol{g} = k_n\gamma^2 + O(k_n n^{-\delta}) + O(n^{-1}k_n^2),$$

which establishes (21). This concludes the proof.

Proof of Theorem 3.4. The proof is given for the case $k_n \to \infty$, the case k_n is fixed can be proved the same way. $\widehat{e}_i = e_i + (a-\widehat{a}) + (b-\widehat{b})x_i$. Let $\boldsymbol{e}$ and $\boldsymbol{x}$ be respectively the nk_n observations in the hypothetical ANOVA constructed by augmenting $(x_i, e_i), i = 1, \ldots, n$ and $(x_i, x_i), i = 1, \ldots, n$, then

$$\begin{aligned}MST - MSE &= \boldsymbol{e}'\boldsymbol{A}\boldsymbol{e} + (a-\widehat{a})^2\boldsymbol{1}_N'\boldsymbol{A}\boldsymbol{1}_N + (b-\widehat{b})^2\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \\ &\quad +2(a-\widehat{a})\boldsymbol{e}'\boldsymbol{A}\boldsymbol{1}_N + 2(b-\widehat{b})\boldsymbol{e}'\boldsymbol{A}\boldsymbol{x} + 2(a-\widehat{a})(b-\widehat{b})\boldsymbol{1}_N'\boldsymbol{A}\boldsymbol{x} \\ &= \boldsymbol{e}'\boldsymbol{A}\boldsymbol{e} + (b-\widehat{b})^2\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} + 2(b-\widehat{b})\boldsymbol{e}'\boldsymbol{A}\boldsymbol{x},\end{aligned}$$

where the second step follows because $\boldsymbol{A}$ is a contrast matrix. $n^{1/2}k_n^{-1/2}\boldsymbol{e}'\boldsymbol{A}\boldsymbol{e}$ converges in distribution to the designated normal distribution. It's sufficient to show that

$$n^{1/2}k_n^{-1/2}(b-\widehat{b})\boldsymbol{e}'\boldsymbol{A}\boldsymbol{x} = o_p(1), \text{ and } n^{1/2}k_n^{-1/2}(b-\widehat{b})^2\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = o(1).$$

We have similarly as in the proof of Theorem 3.3,

$$\begin{aligned}
&\boldsymbol{e}'\boldsymbol{A}\boldsymbol{x} \\
&= \frac{nk_n - 1}{n(n-1)k_n(k_n-1)} \sum_{i=1}^{n} \left[\sum_{j=1}^{n} x_j I(j \in W_i) \right] \left[\sum_{k=1}^{n} e_k I(k \in W_i) \right] \\
&\quad - \frac{k_n}{n(n-1)k_n} \left[\sum_{i=1}^{n} x_i \right] \left[\sum_{i=1}^{n} e_i \right] - \frac{k_n}{n(k_n-1)} \sum_{i=1}^{n} x_i e_i \\
&= \frac{nk_n}{n-1} \left\{ \left[n^{-1} \sum_{i=1}^{n} x_i e_i \right] - \left[n^{-1} \sum_{i=1}^{n} x_i \right] \left[n^{-1} \sum_{i=1}^{n} e_i \right] \right\} + O_P(n^{-1}k_n^2).
\end{aligned}$$

By checking the mean and variance of the first term of the above display, we have $n^{1/2}k_n^{-1/2}(b - \widehat{b})\boldsymbol{e}'\boldsymbol{A}\boldsymbol{x} = O_p(n^{-1}k_n^{3/2}) = o_p(1)$. We similarly prove $n^{1/2}k_n^{-1/2}(b - \widehat{b})^2\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = o(1)$. □