Testing linearity of regression models with dependent errors by kernel based methods

Stefanie Biedermann
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: stefanie.biedermann@ruhr-uni-bochum.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 2 34 3214 559
Tel.: +49 2 34 322 82 84

July 20, 2000

Abstract

In a recent paper González Manteiga and Vilar Fernández (1995) considered the problem of testing linearity of a regression under $MA(\infty)$ structure of the errors using a weighted L^2-distance between a parametric and a nonparametric fit. They established asymptotic normality of the corresponding test statistic under the hypothesis and under local alternatives. In the present paper we extend these results and establish asymptotic normality of the statistic under fixed alternatives. This result is then used to prove that the optimal (with respect to uniform maximization of power) weight function in the test of González Manteiga and Vilar Fernández (1995) is given by the Lebesgue measure independently of the design density.

The paper also discusses several extensions of tests proposed by Azzalini and Bowman (1993), Zheng (1996) and Dette (1999) to the case of non-independent errors and compares these methods with the method of González Manteiga and Vilar Fernández (1995). It is demonstrated that among the kernel based methods the approach of the latter authors is the most efficient from an asymptotic point of view.

Keywords: Test of linearity, nonparametric regression, moving average process, optimal weighted least squares, asymptotic relative efficiency

1 Introduction

Consider the common nonparametric regression model

$$Y_i = m(x_i) + \varepsilon_i \quad i = 1, \ldots, n \tag{1.1}$$

where m denotes the regression function, x_i the ith explanatory variable varying in the interval $[0,1]$ and the ε_i form a triangular array of random errors with zero mean. It is an important question in applied statistics if a linear model describes the data adequately , i.e.

$$(1.2) \qquad H_0 : m(x) = \sum_{j=1}^{p} \vartheta_j g_j(x) \quad \forall\ x \in [0,1],$$

where $g_1, \ldots, g_p$ are given linearly independent functions and $\theta = (\vartheta_1, \ldots, \vartheta_p)^T \in \Theta \subset \mathbb{R}^p$ denotes the vector of parameters. Linear models are attractive among practitioners because they describe the relation between the response and the predictor in a concise way. Much effort has been devoted to the problem of checking linearity in the recent literature, because misspecification of a linear model may lead to serious errors in the subsequent data analysis. For some recent literature we refer to Eubank and Hart (1992), Azzalini and Bowman (1993), Brodeau (1993), Stute, González Manteiga, Presedo Quindimil (1998), Dette and Munk (1998), Alcalá, Christóbal, González Manteiga (1999) or Dette (1999). While most authors consider the case of independent innovations much less progress has been made in the problem of checking linearity in the case of dependent errors.

Recently, González Manteiga and Vilar Fernández (1995) studied the problem of testing the lack of fit of a parametric regression under an $MA(\infty)$ structure of the errors by considering the weighted L^2-distance

$$(1.3) \qquad \hat{T}_n^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \{\hat{m}_n(x_i) - \sum_{j=1}^{p} \hat{\vartheta}_j g_j(x_i)\}^2 w(x_i)$$

where w denotes a (known) weight function, $\hat{\theta}_n = (\hat{\vartheta}_1, \ldots, \hat{\vartheta}_p)^T$ the weighted (with respect to w) LSE in the assumed linear regression and $\hat{m}_n$ is the nonparametric curve estimator of Gasser and Müller (1979). Note that originally a smoothed version of the parametric fit was considered in $T_n^{(1)}$ in order to avoid problems with the bias [see also Härdle and Mammen (1993)]. The differences between the two statistics are minor and will be explained at the end of Section 2. Therefore we will also denote $T_n^{(1)}$ as the statistic introduced by González Manteiga and Vilar Fernández (1995). The statistic (1.3) defines an empirical distance between a parametric and nonparametric estimate of the regression and the null hypothesis (1.2) is rejected for large values of $\hat{T}_n^{(1)}$. González Manteiga and Vilar Fernández (1995) proved asymptotic normality of $\hat{T}_n^{(1)}$ under the hypothesis of linearity and under local alternatives and as a consequence they obtained the consistency of this procedure. A bootstrap procedure of this test was examined by means of a simulation study in Vilar Fernández and González Manteiga (1996).

In this paper we are interested in the asymptotic behaviour of the statistic $\hat{T}_n^{(1)}$ (and several related tests) under fixed alternatives. These results are important for at least two reasons. On the one hand we obtain estimates of the type II error which are of particular interest if the hypothesis of linearity is not rejected. On the other hand we will demonstrate below that these results can be used for the determination of an optimal weight function w in the statistic $\hat{T}_n^{(1)}$ such that the (asymptotic) power at any fixed alternative becomes maximal.

The paper will be organized as follows. In Section 2 we introduce the necessary notation and establish asymptotic normality of $\hat{T}_n^{(1)}$ under fixed alternatives. This result is used to prove that the uniform weight function maximizes the (asymptotic) power of the corresponding test under any fixed alternative and that this property does not depend on the underlying design density. Section 3 discusses generalizations of the tests of Azzalini and Bowman (1993), Zheng (1996) and Dette (1999) to the case of errors with $MA(\infty)$ structure and compares the different methods from a local asymptotic point of view. In particular it is shown that from an asymptotic viewpoint the approach of González Manteiga and Vilar Fernández (1995) yields a most efficient procedure for testing linearity under $MA(\infty)$ structure of the errors. Finally, some of the proofs are given in Section 4.

2 The statistic $T_n^{(1)}$ and its asymptotic distribution under fixed alternatives

Throughout this paper we consider the regression model (1.1) with a fixed design given by

$$\frac{i}{n} = \int_0^{x_i} f(t)dt \tag{2.1}$$

where f is a positive density on the interval $[0,1]$ [see Sacks and Ylvisaker (1970)]. We also assume that

$$M_1^2 = \int_0^1 (m(x) - g^T(x)\theta)^2 f(x)w(x)dx$$

is minimal at a unique point $\theta_0 \in \Theta^0$ where Θ^0 denotes the interior of $\Theta \subset \mathbb{R}^p$ (note that $M_1^2 = 0$ if and only if the hypothesis of linearity is valid). In the general regression model we use the nonparametric curve estimate of Gasser and Müller (1979)

$$\hat{m}_n(x) = \frac{1}{h}\sum_{j=1}^{n} Y_j \cdot \int_{s_{j-1}}^{s_j} K(\frac{x-s}{h})ds \tag{2.2}$$

where $s_0 = 0, s_n = 1, s_{j-1} \le x_j \le s_j \quad (j = 2, \dots, n)$, h is the bandwidth and K a symmetric kernel with compact support, say $[-1,1]$. For the asymptotic analysis of the statistic $T_n^{(1)}$ in (1.3) we require the following basic assumptions [see also González Manteiga and Vilar Fernández (1995)]. The design density, the regression, the weight and kernel function are assumed to be sufficiently smooth, that is

$$g_1, \dots, g_p, w, f, m \in C^{(r)}[0,1], K \in C^{(2)}[0,1] \tag{2.3}$$

where $r \ge 2$ and $C^{(p)}[0,1]$ denotes the set of p-times continuously differentiable functions. Throughout this paper

$$\mathcal{U}_p = \text{span}\{g_1, \dots, g_p\} \tag{2.4}$$

denotes the linear subspace spanned by the linearly independent regression functions $g_1, \dots, g_p$ and obviously the null hypothesis (1.2) is valid if and only if

$$m \in \mathcal{U}_p \ .$$

The errors ε_i are generated by a stationary causal process

$$\varepsilon_i = \sum_{j=0}^{\infty} b_j e_{i-j},$$

where $\{e_i\}$ is a sequence of independent identically distributed random variables with zero mean, zero kurtosis, $\sigma_e^2 =$ Var$(e) < \infty$, such that

$$E[|e_i|^{4+2\delta}] < \infty \tag{2.5}$$

(for some $\delta > 0$) and the autocovariance function $\gamma(k) = E[\varepsilon_1 \varepsilon_{k+1}] = \sigma_e^2 \sum_{j=0}^{\infty} b_j b_{j+k}$ is absolutely summable and additionally

$$\sum_{s=-\infty}^{\infty} |s||\gamma(s)| < \infty. \tag{2.6}$$

Finally, we assume that the bandwidth in (2.2) satisfies

$$nh^{3/2} \to \infty; \quad hn^{(2+\delta)/(2\delta+2)} \to 0; \tag{2.7}$$

and that the weight function has support contained in the interval $[0, 1]$. The following theorem (part b) specifies the asymptotic distribution of the statistic $T_n^{(1)}$ introduced by González Manteiga and Vilar Fernández (1995) under fixed alternatives. Because there is a term missing in the asymptotic bias under the hypothesis of linearity given by the lastnamed authors we also restate it here (part a).

Theorem 2.1. *Assume that (2.1), (2.3) – (2.7) are satisfied and $n \to \infty$.*

(a) Under the hypothesis of linearity we have

$$n\sqrt{h}\left(T_n^{(1)} - \frac{B_1}{nh}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda_0^2) \tag{2.8}$$

where the asymptotic bias and variance are given by

$$\lambda_0^2 = 2\Big(\sum_{s=-\infty}^{\infty} \gamma(s)\Big)^2 \int_{-2}^{2} (K * K)^2(z)\, dz \int_0^1 w^2(x)\, dx, \tag{2.9}$$

$$B_1 = \sum_{s=-\infty}^{\infty} \gamma(s) \int_{-1}^{1} K^2(z)\, dz \int_0^1 w(x)\, dx, \tag{2.10}$$

respectively and $K * K$ denotes the convolution of K with itself.

(b) Under a fixed alternative $m \notin \mathcal{U}_p =$ span$\{g_1, \ldots, g_p\}$ we have

$$\sqrt{n}\left(T_n^{(1)} - M_1^2 - \frac{B_1}{nh}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda_1^2) \tag{2.11}$$

where the asymptotic bias and variance are given by

$$M_1^2 = \int_0^1 w(x)\Delta^2(x)f(x)\,dx, \tag{2.12}$$

$$\lambda_1^2 = 4\sum_{s=-\infty}^{\infty}\gamma(s)\int_0^1 w^2(x)\Delta^2(x)f(x)\,dx, \tag{2.13}$$

$\Delta = m - \mathcal{P}_{\mathcal{U}_p}m$ *and* $\mathcal{P}_{\mathcal{U}_p}$ *denotes the orthogonal projection onto* $\mathcal{U}_p$ *with respect to the inner product* $< q_1, q_2 >= \int_0^1 q_1(x)q_2(x)w(x)f(x)\,dx$.

It is important to note the different rates of convergence under the null hypothesis and alternative in Theorem 2.1. While under the hypothesis of linearity [and under local alternatives converging to the null at a rate $(n\sqrt{h})^{-1/2}$] the variance of $T_n^{(1)}$ is of order $(n^2h)^{-1}$, it is of order n^{-1} under fixed alternatives. The second part of Theorem 2.1 is particularly useful for the analysis of the type II error of the test which rejects the hypothesis whenever

$$n\sqrt{h}\{T_n^{(1)} - B_1/nh\} > u_{1-\alpha}\lambda_0 \tag{2.14}$$

[$u_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution and in practice B_1 and λ_0^2 have to be replaced by consistent estimates]. Because the acceptance of the null hypothesis yields to a data analysis adapted to the linear model this error is often considered as more important than the type I error. By Theorem 2.1 b) the probability of such an error is approximately given by

$$\begin{aligned} P(\text{“rejection”}) &= P(n\sqrt{h}\{T_n^{(1)} - \frac{B_1}{nh}\} > u_{1-\alpha}\lambda_0) \\ &= P\Big(\frac{\sqrt{n}}{\lambda_1}\{T_n^{(1)} - M_1^2 - \frac{B_1}{nh}\} > \frac{\lambda_0}{\lambda_1}\frac{u_{1-\alpha}}{\sqrt{nh}} - \frac{\sqrt{n}}{\lambda_1}M_1^2\Big) \\ &\approx \Phi(\frac{\sqrt{n}}{\lambda_1}M_1^2 - \frac{\lambda_0}{\lambda_1}\frac{u_{1-\alpha}}{\sqrt{nh}}) \approx \Phi(\frac{\sqrt{n}}{\lambda_1}M_1^2) \end{aligned} \tag{2.15}$$

where λ_0^2, M_1^2 and λ_1^2 are defined in (2.9), (2.12), (2.13), respectively. A further important application of the second part of Theorem 2.1 is given in the following corollary, which identifies an optimal weight function such that the asymptotic power becomes maximal.

Corollary 2.2. *Under the assumptions of Theorem 2.1 the asymptotic power*

$$\Phi(\frac{\sqrt{n}M_1^2}{\lambda_1})$$

of the test (2.14) is maximized for the weight function proportional to the Lebesgue density on the interval $[0, 1]$ *uniformly with respect to* $m \notin \mathcal{U}_p$.

Proof. In order to make the dependence of the asymptotic power on the weight function w more explicit we denote the quantities in (2.12) and (2.13) by $M_1^2(w)$, Δ_w and $\lambda_1^2(w)$, respectively, and obtain

$$(2.16) \qquad \left(\frac{M_1^2(w)}{\lambda_1(w)}\right)^2 = \frac{(\int_0^1 w(x)\Delta_w^2(x)f(x)dx)^2}{4\sum_{s=-\infty}^{\infty}\gamma(s)\int_0^1 w^2(x)\Delta_w^2(x)f(x)dx}$$
$$= \frac{(\int_0^1 w(x)\Delta_w(x)\Delta_\lambda(x)f(x)dx)^2}{4\sum_{s=-\infty}^{\infty}\gamma(s)\int_0^1 w^2(x)\Delta_w^2(x)f(x)dx}$$
$$\leq \frac{\int_0^1 \Delta_\lambda^2(x)f(x)dx}{4\sum_{s=-\infty}^{\infty}\gamma(s)} = \left(\frac{M_1^2(\lambda)}{\lambda_1(\lambda)}\right)^2$$

where λ denotes the Lebesgue density and the inequality follows from Cauchy's inequality applied to the factors $w(x)\Delta_w(x)\sqrt{f(x)}$ and $\sqrt{f(x)}\Delta_\lambda(x)$. Discussing equality in (2.16) shows that the optimal weight function has to be constant. Therefore the Lebesgue density (or any multiple) maximizes the asymptotic power independently of the specific alternative.

$\square$

Remark 2.3. We note that González Manteiga and Vilar Fernández (1995) worked with a modified weighted LSE $\tilde{\theta}_n$ in the definition of $T_n^{(1)}$, which minimizes

$$\sum_{i=1}^n \{\hat{m}_n(x_i) - g^T(x_i)\theta\}^2 w(x_i).$$

Theorem 2.1 and Corollary 2.2 remain valid in this case. Under the null hypothesis of linearity this method avoids a bias of order $O(h^{2r})$ [see also Härdle and Mammen (1993)]. However, under fixed alternatives this bias also appears if the smoothed version of the weighted LSE is used. Because the main interest of this paper is the asymptotic behaviour under fixed alternatives we worked with the classical weighted LSE and used a sufficiently small bandwidth [see assumption (2.7)] to obtain the order $o(1)$ for the corresponding term in the bias of the standardized statistic.

3 Related tests of linearity.

In this section we discuss the asymptotic behaviour of several related tests which were recently introduced in the context of independent observations. We begin with a test statistic proposed by Zheng (1996)

$$(3.1) \qquad T_n^{(2)} = \frac{1}{n(n-1)h}\sum_{i\neq j} K\left(\frac{x_i - x_j}{h}\right) w(x_i)w(x_j)\hat{\varepsilon}_i\hat{\varepsilon}_j$$

where $\hat{\varepsilon}_i$ are the residuals formed from a weighted least squares fit, i.e.

$$(3.2) \qquad \hat{\varepsilon}_i = Y_i - \sum_{\ell=1}^p g_\ell(x_i)\hat{\vartheta}_\ell$$

[note that in contrast to Zheng's (1996) work we introduced a weight function in the definition of $T_n^{(2)}$].

Theorem 3.1. *If the assumptions of Theorem 2.1 are satisfied we have under the null hypothesis of linearity*

$$n\sqrt{h}\left(T_n^{(2)} - B_2/nh\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_0^2)$$

where the asymptotic variance and bias are given by

$$\begin{aligned} \mu_0^2 &= 2\Big(\sum_{s=-\infty}^{\infty} \gamma(s)\Big)^2 \int_0^1 f^2(x)\, w^4(x) dx \int_{-1}^1 K^2(z)\, dz \\ B_2 &= K(0) \sum_{s=-\infty, s\neq 0}^{\infty} \gamma(s) \int_0^1 w^2(x) f(x) dx \end{aligned} \tag{3.3}$$

Under a fixed alternative we obtain

$$\sqrt{n}\Big(T_n^{(2)} - M_2^2 - \frac{\tilde{B}_2}{nh}\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_1^2),$$

where the asymptotic bias and variance are given by

$$\begin{aligned} M_2^2 &= \int \Delta^2(x) f^2(x) w^2(x) dx \\ \tilde{B}_2 &= B_2 - K(0) \int_0^1 \Delta^2(x) w^2(x) f(x) dx \end{aligned}$$

$$\mu_1^2 = 4 \sum_{s=-\infty}^{\infty} \gamma(s) \int_0^1 f(x) w^2(x) \{(\Delta f w)(x) - \mathcal{P}_{\mathcal{U}_p}(\Delta f w)(x)\}^2 dx \tag{3.4}$$

Our next example considers the asymptotic behaviour of the test of Dette (1999), who studied a difference of variance estimators as test statistic, i.e.

$$T_n^{(3)} = \hat{\sigma}^2_{\text{LSE}} - \hat{\sigma}^2_{\text{HM}}.$$

Here $\hat{\sigma}^2_{\text{LSE}}$ is the weighted least squares estimator of the variance in the linear regression model and $\hat{\sigma}^2_{\text{HM}}$ is a weighted version of the nonparametric estimator introduced by Hall and Marron (1990) which is defined by

$$\begin{aligned} \hat{\sigma}^2_{HM} &= \frac{1}{\nu} \sum_{i=1}^n \Big(Y_i - \sum\nolimits_{j=1}^n w_{ij} Y_j\Big)^2 w(x_i) \\ \nu &= n - 2\sum_{i=1}^n w_{ii} + \sum_{i=1}^n \sum_{k=1}^n w_{ik}^2 \\ w_{ij} &= \frac{K\left(\frac{x_i - x_j}{h}\right)}{\sum_{l=1}^n K\left(\frac{x_i - x_l}{h}\right)}. \end{aligned} \tag{3.5}$$

Theorem 3.2. *If the assumptions of Theorem 2.1 are satisfied we have under the null hypothesis of linearity*

$$n\sqrt{h}\left(T_n^{(3)} - \frac{B_3}{nh}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \rho_0^2)$$

where the asymptotic bias and variance are given by

$$\begin{aligned} B_3 &= \sum_{s=-\infty, s\neq 0}^{\infty} \gamma(s) \left\{2K(0) - \int_{-1}^{1} K^2(x)\, dx\right\} \int_0^1 w(x)dx \\ \rho_0^2 &= 2\Big(\sum_{s=-\infty}^{\infty} \gamma(s)\Big)^2 \int_{-2}^{2} \{2K(x) - (K * K)(x)\}^2 dx \int_0^1 w^2(x)dx \end{aligned} \tag{3.6}$$

Under a fixed alternative we obtain

$$\sqrt{n}\left(T_n^{(3)} - M_1^2 - \frac{B_3}{nh}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \rho_1^2)$$

where the asymptotic variance is given by

$$\rho_1^2 = 4 \sum_{s=-\infty}^{\infty} \gamma(s) \int_0^1 f(x) w^2(x) \Delta^2(x) dx \tag{3.7}$$

Corollary 3.3. *Under the assumptions of Theorem 2.1 the asymptotic power of the test which rejects H_0, whenever*

$$n\sqrt{h}(T_n^{(3)} - \frac{B_3}{nh}) > u_{1-\alpha}\rho_0$$

is maximized for the weight function proportional to the density of the Lebesgue measure uniformly with respect to $m \notin \mathcal{U}_p$.

A very similar statistic was considered by Azzalini and Bowman (1993)

$$T_n^{(4)} = \frac{\hat{\varepsilon}^T \hat{\varepsilon} - \hat{\varepsilon}^T M \hat{\varepsilon}}{\hat{\varepsilon}^T M \hat{\varepsilon}}$$

where $\hat{\varepsilon} = (\sqrt{w(x_1)}\hat{\varepsilon}_1, \ldots, \sqrt{w(x_n)}\hat{\varepsilon}_n)^T$ is the vector of (weighted) residuals formed from a weighted LSE fit, $M = (I_n - W)^T(I_n - W)$ and $W = (w_{ij})_{ij=1}^n$ is the matrix defined by the weights (3.5). Roughly speaking, this statistic is obtained from the statistic $T_n^{(3)}$ replacing the original observations by residuals from a parametric fit.

Theorem 3.4. *If the assumptions of Theorem 2.1 are satisfied we have under the null hypothesis of linearity*

$$n\sqrt{h}\left(T_n^{(4)} - \frac{B_4}{nh}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \rho_0^2/\gamma^2)$$

where the asymptotic bias is given by

$$B_4 = \frac{\sum_{s=-\infty}^{\infty} \gamma(s)}{\gamma} \left\{ 2K(0) - \int_{-1}^{1} K^2(z)\, dz \right\} \int_0^1 w(x)dx,$$

ρ_0^2 is defined in (3.6) and γ is a constant of proportionality given by

$$\gamma = \gamma(0) \int_0^1 w(x)p(x)dx$$

Under a fixed alternative we obtain

$$\sqrt{n}\left(T_n^{(4)} - \frac{M_1^2}{\gamma} - \frac{B_4}{nh}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \rho_1^2/\gamma^2)$$

where ρ_1^2 is defined in (3.7).

Corollary 3.5. *Under the assumptions of Theorem 2.1 the asymptotic power of the test which rejects H_0, whenever*

$$n\sqrt{h}(T_n^{(4)} - \frac{B_4}{nh}) > u_{1-\alpha}\rho_0/\gamma$$

is maximized for the weight function proportional to the density of the Lebesgue measure uniformly with respect to $m \notin \mathcal{U}_p$.

Remark 3.6. Note that we are not able to derive a result similar to Corollary 3.3 and 3.5 about the optimal weight function for the statistical test proposed by Zheng (1996), because the asymptotic variance under the alternative in Theorem 3.1 is more complicated compared to Theorem 3.2 and 3.4.

We will conclude this section with a brief comparison of the different methods based on $T_n^{(1)} - T_n^{(4)}$. Calculations similar as those used in the derivation of (2.15) show that the asymptotic power of the test based on $T_n^{(i)}$ is given by

$$p_i \approx \Phi(\sqrt{n}\frac{M_1^2}{x_i^{(1)}} - \frac{x_i^{(0)}}{x_i^{(1)}}\frac{u_{1-\alpha}}{\sqrt{nh}}) \qquad i = 1, 3, 4 \tag{3.8}$$

where (for $j = 0, 1$)

$$x_i^{(j)} = \begin{cases} \lambda_j & \text{if } i = 1 \\ \rho_j & \text{if } i = 3, 4 \end{cases}$$

and $\lambda_0^2, \lambda_1^2, \rho_0^2, \rho_1^2$ are defined in (2.9), (2.13), (3.6) and (3.7), respectively. The application of the Lebesgue measure as optimal weight function makes the dominating term in (3.8) for all methods equal to

$$\sqrt{n}\frac{M_1^2}{\lambda_1} = \sqrt{n}\Big(\frac{\int_0^1 \Delta^2(x)f(x)dx}{4\sum_{s=-\infty}^{\infty}\gamma(s)}\Big)^{\frac{1}{2}}. \tag{3.9}$$

Note that derivation of an optimal weight function for Zheng's (1996) statistic is not possible because of the complicated structure of the limiting variance μ_1^2 under fixed alternatives [see Theorem 3.1]. In this case the power of the corresponding test is given by

$$(3.10) \qquad p_2 \approx \Phi(\sqrt{n}\frac{M_2^2}{\mu_1} - \frac{\mu_0 u_{1-\alpha}}{\mu_1\sqrt{nh}} - \frac{y}{\mu_1\sqrt{nh}}) < \Phi(\sqrt{n}\frac{M_2^2}{\mu_1} - \frac{\mu_0 u_{1-\alpha}}{\mu_1\sqrt{nh}})$$

where $y = K(0)\int_0^1 \Delta^2(x)w^2(x)f(x)dx$ and μ_0^2, μ_1^2 are defined in (3.3) and (3.4), respectively. The following result shows that the dominating term in (3.10) is smaller than the term in (3.9). Consequently for any weight function a test of linearity based on $T_n^{(2)}$ is (asymptotically) less efficient than procedures based on $T_n^{(1)}, T_n^{(3)}$ and $T_n^{(4)}$ provided that the Lebesgue measure is used as the optimal weight function in these procedures.

Lemma 3.7. *Under the assumptions of Theorem 2.1 it follows*

$$4\sum_{s=-\infty}^{\infty}\gamma(s)\left(\frac{M_2^2}{\mu_1}\right)^2 = \frac{\int_0^1 \Delta^2(x)w^2(x)f^2(x)dx}{\int_0^1 f(x)w^2(x)\{(\Delta f w)(x) - \mathcal{P}_{\mathcal{U}_p}(\Delta f w)(x)\}^2 dx} \le \int_0^1 \Delta^2(x)f(x)dx$$

for every weight function w, such that the integrals in this inequality exist.

It follows from (3.8) that for the remaining procedures the power is maximized by minimizing the asymptotic variance under the null hypothesis. Our final result shows that $x_i^{(0)}$ becomes minimal for the test of González Manteiga and Vilar Fernández (1995) and consequently this procedure is asymptotically most powerful among the kernel based methods discussed in this paper.

Lemma 3.8. *For any square integrable density K we have*

$$\int (K*K)^2(x)dx \le \int K^2(x)dx \le \int (2K - K*K)^2(x)dx$$

or equivalently

$$\lambda_0^2 \le \mu_0^2 \le \rho_0^2.$$

4 Proofs

Because all proofs are similar, we restrict ourselves exemplarily to a proof of Theorem 3.1, for which the asymptotics is slightly more complicated. For the sake of a transparent notation we only consider the case $w = \lambda$ (here λ denotes the density of the Lebesgue measure on the interval $[0,1]$). Without loss of generality we assume orthonormality of the regression functions $g_1, \ldots, g_p$ with respect to the density f. Introducing the notation $g(x) = (g_1(x), \ldots, g_p(x))^T$ the residuals in (3.2) can be written as

$$(4.1) \qquad \hat{\varepsilon}_i = \varepsilon_i + \Delta(x_i) - g^T(x_i)\{\hat{\theta}_n - \theta_0\}$$

where θ_0 is the unique minimizer of $\int_0^1 (m(x) - g^T(x)\theta)^2 f(x)dx$. Our first Lemma specifies the asymptotic behaviour of $\hat{\theta}_n - \theta_0$ under the null hypothesis and fixed alternatives.

Lemma A.1. *Under the assumptions of Theorem 2.1, $w \equiv 1$ and orthonormal regression functions we have for any $m \in C^{(r)}[0,1]$*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g(x_i)\varepsilon_i + o_p(1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sum_{s=-\infty}^{\infty} \gamma(s) I_p)$$

where I_p denotes the $p \times p$ identity matrix.

Proof. Recalling the notation $\Delta(x_i) = (m - \mathcal{P}_{\mathcal{U}_p} m)(x_i) = m(x_i) - \theta_0^T g(x_i)$ we obtain $Y_i = \Delta(x_i) + g^T(x_i)\theta_0 + \varepsilon_i$ and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n} B_n^{-1} \{ \frac{1}{n} \sum_{i=1}^n g(x_i)\Delta(x_i) + \frac{1}{n} \sum_{i=1}^n g(x_i)\varepsilon_i \}$$

where

$$B_n = \frac{1}{n} \sum_{i=1}^n g(x_i) g^T(x_i) = I_p + O(\frac{1}{n}) \tag{4.2}$$

is the design matrix of the LSE

$$\hat{\theta}_n = B_n^{-1} \frac{1}{n} \sum_{i=1}^n g(x_i) Y_i.$$

For the first term in the sum we note that

$$\frac{1}{n} \sum_{i=1}^n g(x_i)\Delta(x_i) = \int_0^1 g(x)\Delta(x) f(x)dx + O(\frac{1}{n}) = O(\frac{1}{n})$$

where the last estimate follows from the fact that $\theta_0 \in \Theta^0$ is the unique minimizer of

$$\int_0^1 (m(x) - \theta^T g(x))^2 f(x)dx.$$

Observing (4.2) this establishes the first equality of Lemma A.1. The asymptotic normality now follows exactly by the same arguments as given by González Manteiga and Vilar Fernández (1995) in the proof of their Theorem 1.

□

Throughout the proof of Theorem 3.1 we make use of the decomposition

$$T_n^{(2)} = V_{1,n} - 2\{V_{2,n}^{(1)} - V_{2,n}^{(2)}\} + \{V_{3,n}^{(1)} - 2V_{3,n}^{(2)} + V_{3,n}^{(3)}\} \tag{4.3}$$

which is obtained from (4.1) and the notation

$$
(4.4) \qquad
\begin{aligned}
V_{1,n} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{h} K\Big(\frac{x_i - x_j}{h}\Big) \varepsilon_i \varepsilon_j \\
V_{2,n}^{(1)} &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{h} K\Big(\frac{x_i - x_j}{h}\Big) \varepsilon_i g^T(x_j) \{\hat{\theta}_n - \theta_0\} \\
V_{2,n}^{(2)} &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{h} K\Big(\frac{x_i - x_j}{h}\Big) \varepsilon_i \Delta(x_j) \\
V_{3,n}^{(1)} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{h} K\Big(\frac{x_i - x_j}{h}\Big) g^T(x_i) \{\hat{\theta}_n - \theta_0\} g^T(x_j) \{\hat{\theta}_n - \theta_0\} \\
V_{3,n}^{(2)} &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{h} K\Big(\frac{x_i - x_j}{h}\Big) \Delta(x_i) g^T(x_j) \{\hat{\theta}_n - \theta_0\} \\
V_{3,n}^{(3)} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{h} K\Big(\frac{x_i - x_j}{h}\Big) \Delta(x_i) \Delta(x_j)
\end{aligned}
$$

Proof of Part a) of Theorem 3.1. Under the hypothesis of linearity $\Delta \equiv 0$ we have $V_{2,n}^{(2)} = V_{3,n}^{(2)} = V_{3,n}^{(3)} = 0$. The remaining terms are treated essentially in the same way as in González Manteiga and Vilar Fernández (1995) and therefore we only state the main steps here. We have

$$
(4.5) \qquad V_{2,n}^{(1)} = o_p(\frac{1}{n\sqrt{h}}); V_{3,n}^{(1)} = o_p(\frac{1}{n\sqrt{h}})
$$

and for the asymptotic bias and variance of $V_{1,n}$

$$
(4.6) \qquad E[V_{1,n}] = \frac{K(0)}{nh} \sum_{s=-\infty, s\neq 0}^{\infty} \gamma(s) + o\left(\frac{1}{n\sqrt{h}}\right)
$$

$$
(4.7) \qquad Var(V_{1,n}) = \frac{2}{n^2 h} \Big(\sum_{r=-\infty}^{\infty} \gamma(r) \Big)^2 \int_0^1 f^2(x)\, dx \int_{-1}^{1} K^2(z)\, dz + o(\frac{1}{n^2 h})
$$

Note that the derivation of (4.6) requires a finite first moment of the autocovariance function as assumed in (2.6) and the condition $nh^{3/2} \to \infty$ specified in (2.7). These assumptions are necessary but not stated explicitly in González Manteiga and Vilar Fernández (1995). Finally, the asymptotic normality of $n\sqrt{h}(V_{1,n} - E[V_{1,n}])$ follows from a central limit theorem for triangular arrays with $m(n)$ dependent main part [see Niewenhuis (1992)].

Proof of part b) of Theorem 3.1. The statements given in (4.5) of the previous paragraph show

$$
(4.8) \qquad T_n^{(2)} - E[V_{1,n}] = 2\{V_{2,n}^{(2)} - V_{3,n}^{(2)}\} + V_{3,n}^{(3)} + o_p(\frac{1}{\sqrt{n}})
$$

where $V_{3,n}^{(3)}$ is nonrandom and asymptotically equivalent to

$$V_{3,n}^{(3)} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{h} K\left(\frac{x_i - x_j}{h}\right) \Delta(x_i)\Delta(x_j) - \frac{1}{n(n-1)} \sum_{i=1}^{n} \frac{1}{h} K(0)\Delta^2(x_i)$$
$$= \int_0^1 \Delta^2(x) f^2(x)\, dx - \frac{K(0)}{nh} \int_0^1 \Delta^2(x) f(x)\, dx + o\left(\frac{1}{\sqrt{n}}\right).$$

Combining this estimate with (4.6) and (4.8) yields for the statistic of interest

$$T_n^{(2)} - M_2^2 - \frac{\tilde{B}_2}{nh} = 2\{V_{2,n}^{(2)} - V_{3,n}^{(2)}\} + o_p(\frac{1}{\sqrt{n}}) \tag{4.9}$$

For the variance of the dominating term on the right hand side of (4.9) we obtain

$$b_n^2 = Var(V_{3,n}^{(2)} - V_{2,n}^{(2)}) = Var\Big(\frac{1}{n^2(n-1)h} \sum_{i,j,k=1}^{n} K\Big(\frac{x_i - x_j}{h}\Big)\Delta(x_i) g^T(x_j) g(x_k)\varepsilon_k$$
$$-\frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1}^{n} K(\frac{x_i - x_j}{h})\Delta(x_j)\varepsilon_i\Big) + o(\frac{1}{n})$$

where we used the representation of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ of Lemma A.1. Changing the order of summation yields

$$b_n^2 = Var\Big(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \Big\{\frac{1}{n(n-1)h} \sum_{j,k=1}^{n} K(\frac{x_k - x_j}{h})\Delta(x_k) g^T(x_j) g(x_i) \tag{4.10}$$
$$- \frac{1}{(n-1)h} \sum_{j=1}^{n} K(\frac{x_i - x_j}{h})\Delta(x_j)\Big\}\Big) + o(\frac{1}{n})$$
$$= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{r=1}^{n} \gamma(r-i) \Big\{\frac{1}{n(n-1)h} \sum_{j=1}^{n} \sum_{k=1}^{n} K(\frac{x_k - x_j}{h})\Delta(x_k) g^T(x_j) g(x_i)$$
$$- \frac{1}{(n-1)h} \sum_{j=1}^{n} K(\frac{x_i - x_j}{h})\Delta(x_j)\Big\}$$
$$\times \Big\{\frac{1}{n(n-1)h} \sum_{j=1}^{n} \sum_{k=1}^{n} K(\frac{x_k - x_j}{h})\Delta(x_k) g^T(x_j) g(x_r)$$
$$- \frac{1}{(n-1)h} \sum_{j=1}^{n} K(\frac{x_r - x_j}{h})\Delta(x_j)\Big\} + o(\frac{1}{n})$$
$$= \frac{1}{n} \sum_{s=-\infty}^{\infty} \gamma(s) \int_0^1 f(x) \Big\{\frac{1}{h} \int_0^1 \int_0^1 K(\frac{z-y}{h})\Delta(z) g^T(y) g(x) f(y) f(z)\, dz\, dy$$
$$- \frac{1}{h} \int_0^1 K(\frac{x-y}{h})\Delta(y) f(y)\, dy\Big\}^2 dx + o(\frac{1}{n})$$
$$= \frac{1}{n} \sum_{s=-\infty}^{\infty} \gamma(s) \int_0^1 f(x) \Big\{\sum_{l=1}^{p} g_l(x) \int_0^1 (\Delta f)(y) g_l(y) f(y)\, dy - (\Delta f)(x)\Big\}^2 dx + o(\frac{1}{n})$$

Summarizing these calculations gives

$$\text{(4.11)}\quad \lim_{n\to\infty} 4nV(V_{2,n}^{(2)} - V_{3,n}^{(2)}) = 4 \sum_{s=-\infty}^{\infty} \gamma(s) \int_0^1 f(x)\{(\Delta f)(x) - \mathcal{P}_{\mathcal{U}_p}(\Delta f)(x)\}^2 dx = \mu_1^2.$$

In order to establish asymptotic normality we apply Theorem 2.3 of Nieuwenhuis (1992) to the statistic

$$V_{3,n} - V_{2,n} = \sum_{i=1}^{n} X_{i,n} + o(\frac{1}{\sqrt{n}}) \tag{4.12}$$

where [note that we have applied Lemma A.1 in the definition of the $X_{i,n}$].

$$\begin{aligned} X_{i,n} &= c_{i,n}\varepsilon_i \\ X_{i,n,m(n)} &= c_{i,n} \sum_{r=0}^{m(n)} b_r e_{i-r} \\ \bar{X}_{i,n,m(n)} &= c_{i,n} \sum_{r=m(n)+1}^{\infty} b_r e_{i-r} \end{aligned} \tag{4.13}$$

and the constants $c_{i,n}$ are defined by

$$\begin{aligned} c_{i,n} &= \frac{1}{n}\Big\{\frac{1}{n(n-1)h}\sum_{j=1}^{n}\sum_{k=1}^{n} K(\frac{x_k - x_j}{h})\Delta(x_k) g^T(x_j) g(x_i) \\ &- \frac{1}{(n-1)h}\sum_{j=1}^{n} K(\frac{x_i - x_j}{h})\Delta(x_j)\Big\}. \end{aligned} \tag{4.14}$$

We are now establishing conditions (C1), (C2) and (C2*) in Theorem 2.3 of Nieuwenhuis (1992) noting that $b_n^2 = \text{Var}(\sum_{i=1}^n X_{i,n}) = \mu_1^2/(4n) + o(1/n)$ by (4.10). We start with the condition (C2) and obtain

$$\begin{aligned} \frac{1}{j-i} Var\Big(\sum_{k=i+1}^{j} \frac{X_{k,n}}{b_n}\Big) &= \frac{O(n)}{j-i} \sum_{k=i+1}^{j}\sum_{l=i+1}^{j} \gamma(l-k) c_{k,n} c_{l,n} \\ &= O(\frac{1}{n}) \frac{1}{j-i} \sum_{k=i+1}^{j}\sum_{l=i+1}^{j} |\gamma(l-k)| \\ &= O(\frac{1}{n}) \sum_{s<|j-i|} |\gamma(s)|(1 - \frac{|s|}{j-i}) \\ &= O(\frac{1}{n}) \sum_{s=-\infty}^{\infty} |\gamma(s)| \;=\; O(\frac{1}{n}) \end{aligned}$$

where we used $|c_{k,n}| = O(\frac{1}{n})$ (uniformly with respect to $k \in \{1, \ldots, n\}$) in the second estimate. This establishes condition (C2) in Nieuwenhuis (1992), i.e.

$$\max_{i<j\le n} \frac{1}{j-i} Var\Big(\sum_{k=i+1}^{j} \frac{X_{k,n}}{b_n}\Big) = O(\frac{1}{n})$$

because all estimates are independent of i, j. A similar argument yields

$$\begin{aligned}\frac{1}{j-i} Var\Big(\sum_{k=i+1}^{j} \frac{\bar{X}_{k,n,m(n)}}{b_n}\Big) &= O(\frac{1}{n}) \frac{1}{j-i} \sum_{k=i+1}^{j} \sum_{l=i+1}^{j} \sum_{r,s>m(n)} |Cov\,(b_r e_{k-r}, b_s e_{l-s})| \\ &= O(\frac{1}{n}) \sum_{t\in\mathbb{Z}} \sum_{r,s>m(n)} |Cov\,(b_r e_{1-r}, b_s e_{1+t-s})| \;+\; o(\frac{1}{n}) \\ &= O(\frac{1}{n}) \sum_{t\in\mathbb{Z}} \sum_{r,s>m(n)} |b_r||b_s|\,|Cov(e_{1-r}, e_{1+t-s})| \;+\; o(\frac{1}{n}) \\ &= O(\frac{1}{n}) \sigma_e^2 \Big(\sum_{r>m(n)} |b_r|\Big)^2 \;+\; o(\frac{1}{n}) \;=\; o(\frac{1}{n})\end{aligned}$$

which gives the corresponding estimate (C2*), that is

$$\max_{i<j\le n} \frac{1}{j-i} Var\Big(\sum_{k=i+1}^{j} \frac{\bar{X}_{k,n,m(n)}}{b_n}\Big) = o(\frac{1}{n})$$

We finally have to prove the $(2+\delta)$-moment condition (C1) for the arrays $(X_{i,n}/b_n)$ and $(\bar{X}_{i,n,m(n)}/b_n)$. To this end we note that

$$\begin{aligned}(4.15)\quad E\left|\frac{X_{i,n,m(n)}}{b_n}\right|^{2+\delta} &= O(n^{1+\delta/2}) E\Big|\sum_{r=0}^{m(n)} c_{i,n} b_r e_{i-r}\Big|^{2+\delta} \;=\; O\left(\frac{1}{n^{1+\delta/2}}\right) E\Big|\sum_{r=0}^{m(n)} b_r e_{i-r}\Big|^{2+\delta} \\ &= O\Big(\frac{1}{n^{1+\delta/2}}\Big)\Big\{\sum_{r=0}^{m(n)} |b_r| (E|e_1|^{2+\delta})^{1/(2+\delta)}\Big\}^{2+\delta} \\ &= O\left(\frac{1}{n^{1+\delta/2}}\right) E|e_1|^{2+\delta} \Big(\sum_{r=0}^{m(n)} |b_r|\Big)^{2+\delta} \;=\; O\left(\frac{1}{n^{1+\delta/2}}\right)\end{aligned}$$

where we used Minkowski's inequality in the second step. Because this estimate also holds if $m(n)$ is replaced by ∞, we have proved condition (C1) for the array $(X_{i,n}/b_n)$. The corresponding condition for the array $(\bar{X}_{i,n,m(n)}/b_n)$ is now obtained from (4.15) and Minkowski's inequality, which gives

$$E|\bar{X}_{i,n,m(n)}|^{2+\delta} \le \Big\{(E|X_{i,n}|^{2+\delta})^{1/(2+\delta)} + (E|X_{i,n,m(n)}|^{2+\delta})^{1/(2+\delta)}\Big\}^{2+\delta}$$

The asymptotic normality under a fixed alternative now follows from (4.9), (4.11), (4.12) and Theorem 2.3 in Nieuwenhuis (1992).

□

Proof of Lemma 3.7. Assume that the regression functions are orthonormal with respect to the measure $w(x)f(x)dx$, then

$$\mathcal{P}_{\mathcal{U}_p}(\Delta w f) \;=\; \sum_{\ell=1}^{p} g_\ell \cdot \int_0^1 g_\ell(x) w^2(x) f^2(x) \Delta(x) dx$$

$$\Delta wf \;=\; wfm - wf\sum_{\ell=1}^{p} g_\ell \cdot \int_0^1 g_\ell(x)m(x)w(x)f(x)dx$$

and a straightforward calculation shows that

$$\int_0^1 (\Delta fw)(x)\mathcal{P}_{\mathcal{U}_p}(\Delta fw)(x)dx = 0.$$

With this identity we obtain by Cauchy's inequality

$$\begin{aligned} 4\sum_{s=-\infty}^{\infty}\gamma(s)\left(\frac{M_2^2}{\mu_1}\right)^2 &= \frac{(\int_0^1(\Delta wf)(x)\{(\Delta fw)(x)-\mathcal{P}_{\mathcal{U}_p}(\Delta fw)(x)\}dx)^2}{\int_0^1 f(x)w^2(x)\{(\Delta wf)(x)-\mathcal{P}_{\mathcal{U}_p}(\Delta wf)(x)\}^2dx} \\ &\le \int_0^1 \Delta^2(x)f(x)dx \end{aligned}$$

which proves the assertion of the Lemma.

□

Proof of Lemma 3.8. Using Jensen's inequality and Fubini's theorem we obtain

$$\begin{aligned} \int (K*K)^2(x)dx &= \int\{\int K(x+z)K(z)dz\}^2dx \\ &\le \int\int K^2(x+z)K(z)dzdx = \int K^2(x)dx \end{aligned}$$

which yields the first inequality. For the second part we note that

$$\begin{aligned} [\int(2K-K*K)^2(x)dx]^{1/2} &\ge [\int(2K)^2(x)dx]^{1/2} - [\int(K*K)^2(x)dx]^{1/2} \\ &\ge [\int K^2(x)dx]^{1/2} \end{aligned}$$

where we used the first inequality in the last step. This proves Lemma 3.8.

□

Acknowledgements. The authors are grateful to I. Gottschlich who typed most parts of this paper with considerable technical expertise and to L. Mattner, who provided the proof of Lemma 3.8. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, Reduction fo complexity in multivariate data structures) is gratefully acknowledged.

References

J.T. Alcalá, J. A. Christóbal, W. González Manteiga (1999). Goodness-of-fit test for linear models based on local polynomials. Statist. & Probab. Letters 42, 39-46.

A. Azzalini, A. Bowman (1993). On the use of nonparametric regression for checking linear relationships. J. Roy. Statist. Soc., Ser. B., 55, 549-559.

F. Brodeau (1993). Tests for the choice of approximative models in nonlinear regression when the variance is unknown. Statistics, 24, 95-106.

H. Dette, A. Munk (1998). Validation of linear regression models. Ann. Statist. 26, 778-800.

H. Dette (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. Ann. Statist., to appear.

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

T. Gasser, H.-G. Müller (1979). Kernel estimation of regression functions. In: Smoothing techniques for curve Estimation. Lecture Notes in Mathematics 757. Springer-Verlag, New York.

W. Härdle, E. Mammen (1993). Comparing nonparametric versus parametric regression fits. Ann. Statist. 21, 1926-1947.

W. González Manteiga, J.M. Vilar Fernández (1995). Testing linear regression models using non-parametric regression estimators when the errors are correlated. Comput. Statist. & Data Analysis 20, 521-541.

P. Hall, J.S. Marron (1990). On variance estimation in nonparametric regression. Biometrika 77, 415-419.

G. Nieuwenhuis (1992). Central limit theorems for sequences with $m(n)$ dependent main part. J. Statist. Planning Inference 32, 229-241.

J. Sacks, D. Ylvisaker (1970). Designs for regression problems for correlated errors. Ann. Math. Statist. 41, 2057-2074.

W. Stute, W. González Manteiga, M. Presedo Quindimil (1998). Bootstrap approximation in model checks for regression. J. Amer. Statist. Assoc. 93, 141-149.

J.M. Vilar Fernández, W. González Manteiga (1996). Bootstrap test of goodness of fit of a linear model when errors are correlated. Commun. Statist.-Theory Math. 25, 2925-2953.

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