International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 7, 2008, pp 13-29

A Diagnostic Test for Parallelism

R. L. Eubank
Department of Mathematics and Statistics
Arizona State University
Tempe, AZ 85287, USA

Chin-Shang Li
Department of Public Health Sciences
University of California
Davis, California 95616, USA

[Received January 20, 2005 ; Accepted July 16, 2007]

Abstract

A new diagnostic test is developed for detecting departures from parallel mean functions in analysis of covariance and related settings. The new test is derived from an elementary regression perspective using nonparametric smoothing methodology. Its large sample null distribution is shown to be standard normal under certain mild conditions and its finite sample power properties are investigated via a small scale simulation experiment. Within the context of this simulation study, the new test is found to have substantially better empirical power than the classical linear models test when the covariate enters the model nonlinearly and has competitive or better empirical power than another nonparametric smoothing based test that has been proposed by Young and Bowman. The test is applied to a data set concerning cancer and pregnancy where we show that the classical linear models approach apparently leads to a false rejection of the parallel means hypothesis.

Keywords and Phrases: Analysis of covariance, asymptotic distribution, empirical power, smoothing splines.

AMS Classification: 62G08, 62J10.

1 Introduction

A frequently arising issue in the analysis of data concerns the assessment of treatment (or other parametric) effects in the presence of a covariate. Under the assumption of an additive model, treatment effects can be most easily understood when the covariate influences each treatment group in the same way resulting in group means that are parallel to one another. If the covariate's influence can be modeled parametrically, tests for parallelism can be obtained via standard linear or nonlinear models methodology. However, in many cases a suitable parametric form may not be known for the covariate's effect which makes it necessary to treat this aspect of the model nonparametrically. Tests derived from this latter perspective have been proposed by Young and Bowman (1995), for example. In this paper, we use an elementary regression approach to develop a new diagnostic test for parallelism that can be employed when the covariate term in the model cannot be handled parametrically.

The data in Figure 1 provides an illustration of a situation where our methodology appears to be useful. The response in this case is the logarithm of alpha-fetal protein measured in 858 different women on various days of their particular pregnancies. The subjects were monitored for a period of time subsequent to delivery over which incidences of cancer were recorded. The question here is whether there is a cancer effect which is easy to assess if the two mean curves (for the cancer and non-cancer populations) that relate the level of alpha-fetal protein to day of pregnancy are parallel. We will analyze this data in detail in Section 4.

In the next section we derive our test for parallelism and discuss its large sample properties. This is accompanied by a small simulation in Section 3 to explore its finite sample power and level behavior. Section 4 then contains an analysis of the alpha-fetal protein data using our methodology. Technical results and proofs are collected in the Appendix.

2 A test for parallelism

Assume now that we have Y and Z responses of the form

$$y_i = \alpha_1 + f_1(t_i) + \varepsilon_i, \quad i = 1, \ldots, n_1, \tag{1}$$

and

$$z_i = \alpha_2 + f_2(x_i) + \eta_i, i = 1, \ldots, n_2. \tag{2}$$

Here α_1 and α_2 represent unknown parameters of interest while $\varepsilon_1, \ldots, \varepsilon_{n_1}$, $\eta_1, \ldots, \eta_{n_2}$ are presumed to be iid random variables with mean zero and unknown variance σ^2. The t_i and x_i are taken to be nonrandom with $0 \leq t_1 \leq \cdots \leq t_{n_1} \leq 1$, $0 \leq x_1 \leq \cdots \leq x_{n_2} \leq 1$. They can be viewed as values which are chosen at possibly different points for some common covariate X. The functions f_1 and f_2 are unknown apart from satisfying

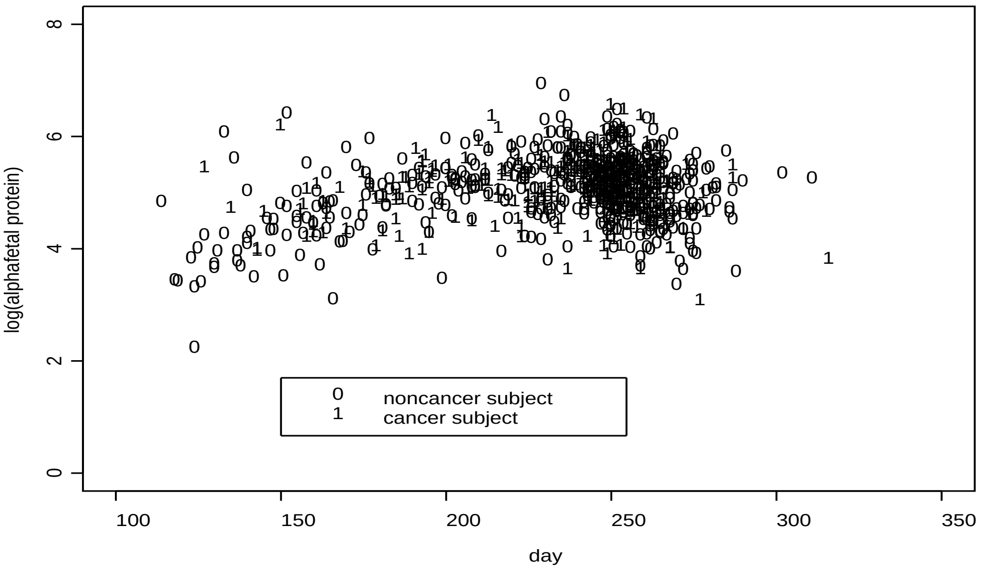

Figure 5: Alpha-fetal protein data

some type of identifiability condition such as $\sum_{i=1}^{n_1} f_1(t_i) = \sum_{i=1}^{n_2} f_2(x_i) = 0$. Under this formulation the two mean functions in (1)-(2) are parallel if the null hypothesis

$$H_0 : f_1(t) = f_2(t) \text{ for all } t \in [0, 1]$$

is true. We will now derive a test for H_0.

The motivation for our test stems from the following simple regression perspective. If we actually knew f_1 and f_2, we could fit models of the form

$$\mathrm{E}y_i = \alpha_1 + \beta_1 f_2(t_i), i = 1, \ldots, n_1,$$

and

$$\mathrm{E}z_i = \alpha_2 + \beta_2 f_1(x_i), i = 1, \ldots, n_2,$$

to the Y and Z data and then test H_0 by testing that $\beta_1 = \beta_2 = 1$. The standard linear models test statistic for such a hypothesis would have the form

$$\widetilde{T} = \left(\tilde{b}_1 - \tilde{b}_2\right) / \sigma \left[\left(n_1(||f_2||)^2_{1n_1}\right)^{-1} + \left(n_2||f_1||^2_{2n_2}\right)^{-1}\right]^{1/2} \tag{3}$$

with

$$||f_2||^2_{1n_1} = n_1^{-1} \sum_{i=1}^{n_1} f_2^2(t_i), \quad ||f_1||^2_{2n_2} = n_2^{-1} \sum_{i=1}^{n_2} f_1^2(x_i) \tag{4}$$

and

$$\tilde{b}_1 = \frac{\sum_{i=1}^{n_1}(y_i - \bar{y})f_2(t_i)}{n_1||f_2||^2_{1n_1}}, \tilde{b}_2 = \frac{\sum_{i=1}^{n_2}(z_i - \bar{z})f_1(x_i)}{n_2||f_1||^2_{2n_2}} \tag{5}$$

for $\bar{y}$ and $\bar{z}$ the two response averages. Of course f_1 and f_2 are unknown which makes (3) uncomputable in practice. Our solution to this problem is to replace f_1 and f_2 by nonparametric estimators in (3)-(5).

The basic idea is to use nonparametric smoothers applied to the Y and Z data separately to estimate the unknown functions f_1 and f_2. Thus, one obtains a nonparametric estimator $\hat{f}_1$ of f_1 by fitting $y_i - \bar{y}$, $i = 1, \ldots, n_1$, and a nonparametric estimator $\hat{f}_2$ of f_2 from $z_i - \bar{z}$, $i = 1, \ldots, n_2$. In our simulations and data analysis we have used cubic smoothing splines to obtain $\hat{f}_1$ and $\hat{f}_2$ which automatically produces estimators with $\sum_{i=1}^{n_1} \hat{f}_1(t_i) = \sum_{i=1}^{n_2} \hat{f}_2(x_i) = 0$ analogous to the identifiability conditions for f_1 and f_2. However, the choice of smoother is unlikely to be crucial and similar empirical results should be obtainable from kernel or series type smoothers, for example.

Given nonparametric estimators $\hat{f}_1$ and $\hat{f}_2$, the test statistics we now propose for H_0 is

$$T = \left(\hat{b}_1 - \hat{b}_2 - \widehat{A}\right) / \hat{\sigma} \left[\left(n_1||\hat{f}_2||^2_{1n_1}\right)^{-1} + \left(n_2||\hat{f}_1||^2_{2n_2}\right)^{-1}\right]^{1/2} \tag{6}$$

with $||\hat{f}_2||^2_{1n_1}$, $||\hat{f}_1||^2_{2n_2}$ defined as in (4),

$$\hat{b}_1 = \frac{\sum_{i=1}^{n_1}(y_i - \bar{y})\hat{f}_2(t_i)}{n_1||\hat{f}_2||^2_{1n_1}}, \ \hat{b}_2 = \frac{\sum_{i=1}^{n_2}(z_i - \bar{z})\hat{f}_1(x_i)}{n_2||\hat{f}_1||^2_{2n_2}} \tag{7}$$

and

$$\widehat{A} = \frac{1}{2}\left(\frac{\sum_{i=1}^{n_1} \hat{f}_1(t_i)\hat{f}_2(t_i)}{n_1||\hat{f}_2||^2_{1n_1}} - \frac{\sum_{i=1}^{n_2} \hat{f}_1(x_i)\hat{f}_2(x_i)}{n_2||\hat{f}_1||^2_{2n_2}}\right). \tag{8}$$

The additional term (i.e., $\hat{A}$) in (6) is there to remove a bias effect arising from estimation of f_1 and f_2. The standard deviation estimator $\hat{\sigma}$ in (6) can be any $\sqrt{n}$-constraint estimator of σ^2. For example, one can use the difference based estimator of Gasser, et al. (1986) to obtain estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ from the Y and Z data, respectively, that can be pooled as in Young and Bowman (1995) to obtain $\hat{\sigma}^2 = [(n_1 - 2)\hat{\sigma}_1^2 + (n_2 - 2)\hat{\sigma}_2^2]/(n_1 + n_2 - 4)$. This is the estimator that was used in our simulations and data analysis.

In the Appendix we give conditions under which T can be expected to have an approximate standard normal distribution when H_0 is true and $\min(n_1, n_2)$ is large. Thus, when these conditions hold one can consider rejecting H_0 at level α if $|T|$ exceeds the $100(1 - \alpha/2)$ percentage point of the standard normal distribution.

Roughly speaking, one can expect a normal approximation to be effective for large samples where the two sets of design points have a large amount of overlap (i.e., common or nearly common points as described in the Appendix) and neither f_1 nor f_2 are the zero function. The need for some shared design space can be illustrated by noting that in the null case where $f_1 = f_2 = f$ we cannot effectively estimate f over $\{t_1, \ldots, t_{n_1}\}$ using the Z data if $x_1 > t_{n_1}$, for example. The condition that f_1, f_2 cannot vanish identically means that one must model carefully and not include a spurious covariate that has no effect on the responses.

The power of the test obtained from T in (6) can be expected to grow, asymptotically, as a function of $\sqrt{n}C_n$ with

$$C_n = \frac{\sum_{i=1}^{n_1} f_1(t_i) f_2(t_i)}{\sum_{i=1}^{n_1} f_2^2(t_i)} - \frac{\sum_{i=1}^{n_2} f_1(x_i) f_2(x_i)}{\sum_{i=1}^{n_2} f_1^2(x_i)}.$$

This is essentially a comparison of "regression" coefficients for the regression of f_1 on f_2 and of f_2 on f_1 under the two designs. In the simplest case where $n_1 = n_2 = n$ and $t_i = x_i = i/n$, $i = 1, \ldots, n$, we have

$$C_n \sim \frac{\int_0^1 f_1(t) f_2(t) dt}{\int_0^1 f_1^2(t) dt} - \frac{\int_0^1 f_1(t) f_2(t) dt}{\int_0^1 f_2^2(t) dt}.$$

Thus, $C_n \doteq 0$ if $\int_0^1 f_1^2(t)dt = \int_0^1 f_2^2(t)dt$ so that the test cannot detect alternatives such as $f_1(t) = \cos \pi t$, $f_2(t) = \cos 2\pi t$ which differ but have the same $L_2[0, 1]$ norm. More generally, T will have trivial asymptotic power against any alternative for which $C_n \to 0$. Since the chances of a real data scenario having $C_n = 0$ seem somewhat remote, we expect T to provide a useful diagnostic tool for most situations that could arise in practice.

3 Empirical results

In this section we report the results of simulation experiments that were conducted to assess the performance of the test for parallelism discussed in the previous section. All of our experiments employed data generated from model (1)-(2) using normal random errors. Various choices were then considered for f_1 f_2, n_1, n_2 and the design points. All the tests we considered are invariant with respect to additive constants so the values of α_1 and α_2 have no influence on our results.

For comparison purposes we considered two other possible test statistics including the classical linear models statistic

$$L = \frac{\left[\sum_{i=1}^{n_1} (y_i - \bar{y}) t_i \left(\sum_{i=1}^{n_1} t_i^2\right)^{-1} - \sum_{i=1}^{n_2} (z_i - \bar{z}) x_i \left(\sum_{i=1}^{n_2} x_i^2\right)^{-1}\right]}{\hat{\sigma}^2 \left[\left(\sum_{i=1}^{n_1} t_i^2\right)^{-1} + \left(\sum_{i=1}^{n_2} x_i^2\right)^{-1}\right]^{1/2}} \tag{9}$$

which arises from least-squares fitting under the assumption that f_1 and f_2 are linear in (1)-(2). We choose $\hat{\sigma}^2$ in (9) to be the pooled Gasser, et al. (1986) estimator discussed in the previous section rather than the more standard residual mean squared error. This is because the latter choice can have a large positive bias when f_1 and f_2 are not linear which can adversely effect both the power and level of the test. This was demonstrated in the simulations of Young and Bowman (1995), for example.

The other test we considered was a version of the Young and Bowman (1995) test. Specifically, we used a statistic of the form

$$B = \left[\sum_{i=1}^{n_1} \left(\tilde{f}_1(t_i) - \tilde{f}(t_i) \right)^2 + \sum_{i=1}^{n_1} \left(\tilde{f}_2(x_i) - \tilde{f}(x_i) \right) \right] / \widehat{\sigma}^2 \tag{10}$$

with $\tilde{f}_1$, $\tilde{f}_2$ and $\tilde{f}$ obtained from smoothing the mean corrected Y, Z and combined (i.e., pooled Y and Z) data, respectively.

All our data smoothing was accomplished using cubic smoothing splines. Given pairs of points (u_i, v_i), $i = 1, \ldots, r$, with $0 \leq u_1 \leq \cdots \leq u_r \leq 1$, a cubic smoothing spline fit to the v_i is defined to be the minimizer of

$$\sum_{i=1}^{r} (v_i - f(u_i))^2 + \lambda \int_0^1 (f''(t))^2 dt, \lambda > 0, \tag{11}$$

over all functions with two absolutely continuous derivatives. The resulting estimator is a natural cubic spline with knots at $u_1, \ldots, u_r$ that can be efficiently computed in $O(r)$ operations. More detailed discussions can be found in Wahba (1990) and Eubank (1999). The parameter λ in (11) governs the smoothness of the fit. Its choice will be discussed subsequently.

To obtain the estimators for f_1 and f_2 in (10) we used smoothing spline fits $\tilde{f}_1$ and $\tilde{f}_2$ to the data $(t_i, y_i - \bar{y})$, $i = 1, \ldots, n_1$, and $(x_i, z_i - \bar{z})$, $i = 1, \ldots, n_2$, respectively. In obtaining these fits we used $\lambda = [2R/n]^4$ with R being the range of the design and n the number of observations (i.e., either n_1 or n_2). The pooled data fit $\tilde{f}$ in (10) was obtained similarly using the combined $(t_i, y_i - y)$ and $(x_i, z_i - \bar{z})$ data. The particular choice of λ used here derives from the bandwidth employed by Young and Bowman (1995) and the relationship between the spline smoothing parameter λ and the bandwidth for its "equivalent" kernel estimator. See, e.g. Eubank (1999, Chapter 5)

In the case of our proposed test (3) the estimators $\hat{f}_1$ and $\hat{f}_2$ were also obtained by cubic spline smoothing. However, data driven choices were used for λ that minimize an unbiased risk or Mallow's type criterion. This was done for the mean corrected Y and Z data separately using Gasser, et al. (1986) variance estimators for the variance term that arises in the smoothing parameter selection criterion. See, e.e., Eubank (1999) for further discussion of smoothing parameter selection.

All our tests were conducted at a nominal 5% level. Critical values for T in (6) and L in (9) were obtained from the standard normal distribution. Critical values for the Young/Bowman type statistic (10) were obtained by simulating from its null distribution in each of the various experimental settings described below.

Our power study involved three basic scenarios. The first two of these are motivated by empirical work reported in Young and Bowman (1995). In one case we took

$$f_1(t) = t, f_2(t) = \beta f_1(t) \tag{12}$$

with $t_i = i/n_1$, $i = 1, \ldots, n_1$, $x_i = i/n_2$, $i = 1, \ldots, n_2$, and examined this over an (increasing) grid of β's starting at the null model case with $\beta = 1$. For σ we used .05 and .1 and generated 500 replicate samples of size $n_1 = n_2 = 30$ or $n_1 = 30$, $n_2 = 50$, for each value of β. A similar type of simulation was conducted using

$$f_1(t) = t^2 - t + 1/6, \quad f_2(t) = \beta f_1(t) \tag{13}$$

over the same type of grid for β, starting at $\beta = 1$, with the same choices of sample sizes, designs and variances as under (12).

Our third simulation experiment was motivated by the alpha-fetal protein data discussed in the introduction. Here we used the distinct covariate (or time) values from the non-cancer and cancer groups (rescaled onto [0,1]) to obtain the t_i and x_i values. This resulted in $n_1 = 141$, $n_2 = 103$ and we again generated 500 replicate pairs of samples at each experimental setting. We also took σ to be .5 because this is the pooled Gasser, et al. standard deviation estimator from the alpha-fetal protein data. The choices of f_1 and f_2 were obtained by fitting cubic polynomials f_1^* and f_2^* to the non-cancer and cancer groups respectively, and then translating these two functions onto [0,1]. Plots of typical Y and Z data sets obtained in this fashion along with f_1^* and f_2^* are shown in Figures 2–3. The choices we used for f_1 and f_2 in model (1) were then

$$f_1 = f_1^*, \quad f_2 = (1 - \beta) f_1^* + \beta f_2^* \tag{14}$$

with $\beta = j/9$, $j = 0, \ldots, 9$. Note that $\beta = 0$ gives the null model case and that (14) differs from (12)-(13) in that f_2 is no longer a constant multiple of f_1 when H_0 is false.

Figures 4–8 show some representative empirical power curves corresponding to some of the cases treated in our simulation. Figures 4–5 give results for the linear case (12) with $\sigma = .1$ and $n_1 = n_2 = 30$ or $n_1 = 30$, $n_2 = 50$. The power for all three tests (T, L and B) are quite similar in this situation and the tests all attain their nominal level when $\beta = 1$. Things change for the quadratic case (13). The quadratic and linear functions are essentially orthogonal here so that the power of the L test reverts to its level asymptotically. This is seen to be true in finite samples as well from Figures 6–7 which give simulation results for the quadratic case with $\sigma = .05$ and $n_1 = n_2 = 30$ or $n_1 = 30, n_2 = 50$. Our proposed test performs well in this case and has substantially better power than the Young/Bowman type test. This is as one would hope since the alternatives being considered are of the precise form our test was designed to detect.

Simulation results for the cubic case are shown in Figure 8. None of the tests perform particularly well in this setting and, in particular, the L test has level difficulties. This problem occurs because of the difference between the number and location of the covariate values for the Y and Z data and because the null regression function

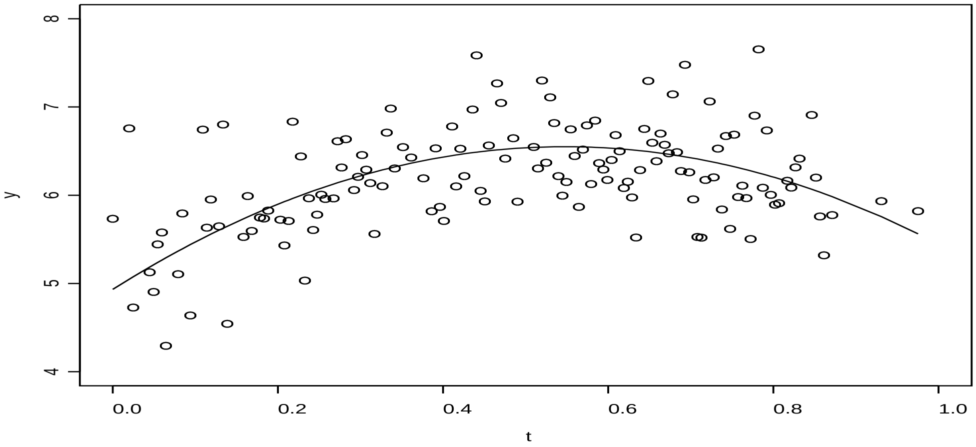

Figure 6: Example Y data for cubic case

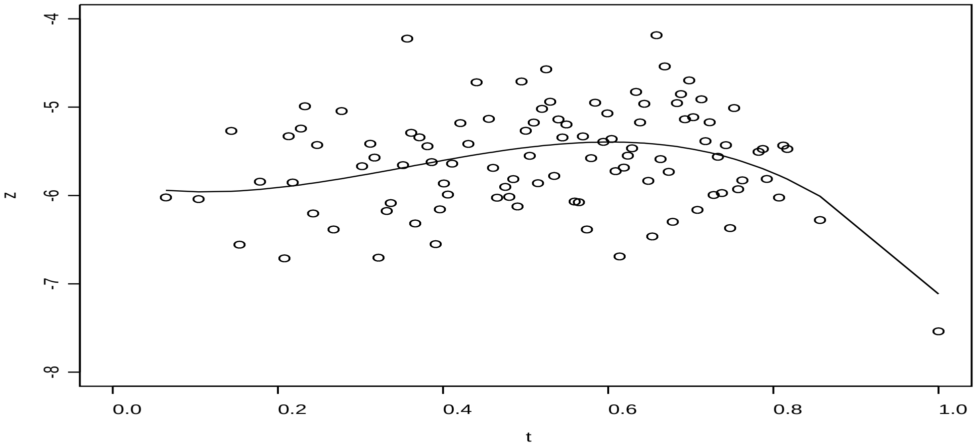

Figure 7: Example Z data for cubic case

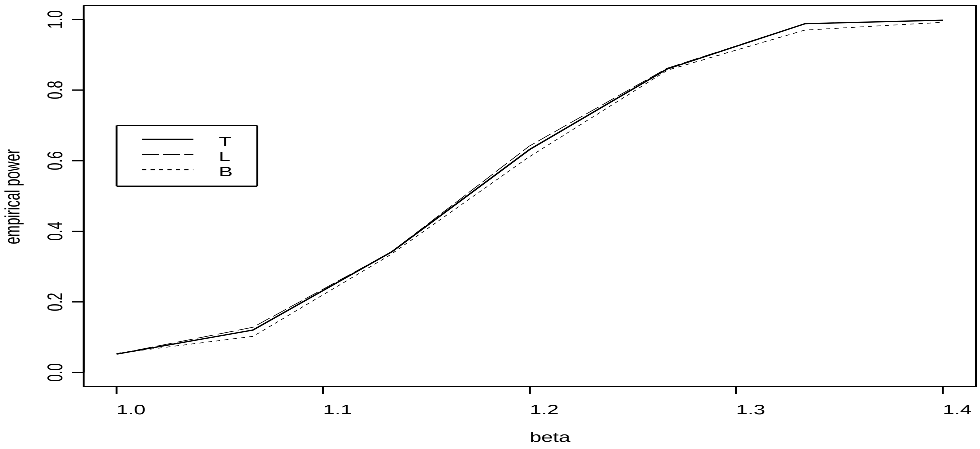

Figure 8: Linear case: $n_1 = n_2 = 30, \sigma = .1$

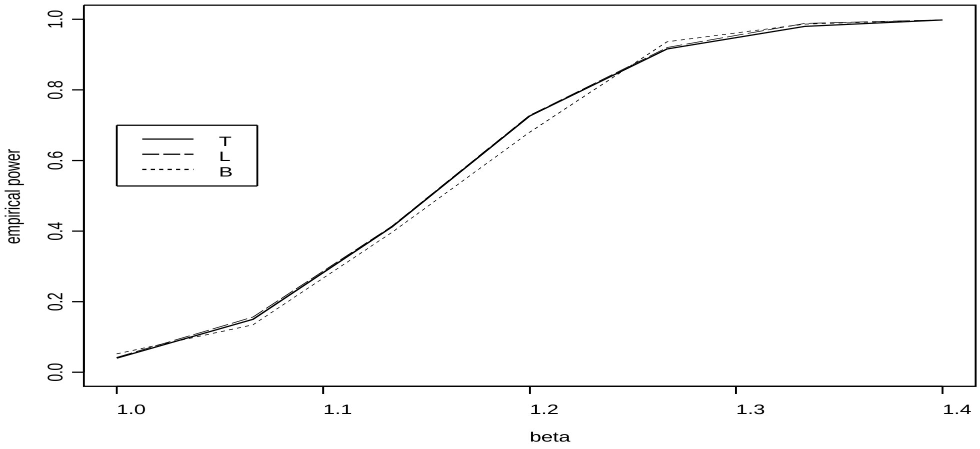

Figure 9: Linear case: $n_1 = 30, n_2 = 50, \sigma = .1$

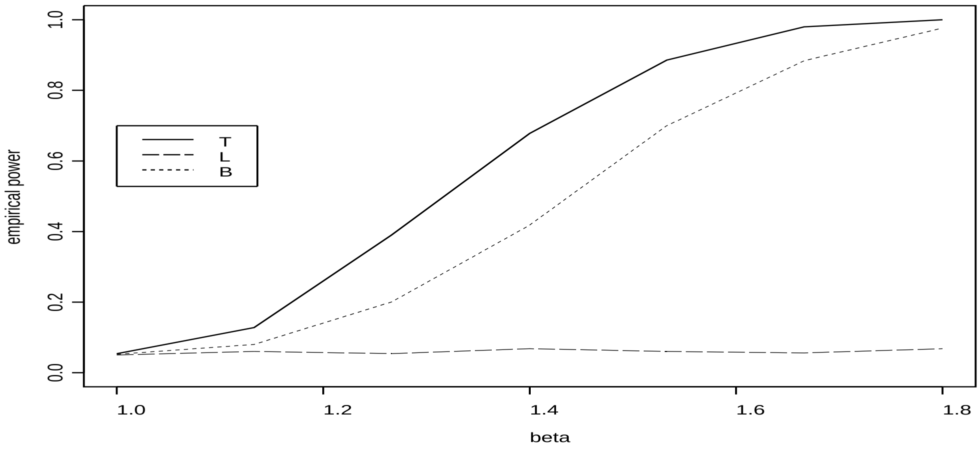

Figure 10: Quadratic case: $n_1 = n_2 = 30, \sigma = .05$

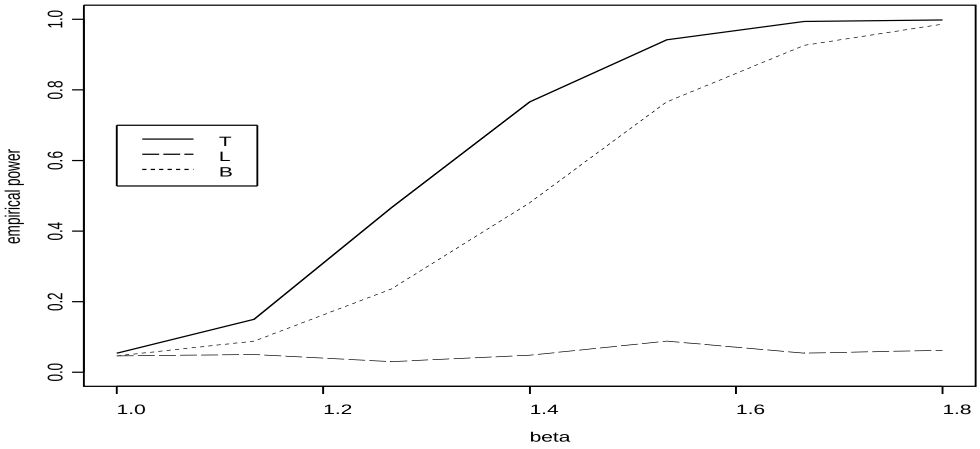

Figure 11: Quadratic case: $n_1 = 30, n_2 = 50, \sigma = .05$

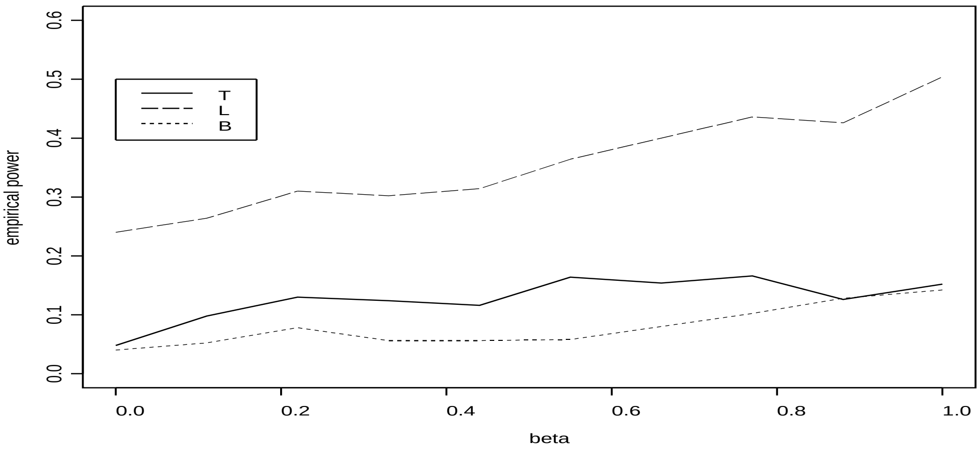

Figure 12: Cubic case: $n_1 = 141, n_2 = 103, \sigma = .5$

is not linear. The presence of nonlinearity entails that the linear fits are only giving approximations to the true regression curve while the design disparity causes the resulting approximating slopes to differ. Consequently, the linear models hypothesis of equal slopes is false even when the two mean functions are identical and, accordingly, the L test fails to provide an effective tests for parallelism in this setting. This simulation also illustrates that our proposed test can perform at least as well as the Young/Bowman test in a case where f_2 is not a constant multiple of f_1.

4 Example

We now return to the alpha-fetal protein data discussed in the introduction and presume that this data can be analyzed using a model of the form (1)-(2). In this context the y_i and z_i represent the log-alpha-fetal protein levels for non-cancer and cancer subjects, respectively, while the t_i and x_i are the corresponding values for the covariate day of pregnancy.

A standard linear models analysis of covariance treatment for this data would involve fitting model (1)-(2) with $f_1(t) = \beta_1 t$ and $f_2(x) = \beta_2 x$. The resulting test statistic for the hypothesis that $\beta_1 = \beta_2$ has the value $L = 2.4$ which might lead one to conclude that the two population mean functions are not parallel. However, in light of the simulation results from the previous section, the large magnitude for L is likely to be a reflection of mean function nonlinearity rather than an indication of departure from paralellism.

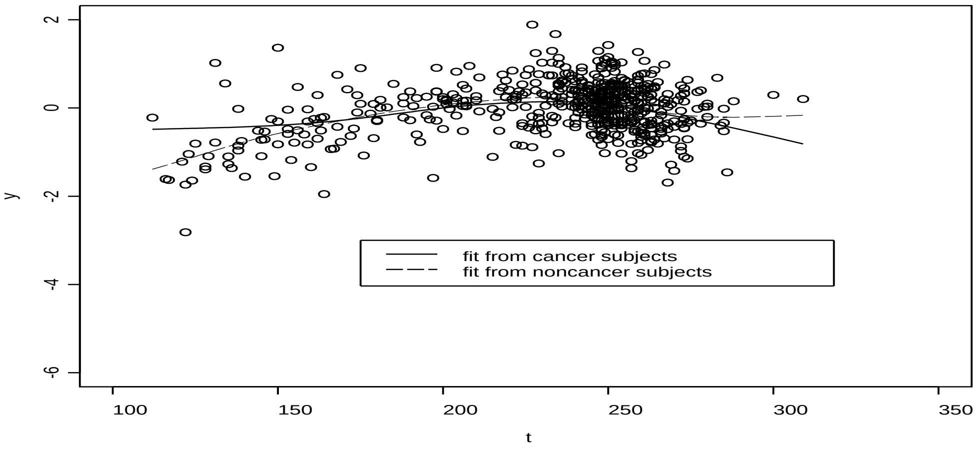

Figure 13: Fits to noncancer subjects

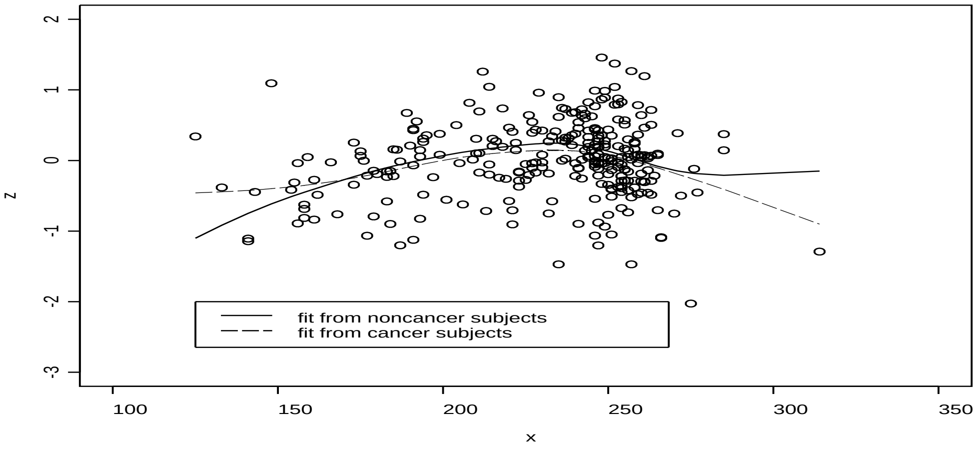

Figure 14: Fits to cancer subjects

Figures 9-10 show smoothing spline fits to the non-cancer and cancer subjects in the alpha-fetal protein data. Figure 9 is a plot of the mean corrected log-alpha-fetal protein values for the non-cancer subjects upon which the fits from the non-cancer and cancer subjects (i.e., $\hat{f}_1$ and $\hat{f}_2$, respectively) have been superimposed. Figure 10 is a similar plot for the cancer subjects. The regression coefficient for the regression of $\hat{f}_2$ on the mean corrected non-cancer subject values is $\tilde{b}_1 = 1.4$ while the regression of $\hat{f}_1$ on the cancer subjects gives $\tilde{b}_2 = .75$. The resulting value of our test statistic T is 1.8 and is not significant at the 5% level. Thus, in terms of our test there is only a mild suggestion that the underlying mean functions are not of a parallel nature.

Appendix

In this section we give technical details and proofs concerning the large sample properties of our proposed test for parallelism. As in Section 1 we have responses

$$y_i = \alpha_1 + f_1(t_i) + \varepsilon_i, i = 1, \dots, n_1, \tag{15}$$

and

$$z_i = \alpha_2 + f_2(x_i) + \eta_i, i = 1, \dots, n_2, \tag{16}$$

with α_1, α_2 unknown constants, f_1 and f_2 unknown functions, $0 \le t_1 \le \cdots \le t_{n_1} \le 1$ and $0 \le x_1 \le \cdots \le x_{n_2} \le 1$ nonstochastic design points and $\varepsilon_1, \dots, \varepsilon_{n_1}$, $\eta_1, \dots, \eta_{n_2}$ independent and identically distributed random variables with zero mean and common variance $\sigma^2 = 1$. Given estimators $\hat{f}_1$ and $\hat{f}_2$ of f_1 and f_2 obtained from the $y_i, i = 1, \dots, n_1$ and $z_i, i = 1, \dots, n_2$, respectively, our test statistic is derived from

$$T = \frac{\left(\frac{\sum_{i=1}^{n_1} (y_i - \bar{y} - \bar{f}(t_i)) \hat{f}_2(t_i)}{n_1 ||\hat{f}_2||_{1n_1}^2} - \frac{\sum_{i=1}^{n_2} (z_i - \bar{z} - \bar{f}(x_i)) \hat{f}_1(x_i)}{n_2 ||\hat{f}_1||_{2n_2}^2} \right)}{\left[\left(n_1 ||\hat{f}_2||_{n_1}^2 \right)^{-1} + \left(n_2 ||\hat{f}_1||_{2n_2}^2 \right)^{-1} \right]^{1/2}} \tag{17}$$

where $\bar{f} = \frac{1}{2}\left(\hat{f}_1 + \hat{f}_2\right)$,

$$||\hat{f}_2||_{1n_1}^2 = n_1^{-1} \sum_{i=1}^{n_1} \hat{f}_2^2(t_i) \tag{18}$$

and

$$||\hat{f}_1||_{2n_2}^2 = n_2^{-1} \sum_{i=1}^{n_2} \hat{f}_1^2(x_i) \tag{19}$$

for $\bar{y} = \sum_{i=1}^{n_1} y_i / n_1$ and $\bar{z} = \sum_{i=1}^{n_2} z_i / n_2$ the response averages.

In Theorem 1 below we state that T in (17) has a limiting standard normal distribution when $f_1 = f_2$ in (15)–(16). This result holds under certain technical conditions that we now list and discuss.

C1: There are positive constants $||f_1||_2$ and $||f_2||_1$ such that

$$\textstyle\sum_{i=1}^{n_1} f_2^2(t_i) = n_1||f_2||_1^2 + o(n_1) \tag{20}$$
$$\textstyle\sum_{i=1}^{n_2} f_1^2(x_i) = n_2||f_1||_2^2 + o(n_2) \tag{21}$$

and

$$\sum_{i=1}^{n_1} f_1(t_i) = O(1) = \sum_{i=1}^{n_2} f_2(x_i). \tag{22}$$

C2: The estimators $\hat{f}_1$ and $\hat{f}_2$ satisfy

$$\sum_{i=1}^{n_1} \hat{f}_2(t_i) = o_p(\min(n_1, n_2)) = \sum_{i=1}^{n_2} \hat{f}_1(x_i), \tag{23}$$

$$\sum_{i=1}^{n_1} (\hat{f}_2(t_i) - f_2(t_i))^2 = o_p\left(\sqrt{\min(n_1, n_2)}\right) = \sum_{i=1}^{n_2} \left(\hat{f}_1(x_i) - f_1(x_i)\right)^2 \tag{24}$$

and

$$\mathrm{E}\sum_{i=1}^{n_1} \left(\hat{f}_2(t_i) - f_2(t_i)\right)^2 = o(\min(n_1, n_2)) = \mathrm{E}\sum_{i=1}^{n_2} \left(\hat{f}_1(x_i) - f_1(x_i)\right)^2. \tag{25}$$

C3: If $f_1 = f_2 \equiv f$, then

$$\begin{aligned} \frac{\sum_{i=1}^{n_1} \left[f(t_i) - \bar{f}(t_i)\right] \hat{f}_2(t_i)}{n_1||\hat{f}_2||_{1n_1}^2} &- \frac{\sum_{i=1}^{n_2} \left[f(x_i) - \bar{f}(x_i)\right] \hat{f}_1(x_i)}{n_2||\hat{f}_1||_{2n_2}^2} \\ &= o_p\left(\min(n_1, n_2)^{-1/2}\right). \end{aligned} \tag{26}$$

Condition (20)–(21) ensures that f_1 and f_2 do not vanish identically over the x and t design spaces. Restriction (22) can be viewed as an identifiability restriction since it essentially means that f_1 and f_2 must integrate to zero over the t and x designs, respectively.

Conditions (24)–(25) are rather mild consistency conditions on the nonparametric smoothers. If, for example, the $\hat{f}_i$ were kernel smoothers with bandwidths b_i, $i = 1, 2$, it is known (Müller 1988, Section 11.2) that under certain restrictions

$$\sup_{u\in[0,1]} |\hat{f}_i(u) - f_i(u)| = O_p\left(\left(\frac{\log n_i}{n_i b_i}\right)^{1/2} + b_i^m\right), i = 1, 2,$$

when f_1, f_2 have $m \geq 1$ continuous derivatives. As a result, the bandwidths need only be chosen so that $(\log n_i/b_i) = o_p(\sqrt{n_i}) = n_i b_i^{2m}$. In particular, an "optimal" global smoothing level of $b_i \propto n^{-1/(2m+1)}$ will satisfy this condition. In the case of deterministic levels of smoothing both (24) and (25) can be replaced by

$$\mathrm{E}\sum_{i=1}^{n_1}(\hat{f}_2(t_i)-f_2(t_i))^2 = o\left(\sqrt{\min(n_1,n_2)}\right) = \mathrm{E}\sum_{i=1}^{n_2}\left(\hat{f}_1(x_i)-f_1(x_i)\right)^2.$$

Condition (26) is a technical restriction that will be discussed later in this section.

With the above preliminaries, our major results can be stated as follows.

Theorem 1. *Assume that conditions C1–C3 hold and that* $min(n_1,n_2)\to\infty$ *with* $n_1/n_2\to\theta\in(0,\infty)$*. Then, if* $f_1=f_2=f$ *for some bounded function* f*,* T *has a limiting standard normal distribution.*

Proof. The proof consists of showing that under the null model where $f_1=f_2=f$, T behaves like

$$\widetilde{T} = \frac{\sum_{i=1}^{n_1}\varepsilon_i f(t_i)/n_1||f||_{1n_1}^2 - \sum_{i=1}^{n_2}\eta_i f(x_i)/n_2||f||_{2n_2}^2}{\sigma\left[\left(n_1||f||_{1n_1}^2\right)^{-1}+\left(n_2||f||_{2n_2}^2\right)^{-1}\right]^{1/2}}$$

with $||f||_{1n_1}^2, ||f||_{2n_2}^2$ defined as in (18)–(19). $\widetilde{T}$ can be seen to have a limiting standard normal distribution using the Lindeberg-Feller Theorem for double arrays (e.g., Serfling 1980, pg 31–32) along with conditions (20)–(22), the iid and finite variance assumption for the ε's and η's, the boundedness of f and the fact that n_1/n_2 is bounded away from 0 and ∞.

We must now show that $T=\widetilde{T}+o_p(1)$. In this regard, observe that

$$\begin{aligned}\sum_{i=1}^{n_1}\left(y_i-\bar{y}-\bar{f}(t_i)\right)\hat{f}_2(t_i) &= \sum_{i=1}^{n_1}\varepsilon_i\hat{f}_2(t_i)-\sum_{i=1}^{n_1}\hat{f}_2(t_i)\left[\bar{\varepsilon}+n_1^{-1}\sum_{i=1}^{n_1}f(t_i)\right]\\ &\quad+\sum_{i=1}^{n_1}(f(t_i)-\bar{f}(t_i))\hat{f}_2(t_i)\end{aligned}$$

with $\bar{\varepsilon}=n_1^{-1}\sum_{i=1}^{n_1}\varepsilon_i$. Conditions (20)–(22) along with $\bar{\varepsilon}=O_p\left(n_1^{-1/2}\right)$ imply that the second term in the last expression is $o_p\left(\min(n_1,n_2)/\sqrt{n_1}\right)$. Similarly,

$$\begin{aligned}\sum_{i=1}^{n_2}(z_i-\bar{z}-\bar{f}(x_i))\hat{f}_1(x_i) &= \sum_{i=1}^{n_2}\eta_i\hat{f}_1(x_i)+\sum_{i=1}^{n_2}(f(x_i)-\bar{f}(x_i))\hat{f}_1(x_i)\\ &\quad+o_p\left(\min(n_1,n_2)/\sqrt{n_2}\right).\end{aligned}$$

Since $\hat{f}_2$ is independent of the ε_i a conditional expectation argument reveals that $\sum_{i=1}^{n_2}\varepsilon_i(\hat{f}_2(t_i)-f(t_i))$ has mean zero and variance $\sigma^2\mathrm{E}\sum_{i=1}^{n_1}(\hat{f}_2(t_i)-f(t_i))^2$. Applying the same approach to $\sum_{i=1}^{n_2}\eta_i\hat{f}_1(x_i)$ along with condition (25) then gives that

$\sum_{i=1}^{n_1} \varepsilon_i \hat{f}_2(t_i) = \sum_{i=1}^{n_1} \varepsilon_i f(t_i) + o_p\left(\sqrt{\min(n_1, n_2)}\right)$ and $\sum_{i=1}^{n_2} \eta_i \hat{f}_1(x_i) = \sum_{i=1}^{n_2} \eta_i f(x_i) + o_p\left(\sqrt{\min(n_1, n_2)}\right)$.

The proof is completed by using (20)–(21) and (24) to obtain $||\hat{f}_2||^2_{1n_1} = ||f||^2_1 + o_p(1)$ and $||\hat{f}_2||^2_{2n_2} = ||f||^2_2 + o_p(1)$. Combining this with condition (26) and $n_1/n_2 \to \theta \in (0, \infty)$ gives the desired result. •

There are a number of situations where one can verify condition C3. Perhaps the simplest of these is when the two designs coincide, i.e., $n_1 = n_2 \equiv n$ and $t_i = x_i$, $i = 1, \ldots, n$. In that case we will have $||f||_1 = ||f||_2 \equiv ||f||$ and find that $\sum_{i=1}^{n} \hat{f}_j^2(t_i) = n_1 ||f||^2 (1 + o_p(n^{-1/4}))$, $j = 1, 2$. The Cauchy-Schwarz and triangle inequalities along with (24) can then be used to show that

$$\sum_{i=1}^{n} (f(t_i) - \bar{f}(t_i)) \hat{f}_j(t_i) = o_p(n^{-3/4}), j = 1, 2,$$

and that

$$\sum_{i=1}^{n} (f(t_i) - \bar{f}(t_i)) \left(\hat{f}_2(t_i) - \hat{f}_1(t_i)\right) = o_p(\sqrt{n}).$$

In combination these results imply C3.

One would hope that similar results obtain when the x and t designs are only close in some asymptotic sense. An example of such a situation can be described as follows. Suppose that $n_2 \geq n_1$ with $\frac{n_1}{n_2} = 1 + o\left(n_2^{-1/2}\right)$ and that for each t_i, $i = 1, \ldots, n_1$, there exists an element of the x design, x_i', such that $t_i - x_i' = o\left(n_1^{-1/2}\right)$. The physical meaning of these conditions is that there are x design points that are close to, and can be "paired" with, all the t design points and the number of remaining x design points (that cannot be paired) is not too large. By adding uniform boundedness conditions on the derivatives of f, $\hat{f}_1$ and $\hat{f}_2$ the closeness of the t and x designs can be exploited to again verify C3.

The basic conclusion that one might infer from all this is that T should be expected to have an approximate normal distribution in practical settings where the x and t designs cover the same interval with a substantial amount of overlap or sharing of common points. In particular, we would expect Theorem 1 to be applicable to the designs for the alpha-fetal protein data discussed in Sections 1 and 4.

A final question concerns the behavior of the test under alternatives where $f_1 \neq f_2$. Suppose that

$$\sum_{i=1}^{n_1} f_1(t_i) f_2(t_i) = n_1 < f_1, f_2 >_1 + o(n_1)$$

and

$$\sum_{i=1}^{n_2} f_1(x_i) f_2(x_i) = n_2 < f_1, f_2 >_2 + o(n_2)$$

for finite constants $< f_1, f_2 >_j$, $j = 1, 2$. Then, asymptotically T is dominated by $\sqrt{n_1}C$ with

$$C = \frac{\left(< f_1, f_2 >_1 / ||f_2||_1^2\right) - \left(< f_1, f_2 >_2 / ||f_1||_2^2\right)}{\left(||f_2||_1^{-2} + \theta ||f_1||_2^{-2}\right)^{1/2}}. \tag{26}$$

This provides us with the following companion result for Theorem 1.

Theorem 2. *Assume that conditions C1–C3 hold and that* $min(n_2, n_2) \to \infty$ *with* $n_1/n_2 \to \theta \in (0, \infty)$. *Then the test obtained by rejecting* $H_0 : f_1 = f_2$ *for large values of* $|T|$ *is consistent against any alternative for which* C *in (26) does not vanish.*

The noncentrality parameter C can be viewed as a comparison of projections of f_1 on f_2 and f_2 on f_1 in different metrics dictated by the two designs. In the simple case $n_1 = n_2 = n$ with $t_i = x_i$, $i = 1, \ldots, n$, C vanishes if and only if $||f_1||^2 = ||f_2||^2$ with $||\cdot||$ now representing the common norm in (20)–(21). Thus, the test will not be able to detect function that differ but have the same norms in this case.

Acknowledgements

Eubank's research was supported by the National Science Foundation.

References

[1] Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing, Second Edition.* New York: MarcelDekker.

[2] Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* 73, 625633.

[3] Müller, H. G. (1988). Nonparametric Regression Analysis of Longitudinal Data. New York: Springer-Verlag.

[4] Serfling, R. (1976). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley.

[5] Wahba, G. (1990). *Spline Models for Observational Data.* CBMS-NSF series. Philadelphia: SIAM.

[6] Young, S. and Bowman, A. (1995). Non-parametric analysis of covariance. *Biometrics* 51, 9201931.