Far East J. Theo. Stat. **16**(1) (2005), 29-54
(A special volume on Biostatistics and related areas)

TESTING LACK OF FIT FOR HETEROSCEDASTIC NONLINEAR REGRESSION MODELS VIA LOCALLY WEIGHTED LEAST SQUARES REGRESSION

CHIN-SHANG LI

Department of Biostatistics
St. Jude Children's Research Hospital
Memphis, TN 38105, U. S. A.
e-mail: chinshang.li@stjude.org

Abstract

In this paper, we propose a data-driven test for assessing the appropriateness of heteroscedastic nonlinear regression models by using local linear regression smoothers in which no boundary-corrected kernels are needed to resolve boundary effects. The method is proposed for selecting a bandwidth by using the asymptotically optimal bandwidth under the parametric null model. This selection method leads to the data-driven test that has a limiting normal distribution under the null hypothesis and is consistent against any fixed alternative. The resulting test can be applied to testing the lack of fit of a postulated generalized linear model and is compared to some existing tests. We will apply esterase radioimmunoassay data to demonstrate the practical use of the proposed test.

1. Introduction

In recent years, nonparametric regression techniques have rapidly become popular tools for testing the validity of postulated parametric regression models. For example, see Fowlkes [11], Cox et al. [5], Barry

2000 Mathematics Subject Classification: 62G10.

Key words and phrases: bandwidth selection, boundary effects, fit comparison, local linear kernel estimator, quasi-likelihood estimator.

Received November 18, 2004

and Hartigan [2], le Cessie and van Houwelingen [17], Müller [22], Eubank and Hart [8], Härdle and Mammen [12], Stute [24], Zheng [27], Dette [7], Li [18], Alcalá et al. [1], Fan et al. [10], and Horowitz and Spokoiny [16], Fowlkes [11] compared parametric and nonparametric (nearest-neighbor) fits to define residuals for testing the adequacy of binary logistic regression models; however, Fowlkes did not propose a formal test statistic for testing the hypothesis. The kernel smoothing method proposed by le Cessie and van Houwelingen [17] used a weighted sum of the smoothed standardized residuals as the goodness-of-fit measure for binary logistic regression models. In addition, Cox et al. [5], Barry and Hartigan [2], Eubank and Hart [8], and Härdle and Mammen [12] proposed methods that use kernel or series-type smoothers to assess the lack of fit of parametric models for which the usual smoothing parameter asymptotics do not apply. In this paper, we will focus on the situation in which this is not the case, and smoothing parameters can be expressed in a standard, asymptotic manner. As will be seen in Section 2, the large-sample properties of the proposed test are substantially different from others that have been previously discussed in the literature.

Li [18] proposed a data-driven test statistic based on the comparison of parametric and nonparametric Gasser-Müller kernel fits for testing the lack of fit of heteroscedastic nonlinear regression models. In Li's test, the problems of boundary effects are resolved by using boundary-corrected kernels as design points in the boundary regions. This test is applied to assess the linearity of the logit link in the logistic regression models and that of the log link in the Poisson regression models. In contrast to Li's [18] work, the present paper proposes a data-driven test statistic based on the comparison of parametric and local-linear kernel (LLK) fits in which no boundary-corrected kernels are needed for the local-linear fitting as design points in the boundary regions. Alcalá et al. [1] also used LLK smoothers, but they did not provide a practical guideline for bandwidth selection in their test statistic. Local-linear smoothers have the attractive property of automatic boundary correction; the order of the bias at the boundary automatically remains the same as that in the interior. Another attractive property is that local-linear smoothers can adapt to various types of designs such as random and fixed designs or highly clustered and nearly uniform designs. For a more detailed discussion of the advantages of local polynomial smoothers, see Fan and Gijbels [9], Bowman and Azzalini [3], and Hart [14]. The simulation study of an unequally spaced design, which is presented in Section 4, demonstrates that the proposed test is superior to the one proposed by Li [18]. However, both tests have essentially the same power when applied to an equally spaced design.

In this paper, we will study the case in which the smoothing parameter asymptotics can apply under the null model and the random errors are heteroscedastic. Assume that we have responses Y_{in} at design points t_{in} following the model:

$$Y_{in} = m(t_{in}) + \varepsilon_{in}, \quad i = 1, \ldots, n, \tag{1}$$

where m is a smooth, unknown regression function, and ε_{in}'s are random errors with zero expectation and finite variance $\operatorname{var}(\varepsilon_{in}) = \sigma^2(t_{in})$, $i = 1, \ldots, n$. Without loss of generality, we assume that the design points $t_{1n}, \ldots, t_{nn}$ are generated by a design density f on $[0, 1]$ via the relation $\int_0^{t_{in}} f(t)dt = i/n$. In the present paper, our main interest is to test the following parametric hypothesis:

$$H_0 : m(\cdot) = m(\cdot; \boldsymbol{\theta}), \tag{2}$$

where $m(\cdot; \boldsymbol{\theta})$ is a specific nonlinear function of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T \in \boldsymbol{\Theta} \subset \mathbf{R}^p$, which is an unknown vector of p parameters to be estimated, and where, under the null model, $\operatorname{var}(\varepsilon_{in}) = \sigma^2(t_{in}; \boldsymbol{\theta})$, $i = 1, \ldots, n$, for $\sigma^2(\cdot; \boldsymbol{\theta})$, a known function. This is not an unrealistic case; for example, this situation occurs when testing a postulated parametric generalized linear model (GLM). In Section 5, we present a real data set to illustrate the practical use of our methodology.

To test whether the null hypothesis H_0 is false, we measure the distance between the parametric and nonparametric fits and use this distance to test the postulated parametric model. Therefore, we must find a parametric estimator $\mathbf{m}(\hat{\boldsymbol{\theta}}) = (m(t_{1n}; \hat{\boldsymbol{\theta}}), \ldots, m(t_{nn}; \hat{\boldsymbol{\theta}}))^T$ of $\mathbf{m}(\boldsymbol{\theta}_0)$, where $\hat{\boldsymbol{\theta}}$ is an estimator of $\boldsymbol{\theta}_0$, and provide a local linear regression estimator $\hat{\mathbf{m}}_h = (\hat{m}_h(t_{1n}), \ldots, \hat{m}_h(t_{nn}))^T$ for $\mathbf{m} = (m(t_{1n}), \ldots, m(t_{nn}))^T$, where h is a bandwidth or smoothing parameter. Thus, we have the following fit-comparison type of test statistic T_h for H_0:

$$T_h = \| \hat{\mathbf{m}}_h - \mathbf{m}(\hat{\boldsymbol{\theta}}) \|^2 = \sum_{i=1}^{n} (\hat{m}_h(t_{in}) - m(t_{in}; \hat{\boldsymbol{\theta}}))^2. \tag{3}$$

The proposed test statistic can provide omnibus, across-the-design comparisons of a fit under the null model and a local-linear fit that should be closer to the true mean function when H_0 is false. A parametric estimation of $\mathbf{m}(\boldsymbol{\theta}_0)$ and a local linear regression estimation of $\mathbf{m}$ will be discussed in detail in the next section.

The remainder of the paper is organized as follows. Section 2 is an investigation of the behavior of a data-driven version of the test statistic T_h in (3) and shows its asymptotic normality under H_0 after appropriate centering and scaling. The resulting test is consistent against any fixed nonpolynomial alternative. In Section 3, we discuss how the proposed data-driven test statistic can be applied to testing the lack of fit of a postulated parametric GLM model. In Section 4, we study the finite-sample behavior of the proposed test by generating Poisson data. We also compare the power of the proposed test with those of Li [18], Härdle and Mammen [12] and Alcalá et al. [1] in the simulation studies by considering simulations of both an equally spaced design and an unequally spaced design. Section 5 illustrates the practical use of the proposed methodology. Finally, the proofs and required conditions are presented in the Appendix.

2. The Proposed LLK-based Test

To give more details about the data-driven version of T_h in (3), we will first introduce how to obtain a parametric estimator of $\boldsymbol{\theta}_0$ and how to provide a local linear regression estimator of $\mathbf{m}$; we will then describe the large-sample properties of the proposed test.

We can use the quasi-likelihood estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$, which is the solution of the following weighted least-squares:

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}}(\mathbf{Y}_n - \mathbf{m}(\boldsymbol{\theta}))^T \Sigma^{-1}(\boldsymbol{\theta})(\mathbf{Y}_n - \mathbf{m}(\boldsymbol{\theta})),$$

where $\mathbf{Y}_n = (Y_{1n}, \ldots, Y_{nn})^T$ and $\Sigma(\boldsymbol{\theta}) = \operatorname{diag}\{\sigma^2(t_{1n}; \boldsymbol{\theta}), \ldots, \sigma^2(t_{nn}; \boldsymbol{\theta})\}$. Therefore, $\hat{\boldsymbol{\theta}}$ is the solution of the following quasi-likelihood estimation equation:

$$\mathbf{A}^T(\boldsymbol{\theta})\Sigma^{-1}(\boldsymbol{\theta})[\mathbf{Y}_n - \mathbf{m}(\boldsymbol{\theta})] = \mathbf{0},$$

where $\mathbf{A}(\boldsymbol{\theta}) = (\mathbf{a}(t_{1n}; \boldsymbol{\theta}), \ldots, \mathbf{a}(t_{nn}; \boldsymbol{\theta}))^T$ for $\mathbf{a}(t_{in}; \boldsymbol{\theta}) = \partial m(t_{in}; \boldsymbol{\theta})/\partial \boldsymbol{\theta}$. According to Seber and Wild [23, p. 44], $\hat{\boldsymbol{\theta}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_0$. Under H_0 and using conditions (C2) and (C3) given in the Appendix and a Taylor's expansion with the Cauchy-Schwarz inequality, one can show that $\hat{\boldsymbol{\theta}}$ satisfies

$$\mathbf{m}(\hat{\boldsymbol{\theta}}) - \mathbf{m}(\boldsymbol{\theta}_0) = \mathbf{P}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_n + \mathbf{r}_n, \tag{4}$$

where $\boldsymbol{\varepsilon}_n = (\varepsilon_{1n}, \ldots, \varepsilon_{nn})^T$, $\mathbf{r}_n = (r_{1n}, \ldots, r_{nn})^T$, a random vector with

$$\max_{1 \le i \le n} |r_{in}| = o_p(n^{-1/2}), \tag{5}$$

and $\mathbf{P}(\boldsymbol{\theta}_0) = \mathbf{A}(\boldsymbol{\theta}_0)[\mathbf{A}^T(\boldsymbol{\theta}_0)\Sigma^{-1}(\boldsymbol{\theta}_0)\mathbf{A}(\boldsymbol{\theta}_0)]^{-1}\mathbf{A}^T(\boldsymbol{\theta}_0)\Sigma^{-1}(\boldsymbol{\theta}_0)$ satisfies

$$\| \mathbf{P}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_n \|^2 = O_p(1). \tag{6}$$

For a detailed proof of (4), see Li [18].

Now, to obtain an LLK estimator $\hat{\mathbf{m}}_h$ of $\mathbf{m}$, assume the existence of the second derivative of the regression function $m(t)$ at the point z. Then, approximate $m(t)$ locally by using a polynomial of order $p = 2$ that is $m(t) \approx m(z) + m'(z)(t - z) \equiv (1, (t - z))\boldsymbol{\beta}_z$, where $\boldsymbol{\beta}_z = (m(z), m'(z))^T$. Thus, at the point z, the LLK estimator $\hat{m}_h(z)$ of $m(z)$ is equal to $\mathbf{e}_1^T\hat{\boldsymbol{\beta}}_z$, where $\mathbf{e}_1 = (1, 0)^T$ is the 2×1 vector, and $\hat{\boldsymbol{\beta}}_z$ is the solution of the following weighted least-squares:

$$\hat{\boldsymbol{\beta}}_z = \operatorname{argmin}_{\boldsymbol{\beta}_z}(\mathbf{Y}_n - \mathbf{X}_z\boldsymbol{\beta}_z)^T\mathbf{K}_{z;h}(\mathbf{Y}_n - \mathbf{X}_z\boldsymbol{\beta}_z),$$

where $\mathbf{X}_z = [(t_{in} - z)^{j-1}]_{1\le i\le n; 1\le j\le 2}$ is an $n \times 2$ design matrix, and $\mathbf{K}_{z;h} = \operatorname{diag}\{K_h(t_{1n} - z), \ldots, K_h(t_{nn} - z)\}$ is an $n \times n$ diagonal matrix of kernel weights in which K_h denotes the rescaling $K_h(\cdot) = K(\cdot/h)/h$ of the kernel function K. This function is symmetric about zero and supported on $[-1, 1]$; the bandwidth h is used to control the size of the local neighborhood. Let $\mathbf{S}_{z;h} = \mathbf{X}_z^T\mathbf{K}_{z;h}\mathbf{X}_z$, and assume that $\mathbf{S}_{z;h}$ is nonsingular. One can then have the solution $\hat{\boldsymbol{\beta}}_z = \mathbf{S}_{z;h}^{-1}\mathbf{X}_z^T\mathbf{K}_{z;h}\mathbf{Y}_n$ by using weighted least-squares theory. Thus, the LLK estimator of $m(z)$ is

$$\hat{m}_h(z) = \mathbf{e}_1^T\mathbf{S}_{z;h}^{-1}\mathbf{X}_z^T\mathbf{K}_{z;h}\mathbf{Y}_n. \tag{7}$$

As a result, the LLK estimator of $\mathbf{m}$ is $\hat{\mathbf{m}}_h = \mathbf{S}_h\mathbf{Y}_n$, and the associated smoother matrix $\mathbf{S}_h$ can be expressed as

$$\mathbf{S}_h = \begin{bmatrix} \mathbf{e}_1^T\mathbf{S}_{t_{1n};h}^{-1}\mathbf{X}_{t_1}^T\mathbf{K}_{t_{1n};h} \\ \vdots \\ \mathbf{e}_1^T\mathbf{S}_{t_{nn};h}^{-1}\mathbf{X}_{t_n}^T\mathbf{K}_{t_{nn};h} \end{bmatrix}. \tag{8}$$

Note that when $p = 1$, the estimator $\hat{m}_h(z)$ in (7) becomes the Nadaraya-Watson estimator. Now, under H_0 and h of exact order $n^{-1/5}$ and by Lemma 1, which are given in the Appendix, one can see that the expected squared error for the estimator across the design is as follows:

$$E\|\mathbf{S}_h\mathbf{Y}_n - \mathbf{m}(\boldsymbol{\theta}_0)\|^2 \sim C(\boldsymbol{\theta}_0)/h + nh^4J_1\{m(\cdot;\, \boldsymbol{\theta}_0)\} \tag{9}$$

with

$$C(\boldsymbol{\theta}_0) = \int_0^1 \sigma^2(t;\, \boldsymbol{\theta}_0)dt\int_{-1}^1 K^2(u)du$$

and

$$J_1\{m(\cdot;\, \boldsymbol{\theta}_0)\} = \left\{\int_0^1 m''(t;\, \boldsymbol{\theta}_0)^2 f(t)dt\right\}\mu_2^2/4 \tag{10}$$

for $m''(t;\, \boldsymbol{\theta}_0) = \partial^2 m(t;\, \boldsymbol{\theta}_0)/\partial t^2$ and $\mu_2 = \int_{-1}^{1} u^2 K(u)du$. If we now assume that $J_1\{m(\cdot;\, \boldsymbol{\theta}_0)\} > 0$, then the asymptotically optimal bandwidth can be obtained by minimizing (9) with respect to h as follows:

$$h_{opt} = h_{opt}(\boldsymbol{\theta}_0) = \left[\frac{C(\boldsymbol{\theta}_0)}{4nJ_1\{m(\cdot;\, \boldsymbol{\theta}_0)\}}\right]^{1/5}, \tag{11}$$

which decays to zero at the exact rate $n^{-1/5}$. From (10), we can see that the condition $J_1\{m(\cdot;\, \boldsymbol{\theta}_0)\} > 0$ is equivalent to the null regression function $m(\cdot;\, \boldsymbol{\theta}_0)$, which is not a polynomial of order 2. Furthermore, the only unknown in $h_{opt}(\boldsymbol{\theta}_0)$ in (11) is the parameter vector $\boldsymbol{\theta}_0$; therefore, under H_0, one can obtain the estimator $\hat{h}_{opt} = h_{opt}(\hat{\boldsymbol{\theta}})$ of $h_{opt}(\boldsymbol{\theta}_0)$ by replacing $\boldsymbol{\theta}_0$ with the quasi-likelihood estimator $\hat{\boldsymbol{\theta}}$. As a result, we have the following data-driven test statistic:

$$T_{\hat{h}_{opt}} = \| \mathbf{S}_{\hat{h}_{opt}} \mathbf{Y}_n - \mathbf{m}(\hat{\boldsymbol{\theta}}) \|^2. \tag{12}$$

Now, we will focus on the limiting distribution of $T_{\hat{h}_{opt}}$ after recentering and rescaling.

Let

$$J_2\{m(\cdot;\, \boldsymbol{\theta}_0)\} = \frac{\mu_2^2}{4}\int_0^1 \left\{ m''(s;\, \boldsymbol{\theta}_0)\sigma(s;\, \boldsymbol{\theta}_0) - \boldsymbol{\Psi}^T(\boldsymbol{\theta}_0)\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}_0)\frac{\mathbf{a}(s;\, \boldsymbol{\theta}_0)}{\sigma(s;\, \boldsymbol{\theta}_0)} \right\}^2 f(s)ds,$$

where

$$\boldsymbol{\Psi}(\boldsymbol{\theta}_0) = \int_0^1 \mathbf{a}(t;\, \boldsymbol{\theta}_0)m''(t;\, \boldsymbol{\theta}_0)f(t)dt,$$

and

$$\boldsymbol{\Omega}(\boldsymbol{\theta}_0) = \int_0^1 \{\mathbf{a}(t;\, \boldsymbol{\theta}_0)\mathbf{a}^T(t;\, \boldsymbol{\theta}_0)/\sigma^2(t;\, \boldsymbol{\theta}_0)\} f(t)dt.$$

Then, we can state the main result of $T_{\hat{h}_{opt}}$.

Theorem 1. *Let*

$$Z_{\hat{h}_{opt}} = \frac{T_{\hat{h}_{opt}} - \mathrm{tr}[(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))^T(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))\Sigma(\hat{\boldsymbol{\theta}})] - n\hat{h}_{opt}^4 J_1\{m(\cdot;\, \hat{\boldsymbol{\theta}})\}}{[2\mathrm{tr}\{[(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))^T(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))\Sigma(\hat{\boldsymbol{\theta}})]^2\} + 4n\hat{h}_{opt}^4 J_2\{m(\cdot;\, \hat{\boldsymbol{\theta}})\}]^{1/2}}. \quad (13)$$

If conditions (C1) *through* (C6), *which are given in the Appendix, hold, then*

(i) $Z_{\hat{h}_{opt}} \xrightarrow{D} N(0, 1)$ *under* H_0,

(ii) *if* H_0 *is false, then the test statistic* $Z_{\hat{h}_{opt}}$ *satisfies* $P(Z_{\hat{h}_{opt}} \geq q_n)$ $\to 1$ *as* $n \to \infty$ *for any sequence* $q_n = o(n^{1/5})$.

In the theorem, we can see that the test obtained by rejecting H_0 if $Z_{\hat{h}_{opt}}$ exceeds the $100(1-\alpha)$th percentile of the standard normal distribution is an asymptotic $100(1-\alpha)\%$ test for H_0 and is consistent against any fixed alternative.

3. Application to GLM

In this section, we will illustrate how the proposed test statistic can be used to assess the validity of a postulated parametric GLM. In the GLM setting, the random variables Y_{in} have a probability density function or mass function of the following form:

$$f(y_{in};\, \vartheta_{in}) = \exp\left\{\frac{y_{in}\vartheta_{in} - b(\vartheta_{in})}{a(\phi)} + c(y_{in}, \phi)\right\}, \quad i = 1, \ldots, n.$$

Here the ϑ_{in}'s are the parameters of interest, and $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are the functions of known form with the dispersion parameter ϕ. For details, see McCullagh and Nelder [20]. Let $\vartheta_{in} = g\{E(Y_{in})\} = g\{m(t_{in})\}$ for $g(\cdot)$, which is a known, monotonic, differentiable-link function. Now, we consider the following null hypothesis:

$$H_0 : g\{m(\cdot)\} = \eta(\cdot;\, \boldsymbol{\theta}),$$

where $\eta(\cdot;\, \boldsymbol{\theta})$ is a known functional form apart from the vector of parameters $\boldsymbol{\theta}$, then under H_0, we have $m(t_{in}; \boldsymbol{\theta}) = b'(\vartheta_{in})$, and $\sigma^2(t_{in}; \boldsymbol{\theta}) = a(\phi)b''(\vartheta_{in})$, which is a function of $\boldsymbol{\theta}$ only. Therefore, testing $H_0 : g\{m(\cdot)\} = \eta(\cdot; \boldsymbol{\theta})$ is equivalent to testing the following hypothesis:

$$H_0 : m(\cdot) = g^{-1}\{\eta(\cdot; \boldsymbol{\theta})\}.$$

This illustrates that testing H_0 is a special case of testing the lack of fit of heteroscedastic nonlinear regression models.

In the GLM setting, we find that the maximum-likelihood estimator of $\boldsymbol{\theta}$ is the solution of $\mathbf{A}^T(\boldsymbol{\theta})\Sigma^{-1}(\boldsymbol{\theta})[\mathbf{Y}_n - \mathbf{m}(\boldsymbol{\theta})] = \mathbf{0}$; therefore, the maximum-likelihood and quasi-likelihood estimators coincide in the setting. As a result, the test statistic $Z_{\hat{h}_{opt}}$ proposed in Theorem 1 can be applied.

4. Simulation

In this section, we will present the Monte Carlo results for the Poisson regression models to compare the power performance of the proposed LLK-based test with that of the nonparametric boundary-corrected kernel (NK)-based test by Li [18], the HM test, which is based on the distance between a kernel nonparametric fit and a kernel-smoothed parametric fit, by Härdle and Mammen [12], and the ACG test by Alcalá et al. [1] by considering the case in which $\eta(t; \boldsymbol{\theta}) = \theta_1 + \theta_2 t$ is a simple linear predictor of $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$. Two design types will be considered: one is an equally spaced design in which design points $t_{in} = (2i - 1)/2n$, $i = 1, \ldots, n$ for $n = 100$ or 200, and the other is an unequally spaced design in which design points with a sample size of 100 or 200 were generated only once from $U[0, 1]$ and then used throughout all of the simulations. Theorem 1 provides a rough idea of the stochastic behavior of $Z_{\hat{h}_{opt}}$, but if the sample is small or moderate, the normal approximation does not work well; see, e.g., Härdle and Mammen [12]. Therefore, in the simulation studies, we will use the *wild bootstrap* related to the proposals of Wu [26] as an alternative to the normal asymptotic method.

To generate bootstrap data, we first define standardized residuals on the linear predictor scale on the basis of the suggestion of Davison and Hinkley (6) as follows:

$$\hat{\varepsilon}_{Lin} = \frac{g(y_{in}) - g(m_{\hat{h}_{opt}}(t_{in}))}{\sqrt{g'^2(m_{\hat{h}_{opt}}(t_{in}))\,\mathrm{var}(m_{\hat{h}_{opt}}(t_{in}))(1 - s_{\hat{h}_{opt}ii})}}, \quad i = 1, \ldots, n,$$

where $s_{\hat{h}_{opt}ii}$ is the ith diagonal element of the LLK smoother matrix $\mathbf{S}_{\hat{h}_{opt}}$. Let ε_{in}^* denote the bootstrap residuals generated based on the residuals $\hat{\varepsilon}_{in} = g'(m_{\hat{h}_{opt}}(t_{in}))\,\mathrm{var}^{1/2}(m_{\hat{h}_{opt}}(t_{in}))\hat{\varepsilon}_{Lin}$. Let $\hat{F}_i$ be an arbitrary distribution such that (i) $E_{\hat{F}_i}(\varepsilon_{in}^*) = 0$, (ii) $E_{\hat{F}_i}(\varepsilon_{in}^{*2}) = \hat{\varepsilon}_{in}^2$, and (iii) $E_{\hat{F}_i}(\varepsilon_{in}^{*3}) = \hat{\varepsilon}_{in}^3$. We use a two-point distribution satisfying the aforementioned three conditions to generate the residuals ε_{in}^*, more specifically, $\varepsilon_{in}^* = \hat{\varepsilon}_{in}(1 - \sqrt{5})/2$ with probability $(5 + \sqrt{5})/10$, and $\varepsilon_{in}^* = \hat{\varepsilon}_{in}(1 + \sqrt{5})/2$ with probability $(5 - \sqrt{5})/10$; see Liu [19] for other constructions. We can then compute the *wild-bootstarp* test statistic $Z^*_{\hat{h}^*_{opt}}$ in the same manner that we compute the LLK-based test statistic $Z_{\hat{h}_{opt}}$.

We can obtain the $(1 - \alpha)$-quantile, z^*_α, by using the Monte Carlo approximation of $Z^*_{\hat{h}_{opt}}$ and reject the null hypothesis if $Z_{\hat{h}_{opt}} > z^*_\alpha$. The nominal level $\alpha = 0.05$ for all tests was used in our simulations. We conducted 1,000 replications for each configuration of the experiment, and we performed the bootstrap resampling 100 times for each sample. To give a fair comparison, we also employed the *wild-bootstrap* algorithm to approximate the finite-sample distributions of the NK-based test statistic, the HM test statistic, and the ACG test statistic.

The H_0 that we wanted to test was

$$H_0 : m(t) = \exp(\theta_1 + \theta_2 t) = m(t; \boldsymbol{\theta}), \tag{14}$$

and the data were generated from models with the following mean functions:

$$m(t;\, \boldsymbol{\theta},\, \beta) = \exp(\theta_1 + \theta_2 t + \beta t^2) \tag{15}$$

for $\beta = 0, 0.125, \ldots, 2.875$, and $n = 100$; $\beta = 0, 0.125, \ldots, 2.25$, and $n = 200$ for the equally spaced design; $\beta = 0, 0.125, \ldots, 3.0$, and $n = 100$; $\beta = 0, 0.125, \ldots, 2.75$, and $n = 200$ for the unequally spaced design. When $\beta = 0$, the generating model (15) becomes the null model (14), and the generating model (15) provides alternatives that become farther from the null model as $|\beta|$ increases.

To compute the LLK-based test statistic $Z_{\hat{h}_{opt}}$, we used the Epanechnikov kernel; to compute the NK-based test statistic, we used the Epanechnikov kernel and associated second-order boundary-corrected kernel from Müller [21]. Hence, let $\hat{\theta}_3 = m(0, \hat{\boldsymbol{\theta}})$, and $\hat{\theta}_4 = m(1, \hat{\boldsymbol{\theta}})$. Then, in our simulations, the data-driven bandwidth $\hat{h}_{opt} = \{0.15(\hat{\theta}_4 - \hat{\theta}_3)/n\hat{\theta}_2 J_1\{m(\cdot;\, \hat{\boldsymbol{\theta}})\}\}^{1/5}$ with $J_1\{m(\cdot;\, \hat{\boldsymbol{\theta}})\} = 0.005\hat{\theta}_2^3(\hat{\theta}_4^2 - \hat{\theta}_3^2)$, and $J_2\{m(\cdot;\, \hat{\boldsymbol{\theta}})\} = 0.01\{\hat{\theta}_2^3(\hat{\theta}_4^3 - \hat{\theta}_3^3)/3 - \boldsymbol{\Psi}^T(\hat{\boldsymbol{\theta}})\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{\Psi}(\hat{\boldsymbol{\theta}})\}$, where $\boldsymbol{\Psi}(\hat{\boldsymbol{\theta}}) = (0.5\hat{\theta}_2(\hat{\theta}_4^2 - \hat{\theta}_3^2),\, 0.25[(2\hat{\theta}_2 - 1)\hat{\theta}_4^2 + \hat{\theta}_3^2])^T$, and

$$\boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \hat{\theta}_2^{-1}(\hat{\theta}_4 - \hat{\theta}_3) & \hat{\theta}_2^{-2}[(\hat{\theta}_2 - 1)\hat{\theta}_4 + \hat{\theta}_3] \\ \hat{\theta}_2^{-2}[(\hat{\theta}_2 - 1)\hat{\theta}_4 + \hat{\theta}_3] & \hat{\theta}_2^{-3}[(\hat{\theta}_2^2 - 2\hat{\theta}_2 + 2)\hat{\theta}_4 - \hat{\theta}_3] \end{bmatrix}.$$

The matrix $\mathbf{P}(\hat{\boldsymbol{\theta}})$ can be obtained by using $\mathbf{a}(t_{in};\, \hat{\boldsymbol{\theta}}) = (m(t_{in};\, \hat{\boldsymbol{\theta}}),\, t_{in}m(t_{in};\, \hat{\boldsymbol{\theta}}))^T$, and $\Sigma(\hat{\boldsymbol{\theta}}) = \operatorname{diag}\{m(t_{1n};\, \hat{\boldsymbol{\theta}}), \ldots, m(t_{nn};\, \hat{\boldsymbol{\theta}})\}$. We computed the HM test statistic by using a local constant (Nadaraya-Watson) smoother and the ACG test statistic by using a local linear smoother with the quartic kernel as Härdle and Mammen, and Alcalá et al. used in their papers. However, they did not provide a practical guideline for bandwidth selection in their test statistics in applications, so we chose smoothing parameters such that when H_0 is true, the empirical-rejection probabilities of the two tests do not differ from the nominal level 0.05 at a significance level of 0.01, when the two test statistics are computed through simulations.

For each sample, the values of the LLK-based, NK-based, HM, and ACG test statistics for testing $\beta = 0$ in (15) were computed and compared with the approximated 0.95-quantiles of their bootstrap distributions. The empirical powers of the equally spaced design are shown in Figure 1; the power performance of the NK-based and LLK-based tests is essentially the same, and the LLK-based test outperforms the HM test with chosen values of smoothing parameter 0.05 and 0.016 for sample size 100 and 200, respectively. More specifically, for $n = 100$ when H_0 is false and $\beta = 0.75, 0.875, \ldots, 1.875,$ the power of the proposed LLK-based test is much higher than that of the HM test, and the difference between the powers of the two tests is significant at the significance level of 0.01. For $n = 200,$ the power of the proposed LLK-based test is significantly higher than that of the NK-based test at the significance level of 0.01 when H_0 is false and $\beta = 0.375, 0.5, \ldots, 1.875.$

For the unequally spaced design, the empirical powers are plotted in Figure 2. When H_0 is true, all tests have empirical-rejection probabilities close to the nominal level 0.05, and none of the empirical-rejection probabilities differs from the nominal level 0.05 at the 0.01 significance level. The proposed LLK-based test outperforms both the NK-based test and the HM test with chosen values of smoothing parameter 0.055 for $n = 100$ and 0.02 for $n = 200.$ Specifically for $n = 100,$ when H_0 is false, the power of the LLK-based test is significantly higher than that of the NK-based test, when $\beta = 1.25, 1.375, \ldots, 1.75,$ and that of the HM test, when $\beta = 0.5, 0.625, \ldots, 2.375,$ at the 0.01 significance level. For $n = 200,$ when H_0 is false, the power of the proposed LLK-based test is significantly higher than that of the NK-based test, when $\beta = 0.5, 0.625, \ldots, 2.25,$ and the proposed LLK-based test has significantly higher power than the HM test, when $\beta = 0.375, 0.5, \ldots, 1.875,$ at the 0.01 significance level. This finding illustrates the advantage of local-linear fitting and shows that the proposed LLK-based test is preferable to the NK-based test when the design points are unequally spaced. Furthermore, in the

LLK-based test, specific boundary-corrected kernels are not needed to resolve boundary effects.

The power performance of the LLK-based test is slightly better than that of the ACG test with selected bandwidths 0.145 for $n = 100$ and 0.138 for $n = 200$ for equally spaced design and with selected bandwidths 0.130 for $n = 100$ and 0.115 for $n = 200$ for unequally spaced design. Furthermore, the proposed LLK-based test is more practical than the ACG test, because no practical guideline for bandwidth selection was provided in the ACG test.

5. Real-life Example

In this section, we illustrate the practical use of the proposed LLK-based test statistic with the esterase radioimmunoassay (RIA) data set (Carroll and Ruppert [4, p. 46]) and include the NK-based test statistic by Li [18]. The data consist of RIA assessment of esterase levels in 106 samples; the level of radioactivity is the response, and the concentration of esterase is the covariate. The data are shown in Figure 3; the t variable is used to denote the concentration of esterase rescaled so that all points fall within the interval $[0, 1]$.

To investigate the relationship between the concentration of esterase and the RIA count, we consider fitting a log-linear model whose linear predictor is a polynomial of order p as follows:

$$H_0 : \log(m(t)) = \eta(t, \boldsymbol{\theta}) \tag{16}$$

for $\eta(t, \boldsymbol{\theta}) = \theta_1 + \theta_2 t + \cdots + \theta_p t^{p-1}$. We first wish to assess the validity of the following parametric log-linear model to the data set:

$$H_{10} : \eta(t, \boldsymbol{\theta}) = \theta_1 + \theta_2 t, \tag{17}$$

and $\sigma^2(t) = \phi \exp(\theta_1 + \theta_2 t)$ under H_{10}. The value of the proposed LLK-based test statistic $Z_{\hat{h}_{opt}}$ for H_{10} is 10.48 and that of the NK-based test statistic by Li [18] for H_{10} is 4.77. The 0.95-quantile of the bootstrap distribution approximating the sample distribution of the LLK-based test statistic under H_{10} in (17) is 3.44; that of the NK-based test statistic is

1.99. Therefore, the two bootstrap tests rejected the postulated parametric model at the significance level 0.05, which suggests that according to both test statistics, the proposed log-linear model in (17) cannot fit the data set well.

We now fit the data set by the following log-linear model:

$$H_{20} : \eta(t, \boldsymbol{\theta}) = \theta_1 + \theta_2 t + \theta_3 t^2, \tag{18}$$

and $\sigma^2(t) = \phi \exp(\theta_1 + \theta_2 t + \theta_3 t^2)$ under H_{20}. The observed LLK-based test statistic for the model in (18) is 0.92, and the approximated 0.95-quantile of the sample distribution of the LLK-based test statistic under H_{20} in (18) is 1.71 from the bootstrap distribution. The observed NK-based test statistic is –0.462, and the 0.95-quantile of the bootstrap distribution approximating the sample distribution of the NK test statistic under H_{20} in (18) is 0.91. As a result, the two bootstrap tests do not reject the proposed log-linear model in (18), which suggests that the parametric model fits the data set well enough. The parametric, LLK-based and NK-based fits are given in Figure 3, in which the fitted parametric log-linear model is $m(t; \hat{\boldsymbol{\theta}}) = \exp(4.65 + 4.57t - 2.65t^2)$.

Appendix

To prove Theorem 1 given in Section 2, we need some required conditions and a series of lemmas. The arguments of the proof of the theorem are similar to those of Theorem 2.1 in Li [18]. Hence, we will give only a brief sketch of the proof. We will give the lemmas and the proofs of the lemmas after the proof of the Theorem. To begin, let us first outline the following required conditions:

(C1) $\lim_{n\to\infty} n^{-1}\mathbf{A}^T(\boldsymbol{\theta})\Sigma^{-1}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) = \boldsymbol{\Omega}(\boldsymbol{\theta})$, which is nonsingular, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}_0$ is an open neighborhood of $\boldsymbol{\theta}_0$, and $\boldsymbol{\Theta}$ is a compact region.

(C2) $\sup_t \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} |\partial m(t; \boldsymbol{\theta})/\partial \boldsymbol{\theta}_j|$ is bounded for $j = 1, \ldots, p$.

(C3) $\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} \sup_t |\partial m(t; \boldsymbol{\theta})/\partial \boldsymbol{\theta}_j - \partial m(t; \boldsymbol{\theta})/\partial \boldsymbol{\theta}_j|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}| = 0$ for $j = 1, \ldots, p$.

(C4) $m''(t;\, \boldsymbol{\theta}) = \partial^2 m(t;\, \boldsymbol{\theta})/\partial t^2$ is Lipschitz continuous of order ζ in t for $\boldsymbol{\theta} \in \Theta$.

(C5) $\sup_t \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} |\, \partial m''(t;\, \boldsymbol{\theta})/\partial \boldsymbol{\theta}_j \,|$ is bounded for $j = 1, \ldots, p$.

(C6) $\max_{1\le i\le n} E|\, \varepsilon_{in}\,|^{20+\delta_1}$ is bounded for some $\delta_1 > 0$.

Proof of Theorem 1. Using Lemma 1, the Cauchy-Schwarz inequality, $\|\, \mathbf{P}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_n \,\|^2 = O_p(1)$, $\max_{1\le i\le n}|\, r_{in}\,| = o_p(n^{-1/2})$, and $\|\, \mathbf{S}_h\boldsymbol{\varepsilon}_n \,\|^2 = O_p(h^{-1})$, one can rewrite $T_{\hat{h}_{opt}}$ in (12) as

$$T_{\hat{h}_{opt}} = \|\, (\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))\boldsymbol{\varepsilon}_n \,\|^2 - 2\boldsymbol{\varepsilon}_n^T(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T(\mathbf{I} - \mathbf{S}_{\hat{h}_{opt}})\mathbf{m}(\boldsymbol{\theta}_0)$$

$$+ n\hat{h}_{opt}^4 J_1\{m(\cdot;\, \boldsymbol{\theta}_0)\} + o_p(n^{1/10}). \tag{19}$$

Furthermore, by using Lemmas 1 through 3, Chebychev's inequality, and Slutsky's Theorem, we can express (19) as

$$T_{\hat{h}_{opt}} - \mathrm{tr}[(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))^T(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))\Sigma(\hat{\boldsymbol{\theta}})] - n\hat{h}_{opt}^4 J_1\{m(\cdot;\, \hat{\boldsymbol{\theta}})\}$$

$$= \sum_{i\ne j} w_{ij;\, h_{opt}}\varepsilon_{in}\varepsilon_{jn} - 2\sum_{i=1}^{n} \ell_i\varepsilon_{in} + O_p(1), \tag{20}$$

where $w_{ij;\, h_{opt}}$ is the $(i,\, j)$-element of $\mathbf{S}_{h_{opt}}^T\mathbf{S}_{h_{opt}}$, and ℓ_{kn} is the kth element of $\mathcal{L}_n = (\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T(\mathbf{I} - \mathbf{S}_{h_{opt}})\mathbf{m}(\boldsymbol{\theta}_0)$. Therefore, by using Lemma 5 and the Central Limit Theorem in Heyde and Brown [15], we can show that

$$\left(\sum_{i\ne j} w_{ij;\, h_{opt}}\varepsilon_{in}\varepsilon_{jn} - 2\sum_{i=1}^{n} \ell_{in}\varepsilon_{in}\right)\Big/ v_n \xrightarrow{D} N(0,\, 1), \tag{21}$$

where $v_n^2 = 2\mathrm{tr}\{[(\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T(\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))\Sigma(\boldsymbol{\theta}_0)]^2\} + 4\|\, \Sigma^{1/2}(\boldsymbol{\theta}_0)\mathcal{L}_n \,\|^2$, which is $\asymp n^{1/5}$ because of Lemmas 1 and 4. Finally, by using Lemmas 3 and 4, we can estimate v_n^2 consistently by $2\mathrm{tr}\{[(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))^T(\mathbf{S}_{\hat{h}_{opt}} -$

$\mathbf{P}(\hat{\boldsymbol{\theta}}))\Sigma(\hat{\boldsymbol{\theta}})]^2\} + 4n\hat{h}_{opt}^4 J_2\{m(\cdot;\, \hat{\boldsymbol{\theta}})\}$, and by using Slutsky's theorem, the proof can be finished from (20) and (21).

The arguments of the proof of consistency are similar to those presented in Li [18]; therefore, the proof is omitted.

Lemma 1. *If* $h \asymp n^{-1/5}$, *then*

(i) $\operatorname{tr}[\mathbf{S}_h^T \mathbf{S}_h\, \Sigma(\boldsymbol{\theta}_0)] \sim \int_0^1 \sigma^2(t;\, \boldsymbol{\theta}_0)dt \int_{-1}^1 K^2(u)du/h$,

(ii) $\operatorname{tr}[(\mathbf{S}_h - \mathbf{P}(\boldsymbol{\theta}_0))^T(\mathbf{S}_h - \mathbf{P}(\boldsymbol{\theta}_0))\Sigma(\boldsymbol{\theta}_0)] \sim \int_0^1 \sigma^2(t;\, \boldsymbol{\theta}_0)dt \int_{-1}^1 K^2(u)du/h$,

(iii) $\|(\mathbf{I} - \mathbf{S}_h)\mathbf{m}(\boldsymbol{\theta}_0)\|^2 = nh^4 J_1\{m(\cdot;\, \boldsymbol{\theta}_0)\} + o(nh^4)$,

(iv) $J_1\{m(\cdot;\, \hat{\boldsymbol{\theta}})\} - J_1\{m(\cdot;\, \boldsymbol{\theta}_0)\} = o_p(n^{-1/10})$,

(v) $\operatorname{tr}\{[\mathbf{S}_h^T \mathbf{S}_h\, \Sigma(\boldsymbol{\theta}_0)]^2\} \sim \int_0^1 \sigma^4(t;\, \boldsymbol{\theta}_0)dt \int_{-2}^2 K^{*2}(z)dz/h$, *where* $K^*(z) = \int_{-1}^1 K(u)K(z-u)du$ *is the convolution kernel with support on* $[-2, 2]$, *and*

(vi) $\operatorname{tr}\{[(\mathbf{S}_h - \mathbf{P}(\boldsymbol{\theta}_0))^T(\mathbf{S}_h - \mathbf{P}(\boldsymbol{\theta}_0))\Sigma(\boldsymbol{\theta}_0)]^2\} = \operatorname{tr}\{[\mathbf{S}_h^T \mathbf{S}_h\, \Sigma(\boldsymbol{\theta}_0)]^2\} + O(1)$.

Proof. Let $\mathbf{S}_{t_{in};h} = \mathbf{X}_{t_{in}}^T \mathbf{K}_{t_{in};h} \mathbf{X}_{t_{in}} = [s_{t_{in};h;j+\ell}]_{0 \le j, \ell \le 1}$, $i = 1, \ldots, n$, $F_n(\cdot)$ be the empirical distribution for the design, and $F(t) = \int_0^t f(s)ds$. Set $B_h = LB_h \cup UB_h$, where $LB_h = \{t : t < h\}$, and $UB_h = \{t : t > 1 - h\}$, $\mu_r = \int_{-1}^1 u^r K(u)du$, $\mu_{r,-\alpha_i} = \int_{-\alpha_i}^1 u^r K(u)du$, and $\mu_{r,\alpha_i} = \int_{-1}^{\alpha_i} u^r K(u)du$. Using $\sup_t |F_n(t) - F(t)| = O(n^{-1})$, one can show that

$$\begin{aligned} s_{t_{in};h;r} &= \sum_{\ell=1}^n (t_{\ell n} - t_{in})^r K_h(t_{\ell n} - t_{in}) \\ &= nh^r f(t_{in})\mu_{r,t_{in}} + o(nh^r) + O(1), \qquad (22) \end{aligned}$$

where $\mu_{r, t_{in}}$ is μ_r if $t_{in} \notin B_h$, $\mu_{r, -\alpha_i}$ if $t_{in} \in LB_h$ (i.e., $t_{in} = \alpha_i h$, $0 \le \alpha_i < 1$), or μ_{r, α_i} if $t_{in} \in UB_h$ (i.e., $t_{in} = 1 - \alpha_i h$, $0 \le \alpha_i < 1$). Therefore, $\mathbf{S}_{t_{in}; h}$ can be expressed as

$$\mathbf{S}_{t_{in}; h} = nf(t_{in})\mathbf{H}\mathbf{M}_{t_{in}}\mathbf{H}\{1 + o(1)\}, \tag{23}$$

where $\mathbf{H} = \text{diag}\{1, h\}$, and $\mathbf{M}_{t_{in}}$ is $\mathbf{M} = [\mu_{j+\ell}]_{0 \le j, \ell \le 1}$ if $t_{in} \notin B_h$, $\mathbf{M}_{-\alpha_i} = [\mu_{j+\ell; -\alpha_i}]_{0 \le j, \ell \le 1}$ if $t_{in} = \alpha_i h$, or $\mathbf{M}_{\alpha_i} = [\mu_{j+\ell; \alpha_i}]_{0 \le j, \ell \le 1}$ if $t_{in} = 1 - \alpha_i h$.

Let $\mathbf{S}^*_{t_{in}; h} = \mathbf{X}^T_{t_{in}}\mathbf{K}_{t_{in}; h}\Sigma(\boldsymbol{\theta}_0)\mathbf{K}_{t_{in}; h}\mathbf{X}_{t_{in}}$, $\mathbf{M}^* = [\nu_{j+\ell}]_{0 \le j, \ell \le 1}$, $\mathbf{M}^*_{-\alpha_i} = [\nu_{j+\ell; -\alpha_i}]_{0 \le j, \ell \le 1}$, and $\mathbf{M}^*_{\alpha_i} = [\nu_{j+\ell; \alpha_i}]_{0 \le j, \ell \le 1}$, where $\nu_r = \int_{-1}^{1} u^r K^2(u)du$, $\nu_{r, -\alpha_i} = \int_{-\alpha_i}^{1} u^r K^2(u)du$, and $\nu_{r, \alpha_i} = \int_{-1}^{\alpha_i} u^r K^2(u)du$. Because $\sup_t |F_n(t) - F(t)| = O(n^{-1})$, we can show that

$$\mathbf{S}^*_{t_{in}; h} = nh^{-1}f(t_{in})\sigma^2(t_{in}; \boldsymbol{\theta}_0)\mathbf{H}\mathbf{M}^*_{t_{in}}\mathbf{H}\{1 + o(1)\}, \tag{24}$$

where $\mathbf{M}^*_{t_{in}}$ is $\mathbf{M}^*$ if $t_{in} \notin B_h$, $\mathbf{M}^*_{-\alpha_i}$ if $t_{in} = \alpha_i h$, or $\mathbf{M}^*_{\alpha_i}$ if $t_{in} = 1 - \alpha_i h$. The proof of (i) can be finished by using the cyclic property of the trace, (23), (24), and $\sup_t |F_n(t) - F(t)| = O(n^{-1})$.

Let $p_{ij}(\boldsymbol{\theta}_0)$ be the (i, j)-element of $\mathbf{P}(\boldsymbol{\theta}_0)$ and $s_{ij; h}$ be that of $\mathbf{S}_h$. One can then obtain the following facts:

$$\max_{i, j} |p_{ij}(\boldsymbol{\theta}_0)| = O(n^{-1}), \tag{25}$$

$$\max_{i, j} |p_{ij}(\boldsymbol{\theta}_0)\sigma^2(t_{jn}; \boldsymbol{\theta}_0)| \le \max_{i, j} |p_{ij}(\boldsymbol{\theta}_0)| \max_j \sigma^2(t_{jn}; \boldsymbol{\theta}_0) = O(n^{-1}), \text{ and} \tag{26}$$

$$\max_i \left| \sum_{j=1}^{n} s_{ij; h}d_j \right| \le \max_j |d_j| O(1) = O(\max_j |d_j|). \tag{27}$$

Then, using $\mathbf{P}(\boldsymbol{\theta}_0)\Sigma(\boldsymbol{\theta}_0)\mathbf{P}^T(\boldsymbol{\theta}_0) = \mathbf{P}(\boldsymbol{\theta}_0)\Sigma(\boldsymbol{\theta}_0) = \Sigma(\boldsymbol{\theta}_0)\mathbf{P}^T(\boldsymbol{\theta}_0)$, (25) through (27), and the cyclic property of the trace, the proof of (ii) is finished.

(iii) By applying a Taylor's expansion to $\mathbf{m}(\boldsymbol{\theta}_0)$, the ith-component of $\mathbf{S}_h\mathbf{m}(\boldsymbol{\theta}_0)$ is

$$\mathbf{e}_1^T\mathbf{S}_{t_{in};h}^{-1}\mathbf{X}_{t_{in}}^T\mathbf{K}_{t_{in};h}\mathbf{m}(\boldsymbol{\theta}_0)$$

$$= m(t_{in};\ \boldsymbol{\theta}_0) + \frac{1}{2}m''(t_{in};\ \boldsymbol{\theta}_0)\mathbf{e}_1^T\mathbf{S}_{t_{in};h}^{-1}(s_{t_{in};h;2},\ s_{t_{in};h;3})^T + o(h^2). \tag{28}$$

Thus, by (22), the ith-component of $(\mathbf{I}-\mathbf{S}_h)\mathbf{m}(\boldsymbol{\theta}_0)$ is

$$-\frac{1}{2}h^2m''(t_{in};\ \boldsymbol{\theta}_0)\mathbf{e}_1^T\mathbf{M}_{t_{in}}^{-1}(\mu_{2,t_{in}},\ \mu_{3,t_{in}})^T\{1+o(1)\}. \tag{29}$$

Consequently, because $\sup_t|F_n(t)-F(t)| = O(n^{-1})$, we can obtain the desired result from (29).

(iv) The proof is immediately done because $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p(n^{-1/2})$, and $\int_0^1 m''(t;\ \hat{\boldsymbol{\theta}})^2 f(t)dt = \int_0^1 m''(t;\ \boldsymbol{\theta}_0)^2 f(t)dt + O_p(n^{-1/2})$.

(v and vi) Let $B_h^* = LB_h^* \cup UB_h^*$ for $LB_h^* = \{t : t < 2h\}$ and $UB_h^* = \{t : t > 1-2h\}$. Let $\mathbf{W}_h = [w_{ij;h}] = \mathbf{S}_h^T\mathbf{S}_h$. Then, for n sufficiently large, by using the result in (23), a Taylor's expansion and $\sup_t|F_n(t)-F(t)| = O(n^{-1})$, one can show

$$w_{ij;h} = \begin{cases} O\left(\frac{1}{nh}\right), & \text{if } t_{in}, t_{jn} \in LB_h^* \text{ or } t_{in}, t_{jn} \in UB_h^*, \\ \frac{K^*\left(\frac{t_{in}-t_{jn}}{h}\right)}{nhf(t_{jn})} + o\left(\frac{1}{nh}\right), & \text{if } (t_{in} \in LB_h^* \text{ and } t_{jn} \notin B_h^*), (t_{in} \notin B_h^* \text{ and } t_{jn} \in LB_h^*), \\ & \quad (t_{in} \in UB_h^* \text{ and } t_{jn} \notin B_h^*), (t_{in} \notin B_h^* \text{ and } t_{jn} \in UB_h^*), \\ & \quad (t_{in}, t_{jn} \notin B_h^*), \text{and satisfies } |t_{in}-t_{jn}| \le 2h, \\ 0, & \text{if } |t_{in}-t_{jn}| > 2h. \end{cases}$$

This proof can be done by using the above fact, the cyclic property of the trace, (25) through (27), and $\sup_t|F_n(t)-F(t)| = O(n^{-1})$.

Lemma 2. *If conditions* (C5) *and* (C6) *hold, then*

(i) $\|(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))\boldsymbol{\varepsilon}_n\|^2 - \mathrm{tr}[(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))\Sigma(\boldsymbol{\theta}_0)]$

$-\| (\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))\boldsymbol{\varepsilon}_n \|^2 + \mathrm{tr}[(\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T (\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))\Sigma(\boldsymbol{\theta}_0)] = o_p(n^{1/10}),$

(ii) $\boldsymbol{\varepsilon}_n^T (\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T (\mathbf{I} - \mathbf{S}_{\hat{h}_{opt}})\mathbf{m}(\boldsymbol{\theta}_0) - \boldsymbol{\varepsilon}_n^T (\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T$

$(\mathbf{I} - \mathbf{S}_{h_{opt}})\mathbf{m}(\boldsymbol{\theta}_0) = o_p(n^{1/10}),$

and

(iii) $\boldsymbol{\varepsilon}_n^T \mathbf{S}_{\hat{h}_{opt}}^T (\mathbf{I} - \mathbf{S}_{\hat{h}_{opt}})\mathbf{m}(\boldsymbol{\theta}_0) - \boldsymbol{\varepsilon}_n^T \mathbf{S}_{h_{opt}}^T (\mathbf{I} - \mathbf{S}_{h_{opt}})\mathbf{m}(\boldsymbol{\theta}_0) = o_p(n^{1/10}).$

Proof. (i) First note that $(\hat{h}_{opt} - h_{opt})/h_{opt} = O_p(n^{-1/2})$ by using a Taylor's expansion and (C5). To finish the proof, we use a partitioning argument similar to one used in Härdle et al. [13]. Since $P(| \hat{h}_{opt} - h_{opt} |/h_{opt} \leq n^{-\gamma}) \to 1$ for $\gamma < 1/2$, we need to consider only those values of h falling in $H_n = \{h : | h - h_{opt} |/h_{opt} \leq n^{-\gamma}\}$. Let $w_{rj; h_1, h_2}$ be the (r, j)-element of $(\mathbf{S}_{h_1} - \mathbf{P}(\boldsymbol{\theta}_0))^T (\mathbf{S}_{h_1} - \mathbf{P}(\boldsymbol{\theta}_0)) - (\mathbf{S}_{h_2} - \mathbf{P}(\boldsymbol{\theta}_0))^T (\mathbf{S}_{h_2} - \mathbf{P}(\boldsymbol{\theta}_0))$. Then, for $h_1, h_2 \in H_n$, one can show that $w_{rj; h_1, h_2} = O(| h_1 - h_2 |/nh_1^2)$ uniformly in r, j by using (23) and the Mean Value Theorem. The remainder of the proof can be finished by using the moment condition (C6) and the inequality (8) from Whittle [25].

The arguments of proving (ii) and (iii) are similar to those for (i), so it is omitted.

Lemma 3. *If* $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p(n^{-1/2})$, *then*

(i) $\mathrm{tr}[(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))^T (\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))\Sigma(\hat{\boldsymbol{\theta}})] - \mathrm{tr}[(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T (\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))\Sigma(\boldsymbol{\theta}_0)] = o_p(n^{1/10})$, *and*

(ii) $\mathrm{tr}\{[(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))^T (\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\hat{\boldsymbol{\theta}}))\Sigma(\hat{\boldsymbol{\theta}})]^2\} - \mathrm{tr}\{[(\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T (\mathbf{S}_{\hat{h}_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))\Sigma(\boldsymbol{\theta}_0)]^2\} = o_p(n^{1/10})$.

Proof. The proof can be done by using $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p(n^{-1/2})$, $\mathbf{A}(\hat{\boldsymbol{\theta}}) - \mathbf{A}(\boldsymbol{\theta}_0) = O_p(n^{-1/2})$, $\mathbf{P}(\hat{\boldsymbol{\theta}}) - \mathbf{P}(\boldsymbol{\theta}_0) = O_p(n^{-3/2})$, and (27).

Lemma 4. *If* $h \asymp n^{-1/5}$ *and conditions* (C1) *and* (C4) *hold, then*

(i) $\|\Sigma^{1/2}(\boldsymbol{\theta}_0)(\mathbf{S}_h - \mathbf{P}(\boldsymbol{\theta}_0))^T(\mathbf{I} - \mathbf{S}_h)\mathbf{m}(\boldsymbol{\theta}_0)\|^2 = nh^4 J_2\{m(\cdot;\boldsymbol{\theta}_0)\} + o(n^{1/10})$, *and*

(ii) $J_2\{m(\cdot;\hat{\boldsymbol{\theta}})\} - J_2\{m(\cdot;\boldsymbol{\theta}_0)\} = o_p(1)$.

Proof. (i) First, using (29), condition (C1), and $\sup_t|F_n(t) - F(t)| = O(n^{-1})$, one can show that $\mathbf{P}^T(\boldsymbol{\theta}_0)(\mathbf{I} - \mathbf{S}_h)\mathbf{m}(\boldsymbol{\theta}_0) = -(h^2\mu_2/2)\Sigma^{-1}(\boldsymbol{\theta}_0)$ $\mathbf{A}(\boldsymbol{\theta}_0)\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}_0)n^{-1}\mathbf{A}^T(\boldsymbol{\theta}_0)\mathbf{m}''(\boldsymbol{\theta}_0) + o(h^2)$, where $\mathbf{m}''(\boldsymbol{\theta}_0) = (m''(t_{1n};\boldsymbol{\theta}_0),$ $\ldots, m''(t_{nn};\boldsymbol{\theta}_0))^T$. By using (23), (C4), and $\sup_t|F_n(t) - F(t)| = O(n^{-1})$, it can be shown that the ith element of $\mathbf{S}_h^T\mathbf{m}''(\boldsymbol{\theta}_0)$ is $m''(t_{in};\boldsymbol{\theta}_0) + O(h^{\zeta})$ if $t_{in} \notin B_h^*$ or $O(1)$ if $t_{in} \in B_h^*$. Hence, the ith element of $\mathbf{S}_h^T(\mathbf{I} - \mathbf{S}_h)\mathbf{m}(\boldsymbol{\theta}_0)$ is $-h^2\mu_2 m''(t_{in};\boldsymbol{\theta}_0)/2 + o(h^2)$ if $t_{in} \notin B_h^*$ or $O(h^2)$ otherwise. Finally, the proof can be finished by using $\sup_t|F_n(t) - F(t)| = O(n^{-1})$, (C4), and $n^{-1}\mathbf{A}^T(\boldsymbol{\theta}_0)\mathbf{m}''(\boldsymbol{\theta}_0) = \boldsymbol{\Psi}(\boldsymbol{\theta}_0) + O(n^{-1})$.

(ii) The proof is immediately finished because $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p(n^{-1/2})$.

Lemma 5. *Let* $U_{kn} = 2\left(\sum_{j=1}^{k-1} w_{kj;h_{opt}}\varepsilon_{jn} - \ell_{kn}\right)\varepsilon_{kn}/v_n$, *where* ℓ_{kn} *is the kth element of* $\mathcal{L}_n = (\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T(\mathbf{I} - \mathbf{S}_{h_{opt}})\mathbf{m}(\boldsymbol{\theta}_0)$, *and* $v_n^2 =$ $2\mathrm{tr}\{[(\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))^T(\mathbf{S}_{h_{opt}} - \mathbf{P}(\boldsymbol{\theta}_0))\Sigma(\boldsymbol{\theta}_0)]^2\} + 4\|\Sigma^{1/2}(\boldsymbol{\theta}_0)\mathcal{L}_n\|^2$. *Then, under* (C6) *for* $0 < \delta \le 1$ *as* $n \to \infty$,

(i) $\sum_{k=1}^n E|U_{kn}|^{2+2\delta} \to 0$, *and*

(ii) $E\left|\sum_{k=1}^n E(U_{kn}^2 \mid \boldsymbol{\varepsilon}_{k-1}) - 1\right|^{1+\delta} \to 0$,

where $\boldsymbol{\varepsilon}_{k-1} = (\varepsilon_{1n}, \ldots, \varepsilon_{k-1,n})^T$ *for* $k = 1, \ldots, n$.

Proof. We take $\delta = 1$ to establish this lemma.

(i) First note that $v_n^2 \asymp n^{1/5}$ because of Lemmas 1 and 4. We can express U_{kn} as $U_{kn} = Q_{kn}\varepsilon_{kn} + L_{kn}\varepsilon_{kn}$, where $Q_{kn} = 2\sum_{j=1}^{k-1} w_{kj;h_{opt}}\varepsilon_{jn}/v_n$, and $L_{kn} = -2\ell_{kn}/v_n$, then using $w_{kj;\,h_{opt}} = O(1/nh_{opt})$, which was given in Lemma 1, the moment condition (C6), and $\ell_{kn} = O(h_{opt}^2)$, it can be seen that $\sum_{k=1}^{n} E[(Q_{kn}\varepsilon_{kn})^4] = o(1)$, and $\sum_{k=1}^{n} E[(L_{kn}\varepsilon_{kn})^4] = o(1)$. Hence, the proof of (i) is complete.

(ii) Using Lemma 1, we can write v_n^2 as $v_n^2 = v_{1n}^2 + v_{2n}^2 + v_{3n}^2 + O(1)$, where $v_{1n}^2 = 2\sum_{k=1}^{n} w_{kk;\,h_{opt}}^2 \sigma^4(t_{kn};\,\boldsymbol{\theta}_0)$, $v_{2n}^2 = 4\sum_{k=2}^{n}\sum_{j=1}^{k-1} w_{kj;\,h_{opt}}^2 \sigma^2(t_{kn};\,\boldsymbol{\theta}_0)\sigma^2(t_{jn};\,\boldsymbol{\theta}_0)$, and $v_{3n}^2 = 4\sum_{k=1}^{n}\ell_{kn}^2\sigma^2(t_{kn};\,\boldsymbol{\theta}_0)$. In addition, $E\left[\sum_{k=1}^{n} E(U_{kn}^2 \mid \boldsymbol{\varepsilon}_{k-1}) - 1\right]^2$ can be expressed as

$$
\begin{aligned}
&E\left[\sum_{k=1}^{n} E(U_{kn}^2 \mid \varepsilon_{k-1}) - 1\right]^2 \\
&= E\left[\sum_{k=1}^{n} (Q_{kn}^2 + L_{kn}^2)\sigma^2(t_{kn};\,\boldsymbol{\theta}_0)\right]^2 \\
&\quad + 4E\left\{\left[\sum_{k=1}^{n} (Q_{kn}^2 + L_{kn}^2)\sigma^2(t_{kn};\,\boldsymbol{\theta}_0)\right]\left[\sum_{i=1}^{n} Q_{in}L_{in}\sigma^2(t_{in};\,\boldsymbol{\theta}_0)\right]\right\} \\
&\quad + 4E\left[\sum_{k=1}^{n} Q_{kn}L_{kn}\sigma^2(t_{kn};\,\boldsymbol{\theta}_0)\right]^2 - 2\sum_{k=1}^{n}\sigma^2(t_{kn};\,\boldsymbol{\theta}_0)E(Q_{kn} + L_{kn})^2 + 1. \qquad (30)
\end{aligned}
$$

Because

$$E\left[\sum_{k=1}^{n} Q_{kn}^2\sigma^2(t_{kn};\,\boldsymbol{\theta}_0)\right]^2 = (v_{1n}^2 + v_{2n}^2)^2/v_n^4 + o(1),$$

$$E\left[\sum_{k=1}^{n} Q_{kn}^2\sigma^2(t_{kn};\,\boldsymbol{\theta}_0)\right] = (v_{1n}^2 + v_{2n}^2)/v_n^2 + o(1),$$

and $\sum_{k=1}^{n} L_{kn}^2\sigma^2(t_{kn};\,\boldsymbol{\theta}_0) = v_{3n}^2/v_n^2$, the first term of (30) is $1 + o(1)$. Using

$\ell_{kn} = O(h_{opt}^2)$ and $w_{kj;\,h_{opt}} = O(1/nh_{opt})$ and knowing that only $O(nh_{opt})$ nonzero $w_{kj;\,h_{opt}}$ exists for any fixed j, we can show that the third term of (30) is $o(1)$. Using the Cauchy-Schwarz inequality and the above result shows that the second term of (30) is also $o(1)$. Finally, the proof can be finished by demonstrating that $\sum_{k=1}^{n} \sigma^2(t_{kn};\, \boldsymbol{\theta}_0) E(Q_{kn} + L_{kn})^2 = 1 + o(1)$.

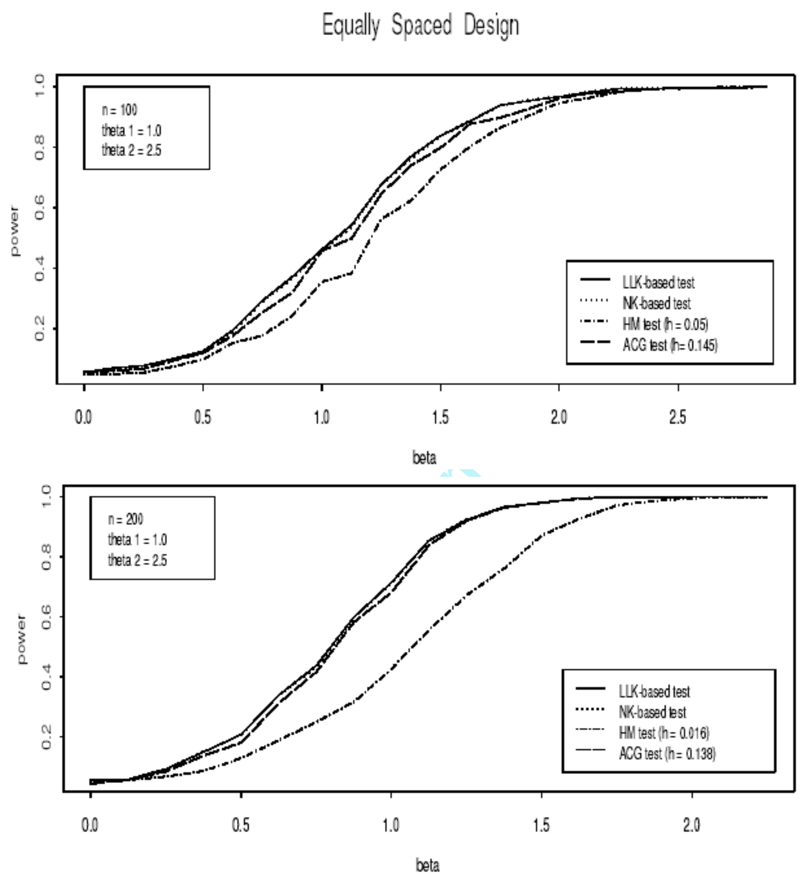

Figure 1. Comparison of the Poisson empirical power function among the local-linear kernel (LLK)-based test, nonparametric boundary-corrected kernel (NK)-based test, Härdle and Mammen test, and Alcalá et al. test for the null model $\log m(t) = \theta_1 + \theta_2 t$ and alternative models $\log m(t) = \theta_1 + \theta_2 t + \beta t^2$ based on the nominal level 0.05.

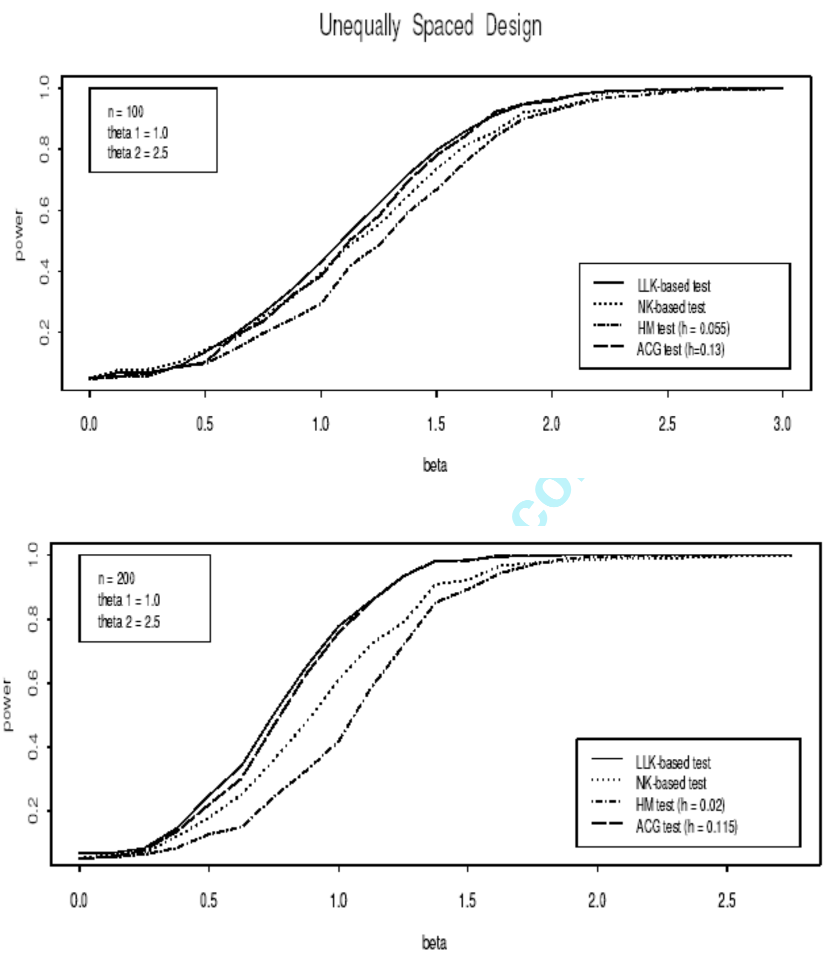

Figure 2. Comparison of the Poisson empirical power function among the local-linear kernel (LLK)-based test, nonparametric boundary-corrected kernel (NK)-based test, Härdle and Mammen test, and Alcalá et al. test for the null model $\log m(t) = \theta_1 + \theta_2 t$ and alternative models $\log m(t) = \theta_1 + \theta_2 t + \beta t^2$ based on the nominal level 0.05.

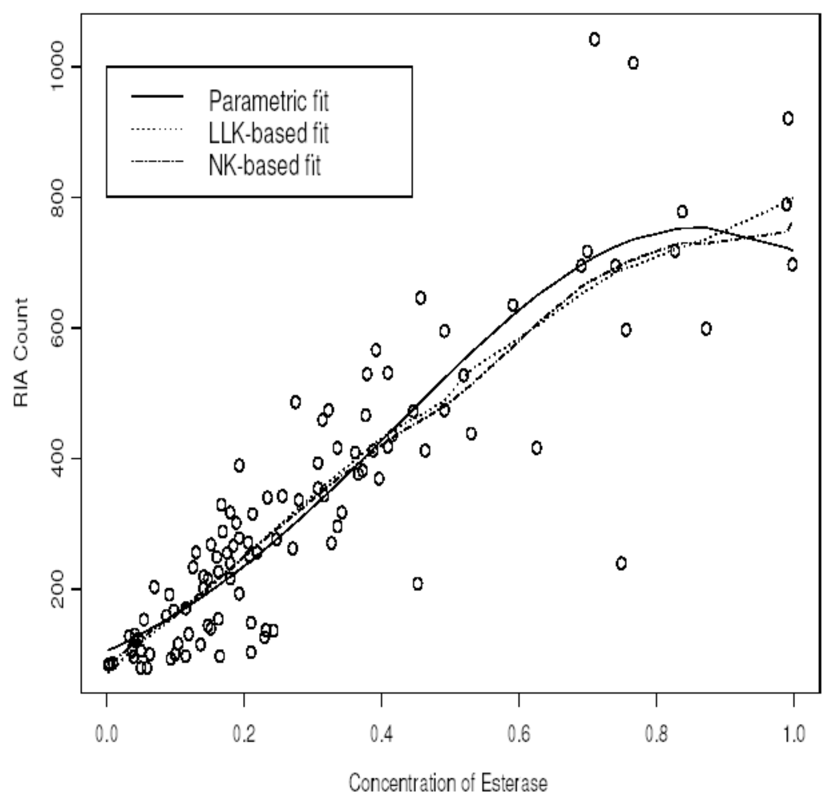

Figure 3. Parametric $(m(t, \hat{\boldsymbol{\theta}}) = \exp(4.65 + 4.57t - 2.65t^2))$, local-linear kernel (LLK)-based fit, and nonparametric boundary-corrected kernel (NK)-based fit for the esterase radioimmunoassay data

Acknowledgements

This research was supported in part by Cancer Center Support Grant CA21765 from the National Institutes of Health and by the American Lebanese Syrian Associated Charities (ALSAC).

References

[1] J. T. Alcalá, J. A. Cristóbal and W. González-Manteiga, Goodness-of-fit test for linear models based on local polynomials, Statist. Probab. Lett. 42 (1999), 39-46.

[2] D. Barry and J. A. Hartigan, An omnibus test for departures from constant mean, Ann. Statist. 18 (1990), 1340-1357.

[3] A. W. Bowman and A. Azzalini, Applied Smoothing Techniques for Data Analysis: The Kernel Approach with *S*-plus Illustrations, Oxford University Press, Oxford, 1997.

[4] R. J. Carroll and D. Ruppert, Transformation and Weighting in Regression, Chapman & Hall, New York, 1988.

[5] D. Cox, E. Koh, G. Wahba and B. Yandell, Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models, Ann. Statist. 16 (1988), 113-119.

[6] A. C. Davison and D. V. Hinkley, Bootstrap Methods and their Application, Cambridge University Press, Cambridge, 1997.

[7] H. Dette, A consistent test for the functional form of a regression based on a difference of variance estimators, Ann. Statist. 27 (1999), 1012-1040.

[8] R. L. Eubank and J. D. Hart, Testing goodness-of-fit in regression via order selection criteria, Ann. Statist. 20 (1992), 1412-1425.

[9] J. Fan and I. Gijbels, Local Polynomial Modelling and its Applications, Chapman & Hall, New York, 1996.

[10] J. Fan, G. Zhang and J. Zhang, Generalized likelihood ratio statistics and Wilks phenomenon, Ann. Statist. 29 (2001), 153-193.

[11] E. B. Fowlkes, Some diagnostics for binary logistic regression via smoothing, Biometrika 74 (1987), 503-515.

[12] W. Härdle and M. Mammen, Comparing nonparametric versus parametric regression fits, Ann. Statist. 21 (1993), 1926-1947.

[13] W. Härdle, P. Hall and J. S. Marron, How far are automatically chosen regression smoothing parameters from their optimum? (with discussion), J. Amer. Statist. Assoc. 83 (1988), 86-99.

[14] J. D. Hart, Nonparametric Smoothing and Lack-of-fit Tests, Springer-Verlag, New York, 1997.

[15] C. C. Heyde and B. M. Brown, On the departure from normality of a certain class of martingales, Ann. Math. Statist. 41 (1970), 2161-2165.

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

[17] S. le Cessie and J. C. van Houwelingen, A goodness-of-fit test for binary regression models based on smoothing methods, Biometrics 47 (1991), 1267-1282.

[18] C. S. Li, Testing lack of fit of regression models under heteroscedasticity, Canad. J. Statist. 27(3) (1999), 485-496.

[19] R. Liu, Bootstrap procedures under some non-i.i.d. models, Ann. Statist. 16 (1988), 1696-1708.

[20] P. McCullagh and J. A. Nelder, Generalized Linear Models, 2nd ed., Chapman & Hall, New York, 1989.

[21] H. G. Müller, Smooth optimum kernel estimators near endpoints, Biometrika 78 (1991), 521-530.

[22] H. G. Müller, Goodness-of-fit diagnostics for regression models, Scand. J. Statist. 19 (1992), 157-172.

[23] G. A. F. Seber and C. J. Wild, Nonlinear Regression, Wiley, New York, 1988.

[24] W. Stute, Nonparametric model checks for regression, Ann. Statist. 25 (1997), 613-641.

[25] P. Whittle, Bounds for the moments of linear and quadratic forms in dependent variables, Theory Probab. Appl. 5 (1960), 302-305.

[26] C. F. J. Wu, Jackknife, bootstrap and other resampling methods in regression analysis (with discussion), Ann. Statist. 14 (1986), 1261-1350.

[27] J. X. Zheng, A consistent test of functional form via nonparametric estimation techniques, J. Econometrics 75 (1996), 263-289.